Real-Time Forecasting with a Large, Mixed Frequency, Bayesian VAR*

Michael W. McCracken† — Federal Reserve Bank of St. Louis

Michael Owyang‡ — Federal Reserve Bank of St. Louis

Tatevik Sekhposyan§ — Texas A&M University

May 17, 2016

Abstract

We assess point and density forecasts from a mixed-frequency vector autoregression (VAR) to obtain intra-quarter forecasts of output growth as new information becomes available. The econometric model is specified at the lowest sampling frequency; high frequency observations are treated as different economic series occurring at the low frequency. We impose restrictions on the VAR to account explicitly for the temporal ordering of the data releases. Because this type of data stacking results in a high-dimensional system, we rely on Bayesian shrinkage to mitigate parameter proliferation. The relative performance of the model is compared to forecasts from various time-series models and the Survey of Professional Forecaster's forecasts.

Keywords: Vector autoregression, Stacked vector autoregression, Mixed-frequency estimation, Bayesian methods, Nowcasting, Forecasting

JEL Codes: C22, C52, C53

---

*We have benefited greatly from conversations with Bob Rasche. We are grateful to Domenico Giannone, Eric Ghysels and participants of seminars at the Bank of Canada, Federal Reserve Board, Ohio State University, University of Houston, 2013 CIRANO-CIREQ Workshop on Data Revisions in Macroeconomic Forecasting and Policy, 2013 $(EC)^2$ Conference, 2014 SNDE Conference, 2014 Midwest Macro Meeting, 2014 EABCN Conference, 2014 IAAE Meeting, 2015 Econometrics Workshop of the St. Louis Fed, 2015 Workshop on High-Dimensional Time Series in Macroeconomics and Finance, 2016 Texas Econometrics Camp for comments and suggestions. Joe McGillicudy and Hannah G. Shell provided research assistance. The views expressed herein do not reflect the official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

†Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166, USA; tel: (314) 444-8594; e-mail: Michael.W.McCracken@stls.frb.org

‡Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166, USA; tel: (314) 444-8558; e-mail: Michael.T.Owyang@stls.frb.org

§Texas A&M University, 3060 Allen Building, 4228 TAMU, College Station, TX 77843, USA; tel.: (979) 862-8857; e-mail: tsekhposyan@tamu.edu

1 Introduction

Information available to forecasters changes over the forecasting horizon. This variation occurs because macroeconomic variables are observed at mixed frequencies (i.e., some variables are monthly, while others are quarterly) and are released at different times during the quarter, each with its specific publication lag. Thus, modeling the information flow in an efficient way is a central issue for practitioners and policymakers.

Many of the models that have been developed to deal with the variation in incoming information are estimated at the highest sampling frequency, treating the low frequency variables as latent processes. Typically, these are modeled in a state-space framework. In this paper, we address the forecaster's information problem by using a mixed-frequency Bayesian VAR estimated at the lowest common data frequency. More specifically, our VAR stacks a variable's three monthly observations in a quarter, modeling each individual monthly release with a separate process. In addition, we incorporate the data release properties with recursive restrictions imposed on the VAR. Ordering the data consistent with the release calendar helps us deal with the publication lags: We model the available information set and consider its predictive content with no particular emphasis to the reference period of the released data.

Our high-dimensional VAR is evaluated in real-time for out-of-sample nowcasting and forecasting of the U.S. real GDP growth. Given that our model is large and unrestricted, we estimate it relying on Bayesian shrinkage.[1] Our postulated model has a couple of advantages over the high-dimentional state-space models as (i) it is robust to mis-specification and (ii) computationally it is much easier to impose "optimal" shrinkage in our case since an analytical closed-form solution to the marginal data density calculations (MDD) is available. Our setup intends to replicate a realistic information flow that policymakers and practitioners would be interested in, and we evaluate the reliability of our models' forecasts relative to commonly used time series alternatives considered in the literature, as well as competitive consensus forecasts provided by the Survey of Professional Forecasters (SPF).

There are a variety of alternative forecasting models that address the variation in data sampling and data release timings—e.g., state-space models and Mixed Data Sampling (MIDAS). State space models are usually specified at a high-frequency and filtering/smoothing algorithms provide updates for the missing values of low frequency variables.[2] MIDAS regressions (Ghysels et al., 2004)

[1] See Foroni, Guérin and Marcellino (2013) for a classical evaluation of a small-scale stacked VAR for the Euro-area in real-time. Blasques et al. (2016) provides a weighted maximum likelihood estimation procedure for mixed frequency dynamic factor models considered in the stacked form. Ghysels (2015) considers the advantages of observation driven stacked VARs when used in the context of causal analysis.

[2] Giannone et al. (2008), Aruoba et al. (2009), and others utilize state-space framework to estimate factor models. They summarize large panels of macro data with a few factors to obtain high-frequency updates for missing observations of (latent) variables of interest. Schorfheide and Song (2015) and Brave et al. (2015) use a mixed-frequency medium-scale VAR to obtain monthly nowcasts and forecasts of quarterly GDP growth.

are typically implemented for univariate autoregressive distributed lag models, allowing the high frequency data to enter the predictive regressions by sparsely parameterized weighting polynomials. Under certain non-restrictive conditions, VARs can be mapped back to these alternatives.[3] In fact, our forecasting model parallels with a multivariate version of MIDAS regressions where we use unrestricted, data-driven weights instead of weighting polynomials. Thus, which model performs the best can not be known ex ante and is an empirical question.[4]

We find that the mixed-frequency nature of the data is important for nowcasting, but that not all intra-quarter information improves forecast accuracy. Our model dominates some competitive alternatives—including a variant of a factor model—when nowcasting real GDP growth. In fact, the accuracy of our model's nowcasts are comparable to those of the SPF's: The root-mean-squared forecast error associated with our nowcasts is only 7 percent higher than that associated with the nowcasts of the SPF. Admittedly, we find considerable time variation in the forecasting performance of the models. For example, our model performs better than the factor model during the post-Great Recession period, but the nowcasts of the SPF have been more accurate since 2000. Controlling for a few outliers by using a median-squared forecast errors loss criterion makes our model even more competitive relative to the SPF.

The remainder of the paper is organized as follows. Section 2 lays out the econometric specification of our proposed model and discusses the data. Section 3 considers the estimation methodology. Section 4 presents the construction of nowcasts and forecasts; Section 5 discusses the forecasting results and compares them to various alternatives. We conclude with Section 6.

2 The Setup

In this section, we describe the data and method for performing our out-of-sample forecasting exercises. In our description of the data, we highlight the timing of their releases. These release dates will be used to generate a number of restrictions on the contemporaneous effects in the VAR outlined in the following subsection.

2.1 Data

Table 1 summarizes the data. The choice of the data series is based on some judgment as to which economic variables are important and are broadly followed by markets and policymakers.[5] For

[3] For instance, De Mol et al. (2008) consider the equivalence of principal components and Bayesian shrinkage. Bai et al. (2013) show the equivalence between the state space models and MIDAS regressions.

[4] Bai et al. (2013) suggest that MIDAS regressions would be preferred to the state-space models when the latter is misspecified or overparameterized. Both, misspecification and overparameterization, are difficult to evaluate ex ante. Schorfheide and Song (2015) show that the nowcasts and forecasts of real GDP growth from the mixed frequency VAR are empirically similar to those obtained from unrestricted MIDAS regressions.

[5] For some discussion on data choice, see McBride (2014).

each series listed in the first column, we indicate the original frequency of the data, as well as the transformation used to achieve stationarity.[6] We specify the approximate release date of the series, their publication lags, the dates of the first available vintages, as well as the economic data release report from which the series are obtained. The data are combined from ALFRED (ArchivaL Federal Reserve Economic Data) database of the St. Louis Fed and Haver Analytics (marked by a †). We have monthly real-time vintages from 1980:01 to 2014:04. However, in the estimation we use vintages starting from 1985:01 onwards to ensure we have sufficient degrees of freedom for our estimation.[7] The earliest observation for any vintage is 1971:01.[8]

INSERT TABLE 1 HERE

To make the information reported in the table more transparent, take the example of employment labeled as "Empl.: Total Nonfarm Payrolls". "M" in the frequency column indicates the monthly nature of the employment data. Employment, with some exceptions, is released on the first Friday of each month and pertains to the previous month. Thus, employment series has one-month publication delay, which captures the difference between the release and reference months.

Over time, new real-time data series become available. As we attain real-time vintages for these variables, we incorporate them in our estimation. As a result, the dimensionality of our model changes six times across our collection of vintages.[9] Prior to 1996, our model is comprised of sixteen monthly and one quarterly variables. In the late 1990s, we add core CPI (1996:12), ISM Manufacturing PMI (1997:03), and New Single Family Houses Sold (1999:07). The latter two series are considered to be leading indicators for real activity. The core CPI, on the other hand, can improve the interest rate—and consequently real output growth—forecasts as it is central to the central bank's objectives. The real-time data for PCE headline and core price indices become available for the 2000:07 vintage. Initial Unemployment Insurance Claims is added in the 2009:06 vintage and the New Orders Index and the Supplier Deliveries Index are added in the 2009:11 vintage. We include these variables because of their timely release.

[6] The transformations are broadly consistent with (unreported) unit root test results. The growth rates are annualized, consistent with quarterly frequency, while variables in percentage differentials are multiplied by 4.

[7] While degrees of freedom is less important for Bayesian estimation, we start the sample in 1985:01 to minimize the impact of the prior on the estimation results. Panel B in Figure 2 shows that considerable amount of uncertainty remains even when starting the estimation in 1985:01.

[8] Although the dataset is real time in the sense that the most recent vintage is used for estimation in each out-of-sample experiment, we do not forecast data-revisions. To accommodate data revisions, additional equations would have to be included to link the revisions across releases (especially for GDP). We leave this exercise for future research.

[9] Retail Sales is a splice of two series RETAIL (Retail Sales) and RSAFS (Retail and Food Services Sales). Post 2006:06, we use the RSAFS data and its revisions, always splicing the data with the historic values of RETAIL for pre-1992:01. Moreover, for Retail Sales, Housing Starts and New Houses Sold, there are vintages that provide revisions for historical data, but no new observations. In these cases, we proxy the new values by repeating the values of the previous month so that the last few values (depending on the number of missing observations) in the faux-vintage are identical. This is consistent with the assumption that we are modeling what is in the forecasters information set.

Since our objective is to model the timing of the releases explicitly, we collect the data into four groups based on the timing of their release dates using a few simplifying assumptions:

1. Group 1 consists of the variables that are released during the first week of the month; Group 2 consists of the variables released during the second and third weeks in the month; Group 3 consists of the non-financial variables released during the last week of the month; Group 4 includes the financial variables.[10]

2. The number of estimated parameters in our model increase faster than in the standard VAR of equal lag order. Introducing variables at higher than monthly frequencies (weekly or daily) will exacerbate this problem. In order to mitigate parameter proliferation, we consider data only in monthly and quarterly frequencies. Financial variables are summarized at a monthly frequency, using end of month values. The exception is the S&P 500 Composite Stock Index, for which we use the average value of the month.

3. Initial Unemployment Insurance Claims is a weekly series and is usually revised substantially. Each weekly release reports the four-week-moving-average of initial claims. We use the first weekly release in the month which reflects the past month's average of initial claims. Based on this assumption, Initial Unemployment Insurance Claims is placed in Group 1 with the variables that are released at the beginning of the month.

We assume that variables released in the same group are simultaneously correlated (for instance in many cases they are part of the same economic report). This assumption may be reasonable from a policymaker's point of view; in many circumstances, the costs associated with acquiring and explaining daily forecasts far exceed the benefits. We also further assume that each group reacts to the groups released before but not to the groups of variables released after. Note that we are modeling the temporal restrictions associated with data releases, and we do not take a stance on causal relationships. Thus, the adopted structure could well represent the data flow if we choose to a (bi)weekly structure for updating the forecasts.

It should be noted that our data are organized based on the release and not reference calendar. Moreover, the longest publication lag we have across our data series is one month. Thus, at the end of each month, we have complete realizations for all the information scheduled to be released in the previous month. As a result, in January, the information set that is used for forecasting typically includes data pertaining to the past month (e.g. December IP) and the current month (e.g. January end of month term spread), as well as the past quarter's realization of GDP.

[10] The release dates required some approximations. For example, the Employment Situation Report can be released on the second Friday of the month when the first Friday falls on the 1st of the month and when the prior month has fewer than 31 days. This happens rarely. In such cases, we treat the reports as in the table since the changes usually preserve the temporal ordering.

We use all these information for predicting the release of the current quarter real GDP growth. Within a quarter, the exercise parallels the direct forecasting approach (which is more robust in the case of misspecification), rather than filling the missing observations consistent with a particular parametric model such as the state-space model.

2.2 Model

Our forecasting model is a VAR, estimated at the lowest sampling frequency. While the data releases are observed in mixed frequencies, we treat multiple releases in the lowest frequency as separate observations modeled in a blocked linear form consistent with the previously defined groups.

Without loss of generality, we define our notation consistent with the specification used in our empirical application, mixing monthly and quarterly variables; alternative frequency mixing is a straightforward extension. Let $x_{t-\tau}$ represent the high frequency (monthly) variable and y_t denote the low frequency (quarterly) variable, for quarters $t = 1, \ ..., \ T$ and months within the quarter $\tau = \{2/3, 1/3, \ 0\}$. In this setup, each τ represents an intra-quarter month: $\tau = 0$ indexes data that are observed during the first month of the t quarter, $\tau = 1/3$ indexes data that are observed during the second month of the t quarter, and $\tau = 2/3$ indexes data that are observed in the last month of the t quarter.

For illustrative purposes, consider the case with one quarterly variable (say, real GDP) and one monthly variable (say, payroll employment) that is released during each of the three intra-quarter months and in the first month of the quarter, when it is released jointly with the advanced release of the GDP (see Table 1 for the temporal ordering), it is released with the group 1 data prior to the release of the quarterly variable that is in group 3. Define the vector $Y_t = [x_{t-1+1/3}, x_{t-1+2/3}, x_t, y_t]'$; then, the mixed-frequency VAR takes the form:

$$A'Y_t = C + \sum_{l=0}^{p} B_l' Y_{t-l} + \varepsilon_t, \ t = 1, \ ..., \ T, \tag{1}$$

where A' and B_l' are $n \times n$ parameter matrices, C is a vector of deterministic components, p is the lag length (equal to four in the empirical implementation), $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I})$, and T is the sample size in quarters.

The model structure is different from a standard VAR. Our model allows each monthly release of a data series within a quarter to have its own data generating process (which ultimately will be described by the data). If our data was only at a monthly frequency, the standard monthly VAR could be written in the stacked form with a set of restrictions on the coefficients across months of the quarter.

The matrix A' is used to impose temporal ordering restrictions on the elements of Y_t depending on the timing of their releases over the quarter. In the example, the three monthly releases of

$x_{t-\tau}$ are observed in a specific order followed by the eventual release of the quarterly value of y_t; thus, $n = 4$ and A' is lower triangular (4×4) matrix. Since in this case the monthly and the quarterly variables, when available jointly, are released in completely different weeks (group 1 versus group 3), we impose a lower triangular A'. If, on the other hand, the monthly variable that we consider belonged to the same group as the quarterly variable (consider the consumer sentiment index instead of payroll employment), the monthly release of the high-frequency variable x_t would be occurring contemporaneously with the quarterly release of y_t: we would leave the contemporaneous correlations of x_t and y_t unrestricted.[11] Accordingly, the A' matrix, instead of being lower triangular, would have the following form:

$$A' = \begin{pmatrix} a_{1,1} & 0 & 0 & 0 \\ a_{2,1} & a_{2,1} & 0 & 0 \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix}, \tag{2}$$

where $a_{i,j}$-s, $i, j = 1, ..., 4$, are unrestricted scalars.

In general, we can write A' in terms of monthly blocks A'_j, where j reflects the month of the quarter, and $\mathbf{a}$ denotes unrestricted matrices of varying dimensions:[12]

$$A' = \begin{pmatrix} A'_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{a} & A'_3 & \mathbf{0} \\ \mathbf{a} & \mathbf{a} & A'_1 \end{pmatrix}. \tag{3}$$

Each monthly block, as suggested above, has its own internal timing. Based on the data and release date assumptions in Table 1, the monthly block A'_m, for $m = 1, 2, 3$, would have the following lower-triangular form

$$A'_m = \begin{pmatrix} \mathbf{a}_{1,1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{a} & \mathbf{a}_{2,1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{a} & \mathbf{a} & \mathbf{a}_{3,3} & \mathbf{0} & \mathbf{0} \\ \mathbf{a} & \mathbf{a} & \mathbf{a} & \mathbf{a}_{4,4} & \mathbf{0} \\ \mathbf{a} & \mathbf{a} & \mathbf{a} & \mathbf{a} & \mathbf{a}_{5,5} \end{pmatrix}, \tag{4}$$

where $\mathbf{a}$ again denotes unrestricted matrices of different dimensions and the $\mathbf{a}_{i,j}$s are unrestricted matrices of dimensions corresponding to the variables in each group. The first month in the quarter $(m = 1)$ is different than the rest since it includes the quarterly release of the GDP: A_1 is one dimension larger than A_2 and A_3. This dimensionality increase is associated with the third

[11] In this case, the formulation of our forecasting problem is such that y_t is not updated in response to the x_t because we assume they are released at the same time.

[12] We omit the subindices and use a generic notation for simplicity.

group of variables: it now contains real GDP growth. For instance, in 2013:02, $\mathbf{a}_{1,1}$ is a matrix of dimensions 7×7 and $\mathbf{a}_{3,1}$ is of dimension 4×4. However, in 2013:04, $\mathbf{a}_{1,1}$ retains its dimensionality, while $\mathbf{a}_{3,1}$ becomes of dimension 5×5. Consequently, the dimension of A'_m increases by one in the first month of the quarter, i.e., when $m = 1$.

3 Estimation

In general, the dimension of VARs becomes large quickly; in our case, the parameter proliferation is exacerbated because adding one more monthly predictor adds three variables to the VAR. In order to address the dimensionality issue, we estimate the VAR with Bayesian methods assuming a shrinkage prior. Bańbura et al. (2010) show that Bayesian VARs forecast with sufficient accuracy, even when the considered systems are fairly large (e.g., 100+ variables).

3.1 The Prior

Let $X_t = [Y'_{t-1} \dots Y'_{t-p}\ 1]'$ and $F_{k \times n} = [B_1 \dots B_p\ C]'$, where n is the number of endogenous variables in the VAR, p is the lag order of the VAR and $k = np + 1$ is the number of parameters in each VAR equation. A general form of the VAR in equation (1) can be re-written as

$$Y'_t A = X'_t F + \varepsilon'_t, \text{ for } t = 1, \dots, T. \tag{5}$$

Let a_i and f_i represent the respective ith columns of A and F, accounting for the ith structural equation in the VAR. We consider a modification of the Sims and Zha (1998) prior. The prior imposes independence across the structural equations and assumes that the vectors a_i and f_i are both jointly normally distributed. More specifically, the prior postulates the following distributions on the parameters. For $i = 1, ..., n$,

$$\begin{aligned} a_i &\sim N(0, \bar{S}), \\ f_i | a_i &\sim N(\bar{P} a_i, \bar{H}). \end{aligned} \tag{6}$$

Sims and Zha (1998) consider a random walk prior which centers the reduced-form parameters on the first autoregressive lag at unity. Since we have transformed the variables to stationary, we instead center the reduced-form autoregressive lag coefficients on zero. Accordingly, $\bar{P}$ is parameterized as a matrix of zeros. Furthermore, $\bar{S}$ is a $n \times n$ diagonal positive definite matrix, where the elements on the ith row are $\bar{S}_{ij} = \frac{\lambda_0^2}{\sigma_j^2}$, for $j = i = 1, \dots, n$, and zero otherwise. On the other hand, $\bar{H}$ is a $k \times k$ positive definite matrix, such that the $(k-1) \times (k-1)$ elements are defined as $diag([\bar{H}_1 \dots \bar{H}_p])$, where $diag(.)$ makes a (block) diagonal matrix and $\bar{H}_l$, $l = 1, \dots, p$, is an $n \times n$

diagonal matrix, where $\bar{H}_{l_{ij}} = \frac{\lambda_0^2\lambda_1^2}{\sigma_j^2 l^{2\lambda_3}}$, for $j = i = 1, \ ..., \ n$, and zero otherwise. The prior variance for the constant, i.e. the kth diagonal, is $\lambda_0^2\lambda_4^2$ and uncorrelated with the rest of the parameters.

For example, suppose $n = 2$ and $p = 2$. It follows from the discussion above that

$$\bar{P} = \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \end{bmatrix}';$$

$$\bar{S} = diag\left(\begin{bmatrix} \frac{\lambda_0^2}{\sigma_1^2} & \frac{\lambda_0^2}{\sigma_2^2} \end{bmatrix}\right),$$

and

$$\bar{H} = diag\left(\begin{bmatrix} \frac{\lambda_0^2\lambda_1^2}{\sigma_1^2} & \frac{\lambda_0^2\lambda_1^2}{\sigma_2^2} & \frac{\lambda_0^2\lambda_1^2}{\sigma_1^2 2^{2\lambda_3}} & \frac{\lambda_0^2\lambda_1^2}{\sigma_2^2 2^{2\lambda_3}} & \lambda_0^2\lambda_4^2 \end{bmatrix}\right).$$

A scaling factor, σ_j^2, is included to mitigate the effect of a difference in the unit of measure across the variables. In practice, σ_j^2 is proxied by the sample variance of the residuals that result from a univariate autoregression of order p for series j. The role of the hyperparameters λ_0, λ_1, λ_3 and λ_4 is presented in Table 2.[13]

INSERT TABLE 2 HERE

Hyperparameters can be selected by borrowing the values from, for example, Sims and Zha (1998) or Clark and Doh (2014). However, we mix the monthly and quarterly series together differently from previous studies. Moreover, as shown in Bańbura et al. (2010) the optimal level of shrinkage increases as the dimensions of the model increase relative to the sample size. Therefore, it is desirable to allow hyperparameter values to be selected in a systematic way. We do so by selecting a new set of hyperparameters for each (recursively expanding) sample used for estimation and, consequently, forecasting. The hyperparameter selection is the key, since the hope is to acquire accurate forecasts with sufficient shrinkage. Thus, we choose the hyperparameters "optimally" in order to maximize the marginal data density. The convenience in our case is that our model and the imposed restrictions results in an analytical solution for the marginal data density, which makes the hyperparameter selection computationally an easier task.

We compute the marginal data density to choose the hyperparameter values λ_0 and λ_1 (hence, their values are "tbd" in Table 2). We set $\lambda_3 = 1$, the same value used by Bańbura et al. (2010), and set $\lambda_4 = 1$, consistent with Sims and Zha (1998). In principle, we could also use the marginal data density to obtain the values for λ_3 and λ_4, but, given that we are conducting a recursive forecasting exercise, this would increase computation time. Moreover, according to Giannone et al.

[13] We adopt the notation of Sims and Zha (1998), who omit λ_2 which differentiates own lag contraction from cross-variable lag contraction. In principle, we could also allow for different shrinkage for monthly and quarterly variables. In this case, the marginal data density calculations are not trivial since the prior becomes assymetric. Whether or not the adjustment of the shrinkage would make a difference for forecasting is an empirical open question that we leave for future work.

(2014), the behavior of the constant is not important for short-run forecasting and $\lambda_3 = 1$ performs well in Bańbura et al. (2010).[14]

To make the out-of-sample forecasting exercise easy to conduct, an analytical solution for the marginal data density is preferred. As our temporal restrictions result in an underidentified system, a closed-form solution for the marginal data density exists and is presented below:[15]

$$p(\mathbf{y}|\mathbf{x}) \propto (2\pi)^{-\frac{n(t+n)}{2}} |\bar{H}|^{-\frac{n}{2}} |\bar{S}|^{-\frac{n}{2}} |(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})|^{-\frac{n}{2}} |\Delta|^{-\frac{t+1}{2}}, \tag{7}$$

where $\Delta = (\bar{S}^{-1} + \bar{P}'\bar{H}^{-1}\bar{P} + \mathbf{Y}'\mathbf{Y} - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi)$, $\Xi = (\mathbf{X}'\mathbf{Y} + \bar{H}^{-1\prime}\bar{P})$, $\mathbf{Y} \equiv [Y_1 \ ... \ Y_t]'$, $\mathbf{X} \equiv [X_1 \ X_2 \ ... \ X_t]'$, while $\mathbf{y} \equiv vec(\mathbf{Y})$ and $\mathbf{x} = \mathbf{I}_n \otimes \mathbf{X}$.

INSERT FIGURE 1 HERE

To choose the hyperparameters λ_0 and λ_1 we conduct a grid search over a grid [0.01 : 0.01 : 0.25] for each hyperparameter. The selected hyperparameter values and the corresponding value of the marginal data density for each iteration of estimation over the sample is depicted in Figure 1. As Panel A in the figure shows, shrinkage is much tighter at the beginning of the out-of-sample exercise, when the data sample is short. As Panel B shows, there are noticable jumps in the marginal data density around late 1990s, early 2000s and around 2009. As we discussed in the subsection 2.1, these are roughly the periods when our dateset improves (we add additional data series), and consequently our model size increases. Panel B certainly suggests that the variables we add to the dataset later in the sample are informative and improve the model fit.

3.2 The Sampler

The model is estimated via Gibbs sampler proposed in Waggoner and Zha (2003).[16] We need to estimate the VAR in equation (5) subject to constraints imposed by matrices Q_i and R_i. The restrictions are imposed such that for $i = 1, ..., n$:

$$\begin{aligned} Q_i a_i &= 0, \\ R_i f_i &= 0. \end{aligned} \tag{8}$$

[14] We estimated the model by choosing all hyperparameter values from the marginal data density. The results are not very sensitive to fixing the hyperparameter values of λ_3 and λ_4.

[15] Details are provided in Appendix A and a numerically more stable implementation is discussed in Appendix B.

[16] There are alternative ways to estimate a VAR postulated in equation (5). We prefer Waggoner and Zha (2003) approach as it provides with an efficient algorithm to estimate a VAR even when the contemporaneous correlation matrix A is overidentified, resulting in a restricted reduced form variance-covariance matrix for the VAR. Since we have modeled the within group releases of data as contemporaneously correlated, A is underidentified in our case. Nevertheless, we prefer to utilize a general framework that is not restrictive in terms of the form of temporal restrictions one can impose. If one imposes many restriction such that the system becomes overparameterized, then the marginal data density calculations in the Appendices A and B become approximations.

For our implementation Q_i is a $n \times n$ matrix parameterized to incorporate the temporal ordering restrictions highlighted in Section 2, while R_i is a $k \times k$ matrix of zeros reflecting the fact that there are no restrictions on the lag dynamics. We further define U_i and V_i to be matrices whose columns form an orthonormal basis for the null spaces of Q_i and R_i. We can then rewrite our parameter space in terms of $\psi_i = U_i f_i$ and $\phi_i = V_i g_i$, where the orthonormal rotation matrices U_i and V_i reduce the parameter space of the VAR, taking into account the linear restrictions on the contemporaneous and lagged dynamics (which is unrestricted in our case) of the system.

As Wagonner and Zha (2003) show, the prior in equation (6), the likelihood of the model in equation (5) and the restrictions in equation (8) imply marginal posterior distributions for ψ_i and ϕ_i represented by:

$$p(\psi_1, ..., \psi_m | \mathbf{X}, \mathbf{Y}) \propto |[U_1\psi_1, \ ..., \ U_m\psi_m]|^t exp\left(-\frac{t}{2}\sum_{i=1}^{n}\psi_i' S_i^{-1}\psi_i\right) \tag{9}$$

$$p(\phi_i | \psi_i, \mathbf{X}, \mathbf{Y}) = \varphi(P_i\psi_i, H_i), \tag{10}$$

where H, P, and S are transformations of the prior mean and variance matrices $\bar{H}$, $\bar{P}$, and $\bar{S}$ and $\varphi(.)$ denotes the pdf of a normal distribution.

The implied joint conditional posterior distribution of ψ_is is independent of ϕ_is, but non-standard. Waggoner and Zha (2003) propose a Gibbs sampler as a strategy to simulate from this joint distribution. Starting from some arbitrary initial values of $\{\psi_1^*, ..., \psi_{i-1}^*, \psi_i^*, \psi_{i+1}^*, ..., \psi_m^*\}$, Gibbs sampler sequentially draws from the conditional posteriors $p(\psi_i|\psi_1, ..., \psi_{i-1}, \psi_{i+1}, ..., \psi_m, \mathbf{X}, \mathbf{Y})$ as a mechanism to draw from the joint. In order not to be redundant and to conserve some space, we refer the reader to Waggoner and Zha (2003) for the details on how to draw from the conditional distributions. Given the simulated values of ψ_is, simulating the ϕ_is is fairly easy as they are drawn from a normal distribution.

As Waggoner and Zha (2003) suggest, the algorithm results in Monte Carlo chains that mix fairly well and converge fairly fast. Thus, our reported results are based on 5,000 accumulated.[17] Once we have the distribution of ϕ_is and ψ_is, the distributions of a_is and f_is, thus of A and F, could be easily recovered with the help of the orthonormal rotation matrices U_i and V_i.

4 Forecasting

To document the forecasting performance of our proposed model in real-time, we consider point and density forecasts, focusing on nowcast and four-quarter-ahead (one-year-ahead) forecasts. To

[17] The initial 100 draws are discarded to eliminate the effect of the initial values. We use all the draws to construct the posterior in order to obtain more efficient estimates. Since draws are correlated over time, we can, in principle, retain only a subset of draws for computational purposes.

simplify the discussion that follows, we first address some timing issues:

1. January, April, July, and October are the months for which we have observations for the previous quarter real GDP growth and all other variables in our dataset.

2. The model is only estimated at the end of the months of January, April, July, and October. The posteriors of the parameter estimates are stored to construct intra-quarter forecasts. The lag length is fixed at $p = 4$ for the entirety of the empirical implementation.

3. We compute two forecasts, Y_{t+1} and Y_{t+4}. During the first month of the quarter, the one-quarter-ahead forecast is effectively forecasting the advance release for first quarter real GDP growth. Similarly, Y_{t+4} is a one-year-ahead forecast.

4. The forecasts are updated as data become available within the quarter. For simplicity, we recompute the forecasts at the end of each month, when a full set of monthly predictors are available. We refer to the forecasts obtained at the end of the first month of the quarter (January, April, July, October) as "Month 1"; forecasts obtained at the end of the second month of the quarter (February, May, August, November) are labeled as "Month 2"; and the rest as "Month 3".

5. In order to understand the importance of the economic releases and to better match with the timing of professional forecasts, which are usually collected in the middle of the month, we also show how the forecasts get updated when a new group of data (based on the groupings in Section 2) become available.

6. All forecasts are evaluated against the advance release of the real GDP growth. For example, forecasts produced in January, February and March (and all the times in between) are evaluated against the advance release of the first quarter GDP in the April vintage of the data.

4.1 Point Forecasts

There is always a question of which measure of central tendency to consider when dealing with optimal forecasts. Gneiting (2012) argues that when the scoring function is in a form of a squared error, the optimal point forecast is the mean of the predictive distribution. Therefore, we construct point forecasts for all 5000 draws of A and F and report the mean of the forecast distributions as a point forecast. This is equivalent to reporting the conditional mean of Y_{t+h}, $h = 1, 4$, for each parameter draw given the information available at time t and immediately following each group of releases.[18]

[18] The point forecasts are constructed under the maintained assumption that $E(\varepsilon_t) = 0$. Alternatively, one could construct the measure using the mean of the simulated predictive distribution.

We provide the exact formulas for forecast construction after every monthly release; they can be easily adjusted for more granular forecasts. Let n_1, n_2 and n_3, be the number of variables released in Month 1, Month 2 and Month 3, respectively, where $n_1+n_2+n_3=n$. For each parameter draw, nowcasts are obtained as:

1. Time t (Month 1): At time t, $E(\varepsilon_{t+h}|\mathbf{X}_t, \mathbf{Y}_t)=0$ for all $h>0$ and hence:

$$\hat{Y}'_{t+1|t} = X'_t F A^{-1}. \tag{11}$$

2. Time $t+1/3$ (Month 2): Let $i_{1/3}=[\mathbf{1}_{n_2\times n_2}, \mathbf{0}_{n_2\times n_3}, \mathbf{0}_{n_2\times n_1}]'$.

   Define $\hat{\varepsilon}_{t+1/3}=[(i'_{1/3}Ai_{1/3})'(i'_{1/3}(Y_{t+1/3}-\hat{Y}_{t+1|t})), \mathbf{0}_{n_2\times n_3}, \mathbf{0}_{n_2\times n_1}]'$. $E(\varepsilon_{t+h}|\mathbf{X}_{t+1/3}, \mathbf{Y}_{t+1/3})=0$ for all $h>1/3$. However, when $h=1/3$, that is no longer the case. Because $Y_{t+1/3}$ is observed, we find that $E(\varepsilon_{t+1}|\mathbf{X}_{t+1/3}, \mathbf{Y}_{t+1/3})=\varepsilon_{t+1/3}$, which we proxy with $\hat{\varepsilon}_{t+1/3}$. Substituting this into the forecasting formula above gives us:

$$\hat{Y}'_{t+1|t+1/3} = \hat{Y}'_{t+1|t} + \hat{\varepsilon}'_{t+1/3}A^{-1}, \tag{12}$$

   which updates the forecast at time t with an add factor $\hat{\varepsilon}'_{t+1/3}A^{-1}$.

3. Time $t+2/3$ (Month 3): Let $i_{2/3}=[\mathbf{0}_{n_3\times n_2}, \mathbf{1}_{n_3\times n_3}, \mathbf{0}_{n_3\times n_1}]'$.

   Define $\hat{\varepsilon}_{t+2/3}=[\hat{\varepsilon}_{t+1/3}, (i'_{2/3}Ai_{2/3})'(i'_{1/3}(Y_{t+2/3}-\hat{Y}_{t+1|t+1/3})), \mathbf{0}_{n3\times n1})'$. With a logic similar to that in the previous step we get:

$$\hat{Y}'_{t+1|t+2/3} = \hat{Y}'_{t+1|t} + \hat{\varepsilon}'_{t+2/3}A^{-1}, \tag{13}$$

4. Time $t+1$ (Month 4): We go back to step 1, as Month 4 has the same structure as Month 1.

Now, in order to obtain forecasts for an arbitrary $h>1$ (in the case of the one-year-ahead forecasts, $h=4$), we iterate the forecasts forward. Thus, the forecasts at the end of each month can be calculated as $\hat{Y}'_{t+h|t}=\hat{X}'_{t+1|t}(FA^{-1})^{h-1}$; $\hat{Y}'_{t+h|t+1/3}=\hat{X}'_{t+1|t+1/3}(FA^{-1})^{h-1}$ and $\hat{Y}'_{t+h|t+2/3}=\hat{X}'_{t+1|t+2/3}(FA^{-1})^{h-1}$ and $\hat{X}_{t+1|t}=[\hat{Y}'_{t+1|t} \ ... \ Y'_{t-p-1} \ 1]'$, $\hat{X}_{t+1|t+1/3}=[\hat{Y}'_{t+1|t+1/3} \ ... \ Y'_{t-p-1} \ 1]'$ and $\hat{X}_{t+1|t+2/3}=[\hat{Y}'_{t+1|t+2/3} \ ... \ Y'_{t-p-1} \ 1]'$.[19]

[19] Admittedly, we can construct this observation-driven forecasting exercise a bit differently. For instance, we can estimate the reduced form representation of the VAR in equation (1) and construct the updating exercise with a release-based selection matrix that takes the relevant components in the reduced form variance-covariance matrix for updating. This exercise would involve estimating a Bayesian VAR using standard methods (perhaps even direct sampling). The difference between that approach and ours is that our estimation is consistent with the data release restrictions, and the restrictions are not merely imposed on the updating equations.

4.2 Predictive Distribution

In addition to point forecasts, we are also interested in the evaluation of density forecasts. The predictive density for $h = 1$ is defined as:

$$p(Y_{t+1}|\mathbf{X}_t, \mathbf{Y}_t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(Y_{t+1}|\mathbf{X}_t, \mathbf{Y}_t, F, A)p(F|A, \mathbf{X}_t, \mathbf{Y}_t)p(A|\mathbf{X}_t, \mathbf{Y}_t)dFdA, \qquad (14)$$

where $p(F|A, \mathbf{X}_t, \mathbf{Y}_t)$ and $p(A|\mathbf{X}_t, \mathbf{Y}_t)$ are the posterior distributions of the parameters F and A. In order to construct the predictive distributions, we follow the algorithm below:

1. Draw 500 observations of ϵ_t.
2. Given the posteriors of $p(F|A, \mathbf{X}_t, \mathbf{Y}_t)$, $p(A|\mathbf{X}_t, \mathbf{Y}_t)$, we construct the corresponding $\hat{Y}_{t+1}$ for the forecasting model in equation (5).
3. In total, we have 5000 parameter draws. Thus, for each observation of ϵ_t, we construct averages of the prediction $\hat{Y}_{t+1|t}$ across the draws.
4. Calculate percentiles for the distributions with 500 observations.
5. When calculating the predictive densities within the quarter, we simulate only the shocks that have not yet been realized. In other words, when computing density forecast at the end of month two, we simulate only month three shocks.

4.3 Evaluation

We use a few different evaluation criteria for our forecasting exercise.[20] First, we evaluate the unconditional predictive performance of our model (and the alternatives) using the root-mean-squared forecast error for an out-of-sample evaluation period $P = T - R - h$:

$$rMSE = \sqrt{\frac{1}{P}\sum_{s=1}^{P}(Y_{s+h} - \hat{Y}_{s+h|s})^2}, \; s = R, ..., T - h. \qquad (15)$$

Next, we compute the root-median-squared forecast error to obtain a performance measure. To compare our model to alternatives that are at time-smoothed consensus forecasts, we want an measure that is robust to outliers. The root-median-squared forecast error is:

$$rMedSE = \sqrt{median(Y_{s+h} - \hat{Y}_{s+h|s})^2}, \; s = R, ..., T - h. \qquad (16)$$

[20] We do not assess whether our model is statistically better than the alternatives since there are no readily available tests of conditional and unconditional equal predictive ability for recursively estimated Bayesian models in a real-time framework.

In addition, we consider conditional predictive ability tests by tracing measures in equations (15) and (16) over time with a fixed, rolling window of 10 years. To determine how the predictive densities change over time, we examine the evolution of the distance between the 97.5th and 2.5th percentiles over time.

5 Forecasting Results

We compare the forecasting performance of our model to some alternative time-series models and the Survey of Professional Forecasters' forecasts. For ease of exposition, we concentrate on the forecasting performance of four macroeconomic variables: employment growth ("EMP"), CPI inflation ("CPI"), the change in the Federal Funds Rate ("FFR") and real GDP growth ("GDP"); our emphasis throughout will be on the growth rate of real GDP.

INSERT FIGURE 2 HERE

First, Figure 2 documents the properties of the forecasts from our model in absolute sense. Panel A shows the nowcasts from our model at the end of Month 1, Month 2 and Month 3 together with the realization; Panel B shows the evolution of the distance between the 97.5 and 2.5 percentiles of the predictive distribution constructed as in Section 4.2. As we can see in Panel A, the nowcasts at the end of Month 1 generally replicate the fluctuations in the data but do not get the peaks: these forecasts are rather smooth. Month 2 predictions pick up the peaks, suggesting that the additional information helps. On the other hand, Month 3 information can magnify and, at times, misalign the timing of these peaks.

Panel B shows that there is considerable uncertainty in our nowcasts at the beginning of the out-of-sample period. Towards the end of the out-of-sample period, the distance between the 97.5 and 2.5 percentiles of the predictive distribution becomes more reasonable: around 5 percentage points. Moreover, it appears that within-quarter information makes the distribution more concentrated around the mean. However, there are no significant shifts in this distribution measures: it appears that the uncertainty this model characterizes is dominated by parameter estimation uncertainty and it does not capture the evolution of uncertainty that one might associate with the business cycle dynamics. This is in line with previous research, such as Clark (2011), which shows that allowing for stochastic volatility is very important for producing correctly calibrated density forecasts.

Quarterly Models

We first evaluate the predictive performance of our quarterly model and compare the forecasting performance of our model for $h = 1$ in Month 1 of each quarter to univariate autoregressive models and a simple VAR estimated with OLS. The intent of this exercise is to compare the predictive performance of our large VAR to that of parsimonious time series models, known to be competitive benchmarks for forecasting exercises.

The autoregressive models ("AR-Quarterly") are an AR(4) for the real GDP growth and an AR(12) for the monthly variables. We estimate monthly and quarterly ARs with data from the January, April, July and October vintages. For the monthly variables, we construct three-step-ahead forecasts (with no new information) and compare the quarterly average of the monthly forecast to the quarterly average of the monthly realization. We construct forecast errors based on the averages of three intra-quarter monthly forecasts and realizations to make them comparable with our Bayesian VAR ("BVAR-Quarterly"). For the BVAR, we obtain three direct, one-step-ahead forecasts per quarter (as the monthly variable appears in the blocking model three times over the quarter) and average them to obtain a quarterly value. For the single-frequency quarterly VAR ("VAR-Quarterly"), the model is a four-variable VAR(4) in employment growth, CPI inflation, changes in the Federal Funds Rate and real output growth as a forecasting model.[21] We average the monthly variables to a quarterly frequency by considering simple averages across the months in a quarter.

INSERT TABLE 3 HERE

In Table 3, we report the rMSE for the "AR-Quarterly" (in bold); for the other the models, we report the ratio of their respective rMSEs to that of the "AR-Quarterly". Ratios less than 1 (marked in italics) indicate improvements. As the table suggests, the performance of the small-scale VAR ("VAR-Quarterly") is inferior to the autoregressive model. For employment growth and CPI inflation, the performance of our "BVAR-Quarterly" is also inferior to that of the autoregressive benchmark. However, our quarterly BVAR performs slightly better (2 percent) than the parsimonious AR(4) for the real GDP growth. This is an encouraging result, since our estimated system is fairly large. For the changes in the FFR, our "BVAR-Quarterly" provides a 14 percent improvement over the AR.

Mixed Models

Table 4 reports the results from our mixed frequency BVAR and compares them to a mixed-frequency AR model. The BVAR results correspond to those discussed in Section 4 and are reported in the last three columns of the table as a ratio of rMSE-s relative to those from an autoregressive model. The results reported under the column labeled as "AR-Mixed" correspond to a monthly AR(12) for each reported variable. We estimate the AR(12) only when we have a balanced panel

[21] We considered an optimal lag selection mechanism with BIC. Using information criterium for lag selection improves the FFR forecast about 50%, but it does not change the rest of the rMSE-s much. For real output growth the rMSFE improves by about 0.05.

and use these estimates to update the forecasts as new monthly observations become available throughout the quarter. This is in contrast to the quarterly ARs in the previous section, where the iterated forecasts were obtained without using the interquarter information. For the GDP growth, AR-Quarterly and AR-Mixed are identical models. In the case of the mixed models, we are reporting one-step-ahead monthly forecasts and comparing them to one-step-ahead forecasts for these monthly variables obtained via direct forecasting as implied by our mixed BVAR framework.

INSERT TABLE 4 HERE

The nowcasting performance of the BVAR improves as new information becomes available over the course of a month. At the end of the second month of the quarter, the nowcasting performance improves by 15 percent relative to the previous month. However, the new information realized in third month of the quarter does not further improve the nowcasting performance of the model. When we look at the forecasting performance of our model for the monthly variables, the story is mixed. Our BVAR does better than the parsimonious autoregressive benchmarks for forecasting the Month 1 changes in the FFR and Month 3 employment growth. For the rest of the cases, it does no better than the benchmark. Moreover, for all variables except the FFR, the forecasting performance improves as new information becomes available, i.e., BVAR - M3 has a lower rMSE than BVAR - M2.

Two points are worth noting. First, the shrinkage employed in the estimation is optimized for a quarterly set-up. It is possible that one could further improve the forecasts of the monthly variables if alternatives are considered. Second, there is a tradeoff between the estimation risk that a variable introduces versus the information content it provides for improving the nowcasts as the quarter goes by. For instance, we have considered an alternative specification of a model where we do not incorporate the financials given that they are volatile and perhaps embed a lot of noise. In this specification, the BVAR-M1 forecasts improve, indicating a reduction in the estimation risk, however, the BVAR-M2 and BVAR-M3 forecasts deteriorate. This implies that the information in the financials is useful for intra-quarter updates.

Table 5 shows that the information within the quarter does not affect the four-quarter-ahead forecasts, consistent with the results in Schorfheide and Song (2015).

INSERT TABLE 5 HERE

An additional alternative we consider for forecasting the real GDP growth is by factors extracted from our large dataset. More specifically, we consider a model inspired by Giannone et al. (2008), where we extract factors from a monthly dataset and use these monthly factors to forecast the quarterly real GDP growth. For $t = 1, \ ..., \ T$ and $\tau = \{2/3, 1/3, 0\}$, the model is constructed as follows:

$$\mathcal{X}_{t-\tau} = \Lambda \mathcal{F}_{t-\tau} + \xi_{t-\tau} \tag{17}$$

$$\mathcal{F}_{t-\tau} = \Pi \mathcal{F}_{t-\tau} + u_{t-\tau} \tag{18}$$

$$\hat{\mathcal{Y}}_t = \alpha + \beta \hat{\mathcal{F}}_t, \tag{19}$$

where $\mathcal{X}_{t-\tau}$ is a $(n-1)/3 \times 1$ vector of monthly variables, $\mathcal{F}_{t-\tau}$ is a $r \times 1$ vector of common factors and $\mathcal{Y}_t$ is the real GDP growth. The parameters Λ, Π, α and β capture the parameters, while $\xi_{t-\tau}$ and $u_{t-\tau}$ are Gaussian white noise error terms. We estimate the number of factors $\hat{\mathcal{F}}_{t-\tau}$ via principal components, choosing the number of factors r such that they jointly explains at least 50 percent of the variation contained in the $(n-1)/3$ macroeconomic data series at any t.[22]

INSERT FIGURE 3 HERE

We estimate the model at the end of the Month 1 of each quarter (when there is a balanced panel). Our forecast of the quarterly growth rate of the real GDP is determined by its projection over the predicted value of the factor at the end of each quarter. In order to get the value of the extracted factor at the end of each quarter, we consider equation (18). Thus, at the end of $t-1$, we consider a three-step-ahead, iterated forecast of the factor to get to a time-t value. Accordingly, at times $t-\tau$, $\tau = 2/3, 1/3$, we consider two- and one-step-ahead forecasts, respectively. The difference between how we structure the factor model relative to the traditional approaches is that we still keep the data ordered based on the release calendar, not the reference calendar, to make the results from the various considered approaches comparable.

The ratio of the rMSE-s for Month 1, Month 2 and Month 3 over the full sample are 0.99, 0.81 and 0.63, respectively. Thus, our model on average performs as well as the factor model at the end of Month 1. However, it does better than the factor model at the end of Month 2 and 3. The improvement goes up to 37 percent in Month 3. The evolution of the real output growth forecasts from the factor model in equations (17-19) and the BVAR presented in Sections 2-4 are depicted in Figure 3. The figure shows the ratio of the rMSE-s with a rolling window of 10 years. The dates on the horizontal axes correspond to the middle of the 10 year rolling window. As Figure 3 indicates, at the end of Month 1 the performance of both models is comparable: the rMSE ratios hover around one. However, there is time-variation in the relative forecasting performance of the models in Month 2 and Month 3: our model performs better than the factor model at the beginning and at the end of the sample period. However, between 1996 and 2004, the nowcasts from the factor model dominate the ones we obtain from the BVAR.

[22] The choice of 50 percent has been motivated by parsimony. Alternatively, one could try using information criteria such as those proposed in Bai and Ng (2002), though the information criteria usually select larger models.

Survey of Professional Forecasters

The Survey of Professional Forecasters (SPF) provides forecasts of real GDP growth as well as a variety of economic fundamentals at the quarterly frequency. These include nowcasts as well as forecasts up to four quarters ahead. We use the forecasts of (annualized) quarterly real GNP/GDP growth rates. The series start in 1968:IV. We use the mean forecast across 34 professional forecasters starting 1985:I provided by the Federal Reserve Bank of Philadelphia. The choice of the mean is motivated with our loss function of squared forecast error, as well as with the intention to keep our model forecasts (based on the means of predictive distributions) comparable with that of the survey. We should also note that SPF is conducted with the advance release of the GDP as a target variable in mind, putting it on fair grounds with the forecasting exercise conducted in our paper.[23]

INSERT TABLE 6 HERE

The timing of the SPF is crucial since it would be important to match the timing of our forecasts with that of the SPF. Usually, the professional forecasters respond to the survey in between the second and third weeks of the month.[24] Consistent with the temporal restrictions discussed in Section 2, the forecasts obtained after either Group 1 or Group 2 data are released match with the SPF timing the best. Table 6 reports the rMSE's for nowcasts constructed multiple times throughout the quarter, after each groups' data (as in Table 1) has been released. Group 5 coincides with the end of the month. We also report the rMSE for the SPF's mean forecast for real output growth. As the table suggests, the nowcasting performance of our model improves as the first two groups of data in the second month of the quarter become available. After that, the marginal impact of each group on the nowcast is negligible. The rMSE of our BVAR is 1.44 after the Group 2 data in the second month is released. This is only about 7 percent worse than that of the SPF with the rMSE of 1.35.

INSERT FIGURE 4 HERE

Figure 4 sheds further light on the importance of the group releases for the performance of our BVAR relative to the SPF. We divide the figure into two parts: Panel A shows the smoothed rMSE-s calculated with a fixed rolling windows of ten years for the nowcasts obtained at the end

[23] Certainly, there are other competitive benchmarks one can use. For instance, Greenbook/Teelbook forecasts are the staff forecasts of the Federal Reserve Board and Blue Chip Economic Indicators (BCEI) are monthly professional forecasts that are sometimes considered as benchmarks. The Greenbook/Teelbook forecasts are available to the public with a five-year lag, while the BCEI are proprietary forecasts. Thus, we settle for SPF available at https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/rgdp. SPF is known to be a competitive benchmark as suggested by the results in Giannone et al. (2008).

[24] The exact deadlines for returning the surveys are available at http://www.philadelphiafed.org/research-anddata/real-time-center/survey-of-professional-forecasters/spf-release-dates.txt for the period starting from 1990 onwards.

of Month 1 and after the release of two group variables. As a reminder, the first group captured mainly soft survey data and the information from the Employment Situation Report, while group two had more diverse information such as industrial production, inflation and housing information as well as Philadelphia Fed's Business Outlook Survey. We show the behavior of the nowcasts after marginal releases of the 10 sequential groups in Panel B. In order to make the picture legible, we only plot the bounds corresponding to the point-wise maximum and minimum across the 10 various nowcasts. First, Panel B confirms the results of Table 6: there is not much action after the Group 2 in the Month 2 has been realized. As we can see from the figure, the bounds are narrow and they tightly replicate the behavior of the Group 2 nowcast presented in Panel A. Panel A, on the other hand, shows a few pieces of interesting information. First, our nowcasts are in general worse than that of the SPF even when we consider the nowcasting performance conditionally over time and not on average. There is a brief period in time between 1996 and 2002 that our model performs similar to the SPF, but this corresponds to a period where the SPF is particularly bad relative to other periods.

INSERT FIGURE 5 HERE

Figure 5 shows the nowcasts based on our BVAR relative to the SPF based on a rMedSE criterion. As suggested by the figure, if one uses a median, which is robust to outliers, rather than a mean, then the relative ranking of the nowcasts changes. In fact, Panel A shows that according to rMedSE criterion our nowcasting model has done better than the SPF for most of the out-of-sample, though the performance has declined recently since mid-2005. Panel B suggests more variability in the marginal impact of data releases for the data that becomes available after the mid-second month of the quarter.

INSERT FIGURE 6 HERE

Figure 6 shows the boxplot of the nowcasts from the BVAR after the Group 2 data has been released. The figure shows the median for forecast errors grouped by a rolling window of ten years worth of observations. In addition, it shows the 25th and 75th percentiles. Points are marked as outliers (by red +) if they are larger than Q3+1.5*(Q3-Q1) or smaller than Q1-1.5*(Q3-Q1), where Q1 and Q3 are the 25th and 75th percentiles, respectively. The default 1.5 corresponds to approximately 99.3 coverage if the data are normally distributed. The plotted whisker extends to the most extreme data value that is not an outlier. The figure documents the existence of outliers but also a fairly stable median value. There is only a slight trend in the median at the end of the out-of-sample period.

There is an important aspect of discussion here. From the figures, one could conclude that the predictive content of the variables might have changed. Though that could be the case. However, it is also true that our model structure changes over time. It is possible that once new variables become available in our dataset and they are added to the model, they diminish the importance of the variables that were previously included in the model.[25]

6 Conclusion

We present a stacked VAR as an alternative for high dimensional forecasting. We show that our VAR, estimated by Bayesian shrinkage, performs well for GDP nowcasting: it outperforms the considered time series models and does comparably well relative to the Survey of Professional Forecasters' forecasts. Nevertheless, there is considerable time variation in the forecasting performance.

References

[1] Aruoba, Borağan S.; Diebold, Francis X.; and Scotti, Chiara (2009). "Real-Time Measurement of Business Conditions." *Journal of Business and Economic Statistics,* 27(4), 417-427.

[2] Bańbura, Marta; Giannone, Domenico; and Reichlin, Lucrezia (2010). "Large Bayesian Vector Auto Regressions."*Journal of Applied Econometrics,* 25, 71-92.

[3] Bai, Jennie; Ghysels, Eric; and Wright, Jonathan (2013). "State Space Models and MIDAS Regressions." *Econometric Reviews,* 32(7), 779-813.

[4] Bai, Jushan; and Ng, Serena (2002). "Determining the Number of Factors in Approximate Factor Models," *Econometrica,* 70(1), 191-221.

[5] Brave, S., R. A. Butters, and A. Justiniano (2015). "Forecasting Economic Activity with Mixed Frequency Bayesian VARs," *mimeograph.*

[6] Blasques, F., Koopman, S.J., Mallee, M. and Z. Zhang (2016)."Weighted Maximum Likelihood for Dynamic Factor Analysis and Forecasting with Mixed Frequency Data," *Journal of Econometrics*, forthcoming.

[7] Clark, Todd, E. (2011). "Real-Time Density Forecasts from VARs with Stochastic Volatility." *Journal of Business and Economic Statistics* 29 (3), 327-341.

[8] Clark, Todd E.; and Doh, Taeyoung (2014). "Evaluating Alternative Models of Trend Inflation." *International Journal of Forecasting,* 30(3), 426-448.

[25] We refer the reader to Gilbert et al. (2015) for an interesting discussion on the changes in the marginal predictive content of the variables depending on their informational context, i.e. what other variables are available in the public domain when the new variable is released.

[9] De Mol, Christine; Giannone, Domenico; and Reichlin, Lucrezia (2008). "Forecasting Using a Large Number of Predictors: is Bayesian Shrinkage a Valid Alternative to Principal Components?" *Journal of Econometrics,* 146(2), 318-328.

[10] Foroni, Claudia; Guérin, Pierre; and Marcellino, Massimiliano (2015). "Markov-Switching Mixed-Frequency VAR Models." *International Journal of Forecasting,* 31(3), 692- 711.

[11] Giannone, Domenico; Lenza, Michele; and Primiceri, Giorgio E. (2015). "Prior Selection for Vector Autoregressions." *Review of Economics and Statistics,* 97(2), 412-435.

[12] Giannone, Domenico; Lenza, Michele; and Primiceri, Giorgio E. (2014). "Priors for the Long Run." *mimeo.*

[13] Giannone, Domenico; Reichlin, Lucrezia; and Small, David (2008). "Nowcasting: The Real-time Informational Content of Macroeconomic Data." *Journal of Monetary Economics,* 55(4), 665-676.

[14] Ghysels, Eric (2015). "Macroeconomics and the Reality of Mixed Frequency Data." *Journal of Econometrics*, forthcoming.

[15] Gilbert, Thomas; Scotti, Chiara; Strasser, Georg; and Vega, Clara (2015). "Is the Intrinsic Value of Macroeconomic News Announcements Related to their Asset Price Impact?" *mimeograph.*

[16] Ghysels, Eric; Santa-Clara, Pedro; and Valkanov, Rossen (2004). "The MIDAS Touch: Mixed Data Sampling Regressions." Discussion paper UNC and UCLA.

[17] Gneiting, Tilmann (2011). "Making and Evaluating Point Forecasts." *Journal of the American Statistical Association,* 106(494), 746-762.

[18] Koop, Gary (2003). *Bayesian Econometrics*, John Wiley.

[19] McBride, Bill (2014). "Ranking of Economic Data." *Calculated Risk*, April 27, web.

[20] Schorfheide, Frank; and Song, Dongho (2015). "Real-Time Forecasting with a Mixed-Frequency VAR." *Journal of Business and Economic Statistics* 33(3), 366-380.

[21] Sims, Christopher A.; and Zha, Tao (1998). "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review,* 39(4), 949-968.

[22] Waggoner, Daniel F.; and Zha, Tao (2003). "A Gibbs Sampler for Structural Vector Autoregressions." *Journal of Economic Dynamics and Control,* 28(2), 349-366.

A Derivation of the Marginal Data Density

Below we consider the derivation of the marginal data density for the model in equation (5) and the priors in equation (6). In the derivation below $\otimes$ represents Kronecker product. Consider a slightly rewritten version of the model for convenience:

$$A'Y_t = F'X_t + \epsilon_t, \tag{20}$$

where Y_t is an $n \times 1$ vector of endogenous variables at time t, X_t is a $k \times 1$ vector of predetermined and exogenous variables at time t, $X_t = [Y'_{t-1} \ ... \ Y'_{t-p} \ 1]'$, where $k = np + 1$. $\epsilon_t | X_t \sim N(\mathbf{0}, \mathbf{I}_n)$, and a_i and f_i are the respective ith columns of A and F.

We can write the reduced form as:

$$Y_t = (FA^{-1})'X_t + (A^{-1})'\epsilon_t. \tag{21}$$

Thus, $Y_t | X_t \sim N((FA^{-1})'X_t, ((A^{-1})'A^{-1}))$.

Let $\mathbf{Y} \equiv [Y_1 \ ... \ Y_t]'$, which is of a dimension of $t \times n$ and $\mathbf{X} \equiv [X_1 \ X_2 \ ... \ X_t]'$, which is of dimension $t \times k$, and $\epsilon \equiv [\epsilon_1 \ ... \ \epsilon_t]'$, which is of the same dimensionality as $\mathbf{Y}$. Let $\mathbf{y} \equiv vec(\mathbf{Y})$ and $\mathbf{e} \equiv vec(\epsilon A^{-1})$, where $vec(.)$ vectorizes the respective matrices making $(tn \times 1)$ vectors. Let $\mathbf{x} \equiv \mathbf{I}_n \otimes \mathbf{X}$, which is a $(tn \times kn)$ matrix. Consequently, $\mathbf{e} \sim N(0, ((A^{-1})'A^{-1}) \otimes \mathbf{I}_t)$. Let $\mathbf{f} = vec(FA^{-1}) = ((A^{-1})' \otimes \mathbf{I}_k)vec(F)$. Similarly, $\mathbf{e} \equiv vec(\epsilon A^{-1}) = ((A^{-1})' \otimes \mathbf{I}_t)vec(\epsilon)$.

The system can be rewritten as:

$$\mathbf{y} = \mathbf{xf} + \mathbf{e}. \tag{22}$$

The likelihood function is:[26]

$$\begin{aligned}
p(\mathbf{y}|\mathbf{x}, F, A) &= (2\pi)^{-\frac{tn}{2}} |((A^{-1})'A^{-1}) \otimes \mathbf{I}_t)^{-1}|^{\frac{1}{2}} exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{xf})'((A^{-1})'A^{-1}) \otimes \mathbf{I}_t)^{-1}(\mathbf{y} - \mathbf{xf})\right] \\
&= (2\pi)^{-\frac{tn}{2}} |AA' \otimes \mathbf{I}_t|^{\frac{1}{2}} exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{xf})'(AA' \otimes \mathbf{I}_t)(\mathbf{y} - \mathbf{xf})\right] \\
&= (2\pi)^{-\frac{tn}{2}} |AA'|^{\frac{t}{2}} exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{x}((A^{-1})' \otimes \mathbf{I}_k)vec(F))'(AA' \otimes \mathbf{I}_t)(\mathbf{y} - \mathbf{x}((A^{-1})' \otimes \mathbf{I}_k)vec(F))\right] \\
&= (2\pi)^{-\frac{tn}{2}} |AA'|^{\frac{t}{2}} exp\left[\begin{array}{c} -\frac{1}{2}(\mathbf{y} - (\mathbf{I}_n \otimes \mathbf{X})((A^{-1})' \otimes \mathbf{I}_k)vec(F))'(AA' \otimes \mathbf{I}_t) \\ (\mathbf{y} - (\mathbf{I}_n \otimes \mathbf{X})((A^{-1})' \otimes \mathbf{I}_k)vec(F)) \end{array}\right] \\
&= (2\pi)^{-\frac{tn}{2}} |AA'|^{\frac{t}{2}} exp\left[-\frac{1}{2}(\mathbf{y} - ((A^{-1})' \otimes \mathbf{X})vec(F))'(AA' \otimes \mathbf{I}_t)(\mathbf{y} - ((A^{-1})' \otimes \mathbf{X})vec(F))\right].
\end{aligned}$$

Since the prior imposes independence across the structural equations, we can re-write the priors

[26] In the derivation we use the property of the Kronecker product that $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

in equation (6) in the following vectorized notation:

$$vec(A) \sim N(0, \mathbf{I}_n \otimes \bar{S}),$$
$$vec(F|A) \sim N(vec(\bar{P}A), \mathbf{I}_n \otimes \bar{H}). \tag{23}$$

It follows that

$$p(A) = \prod_{i=1}^{n} p(a_i) = p(vec(A)) = (2\pi)^{-\frac{nn}{2}} |(\mathbf{I}_n \otimes \bar{S})^{-1}|^{\frac{1}{2}} exp\left[-\frac{1}{2} vec(A)'(\mathbf{I}_n \otimes \bar{S})^{-1} vec(A)\right].$$

With the property of the trace operator $tr(ABC) = vec(A')'(I \otimes B)vec(C)$, we can rewrite:

$$p(A) = (2\pi)^{-\frac{nn}{2}} |\bar{S}^{-1}|^{\frac{n}{2}} exp\left[-\frac{1}{2} tr(A'\bar{S}^{-1}A)\right] = (2\pi)^{-\frac{nn}{2}} |\bar{S}^{-1}|^{\frac{n}{2}} exp\left[-\frac{1}{2} tr(\bar{S}^{-1}AA')\right] \tag{24}$$

$$\begin{aligned} p(F|A) &= \prod_{i=1}^{n} p(f_i|a_i) = p(vec(F|A)) \\ &= (2\pi)^{-\frac{kn}{2}} |\bar{H}^{-1}|^{\frac{n}{2}} exp\left[-\frac{1}{2}(vec(F) - vec(\bar{P}A))'(\mathbf{I}_n \otimes \bar{H})^{-1}(vec(F) - vec(\bar{P}A))\right] \end{aligned} \tag{25}$$

Note that we derive the marginal data density for an arbitrary prior characterized by $\bar{P}$ as opposed to it being a matrix of zeros as it is in the case of our empirical application. In the case of zeros the expressions simplify.

The marginal likelihood or marginal data density is then given by

$$p(\mathbf{y}|\mathbf{x}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\mathbf{y}|\mathbf{x}, A, F) p(F|A) p(A) dF dA. \tag{26}$$

We can plug in the expressions of the respective probabilities to get:

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (2\pi)^{-\frac{tn}{2}} |AA'|^{\frac{t}{2}} exp\left[-\frac{1}{2}(\mathbf{y} - ((A^{-1})' \otimes \mathbf{X})vec(F))'(AA' \otimes \mathbf{I}_t)(\mathbf{y} - ((A^{-1})' \otimes \mathbf{X})vec(F))\right] \\ & \times (2\pi)^{-\frac{kn}{2}} |\bar{H}^{-1}|^{\frac{n}{2}} exp\left[-\frac{1}{2}(vec(F) - vec(\bar{P}A))'(\mathbf{I}_n \otimes \bar{H})^{-1}(vec(F) - vec(\bar{P}A))\right] \\ & \times (2\pi)^{-\frac{nn}{2}} |\bar{S}^{-1}|^{\frac{n}{2}} exp\left[-\frac{1}{2} tr(\bar{S}^{-1}AA')\right] dF dA \\ & = (2\pi)^{-\frac{n(t+k+n)}{2}} |\bar{H}^{-1}|^{\frac{n}{2}} |\bar{S}^{-1}|^{\frac{n}{2}} \int_{-\infty}^{\infty} |AA'|^{\frac{t}{2}} exp\left[-\frac{1}{2} tr(\bar{S}^{-1}AA')\right] && (27) \\ & \times \int_{-\infty}^{\infty} exp\left[-\frac{1}{2}(\mathbf{y} - ((A^{-1})' \otimes \mathbf{X})vec(F))'(AA' \otimes \mathbf{I}_t)(\mathbf{y} - ((A^{-1})' \otimes \mathbf{X})vec(F))\right] && (28) \\ & \times exp\left[-\frac{1}{2}(vec(F) - vec(\bar{P}A))'(\mathbf{I}_n \otimes \bar{H})^{-1}(vec(F) - vec(\bar{P}A))\right] dF dA && (29) \end{aligned}$$

The last integral could be simplified. We proceed by simplifying the expressions under the exponentials separately.

The first expression under the exponential could be written as:

$$
\begin{gathered}
(\mathbf{y}-((A^{-1})'\otimes\mathbf{X})vec(F))'(AA'\otimes\mathbf{I}_t)(\mathbf{y}-((A^{-1})'\otimes\mathbf{X})vec(F)) = \\
\mathbf{y}'(AA'\otimes\mathbf{I}_t)\mathbf{y}-2\mathbf{y}'(AA'\otimes\mathbf{I}_t)((A^{-1})'\otimes\mathbf{X})vec(F)+vec(F)'(A^{-1}\otimes\mathbf{X}')(AA'\otimes\mathbf{I}_t)((A^{-1})'\otimes\mathbf{X})vec(F) = \\
\mathbf{y}'(AA'\otimes\mathbf{I}_t)\mathbf{y}-2\mathbf{y}'(AA'(A^{-1})'\otimes\mathbf{I}_t\mathbf{X})vec(F)+vec(F)'(A^{-1}AA'(A^{-1})'\otimes\mathbf{X}'\mathbf{I}_t\mathbf{X})vec(F) = \\
\mathbf{y}'(AA'\otimes\mathbf{I}_t)\mathbf{y}-2\mathbf{y}'(A\otimes\mathbf{X})vec(F)+vec(F)'(\mathbf{I}_n\otimes\mathbf{X}'\mathbf{X})vec(F) = \\
tr(\mathbf{Y}'\mathbf{Y}AA')-2\mathbf{y}'(A\otimes\mathbf{X})vec(F)+vec(F)'(\mathbf{I}_n\otimes\mathbf{X}'\mathbf{X})vec(F)
\end{gathered}
$$

The second expression under the exponential could be written as:

$$
\begin{gathered}
(vec(F)-vec(\bar{P}A))'(\mathbf{I}_n\otimes\bar{H})^{-1}(vec(F)-vec(\bar{P}A)) = \\
vec(F)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)-2vec(\bar{P}A)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)+vec(\bar{P}A)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(\bar{P}A) = \\
vec(F)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)-2vec(\bar{P}A)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)+tr(A'\bar{P}'\bar{H}^{-1}\bar{P}A) = \\
vec(F)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)-2vec(\bar{P}A)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)+tr(\bar{P}'\bar{H}^{-1}\bar{P}AA')
\end{gathered}
$$

The marginal data density becomes:

$$
\begin{aligned}
p(\mathbf{y}|\mathbf{x}) = (2\pi)^{-\frac{n(t+k+n)}{2}}|\bar{H}^{-1}|^{\frac{n}{2}}|\bar{S}^{-1}|^{\frac{n}{2}}\int_{-\infty}^{\infty}|AA'|^{\frac{t}{2}}exp\left[-\frac{1}{2}tr(\bar{S}^{-1}AA'+\bar{P}'\bar{H}^{-1}\bar{P}AA'+\mathbf{Y}'\mathbf{Y}AA')\right] \\
\times\int_{-\infty}^{\infty}exp\left[-\frac{1}{2}(vec(F)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)-2vec(\bar{P}A)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F))\right] \\
\times\, exp\left[-\frac{1}{2}(vec(F)'(\mathbf{I}_n\otimes\mathbf{X}'\mathbf{X})vec(F)-2\mathbf{y}'(A\otimes\mathbf{X})vec(F))\right]dFdA
\end{aligned}
\tag{30}
$$

We can rewrite the last integral as:

$$
\begin{aligned}
&\int_{-\infty}^{\infty}exp\left[\begin{array}{c}-\frac{1}{2}(vec(F)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)+vec(F)'(\mathbf{I}_n\otimes\mathbf{X}'\mathbf{X})vec(F))+\\ vec(\bar{P}A)'(\mathbf{I}_n\otimes\bar{H})^{-1}vec(F)+\mathbf{y}'(A\otimes\mathbf{X})vec(F)\end{array}\right]dF = \\
&\int_{-\infty}^{\infty}exp\left[-\frac{1}{2}(vec(F)'(\mathbf{I}_n\otimes\bar{H}^{-1}+\mathbf{I}_n\otimes\mathbf{X}'\mathbf{X})vec(F))+(vec(\bar{P})'(A\otimes\bar{H}^{-1})+\mathbf{y}'(A\otimes\mathbf{X}))vec(F)\right]dF
\end{aligned}
$$

Let $\Sigma=\mathbf{I}_n\otimes\bar{H}^{-1}+\mathbf{I}_n\otimes\mathbf{X}'\mathbf{X}=\mathbf{I}_n\otimes(\mathbf{X}'\mathbf{X}+\bar{H}^{-1})$ and $\lambda=vec(\bar{P})'(A\otimes\bar{H}^{-1})+\mathbf{y}'(A\otimes\mathbf{X})$, where λ is a $1\times kn$ vector and Σ is a $kn\times kn$ matrix.

The integral could be written in the following form:

$$
\int_{-\infty}^{\infty}exp\left(-\frac{1}{2}vec(F)'\Sigma vec(F)+\lambda vec(F)\right)dF
\tag{31}
$$

This integral is a Gaussian integral and is equal to $(2\pi)^{\frac{kn}{2}}|\Sigma|^{-1/2}exp\left(\frac{1}{2}\lambda\Sigma^{-1}\lambda'\right)$

Let us simplify some relevant expressions that enter this integral. First,[27]

$$|\Sigma| = |\mathbf{I}_n \otimes (\mathbf{X}'\mathbf{X} + \bar{H}^{-1})| = |\mathbf{I}_n|^k |\mathbf{X}'\mathbf{X} + \bar{H}^{-1}|^n = |\mathbf{X}'\mathbf{X} + \bar{H}^{-1}|^n$$

Further, let $\Xi = (\mathbf{X}'\mathbf{Y} + \bar{H}^{-1\prime}\bar{P})$. We can rewrite λ as:[28]

$$\lambda = vec(\bar{P})'(A \otimes \bar{H}^{-1}) + \mathbf{y}'(A \otimes \mathbf{X}) = vec(\bar{H}^{-1\prime}\bar{P}A)' + vec(\mathbf{X}'\mathbf{Y}A)' = vec(\mathbf{X}'\mathbf{Y}A + \bar{H}^{-1\prime}\bar{P}A)'$$
$$= vec((\mathbf{X}'\mathbf{Y} + \bar{H}^{-1\prime}\bar{P})A)' = vec(\Xi A)'$$

Then,

$$\lambda\Sigma^{-1}\lambda' = vec(\Xi A)'\Sigma^{-1}vec(\Xi A) = tr(A'\Xi'\Sigma^{-1}\Xi A) = tr(\Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi AA').$$

Thus, the Gaussian integral becomes:

$$(2\pi)^{\frac{kn}{2}}|\Sigma|^{-1/2}exp\left(\frac{1}{2}\lambda\Sigma^{-1}\lambda'\right) = (2\pi)^{\frac{kn}{2}}|\mathbf{X}'\mathbf{X} + \bar{H}^{-1}|^{-\frac{n}{2}}exp\left(\frac{1}{2}tr(\Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi AA')\right)$$

Let $\Delta = (\bar{S}^{-1} + \bar{P}'\bar{H}^{-1}\bar{P} + \mathbf{Y}'\mathbf{Y} - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi)$.

The marginal data density simplifies to:

$$p(\mathbf{y}|\mathbf{x}) = (2\pi)^{-\frac{n(t+k+n)}{2}}|\bar{H}^{-1}|^{\frac{n}{2}}|\bar{S}^{-1}|^{\frac{n}{2}}\int_{-\infty}^{\infty}|AA'|^{\frac{t}{2}}exp\left[-\frac{1}{2}tr(\bar{S}^{-1}AA' + \bar{P}'\bar{H}^{-1}\bar{P}AA' + \mathbf{Y}'\mathbf{Y}AA')\right]$$
$$\times (2\pi)^{\frac{kn}{2}}|\mathbf{X}'\mathbf{X} + \bar{H}^{-1}|^{-\frac{n}{2}}exp\left(\frac{1}{2}tr(\Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi AA')\right)dA$$
$$= (2\pi)^{-\frac{n(t+n)}{2}}|\bar{H}^{-1}|^{\frac{n}{2}}|\bar{S}^{-1}|^{\frac{n}{2}}|\mathbf{X}'\mathbf{X} + \bar{H}^{-1}|^{-\frac{n}{2}}$$
$$\times \int_{-\infty}^{\infty}|AA'|^{\frac{t}{2}}exp\left[-\frac{1}{2}tr(\bar{S}^{-1}AA' + \bar{P}'\bar{H}^{-1}\bar{P}AA' + \mathbf{Y}'\mathbf{Y}AA' - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi AA')\right]dA$$
$$= (2\pi)^{-\frac{n(t+n)}{2}}|\bar{H}^{-1}|^{\frac{n}{2}}|\bar{S}^{-1}|^{\frac{n}{2}}|(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})|^{-\frac{n}{2}}\int_{-\infty}^{\infty}|AA'|^{\frac{t}{2}}exp\left[-\frac{1}{2}tr(\Delta AA')\right]dA$$

We can do a change of a variable and write the integral above in terms of dAA'. In order to do so, we need to figure out the Jacobian matrix, which $|dAA'/dA|$.

We can write

$$vec(AA') = (A \otimes \mathbf{I}_n)vec(A) = (\mathbf{I}_n \otimes A)vec(A') = (\mathbf{I}_n \otimes A)\mathbf{T}_{nn}vec(A),$$

where $\mathbf{T}_{nn}$ is a $nn \times nn$ permutation matrix such that $\mathbf{T}_{nn}vec(A) = vec(A')$ and $(A \otimes B)\mathbf{T}_{nn} = \mathbf{T}_{nn}(B \otimes A)$.

[27] If A is a $n \times n$ matrix and B is a $m \times m$ matrix. Then, $|A \otimes B| = |A|^m|B|^n$.

[28] We use the property $vec(ABC) = (C' \otimes A)vec(B)$.

The derivative of dAA'/dA by the product rule is:

$$dAA'/dA = (A \otimes \mathbf{I}_n) + (\mathbf{I}_n \otimes A)\mathbf{T}_{nn} = (A \otimes \mathbf{I}_n) + \mathbf{T}_{nn}(A \otimes \mathbf{I}_n) = (A \otimes \mathbf{I}_n)(\mathbf{I}_{nn} + \mathbf{T}_{nn})$$

The Jacobian is:

$$|(A \otimes \mathbf{I}_n)(\mathbf{I}_{nn} + \mathbf{T}_{nn})| = |(A \otimes \mathbf{I}_n)||(\mathbf{I}_{nn} + \mathbf{T}_{nn})| = |A|^n|(\mathbf{I}_{nn} + \mathbf{T}_{nn})|$$

Thus, the last integral in the marginal data density calculation becomes:

$$\begin{aligned}
&\int_{-\infty}^{\infty} |AA'|^{\frac{t}{2}} exp\left[-\frac{1}{2}tr(\Delta AA')\right] dA = \int_{-\infty}^{\infty} |A|^t exp\left[-\frac{1}{2}tr(\Delta AA')\right] dA \\
&= |(\mathbf{I}_{nn} + \mathbf{T}_{nn})|^{-1} \int_{-\infty}^{\infty} |A|^{t-n} exp\left[-\frac{1}{2}tr(\Delta AA')\right] dAA' \\
&= |(\mathbf{I}_{nn} + \mathbf{T}_{nn})|^{-1} \int_{-\infty}^{\infty} |AA'|^{\frac{t-n}{2}} exp\left[-\frac{1}{2}tr(\Delta AA')\right] dAA'. \qquad (32)
\end{aligned}$$

Suppose A is such that $H = AA'$ is a symmetric, nonnegative-definite matrix-valued random variable. The integral above then has the kernel of a Wishart distribution, where $H = AA'$ is the random variable. Thus, $H \sim W(\nu, \Delta^{-1})$ (see Definition B.27 in Koop, 2003), where $\nu = t + 1$ is the degrees of freedom and $\Delta^{-1})$ is a positive definite matrix.

The constant adjustment we need is:

$$c_w = 2^{\frac{(t+1)n}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{t+2-i}{2}\right)$$

Thus, the marginal data density can be rewritten as:

$$p(\mathbf{y}|\mathbf{x}) = (2\pi)^{-\frac{n(t+n)}{2}} c_w |(\mathbf{I}_{nn} + \mathbf{T}_{nn})^{-1}||\bar{H}|^{-\frac{n}{2}} |\bar{S}|^{-\frac{n}{2}} |(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})|^{-\frac{n}{2}} |\Delta|^{-\frac{t+1}{2}}, \qquad (33)$$

where $\Delta = (\bar{S}^{-1} + \bar{P}'\bar{H}^{-1}\bar{P} + \mathbf{Y}'\mathbf{Y} - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi)$ and $\Xi = (\mathbf{X}'\mathbf{Y} + \bar{H}^{-1\prime}\bar{P})$.

B Numerical Considerations for Marginal Data Density

As in Giannone et al. (2015) we work with a modified version of the marginal data density for numerical stability purposes. We consider re-writing $|(\mathbf{X}'\mathbf{X}+\bar{H}^{-1})|^{-\frac{n}{2}}$ and Δ in the following ways.

Let C_H be the Choleski factor of $\bar{H}$, such that $C_H C_H' = \bar{H}$. Then, we can rewrite:

$$
\begin{aligned}
&|(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})|^{-\frac{n}{2}} = |\bar{H}|^{\frac{n}{2}} |C_H C_H'|^{-\frac{n}{2}} |(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})|^{-\frac{n}{2}} \\
&= |\bar{H}|^{\frac{n}{2}} |C_H|^{-\frac{n}{2}} |C_H'|^{-\frac{n}{2}} |(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})|^{-\frac{n}{2}} = |\bar{H}|^{\frac{n}{2}} |C_H|^{-\frac{n}{2}} (\mathbf{X}'\mathbf{X} + \bar{H}^{-1})|^{-\frac{n}{2}} |C_H'|^{-\frac{n}{2}} \\
&= |\bar{H}|^{\frac{n}{2}} |C_H (\mathbf{X}'\mathbf{X} + \bar{H}^{-1}) C_H'|^{-\frac{n}{2}} = |\bar{H}|^{\frac{n}{2}} |(C_H \mathbf{X}'\mathbf{X} C_H' + C_H \bar{H}^{-1} C_H')|^{-\frac{n}{2}} \\
&= |\bar{H}|^{\frac{n}{2}} |(C_H \mathbf{X}'\mathbf{X} C_H' + \mathbf{I}_k)|^{-\frac{n}{2}},
\end{aligned}
$$

where the last passage is true due to the fact that $\bar{H}$ is a diagonal matrix.

Similarly, let C_Δ be the Choleski factor of $\bar{S}$, such that $C_\Delta C_\Delta{}' = \bar{S}$:

$$
\begin{aligned}
|\Delta|^{-\frac{t+1}{2}} &= |\bar{S}^{-1} + \bar{P}'\bar{H}^{-1}\bar{P} + \mathbf{Y}'\mathbf{Y} - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi|^{-\frac{t+1}{2}} \\
&= |\bar{S}|^{\frac{t+1}{2}} |\mathbf{I}_n + C_\Delta(\bar{P}'\bar{H}^{-1}\bar{P} + \mathbf{Y}'\mathbf{Y} - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi)C_\Delta{}'|^{-\frac{t+1}{2}}
\end{aligned}
$$

This alternative reformulations of the determinants are more stable since the original expression is decomposed into a component directly coming from a prior, i.e. $\bar{H}$ and $\bar{S}$. Note only the researcher has a direct control over the specification of this prior, but also these are diagonal matrices for the form of the prior we consider. Thus, its inversion is very easy. The second determinant created in either of the decompositions involves a sum or two matrices, one of them being the identity matrix. The identify matrix assures that there is at least one positive definite matrix in the sum, thus yielding more numerical stability in the calculation of the determinant in large datasets. Thus,

$$
\begin{aligned}
p(\mathbf{y}|\mathbf{x}) = (2\pi)^{-\frac{n(t+n)}{2}} c_w |(\mathbf{I}_{nn} + \mathbf{T}_{nn})^{-1}|\bar{S}|^{-\frac{n-(t+1)}{2}} |(C_H \mathbf{X}'\mathbf{X} C_H' + \mathbf{I}_k)|^{-\frac{n}{2}} \\
\times |\mathbf{I}_n + C_\Delta(\bar{P}'\bar{H}^{-1}\bar{P} + \mathbf{Y}'\mathbf{Y} - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi)C_\Delta{}'|^{-\frac{t+1}{2}}.
\end{aligned} \tag{34}
$$

Instead of the marginal data density, we grid search the natural logarithmic transformation of it written in the following form:

$$
\begin{aligned}
ln(p(\mathbf{y}|\mathbf{x})) \propto -\frac{n(t+n)}{2} ln(2\pi) + ln(c_w) - \frac{n-(t+1)}{2} ln(abs(|\bar{S}|)) - \frac{n}{2} ln(abs(|(C_H \mathbf{X}'\mathbf{X} C_H' + \mathbf{I}_k)|)) \\
\times -\frac{t+1}{2} ln(abs(|\mathbf{I}_n + C_\Delta(\bar{P}'\bar{H}^{-1}\bar{P} + \mathbf{Y}'\mathbf{Y} - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi)C_\Delta{}'|)),
\end{aligned} \tag{35}
$$

where $abs(.)$ takes the absolute value of the arguments, in this case the determinants, of interest.

The set of hyperparameter values that make the determinant $|\mathbf{I}_n + C_\Delta(\bar{P}'\bar{H}^{-1}\bar{P} + \mathbf{Y}'\mathbf{Y} - \Xi'(\mathbf{X}'\mathbf{X} + \bar{H}^{-1})^{-1}\Xi)C_\Delta{}'|$ zero are not permissible. Thus, we incorporate a check in the grid search and assign $-\infty$ to the log marginal data density value to make sure that the set of hyperparameters that generate a zero determinant are not selected when we pick the maximum of the marginal data density. Note that similar checks for the other determinants under the logarithm are not necessary since they come directly from priors or involve sum of matrices where at least one of them is a positive definite matrix. Thus, the positive domain condition is satisfied for those by construction.

C Tables

Table 1: Description of Data Series

Name	Freq.	Transf.	Release Date	Publ. Lag	Vintage	Report
		Group 1				
ISM Manufacturing PMI (s.a.)	M	level	1st business day	one month	1997:03	ISM
ISM Supplier Deliveries Index (s.a.)	M	level	1st business day	one month	2009:11	ISM
ISM New Orders Index (s.a)	M	level	1st business day	one month	2009:11	ISM
Initial Unempl Insurance Claims (s.a.)	W	Δ ln	Thursdays, weekly	one week	2009:06	DL
Civilian Unempl. Rate, 16+ (s.a.)	M	Δ	1st Friday	one month	1980:01	ES
Empl.: Total Nonfarm Payrolls (s.a.)	M	Δ ln	1st Friday	one month	1980:01	ES
Avg Weekly Manufacturing Hours (s.a.)	M	level	1st Friday	one month	1980:01	ES
		Group 2				
Industrial Production (s.a.)	M	Δ ln	after 2 weeks	one month	1980:01	Board
Retail Sales (s.a)	M	Δ ln	after 2 weeks	one month	1980:01	DC
New Res. Construction - Housing Starts (s.a)	M	Δ ln	12th workday	one month	1980:01	DC
Philadelphia Fed Business Outlook Survey (s.a.)*	M	level	3rd Thurs	curr. month	1980:01	PF †
CPI Headline (s.a.)	M	Δ ln	varying, mid-month	one month	1980:01	DL
CPI Core (s.a.)	M	Δ ln	varying, mid-month	one month	1996:12	DL
New (1-Family) Houses Sold (s.a.)	M	Δ ln	17th workday	one month	1999:07	DC
		Group 3				
Consumer Sentiment Index (n.s.a)*	M	Δ	last Friday	curr. month	1980:01	UM†
GDP AdvanceEstimate (s.a.)*	Q	Δ ln	last week of month	one month	1980:01	DC
PCE Headline (s.a.)	M	Δ ln	day after GDP	one month	2000:07	DC
PCE Core (s.a.)	M	Δ ln	day after GDP	one month	2000:07	DC
Personal Income (s.a)	M	Δ ln	day after GDP	one month	1980:01	DC
		Group 4				
Federal Funds (Effective) Rate*	D	Δ	last day	curr. month	1980:01	Board†
Term Spread (10-Year Tr. Note - 3-Month Tr. Bill)*	D	level	last day	curr. month	1980:01	Board†
WTI Oil Price*	D	Δ ln	last day	curr. month	1980:01	WSJ†
S&P 500 Stock Index*	D	Δ ln	last day	curr. month	1980:01	WSJ†
Credit Spread (Moodies Seasoned Baa - Aaa)*	D	level	last day	curr. month	1980:01	Board†
Nom. Trade Weighted Exch. Rate vs Major Curr*	M	Δ ln	last day	curr. month	1980:01	Board†

Notes: Reports are abbreviated as follows: Board - Federal Reserve Board of Governors, DC - Department of Commerce, DL - Department of Labor, ES - Employment Situation, ISM - ISM Report on Business-Manufacturing, PF - Philadelphia Fed, UM - University of Michigan, WSJ - Wall Street Journal. (Q) indicates quarterly frequency for the data, while (M), (W) and (D) stand for monthly, weekly and daily frequencies, respectively. "*" next to the name of the data series indicates that these variables do not get revised. "s.a." indicates seasonally adjusted series: if not marked, then series are seasonally unadjusted. The data series are collected as vintage series from ALFRED (ArchivaL Federal Reserve Economic Data). The data marked with † is collected from Haver Analytics.

Table 2: Description of Hyperparameters

λ_0	tbd	controls the overall tightness of the beliefs
λ_1	tbd	tightens the prior around the mean
λ_3	1	rate of contraction with an increase in lag length
λ_4	1	controls the tightness of the constant

Table 3: Evaluation of Quarterly Models

	EMP	CPI	FFR	GDP
AR-Quarterly	**0.33**	**0.94**	**0.78**	**1.78**
VAR-Quarterly	1.24	1.05	1.72	1.26
BVAR-Quarterly	1.41	1.14	*0.86*	*0.98*

Notes: We report the out-of-sample rMSE for the AR-Quarterly model (in bold), while for the VAR-Quarterly and BVAR-Quarterly we report the rMSEs relative to that of the AR. Numbers in italics indicate improvements.

Table 4: Evaluation of Mixed Model

	AR-Mixed	BVAR - M1	BVAR - M2	BVAR - M3
EMP-M1	**0.45**	1.19		
CPI-M1	**1.07**	1.12		
FFR-M1	**0.85**	*0.99*		
EMP-M2	**0.48**	1.27	1.15	
CPI-M2	**1.14**	1.23	1.04	
FFR-M2	**0.83**	1.04	1.55	
EMP-M3	**0.53**	1.07	*0.94*	*0.94*
CPI-M3	**1.03**	1.25	1.20	1.01
FFR-M3	**0.70**	1.11	1.30	1.28
GDP	**1.78**	*0.98*	*0.83*	*0.84*

Notes: We report the out-of-sample rMSEs for the AR model (in bold), while for the mixed-BVAR we report the rMSE relative to that of the AR. "BVAR - M1," "BVAR - M2" and "BVAR - M3" correspond to forecasts obtained at the end of Month 1, Month 2 and Month 3 in the quarter.

Table 5: Evaluation of BVAR for $h = 4$ Forecasts

	BVAR - M1	BVAR - M2	BVAR - M3
EMP-M1	0.69	0.69	0.68
CPI-M1	1.31	1.30	1.29
FFR-M1	0.86	0.85	0.85
EMP-M2	0.70	0.70	0.70
CPI-M2	1.46	1.45	1.45
FFR-M2	0.91	0.90	0.90
EMP-M3	0.69	0.69	0.69
CPI-M3	1.34	1.33	1.33
FFR-M3	0.77	0.77	0.77
GDP	2.06	2.08	2.07

Notes: We report the out-of-sample rMSEs for the "BVAR - M1," "BVAR - M2"and "BVAR - M3".

Table 6: Output Growth Nowcasts

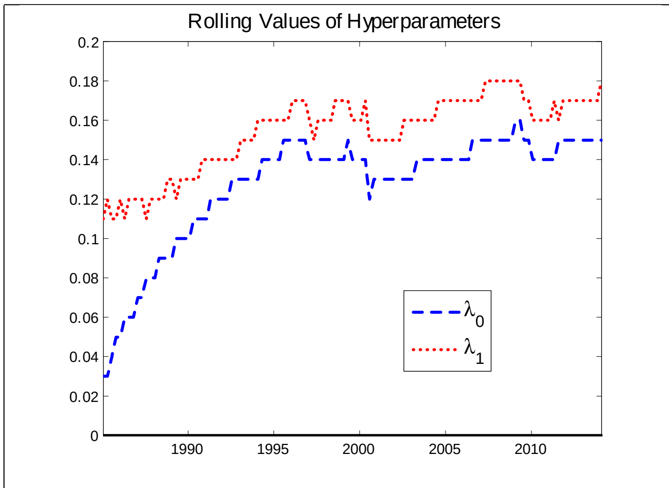
	BVAR	SPF
End of Month 1	1.75	
Group 1	1.61	
Group 2	1.44	1.35
Group 3	1.46	
End of Month 2	1.47	
Group 1	1.45	
Group 2	1.48	
Group 3	1.47	
End of Month 3	1.50	
Group 1	1.47	
Group 2	1.51	

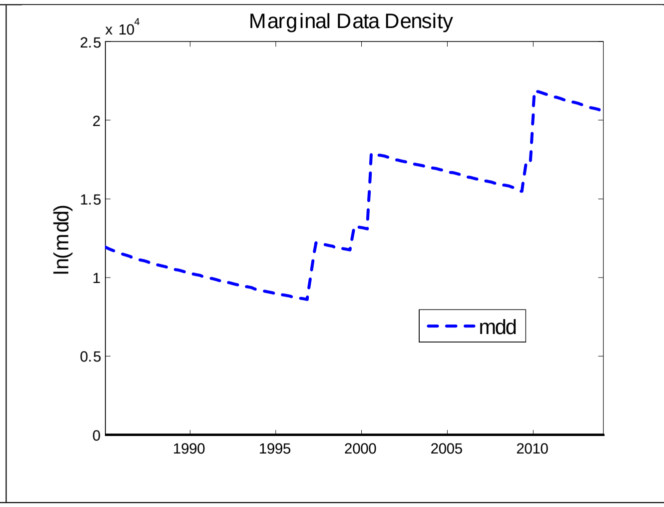
Notes: We report the out-of-sample rMSEs for the mixed frequency nowcasts of the BVAR after each group's data (as in Table 1) has been released. We also report the rMSE for the SPF's mean forecast.

D Figures

Figure 1: BVAR Hyperparameter Selection

Panel A: Hyperparameter values	Panel B: Value of Marginal Data Density

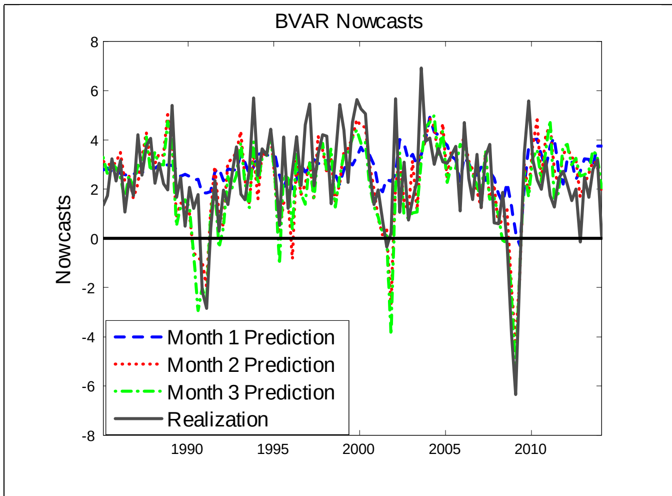


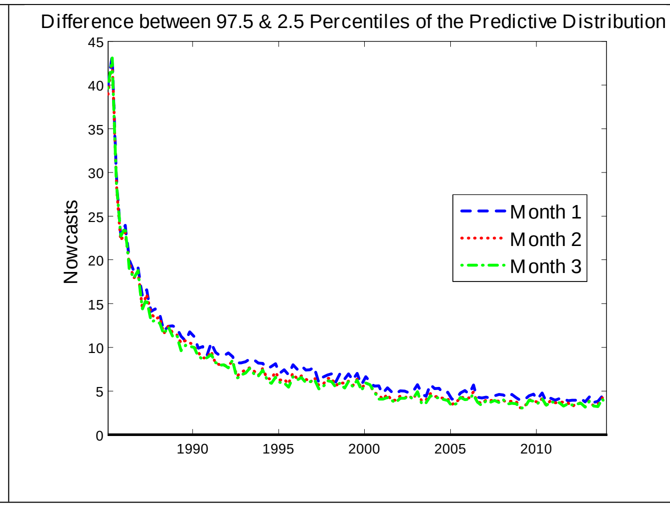


Note: Panel A depicts the hyperparameter values selected via grid search over grids [0.01:0.01:0.25], while Panel B shows the corresponding value of the marginal data density.

Figure 2: BVAR Nowcasts

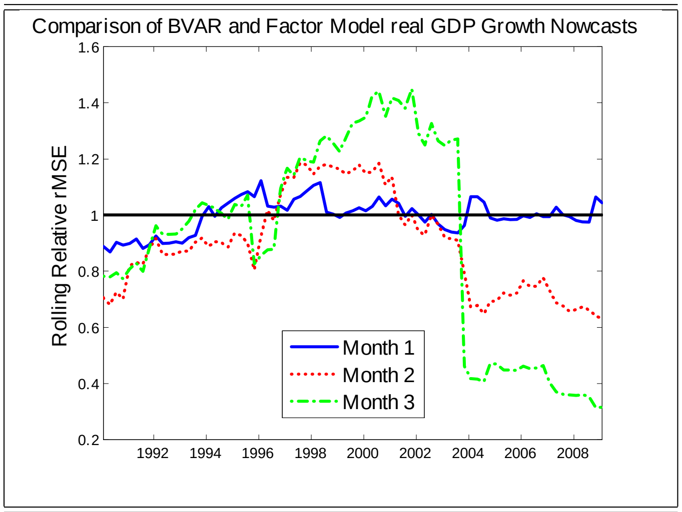
Panel A: Point Nowcasts	Panel B: Nowcast Distributions





Note: Panel A depicts the time series of nowcasts, while Panel B shows the evolution of the distance between 97.5 and 2.5 percentiles of the predictive distribution over time.

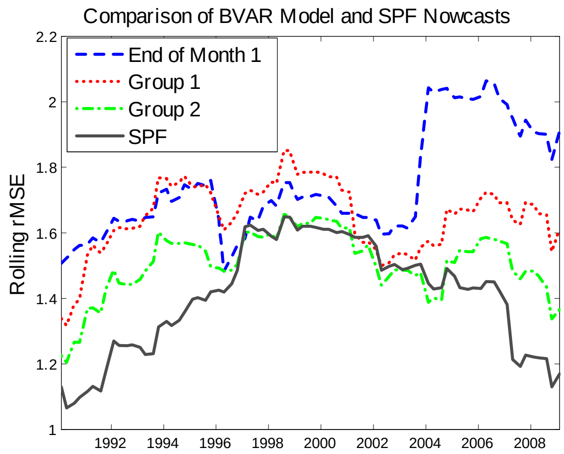
Figure 3: BVAR versus Factor Model

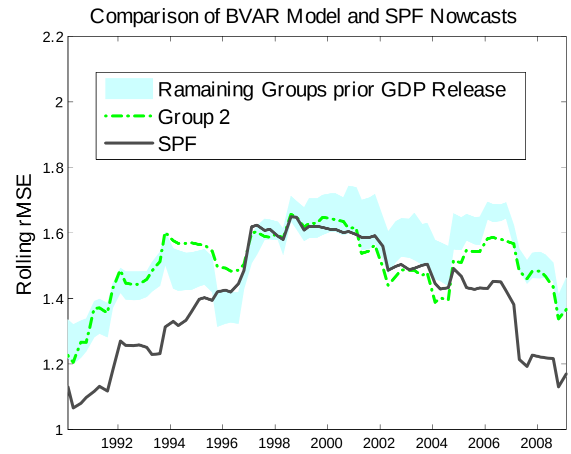


Notes: Ratio less than 1 indicates that the BVAR dominates the factor model. Horizontal axis reports the mid-point of the 10-year rolling window.

Figure 4: BVAR versus SPF rMSE

Panel A: Up to Group 2	Panel B: The Rest of the Groups

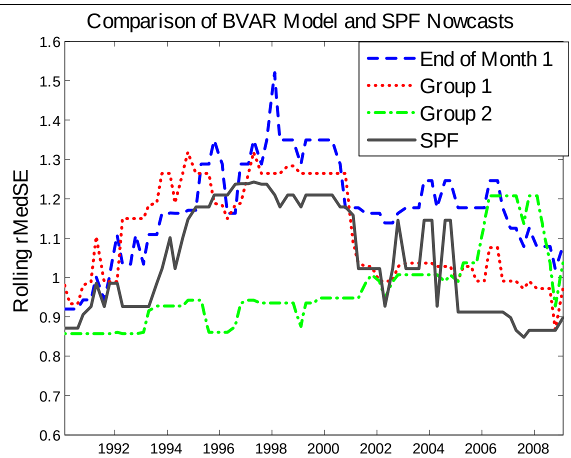


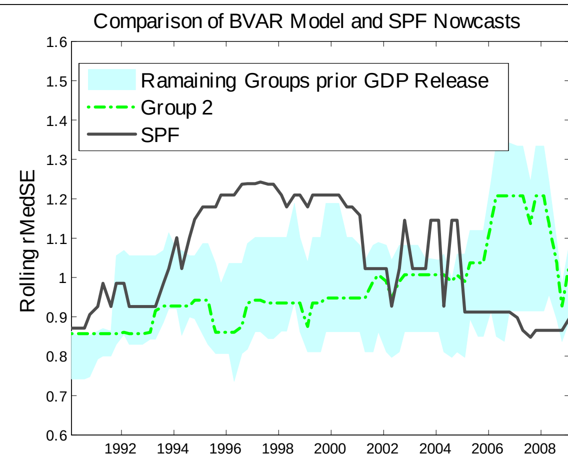


Notes: Panel A depicts rolling rMSEs, while Panel B, in addition, shows the bounds of the rMSEs for the nowcasts obtained after the middle of the quarter. Horizontal axes report the mid-point of the 10-year rolling window.

Figure 5: BVAR versus SPF rMedSE

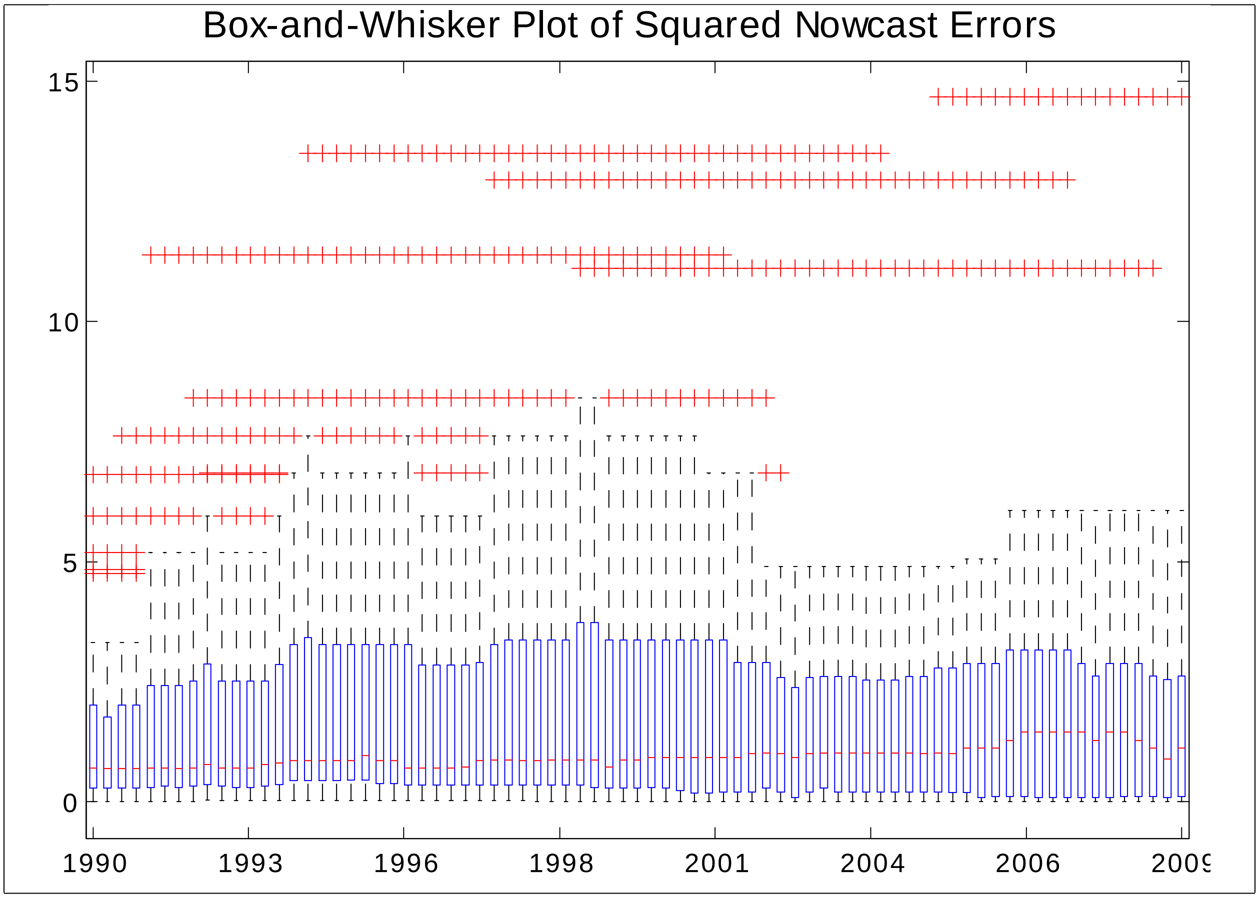
Panel A: Up to Group 2	Panel B: The Rest of the Groups





Notes: Panel A depicts rolling rMSEs, while Panel B, in addition, shows the bounds of the rMSEs for the nowcasts obtained after the middle of the quarter. Horizontal axis reports the mid-point of the 10-year rolling window.

Figure 6: Rolling Windows of Squared Nowcast Errors after Group 2 Data Release



Notes: Horizontal axis reports the mid-point of the 10-year rolling window.