International Journal of Forecasting 35 (2019) 1692–1707

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting economic activity with mixed frequency BVARs

Scott A. Brave [a], R. Andrew Butters [b,*], Alejandro Justiniano [a]

[a] Economic Research, Federal Reserve Bank of Chicago, 230 S. La Salle St., Chicago, IL, 60604, USA
[b] Business Economics and Public Policy, Kelley School of Business, Indiana University, 1309 E. Tenth Street, Bloomington, IN, 47405, USA

ARTICLE INFO

Keywords:
Mixed frequency
Bayesian VAR
Real-time data
Nowcasting
Forecasting
Economic activity

ABSTRACT

Mixed frequency Bayesian vector autoregressions (MF-BVARs) allow forecasters to incorporate large numbers of time series that are observed at different intervals into forecasts of economic activity. This paper benchmarks the performances of MF-BVARs for forecasting U.S. real gross domestic product growth against surveys of professional forecasters and documents the influences of certain specification choices. We find that a medium–large MF-BVAR provides an attractive alternative to surveys at the medium-term forecast horizons that are of interest to central bankers and private sector analysts. Furthermore, we demonstrate that certain specification choices influence its performance strongly, such as model size, prior selection mechanisms, and modeling in levels versus growth rates.
© 2019 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Central bankers and private sector analysts share a demand for timely forecasts of economic activity. To both encapsulate and reflect the most recent events, these forecasts must blend information that is collected from a wide array of sources and observed at different intervals. Researchers have endeavored to meet these demands by spending a considerable amount of effort on the development of methods that are able to handle both (i) data observed at different frequencies, and (ii) large numbers 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 methodology that is distinctly well-suited for this purpose.

In a pivotal contribution to the MF-BVAR methodology, Schorfheide and Song (2015) showed how to conduct exact posterior Bayesian inference using a data-driven approach to inform the degree of shrinkage. They then demonstrated that, in real-time, an 11-variable MF-BVAR with monthly and quarterly variables achieves 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] However, their work, which focused on a small system of variables, left unanswered the question of how well the MF-BVAR approach may be expected to perform in the data-rich setting that is typically faced by central bankers and private sector analysts, and how robust this performance might be around business cycle turning points.[2] The current paper closes this gap by answering two questions: (1) *How well does a medium–large MF-BVAR do in forecasting economic activity?* and (2) *How stable is this relative performance over time and across different specification choices?*

* Corresponding author.
E-mail address: rabutter@indiana.edu (R.A. Butters).

1 The use of Bayesian methods for forecasting economic activity has a celebrated tradition dating back to the work of Doan, Litterman, and Sims (1984) and Litterman (1986), who were the first to document the fact that Bayesian shrinkage could improve upon the forecast accuracy of a small vector autoregression.

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

https://doi.org/10.1016/j.ijforecast.2019.02.010
0169-2070/© 2019 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

To this end, we estimate a 37-variable MF-BVAR in order 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 because 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 that are 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 that are typically included in the information sets 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 to 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 rely primarily 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 irregularly like other common benchmarks (e.g. the Survey of Professional Forecasters or the Greenbook). Convenience issues aside, however, BCC forecasts also tend to outperform other modeling approaches for forecasting real GDP growth.[4] Encouragingly, we find that our MF-BVAR forecasts are on a par with or dominate the BCC predictions. At the nearest horizons (nowcasts and forecasts one to two quarters out), the differences between the two are not statistically distinguishable. However, three and four quarters out, the gains in predictive accuracy from using our model can be as large as 10%–15% and are statistically significant. We ensure the robustness of these findings by also comparing the performance of our model against that of 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 20%. Our findings highlight how a richer information set than those analyzed previously 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 using our model not only during the sharp contraction of the Great Recession, but also in the ensuing sluggish recovery. We interpret this result as being complementary to similar work (Carriero, Galvao, & Kapetanios, 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 relative to these alternatives, we deconstruct its performance gains by examining how key specification choices affect its performance. For this part of the analysis, we look at the roles of model size, the 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 generally improves its forecasting performance, although the magnitudes 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 accrue mostly 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, Lenza, & Primiceri, 2015). We find that using the MDD to select model hyperparameters, including the prior on the residual variances, delivers performance gains of roughly 20% at the nowcast horizon relative to the more traditional/default hyperparameter values (Carriero, Clark, & Marcellino, 2015, Giannone et al., 2015), with gains of roughly 10% 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 of roughly 10%–15% 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.

[3] Information on the CFNAI can be obtained from the Federal Reserve Bank of Chicago at https://www.chicagofed.org/publications/cfnai/index. Brave and Butters (2010, 2014) discuss the use of real-time data from the index for 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, our analysis includes several additional data series that are not found in FRED-MD.

[4] Chauvet and Potter (2013) find that the BCC forecasts outperform a wide range of models (e.g. univariate methods, dynamic factor methods and dynamic stochastic general equilibrium (DSGE) models) 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).

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 for incorporating a wide array of information on economic activity that is observed at different frequencies into the regularly produced forecasts at central banks and other institutions that are 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 with professional forecasters over very short horizons while still providing superior performances at longer horizons. Combining this result with the established dominance of the MF-BVAR relative to traditional quarterly VARs in the near term (Schorfheide & Song, 2015)[5] suggests that the MF-BVAR might 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 from the mixed frequency nature of the information in the near term. Furthermore, by being the first to document the way in which the performances of MF-BVARs respond to key specification choices, our analysis offers further insights for practitioners who are seeking to apply this methodology.

In terms of related literature, various similar methods have been proposed for dealing with data observed at different intervals, including dynamic factor models (Aruoba, Diebold, & Scotti, 2009; Mariano & Murasawa, 2003, 2010) and the mixed data sampling (MIDAS) regressions first proposed by Ghysels, Santa-Clara, and Valkanov (2004) and extended to the VAR case by Foroni, Ghysels, and Marcellino (2013), Ghysels (2016), and McCracken, Owyang, and Sekhposyan (2015). We contribute to this body of 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 & Potter, 2013). Similarly, our work also contributes to the growing body of literature investigating the effects of specification choices on the forecasting performances of traditional, single frequency BVARs. For example, past investigations have explored the effect of the model size (Bańbura, Giannone, & Reichlin, 2010; Koop, 2013), the choice of prior (Carriero et al., 2015; Giannone et al., 2015; Koop, 2013), and specification choices more generally (Carriero et al., 2015). Interestingly, our work for the mixed frequency case has found a somewhat larger sensitivity to some specification choices than 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, Section 3 outlines the dataset that we use for our evaluation and makes explicit the associated timing of the real-time information flow and the set of hyperparameters used. 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. Section 6 concludes and offers suggestions for future work.

2. Methodology

The MF-BVAR accommodates mixed frequency data by casting its system of equations into a state space framework that is modeled at the level of the highest frequency variable. Exact posterior inference is then conducted within this framework by using a Gibbs sampling procedure to handle the latent values of lower-frequency variables. Bayesian shrinkage operates on both the individual dynamics of each variable and the overall co-movements among them. We keep the exposition in this section succinct by presenting 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 the online appendix 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 the monthly (high) and quarterly (low) frequencies. We accommodate the mixed frequency nature of y_t by adopting the convention of modeling the underlying base (monthly) frequency of each time series, stacked in the partially latent vector x_t. We match the realized values of those series in y_t that are observed at a quarterly frequency by aggregating the corresponding elements of x_t using the appropriate accumulator (Harvey, 1989) depending on each time series' temporal aggregation properties (see the online appendix). 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 by $\{x_t^{latent}\}_{t=1}^{T}$, which stacks the monthly values of the quarterly time series as well as any missing observations for the monthly time series.

[5] The online appendix reports the performance of the medium–large MF-BVAR relative to that of a quarterly BVAR(1) alternative of the same sized model. The pattern that we find is almost identical to that reported by Schorfheide and Song (2015) for their 11-variable model. The MF-BVAR registers large gains over the quarterly BVAR at the nowcast and one-quarter-ahead horizons, while the two models perform very similarly at further horizons.

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

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

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

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

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 a 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

$$s'_t = \left[x'_t,\ \ldots,\ x'_{t-p},\ \zeta'_t\right],$$

which includes both lags of the time series at the monthly frequency and ζ_t, a vector of accumulators. Each accumulator maintains the appropriate combination of current and past x_ts in order 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. Note that even if the VAR parameters are assumed to be time-invariant (as is the case in our empirical application), the state space system matrices are indexed by t due to the deterministic time variation that is required for calculating the accumulators, ζ_t. The remaining entries of the transition matrix, T_t, correspond to either ones and zeros in order to preserve the lag structure, or some scaled replication of the coefficients in order 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 that has been augmented to accommodate the accumulators in the state.[8]

The matrix Z_t consists solely of selection rows that are 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, it will have the full n selection rows in those months in which the quarterly series are observed. For the remaining periods, in which only the monthly frequency series are observed, we include a subset of these selection rows. Furthermore, not all of the monthly series will be available towards the end of the sample in our empirical application. 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 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 the online appendix for more details).

We address the curse of dimensionality by using the following informative prior distributions on (Φ, Σ), belonging to the normal-inverse Wishart family, that preserve conjugacy:

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

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

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 $n + 4$. We operationalize the prior by using the data augmentation, or "shrinkage through dummy observations", approach that was introduced by Sims and Zha (1998) and is often used in the single-frequency BVAR context (e.g. Bańbura et al., 2010, see also the online appendix). We select the set of hyperparameters that maximizes the marginal data density (MDD), or, explicitly,

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

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 presample set of observations for initializing 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 us to use optimization routines to find $\Lambda^\star$ quickly and efficiently (Giannone et al., 2015).

With mixed frequency data, though, the MDD must be approximated to account for the unobserved higher-frequency paths of lower-frequency time series, and the

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

[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 is explained in the online appendix.

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

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 & Song, 2015). Unfortunately, computational considerations then prevent us from using optimization techniques and force the use of sparse grids over which to evaluate each possible combination of the hyperparameters (Carriero, Kapetanios, & Marcellino, 2012b). Adding to the computational burden, and contrary to the work of 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 the work of Giannone et al. (2015) and Schorfheide and Song (2015). In the first step, we gauge the general patterns of the MDD by means 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 in order 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 that are deemed to be 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 a discussion of this approach can be found in the online appendix.

3. Real-time data and forecasts

Here, we describe the salient features of our baseline MF-BVAR and the real-time data that are used to evaluate its forecast performance. We begin in Section 3.1 by discussing the choice of variables that define our baseline model and the historical vintage data that we use to assess its out-of-sample forecast performance. Next, Section 3.2 details the timing of our forecasts and provides a rationale as to why surveys of professional forecasters provide a credible benchmark against which to judge the MF-BVAR's out-of-sample forecast performance. Then, Section 3.3 describes our out-of-sample forecasting exercise and a formal evaluation of point forecast accuracy. Finally, Section 3.4 discusses the results of our data-driven procedure for selecting hyperparameters, which then are used ultimately 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 that are available in real-time and have been used previously to forecast GDP. Here, we made use of the literature on forecasting with many predictors (e.g. Bańbura et al., 2010; Koop, 2013) in order to obtain an initial set of key indicators, such as industrial production, payroll employment, etc. Our second criterion then involved limiting our search to indicators that correspond to headline numbers within subcomponents of GDP. This was done so as to avoid having an over-representation of series that have a diverse set of disaggregations available. For example, in the cases of industrial production and payroll employment, we included the aggregate series but not the industry breakdown. For our two final criteria, we included only those variables that had long histories and for which real-time vintages existed going back at least as far as the starting point of our most comprehensive data source.

Applying these selection criteria was straightforward for most monthly time series, as full real-time vintage data were already available either through Haver Analytics or via publicly available sources such as the St. Louis and Philadelphia Federal Reserve Banks' ALFRED database and Real-time Dataset for Macroeconomists, respectively. However, there were several series (e.g. real nonresidential and residential private construction spending and real public construction spending) for which the only way to obtain a full history of real-time vintage data was to make use of 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 sizes, including specifications that encompass large numbers 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 the performances of forecasts from MF-BVARs 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 seven 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 (with mnemonics for reporting purposes), categorized under major subcomponents of GDP: personal consumption expenditures, business fixed investment, residential investment, changes in the valuation of inventories, government spending and net exports. Section 5.1 uses these groupings to organize comparisons across models of different sizes and to assess the informational contributions 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 the online appendix.

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

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)***[a]		
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)***[a]		
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, the variables are transformed to logs unless they are already expressed as percentage rates, in which case they are divided by 100 to retain a comparable scale. In contrast, for specifications in growth rates, the transformation used is 100 times the log difference or the difference in percentage rates.
[a] Schorfheide and Song (2015) (SS (2015)) use Fixed Investment in their 11-variable MF-BVAR, which is a combination of business fixed investment and residential investment. The four other indicators used by Schorfheide and Song (2015) that are not listed here include the Consumer Price Index, the Federal Funds Rate, the 10-year Treasury Bond Yield, and the S&P 500 Index.

3.2. Surveys and forecast origins

We compare the performance of our MF-BVAR against those of 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 either a quarterly schedule (e.g. the Survey of Professional Forecasters), or irregularly 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 tended historically to outperform other modeling approaches in forecasting real GDP growth (Chauvet & Potter, 2013), and compare favorably to other professional forecasters (Reifschneider & Tulip, 2007) on this dimension.

We keep track of the forecast timing by labeling the forecast origins as R1, R2, or R3, according to the last available GDP release (i.e., the first, second, or third release, as labeled by the Bureau of Economic Analysis) at the time when the forecast was made. This convention helps us to keep track of the information set that was 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 the first quarter's GDP is published at the end of April, the second release is published in May, and so on. For a further description of our forecast timing, see the online appendix.

We clarify the labeling of forecast origins by using the second quarter's GDP as an example. The first release of the previous quarter's GDP (Q1) is published at the end of April, 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, along with projections for longer 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 starting 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 origins and nowcasts for other quarters.

The staggered release of the monthly variables in our dataset (see the online appendix) and the uncertain timing of the submission of survey responses create ambiguity as to how best to 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 and use whatever data were available at the time when the CFNAI was constructed, which tends to be closer to the middle of the month. We then assessed the sensitivity of our results to this timing assumption by also comparing 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.), which falls more closely in line with the production schedule of the CFNAI in these months, thus 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 to the first quarter of 2016. The beginning of the sample is imposed by the availability of the real-time vintages from the CFNAI archives. This results in an evaluation sample of 138 forecasts, each of which corresponds to a different real-time vintage.[14] The first vintage covers the period from January 1973 to October 2004, with the first four years of data being 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 the median forecast at each horizon to the realized GDP growth, where we gauge the sensitivity of our findings to the use of sequential real-time NIPA releases (first, second, and third) for realizations of GDP. We measure forecast gains over the Blue Chip Consensus and the Survey of Professional Forecasters, as well as the MF-BVAR of Schorfheide and Song (2015) in Section 4, by reporting percentage root mean squared forecast error (RMSFE) gains/losses, given by

$$RMSFEG_{b,a}^{h} = 100 * \left(1 - \frac{RMSFE_b^h}{RMSFE_a^h}\right),$$

where for horizon $h = 0, \ldots, 4$ (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 using a one-sided Diebold and Mariano (1995) test of equal mean squared forecast error that is consistent with the sign of the percentage gain/loss.[16] We report p-values using Student's t critical values, and also incorporate the small-sample size correction that was recommended by Harvey, Leybourne, and Newbold (1997). Heteroskedasticity and autocorrelation consistent (HAC) standard errors are constructed for this purpose using the Bartlett kernel with a bandwith set equal to 50% of

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), meaning that the "nowcast" for GDP in this instance would be described more accurately as a backcast, while the one-quarter-ahead forecast might be more reflective of a nowcast.

13 To err further on the side of caution, we also adopt an alternative timing assumption in which, for variables that are typically released after the Blue Chip survey has been 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 the online appendix for the details.

14 We lose 12 out-of-sample observations for the four-quarter-ahead forecast horizon, 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, hence resulting in a delayed release schedule for the CFNAI that would have given the MF-BVAR an informational advantage. The results are almost identical if this quarter is included.

15 More precisely, the first six months of 1973 are used to obtain mean values for the dummy priors, while data from June 1973 to 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.

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, 2011b), we follow the conservative approach of Carriero, Clark, and Marcellino (2012a) in reporting one-sided test results.

the sample size, as a compromise between the standard bandwidth settings of either the number of lags plus the forecast horizon or the sample size, as suggested by Kiefer and Vogelsang (2005). An inspection of the cross-correlograms of the mean squared error differences of our tests suggests that the former bandwidth is far too short, but the results are robust to using 25% of the sample size as the bandwidth.

3.4. Hyperparameter selection

We operationalize our two-step procedure for selecting the hyperparameters (Λ^p, Λ^σ) of our prior by first recursively estimating a bi-variate MF-VAR with a related monthly variable that is 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 the online appendix. 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 that surround 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 in order 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. While feasible, our approach 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 autoregressions. As we show in Section 5.2, including Λ^σ in the initial optimization improves the predictive accuracy at shorter forecast horizons, particularly for larger models.

We facilitate both of these steps in the elicitation of optimal hyperparameters by imparting a set of hyperpriors (Giannone et al., 2015). In the case of λ_1 (tightness), we use a gamma density with a mean of three and a standard deviation of two. 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 one. Finally, for each of the innovation standard deviations in Λ^σ, we use a gamma density with a mean of one 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 that are found commonly in the literature for persistent variables (equal to five for the tightness and one for all of the 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 $\Lambda^\star$ that are 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 one. The wide array of volatilities in our dataset is 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 mnemonics (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 performances of our model's forecasts are unlikely to be impacted significantly by any 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, Carriero, & Marcellino, 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 4000 draws are used, with the first 2000 draws being 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

When discussing the results of our baseline MF-BVAR, we begin in Section 4.1 by benchmarking our MF-BVAR's forecasting performance against those of the Blue Chip Consensus and the Survey of Professional Forecasters. Next, we compare its predictive accuracy in Section 4.2 against that of the smaller scale 11-variable MF-BVAR

[17] See the online appendix for further details. The results are quite robust to the use of alternative diffuse priors, possibly reflecting the greater number of monthly as opposed to quarterly series that are included in our model.

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

[19] The notation of Giannone et al. (2015) for the hyperparameters corresponds to the inverse of ours, such that an overall tightness of four in our context is equal to 0.25 in their case. We also experimented with estimating the inverse of our hyperparameters, as in their paper, and obtained broadly similar results, provided that the priors were adjusted accordingly to represent the same broad coverage of the hyperparameter domain.

Table 2
Prior and posterior estimates of the hyperparameters of the 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: The table reports the key hyperparameters that are 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 Schorfheide and Song (2015) and highlight the relative ability of our MF-BVAR to capture important turning points during our sample period. Additional survey- and model-based forecast comparisons can be found in the online appendix.

4.1. Benchmarking to surveys

Fig. 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 along with 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 the use of alternative real-time releases for GDP (first, second, third) to evaluate forecast accuracy.

The key insight from Fig. 1 is that *medium*-run predictions from our baseline MF-BVAR outperform the Blue Chip Consensus mean forecasts in real-time over our sample period. The MF-BVAR delivers gains as large as 10%–15% for quarter-over-quarter real GDP growth rates at the three- and four-quarter-ahead horizons, which are statistically significant at standard confidence levels using 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 from that of the Blue Chip Consensus.

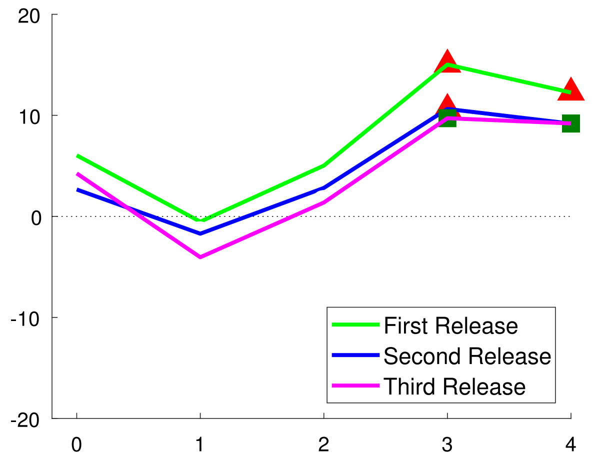

Fig. 1. Percentage gains in RMSFE relative to Blue Chip Consensus. Notes: This figure displays root mean squared forecast error (RMSFE) gains, as described in Section 3.3, for our baseline MF-BVAR relative to the Blue Chip Consensus (BCC) mean forecasts for quarter-over-quarter real GDP growth over the period 2004Q3–2016Q1 at forecast horizons of 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 errors between the MF-BVAR and BCC forecasts at the 10% and 5% levels, respectively.

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*	−15.53**
1	−0.32	−0.77	−4.48
2	7.25	4.55	−4.04
3	11.82**	8.72**	11.41*
4	10.91**	7.80	8.77*

Notes: The entries in the 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 period 2004Q3–2016Q1 at forecast horizons of 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.
*Denote rejections of the null of equal mean squared forecast errors between the MF-BVAR and BCC forecasts at the 10% levels.
**Denote rejections of the null of equal mean squared forecast errors between the MF-BVAR and BCC forecasts at the 5% levels.

These results suggest that the MF-BVAR is a viable approach to nowcasting and forecasting GDP in real-time using 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 probably puts us at a disadvantage relative to the potentially much larger information set that is available to professional forecasters. Furthermore, as is well-known, the means and medians of professional surveys embed combinations of forecasts that in general are difficult to beat using a single model.

Our findings are even more encouraging when we break down the results in Table 3 by forecast origin using the second real-time release of GDP to evaluate the forecast accuracy (the results are qualitatively similar using the first or third releases instead). 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 RMSFE gains of roughly 5%–12% that are statistically significant in many cases. 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 Fig. 1, the gains at the nowcast and two quarter out horizons are countered by losses at forecast origin $R3$. For the nowcast, in particular, this loss is large and statistically significant. Thus, it would appear that the $R3$ forecast origin (more appropriately called a "backcast") is where the informational constraints that we impose on our baseline MF-BVAR are particularly binding for the nowcast to two-quarter-ahead horizons. At horizons of three to four quarters out, though, the RMSFE gains that we find by forecast origin remain similar to what we find in Fig. 1, in both magnitude and statistical significance.

As a further robustness check on the informational content of our baseline MF-BVAR, Fig. 2 shows RMSFE percentage gains in comparison to the Survey of Professional Forecasters median forecasts. Given our timing assumptions, our evaluation against this survey uses the forecast origin $R1$ (second month of the quarter). Once again, we gauge the way in which the assessment of the forecast accuracy 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 sizes from three to four quarters out and comparable performances at shorter forecast horizons for real GDP growth.

These results are reassuring, given that Section 3.2 noted that the timing assumptions underlying our vintage data are likely to be aligned more closely with the information set of the Survey of Professional Forecasters than that of the Blue Chip Consensus.[20] Finally, it is also worth noting that the results of comparisons to either the BCC or the SPF are most favorable to the MF-BVAR when using the first real-time release of GDP, and almost identical with the other two. Thus, we present results for the remainder of the paper using the second real-time release of GDP to avoid overstating our findings.

4.2. Benchmarking to Schorfheide & Song

Next, we compare the performance of our baseline MF-BVAR against that of the smaller-scale 11-variable

20 In the online appendix, we consider the issue of data availability in a slightly different way by repeating the results in this section using an alternative timing assumption which aims to align the information set of our baseline MF-BVAR more closely to the timing of the Blue Chip Consensus survey.

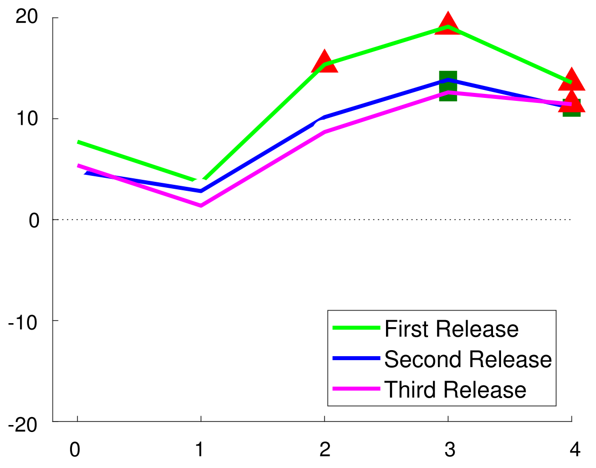

Fig. 2. Percentage gains in RMSFE relative to the Survey of Professional Forecasters. Notes: This figure displays root mean squared forecast error (RMSFE) gains, as described in Section 3.3, for our baseline MF-BVAR relative to the Survey of Professional Forecasters (SPF) median forecasts for quarter-over-quarter real GDP growth over the period 2004Q3–2016Q1 at forecast horizons of 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 errors between the MF-BVAR and the SPF forecasts at the 10% and 5% levels, respectively.

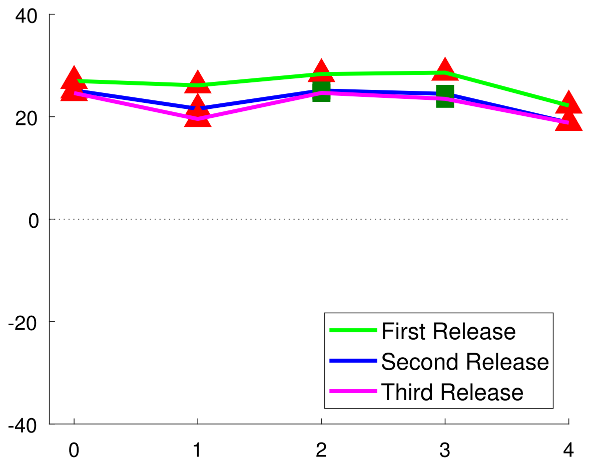

Fig. 3. Percentage gains in RMSFE relative to Schorfheide and Song (2015). Notes: The figure displays root mean squared forecast error (RMSFE) gains, as described in Section 3.3, for our baseline MF-BVAR relative to the 11-variable MF-BVAR of Schorfheide and Song (2015) for forecasts of quarter-over-quarter real GDP growth over the 2004Q3–2016Q1 period at forecast horizons of 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 Schorfheide and Song (2015). 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 rejections of the null of equal mean squared forecast errors between the MF-BVAR and Schorfheide and Song (2015) forecasts at the 10% and 5% levels, respectively.

MF-BVAR of Schorfheide and Song (2015). We include this comparison in order to augment our findings relative to surveys with a model-based forecast, and in particular, to emphasize the magnitudes and timings of the gains that accrue from considering a larger set of monthly real activity variables in real-time than has been shown previously in the MF-BVAR context. The estimation and evaluation of both models are identical to what is described in Section 3, with the hyperparameters being obtained in both instances using our two-step procedure and re-optimized with the MDD every three years. The real variables from 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 strong differences still exist from their original implementation, due to differences in sample periods and hyperparameter selection.

Fig. 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%, 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 due in part to the additionally incorporated real activity series in our model relative to their smaller-scale model, including prominent real, financial and price series[21]; note that the next section discusses the types of real activity variables that drive these information gains at various forecast horizons.

We 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) by also reporting the results of the fluctuation test of Giacomini and Rossi (2009) as implemented by Rossi and Sekhposyan (2010) for each model's 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 performances of the two models vary over time, it should reflect these turning points in relative performance and facilitate appropriate statistical inferences of these differences. Fig. 4 displays the fluctuation test statistic as calculated over rolling windows of 36 months, together with the 90% confidence bands as calculated by 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% of the forecast evaluation sample.[22]

21 In our sample, the BCC outperforms the Schorfheide and Song (2015) model by amounts from a high of 30% at the nowcast horizon to a low of 11% 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, but the

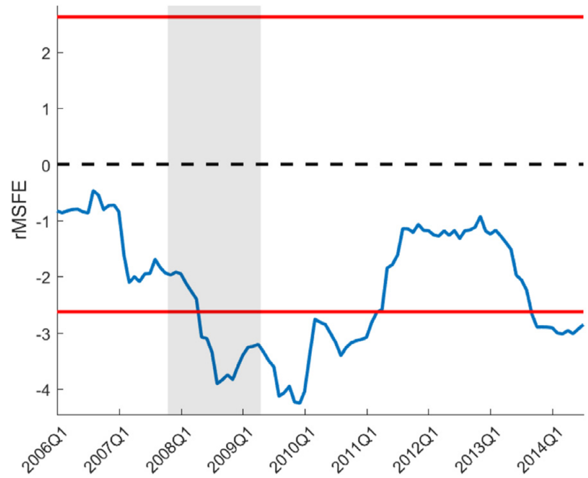

Fig. 4. Results from the fluctuation test of Giacomini and Rossi (2009): baseline MF-BVAR vs. Schorfheide and Song (2015). Notes: The figure shows the fluctuation test statistic (blue solid line) of Giacomini and Rossi (2009) for the nowcasts of the baseline MF-BVAR and the 11-variable MF-BVAR of Schorfheide and Song (2015). The test is 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 performances 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).

Negative values of the test statistic reflect windows where the root mean squared forecast error for the baseline MF-BVAR is lower than that of the 11-variable MF-BVAR of Schorfheide and Song (2015). We facilitate comparisons with the business cycle by also shading the time periods associated with the Great Recession (as defined by NBER).

Several immediate conclusions can be drawn from Fig. 4. First, our baseline MF-BVAR outperforms the smaller MF-BVAR over any 36-month period in our evaluation period. However, fluctuations in relative performance between the two models over the evaluation period do exist. 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% confidence level, as is shown in Fig. 4. The performances of the models then diverge leading into the Great Recession and the subsequent recovery, when 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 the weakness of the ensuing recovery. These results are consistent with a larger information set being more important during business cycle turning points, a result that has been 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, as the baseline MF-BVAR also achieved 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 the various specification choices that are faced by practitioners who seek to apply this methodology. To do so, we deconstruct the forecast performance of our baseline MF-BVAR 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 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 that encompassed the NIPA expenditure categories. However, the computational demands of larger systems like ours make it reasonable to question whether or not this is necessary, or whether considerably smaller models informed by a few variables could perform just as well. We answer this question by proceeding in two steps. First, we identify a suitable "small-scale" model to serve as the counterpoint for our baseline MF-BVAR.[23] Second, we expand the size of this model iteratively by incorporating an additional set of variables from one of the subcomponents of GDP. By parsing the model size comparisons along subcomponents of GDP, we identify heterogeneous gains for certain types of variables across forecast horizons. While the ordering of subcomponents used here is arbitrary, the results that we report hold qualitatively across alternative orderings as well.

Table 4 reports the gains in RMSFE that are achieved using our baseline 37-variable MF-BVAR relative to each of the smaller scale models obtained by expanding the set of real activity variables sequentially according to the expenditure categories in Table 1. The entries in the table are almost uniformly positive, suggesting that forecast performance gains accrue as the 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

results are qualitatively similar for the bandwidth selection criterion used by Rossi and Sekhposyan (2010).

23 One could also interpret this "small-scale" model comparison as providing reassurance that the performance differences between the 11-variable model of Schorfheide and Song (2015) and the medium–large baseline MF-BVAR that have been reported were not driven by the additional price and financial series that were included in that comparison.

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: The entries in the 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 of 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 of 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 subcomponent of GDP that is omitted from 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.

*Denote rejection of the null of equal mean squared forecast errors between our baseline MF-BVAR and alternative models at the 10% levels.

**Denote rejection of the null of equal mean squared forecast errors between our baseline MF-BVAR and alternative models at the 5% levels.

expenditures, nonfarm payroll employment, and the unemployment rate. However, our baseline MF-BVAR still delivers large and statistically significant RMSFE gains at all horizons, ranging from 13% to 25%.

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

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

It is not until we add variables that are related to residential investment (+RES) that the gap in forecast performance with our baseline MF-BVAR begins to shrink appreciably at medium-term horizons as well. However, this occurs at the same time as a small deterioration in performance for the nowcast. An inspection of the forecast paths from this and previous models suggests that data on residential investment helps to capture the slow recovery following the Great Recession better, 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 that are lost by the model with the additional residential investment indicators, while essentially delivering forecasts that are very similar to those of our baseline MF-BVAR at other forecast horizons. Finally, the addition of variables that are 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

Thus far, the estimation strategy 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 model's one-step-ahead forecast performance (Geweke, 2001). What is less clear is how this approach performs at different forecast horizons. Furthermore, the computational costs of 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?

We answer this question by estimating 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 that are collected in the vector Λ^σ are no longer included in the optimization of the MDD. Instead, as is typical in the literature, they are chosen by using univariate autoregressions with sample data.[25] In general, the optimal hyperparameters for the residual variances tend to be quite similar to the values implied by the traditional univariate autoregression approach, at least for the monthly variables that do not suffer from aggregation issues (see footnote 25). While estimates of the optimal

24 Clearly, for instance, the size of the gap and the statistical significance of the 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 according to the number of series that enter the BVAR to reflect the implementation of the prior.

Table 5
Alternative hyperparameters for MF-BVAR.

Λ^p	Description	Specification		
		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 that are 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.
In all cases, posterior mode estimates are reported using the first data vintage.

hyperparameters for the residual variances come in both above and below the default values, the average difference between the two across all of the monthly series and each of the four re-optimized vintages is 8%. As a consequence, any differences in forecast performance that arise from optimizing the residual variances are likely to be manifested primarily in the optimization of the residual variances of the *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 by 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 that correspond 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 that are 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, though there is a substantially smaller role for other forecast horizons. Relative to the *Hybrid* case, the baseline specification records a 12% 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% 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 using the baseline are likely to be 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 the online appendix for further discussion).

Table 6
Percentage gains in RMSFE relative to alternative specifications.

Hyperparameters	Specification		
	Levels		Growth rates
	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: The entries in the table correspond to root mean squared forecast error (RMSFE) gains for forecasts of quarter-over-quarter real GDP growth over the period 2004Q3–2016Q1, as described in Section 3.3, from our baseline MF-BVAR in levels with optimal hyperparameters that are 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 RMSFEs for quarter-over-quarter real GDP growth at forecast horizons of 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.
*Denote rejection of the null of equal mean squared forecast error between our baseline MF-BVAR and the alternative prior specifications at the 10% levels.
**Denote rejection of the null of equal mean squared forecast error between our baseline MF-BVAR and the alternative prior specifications at the 5% levels.

Finally, the prior has less influence on the relative forecasting performance at further out horizons in both cases, 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 the 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, which explains why the differences in forecast accuracy between our baseline and the *Fixed* settings are so small at longer horizons.

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, it would also be natural to work with a stationary specification in growth rates, given that the ultimate forecast of interest is the *growth* rate of real GDP. 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.

When choosing hyperparameters for this alternative model using the marginal data density, 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 using 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 zero, but the optimal value for this hyperparameter, λ_3, came in routinely at zero. As a consequence, this form of shrinkage was shut down by setting $\lambda_3 = 0$. As with our baseline MF-BVAR in levels, all of the other hyperparameters (including the Λ^σ) are elicited using the MDD.

The last column of Table 6 reports RMSFE percentage gains of 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 obtained 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%–15%, with the 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 the capture of the non-stationary behavior that is common to many indicators of economic activity.

6. Conclusion

This paper has 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, the other modeling choices that are inherent in 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 that is observed at different frequencies into the forecasts produced regularly at central banks and other institutions which are charged with tracking the economy. Future work should seek to determine the scalability of these findings to other key series that are of interest to central bankers and private sector analysts (e.g. inflation). In addition, it would also be valuable to provide direct evaluations of how MF-BVARs compare to other popular forecasting methods in regard to predictive accuracy in mixed frequency data-rich environments.

Acknowledgments

This research did not receive any specific grants 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. In memoriam: Scott and Andrew would like to dedicate this paper to our dear departed friend and colleague Alejandro Justiniano. May he rest in peace knowing that his influence on us and everyone that he interacted with in the economics profession will leave a lasting impression on the discipline that he so passionately pursued.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2019.02.010.

References

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

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

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

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

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

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

Carriero, A., Galvao, A. B., & Kapetanios, G. (2016). A comprehensive evaluation of macroeconomic forecasting methods. EMF Research Papers 10, Economic Modelling and Forecasting Group. URL https://ideas.repec.org/p/wrk/wrkemf/10.html.

Carriero, A., Kapetanios, G., & Marcellino, M. (2012b). Forecasting government bond yields with large Bayesian vector autoregressions. *Journal of Banking & Finance*, *36*(7), 2026–2047, URL http://www.sciencedirect.com/science/article/pii/S0378426612000702.

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

Clark, T. E., Carriero, A., & Marcellino, M. (2016). Large vector autoregressions with stochastic volatility and flexible priors. Working Paper 1617, Federal Reserve Bank of Cleveland. URL https://ideas.repec.org/p/fip/fedcwp/1617.html.

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

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

Del Negro, M., & Schorfheide, F. (2011). Bayesian Macroeconometrics. In J. Geweke, G. Koop, & H. Van Dijk (Eds.), *The Oxford Handbook of Bayesian Econometrics* (pp. 293–389). Oxford University Press.

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

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

Durbin, J., & Koopman, S. J. (2012). Time series analysis by state space methods, (2nd ed.). Oxford University Press.

Foroni, C., Ghysels, E., & Marcellino, M. (2013). Mixed frequency vector autoregressive models. In *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher a. Sims* (pp. 247–272). Emerald Group Publishing Limited, Ch. 7.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2014). Bayesian data analysis, *Texts in Statistical Science*, (3rd ed.). London: Chapman and Hall/CRC.

Geweke, J. (2001). 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, URL http://www.sciencedirect.com/science/article/pii/S0304407616300653.

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

Giacomini, R., & Rossi, B. (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies*, *76*(2), 669–705, URL http://www.jstor.org/stable/40247617.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *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., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.

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

Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, *21*(6), 1130–1164.

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

McCracken, M. W., & Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. Working Papers 2015-12, Federal Reserve Bank of St. Louis URL https://ideas.repec.org/p/fip/fedlwp/2015-012.html.

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

Reifschneider, D. L., & Tulip, P. (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.) URL https://ideas.repec.org/p/fip/fedgfe/2007-60.html.

Rossi, B., & Sekhposyan, T. (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, URL http://www.sciencedirect.com/science/article/pii/S0169207009001332.

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

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

Scott A. Brave is a policy economist in the economic research department of the Federal Reserve Bank of Chicago. Brave received a B.A. in economics with honors from the University of Chicago and an M.B.A. with concentrations in economics, statistics and finance from the University of Chicago Booth School of Business.

R. Andrew Butters is an Assistant Professor in the Business Economics & Public Policy department at the Kelley School of Business at Indiana University. Prior to joining Kelley, he was an Associate Economist in the Economic Research department at the Federal Reserve Bank of Chicago. His research has appeared in journals, such as International Journal of Central Banking and Economics Perspectives. He received his B.A. from the University of North Carolina at Chapel Hill, and his Ph.D. from the Kellogg School of Management at Northwestern University.

Alejandro Justiniano is a senior economist and research advisor in the economic research department at the Federal Reserve Bank of Chicago. His research focuses on empirical methods for closed and open economy macroeconomics. Justiniano's research has been published in the Journal of Economic Dynamics and Control and the American Economic Review. Before joining the Chicago Fed in November 2007, he was an economist at the Board of Governors of the Federal Reserve for two years. Before that, he also spent two years at the International Monetary Fund. Justiniano received a B.A. from the University of Maryland College Park and a Ph.D in economics from Princeton University.