Statistica Sinica Preprint No: SS-2021-0206	
Title	A Bayesian Framework for Sparse Estimation in High-Dimensional Mixed Frequency Vector Autoregressive Models
Manuscript ID	SS-2021-0206
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0206
Complete List of Authors	Nilanjana Chakraborty, Kshitij Khare and George Michailidis
Corresponding Authors	George Michailidis
E-mails	gmichail@ufl.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

A Bayesian framework for sparse estimation in high-dimensional mixed frequency Vector Autoregressive models

Nilanjana Chakraborty, Kshitij Khare and George Michailidis

Abstract:

The paper considers a Vector Autoregressive model for high-dimensional mixed frequency data, where selective time series are collected at different frequencies. The high-frequency ones are expanded and modeled as multiple time series to match the low frequency sampling of the corresponding low-frequency series. This leads to an expansion of the parameter space, and poses challenges for estimation and inference in settings with limited number of observations. We address them by considering specific structural relationships in the representation of the high-frequency series, together with sparsity of the model parameters through introduction of spike-and-Gaussian slab prior distributions. In contrast to existing observation-driven methods, the proposed Bayesian approach accommodates general sparsity patterns, and makes a data-driven choice of them. Under certain regularity conditions, we establish consistency for the posterior distribution under high-dimensional scaling. Applications on synthetic and real data illustrate the efficacy of the resulting estimates and corresponding credible intervals.

Key words and phrases: High dimensional data, mixed frequencies, nowcasting, pseudo-likelihood, spike and slab prior, strong selection consistency.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

1. Introduction

Monitoring and frequent assessment of macroeconomic indicators is important for evaluating up-to-date economic conditions and also for policy making. A key challenge stems from delays when the data become available. For example, the Gross Domestic Product (GDP) and its components that summarize the state of the economy is only available on a quarterly basis. In addition, preliminary published estimates of GDP are often revised afterwards, especially around turning points of the business cycle. However, other important economic indicators are collected at higher frequencies - monthly, such as the unemployment rate and industrial production, or even daily, such as stock market indices and interest rates.

The increasing availability of such "mixed frequency" data has led to the development of novel methods for their analysis, and also in providing forecasts/"nowcasts" for key low-frequency variables. Nowcasting, that refers to the prediction of the present or the very near future or past (Banbura et al. (2013)), focuses on providing forecasts of low-frequency series by leveraging higher frequency information, has become an important tool for policy makers (see Uematsu and Tanaka (2019); Carriero et al. (2015)).

There have been different modeling approaches dealing with mixed frequency time series data. A recent strand of inquiry has focused on employ-

ing Vector Autoregressive (VAR) models (Lütkepohl (2005)) for forecasting purposes. However, standard VAR models aim to capture lead-lag relationships between time series, observed at the same frequency. Hence, to accommodate mixed frequency data, modifications are required and the following three broad strategies have been developed for this task. The first strategy is to simply aggregate the high-frequency data to the coarsest frequency and then apply the VAR model that can be estimated with standard frequentist or Bayesian methods (Lütkepohl (2005)). This strategy discards information contained in the higher frequency series and also can not provide nowcasts. Different weighting schemes have been proposed in the literature for aggregation purposes - see, e.g., Schorfheide and Song (2015); Mariano and Murasawa (2003). The second strategy is to treat the low-frequency series as high-frequency ones with *missing observations*. Hence, the goal becomes to *impute* the missing data and estimate the resulting VAR model at the high-frequency. The Expectation-Maximization algorithm is used to obtain maximum likelihood estimates of the model parameters in a frequentist setting2 (Mariano and Murasawa (2003, 2010); Foroni and Marcellino (2014)), whereas the Data Augmentation algorithm (Eraker et al. (2014); Schorfheide and Song (2015); Ankargren and Jonéus (2019)) or a variational Bayes approach (Gefang et al. (2020)) is used for

inference purposes in a Bayesian setting. Note that the state space based approach adds additional parameters to impute/estimate, which may not be desirable in high-dimensional settings (see also Remark S2 in the Supplement). The third strategy, proposed in Ghysels (2016) (see also McCracken et al. (2015)) adopts an observation-driven viewpoint and breaks each high-frequency series into an appropriate number of low-frequency series; for example, in the presence of monthly and quarterly variables, each monthly series is expanded to three new variables for the first, second and third month in each quarter. Thus, the resulting VAR model contains $k_1 + q * k_2$ time series, where k_1 and k_2 denote the number of high and low-frequency variables, respectively, and the high-frequency data are sampled q times more often than the low-frequency series. Subsequently, standard frequentist or Bayesian techniques can be employed for estimating its parameters.

A number of papers (Sims (1992); Leeper et al. (1996); Bańbura et al. (2010); Kilian and Lütkepohl (2017)) have argued for the inclusion of many time series in VAR models to improve forecasting performance. However, this renders the problem of estimating the VAR model parameters *high dimensional*, since the number of parameters grows *quadratically* in the number of time series considered. The problem is compounded in the presence of mixed frequency data, under the third strategy discussed above.

Regularized estimation methods that aim to reduce the effective number of parameters have recently been developed in the statistics and econometrics literature (Basu and Michailidis (2015); Bańbura et al. (2010); Ghosh et al. (2021)) for VAR models. However, frequentist/Bayesian regularization techniques for high-dimensional settings have not been applied to mixed frequency VAR models, with the exception of Ghysels (2016), wherein parameter reduction is achieved through either: (a) *pre-specifying* a significant portion of the $q * k_1 + k_2$-dimensional VAR coefficient matrix to zero, or (b) reduction of the effective dimension of a $d*q*k_1*k_2$ sub-block of parameters to $k_1 * k_2$ (d is the number of lags). The first approach is quite restrictive, while the second one leaves a significant portion of the $d * (q * k_1 + k_2)^2$ parameters unconstrained and the issue of over-parametrization still pertains in high dimensional settings, when both k_1 and/or k_2 are comparable to or larger than the sample size; see Remark S1 in the Supplement.

This paper develops a novel Bayesian approach for high-dimensional mixed frequency VAR models that achieves parameter reduction through a combination of regularization (sparsity) and structural relationships between relevant parameters. The starting point is the formulation of the $q * k_1 + k_2$ dimensional VAR model as in Ghysels (2016). However, a different approach for parameter reduction is adopted that effectively reduces the

parameters from $d*(q*k_1+k_2)^2$ to $d*(k_1+k_2)^2$; see Remark S1 for a careful comparison. Another notable feature is the use of a pseudo-likelihood, based on treating the error covariance matrix as a diagonal matrix, that leads to a significant computational speed-up. Finally, simple modifications enable the approach to provide nowcasts for key low-frequency variables.

Section 2 introduces the mixed frequency VAR model along with the parameter reducing structure of its transition matrices. Section 3 motivates and presents the pseudo-likelihood approach and introduces regularization of the model parameters, while Section 4 establishes the theoretical properties. Performance of these models is investigated through extensive simulations in Section 6 and illustrated on a U.S. macroeconomic data set in Section 7. The proofs of the main theorems and lemmas, along with auxiliary derivations and additional numerical results for both synthetic and real data are presented in the Supplement.

2. A VAR model for multivariate mixed frequency data

Suppose data are observed at multiple frequencies. The proposed methodology can be adapted to any combination of such mixed frequency data (e.g., daily-weekly, or weekly-monthly), but for ease of exposition, we focus on monthly and quarterly sampled time series. Suppose we have k_1 variables

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

observed at a monthly frequency and k_2 variables observed at a quarterly frequency. Similar to the Ghysels (2016) paper, the proposed approach "breaks" each monthly time series into 3 quarterly ones, and considers a joint VAR model for all the resulting quarterly time series in the data.

Specifically, let T be the number of quarters for which data is available. For every $1 \leq i \leq k_1$, the data for the i^{th} monthly (high-frequency) variable is broken into three quarterly series as follows: (a) $\{y_{i,\mathrm{H}}^{t}\}_{t=1}^{T}$ represents the quarterly time series consisting of the i^{th} monthly variable values at the end of the last month of each quarter, (b) $\{y_{i,\mathrm{H}}^{t-1/3}\}_{t=1}^{T}$ represents the quarterly time series consisting of the i^{th} monthly variable values at the end of the second month of each quarter, and (c) $\{y_{i,\mathrm{H}}^{t-2/3}\}_{t=1}^{T}$ represents the quarterly series consisting of the i^{th} monthly variable values at the end of the first month of each quarter. For every $1 \leq j \leq k_2$, the data for the j^{th} quarterly (low-frequency) variable is represented by the single quarterly series $\{y_{j,\mathrm{L}}^{t}\}_{t=1}^{T}$. We can model the ' jointly through the VAR model of lag-d

$$\bar{\boldsymbol{y}}^t = \sum_{u=1}^{d} \bar{\mathbf{W}}_u \bar{\boldsymbol{y}}^{t-u} + \bar{\boldsymbol{\varepsilon}}^t, \quad \text{where} \tag{2.1}$$

$$\bar{\boldsymbol{y}}^t = \left[y_{1,\mathrm{H}}^{t-2/3} \; y_{1,\mathrm{H}}^{t-1/3} \; y_{1,\mathrm{H}}^{t} \; y_{2,\mathrm{H}}^{t-2/3} \; y_{2,\mathrm{H}}^{t-1/3} \; y_{2,\mathrm{H}}^{t} \; \cdots y_{k_1,\mathrm{H}}^{t-2/3} \; y_{k_1,\mathrm{H}}^{t-1/3} \; y_{k_1,\mathrm{H}}^{t} \; y_{1,\mathrm{L}}^{t} \; \cdots y_{k_2,\mathrm{L}}^{t} \right]',$$

and the errors $\{\bar{\boldsymbol{\varepsilon}}^t\}_{t=1}^{T}$ are assumed to be independent and identically multivariate normally distributed with mean zero and covariance matrix $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$.

The temporal dependence structure of the model is characterized by the $(3k_1 + k_2) \times (3k_1 + k_2)$ transition matrices $\bar{\mathbf{W}}_1, \bar{\mathbf{W}}_2, \cdots, \bar{\mathbf{W}}_d$.

To account for the fact that some of the $(3k_1 + k_2)$ quarterly time series are chunks of the same monthly time series, and for parameter reduction in high-dimensional settings, wherein $(3k_1+k_2)$ is comparable to or larger than T, we impose the following structure on each VAR transition matrix $\bar{\mathbf{W}}_u$, whose primary purpose is to reduce the dimensionality of the parameter space. We specify $\bar{\mathbf{W}}_u = ((w^u_{rs}))_{1\le r,s\le(3k_1+k_2)}$, for every $1 \le r \le (3k_1 + k_2)$ and $1 \le j \le k_1$, as follows:

$$w^u_{r,3j-1} = \theta w^u_{r,3j} \quad \text{and} \quad w^u_{r,3j-2} = \theta w^u_{r,3j-1} = \theta^2 w^u_{r,3j}, \tag{2.2}$$

with $\theta \in (0, 1)$. Note that for $1 \le j \le k_1$, the entries $w^u_{r,3j}$, $w^u_{r,3j-1}$, $w^u_{r,3j-2}$ capture the (linear) effect of $y^{t-u}_{j,\mathrm{H}}$, $y^{t-u-1/3}_{j,\mathrm{H}}$, $y^{t-u-2/3}_{j,\mathrm{H}}$, respectively, on the r^{th} quarterly time series at time t. Thus, the effect of the j^{th} high-frequency variable is *dampened* by a factor of θ when moving back one month (or a third of a quarter) in time, as expressed in mathematical terms in (2.2). Similarly, for every $1 \le i \le k_1$, and $1 \le s \le (3k_1 + k_2)$, we specify

$$w^u_{3i-1,s} = \theta w^u_{3i,s} \quad \text{and} \quad w^u_{3i-2,s} = \theta w^u_{3i-1,s} = \theta^2 w^u_{3i,s}. \tag{2.3}$$

Note that for $1 \le j \le k_1$, the entries $w^u_{3i,s}$, $w^u_{3i-1,s}$, $w^u_{3i-2,s}$ capture the (linear) effect of the u^{th} lagged value of the s^{th} quarterly time series on $y^t_{i,\mathrm{H}}$,

$y_{i,\mathrm{H}}^{t-1/3}$, $y_{i,\mathrm{H}}^{t-2/3}$ respectively. We assume that the effect of the value of any of the $(3k_1 + k_2)$ quarterly time series at time $t-1$ on $y_{i,\mathrm{H}}^{t-1/3}$ is equal to the corresponding effect on $y_{i,\mathrm{H}}^{t}$ *dampened* by the same factor θ, and this is expressed in mathematical terms in (2.3). Note that the dampening rate is the same when comparing the effect of $y_{i,\mathrm{H}}^{t-1}$ on $y_{i,\mathrm{H}}^{t-2/3}$ to the effect of $y_{i,\mathrm{H}}^{t-4/3}$ on $y_{i,\mathrm{H}}^{t-2/3}$ or, when comparing the effect of $y_{i,\mathrm{H}}^{t-1}$ on $y_{i,\mathrm{H}}^{t-2/3}$ to the effect of $y_{i,\mathrm{H}}^{t-1}$ on $y_{i,\mathrm{H}}^{t-1/3}$. It can be seen that the effects are *dampened* at the same rate when moving one month back, are invariant of the specific time instance and only depend on the time difference.

To further illustrate the effect of this structural relationship, we consider the one-step ahead forecasts generated by this model for lag $d = 1$. Suppose we have data until quarter $t-1$. The model can be used to simultaneously generate up to 3 month ahead forecasts for each monthly variable as follows: for $1 \le i \le k_1$, the forecasted values of the i^{th} monthly variable at time $t-2/3$ (end of 1st month of quarter t) can be expressed in terms of available lagged values of other monthly and quarterly variables as follows (see similar expressions for $y_{i,\mathrm{H}}^{t-1/3}$ and $y_{i,\mathrm{H}}^{t}$ in the Supplement (Sec. S2.1).)

$$\underbrace{a_{i,1}\theta^2 y_{1,\mathrm{H}}^{t-5/3} + a_{i,1}\theta y_{1,\mathrm{H}}^{t-4/3} + a_{i,1} y_{1,\mathrm{H}}^{t-1}}_{\substack{\text{effect of lagged values } \left(y_{1,\mathrm{H}}^{t-5/3}, y_{1,\mathrm{H}}^{t-4/3}, y_{1,\mathrm{H}}^{t-1}\right) \\ \text{on } y_{i,\mathrm{H}}^{t} \text{with dampening by } \theta \text{ as time gap increases}}} + \cdots + \underbrace{a_{i,k_1}\theta^2 y_{k_1,\mathrm{H}}^{t-5/3} + a_{i,k_1}\theta y_{k_1,\mathrm{H}}^{t-4/3} + a_{i,k_1} y_{k_1,\mathrm{H}}^{t-1}}_{\substack{\text{effect of lagged values } \left(y_{k_1,\mathrm{H}}^{t-5/3}, y_{k_1,\mathrm{H}}^{t-4/3}, y_{k_1,\mathrm{H}}^{t-1}\right) \\ \text{on } y_{i,\mathrm{H}}^{t} \text{with dampening by } \theta \text{ as time gap increases}}} +$$

$$\underbrace{a_{i,k_1+1} y_{1,\mathrm{L}}^{t-1} + \cdots + a_{i,k_1+k_2} y_{k_2,\mathrm{L}}^{t-1}}_{\text{effect of lagged values of low-frequency variables on } y_{i,\mathrm{H}}^{t}} + \varepsilon_{i,\mathrm{H}}^{t-2/3} \qquad (2.4)$$

where $a_{i,j}$ represents the effect of $y_{j,\mathrm{H}}^{t-1}$ on $y_{i,\mathrm{H}}^{t-2/3}$ for $1 \leq i \leq k_1$ and $1 \leq j \leq k_1$, and a_{i,k_1+j} represents the effect of $y_{j,\mathrm{L}}^{t-1}$ on $y_{i,\mathrm{H}}^{t-2/3}$ for $1 \leq i \leq k_1$ and $1 \leq j \leq k_2$. In other words, $a_{i,j}$ represents the effect of the value of the j^{th} monthly variable at the current month on the value of the i^{th} monthly variable in the next month for $1 \leq i, j \leq k_1$, and a_{i,k_1+j} represents the effect of the value of the j^{th} quarterly variable at the current quarter on the value of the i^{th} monthly variable in the next month for $1 \leq i \leq k_1$ and $1 \leq j \leq k_2$. Note that *all these forecasts are based on data up to time $t-1$ and use all the latest values for all the variables.* For mid-quarter forecasts, when data are available until the end of the first/second month of the next quarter, a modification of this approach is proposed in Section 5.

Now, for $1 \leq j \leq k_2$, the forecast for the j^{th} quarterly variable at quarter t can be expressed in terms the lagged values of other monthly and quarterly variables as follows:

$$\underbrace{a_{k_1+j,1}\theta^2 y_{1,\mathrm{H}}^{t-5/3} + a_{k_1+j,1}\theta y_{1,\mathrm{H}}^{t-4/3} + a_{k_1+j,1} y_{1,\mathrm{H}}^{t-1}}_{\substack{\text{effect of lagged } y_{1,\mathrm{H}} \text{values } \left(y_{1,\mathrm{H}}^{t-5/3}, y_{1,\mathrm{H}}^{t-4/3}, y_{1,\mathrm{H}}^{t-1}\right) \text{ on } y_{j,\mathrm{L}}^{t} \\ \text{with dampening by } \theta \text{ as time gap increases}}} + \cdots +$$

$$\underbrace{a_{k_1+j,k_1}\theta^2 y_{k_1,\mathrm{H}}^{t-5/3} + a_{k_1+j,k_1}\theta y_{k_1,\mathrm{H}}^{t-4/3} + a_{k_1+j,k_1} y_{k_1,\mathrm{H}}^{t-1}}_{\substack{\text{effect of lagged } y_{k_1,\mathrm{H}} \text{values} \left(y_{k_1,\mathrm{H}}^{t-5/3}, y_{k_1,\mathrm{H}}^{t-4/3}, y_{k_1,\mathrm{H}}^{t-1}\right) \\ \text{on } y_{j,\mathrm{L}}^{t} \text{ with dampening by } \theta \text{ as time gap increases}}} +$$

$$\underbrace{a_{k_1+j,k_1+1} y_{1,\mathrm{L}}^{t-1} + \cdots + a_{k_1+j,k_1+k_2} y_{k_2,\mathrm{L}}^{t-1}}_{\text{effect of lagged values of low-frequency variables on } y_{j,\mathrm{L}}^{t}} + \varepsilon_{j,\mathrm{L}}^{t} \tag{2.5}$$

where $a_{k_1+j,i}$ represents the effect of $y_{i,\mathrm{H}}^{t-1}$ on $y_{j,\mathrm{L}}^{t}$ for $1 \leq j \leq k_2$ and $1 \leq$

$i \leq k_1$, and $a_{k_1+j,k_1+j'}$ represents the effect of $y^{t-1}_{j',\mathrm{L}}$ on $y^t_{j,\mathrm{L}}$ for $1 \leq j, j' \leq k_2$ and $1 \leq i \leq k_1$. Equations (2.4) and (2.5) are visualized in Figure 1 of the Supplement (Sec. S2.2).

Let $\mathbf{A}_u = ((a^u_{i,j}))_{1\leq i,j\leq k_1+k_2}$, where $a^u_{i,j}$'s are as described in (2.4), (2.5) and (S2.2) and (S2.3) of the supplement , and $\mathbf{A}^u_{11}, \mathbf{A}^u_{12}, \mathbf{A}^u_{21}, \mathbf{A}^u_{22}$ are sub-matrices of $\mathbf{A}_u$ such that

$$\mathbf{A}_u = \begin{bmatrix} \mathbf{A}^u_{11\,k_1\times k_1} & \mathbf{A}^u_{12\,k_1\times k_2} \\ \mathbf{A}^u_{21\,k_2\times k_1} & \mathbf{A}^u_{22\,k_2\times k_2} \end{bmatrix}, \quad \text{and}$$

$$\boldsymbol{y}^t = \left[y^t_{1,\mathrm{H}}\, y^t_{2,\mathrm{H}}\, \cdots\, y^t_{k_1,\mathrm{H}}\, y^{t-1/3}_{1,\mathrm{H}}\, y^{t-1/3}_{2,\mathrm{H}}\, \cdots\, y^{t-1/3}_{k_1,\mathrm{H}}\, y^{t-2/3}_{1,\mathrm{H}}\, y^{t-2/3}_{2,\mathrm{H}}\, \cdots\, y^{t-2/3}_{k_1,\mathrm{H}}\, y^t_{1,\mathrm{L}}\, y^t_{2,\mathrm{L}}\, \cdots\, y^t_{k_2,\mathrm{L}} \right]'$$

denotes a permuted version of $\bar{\boldsymbol{y}}^t$. It follows by straightforward calculations that the VAR model in (2.1) can be equivalently represented as

$$\boldsymbol{y}^t = \sum_{u=1}^{d} \mathbf{W}_u \boldsymbol{y}^{t-u} + \boldsymbol{\varepsilon}^t, \quad \text{where} \tag{2.6}$$

$$\mathbf{W}_u = \left[\begin{array}{ccc|c} \theta^2 & \theta^3 & \theta^4 & \\ \theta & \theta^2 & \theta^3 & \left(\begin{array}{c}\theta^2\\ \theta \\ 1\end{array}\right) \otimes \mathbf{A}^u_{12} \\ 1 & \theta & \theta^2 & \\ \hline \multicolumn{3}{c|}{\left(\begin{array}{ccc} 1 & \theta & \theta^2 \end{array}\right) \otimes \mathbf{A}^u_{21}} & \mathbf{A}^u_{22} \end{array}\right] \tag{2.7}$$

and $\{\boldsymbol{\varepsilon}^t\}_{t=1}^T$ are i.i.d. multivariate normal with mean zero and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, with $\boldsymbol{\varepsilon}^t, \{\mathbf{W}_u\}_{u=1}^d, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ being permuted versions of $\bar{\boldsymbol{\varepsilon}}^t, \{\bar{\mathbf{W}}_u\}_{u=1}^d, \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$, respectively. The extension of this model is given in Section 5 to incorporate the nowcast of quarterly variables across the monthly horizon.

3. Regularized Bayesian inference

The likelihood of $\mathbf{A}, \theta, \mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ using the model (2.6) is given by

$$L\big(\{\mathbf{A}_u\}_{u=1}^{d}, \mathbf{\Sigma}_{\boldsymbol{\varepsilon}} \mid \boldsymbol{y}^t, t = d, \cdots, T\big) = \frac{1}{\sqrt{2\pi|\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}|^{(T-d+1)}}} \exp\left(-\frac{1}{2} tr \sum_{t=d}^{T} \left((\boldsymbol{y}^t - \sum_{u=1}^{d} \mathbf{W}_u \boldsymbol{y}^{t-u})' \mathbf{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} (\boldsymbol{y}^t - \sum_{u=1}^{d} \mathbf{W}_u \boldsymbol{y}^{t-u}) \right)\right) \tag{3.8}$$

Our main interest is in estimating the transition matrices $\mathbf{W}_u$ (which are function of $\mathbf{A}_u$'s and θ), and thus we treat $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ as an unknown nuisance parameter. To that end, we define a pseudo-likelihood function which is equal to the joint density of the data under the assumption that $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ (up to permutation) is block diagonal. In particular, suppose we assume that $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ is given by

$$\mathbf{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{Q}' \text{diag}\left(\mathbf{\Sigma}_{\mathbf{1},\text{H}}, \cdots, \mathbf{\Sigma}_{\mathbf{k_1},\text{H}}, \sigma_{\mathbf{1}}^{\mathbf{2}}, \cdots, \sigma_{\mathbf{k_2}}^{\mathbf{2}}\right) \mathbf{Q} \tag{3.9}$$

where, $\mathbf{\Sigma}_{\mathbf{i},\text{H}}$ is the 3×3 variance-covariance matrix of $(\boldsymbol{\varepsilon}_{i,\text{H}}^{t}\ \boldsymbol{\varepsilon}_{i,\text{H}}^{t-1/3}\ \boldsymbol{\varepsilon}_{i,\text{H}}^{t-2/3})'$, and σ_j^2 is the variance of $\boldsymbol{\varepsilon}_{j,\text{L}}^{t}$ for $1 \le i \le k_1$, and $1 \le j \le k_2$, and $\mathbf{Q}$ is a permutation matrix such that

$$\boldsymbol{y}^t = \mathbf{Q} \left[y_{1,\text{H}}^{t}\ y_{1,\text{H}}^{t-1/3}\ y_{1,\text{H}}^{t-2/3}\ \cdots\ y_{k_1,\text{H}}^{t}\ y_{k_1,\text{H}}^{t-1/3}\ y_{k_1,\text{H}}^{t-2/3}\ y_{1,\text{L}}^{t}\ y_{2,\text{L}}^{t}\ \cdots\ y_{k_2,\text{L}}^{t} \right]'.$$

The above form for $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ only captures correlations between the monthly components of each high-frequency variable and essentially ignores the cross

correlations between the high and low-frequency variables. Our pseudo-likelihood function L_{pseudo} is defined as the joint density of the data with $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ structured as in (3.9), and based on the block diagonal structure of $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$, can be shown to have the form

$$\begin{aligned} & L_{\text{pseudo}}\left(\{\mathbf{A}_u\}_{u=1}^{d}, \theta, \{\mathbf{\Sigma}_{i,\text{H}}\}_{i=1}^{k_1}, \{\sigma_j^2\}_{j=1}^{k_2} \mid \boldsymbol{y}^t, t=d,\cdots,T\right) \\ \propto & \prod_{i=1}^{k_1}\Bigg\{ \exp\left[-\frac{1}{2}\sum_{t=d}^{T}\left(\boldsymbol{y}_{i,\text{H}}^t - \left(\sum_{u=1}^{d}\mathbf{W}_u\boldsymbol{y}^{t-u}\right)_{i,\text{H}}\right)' \mathbf{\Sigma}_{i,\text{H}}^{-1}\left(\boldsymbol{y}_{i,\text{H}}^t - \left(\sum_{u=1}^{d}\mathbf{W}_u\boldsymbol{y}^{t-u}\right)_{i,\text{H}}\right)\right] \\ & \times \left|\mathbf{\Sigma}_{i,\text{H}}\right|^{-\frac{T-d+1}{2}}\Bigg\} \times \prod_{j=1}^{k_2}\Bigg\{ \exp\left[-\frac{1}{2}\sum_{t=d}^{T}\frac{\left(y_{j,\text{L}}^t - \left(\sum_{u=1}^{d}\mathbf{W}_u\boldsymbol{y}^{t-u}\right)_{3k_1+j,\text{L}}\right)^2}{\sigma_j^2}\right](\sigma_j^2)^{-\frac{T-d+1}{2}}\Bigg\} \end{aligned} \tag{3.10}$$

with $\boldsymbol{y}_{i,\text{H}}^t = (y_{i,\text{H}}^t \; y_{i,\text{H}}^{t-1/3} \; y_{i,\text{H}}^{t-2/3})'$ for $1 \le i \le k_1$.

The pseudo-likelihood function L_{pseudo} has a simpler product form (and significantly fewer parameters) than the likelihood function in (3.8) and also has $6k_1 + k_2$ parameters for $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$, as opposed to the $(3k_1 + k_2)(3k_1 + k_2 + 1)/2$ parameters for the likelihood function. Consequently, it leads to significant simplifications regarding methodology, computation and theoretical analysis. The latter comes at a cost, since ignoring the cross correlations among the error blocks decreases statistical efficiency, and raises potential questions regarding the validity and consistency of the resulting estimates. However, our main theoretical results in Section 4 establish that the pseudo-likelihood based Bayesian approach leads to consistent estimates under high-dimensional scaling, even when the true error covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ is not block diagonal (but satisfies some mild assumptions on uniform

boundedness of its eigenvalues). This result, combined with the significant computational simplifications, make the pseudo-likelihood based approach highly preferable in high-dimensional settings. A regression interpretation of the pseudo-likelihood function is provided in the Supplement (Sec. S2.3).

3.1 Specification of prior distributions on the model parameters

Next, we specify prior distributions for the parameters $\{\mathbf{A}_u\}_{u=1}^{d}, \theta, \{\mathbf{\Sigma}_{i,\mathrm{H}}\}_{i=1}^{k_1}$, and $\left\{\sigma_j^2\right\}_{j=1}^{k_2}$. As previously mentioned, we consider a high-dimensional setting, wherein the dimension $p = 3k_1 + k_2$ of the VAR model increases with the sample size T. To that end, a further parameter reduction in the $(k_1 + k_2)^2$ parameters in the matrix $\mathbf{A}_u = ((a_{ij}^u))_{1\leq i,j\leq k_1+k_2}$ for $i = 1, \cdots, d$ is obtained by a *sparsity* inducing prior distribution. To facilitate its introduction, we define binary variables $\gamma_{ij}^u = 1_{a_{ij}^u \neq 0} \quad \forall 1 \leq i, j \leq k_1 + k_2$, that indicate the entries of $\mathbf{A}_u$ that are "active" (non-zero). The matrix $\mathcal{G}_u = \left(\gamma_{ij}^u\right)_{1\leq i,j\leq k_1+k_2}$ captures the pattern of active entries (non-zeros) in $\mathbf{A}_u$. While we specify Gaussian mixture prior distributions for all entries of $\mathbf{A}_u, u = 1, \cdots, d$, the slightly different nature of the factors for the high-frequency and low-frequency variables in L_{pseudo} (see (S2.9) of the supplement) leads us to specify different variance terms for entries in the first k_1 and the last k_2 rows of $\mathbf{A}_u$ for analytical convenience. Let, $\mathcal{G} = (\mathcal{G}_1, \cdots, \mathcal{G}_d)$

be the collection of all activity matrices. In particular, given the activity graph $\mathcal{G}$, we specify for $u = 1, \cdots, d$.

$$\begin{aligned} a^u_{ij} \mid \sigma^2_{i-k_1}, \mathcal{G}_u &\sim (1-\gamma^u_{ij})\, 1_{\{a^u_{ij}=0\}} + \gamma^u_{ij} \mathcal{N}(0, \sigma^2_{i-k_1}\tau^2) \;\; \forall\, k_1+1 \le i \le k_1+k_2,\; 1 \le j \le k_1+k_2 \\ a^u_{ij} \mid \boldsymbol{\Sigma}_{i,\mathrm{H}}, \mathcal{G}_u &\sim (1-\gamma^u_{ij})\, 1_{\{a^u_{ij}=0\}} + \gamma^u_{ij} \mathcal{N}(0, \tilde{\sigma}^2_i\tau^2) \;\; \forall\, 1 \le i \le k_1,\; 1 \le j \le k_1+k_2 \\ \gamma^u_{ij} &\overset{iid}{\sim} \text{Bernoulli}(q) \end{aligned} \tag{3.11}$$

where σ^2_j is as defined in (3.9) for $1 \le j \le k_2$, and $\tilde{\sigma}^2_i$ is the reciprocal of the $(3,3)$ entry of $\mathbf{C}'\boldsymbol{\Sigma}^{-1}_{i,\mathrm{H}}\mathbf{C}$ for a non-singular 3×3 matrix $\mathbf{C}$ whose last column is $\boldsymbol{\delta} = (\theta^2\ 1\ \theta)'$ and first two rows are arbitrarily fixed. While any reasonable scalar function of $\boldsymbol{\Sigma}_{i,\mathrm{H}}$ could potentially be used for $\tilde{\sigma}^2_i$, the choice $\mathbf{C}'\boldsymbol{\Sigma}^{-1}_{i,\mathrm{H}}\mathbf{C}$ leads to closed form expressions for the computation of posterior graph selection probabilities and simplifies both computations and theoretical analysis. For the variance parameters $\{\boldsymbol{\Sigma}_{i,\mathrm{H}}\}_{i=1}^{k_1}$ and $\{\sigma^2_j\}_{j=1}^{k_2}$ we use independent Inverse-Wishart$(\omega, \mathbf{V})$ and independent Inverse-Gamma(α, β) priors respectively. These are standard conditionally conjugate choices for the variance parameters in Bayesian regression problems. A Uniform$[0,1]$ prior is used for θ. The default choice of hyper-parameters that we use is $q = 1/p$ (based on Assumption A3 in the supplement, that is needed for theoretical results), α, β and $\omega = 1, \tau^2 = 0.5$ and $\mathbf{V}$ to be a small multiple of $\mathbf{I}$. This choice can be appropriately modified if prior information is available. If computational resources and time is available, another option is to choose hyper-parameters using cross-validation.

A recent empirical study (Cross et al., 2020) suggests that a purely sparse model may not be most appropriate for macroeconomic data. Giannone et al. (2021) propose a variant of the spike-and-slab priors which combines both sparsity and shrinkage by using Uniform$[0,1]$ priors on q and $R^2 := \frac{qp\tau^2}{qp\tau^2+1}$. This leads to a negative correlation between the inclusion probability q and the slab variance τ^2, which provides a desirable balance between sparsity and shrinkage. However, one needs discretization to sample from the highly non-standard conditional posterior distributions of q and τ^2. While we do not explore this approach due to the additional computational burden in high-dimensional settings, we note that such priors and the relevant additional sampling steps for posterior computation can be easily incorporated in our framework if strongly desired by the practitioner.

3.2 Computation of the posterior distribution

Let us denote $\mathbf{A} = \{\mathbf{A}_1, \cdots, \mathbf{A}_d\}$. The joint posterior of $\mathbf{A}$, θ, $\{\mathbf{\Sigma}_{i,\mathrm{H}}\}_{i=1}^{k_1}$, $\left\{\sigma_j^2\right\}_{j=1}^{k_2}$ turns out to be intractable for closed form computations or direct i.i.d. sampling. However, the full conditional posterior distributions of all the parameters (except θ) are standard and easy to sample from, as shown in the Supplement (Sec. S3). While the full conditional posterior distribution of θ is not standard, given that θ is a scalar with a bounded range $(0,1)$,

we generate samples from the full conditional posterior distribution of θ by an efficient discrete approximation (see Supplemental Section S3). We then use a Gibbs sampling algorithm to generate approximate samples from the posterior distribution. The details of the Gibbs sampler are provided in the Supplement (Sec. S2.4). The Markov chain output of this algorithm can be used to approximate posterior quantities.

4. High dimensional theoretical results

For simplicity of exposition, we consider a $d = 1$ VAR model. The extension to VAR(d) models is rather straightforward using similar ideas. We let the dimension $p = p_n$ of the VAR model vary with the sample size n where $p = 3k_1 + k_2$. We assume that the data are generated according to the following true VAR model. For sample size $n \geq 1$, let $\mathcal{Y}_n := (\boldsymbol{y}^{n,0}, \cdots, \boldsymbol{y}^{n,n})$ be the set of observations obtained from $\boldsymbol{y}^{n,k} = \mathbf{W}_{0n}\boldsymbol{y}^{n,k-1} + \boldsymbol{\varepsilon}^{n,k}$ for $1 \leq k \leq n$. The errors $\{\boldsymbol{\varepsilon}^{n,k}\}_{k=1}^{n}$ are i.i.d. $\mathcal{N}_{p_n}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon},0n})$. Let $\{\mathbf{W}_{0n}\}_{n\geq 1}$ denote the sequence of the true coefficient matrices and $\{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon},0n}\}_{n\geq 1}$ denotes the sequence of the true error covariance matrices. Let $\mathbb{P}_0$ denote the probability measure underlying the true model described above, and $\mathcal{G}_0 = \mathcal{G}_{0,n}$ the true underlying activity graph for the sparse coefficient matrix $\mathbf{W}_0$. The quantities $\mu_{\min}(\mathcal{A})$, $\mu_{\max}(\mathcal{A})$ are as defined in Supplemental Section S2.8

and $\mathbf{C}_X$ as defined in (S2.5) of the supplement, with $\mathbf{W}_{0n}$ and $\mathbf{\Sigma}_{\boldsymbol{\varepsilon},0n}$ as the underlying parameter values. For $1 \leq i \leq k_1 + k_2$, let $\nu_i = \nu_i(\mathcal{G})$ denote the number of non-zero/active entries in the i^{th} row of $\mathcal{G}$. Then the maximum number of non-null entries within the rows of $\mathbf{A}_{0n}$ is given by $b_n = \max_{1 \leq i \leq k_1+k_2} \nu_i(\mathcal{G}) + 1$. The minimum signal strength of $\mathbf{A}_{0n}$ is given by $s_n^2 := \inf_{(i,j):a_{ij} \neq 0} |a_{ij}|$. The total number of non-zero entries in $\mathbf{A}_{0n}$ is denoted by $\delta_n = \sum_{i=1}^{k_1+k_2} \nu_i(\mathcal{G})$. For ease of exposition, we will henceforth denote $\mathbf{W}_{0n}$ as $\mathbf{W}_0$, $\mathbf{A}_{0n}$ as $\mathbf{A}_0$ and $\mathbf{\Sigma}_{\boldsymbol{\varepsilon},0n}$ as $\mathbf{\Sigma}_{\boldsymbol{\varepsilon},0}$ and highlight their dependence on n as needed. Standard regularity assumptions on the true model parameters are provided and discussed in the Supplement (Sec S2.9).

Let $\pi_{\text{pseudo}}(\cdot \mid \mathcal{Y})$ denote the (pseudo) posterior probability mass function on the space of activity graphs, whose expression is derived in the Supplement (Sec. S4). Next, we establish that under high dimensional scaling and with the Assumptions , the posterior distribution of the activity graph concentrates around the true activity graph.

Theorem 1 (Strong Model Selection Consistency)**.** *For the mixed frequency VAR model posited in (2.6) with lag $d = 1$ and the prior distributions on $\boldsymbol{A}, \mathcal{G}, \mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ specified in Section 3.1 and fixed θ, satisfying Assumptions A1-A4 (provided in the Supplement), the following holds; i.e. the (pseudo) posterior probability assigned to the true activity graph $\mathcal{G}_0$ converges to 1 as*

the sample size increases to ∞.

$$\pi_{\text{pseudo}}(\mathcal{G}_0 \mid \mathcal{Y}) \xrightarrow{\mathbb{P}_0} 1 \qquad \text{as } n \to \infty.$$

The above model selection consistency result can be immediately leveraged to obtain the following estimation consistency result.

Theorem 2 (Estimation Consistency Rate)**.** *For the mixed frequency VAR model posited in (2.6) with* $d = 1$ *and the prior distributions on* $\boldsymbol{A}, \mathcal{G}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ *specified in Section 3.1 and fixed* θ*, satisfying Assumptions A1-A4, there exists a constant* K *(not depending on* n*) such that*

$$\mathbb{E}_0 \left[\Pi_{\text{pseudo}} \left(\|\boldsymbol{A} - \boldsymbol{A}_0\|_F > K \, \frac{1 + \mu_{\max}(\mathcal{A})}{\mu_{\min}(\mathcal{A})} \sqrt{\frac{\delta_n \log p}{n}} \Big| \mathcal{Y} \right) \right] \to 0 \text{ as } n \to \infty,$$

wherein Π_{pseudo} refers to the (pseudo) posterior probability distribution.

The proofs of Theorems 1 and 2 are provided in Supplemental Sections S5.2 and S5.3, respectively. The proof of Theorem 1 involves a careful analysis of the ratio $\frac{\pi_{\text{pseudo}}(\mathcal{G}|\mathcal{Y})}{\pi_{\text{pseudo}}(\mathcal{G}_0|\mathcal{Y})}$ for $\mathcal{G} \neq \mathcal{G}_0$, that can be written as a product of $(k_1 + k_2)$ terms (see equation (S5.28) in the Supplemental document). The main challenge and novelty in the proof is the analysis of the k_1 terms $\{B(\boldsymbol{m}_i, \boldsymbol{t}_i)\}_{i=1}^{k_1}$ corresponding to the multivariate response high-frequency regressions. See Remark S3 in the Supplement for a detailed discussion. Also, we examine the behavior of the quantity $\frac{1+\mu_{\max}(\mathcal{A})}{\mu_{\min}(\mathcal{A})} \sqrt{\frac{b_n \log(3k_1+k_2)}{n}}$ from

Assumption A1 under different asymptotic regimes for $k_1 = k_{1,n}$ and $k_2 = k_{2,n}$ in Remark S4 of the Supplement.

5. Obtaining mid-quarter forecasts (Nowcasts)

In the model development in Section 2, we consider data from Quarter 1 to Quarter T to predict the monthly variables at time $T + 1/3, T + 2/3$ and $T + 1$ and quarterly variables at time $T + 1$ and then do so recursively for farther in the future time points. Next, suppose that new data on the monthly variables arrive at time $T + 1/3$ which is in the middle of quarter $T + 1$. Then, the model can be modified so as to provide a nowcast for the quarterly variables at time $T + 1$ that corresponds to the forecast horizon $h = 2/3$. Let us define a new $(3k_1 + k_2)$ dimensional process

$$\boldsymbol{y}^t = \left[y_{1,\mathrm{H}}^{t+1/3} \; y_{2,\mathrm{H}}^{t+1/3} \; \cdots \; y_{k_1,\mathrm{H}}^{t+1/3} \; y_{1,\mathrm{H}}^{t} \; \cdots \; y_{k_1,\mathrm{H}}^{t} \; y_{1,\mathrm{H}}^{t-1/3} \; \cdots \; y_{k_1,\mathrm{H}}^{t-1/3} \; y_{1,\mathrm{L}}^{t} \; y_{2,\mathrm{L}}^{t} \; \cdots \; y_{k_2,\mathrm{L}}^{t} \right]'. \tag{5.12}$$

Essentially, we start counting quarters *backwards* from time $T+1/3$. Hence, the most recent "quarter" covers months $T + 1/3, T, T - 1/3$, the previous "quarter" months $T - 2/3, T - 1, T - 4/3$, and so forth. Each monthly series is broken into three quarterly series, and each quarterly series still has exactly one observation in each newly defined "quarter". A VAR model using the methodology described in Section 3 can now be estimated to

provide nowcasts of the quarterly variables at time $T+1$ and so on. If data on monthly variables is available until the second month of the next quarter, i.e., at time $T+2/3$, we can make similar modifications to the model by considering

$$\boldsymbol{y}^t = \left[y_{1,\mathrm{H}}^{t+2/3} \; y_{2,\mathrm{H}}^{t+2/3} \; \cdots \; y_{k_1,\mathrm{H}}^{t+2/3} \; y_{1,\mathrm{H}}^{t+1/3} \; \cdots \; y_{k_1,\mathrm{H}}^{t+1/3} \; y_{1,\mathrm{H}}^{t} \; \cdots \; y_{k_1,\mathrm{H}}^{t} \; y_{1,\mathrm{L}}^{t} \; y_{2,\mathrm{L}}^{t} \; \cdots \; y_{k_2,\mathrm{L}}^{t} \right]' \tag{5.13}$$

which will lead to the nowcasts of the quarterly variables corresponding to the forecast horizon $h = 1/3$. We demonstrate the performance of the proposed nowcasting methodology on simulated data (Section 6.1), as well as real macroeconomic data (Section 7).

6. Performance Evaluation based on Simulation Studies

We first illustrate the model selection and estimation performance of the Bayesian mixed frequency (BMF henceforth) model. We consider five VAR(1) models of different sizes, e.g. **Setting 1:** $k_1 = 3, k_2 = 30$, **Setting 2:** $k_1 = 5, k_2 = 50$, **Setting 3:** $k_1 = 10, k_2 = 50$, **Setting 4:** $k_1 = 20, k_2 = 20$, **Setting 5:** $k_1 = 30, k_2 = 10$, each with three different values of θ: $\theta = 0.2, \theta = 0.5$ and $\theta = 0.8$ over the range of $(0, 1)$. We generate $n = 100, 150$ time points for the VAR models corresponding to settings 1,4 and 5 and $n = 200, 400$ time points for those in settings 2 and 3.

Data Generation: The true transition matrix $\mathbf{A}$ is generated with non-zero entries drawn from Unif $(0, 10) \cup$ Unif $(-10, 0)$. The edge density of the activity graph of $\mathbf{A}$ is fixed at 4% for each VAR model; however for settings 1 and 2, we do not impose sparsity on the $\mathbf{A}_{11}$ block as the high-frequency block is small. Based on the discussion in the Supplemental Section S6.3 the spectral radius of $\mathbf{A}$ is set to 0.8 for $\theta = 0.2$ and 0.7 for $\theta = 0.5, 0.8$ for setting 1. To generate $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ as given in (3.10), we generate $\sigma_j^2, j = 1 \cdots k_2$ from Unif $(0.1, 1)$. We then generate $\mathbf{\Sigma}_{i,\mathrm{H}}, i = 1 \cdots k_1$ that has diagonal elements $(\mathbf{\Sigma}_{i,\mathrm{H}})_{ii} = \sigma_{i,\mathrm{H}}^2$ and the off-diagonal elements $(\mathbf{\Sigma}_{i,\mathrm{H}})_{ij} = \rho_i{}^{|i-j|}\sigma_{i,\mathrm{H}}^2$ where we generate $\sigma_{i,\mathrm{H}}^2, i = 1 \cdots k_1$ from Unif $(0.1, 1)$ and take $\rho_i = 0.1$ for all $i = 1, \cdots, k_1$. The error covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ is generated and rescaled to ensure that the process is stable with signal-to-noise ratio SNR $= 2$. For all the models the initial activity graph $\mathcal{G}_0$ is selected based on an l_1 penalized least squares estimate that does not use $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$.

Algorithm details: For each data set generated, we apply the BMF approach described in the algorithm in the Supplement (Sec. S2.4). To perform Gibbs sampling, we use 1000 burn-in and 2000 further iterations. For entries of $\mathbf{A}$, those which are estimated as zero more than 1000 times out of 2000 main iterations, are set as zero in final estimates; otherwise we use their posterior mean as the final estimate which is calculated by taking the

average over those iterations where their estimates had a non-zero value.

Model Selection results: Sensitivity (SN) and specificity (SP) are the criteria to evaluate the performance of support recovery for $\mathbf{A}$:

Sensitivity (SN) $= \frac{\text{True Positive (TP)}}{\text{TP+False Negative (FN)}}$ Specificity (SP) $= \frac{\text{True Negative (TN)}}{\text{TN+False Positive (FP)}}$.

The selection performance of the BMF model under the aforementioned settings is illustrated in Table 2 of the Supplement (Sec. S6.4). These measures are also separately reported for each sub-block of $\mathbf{A}$. It can be seen that both SN and SP improve as sample size increases, an expected result. Further, for most settings considered these metrics take values close to 1, except settings with a large number of high-frequency variables.

Estimation Consistency: We use the Relative Error $= \left\|\mathbf{A}_0 - \hat{\mathbf{A}}\right\|_{\mathrm{F}} / \left\|\mathbf{A}_0\right\|_{\mathrm{F}}$ as a measure of estimation quality of the transition matrix $\mathbf{A}$ and its sub-blocks. We report the relative error of $\mathbf{A}$ and $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$ along with the estimated θ for BMF in Table 3 of the Supplement (Sec. S6.4). It can be seen that the estimation error decreases with an increase in sample size n. Further, performance is better for smaller model dimension (k_1, k_2). Finally, the estimates of θ are very well calibrated.

6.1 Nowcasting/forecasting Performance in simulations

Comparisons of BMF with (a) the state space model of Schorfheide and Song (2015) (MFBVAR henceforth) and (b) MIDAS regression models of

Ghysels et al. (2007) (implemented in R-packages `mfbvar` and `midasr` respectively) are provided next, under different data generation mechanisms.

Data generated according to BMF: We first generate data $\boldsymbol{y}^1, \cdots, \boldsymbol{y}^{T+H}$ (henceforth denoted as $\boldsymbol{y}^{1:T+H}$) from BMF, where H denotes the maximum forecast horizon, under the aforementioned 5 different settings using the data generation mechanism in Section 6. We use the smallest sample size corresponding to each setting for performing all the nowcasting/forecasting exercises in the paper. The model is trained on data $\boldsymbol{y}^{1:T}$ until the end of quarter T, and the $\boldsymbol{y}^{T+1:T+H}$ portion is used to evaluate its predictive ability. The model parameters are estimated using the Gibbs sampler developed in Section S2.4. The Gibbs sampling draws are then used to compute the posterior predictive distribution of $\boldsymbol{y}$ and its median, denoted by $\hat{\boldsymbol{y}}^{T+h}$, is used as an estimate for $\boldsymbol{y}^{T+h}$ for any particular forecast horizon h. To leverage the contemporaneous relationships for posterior predictive distribution evaluation, we also obtain a sparse estimate of the precision matrix of the errors *without* any block-diagonal structure, using the graphical lasso algorithm Friedman et al. (2008). The procedure is described in the Supplement (Sec. S2.6). Also, at the end of the first or second months of quarter $T+1$, when fresh information becomes available for the monthly variables, we use it to obtain mid-quarter forecasts corresponding to $h = 1/3, 2/3$, etc., as

described in Section 5. For the generated data, nowcasts/forecasts for MFBVAR can be obtained under the following 3 choices for the prior distribution for the model parameters: Minnesota (Minn), Steady-state (SS) and Hierarchical steady-state (Hier. SS). As MFBVAR evolves at the monthly level, it can directly generate mid-quarter nowcasts for $h = 1/3, 2/3$, etc. Similarly MIDAS regression models are fitted to the same data and estimated (1) without restricting the parameters (as in U-MIDAS) and using Ordinary Least Squares (OLS) and also (2) with the exponential Almon lag polynomial constraint on parameters and using Non-linear Least Squares (MIDAS Res.). The nowcasts/forecasts of the quarterly variables are obtained using these two variants of MIDAS regression models across the forecast horizon. We consider a random walk model with drift as the *benchmark* model. For each of these six models -BMF, 3 variants of MFBVAR and 2 variants of MIDAS- we compute the Root-Mean-Squared-Error (RMSE) for the vector of quarterly forecasts relative to the Naive VAR model which is a ratio of respective RMSE values, for $h = 1/3, 2/3, 1, 4/3, 5/3, 2$. For compactness, in Table 1 we report the relative RMSE for BMF, MIDAS models and the best performing MFBVAR model corresponding to the Minnesota prior, when $\theta = 0.5$. The results for $\theta = 0.2, 0.8$ are very similar to the results for $\theta = 0.5$ for all simulation settings examined in the paper and hence we skip

them. All the forecasting exercises in the paper are performed using the smaller sample size for each setting. The results for the two other priors of MFBVAR are provided in Table 4 of the Supplement (Sec. S6.5). A relative RMSE value smaller than 1 implies that the particular model outperforms the benchmark model. The RMSE values are based on averages over 100 replicates. It can be seen from Table 1 that for all the scenarios, BMF outperforms the best performing MFBVAR model and MIDAS models and for the first 3 settings, MFBVAR has relative RMSEs at least twice that of BMF. We also evaluate the probabilistic forecasts of BMF and MFBVAR in terms of their posterior predictive distribution by calculating continuously ranked probability scores (CRPS) and log predictive scores (LPS), as implemented in the functions `'crps.numeric'` and `'logs.numeric'` of the R-package `scoringRules` respectively. These values are reported in Table 5 of the Supplement (Sec. S6.5). The results show that BMF performs significantly better than MFBVAR even in its evaluation for predictive densities.

We repeat the same forecasting exercise with data generated from the proposed BMF model but when the true error covariance matrix does not have a block-diagonal structure. The results (in Supplement Sec. S6.6) provide empirical evidence of the superior forecasting performance of BMF even when the true error covariance doesn't have a block-diagonal structure.

Table 1: Relative RMSE values (benchmarked to a random walk model with drift) using data generated from the proposed BMF model

	Setting 1 ($k_1 = 3, k_2 = 30$)						Setting 2 ($k_1 = 5, k_2 = 50$)					
	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$
BMF	0.54	0.54	0.53	0.66	0.66	0.66	0.55	0.53	0.53	0.65	0.63	0.63
MFBVAR	1.64	1.47	1.36	1.48	1.37	1.22	2.03	2.03	2.02	1.92	1.87	1.83
U-MIDAS	0.92	0.92	0.92	0.96	0.95	0.95	0.93	0.92	0.92	0.91	0.92	0.93
MIDAS(Res.)	0.92	0.91	0.90	0.92	0.93	0.92	0.64	0.63	0.62	0.66	0.65	0.65
	Setting 3 ($k_1 = 10, k_2 = 50$)						Setting 4 ($k_1 = 20, k_2 = 20$)					
	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$
BMF	0.51	0.49	0.47	0.58	0.58	0.55	0.80	0.64	0.61	0.95	0.70	0.70
MFBVAR	1.39	1.31	1.27	1.12	1.07	1.02	0.99	0.92	0.85	0.87	0.80	0.80
U-MIDAS	0.59	0.57	0.57	0.57	0.56	0.57	1.35	1.38	1.35	1.48	1.41	1.58
MIDAS(Res.)	0.57	0.57	0.55	0.54	0.54	0.54	0.90	0.92	0.86	1.00	0.99	1.00
	Setting 5 ($k_1 = 30, k_2 = 10$)											
	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$						
BMF	0.69	0.67	0.66	0.67	0.68	0.67						
MFBVAR	0.92	0.82	0.77	0.71	0.69	0.71						
U-MIDAS	1.13	1.18	1.13	0.99	1.09	1.10						
MIDAS(Res.)	0.79	0.82	0.81	0.92	0.85	0.86						

Data generated by Ghysels (2016)'s model: The forecasting exercise is undertaken under a neutral data generation setting, wherein data are generated from a MIDAS model and the true coefficient matrix $\mathbf{W}$ uses the structure specified in (2.9) of their paper. We examine and compare the 6 models previously described using data up to time T, together with two other versions of BMF that use information on monthly variables until time $T+1/3$ or $T+2/3$ in order to provide nowcasts corresponding to $h = 1/3, 2/3$ and so forth as discussed in Section 5. Using the parameter estimates obtained from these models, we predict $\boldsymbol{y}$ for the future time points and subsequently obtain RMSE values across the forecasting horizon in a similar

manner as discussed for Table 1. For compactness of presentation, in Table 2 we report the relative RMSE values for BMF, MIDAS models and the best performing MFBVAR model, for $h = 1/3, 2/3, 1, 4/3, 5/3$ and 2. It turns out that for $(k_1, k_2) = (3, 30)$ the hierarchical steady state prior works the best, whereas for the other settings the Minnesota prior is the best performing prior. The results show that BMF performs extremely well compared to the best performing MFBVAR model as well as the MIDAS models, even in the neutral setting across all combinations of (k_1, k_2). As before, we obtain the CRPS and LPS values, and BMF again outperforms MFBVAR. These results along with relative RMSE values for MFBVAR with other priors are provided in Table 8 and 9 of the Supplement (Sec. S6.7).

Data generated by Schorfheide and Song (2015)'s model: We perform a similar forecasting exercise in another neutral setting, where data are generated from the state space model. The details of the data generation process and the corresponding results are provided in the Supplement (Sec. S6.8). The results depict that even in this setting BMF outperforms other competing models significantly across all forecasting horizons. Overall, the proposed BMF model exhibits a strong performance compared to other methods across all the true data generation mechanisms considered

Table 2: Relative RMSE values (benchmarked to a random walk model with drift) using data generated under a neutral data generating setting

	Setting 1 ($k_1 = 3, k_2 = 30$)						Setting 2 ($k_1 = 5, k_2 = 50$)					
	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$
BMF	0.68	0.70	0.70	0.73	0.73	0.72	0.73	0.73	0.73	0.70	0.70	0.70
MFBVAR	2.45	2.24	2.27	2.26	2.18	2.16	2.71	2.56	2.52	2.24	2.28	2.20
U-MIDAS	0.90	0.90	0.90	0.90	0.93	0.93	0.83	0.83	0.83	0.90	0.89	0.88
MIDAS(Res.)	0.84	0.83	0.83	0.91	0.91	0.91	0.89	0.88	0.88	0.86	0.85	0.85
	Setting 3 ($k_1 = 10, k_2 = 50$)						Setting 4 ($k_1 = 20, k_2 = 20$)					
	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$
BMF	0.71	0.71	0.71	0.72	0.72	0.72	0.62	0.63	0.63	0.67	0.67	0.67
MFBVAR	2.36	2.21	2.09	2.12	2.05	1.91	1.24	1.01	0.93	0.89	0.84	0.79
U-MIDAS	0.88	0.89	0.89	0.93	0.92	0.93	1.45	1.70	1.52	1.74	1.64	1.67
MIDAS(Res.)	0.86	0.86	0.85	0.90	0.88	0.89	1.02	1.11	1.04	1.07	1.08	1.13
	Setting 5 ($k_1 = 30, k_2 = 10$)											
	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 5/3$	$h = 2$						
BMF	0.60	0.60	0.60	0.73	0.73	0.73						
MFBVAR	0.73	0.67	0.62	0.77	0.72	0.73						
U-MIDAS	1.07	1.09	1.17	1.28	1.24	1.23						
MIDAS(Res.)	0.89	0.88	0.83	0.94	0.91	0.92						

in the simulation experiments .

7. Application to Macroeconomic data

We apply BMF to a macroeconomic data set (Data 3 henceforth) comprising of 77 monthly and 9 quarterly variables on industrial production, the status of the labor force (participation and unemployment), the price index, assets and liabilities of households and business, monetary policy and the financial markets of the Unites States. The collection of 77 monthly variables is

The codes used for all the simulations and empirical data analysis presented in the paper are available at https://github.com/nchak431/BMF.

obtained from Ankargren and Jonéus (2019) by discarding those monthly variables which were discontinued or had significant missing data issues. The data are obtained from the FRED-MD (McCracken and Ng (2016)) and FRED-QD (McCracken and Ng (2020)) databases for January 1960 to December 2016 and the quarterly and monthly variables used in Data 3 are listed in Tables 13 and 14 of the Supplement. The time series are processed to ensure their stationarity following the recommendations in McCracken and Ng (2016, 2020) (see details in Table 15 of the Supplement).

BMF is fitted with lag $d = 1$ as the partial autocorrelation function (PACF) of most individual series exhibited strong first lag autocorrelation and higher order autocorrelations were quite weak. We use a non-informative (flat) prior distribution on the edge selection probability q in the Gibbs sampler, and then sample q in each iteration from its full conditional Beta distribution. For other hyper-parameters the default choices were used as discussed at the end of Section 3.1. We used 1000 burn-in and 2000 further iterations of the Gibbs sampler for posterior computation. The estimated connectivity pattern between the time series under consideration is presented in Figure 4 of the Supplement (Sec. S7), with the vertices of the network corresponding to the 86 macroeconomic indicators and the edges capturing Granger causal effects. The edge density of the estimated

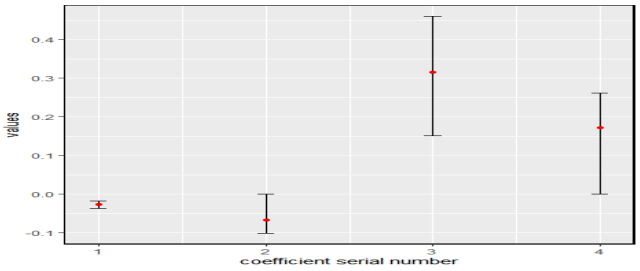

Figure 1: 95% posterior credible intervals by Bayesian MF model for 4 selected edges of the estimated transition matrix $\mathbf{A}$.

transition matrix is given in Table 16 (Sec. S7). Next, we examine 95% posterior credible intervals of 4 selected edges of the estimated transition matrix, one from each sub-block of $\mathbf{A}$, which is depicted in Figure 1. Their final estimates are marked by the red circle. The four intervals correspond to the effect of (i) total reserves of depository institutions on the consumer price index, (ii) total assets for households and nonprofit organizations on securities in bank credit, (iii) real personal consumption expenditures on GDP and (iv) real government consumption expenditures and gross investment on real hourly compensation for all employed persons, respectively.

The data are also used for forecasting/nowcasting purposes and comparisons to competing models. We first aim to investigate how the forecast of the key quarterly variables using the proposed BMF model improves, as more monthly indicators are added to it. To this end, we create two data

sets containing subsets of monthly variables from Data 3 and the same 9 quarterly variables. The smallest subset contains 24 monthly variables (Data 1 henceforth) and the second subset contains 49 monthly variables (Data 2 henceforth) obtained after including some more monthly variables in Data 1. The specific monthly variables present in Data 1 and 2 are indicated in Table 14 of the Supplement (Sec. S7). We use all these 3 data sets for nowcasting and track their respective performance. For this exercise, we consider an increasing sequence of estimation samples starting with 1960Q1-2004Q3 to 1960Q1-2016Q4, with the starting point 1960Q1 being fixed. Subsequently we compute forecasts of the quarterly variables across a variety of forecast horizons, $h = 1/3, 2/3, 1, 4/3, 5/3$ and 2 using each of the estimation samples. As discussed in Section 5, we modify BMF to obtain nowcasts for $h = 1/3, 2/3$, etc. To evaluate the forecasting performance of the models, we calculate RMSE values as well as CRPS and logarithmic score for the fitted BMF using all the 3 data sets, taking average over all the estimation samples. Table 3 reports these metrics (averaged over all the quarterly variables) across different horizons $h = 1/3, 2/3, 1, 4/3, 5/3$ and 2 for the 3 data sets. Table 3 shows that the performance of BMF in terms of nowcasting of quarterly variables improves as monthly indicators are included, an expected result and hence the model performs best under

Table 3: RMSE, CRPS and Log score values for comparing the nowcasting performance of BMF with different no. of monthly variables

	Data 1			Data 2			Data 3		
	RMSE	CRPS	log score	RMSE	CRPS	log score	RMSE	CRPS	log score
$h = 1/3$	1.00	0.48	1.10	0.98	0.47	1.07	0.96	0.46	1.05
$h = 2/3$	1.03	0.50	1.13	1.01	0.49	1.12	0.98	0.48	1.09
$h = 1$	1.24	0.57	1.27	1.14	0.55	1.26	1.05	0.52	1.23
$h = 4/3$	1.25	0.60	1.36	1.28	0.60	1.38	1.25	0.59	1.37
$h = 5/3$	1.38	0.62	1.40	1.36	0.62	1.40	1.30	0.60	1.38
$h = 2$	1.38	0.62	1.40	1.39	0.63	1.42	1.39	0.62	1.40

Data 3. Next we compare the forecasting performance of BMF with (a) the state-space based MFBVAR model using the 3 available prior distributions, (b) unrestricted and restricted MIDAS model and (c) a quarterly VAR model, wherein we aggregate monthly data to quarterly level and estimate it at that frequency. We study two quarterly level VAR models that differ on the weighting schemes of aggregating monthly observations; one uses equal weights for all 3 monthly observations in each quarter to obtain quarterly averages, and the other uses skewed weights $(1/2, 1/3, 1/6)$, with a higher one to the most recent monthly observation in each quarter. We estimate the quarterly VAR models using OLS. As earlier, we consider a random walk model with drift as the benchmark. We perform this exercise separately for each of the 3 data sets constructed. For each of the 8 models- BMF, 3 variants of MFBVAR, 2 MIDAS models and 2 quarterly

VAR models- we obtain the forecasts for the quarterly variables and calculate their relative RMSE values wrt the benchmark random walk model, taking the average over all the estimation samples over the period $T_0 =$ 2004Q3 to $T_1 =$ 2016Q4. Note that the MFBVAR and MIDAS implementations in `R` fail to run when the total number of variables $(3k_1 + k_2)$ is greater than the sample size n. This is exactly what happens when $k_1 = 77$ monthly variables are used and hence we can't obtain results for these models for Data 3. Table 4 shows aggregate relative RMSE values (averaged over all the quarterly variables) for all models under consideration across different horizons $h = 1/3, 2/3, 1, 4/3, 5/3$ and 2 for Data 1 and Data 2. It can be clearly seen that BMF outperforms the other ones across the full forecasting horizon and exhibits notably small forecast errors. The results for Data 3 corresponding to BMF and the quarterly VAR models are provided in Table 17 of the supplement.

We also employ the Diebold-Mariano-West (DMW) test (Diebold and Mariano (1995)) for examining differences in predictive accuracy of the proposed BMF model with respect to MFBVAR models for forecasting GDP using Data 1 and 2. The null hypothesis of the test states that both the models have same predictive accuracy while the alternative hypothesis states that BMF has better predictive accuracy compared to the MFBVAR model. As

mentioned, the MFBVAR models fail to run for Data 3. Hence, we can not perform the test for Data 3. The results of the DMW test for Data 1 and 2 are provided in the Supplement (Sec. S7.1). It can be seen from these results that for short-term forecasting horizons $h = 1/3, 2/3, 1$, the proposed BMF model shows strong evidence of superior predictive performance compared to the MFBVAR model with all 3 choices of prior distributions. Further, we compute CRPS and LPS to evaluate probabilistic forecasts for both BMF and MFBVAR using Data 1 and Data 2. Table 5 shows that BMF performs significantly better compared to MFBVAR in terms of evaluation of the proposed framework for its posterior predictive distribution. We also perform a similar forecasting analysis using the first published estimates of the variables instead of using the current vintage, to respect the release calendar as much as possible for the forecasts. The detailed analysis is provided in the Supplement (Sec. S7.2).

Finally, we investigate the performance of our proposed model in predicting recession episodes. To this end, we evaluate real-time forecasts of real GDP for Q4 2008 and Q1 2009 during the 2008–2009 Great Financial Crisis, using information on key monthly and quarterly variables from 1960 to 2008. We discard those monthly variables from Data 3 for which the data for the last time point in the training data set was not available within two

Table 4: Relative RMSE values for comparing the nowcasting/forecasting performance of competing models

Data 1	BMF	MFBVAR(Minn)	MFBVAR(SS)	MFBVAR(Hier. SS)	U-MIDAS	MIDAS(Res.)	QVAR(Equal wt.)	QVAR(Skewed)
$h = 1/3$	0.66	1.01	1.22	1.04	1.37	1.21	0.80	0.92
$h = 2/3$	0.68	0.94	1.01	0.95	1.12	0.89	0.78	0.74
$h = 1$	0.82	1.00	1.07	1.01	1.28	1.10	0.97	1.04
$h = 4/3$	0.73	0.81	0.87	0.83	1.24	1.04	0.78	0.78
$h = 5/3$	0.80	0.89	0.97	0.91	1.04	0.81	0.78	0.78
$h = 2$	0.80	0.89	0.99	0.93	1.15	0.94	0.94	1.08
Data 2								
$h = 1/3$	0.65	1.02	1.07	1.12	2.22	1.20	0.92	0.93
$h = 2/3$	0.67	0.95	0.96	1.07	2.06	1.13	0.94	0.87
$h = 1$	0.75	0.97	1.00	1.03	2.13	1.24	1.04	1.04
$h = 4/3$	0.74	0.81	0.84	0.87	1.95	1.09	0.87	0.83
$h = 5/3$	0.79	0.88	0.94	0.94	1.85	1.04	0.85	0.86
$h = 2$	0.81	0.88	0.92	0.94	1.94	1.09	1.00	1.03

Table 5: Nowcasting/forecasting performance of BMF and MFBVAR using CRPS and LPS values

CRPS	Data 1		Data 2	
	BMF	MFBVAR	BMF	MFBVAR
$h = 1/3$	0.48	0.70	0.47	0.73
$h = 2/3$	0.50	0.67	0.49	0.68
$h = 1$	0.57	0.70	0.55	0.70
$h = 4/3$	0.60	0.67	0.60	0.69
$h = 5/3$	0.62	0.70	0.62	0.70
$h = 2$	0.62	0.70	0.63	0.70
Log score				
$h = 1/3$	1.10	1.44	1.07	1.48
$h = 2/3$	1.13	1.42	1.12	1.44
$h = 1$	1.27	1.47	1.26	1.49
$h = 4/3$	1.36	1.48	1.38	1.50
$h = 5/3$	1.40	1.48	1.40	1.50
$h = 2$	1.40	1.49	1.42	1.49

months of the actual date, for the sake of real-time forecasts. Thus, we end up with a data set of 42 monthly and 7 quarterly variables and use them to obtain from 6-month to 1-month ahead forecasts of GDP growth for Q4 2008 and Q1 2009. The main objective is to see when the models start predicting downturns in economic activity, as fresh information (initial estimates) becomes available for the relevant variables. Table 6 provides predictions of GDP growth by both BMF and MFBVAR for Q4 2008 and Q1 2009. We compare these forecasts with the latest available estimate of true GDP growth, as well as the initial GDP estimate published by FRED for these two quarters and provide squared errors w.r.t both these versions of true GDP values in Table 6. For each quarter the values corresponding to

specific month names in Table 6 denotes the prediction/error using training data till the end of that month. This analysis indicates that MFBVAR performs better for 4 to 6 months ahead forecasts, while BMF provides more accurate 1 to 3 months ahead forecasts, as we approach closer to the time point of interest. Also note that we observe a consistent decreasing trend in the predictions, as we keep on adding new monthly data and the forecast horizon becomes smaller (both for Q4 2008 and Q1 2009 predictions). The only exception to this phenomenon is an increase in the predicted GDP growth (for both BMF and MFBVAR) for Q4 2008 when we add data for November 2008. We closely examined the November 2008 values for the monthly variables used in the model, and found that the value of these 3 variables, 'FEDFUNDS', 'TB3MS', 'TOTRESNS', changed significantly in November compared to their historical values. Also, the Federal Reserve slashed its rate to practically 0 around this time and the 3-month bill followed suit. These are probable causes for the above noted discrepancy.

Table 6: Performance of BMF and MFBVAR in predicting 2008–2009 recession

	Q1 2009(Final Truth=-1.96,Initial Truth=-2.36)						Q4 2008(Final Truth=-3.01,Initial Truth=-1.76)					
	Sep'08	Oct'08	Nov'08	Dec'08	Jan'09	Feb'09	June'08	Jul'08	Aug'08	Sep'08	Oct'08	Nov'08
Prediction by BMF	-0.73	-1.61	-1.81	-1.98	-2.72	-3.00	-0.33	-0.30	-0.36	-2.20	-2.31	-0.89
Prediction by MFBVAR	-0.78	-1.67	-2.73	-5.26	-6.02	-4.08	-0.68	-0.74	-0.97	-1.48	-1.23	-0.05
Error (BMF, using final truth)	1.52	0.12	0.02	0.0004	0.58	1.08	7.14	7.32	7.02	0.65	0.48	4.46
Error (MFBVAR, using final truth)	1.41	0.09	0.59	10.89	16.46	4.49	5.42	5.16	4.16	2.33	3.14	9.35
Error (BMF, using initial truth)	2.66	0.56	0.30	0.14	0.13	0.41	2.03	2.13	1.97	0.20	0.30	0.75
Error (MFBVAR, using initial truth)	2.51	0.48	0.14	8.42	13.39	2.96	1.17	1.05	0.63	0.08	0.28	3.28

Acknowledgements

The authors would like to thank the AE and two anonymous referees for their useful comments and suggestions. The work of George Michailidis and Kshitij Khare was supported in part by NSF grant DMS 1821220.

References

Ankargren, S. and P. Jonéus (2019). Estimating large mixed-frequency bayesian var models. *arXiv preprint arXiv:1912.02231*.

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Nowcasting and the real-time data flow. handbook of economic forecasting, 2, g. elliott and a. timmermann eds.

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

Basu, S. and G. Michailidis (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics 43*(4), 1535–1567.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Realtime nowcasting with a bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society. Series A,(Statistics in Society) 178*(4), 837.

Cross, J., C. Hou, and A. Poon (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *Int. Jour. Forecasting 36*(3), 899–915.

Diebold, F. X. and R. S. Mariano (1995). Com paring predictive accu racy. *Journal of Business and Economic Statistics 13*(3), 253–263.

Eraker, B., C. W. Chiu, A. T. Foerster, T. B. Kim, and H. D. Seoane (2014). Bayesian mixed frequency vars. *Journal of Financial Econometrics 13*(3), 698–721.

Foroni, C. and M. Marcellino (2014). Mixed-frequency structural models: identification, estimation, and policy analysis. *Journal of Applied Econometrics 29*(7), 1118–1144.

Friedman, J., T. Hastie, and R. Tibshirani (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*(3), 432–441.

Gefang, D., G. Koop, and A. Poon (2020). Computationally efficient inference in large bayesian mixed frequency vars. *Economics Letters 191*, 109120.

Ghosh, S., K. Khare, and G. Michailidis (2021). Strong selection consistency of bayesian vector autoregressive models based on a pseudo-likelihood approach. *Annals of Statistics*.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics 193*(2), 294–314.

Ghysels, E., A. Sinko, and R. Valkanov (2007). Midas regressions: Further results and new directions. *Econometric reviews 26*(1), 53–90.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity.

Kilian, L. and H. Lütkepohl (2017). *Structural vector autoregressive analysis*. CUP.

Leeper, E. M., C. A. Sims, T. Zha, R. E. Hall, and B. S. Bernanke (1996). What does monetary policy do? *Brookings Papers on Economic Activity 1996*(2), 1–78.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

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

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

McCracken, M. and S. Ng (2020). Fred-qd: A quarterly database for macroeconomic research. Technical report, National Bureau of Economic Research.

McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

McCracken, M. W., M. Owyang, and T. Sekhposyan (2015). Real-time forecasting with a large, mixed frequency, bayesian var. *FRB St. Louis Working Paper* (2015-30).

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

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European economic review 36*(5), 975–1000.

Uematsu, Y. and S. Tanaka (2019). High-dimensional macroeconomic forecasting and variable selection via penalized regression. *The Econometrics Journal 22*(1), 34–56.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Authors affiliation: Department of Statistics, University of Florida; E-mail: {nchakraborty,kdkhare,gmichail}@ufl.edu

Statistica Sinica