High-Dimensional Conditionally Gaussian State Space Models with Missing Data*

Joshua C. C. Chan
Purdue University

Aubrey Poon
Örebro University

Dan Zhu
Monash University

May 2023

Abstract

We develop an efficient sampling approach for handling complex missing data patterns and a large number of missing observations in conditionally Gaussian state space models. Two important examples are dynamic factor models with unbalanced datasets and large Bayesian VARs with variables in multiple frequencies. A key observation underlying the proposed approach is that the joint distribution of the missing data conditional on the observed data is Gaussian. Furthermore, the inverse covariance or precision matrix of this conditional distribution is sparse, and this special structure can be exploited to substantially speed up computations. We illustrate the methodology using two empirical applications. The first application combines quarterly, monthly and weekly data using a large Bayesian VAR to produce weekly GDP estimates. In the second application, we extract latent factors from unbalanced datasets involving over a hundred monthly variables via a dynamic factor model with stochastic volatility.

JEL classification: C11, C32, C55

Keywords: mixed-frequency, unbalanced panel, vector autoregression, dynamic factor model, stochastic volatility

*A previous version of the paper was circulated under the title "Efficient Estimation of State-Space Mixed-Frequency VARs: A Precision-Based Approach". We thank Otilia Boldea, Francis Diebold, Gary Koop, James Mitchell, Michele Piffer, as well as Editor Serena Ng, an anonymous Associate Editor and two anonymous reviewers, for many constructive comments and valuable suggestions.

1 Introduction

Large-scale time-series models are increasingly used in empirical macroeconomics to exploit the wide availability of large datasets. This trend promises a more timely and comprehensive analysis but also brings new challenges. First, large datasets are typically compiled from multiple sources, and consequently, they often involve complex missing data patterns. One prominent example is mixed-frequency data to incorporate real-time information, as opposed to the traditional approach of using only variables at the same (lower) frequency. For instance, Aruoba, Diebold, and Scotti (2009) combine daily, weekly, monthly and quarterly variables to construct a business conditions index to track economic activity. Schorfheide and Song (2015) use GDP, which is only available quarterly, and other quarterly and monthly variables to obtain GDP estimates at the monthly frequency. In both cases, the high-frequency observations of the low-frequency variables are treated as missing data. As such, there are a large number of missing observations.

Second, extracting information from large datasets generally requires large-scale time-series models. Factor models have been the workhorse for this purpose, and thanks to the seminal work of Bańbura, Giannone, and Reichlin (2010) and Koop (2013), large Bayesian vector autoregressions (VARs) have now become a popular alternative. In addition, since there is a large body of empirical evidence that shows allowing for various flexible features, such as heteroskedasticity, heavy-tailed distributions and outliers detection, are vitally important for improving in-sample model-fit, and out-of-sample forecast performance (see, e.g., Clark, 2011; D'Agostino, Gambetti, and Giannone, 2013; Cross and Poon, 2016; Stock and Watson, 2016; Chan, 2020), these features are increasingly incorporated into dynamic factor models and large Bayesian VARs (recent examples include Kastner and Huber, 2020; Antolin-Diaz, Drechsel, and Petrella, 2021; Carriero, Clark, Marcellino, and Mertens, 2022). While there are many recent advances in speeding up the estimation of these flexible large-scale models with complete data, efficient algorithms that can handle complex missing data patterns with a large number of missing observations are by comparison underdeveloped.

We tackle these challenges by developing an efficient sampling approach for drawing all the missing observations in one step. To make our approach widely applicable, it is developed under a general framework of conditionally Gaussian state space models. As such, it applies to many of the popular large-scale models, such as dynamic factor models with

stochastic volatility or mixed-frequency VARs with non-Gaussian errors. In addition, the setup can easily handle a wide variety of complex missing data patterns, including unbalanced panels, mixed-frequency settings, and a 'ragged edge' at the end of the sample due to non-synchronous data releases. Thanks to the modular nature of Markov chain Monte Carlo (MCMC) methods, the proposed approach can be straightforwardly implemented in conjunction with any efficient samplers for conditionally Gaussian state space models with complete data. This paper therefore complements existing works on fast estimation of flexible large-scale models and extend them to missing data settings.

We first derive the mean and precision (i.e., inverse covariance) matrix of the joint distribution of the missing observations conditional on the observed data (as well as the model parameters and latent variables), which fully characterize the Gaussian distribution. Furthermore, we show that the precision matrix of this conditional distribution is sparse—in fact, for many of the common missing data patterns, it is banded, i.e., it is sparse, and its non-zero elements are arranged along a diagonal band. These special structures can be exploited to vastly speed up computations. In particular, the precision-based sampler of Chan and Jeliazkov (2009) can be applied to draw all the missing observations in one step. This approach is much more efficient compared to standard Kalman-filter-based methods, especially when there are a large number of missing observations or when the state vector is high-dimensional.[1] In addition, it is straightforward to implement: one only needs to represent the data vector into observed and missing data by defining some appropriate selection matrices. The proposed approach can easily handle many complex missing data patterns, such as settings with variables in multiple frequencies.

In addition, our setup can also accommodate settings in which additional information is available to help sharpen inference on the missing observations. This feature is crucial in mixed-frequency applications where linear combinations of high-frequency missing obser-

[1]The precision-based sampling approach of Chan and Jeliazkov (2009) and McCausland, Miller, and Pelletier (2011) are designed for linear Gaussian state space models with complete data. It builds upon earlier work on Gaussian Markov random fields (Rue, 2001) and nonparametric regression (Chib and Jeliazkov, 2006; Chib, Greenberg, and Jeliazkov, 2009). Due to its ease of implementation and computational efficiency, this approach is increasingly used in a wide range of empirical applications. Recent examples include modeling trend inflation (Chan, Koop, and Potter, 2013; Chan, 2017; Hou, 2020), time-varying Phillips curve (Fu, 2020) and dividend growth (Pettenuzzo, Sabbatucci, and Timmermann, 2023); estimating the output gap (Grant and Chan, 2017a,b); macroeconomic forecasting (Cross and Poon, 2016; Cross, Hou, and Poon, 2019); and fitting various moving average models (Chan, 2013; Chan, Eisenstat, and Koop, 2016; Dimitrakopoulos and Kolossiatis, 2020; Zhang, Chan, and Cross, 2020) and dynamic factor models (Kaufmann and Schumacher, 2019; Beyeler and Kaufmann, 2021).

vations need to be mapped to the observed values of low-frequency variables. Our paper is related to the recent work by Eckert, Kronenberg, Mikosch, and Neuwirth (2020) and Hauber and Schumacher (2021), who also consider a precision-based sampling approach for settings with missing observations. However, they focus on dynamic factor models and the latter does not consider mixed-frequency settings. In contrast, our approach is more general and is applicable to any conditionally Gaussian state space models under a wide variety of missing data patterns.

We conduct a series of Monte Carlo experiments to illustrate the numerical accuracy and computation speed of the proposed precision-based approach. In particular, we estimate mixed-frequency VARs using the proposed samplers and standard filtering methods under a variety of settings. We show that the proposed precision-based approach is much more computationally efficient compared to traditional Kalman-filter-based methods. In addition, it scales well to high-dimensional settings, making it possible to estimate VARs with a large number of low-frequency variables.

To demonstrate the versatility of the proposed precision-based approach, we consider two empirical macroeconomic applications with widely different missing data patterns. In the first application, we use a large mixed-frequency Bayesian VAR with stochastic volatility to generate weekly estimates of real GDP. These high-frequency GDP estimates are useful for a range of purposes, such as monitoring the current state of the economy and delivering timely nowcasts of key macroeconomic variables. To obtain the weekly GDP estimates, we fit a Bayesian VAR using 22 variables in 3 different frequencies: 16 weekly variables, 5 monthly variables and the quarterly real GDP. All variables are modeled at the weekly frequency, and the weekly observations of the monthly and quarterly variables are treated as missing data. Even though the missing data pattern is complex—e.g., there are different numbers of weeks in different months and quarters—and there are a large number of missing observations, the proposed approach is computationally efficient and easy to implement.

In the second application, we use a dynamic factor model with stochastic volatility to extract latent factors in real-time from the FRED-MD datasets of McCracken and Ng (2016). Each data vintage of FRED-MD contains 128 monthly variables, but many have missing values from two sources: missing observations at the beginning of the sample for some recently constructed variables and missing values at the end of the sample due

to publication lags. We implement the proposed approach to sample the missing observations under the dynamic factor model and obtain the latent factors. Our results show that the first factor tracks the broad economic conditions well, even during the pronounced downturn at the onset of the COVID-19 pandemic and the subsequent rebound. In addition, our results suggest that using only variables without missing values can potentially misrepresent the dynamics of the latent factors, highlighting the importance of incorporating the information from variables with missing values.

The remainder of the paper is organized as follows. Section 2 discusses the proposed precision-based sampling approach for drawing the missing observations in a general state space framework. Section 3 conducts a series of Monte Carlo experiments comparing the proposed sampling approach against standard Kalman-filter based techniques in a variety of mixed-frequency settings. Section 4 demonstrates how the proposed sampling approach can be applied to two popular empirical macroeconomic applications. Finally, Section 5 concludes.

2 A General State Space Framework

This section introduces the proposed precision-based approach for sampling the missing data conditional on a variety of information sets under a general state space framework. More specifically, we first derive the joint conditional distribution of the missing data given the observed data and model parameters, which we show is Gaussian. We then discuss an efficient algorithm to generate samples from this typically high-dimensional Gaussian distribution. In addition, since in many applications, such as mixed-frequency settings, one has additional information on the missing data, we demonstrate how this additional information can be incorporated to update the conditional distribution of the missing data.

2.1 The Conditional Distribution of the Missing Data

Our general setup is the following conditionally Gaussian state space model for an $n \times 1$ vector of variables $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ over $t = 1, \ldots, T$:

$$\mathbf{y}_t = \mathbf{W}_t \boldsymbol{\alpha}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_n, \boldsymbol{\Sigma}_t), \tag{1}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\gamma} + \boldsymbol{\Phi}_1 \boldsymbol{\alpha}_{t-1} + \cdots + \boldsymbol{\Phi}_q \boldsymbol{\alpha}_{t-q} + \boldsymbol{\varepsilon}_t^{\boldsymbol{\alpha}}, \qquad \boldsymbol{\varepsilon}_t^{\boldsymbol{\alpha}} \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{\Omega}_t), \tag{2}$$

where $\mathbf{0}_m$ denotes an $m \times 1$ vector of zeros, $\boldsymbol{\beta}$ is a vector of time-invariant parameters, $\boldsymbol{\alpha}_t$ is a vector of time-varying parameters, $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\Omega}_t$ are the covariance matrices for the observation and state equations, respectively. The covariate matrices $\mathbf{W}_t$ and $\mathbf{X}_t$ could include lagged values of $\mathbf{y}_t$. This framework encompasses a wide range of commonly-used models, including dynamic factor models and vector autoregressions.[2]

Note that it also includes many different types of error processes as special cases. For instance, one can specify $\boldsymbol{\Sigma}_t$ as the multivariate stochastic volatility processes in Cogley and Sargent (2005), Primiceri (2005), Carriero, Clark, and Marcellino (2016) or Kastner (2019). In addition, one can also consider various types of non-Gaussian errors, such as the t distribution by setting $\boldsymbol{\Sigma}_t = \lambda_t \mathbf{Q}$, where $\mathbf{Q}$ is a covariance matrix and $\lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$, or an outlier component of the type in Stock and Watson (2016) by specifying $\boldsymbol{\Sigma}_t = o_t^2 \mathbf{Q}$, where o_t follows a 2-part distribution with a point mass at 1 and a uniform distribution on the interval $(2, 10)$. Naturally, any combination of the above multivariate stochastic volatility processes or non-Gaussian errors, such as those in Chan (2020) and Carriero, Clark, Marcellino, and Mertens (2022), is also possible.

We are interested in settings in which some elements of $\mathbf{y}_t$ are missing. More specifically, let $\mathbf{y}_t^o$ denote the $n_t^o \times 1$ subvector of observed values in $\mathbf{y}_t$, and let $\mathbf{y}_t^m$ represent the $n_t^m \times 1$ subvector of missing values in $\mathbf{y}_t$ such that $n_t^o + n_t^m = n$. Note that here n_t^o and

[2]We are interested in settings in which some elements of $\mathbf{y}_t$ are missing. Throughout this paper we assume that missingness depends on the data only through the observed subvector in $\mathbf{y}_t$, or what is known as the 'missing at random' assumption (Little and Rubin, 2019). For other approaches that do not assume missing at random, see Bai and Ng (2021) and the references therein. Under this missing at random setting, one can relax the assumption of serially independent measurement errors $\boldsymbol{\varepsilon}_t$ in (1) by re-defining the state vector $\boldsymbol{\alpha}_t$ to include the serially correlated measurement errors. Alternatively, one can incorporate this serial correlation by deriving the covariance matrix of the joint data $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')$, as is done in the application in Section 4.2 involving a dynamic factor model with autoregressive measurement errors. As shown in Ng and Scanlan (2023), allowing for dynamics in measurement errors can provide substantially more accurate imputed values of the missing data.

n_t^m can be time-varying, and hence this setup can accommodate a wide range of missing data patterns, such as unbalanced panels and ragged edge. In addition, for settings with variables of mixed frequencies, it is common to express the time index in the highest frequency and treat some of the low-frequency variables as missing. For example, in models with both monthly and quarterly variables, the monthly values of the quarterly stock variables are only observed every 3 months and the rest are treated as missing.[3] Finally, let $N^o = \sum_{t=1}^T n_t^o$ and $N^m = \sum_{t=1}^T n_t^m$ denote the total numbers of observed and missing values with $N^o + N^m = Tn$. For later reference, stack $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')' \in \mathbb{R}^{Tn}$, $\mathbf{y}^o = (\mathbf{y}_1^{o\prime}, \ldots, \mathbf{y}_T^{o\prime})' \in \mathbb{R}^{N^o}$ and $\mathbf{y}^m = (\mathbf{y}_1^{m\prime}, \ldots, \mathbf{y}_T^{m\prime})' \in \mathbb{R}^{N^m}$ vectors.

In principle one could rewrite the measurement equation (1) in terms of only the observed values $\mathbf{y}^o$. However, this approach would typically result in a non-standard state space model that is difficult to estimate efficiently. One popular alternative is data augmentation, i.e., treating the missing values $\mathbf{y}^m$ as latent variables to be augmented or sampled. Once we obtain $\mathbf{y}^m$ from its conditional distribution given the observed values $\mathbf{y}^o$, $\mathbf{y}$ is completely observed and we can then leverage the many efficient methods for estimating large-scale state space models for complete data. This crucial step of sampling $\mathbf{y}^m$ is typically done using standard Kalman filtering and smoothing algorithms. However, the main drawback of this approach is that it tends to be computationally intensive in high-dimensional settings when there are a large number of missing observations. This significant computational burden is a key obstacle in practice for using high-dimensional state space models with missing data, despite the increasing popularity of large-scale dynamic factor models and VARs. In addition, when the missing data pattern is complex, the implementation of Kalman filter based algorithms also becomes more cumbersome. To overcome these computational and implementation issues, we develop an efficient method to jointly sample $\mathbf{y}^m$ given $\mathbf{y}^o$, the latent Gaussian states $\boldsymbol{\alpha}$, and model parameters and other latent variables, which we denote as $\boldsymbol{\theta}$. The proposed method is conceptually simple and easy to implement, even with complex missing data patterns.

In what follows, we first derive the joint conditional distribution of $\mathbf{y}^m$ given $\mathbf{y}^o$. To that end, we write $\mathbf{y}$ in terms of $\mathbf{y}^o$ and $\mathbf{y}^m$:

$$\mathbf{y} = \mathbf{S}^o \mathbf{y}^o + \mathbf{S}^m \mathbf{y}^m, \tag{3}$$

[3]For flow variables, their observed values can be viewed as additional information that can be mapped to the missing high-frequency values; this case will be further discussed in the next subsection.

where $\mathbf{S}^o$ and $\mathbf{S}^m$ are, respectively, $Tn \times N^o$ and $Tn \times N^m$ selection matrices. In particular, $[\mathbf{S}^o, \mathbf{S}^m]$ is a permutation matrix that has exactly one entry of 1 in each row and each column, and 0 elsewhere. It follows that both $\mathbf{S}^o$ and $\mathbf{S}^m$ are of full column rank.

As a simple illustration, suppose $T = 2, n = 3$, and $y_{3,1}, y_{1,2}$ and $y_{3,2}$ are missing. Then, $\mathbf{y}^o = (y_{1,1}, y_{2,1}, y_{2,2})'$, $\mathbf{y}^m = (y_{3,1}, y_{1,2}, y_{3,2})'$ and

$$\begin{bmatrix} y_{1,1} \\ y_{2,1} \\ y_{3,1} \\ y_{1,2} \\ y_{2,2} \\ y_{3,2} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathbf{S}^o} \begin{bmatrix} y_{1,1} \\ y_{2,1} \\ y_{2,2} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}^m} \begin{bmatrix} y_{3,1} \\ y_{1,2} \\ y_{3,2} \end{bmatrix}.$$

For a second illustration, suppose $y_{1,t}$ is only observed every 3 periods at $t = 3, 6, 9, \ldots$, whereas $y_{2,t}, \ldots, y_{n,t}$ are observed every period for $t = 1, \ldots, T$. Then, $\mathbf{S}^o$ is block-diagonal consisting of diagonal blocks $\mathbf{S}_1^o, \mathbf{S}_2^o, \ldots, \mathbf{S}_T^o$, i.e., $\mathbf{S}^o = \text{diag}(\mathbf{S}_1^o, \mathbf{S}_2^o, \ldots, \mathbf{S}_T^o)$ and $\mathbf{S}^m = \text{diag}(\mathbf{s}_1^m, \mathbf{s}_2^m, \ldots, \mathbf{s}_T^m)$, where $\mathbf{S}_t^o = \mathbf{I}_n$ and $\mathbf{s}_t^m = \emptyset$ if t is divisible by 3; otherwise

$$\mathbf{S}_t^o = \begin{bmatrix} \mathbf{0}'_{n-1} \\ \mathbf{I}_{n-1} \end{bmatrix}, \quad \mathbf{s}_t^m = \begin{bmatrix} 1 \\ \mathbf{0}_{n-1} \end{bmatrix}.$$

In general, the selection matrices $\mathbf{S}^o$ and $\mathbf{S}^m$ are sparse and can be constructed easily even for complex missing data patterns.

Next, we aim to stack (1) over $t = 1, \ldots, T$, and we consider two cases. If the right-hand side of (1) does not contain any lagged values of $\mathbf{y}_t$, then, using the expression in (3), one can rewrite the model more compactly as

$$\mathbf{S}^o\mathbf{y}^o + \mathbf{S}^m\mathbf{y}^m = \mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_{Tn}, \boldsymbol{\Sigma}), \tag{4}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_T')'$ and $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_T)$. If the right-hand side of (1) contains lagged values of $\mathbf{y}_t$ (e.g., in the case of VARs), then, moving the lagged values to the right-hand side and using the expression in (3), the model can be expressed more generally

as

$$\mathbf{G}^o\mathbf{y}^o + \mathbf{G}^m\mathbf{y}^m = \mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_{Tn}, \boldsymbol{\Sigma}), \tag{5}$$

where $\mathbf{G}^o$ and $\mathbf{G}^m$ are appropriately defined according to the specification—typically they are products of certain difference matrices and selection matrices, as illustrated in Example 2.

For both cases, the system can be written in the form in (5). All the algorithms that follow can be directly applied to this stacked system. Furthermore, we assume $\mathbf{G}^m$ has full column rank, which is satisfied for most commonly-used models (and can be easily verified in practice). Below we provide two examples to show how a dynamic factor model and a VAR(p) can be expressed in the form of (5).

Example 1 Consider the following dynamic factor model with stochastic volatility:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{A}_1\mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathcal{N}(\mathbf{0}_n, \boldsymbol{\Sigma}_t), \\ \mathbf{f}_t &= \boldsymbol{\Phi}_1\mathbf{f}_{t-1} + \cdots + \boldsymbol{\Phi}_q\mathbf{f}_{t-q} + \boldsymbol{\varepsilon}_t^{\mathbf{f}}, & \boldsymbol{\varepsilon}_t^{\mathbf{f}} &\sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{\Omega}_t), \end{aligned}$$

where $\boldsymbol{\Sigma}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{n,t}})$, $\boldsymbol{\Omega}_t = \text{diag}(\text{e}^{h_{n+1,t}}, \ldots, \text{e}^{h_{n+k,t}})$ are diagonal matrices with time-varying variances, and $\mathbf{y}_t$ is split into two subvectors $\mathbf{y}_t^o$ and $\mathbf{y}_t^m$. Using the identity in (3), the observation equation of this dynamic factor model can be expressed in the form in (5) as:

$$\mathbf{S}^o\mathbf{y}^o + \mathbf{S}^m\mathbf{y}^m = \mathbf{H}_{\mathbf{A}_1}\mathbf{f} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_{Tn}, \boldsymbol{\Sigma}),$$

where $\mathbf{H}_{\mathbf{A}_1} = (\mathbf{I}_T \otimes \mathbf{A}_1)$, $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_T)$ and $\otimes$ denotes the Kronecker product. One can consider a more general dynamic factor model in which the observation equation contains lagged values of the dynamic factors, say, $\mathbf{f}_{t-1}, \ldots, \mathbf{f}_{t-p}$. In this case one can simply redefine the matrix $\mathbf{H}_{\mathbf{A}_1}$ to include them in the observation equation.

Example 2 The next example is a VAR(p) with an outlier component:

$$\mathbf{y}_t = \mathbf{b}_0 + \mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_2\mathbf{y}_{t-2} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_n, \boldsymbol{\Sigma}_t), \tag{6}$$

where $\boldsymbol{\Sigma}_t = o_t^2\mathbf{Q}$, $\mathbf{Q}$ is a covariance matrix and o_t follows a 2-part distribution with a point mass at 1 and a uniform distribution on the interval $(2, 10)$. Then, stacking (6)

over $t = 1, \ldots, T$, we obtain

$$\mathbf{H_B y} = \mathbf{c_B} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_{Tn}, \boldsymbol{\Sigma}), \tag{7}$$

where

$$\mathbf{c_B} = \begin{bmatrix} \mathbf{b}_0 + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{y}_{1-j} \\ \mathbf{b}_0 + \sum_{j=2}^{p} \mathbf{B}_j \mathbf{y}_{2-j} \\ \vdots \\ \mathbf{b}_0 + \mathbf{B}_p \mathbf{y}_0 \\ \mathbf{b}_0 \\ \vdots \\ \mathbf{b}_0 \end{bmatrix}, \; \mathbf{H_B} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n\times n} \\ -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n\times n} \\ -\mathbf{B}_2 & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & & & \mathbf{0}_{n\times n} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ -\mathbf{B}_p & \cdots & & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & & \vdots \\ \mathbf{0}_{n\times n} & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} & -\mathbf{B}_p & \cdots & -\mathbf{B}_2 & -\mathbf{B}_1 & \mathbf{I}_n \end{bmatrix}. \tag{8}$$

Again, using the identity in (3), we obtain

$$\mathbf{H_B S}^o \mathbf{y}^o + \mathbf{H_B S}^m \mathbf{y}^m = \mathbf{c_B} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_{Tn}, \boldsymbol{\Sigma}),$$

which is in the form of (5) with $\mathbf{G}^o = \mathbf{H_B S}^o$, $\mathbf{G}^m = \mathbf{H_B S}^m$, $\mathbf{X} = \mathbf{I}_{Tn}$ and $\boldsymbol{\beta} = \mathbf{c_B}$.[4]

Using the expression in (5), next we derive the conditional distribution of $\mathbf{y}^m$ given $\mathbf{y}^o$, $\boldsymbol{\alpha}$, and other latent variables and model parameters, which we collectively denote as $\boldsymbol{\theta}$. Intuitively, since the joint distribution of $(\mathbf{y}^o, \mathbf{y}^m)$ is Gaussian conditional on $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$, the conditional distribution of $\mathbf{y}^m$ given $\mathbf{y}^o, \boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ is also Gaussian by the properties of the Gaussian distribution. More precisely, it follows from (5) that $p(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta})$ can be expressed as

$$\begin{aligned} p(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta}) &\propto \exp\left\{ -\frac{1}{2} \left(\mathbf{G}^m \mathbf{y}^m + \mathbf{G}^o \mathbf{y}^o - \mathbf{W}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}\right)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{G}^m \mathbf{y}^m + \mathbf{G}^o \mathbf{y}^o - \mathbf{W}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta}\right) \right\} \\ &\propto \exp\left\{ -\frac{1}{2} \left[\mathbf{y}^{m\prime} \mathbf{G}^{m\prime} \boldsymbol{\Sigma}^{-1} \mathbf{G}^m \mathbf{y}^m - 2 \mathbf{y}^{m\prime} \mathbf{G}^{m\prime} \boldsymbol{\Sigma}^{-1} (\mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} - \mathbf{G}^o \mathbf{y}^o) \right] \right\}. \end{aligned}$$

[4]For notational convenience, in the derivation we condition on the initial conditions $\mathbf{y}_0, \ldots, \mathbf{y}_{1-p}$. These initial conditions could potentially have missing data. There are two ways to deal with these missing values. First, they can be sampled in a separate step conditional on $\mathbf{y}$ and the model parameters. Since p is much smaller than T in most applications, this extra step is typically computationally trivial. Alternatively, one can jointly sample the missing data in the initial conditions and the sample by redefining $\mathbf{y}$ and the associated matrices. The details of this alternative approach are given in Appendix D.

Next, let $\mathbf{K}_{\mathbf{y}^m} = \mathbf{G}^{m\prime}\boldsymbol{\Sigma}^{-1}\mathbf{G}^m$, which is an $N_m \times N_m$ non-singular matrix—since $\mathbf{G}^m$ has full column rank—and is thus invertible. Furthermore, let $\boldsymbol{\mu}_{\mathbf{y}^m} = \mathbf{K}_{\mathbf{y}^m}^{-1}\mathbf{G}^{m\prime}\boldsymbol{\Sigma}^{-1}(\mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} - \mathbf{G}^o\mathbf{y}^o)$. Then, by completing the square in $\mathbf{y}^m$, one can write the conditional distribution of $\mathbf{y}^m$ as

$$\begin{aligned} p(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta}) &\propto \exp\left\{-\frac{1}{2}\left(\mathbf{y}^{m\prime}\mathbf{K}_{\mathbf{y}^m}\mathbf{y}^m - 2\mathbf{y}^{m\prime}\mathbf{K}_{\mathbf{y}^m}\boldsymbol{\mu}_{\mathbf{y}^m}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\mathbf{y}^m - \boldsymbol{\mu}_{\mathbf{y}^m}\right)'\mathbf{K}_{\mathbf{y}^m}\left(\mathbf{y}^m - \boldsymbol{\mu}_{\mathbf{y}^m}\right)\right\}. \end{aligned}$$

Thus, we have shown that the joint conditional distribution of the missing data given the observed data is Gaussian with mean vector $\boldsymbol{\mu}_{\mathbf{y}^m}$ and precision matrix $\mathbf{K}_{\mathbf{y}^m}$:

$$(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}^m}, \mathbf{K}_{\mathbf{y}^m}^{-1}\right). \tag{9}$$

Since both $\mathbf{G}^m$ and $\boldsymbol{\Sigma}$ are band matrices, so is the precision matrix $\mathbf{K}_{\mathbf{y}^m}$. Therefore, we can use the precision-based sampler of Chan and Jeliazkov (2009) to draw $\mathbf{y}^m$ efficiently. We summarize the sampler in Algorithm 1.

Algorithm 1 Sampling $(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}^m}, \mathbf{K}_{\mathbf{y}^m}^{-1}\right)$.

Given $\boldsymbol{\mu}_{\mathbf{y}^m}$ and $\mathbf{K}_{\mathbf{y}^m}$, complete the following steps.

1. Obtain the Cholesky factor $\mathbf{C}$ of $\mathbf{K}_{\mathbf{y}^m}$ such that $\mathbf{K}_{\mathbf{y}^m} = \mathbf{C}\mathbf{C}'$.
2. Solve $\mathbf{C}'\mathbf{v} = \mathbf{x}$ for $\mathbf{v}$ by backward substitution, where $\mathbf{x} \sim \mathcal{N}(\mathbf{0}_{N^m}, \mathbf{I}_{N^m})$.
3. Return $\mathbf{y}^m = \boldsymbol{\mu}_{\mathbf{y}^m} + \mathbf{v}$.

This paper focuses on Bayesian estimation using MCMC methods. But the above results are also useful for other estimation methods. For example, the conditional distribution in (9) can be used in conjunction with the expectation-maximization algorithm to obtain the maximum likelihood estimate of $\boldsymbol{\theta}$.

The setup so far is suitable for applications with missing data patterns such as unbalanced panels and ragged edges. In many situations, however, the researcher has additional information on the missing data. A prominent example is a mixed-frequency model in which the high-frequency observations of the low-frequency flow variables are treated

as missing data, and these missing observations are linked to multiple low-frequency observations. Next, we show how one can incorporate additional information to update the conditional distribution of the missing data.

2.2 Conditioning on Additional Information

The previous section discussed how one can efficiently sample the vector of missing data $\mathbf{y}^m$ conditional on the observed data $\mathbf{y}^o$, the Gaussian states $\boldsymbol{\alpha}$ and the model parameters and other latent variables $\boldsymbol{\theta}$. Additional information is available in many applications, and it is often desirable or necessary to incorporate the new information into the analysis. For example, in a mixed-frequency setting with both monthly and quarterly variables, a common approach is to treat the monthly observations of the quarterly flow variables as missing, and these missing values are then mapped to the observed values via some inter-temporal constraints. Another example is ragged edge settings where the latest values of some variables are not yet released, but high-quality nowcasts (e.g., from surveys of professional forecasters) are available. Below we show how one can modify the proposed sampling approach to handle various settings with additional information.

To keep the proposed framework general that can handle a wide variety of information settings, suppose there is an additional $M \times 1$ vector of observables, say, $\mathbf{z}$, that is available for sharpening the inference on $\mathbf{y}^m$. We consider two types of mappings that connect $\mathbf{z}$ to $\mathbf{y}^m$. In the first case, suppose $\mathbf{z}$ can be mapped to the missing data exactly via the linear system:

$$\mathbf{z} = \mathbf{M}\mathbf{y}^m, \tag{10}$$

where $\mathbf{M}$ is an $M \times N^m$ matrix specifying the M exact linear relationships. We refer to this type of additional information as hard constraints. One important example is the inter-temporal constraints for mixed-frequency settings based on a log-linear approximation proposed in Mariano and Murasawa (2003, 2010). More specifically, suppose $y_{i,t}^m$ is the missing monthly value of the i-th variable at month t. Let $z_{i,t/3}$ denote the corresponding observed quarterly value (note that $z_{i,t/3}$ is only observed for every third month). Then, a standard log-linear approximation to an arithmetic average of the quarterly variable

can be expressed as:

$$z_{i,t/3} = \frac{1}{3}y^m_{i,t} + \frac{2}{3}y^m_{i,t-1} + y^m_{i,t-2} + \frac{2}{3}y^m_{i,t-3} + \frac{1}{3}y^m_{i,t-4} \tag{11}$$

for $t = 3, 6, 9, \ldots$. By stacking (11) and defining $\mathbf{M}$ appropriately, the exact linear restrictions in (11) can be written in the form in (10). For balanced monthly and quarterly variables, $M = N^m/3$.

Even though the mapping considered in (11) is technically based on a log-linear approximation, in most applied work it is treated as an exact linear relationship. A more appropriate approach might be to explicitly allow for measurement or approximation errors. In addition, there are other situations where allowing for measurement errors is appropriate (e.g., mapping nowcasts from professional forecasters to the underlying endogenous variables). Hence, we consider an alternative mapping that includes measurement errors of the form:

$$\mathbf{z} = \mathbf{M}\mathbf{y}^m + \boldsymbol{\varepsilon}^z, \quad \boldsymbol{\varepsilon}^z \sim \mathcal{N}(\mathbf{0}_M, \mathbf{O}), \tag{12}$$

where $\mathbf{O}$ is a fixed diagonal covariance matrix that encodes the magnitude of the measurement errors. We refer to this type of additional information as soft constraints.

After providing a general setting to incorporate additional information, next we discuss how the sampling of the missing data can be modified given this new information set. First, we consider the case of hard constraints. Recall that the missing data conditional only on the observed data and model parameter is Gaussian as specified in (9). Therefore, sampling the missing data conditioning on the exact linear restrictions in (10) amounts to drawing from the degenerate Gaussian distribution $\mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}^m}, \mathbf{K}^{-1}_{\mathbf{y}^m}\right) 1(\mathbf{M}\mathbf{y}^m = \mathbf{z})$, where $1(\cdot)$ is the indicator function. There are efficient algorithms that can be used to sample from $\mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}^m}, \mathbf{K}^{-1}_{\mathbf{y}^m}\right)$ so that $\mathbf{M}\mathbf{y}^m = \mathbf{z}$, such as Algorithm 2.6 in Rue and Held (2005) and Algorithm 2 in Cong, Chen, and Zhou (2017). In particular, we can first sample $\mathbf{u} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}^m}, \mathbf{K}^{-1}_{\mathbf{y}^m}\right)$ using Algorithm 1. Then, we update the condition set augmented with $\mathbf{z} = \mathbf{M}\mathbf{y}^m$ by computing

$$\mathbf{y}^m = \mathbf{u} + \mathbf{K}^{-1}_{\mathbf{y}^m}\mathbf{M}'(\mathbf{M}\mathbf{K}^{-1}_{\mathbf{y}^m}\mathbf{M}')^{-1}(\mathbf{z} - \mathbf{M}\mathbf{u}).$$

It can be shown that $\mathbf{y}^m$ has the distribution $(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{M}\mathbf{y}^m = \mathbf{z})$. Algorithm 2 describes an efficient implementation in Rue and Held (2005) that avoids explicitly com-

puting the inverse of $\mathbf{K}_{\mathbf{y}^m}$ or $\mathbf{M}\mathbf{K}_{\mathbf{y}^m}^{-1}\mathbf{M}'$. Using this implementation, the additional computational cost for conditioning on $\mathbf{z} = \mathbf{M}\mathbf{y}^m$ is relatively low for $M \ll N^m$. For large M, this algorithm would involve a few large, dense matrices, and the computations could be more intensive.

Algorithm 2 Sampling $(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{z} = \mathbf{M}\mathbf{y}^m)$ with hard constraints, where $(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\theta}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}^m}, \mathbf{K}_{\mathbf{y}^m}^{-1}\right)$.

Given the parameters $\boldsymbol{\mu}_{\mathbf{y}^m}$ and $\mathbf{K}_{\mathbf{y}^m}$, complete the following steps.

1. Obtain the Cholesky factor $\mathbf{C}$ of $\mathbf{K}_{\mathbf{y}^m}$ such that $\mathbf{K}_{\mathbf{y}^m} = \mathbf{C}\mathbf{C}'$.
2. Sample $\mathbf{u} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}^m}, \mathbf{K}_{\mathbf{y}^m}^{-1}\right)$ using Algorithm 1.
3. Solve $\mathbf{C}\mathbf{C}'\mathbf{U} = \mathbf{M}'$ for $\mathbf{U}$ by forward and backward substitution.
4. Solve $\mathbf{M}\mathbf{U}\mathbf{V} = \mathbf{U}'$ for $\mathbf{V}$.
5. Return $\mathbf{y}^m = \mathbf{u} + \mathbf{V}'(\mathbf{z} - \mathbf{M}\mathbf{u})$.

Next, we consider the case of soft constraints. Essentially, we update the conditional distribution of the missing data $\mathbf{y}^m$ given the new information specified in (12). Therefore, one can view the original Gaussian distribution of $\mathbf{y}^m$ in (9) as the 'prior distribution' and the new information in (12) as the 'likelihood'. Then, by standard Bayesian updating, we obtain

$$(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{z}) \sim \mathcal{N}\left(\overline{\boldsymbol{\mu}}_{\mathbf{y}^m}, \overline{\mathbf{K}}_{\mathbf{y}^m}^{-1}\right), \tag{13}$$

where

$$\overline{\mathbf{K}}_{\mathbf{y}^m} = \mathbf{M}'\mathbf{O}^{-1}\mathbf{M} + \mathbf{K}_{\mathbf{y}^m}, \quad \overline{\boldsymbol{\mu}}_{\mathbf{y}^m} = \overline{\mathbf{K}}_{\mathbf{y}^m}^{-1}\left(\mathbf{M}'\mathbf{O}^{-1}\mathbf{z} + \mathbf{K}_{\mathbf{y}^m}\boldsymbol{\mu}_{\mathbf{y}^m}\right).$$

Since for most applications the matrices $\mathbf{M}, \mathbf{O}$ and $\mathbf{K}_{\mathbf{y}^m}$ are all banded, so is the precision matrix $\overline{\mathbf{K}}_{\mathbf{y}^m}$. Hence, the precision-based sampler in Algorithm 1 can be directly applied to sample $\mathbf{y}^m$ efficiently; we simply replace $\mathbf{K}_{\mathbf{y}^m}$ and $\boldsymbol{\mu}_{\mathbf{y}^m}$ by $\overline{\mathbf{K}}_{\mathbf{y}^m}$ and $\overline{\boldsymbol{\mu}}_{\mathbf{y}^m}$, respectively.

Compared to Algorithm 2 for the case of hard constraints, sampling from (13) is much faster and scales well to high-dimensional settings. For approximate inter-temporal restrictions such as Mariano and Murasawa (2003, 2010), the latter sampler is naturally preferable. For other exact inter-temporal restrictions, empirically one can approximate these hard constraints by setting the diagonal elements of $\mathbf{O}$ to be very small (e.g., 10^{-8}).

2.3 Joint Sampling of the Missing Data and Gaussian States

So far we have focused on sampling the missing data $\mathbf{y}^m$ conditional on the observed data $\mathbf{y}^o$ and the Gaussian states $\boldsymbol{\alpha}$. Ideally one would like to draw $(\mathbf{y}^m, \boldsymbol{\alpha})$ together from their joint distribution, because grouping variables into larger blocks makes the posterior sampler more efficient in terms of lower autocorrelations in the MCMC draws. Below we show how one can sample from the joint distribution $p(\mathbf{y}^m, \boldsymbol{\alpha} \,|\, \mathbf{y}^o, \boldsymbol{\theta})$; drawing from $p(\mathbf{y}^m, \boldsymbol{\alpha} \,|\, \mathbf{y}^o, \boldsymbol{\theta}, \mathbf{z})$ with additional information $\mathbf{z}$ can be done similarly.

To that end, first note that the joint distribution of $(\mathbf{y}^m, \boldsymbol{\alpha})$ can be decomposed as

$$p(\mathbf{y}^m, \boldsymbol{\alpha} \,|\, \mathbf{y}^o, \boldsymbol{\theta}) = p(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\theta}) p(\boldsymbol{\alpha} \,|\, \mathbf{y}^o, \mathbf{y}^m, \boldsymbol{\theta}).$$

Hence, one can sample $(\boldsymbol{\alpha}, \mathbf{y}^m)$ jointly by first drawing $\mathbf{y}^m \sim p(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\theta})$ marginally of $\boldsymbol{\alpha}$. Then, given the sampled missing data $\mathbf{y}^m$, one obtains $\boldsymbol{\alpha}$ from the conditional distribution $p(\boldsymbol{\alpha} \,|\, \mathbf{y}^o, \mathbf{y}^m, \boldsymbol{\theta})$. Since the second step can be done using standard algorithms, in what follows we focus on the first step.

Stack the state equation (2) over $t = 1, \ldots, T$:

$$\mathbf{H}_{\boldsymbol{\alpha}} \boldsymbol{\alpha} = \mathbf{m}_{\boldsymbol{\alpha}} + \boldsymbol{\varepsilon}^{\boldsymbol{\alpha}}, \quad \boldsymbol{\varepsilon}^{\boldsymbol{\alpha}} \sim \mathcal{N}(\mathbf{0}_{Tk}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega} = \operatorname{diag}(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_T)$,

$$\mathbf{m}_{\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{\gamma} + \sum_{j=1}^{q} \boldsymbol{\Phi}_j \boldsymbol{\alpha}_{1-j} \\ \boldsymbol{\gamma} + \sum_{j=2}^{q} \boldsymbol{\Phi}_j \boldsymbol{\alpha}_{2-j} \\ \vdots \\ \boldsymbol{\gamma} + \boldsymbol{\Phi}_q \boldsymbol{\alpha}_0 \\ \boldsymbol{\gamma} \\ \vdots \\ \boldsymbol{\gamma} \end{bmatrix}, \; \mathbf{H}_{\boldsymbol{\alpha}} = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k\times k} & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{k\times k} \\ -\boldsymbol{\Phi}_1 & \mathbf{I}_k & \mathbf{0}_{k\times k} & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{k\times k} \\ -\boldsymbol{\Phi}_2 & -\boldsymbol{\Phi}_1 & \mathbf{I}_k & \mathbf{0}_{k\times k} & \cdots & & & \mathbf{0}_{k\times k} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ -\boldsymbol{\Phi}_q & \cdots & & -\boldsymbol{\Phi}_1 & \mathbf{I}_k & \mathbf{0}_{k\times k} & & \vdots \\ \mathbf{0}_{k\times k} & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{k\times k} & \cdots & \mathbf{0}_{k\times k} & -\boldsymbol{\Phi}_p & \cdots & -\boldsymbol{\Phi}_2 & -\boldsymbol{\Phi}_1 & \mathbf{I}_k \end{bmatrix}.$$

Since $\mathbf{H}_{\boldsymbol{\alpha}}$ is a square matrix with unit determinant, it is invertible and we have $\boldsymbol{\alpha} = \mathbf{H}_{\boldsymbol{\alpha}}^{-1} \mathbf{m}_{\boldsymbol{\alpha}} + \mathbf{H}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\varepsilon}^{\boldsymbol{\alpha}}$. Substitute this expression into (5), we obtain

$$\mathbf{G}^o \mathbf{y}^o + \mathbf{G}^m \mathbf{y}^m = \mathbf{W} \mathbf{H}_{\boldsymbol{\alpha}}^{-1} \mathbf{m}_{\boldsymbol{\alpha}} + \mathbf{X}\boldsymbol{\beta} + \widetilde{\boldsymbol{\varepsilon}}, \quad \widetilde{\boldsymbol{\varepsilon}} \sim \mathcal{N}(\mathbf{0}_{Tn}, \widetilde{\boldsymbol{\Sigma}}), \tag{14}$$

where $\widetilde{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon} + \mathbf{W}\mathbf{H}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\varepsilon}^{\boldsymbol{\alpha}}$ and $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \mathbf{W}\mathbf{H}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\Omega}\mathbf{H}_{\boldsymbol{\alpha}}^{-1\prime}\mathbf{W}'$. The representation in (14) is conceptually convenient, because it implies that the conditional distribution of $\mathbf{y}^m$ given $\mathbf{y}^o$ and $\boldsymbol{\theta}$ but marginal of $\boldsymbol{\alpha}$ is also Gaussian:

$$(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\theta}) \sim \mathcal{N}\left(\widetilde{\boldsymbol{\mu}}_{\mathbf{y}^m}, \widetilde{\mathbf{K}}_{\mathbf{y}^m}^{-1}\right),$$

where

$$\widetilde{\mathbf{K}}_{\mathbf{y}^m} = \mathbf{G}^{m\prime}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{G}^m, \quad \widetilde{\boldsymbol{\mu}}_{\mathbf{y}^m} = \widetilde{\mathbf{K}}_{\mathbf{y}^m}^{-1}\mathbf{G}^{m\prime}\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{W}\mathbf{H}_{\boldsymbol{\alpha}}^{-1}\mathbf{m}_{\boldsymbol{\alpha}} + \mathbf{X}\boldsymbol{\beta} - \mathbf{G}^o\mathbf{y}^o).$$

However, since $\widetilde{\boldsymbol{\Sigma}}$ is a dense $Tn \times Tn$ matrix, in order for this approach to be practical, we need to find an efficient way to compute $\mathbf{G}^{m\prime}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{G}^m$ and $\mathbf{G}^{m\prime}\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{W}\mathbf{H}_{\boldsymbol{\alpha}}^{-1}\mathbf{m}_{\boldsymbol{\alpha}} + \mathbf{X}\boldsymbol{\beta} - \mathbf{G}^o\mathbf{y}^o)$ that does not require brute-force inversion of $\widetilde{\boldsymbol{\Sigma}}$. Fortunately, one can use the Woodbury formula to obtain

$$\begin{aligned}
\widetilde{\boldsymbol{\Sigma}}^{-1} &= \left(\boldsymbol{\Sigma} + \mathbf{W}\mathbf{H}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\Omega}\mathbf{H}_{\boldsymbol{\alpha}}^{-1\prime}\mathbf{W}'\right)^{-1} \\
&= \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{H}_{\boldsymbol{\alpha}}^{-1}\left(\boldsymbol{\Omega}^{-1} + \mathbf{H}_{\boldsymbol{\alpha}}^{-1\prime}\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{H}_{\boldsymbol{\alpha}}^{-1}\right)^{-1}\mathbf{H}_{\boldsymbol{\alpha}}^{-1\prime}\mathbf{W}'\boldsymbol{\Sigma}^{-1} \\
&= \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{W}\left(\mathbf{H}_{\boldsymbol{\alpha}}'\boldsymbol{\Omega}^{-1}\mathbf{H}_{\boldsymbol{\alpha}} + \mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W}\right)^{-1}\mathbf{W}'\boldsymbol{\Sigma}^{-1}.
\end{aligned}$$

Note that in the above expression the matrix $\mathbf{H}_{\boldsymbol{\alpha}}'\boldsymbol{\Omega}^{-1}\mathbf{H}_{\boldsymbol{\alpha}} + \mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W}$ is banded because $\mathbf{H}_{\boldsymbol{\alpha}}$ and $\mathbf{W}$ are banded and $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are block-diagonal. Hence, one can compute $\widetilde{\mathbf{K}}_{\mathbf{y}^m}$ and $\widetilde{\boldsymbol{\mu}}_{\mathbf{y}^m}$ efficiently. Finally, given $\widetilde{\mathbf{K}}_{\mathbf{y}^m}$ and $\widetilde{\boldsymbol{\mu}}_{\mathbf{y}^m}$, one can sample $(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\theta})$ using the algorithm in Chan and Jeliazkov (2009).

In general, the precision matrix $\widetilde{\mathbf{K}}_{\mathbf{y}^m}$ is not banded. Consequently, sampling from $(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\theta})$ is typically more computationally intensive than sampling from $(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta})$. The gain, however, is higher computational efficiency in terms of lower autocorrelations of the MCMC draws, as the former approach allows for joint sampling of $(\mathbf{y}^m, \boldsymbol{\alpha})$. There is therefore a trade-off between computational time and computational efficiency.

3 A Monte Carlo Study

In this section we conduct a series of Monte Carlo experiments to assess the speed and accuracy of the proposed precision-based methods for drawing the latent missing ob-

servations relative to Kalman-filter based methods. In the first subsection we consider mixed-frequency settings in which the missing data are the high-frequency observations of the low-frequency variables. We then consider unbalanced panels in the following subsection.

All datasets are generated from the following VAR:

$$\mathbf{y}_t = \mathbf{b}_0 + \mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_2\mathbf{y}_{t-2} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_n, \boldsymbol{\Sigma}),$$

where $\mathbf{y}_t = (\mathbf{y}_t^{o\prime}, \mathbf{y}_t^{m\prime})'$ is an $n \times 1$ vector of mixed-frequency data, $\mathbf{y}_t^o$ is an $n^o \times 1$ vector of (observed) high-frequency variables and $\mathbf{y}_t^m$ is an $n^m \times 1$ vector of (missing) high-frequency observations of the low-frequency variables. In addition, low-frequency variables $z_{i,t/3}, i = 1, \ldots, n^m$, are observed at $t = 3, 6, 9, \ldots$, which can be used to inform the values of the missing $\mathbf{y}_t^m$ via (10) or (12). For the baseline case we set $p = 5$. Furthermore, we generate the model parameters as follows. We set $\mathbf{b}_0 = 0.01 \times \mathbf{1}_n$. The diagonal elements of the first VAR coefficient matrix are iid uniform $\mathcal{U}(0, 0.5)$ and the off-diagonal elements are $\mathcal{U}(-0.2, 0.2)$. All elements of the higher VAR coefficient matrix are iid $\mathcal{N}(0, 0.05^2/l^2)$, where l is the lag length. The error covariance matrix $\boldsymbol{\Sigma}$ is generated from the inverse-Wishart distribution $\mathcal{IW}(n + 10, 0.07\mathbf{I}_n + 0.03\mathbf{1}_n\mathbf{1}_n')$.

For each simulated dataset $r = 1, \ldots, R$, we estimate the missing observations $\mathbf{y}_t^m$ using 4 methods: the precision-based sampler with the hard inter-temporal constraints in (10), the precision-based sampler with the soft constraints in (12), the simulation smoother of Carter and Kohn (1994) as implemented in the code provided by Schorfheide and Song (2015), and the simulation smoother of Durbin and Koopman (2002).[5] For both the simulation smoothers of Carter and Kohn (1994) and Durbin and Koopman (2002), we impose the hard constraints. For the precision-based sampler with soft constraints, we set the diagonal elements of the measurement error covariance matrix $\mathbf{O}$ to be 10^{-8}. Hence, we view it as an approximation of the hard constraints so that results from the 4 methods are comparable.

[5]The implementation in Schorfheide and Song (2015) uses a compact state-space representation to draw the missing observations. This representation removes the monthly observations from the state vector that appears in the measurement equation. Consequently, it reduces the dimension of the state vector and is generally more efficient. In contrast, the simulation smoother of Durbin and Koopman (2002) requires the model in a standard companion form (see Appendix B for details) that results in a higher dimensional state vector.

Finally, we implement a standard normal prior for the VAR coefficients and an inverse-Wishart prior for the error covariance matrix. More specifically, let $\boldsymbol{\beta} = \text{vec}\left([\mathbf{b}_0, \mathbf{B}_1, \ldots, \mathbf{B}_p]'\right)$ denote the $k \times 1$ vector of all VAR coefficients with $k = n(np+1)$. Then, the priors are $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}_k, \mathbf{I}_k)$ and $\boldsymbol{\Sigma} \sim \mathcal{IW}(5, \mathbf{I}_n)$.

3.1 Missing Observations of Low-Frequency Variables

We consider DGPs of different dimensions with $T = 300$: small ($n = 6, n^o = 5, n^m = 1$), medium ($n = 11, n^o = 10, n^m = 1$) and large ($n = 16, n^o = 15, n^m = 1$). We also investigate settings with a larger number of unobserved variables with $n^m = 5$. For each design, we generate $R = 100$ datasets from the mixed-frequency VAR as described above. We then fit each dataset using a Gibbs sampler that draws sequentially the missing observations and model parameters. In particular, we use 4 different methods to sample the missing observations. To assess the accuracy of the different methods, we compute the mean squared error (MSE) of the estimated missing observations against the actual values. More specifically, for each dataset with missing observations $\mathbf{y}_1^{m(i)}, \ldots, \mathbf{y}_T^{m(i)}$, $i = 1, \ldots, R,$, and each method with posterior mean vector $\widehat{\mathbf{y}}^{m(i,j)}, j = 1, \ldots, 4$, we compute $\text{MSE}_i(\widehat{\mathbf{y}}^{m(i,j)}) = \sum_{t=1}^{T} ||\mathbf{y}_t^{m(i)} - \widehat{\mathbf{y}}_t^{m(i,j)}||^2/T$, where $||\cdot||$ is the ℓ^2-norm. We further summarize the accuracy by averaging over the R MSEs for each design, and the results are reported in Table 1.

We also report the computation time, based on 15,000 MCMC draws with a burn-in period of 5,000 draws.[6] Since all four methods aim to draw from the same distribution—namely, the conditional distribution of the missing observations given the observed data and model parameters—in principle they should give the same MSEs. Indeed, they give very similar MSEs in our simulations; the small discrepancies are mainly due to numerical and simulation errors. In terms of runtime, it is clear that the proposed precision-based methods are more computationally efficient compared to the Kalman-filter based method across a range of settings.

[6]The computation time is based on a standard desktop with an Intel Xeon W-2223 @ 3.6GHz processor and 16 GB of RAM and the code is implemented in MATLAB.

Table 1: Mean squared errors of the estimated missing observations and computation time using four methods: the proposed precision-based method with hard inter-temporal constraints (P-hard), the precision-based method with soft inter-temporal constraints (P-soft), the simulation smoother of Carter and Kohn (1994) implemented in Schorfheide and Song (2015) (CK) and the simulation smoother of Durbin and Koopman (2002) (DK).

		MSE				Computation time (minutes)			
n^m	n^o	P-hard	P-soft	DK	CK	P-hard	P-soft	DK	CK
1	5	0.004	0.004	0.004	0.005	0.7	0.6	7	5
1	10	0.004	0.004	0.004	0.004	3	3	31	9
1	15	0.004	0.004	0.004	0.004	13	13	61	16
5	5	0.005	0.005	0.005	0.005	8	4	23	24
5	10	0.005	0.005	0.005	0.005	16	12	51	35
5	15	0.004	0.004	0.004	0.005	38	35	106	51

Table 1 reports the runtime of full MCMC estimation. When the dimension of the VAR increases, the computational complexity of simulating the VAR coefficients dominates, and it might give the impression that the runtime of the four methods converges. To better understand how the proposed methods perform across a wider range of settings, next we compare only the runtime of sampling the missing observations.

First, Figure 1 reports the runtime of sampling ten draws of the missing observations using the four methods for a range of n^o and n^m. Since the differences in computation time are substantial, they are shown in log scale. It is clear that both precision-based methods compare favorably to the Kalman-filter based methods, and both scale well to high dimensional settings. In addition, the variant with soft restrictions is especially efficient when there are a large number of variables with missing observations.

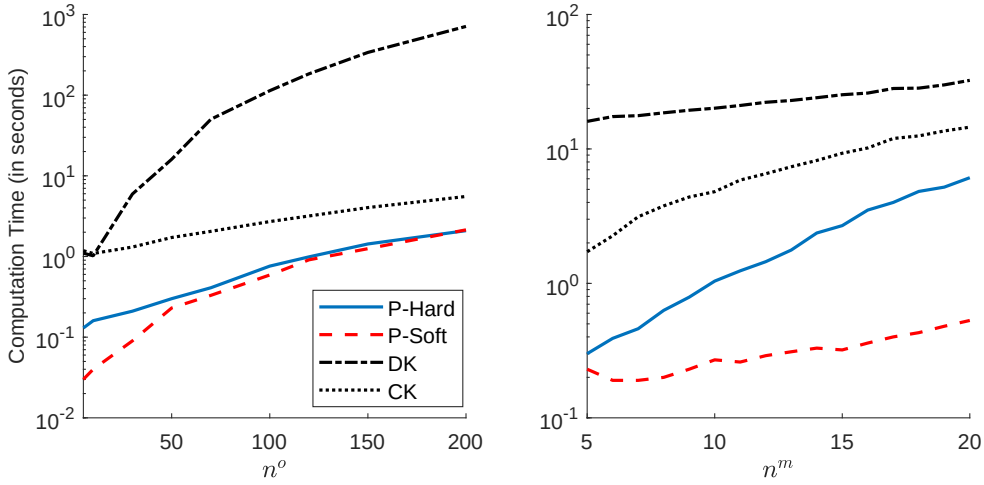

Figure 1: Computation time of obtaining 10 draws against n^o and n^m, the numbers of observed and partially unobserved variables, respectively, with $T = 300$ and $p = 5$. The four methods are: precision-based sampler with hard inter-temporal constraints (P-hard), precision-based sampler with soft constraints (P-soft), the simulation smoother of Carter and Kohn (1994) implemented in Schorfheide and Song (2015) (CK) and the simulation smoother of Durbin and Koopman (2002).

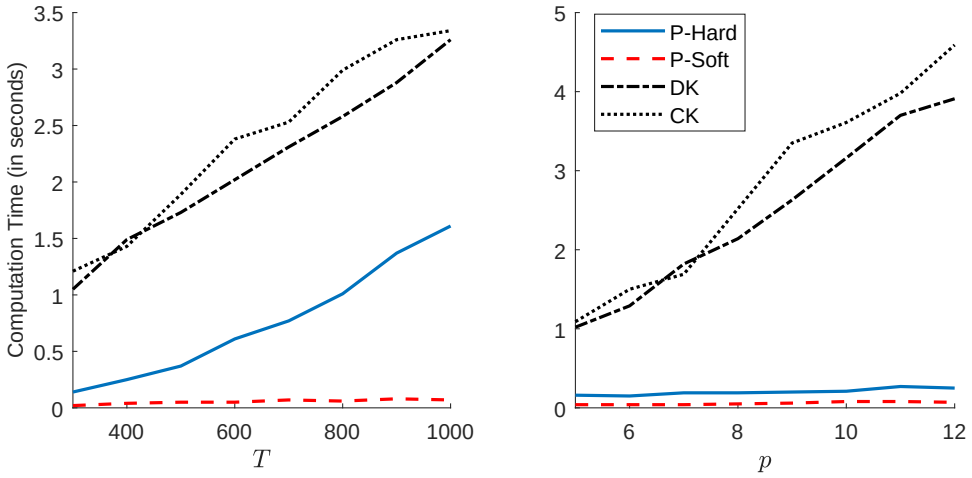

Figure 2: Computation time of obtaining 10 draws against T and p, the numbers of time periods and lags, respectively, with $n^m = 5$ and $n^o = 10$. The four methods are: precision-based sampler with hard inter-temporal constraints (P-hard), precision-based sampler with soft constraints (P-soft), the simulation smoother of Carter and Kohn (1994) implemented in Schorfheide and Song (2015) (CK) and the simulation smoother of Durbin and Koopman (2002).

Next, Figure 2 reports the runtime of sampling ten draws of the missing observations for a range of sample sizes T and lag lengths p. While both precision-based methods perform well, the version with soft constrains does substantially better and scales well to very large T and p. It is also worth mentioning that to apply the Kalman filter, one needs to redefine the states so that the observation equation depends only on the current (redefined) state. When p is large, the dimension of this new state vector is large. That is one reason why the Kalman-filter based methods become more computationally intensive when p is large. In contrast, the computational costs of the precision-based methods remain low even for long lag lengths.

3.2 Unbalanced Panels

In this section we illustrate the performance of the proposed precision-based methods in settings involving unbalanced panels where different variables are missing in different time periods. To that end, we conduct a series of simulations using a medium-size VAR ($n = 13$, $n^o = 10$, $n^m = 3$) with a sample size of $T = 300$. We simulate the data using the same DGP as described above, but here we assume that the low-frequency variables are also missing for selected time periods (i.e., in addition to their missing high-frequency observations). More specifically, the first low-frequency variable is completely missing for the first 30 periods; the second is missing from $t = 150, \ldots, 180$; and the third is missing for the last 30 periods.

Table 2: Mean squared errors of the estimated missing observations and computation time averaging over $R = 100$ replications using four methods: the proposed precision-based method with hard inter-temporal constraints (P-hard), the precision-based method with soft inter-temporal constraints (P-soft), the simulation smoother of Carter and Kohn (1994) implemented in Schorfheide and Song (2015) (CK) and the simulation smoother of Durbin and Koopman (2002) (DK).

		MSE				Computation time (minutes)			
n^m	n^o	P-hard	P-soft	DK	CK	P-hard	P-soft	DK	CK
3	10	0.004	0.004	0.004	0.009	8	7	40	14

Table 2 reports the MSEs and computation time for estimating the missing observations of the three low-frequency variables. Similar to the previous simulated experiments, here we also find that the MSEs across the four methods are very similar. As before, the

proposed precision-based methods are computationally more efficient than both Kalman-filter based simulation smoothers.

Next, Figure 3 plots the posterior means of the missing observations of the low-frequency variables obtained using the four methods against the actual simulated data. In the figure we highlight the time periods in which each low-frequency variable is completely missing. All four methods produce very similar posterior estimates of the missing observations during these periods as expected. We therefore conclude that the precision-based methods can handle any arbitrary missing data pattern as well as standard filtering and smoothing methods, but they are more computationally efficient.

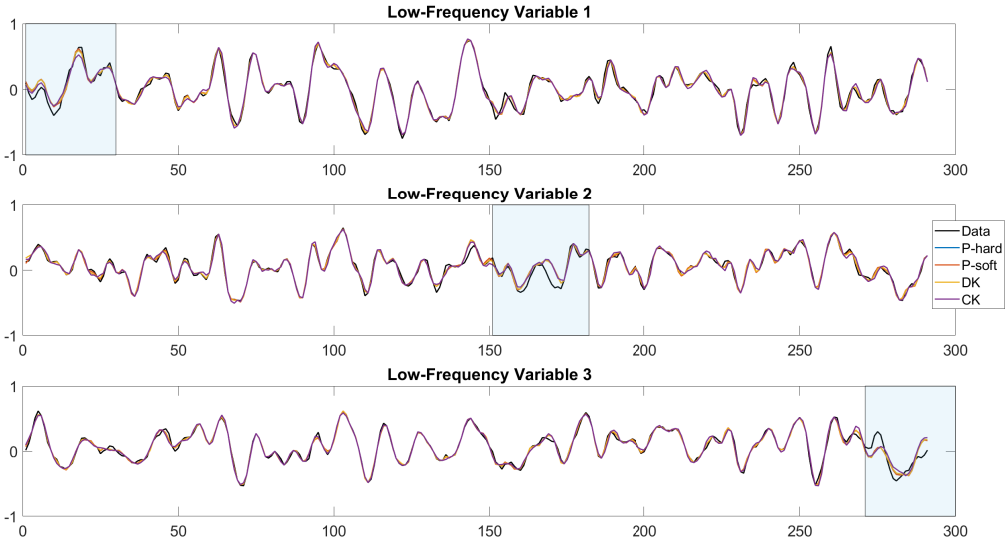

Figure 3: Posterior means of the low-frequency variables obtained using four methods: the proposed precision-based method with hard inter-temporal constraints (P-hard), the precision-based method with soft inter-temporal constraints (P-soft), the simulation smoother of Carter and Kohn (1994) implemented in Schorfheide and Song (2015) (CK) and the simulation smoother of Durbin and Koopman (2002) (DK). The box in each panel highlights the time periods in which the low-frequency variable is completely missing.

4 Empirical Applications

We demonstrate the proposed precision-based approach via two empirical applications involving two popular large-scale models and widely different missing data patterns. In the first application, we use a large mixed-frequency VAR with stochastic volatility to generate weekly estimates of real GDP. In the second application, we extract latent factors from unbalanced datasets using a dynamic factor model with stochastic volatility.

4.1 A Weekly State-Space Mixed-Frequency VAR

In the first application we illustrate how the proposed precision-based approach can be used to estimate a Bayesian VAR with 22 variables in weekly, monthly and quarterly frequencies. Following the seminal work by Schorfheide and Song (2015), we model all variables at the highest observed frequency, and treat the high-frequency observations of the low-frequency variables as missing data. One key advantage of this approach is that it produces interpolated (historical) estimates of the low-frequency variables at a higher frequency. For example, Schorfheide and Song (2015) use monthly and quarterly variables to obtain monthly GDP estimates, which are required as inputs for a variety of applications. In addition, this approach can also deliver more timely nowcasts by incorporating information in higher-frequency variables. Recent applications of this approach include Brave, Butters, and Justiniano (2019) and Koop, McIntyre, Mitchell, and Poon (2020, 2022).

We extend this line of work by including weekly macroeconomic and financial variables. By modeling all variables in weekly frequency, we are able to obtain weekly GDP estimates. Fitting a large mixed-frequency VAR, however, is computationally intensive as there are a large number of missing observations. Moreover, since the missing data pattern is irregular (e.g., each quarter or month does not always have the same number of weeks), the implementation of conventional methods is also more complex. In contrast, the proposed method can easily handle the irregular missing data pattern and it scales well to high dimensions.

The US dataset consists of 16 weekly variables (including raw steel production, retail sales, initial claims for unemployment benefits and various financial variables), 5 monthly

variables (such as industrial production, CPI and labor market variables) and 1 quarterly variable (real GDP) from January 2013 to August 2022. Seven of the weekly variables are obtained from Lewis, Mertens, Stock, and Trivedi (2022), which they use to construct their Weekly Economic Index (WEI); other variables are sourced from the FRED database at the Federal Reserve Bank of St. Louis. More details about the data can be found in Appendix A.

Since our sample includes the COVID-19 pandemic, it is empirically important to allow for some form of heteroskedasticity or non-Gaussian errors, as demonstrated in recent papers such as Hartwig (2021), Lenza and Primiceri (2022) and Carriero, Clark, Marcellino, and Mertens (2022). We therefore incorporate the common stochastic volatility of Carriero, Clark, and Marcellino (2016) into a mixed-frequency VAR as follows:

$$\mathbf{y}_t = \mathbf{b}_0 + \mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_2\mathbf{y}_{t-2} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{e}^{h_t}\boldsymbol{\Sigma}),$$

where $\mathbf{y}_t = (\mathbf{y}_t^{o\prime}, \mathbf{y}_t^{m\prime})'$ is an $n \times 1$ vector of mixed-frequency data, and $\mathbf{y}_t^o$ and $\mathbf{y}_t^m$ are the vectors of observed and missing variables, respectively. Note that the error covariance matrix is scaled by the common log-volatility h_t, which is modeled as a stationary AR(1) process:

$$h_t = \phi h_{t-1} + u_t^h, \quad u_t^h \sim \mathcal{N}(0, \sigma_h^2),$$

for $t = 2, \ldots, T$, where $|\phi| < 1$ and the initial condition is specified as $h_1 \sim \mathcal{N}(0, \sigma_h^2/(1-\phi^2))$. Here the unconditional mean of the AR(1) process is assumed to be zero for identification.[7]

Next, we describe the priors on the model parameters $\boldsymbol{\beta} = \mathrm{vec}([\mathbf{b}_0, \mathbf{B}_1, \ldots, \mathbf{B}_p]')$, $\boldsymbol{\Sigma}, \phi$ and σ_h^2. Specifically, we assume the following independent priors:

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}}),\ \boldsymbol{\Sigma} \sim \mathcal{IW}(\nu_0, \mathbf{S}_0),\ \phi \sim \mathcal{N}(\phi_0, V_\phi)1(|\phi| < 1),\ \sigma_h^2 \sim \mathcal{IG}(\eta_0, S_0),$$

where $1(\cdot)$ denotes the indicator function. The prior mean vector and covariance matrix

[7] In preliminary work we also implemented a version of the model with the more flexible Cholesky stochastic volatility of Cogley and Sargent (2005) and Carriero, Clark, and Marcellino (2019), in which each variable has its own stochastic volatility process. However, given the large number of missing observations, we found that it was hard to pin down some of the stochastic volatility processes in simulations. In contrast, the common stochastic volatility worked well. An additional advantage of the common stochastic volatility is that it is order-invariant, whereas the Cholesky stochastic volatility is not.

of the VAR coefficients, $\boldsymbol{\beta}_0$ and $\mathbf{V}_{\boldsymbol{\beta}}$ respectively, are chosen in the spirit of the Minnesota prior pioneered by Doan, Litterman, and Sims (1984) and Litterman (1986). More specifically, we set $\boldsymbol{\beta}_0 = \mathbf{0}_k$ to shrink the coefficients to zero. For the prior covariance matrix $\mathbf{V}_{\boldsymbol{\beta}}$, it is assumed to be diagonal such as

$$\begin{aligned}
\text{Var}(B_{l,ii}) &= \frac{\kappa_1}{l^2},\ l = 1, \ldots, p,\ i = 1, \ldots, n, \\
\text{Var}(B_{l,ij}) &= \frac{\kappa_2 s_i^2}{l^2 s_j^2},\ l = 1, \ldots, p,\ i, j = 1, \ldots, n,\ i \neq j, \\
\text{Var}(b_{0,i}) &= 100 s_i^2,\ i = 1, \ldots, n,
\end{aligned}$$

where $B_{l,ij}$ is the (i, j) element of $\mathbf{B}_l$, s_r^2 denotes the sample variance of the residuals from an AR(4) model for the variable r for $r = 1, \ldots, n$. To implement cross-variable shrinkage, i.e., shrinking coefficients on lags of other variables more strongly to zero than on own lags, we set $\kappa_1 = 0.04$ and $\kappa_2 = 0.01$. Finally, we set $\nu_0 = n + 3, \mathbf{S}_0 = \mathbf{I}_n$, $\eta_0 = 10$ and $S_0 = 0.004$ so that the prior means of $\boldsymbol{\Sigma}$ and σ_h^2 are $0.5\mathbf{I}_n$ and 0.021^2.

We model all variables in weekly frequency, and the missing observations of the monthly or quarterly variables are linked to their corresponding observed values via inter-temporal constraints similar to those in Mariano and Murasawa (2003, 2010). Compared to standard mixed-frequency settings involving only quarterly and monthly variables, the inter-temporal constraints here are more complex, as there might be different numbers of weeks in different months or quarters. More specifically, given the releasing date t of a monthly/quarterly variable, let $n_{i,t}^w$ denote the number of weeks between t and the last releasing date. Then, each observed monthly/quarterly variable $z_{i,t}$ is linked to the missing observations $y_{i,t}^m$ via the inter-temporal constraint:

$$z_{i,t} = \sum_{s=1}^{2n_{i,t}^w - 1} \left(1(s \leqslant n_{i,t}^w) \frac{s}{n_{i,t}^w} + 1(s > n_{i,t}^w) \frac{2n_{i,t}^w - s}{n_{i,t}^w} \right) y_{i,t-s+1}^m + \varepsilon_{i,t}^z,$$

where $\varepsilon_{i,t}^z \sim \mathcal{N}(0, o_i)$ captures the log-linear approximation error, and we set $o_i = 10^{-8}$.

The mixed-frequency VAR is estimated using MCMC methods. In particular, given the model parameters and the common stochastic volatility, we use the proposed sampler as described in Section 2 to sample the missing observations of the monthly and quarterly variables. Then, given these missing observations, standard algorithms can be used to

sample the model parameters and the common stochastic volatility.[8] To gauge the efficiency of the posterior sampler, we compute the inefficiency factors associated with the missing data and the model parameters. All inefficiency factors are less than 100 (see Appendix C for details), and they are comparable to those of large VARs without missing data.

The raw weekly GDP growth estimates are rather volatile, which is expected given that they are measured in weekly frequency. For easier comparison and interpretation, we convert these weekly estimates to the more familiar quarterly growth rates. More specifically, given the estimated week-on-week GDP growth rates $y^m_{t,j}, t = 1, \ldots, T$, we use the inter-temporal constraints to convert them to quarterly growth rates:

$$y^*_{t,j} = \sum_{s=1}^{n^w_j} \frac{s}{n^w_j} y^m_{t-s,j} + \sum_{s=n^w_j+1}^{2n^w_j-1} \frac{2n^w_j - s}{n^w_j} y^m_{t-s,j}, \tag{15}$$

where we fix $n^w_j = 13$ weeks. Hence, $y^*_{t,j}$ may be interpreted as the cumulative GDP growth over the past 13 weeks. Figure 4 plots the posterior means and the associated 68% credible interval of the these aggregate weekly GDP growth rates. In the graph we also mark the observed quarterly GDP growth rates in black crosses. As expected, all the observed quarterly GDP values lie on the aggregate weekly GDP estimates—the inter-temporal constraints ensure that the weekly GDP estimates are aggregated to the observed quarterly value.

The most prominent feature of the aggregate weekly GDP estimates is the drastic drop at the onset of the COVID-19 pandemic and the subsequent rebound. In particular, the US real GDP decreased by about 37% in 2020:Q2 when the pandemic forced widespread business closures. When the economy gradually opened up in 2020:Q3, GDP bounced back sharply by about 30%. One key advantage of modeling GDP in weekly frequency is that, in between the quarterly GDP release dates, the model is able to provide GDP estimates on a weekly basis by incorporating information in other weekly and monthly variables.

[8]For example, the common stochastic volatility can be sampled using the methods in Carriero, Clark, and Marcellino (2016) or Chan (2020). The VAR coefficients is jointly Gaussian and can be sampled jointly or equation by equation as in Carriero, Chan, Clark, and Marcellino (2022).

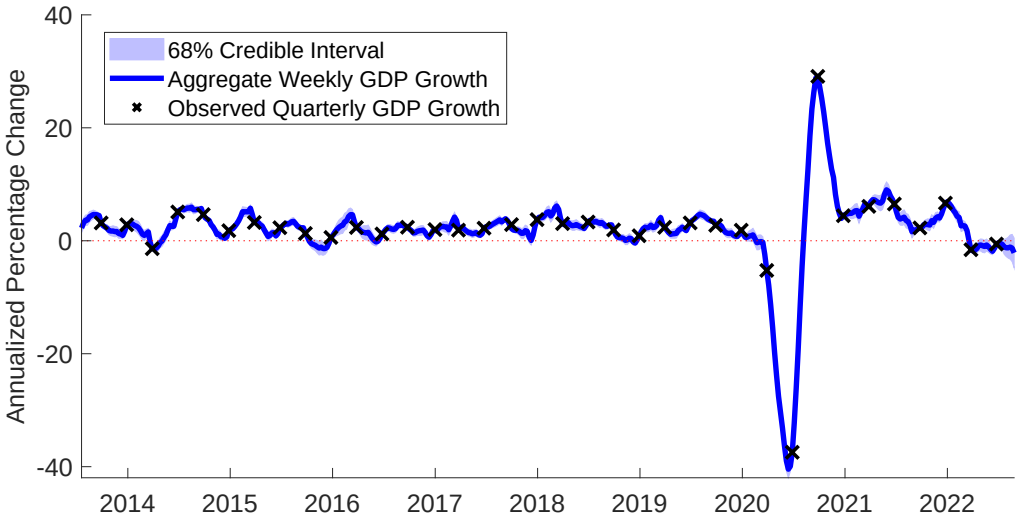

Figure 4: Aggregate weekly estimates of GDP growth over a quarter. The solid line denotes the posterior means of the aggregate weekly GDP growth (annualized), the shaded area is the associated 68% credible intervals, and the crosses denote the actual observed quarterly GDP values.

Next, we compare the aggregate weekly GDP estimates to two high-frequency indicators that are designed to track real economic activity. The first is the Weekly Economic Index of Lewis, Mertens, Stock, and Trivedi (2022), which is updated weekly by the Federal Reserve Bank of New York. The second is the Business Conditions Index of Aruoba, Diebold, and Scotti (2009), which is maintained by the Federal Reserve Bank of Philadelphia. While both indicators incorporate a range of macroeconomic and financial variables at high observation frequency, they are latent factors from dynamic factor models. In contrast, our mixed-frequency VAR provides GDP estimates directly and are easier to interpret.

We obtain the Weekly Economic Index and Business Conditions Index in weekly frequency from the Federal Reserve Banks of New York and Philadelphia, respectively. They are then converted to quarterly growth using (15) for easier comparison. Figure 5 plots the aggregate weekly GDP estimates as well as the two indicators. It is clear from the figure that the aggregate weekly GDP track the Business Conditions Index closely, even during the extreme turning points in 2020:Q2 and 2020:Q3.[9] In contrast, the Weekly Economic

[9]While the Business Conditions Index is constructed so that its average value is zero, the average

Index displays noticeably different dynamics during the onset of the COVID-19 pandemic and the immediate rebound. In particular, the Weekly Economic Index suggests that the US economy experienced a sluggish recovery from the widespread lock-down in 2020:Q2. In contrast, the aggregate weekly GDP and the Business Conditions Index indicate a sharper rebound. One potential driver for this difference is that both the aggregate weekly GDP and the Business Conditions Index incorporate quarterly GDP data, whereas the Weekly Economic Index does not. Consequently, the latter could potentially capture only the economic activity of specific sectors of the US economy, whereas the former two track the whole US economy through the information in the GDP data.

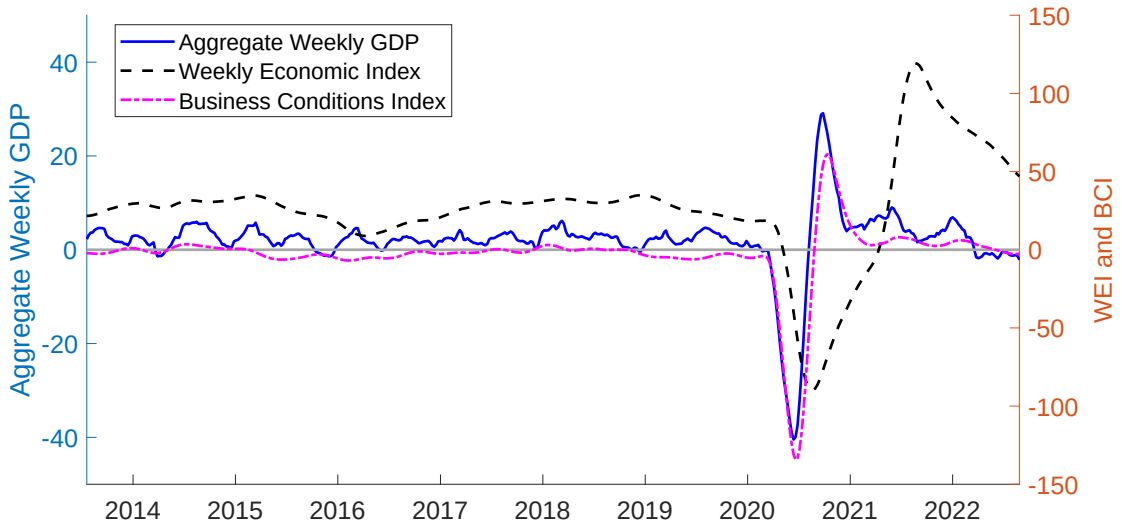

Figure 5: Plots of the aggregate weekly GDP estimates (solid line), the Weekly Economic Index of Lewis, Mertens, Stock, and Trivedi (2022) (dash line) and the Business Conditions Index of Aruoba, Diebold, and Scotti (2009) (dash-dotted line).

4.2 A Dynamic Factor Model with Unbalanced Datasets

To demonstrate the versatility of the proposed approach, in the second application we consider a different type of missing data pattern in the context of another popular model for handling large datasets: a dynamic factor model. More specifically, we extract common factors from the FRED-MD datasets of McCracken and Ng (2016) using a dynamic

GDP growth is about 2% over the sample period. Hence, there are differences in the level of the two series.

factor model with stochastic volatility. We focus on the COVID-19 pandemic period and estimate the common factors in real-time using vintages from March 2020 to September 2022. Each vintage contains 128 monthly variables, but many of these variables have missing values. These missing values mainly come from two sources: missing observations at the beginning of the sample for some more recently constructed variables and missing values at the end of the sample due to publication lags, which is often referred to as ragged edge.

Let $\mathbf{y}_t = (\mathbf{y}_t^{o\prime}, \mathbf{y}_t^{m\prime})'$ denote the $n \times 1$ vector of monthly variables, where $\mathbf{y}_t^o$ is the n_t^o-vector of observed variables and $\mathbf{y}_t^m$ is the n_t^m-vector of variables with missing values at time t. Furthermore, let $\mathbf{f}_t$ represent the $k \times 1$ vector of latent dynamic factors. Due to the COVID-19 outliers, we incorporate stochastic volatility in both the factors and the idiosyncratic errors and consider the following dynamic factor model:

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{A}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \\
\boldsymbol{\varepsilon}_t &= \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \cdots + \boldsymbol{\Psi}_p\boldsymbol{\varepsilon}_{t-p} + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}_n, \boldsymbol{\Sigma}_t), \\
\mathbf{f}_t &= \boldsymbol{\Phi}_1\mathbf{f}_{t-1} + \cdots + \boldsymbol{\Phi}_q\mathbf{f}_{t-q} + \boldsymbol{\varepsilon}_t^{\mathbf{f}}, \quad \boldsymbol{\varepsilon}_t^{\mathbf{f}} \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{\Omega}_t),
\end{aligned}
$$

where $\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_p, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_q$ are diagonal matrices, $\boldsymbol{\Sigma}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{n,t}})$ and $\boldsymbol{\Omega}_t = \text{diag}(\text{e}^{h_{n+1,t}}, \ldots, \text{e}^{h_{n+k,t}})$. The $n + k$ log-volatility processes are assumed to follow independent random walks:

$$
h_{i,t} = h_{i,t-1} + \varepsilon_{i,t}^h, \quad \varepsilon_{i,t}^h \sim \mathcal{N}(0, \sigma_{h,i}^2), \; i = 1, \ldots, n + k,
$$

where the initial conditions $h_{1,0}, \ldots, h_{n+k,0}$ are treated as unknown parameters. Following Antolin-Diaz, Drechsel, and Petrella (2021), we set $p = q = 2$.

Using the PC_p criterion proposed in Bai and Ng (2002), McCracken and Ng (2016) find that the optimal number of factors is 8 for their datasets. We therefore set $k = 8$. For identification purposes, the factor loading matrix $\mathbf{A}$ is assumed to be lower triangular with the diagonal elements set to be 1. We then pick the first 8 variables carefully to aid the interpretation of the latent factors. In particular, we draw on the results in McCracken and Ng (2016) and use the variables that load most heavily on each of the factors.[10]

[10]These 8 variables are 'usgood', 't10yffm', 'cusr0000sac', 'aaa', 'gs5', 'ipcongd', 'S&P: indust' and 'exszusx'. These variables do not have missing values in the vintages we consider.

Their results show that the first factor explains a significant portion of the variation in industrial production and many labor market variables, suggesting that it captures the broad economic conditions. The second factor explains particularly well the variations in interest rate spreads, whereas the third and fourth factors have good explanatory power for variations in prices and interest rates, respectively.

Next, we specify the prior distributions on the model parameters. Let $\mathbf{a}$ denote the free elements of the factor loadings matrix $\mathbf{A}$, and let $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ represent the vectors consisting of the diagonal elements of $\boldsymbol{\Psi}_i, i = 1, \ldots, p$ and $\boldsymbol{\Phi}_j, j = 1, \ldots, q$, respectively. Then, consider the following independent priors on $\mathbf{a}$, $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$:

$$\mathbf{a} \sim \mathcal{N}(\mathbf{a}_0, \mathbf{V}_\mathbf{a}), \quad \boldsymbol{\psi} \sim \mathcal{N}(\boldsymbol{\psi}_0, \mathbf{V}_{\boldsymbol{\psi}})1(|\boldsymbol{\psi}| < 1), \quad \boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\phi}_0, \mathbf{V}_{\boldsymbol{\phi}})1(|\boldsymbol{\phi}| < 1),$$

where the indicator functions ensure the elements ψ_i and ϕ_j, $i = 1, \ldots, np, j = 1, \ldots, kq$, are less then 1 in absolute value. We set the prior means $\mathbf{a}_0, \boldsymbol{\psi}_0$ and $\boldsymbol{\phi}_0$ to be zero, and the prior covariance matrices to be $\mathbf{V}_\mathbf{a} = \mathbf{I}_r$ with $r = rn - r(r+1)/2$, $\mathbf{V}_{\boldsymbol{\psi}} = 0.01\mathbf{I}_{np}$ and $\mathbf{V}_{\boldsymbol{\phi}} = 0.01\mathbf{I}_{nq}$. Finally, for the parameters in the stochastic volatility equations, we assume

$$\sigma^2_{h,i} \sim \mathcal{IG}(\nu_{h,i}, S_{h,i}), \quad h_{i,0} \sim \mathcal{N}(m_{h,i}, V_{h_{i,0}}), \ i = 1, \ldots, n + k,$$

where we set $m_{h,i} = 0, V_{h_{i,0}} = 0.01$, $\nu_{h,i} = 3$ and $S_{h,i} = 1$ so that the prior means of $h_{i,0}$ and $\sigma^2_{h,i}$ are 0 and 0.5, respectively.

This dynamic factor model with an unbalanced panel can be estimated using MCMC methods. More specifically, given the model parameters and the latent factors, we simulate the missing values using the proposed sampler as described in Section 2. Then, given the sampled missing values, the model parameters and latent factors can be drawn from their full conditional distributions using standard algorithms; see, e.g., see Antolin-Diaz, Drechsel, and Petrella (2021) and Chan (2022). To assess the efficiency of the posterior sampler, we compute the inefficiency factors associated with the missing data and the model parameters (see Appendix C for details). In particular, all inefficiency factors are less than 100, and they are comparable to those of a dynamic factor model with a balanced panel.

We estimate the dynamic factor model using FRED-MD data vintages from March 2020 to September 2022 that cover the COVID-19 pandemic period. For each data vintage, we

first transform the series according to the recommendation in McCracken and Ng (2016). Following common practice, we then standardize each series so that it has 0 mean and unit variance. As mentioned earlier, each vintage contains 128 monthly variables, but many have missing values at the beginning or at the end of the sample. For comparison, we also estimate a version of the dynamic factor model using only variables without missing values, i.e., in each vintage we omit any variables that have missing values. Across the data vintages we consider, on average about 24 variables have missing values and are omitted from the estimation of the factors.

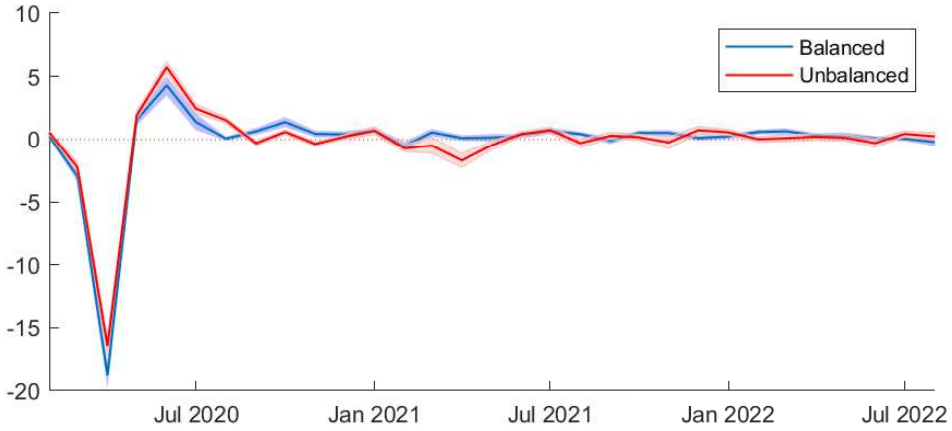

Figure 6: Filtered estimates of the first factor from the dynamic factor model with balanced and unbalanced datasets. The black line denotes the posterior means of the first factor obtained using all 128 variables (with missing data), and the gray shaded area represents the associated 68% credible intervals. The blue line denotes the posterior means obtained using only variables without missing values, and the blue shaded area represents the associated 68% intervals.

Figure 6 plots the filtered estimates of the first factor from the dynamic factor model with balanced and unbalanced datasets. These two estimates are broadly similar, and they track well the pronounced downturn at the onset of the COVID-19 pandemic and the subsequent rebound, confirming that the first factor captures the broad economic conditions. However, they also show noticeable differences, especially at the peak and trough associated with the economy-wide reopening after the lock-down. In particular, the factor estimates obtained using all 128 variables show a less severe down-turn (-16.4 vs -18.8) and a more pronounced uptick afterward (5.7 vs 4.3). These differences may be attributed to the missing values in a number of orders and inventories variables,

such as new orders for consumer goods ('acogno') and for capital goods ('andenox') at the beginning of the sample and total business inventories ('businvx') at the end of the sample.

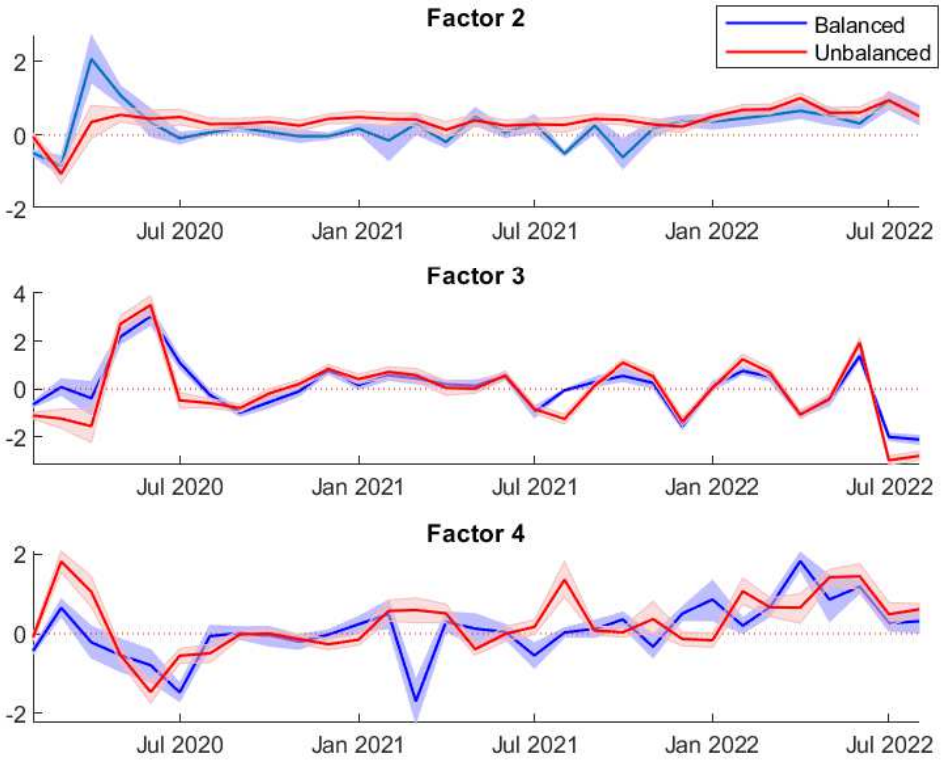

Figure 7: Filtered estimates of the second, third and fourth factors from the dynamic factor model with balanced and unbalanced datasets. The black line denotes the posterior means of the factor obtained using all 128 variables (with missing data), and the gray shaded area represents the associated 68% credible intervals. The blue line denotes the posterior means obtained using only variables without missing values, and the blue shaded area represents the associated 68% intervals.

Next, we report in Figure 7 the filtered estimates of the second, third and fourth factors. There are more substantial differences between the factor estimates obtained using the balanced vs unbalanced datasets. In particular, the most striking differences are those for the fourth factor, which explains variations in interest rates particularly well. These differences might reflect the fact that many of the variables with missing values load heavily on the fourth factor. For example, a number of new private housing permits

variables have missing values, and they presumably contain useful information on interest rates. Ignoring those variables is likely to give an incomplete picture on the development of interest rates.[11]

All in all, these results suggest that omitting variables that have missing values from the empirical analysis can potentially misrepresent the dynamics of broad economic conditions and co-movements in interest rates or prices. This further underlines the utility of the proposed approach to impute missing values that is flexible and works well in large-scale models.

5 Conclusion

We have introduced a novel and efficient approach—that is applicable to any conditionally Gaussian state space models and datasets with arbitrary missing data patterns—for sampling all the missing observations in one step. We have showed via a series of Monte Carlo simulations that the proposed approach is more computationally efficient than standard Kalman-filter based methods under a wide variety of settings. We also demonstrated how the proposed approach can be applied to two empirical macroeconomic applications involving a large mixed-frequency VAR and a dynamic factor model with unbalanced datasets. Both empirical applications illustrated the usefulness of incorporating more information (from high-frequency indicators or variables with missing values) into macroeconomic analysis.

[11] Filtered estimates of the fifth to eighth factors are reported in Appendix C. We also find substantial differences in the estimates from the balanced vs unbalanced datasets.

Appendix A: Dataset for the Mixed-Frequency VAR

This appendix provides details of the 22-variable dataset of the mixed-frequency VAR application. Specifically, Table 3 describes the 22 variables and their transformations. The first seven weekly variables are obtained from Lewis, Mertens, Stock, and Trivedi (2022), and the rest are sourced from the FRED database. The sample period is from January 2013 to August 2022.

Table 3: The list of variables and the corresponding transformation used in the mixed-frequency VAR application.

Variable	Frequency	FRED mnemonic	Transformation
Fuels	Weekly	-	$100\Delta\ln\left(\frac{x_t}{x_{t-52}}\right)$
Raw Steel Production	Weekly	-	$100\Delta\ln\left(\frac{x_t}{x_{t-52}}\right)$
Retail Sales Average	Weekly	-	Level
Electric Utility Output	Weekly	-	$100\Delta\ln\left(\frac{x_t}{x_{t-52}}\right)$
US Railroad Traffic	Weekly	-	$100\Delta\ln\left(\frac{x_t}{x_{t-52}}\right)$
Continued Claims	Weekly	CCSA	$\ln x_t$
Initial Claims	Weekly	ICSA	$\ln x_t$
US Regular All Formulations Gas Price	Weekly	GASREGW	$100\Delta\ln\left(\frac{x_t}{x_{t-52}}\right)$
Crude Oil Prices	Weekly	WCOILWTICO	$\ln x_t$
National Financial Conditions Index	Weekly	NFCI	Level
S&P 500	Weekly	S&P 500	$100\Delta\ln\left(\frac{x_t}{x_{t-1}}\right)$
Yield on 1-year U.S. Treasury	Weekly	DGS1	Level
Yield on 10-year U.S. Treasury	Weekly	DGS10	Level
Moody's Seasoned Baa Corporate Bond Yield	Weekly	WBAA	Level
Moody's Seasoned Aaa Corporate Bond Yield	Weekly	WAAA	Level
VIX	Weekly	VIXCLS	Level
Industrial Production	Monthly	INDPRO	$100\Delta\ln\left(\frac{x_t}{x_{t-1}}\right)$
Consumer Price Index for All Urban Consumers	Monthly	CPIAUCSL	$100\Delta\ln\left(\frac{x_t}{x_{t-1}}\right)$
Unemployment Rate	Monthly	UNRATE	Level
All Employees, Total Nonfarm	Monthly	PAYEMS	$100\Delta\ln\left(\frac{x_t}{x_{t-1}}\right)$
Average Weekly Hours: Manufacturing	Monthly	AWHMAN	$\frac{x_t}{10}$
Real Gross Domestic Product	Quarterly	GDPC1	$400\Delta\ln\left(\frac{x_t}{x_{t-1}}\right)$

Appendix B: The Durbin-Koopman Simulation Smoother

To implement the Durbin-Koopman simulation smoother, we first need to rewrite the VAR(5) into its companion form:

$$\mathbf{s}_t = \mathbf{F}_0 + \mathbf{F}_1 \mathbf{s}_{t-1} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}_{5n}, \mathbf{\Omega}), \tag{16}$$

where $\mathbf{s}_t = (\mathbf{y}_t^{'}, \mathbf{y}_{t-1}^{'}, \mathbf{y}_{t-2}^{'}, \mathbf{y}_{t-3}^{'}, \mathbf{y}_{t-4}^{'})$ is the $5n \times 1$ state vector and

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0}_{n\times 4n} \\ \mathbf{0}_{4n\times n} & \mathbf{0}_{4n\times 4n} \end{bmatrix}, \quad \mathbf{F}_0 = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{0}_{4n\times 1} \end{bmatrix}, \quad \mathbf{F}_1 = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{B}_3 & \mathbf{B}_4 & \mathbf{B}_5 \\ \mathbf{I}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{I}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{I}_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{I}_n & \mathbf{0}_n \end{bmatrix}.$$

When both the high- and low-frequency variables are observed, the measurement equation is given by

$$\mathbf{y}_t = \mathbf{\Lambda}_1 \mathbf{s}_t. \tag{17}$$

When only the high-frequency variables are observed, it becomes

$$\mathbf{y}_t^o = \mathbf{\Lambda}_2 \mathbf{s}_t. \tag{18}$$

The matrices $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ incorporate the inter-temporal constraints:

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \mathbf{I}_{n^o} & \mathbf{0}_{n^m} & \mathbf{0}_{n^o} & \mathbf{0}_{n^m} & \mathbf{0}_{n^o} & \mathbf{0}_{n^m} & \mathbf{0}_{n^o} & \mathbf{0}_{n^m} & \mathbf{0}_{n^o} & \mathbf{0}_{n^m} \\ \mathbf{0}_{n^m} & \frac{1}{3}\mathbf{I}_{n^m} & \mathbf{0}_{n^o} & \frac{2}{3}\mathbf{I}_{n^m} & \mathbf{0}_{n^o} & \mathbf{I}_{n^m} & \mathbf{0}_{n^o} & \frac{2}{3}\mathbf{I}_{n^m} & \mathbf{0}_{n^o} & \frac{1}{3}\mathbf{I}_{n^m} \end{bmatrix},$$
$$\mathbf{\Lambda}_2 = \begin{bmatrix} \mathbf{I}_{n^o} & \mathbf{0}_{n^m} & \mathbf{0}_{n^o} & \mathbf{0}_{n^m} & \mathbf{0}_{n^o} & \mathbf{0}_{n^m} & \mathbf{0}_{n^o} & \mathbf{0}_{n^m} & \mathbf{0}_{n^o} & \mathbf{0}_{n^m} \end{bmatrix}.$$

We follow Jarociński (2015) and implement the Durbin-Koopman simulation smoother on (16), (17) and (18).

Appendix C: Additional Results

This appendix presents additional empirical results from the two applications. We first report in Figure 8 the filtered estimates of factors 5-8 in the dynamic factor model using both balanced and unbalanced FRED-MD datasets.

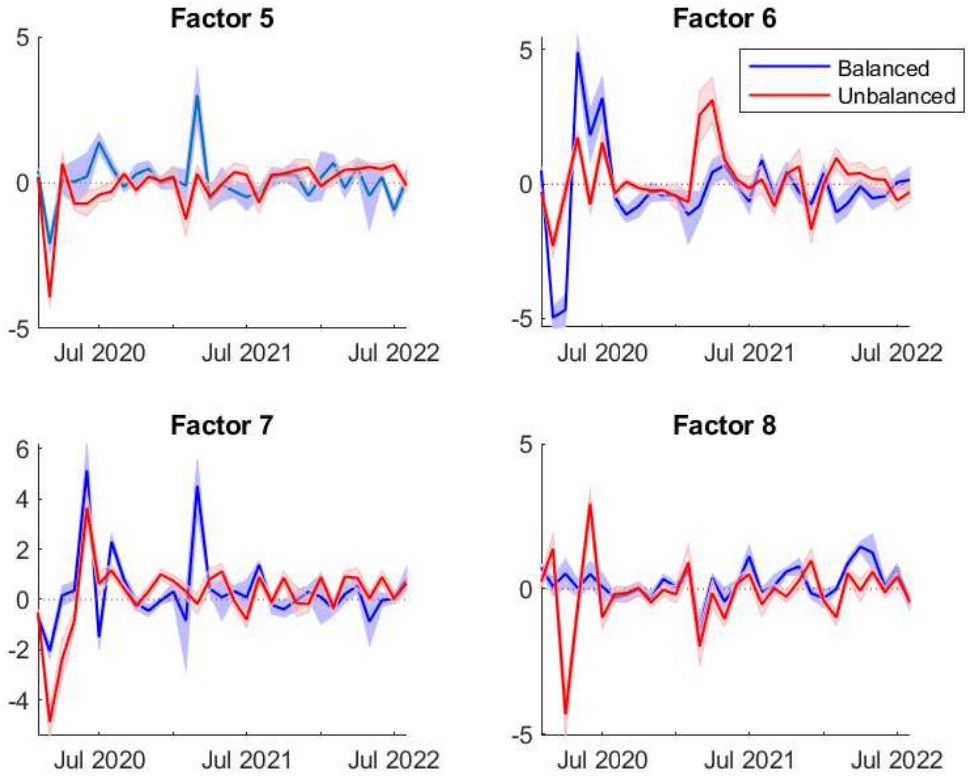

Figure 8: Filtered estimates of the fifth to eighth factors from the dynamic factor model with balanced and unbalanced datasets. The black line denotes the posterior means of the factor obtained using all 128 variables (with missing data), and the gray shaded area represents the associated 68% credible intervals. The blue line denotes the posterior means obtained using only variables without missing values, and the blue shaded area represents the associated 68% intervals.

Next, Figures 9-11 report the inefficiency factors of the MCMC samples from the two empirical applications. Specifically, Figure 9 plots the inefficiency factors of the missing data and model parameters from the weekly state-space mixed-frequency VAR. To present the information more succinctly, boxplots of the inefficiency factors are used. Each boxplot

summarizes the quantiles of each group of model parameters or latent variables. The middle line of each box denotes the median, while the lower and upper lines represent, respectively, the 25- and the 75-percentiles. The whiskers extend to the maximum and minimum. For example, the boxplot associated with $\mathbf{y}^m$ indicates that the median of the inefficiency factors of the missing data is about 65, and the maximum is about 90. Similarly, the boxplot associated with $\boldsymbol{\beta}$ shows that 75% of the inefficiency factors of the VAR coefficients are below 60, with the maximum being just below 100.

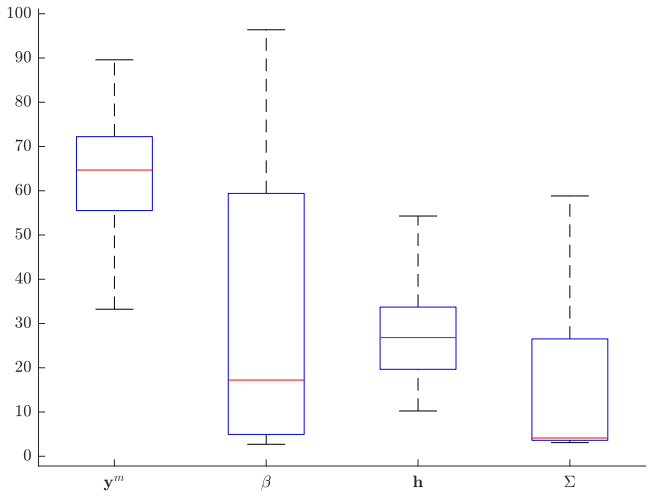

Figure 9: Boxplots of the inefficiency factors corresponding to the posterior draws of the missing data $\mathbf{y}^m$ and model parameters, $\boldsymbol{\beta}, \mathbf{h}$ and $\boldsymbol{\Sigma}$, from the weekly state-space mixed-frequency VAR.

Figures 10 and 11 plot the inefficiency factors from the dynamic factor model with an unbalanced panel and a balanced panel, respectively. Again, each of the boxplot summarizes the quantiles of each group of model parameters or latent variables.

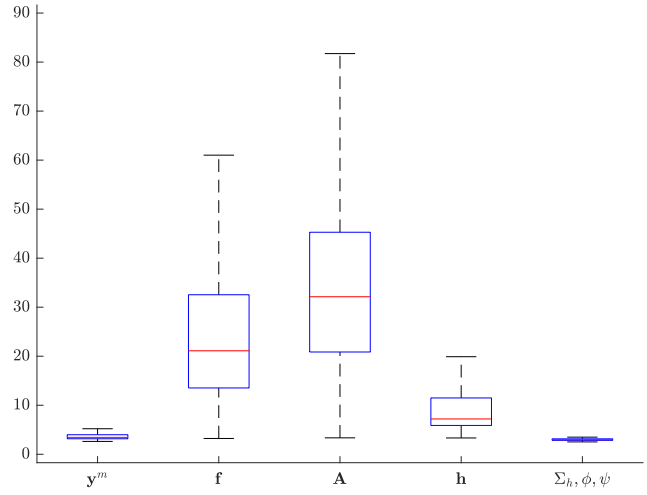

Figure 10: Boxplots of the inefficiency factors corresponding to the posterior draws of the missing data $\mathbf{y}^m$ and model parameters, $\mathbf{f}$, $\mathbf{A}$, $\mathbf{h}$, $\boldsymbol{\sigma}_h^2, \boldsymbol{\psi}$ and $\boldsymbol{\phi}$ from the dynamic factor model with an unbalanced panel.

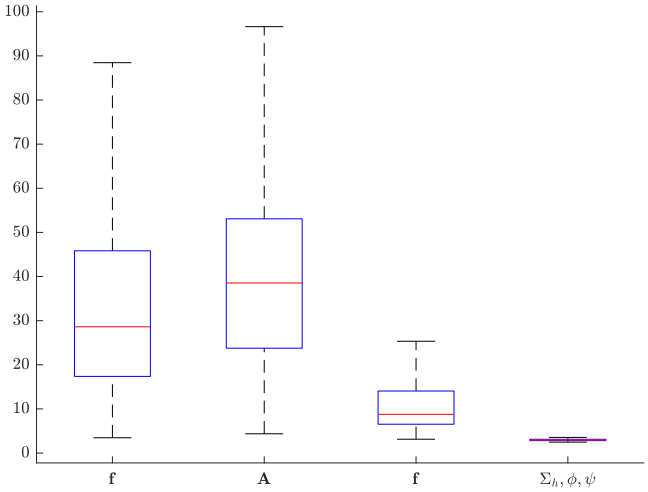

Figure 11: Boxplots of the inefficiency factors corresponding to the posterior draws of $\mathbf{f}$, $\mathbf{A}$, $\mathbf{h}$, $\boldsymbol{\sigma}_h^2$, $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ from the dynamic factor model with a balanced panel.

Appendix D: Joint Sampling of the Missing Data in the Sample and the Initial Conditions

In VAR applications, the initial conditions $\mathbf{y}_0, \ldots, \mathbf{y}_{1-p}$ often contain missing values. This appendix describes an alternative approach of jointly sampling the missing data in both the sample and the initial conditions. It amounts to redefining a few relevant matrices to include the initial conditions. Then, the algorithms described in the main text can be directly applied.

To that end, redefine $\mathbf{y} = (\mathbf{y}'_{1-p}, \ldots, \mathbf{y}'_0, \mathbf{y}'_1, \ldots, \mathbf{y}'_T)'$ to denote the $(T+p)n \times 1$ vector of initial conditions and sample observations. Similarly, let $\mathbf{y}^o$ and $\mathbf{y}^m$ represent, respectively, the N^o-vector of observables and the N^m-vector of missing data, with $N^o + N^m = (T+p)n$. Next, we can write $\mathbf{y}$ in terms of $\mathbf{y}^o$ and $\mathbf{y}^m$: $\mathbf{y} = \mathbf{S}^o\mathbf{y}^o + \mathbf{S}^m\mathbf{y}^m$, where $\mathbf{S}^o$ and $\mathbf{S}^m$ are, respectively, $(T+p)n \times N^o$ and $(T+p)n \times N^m$ selection matrices. Then, stacking (6) over $t = 1, \ldots, T$, we obtain

$$\mathbf{H}_\mathbf{B}\mathbf{y} = \mathbf{c}_\mathbf{B} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_{Tn}, \boldsymbol{\Sigma}), \tag{19}$$

where $\mathbf{c}_\mathbf{B} = \mathbf{1}_T \otimes \mathbf{b}_0$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \ldots, \boldsymbol{\varepsilon}'_T)'$, $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_T)$, and

$$\mathbf{H}_\mathbf{B} = \begin{bmatrix} -\mathbf{B}_p & \cdots & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & \cdots & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & -\mathbf{B}_p & \cdots & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} & -\mathbf{B}_p & \cdots & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \cdots & \cdots & \mathbf{0}_{n\times n} & -\mathbf{B}_p & \cdots & -\mathbf{B}_1 & \mathbf{I}_n \end{bmatrix}.$$

Note that (19) has exactly the same form as (7), with the new definitions of $\mathbf{y}$, $\mathbf{c}_\mathbf{B}$ and $\mathbf{H}_\mathbf{B}$. As such, the algorithms in the main text can be directly applied to sample the missing data $\mathbf{y}^m$ in $\mathbf{y}$. For example, the joint conditional distribution of the missing data given the observed data is $(\mathbf{y}^m \,|\, \mathbf{y}^o, \boldsymbol{\alpha}, \boldsymbol{\theta}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}^m}, \mathbf{K}_{\mathbf{y}^m}^{-1}\right)$, where $\mathbf{K}_{\mathbf{y}^m} = \mathbf{G}^{m\prime}\boldsymbol{\Sigma}^{-1}\mathbf{G}^m$, $\boldsymbol{\mu}_{\mathbf{y}^m} = \mathbf{K}_{\mathbf{y}^m}^{-1}\mathbf{G}^{m\prime}\boldsymbol{\Sigma}^{-1}(\mathbf{c}_\mathbf{B} - \mathbf{G}^o\mathbf{y}^o)$, with $\mathbf{G}^o = \mathbf{H}_\mathbf{B}\mathbf{S}^o$ and $\mathbf{G}^m = \mathbf{H}_\mathbf{B}\mathbf{S}^m$.

References

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2021): "Advances in nowcasting economic activity: Secular trends, large shocks and new data," .

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

Bai, J., and S. Ng (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70(1), 191–221.

——— (2021): "Matrix completion, counterfactuals, and factor analysis of missing data," *Journal of the American Statistical Association*, 116(536), 1746–1763.

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

Beyeler, S., and S. Kaufmann (2021): "Reduced-form factor augmented VAR—Exploiting sparsity to include meaningful factors," *Journal of Applied Econometrics*, forthcoming.

Brave, S. A., R. A. Butters, and A. Justiniano (2019): "Forecasting economic activity with mixed frequency BVARs," *International Journal of Forecasting*, 35(4), 1692–1707.

Carriero, A., J. C. C. Chan, T. E. Clark, and M. G. Marcellino (2022): "Corrigendum to "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors"," *Journal of Econometrics*, 227(2), 506–512.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2022): "Addressing COVID-19 outliers in BVARs with stochastic volatility," *The Review of Economics and Statistics*, Forthcoming.

Carriero, A., T. E. Clark, and M. G. Marcellino (2016): "Common drifting volatility in large Bayesian VARs," *Journal of Business and Economic Statistics*, 34(3), 375–390.

——— (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212(1), 137–154.

Carter, C. K., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Chan, J. C. C. (2013): "Moving Average Stochastic Volatility Models with Application to Inflation Forecast," *Journal of Econometrics*, 176(2), 162–172.

——— (2017): "The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling," *Journal of Business and Economic Statistics*, 35(1), 17–28.

——— (2020): "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business and Economic Statistics*, 38(1), 68–79.

——— (2022): "Comparing stochastic volatility specifications for large Bayesian VARs," *Journal of Econometrics*, forthcoming.

Chan, J. C. C., E. Eisenstat, and G. Koop (2016): "Large Bayesian VARMAs," *Journal of Econometrics*, 192(2), 374–390.

Chan, J. C. C., and I. Jeliazkov (2009): "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1), 101–120.

Chan, J. C. C., G. Koop, and S. M. Potter (2013): "A new model of trend inflation," *Journal of Business and Economic Statistics*, 31(1), 94–106.

Chib, S., E. Greenberg, and I. Jeliazkov (2009): "Estimation of Semiparametric Models in the Presence of Endogeneity and Sample Selection," *Journal of Computational and Graphical Statistics*, 18, 321–348.

Chib, S., and I. Jeliazkov (2006): "Inference in semiparametric dynamic models for binary longitudinal data," *Journal of the American Statistical Association*, 101(474), 685–700.

Clark, T. E. (2011): "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility," *Journal of Business and Economic Statistics*, 29(3), 327–341.

Cogley, T., and T. J. Sargent (2005): "Drifts and volatilities: Monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8(2), 262–302.

Cong, Y., B. Chen, and M. Zhou (2017): "Fast simulation of hyperplane-truncated multivariate normal distributions," *Bayesian Analysis*, 12(4), 1017–1037.

Cross, J., C. Hou, and A. Poon (2019): "Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity," *International Journal of Forecasting*, 36(3), 899–915.

Cross, J., and A. Poon (2016): "Forecasting structural change and fat-tailed events in Australian macroeconomic variables," *Economic Modelling*, 58, 34–51.

D'Agostino, A., L. Gambetti, and D. Giannone (2013): "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28, 82–101.

Dimitrakopoulos, S., and M. Kolossiatis (2020): "Bayesian analysis of moving average stochastic volatility models: modeling in-mean effects and leverage for financial time series," *Econometric Reviews*, 39(4), 319–343.

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

Durbin, J., and S. J. Koopman (2002): "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89, 603–615.

Eckert, F., P. Kronenberg, H. Mikosch, and S. Neuwirth (2020): "Tracking economic activity with alternative high-frequency data," *KOF Working Papers*, 488, 1–35.

Fu, B. (2020): "Is the slope of the Phillips curve time-varying? Evidence from unobserved components models," *Economic Modelling*, 88, 320–340.

Grant, A. L., and J. C. C. Chan (2017a): "A Bayesian Model Comparison for Trend-Cycle Decompositions of Output," *Journal of Money, Credit and Banking*, 49(2-3), 525–552.

——— (2017b): "Reconciling output gaps: Unobserved components model and Hodrick-Prescott filter," *Journal of Economic Dynamics and Control*, 75, 114–121.

Hartwig, B. (2021): "Bayesian VARs and Prior Calibration in Times of COVID-19," *Available at SSRN 3792070*.

Hauber, P., and C. Schumacher (2021): "Precision-based sampling with missing observations: A factor model application," Discussion Papers 11/2021, Deutsche Bundesbank.

Hou, C. (2020): "Time-varying relationship between inflation and inflation uncertainty," *Oxford Bulletin of Economics and Statistics*, 82(1), 83–124.

Jarociński, M. (2015): "A note on implementing the Durbin and Koopman simulation smoother," *Computational Statistics and Data Analysis*, 91, 1–3.

Kastner, G. (2019): "Sparse Bayesian time-varying covariance estimation in many dimensions," *Journal of econometrics*, 210(1), 98–115.

Kastner, G., and F. Huber (2020): "Sparse Bayesian vector autoregressions in huge dimensions," *Journal of Forecasting*, 39(7), 1142–1165.

Kaufmann, S., and C. Schumacher (2019): "Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification," *Journal of Econometrics*, 210(1), 116–134.

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

Koop, G., S. McIntyre, J. Mitchell, and A. Poon (2020): "Regional output growth in the United Kingdom: More timely and higher frequency estimates from 1970," *Journal of Applied Econometrics*, 35(2), 176–197.

——— (2022): "Reconciled Estimates of Monthly GDP in the US," *Journal of Business and Economic Statistics*, Forthcoming.

Lenza, M., and G. E. Primiceri (2022): "How to Estimate a VAR after March 2020," *Journal of Applied Econometrics*, Forthcoming.

Lewis, D. J., K. Mertens, J. H. Stock, and M. Trivedi (2022): "Measuring real activity using a weekly economic index," *Journal of Applied Econometrics*, 37(4), 667–687.

Litterman, R. (1986): "Forecasting With Bayesian Vector Autoregressions — Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25–38.

Little, R. J., and D. B. Rubin (2019): *Statistical analysis with missing data*, vol. 793. John Wiley & Sons, third edn.

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.

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

McCausland, W. J., S. Miller, and D. Pelletier (2011): "Simulation smoothing for state-space models: A computational efficiency analysis," *Computational Statistics and Data Analysis*, 55(1), 199–212.

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

Ng, S., and S. Scanlan (2023): "Constructing High Frequency Economic Indicators by Imputation," *arXiv preprint arXiv:2303.01863*.

Pettenuzzo, D., R. Sabbatucci, and A. Timmermann (2023): "Dividend suspensions and cash flows during the Covid-19 pandemic: A dynamic econometric model," *Journal of Econometrics*, Forthcoming.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

Rue, H. (2001): "Fast sampling of Gaussian Markov random fields with applications," *Journal of the Royal Statistical Society: Series B (Methodological)*, 63(2), 325–338.

Rue, H., and L. Held (2005): *Gaussian Markov Random Fields: Theory and Applications*. CRC press.

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

Stock, J. H., and M. W. Watson (2016): "Core inflation and trend inflation," *Review of Economics and Statistics*, 98(4), 770–784.

Zhang, B., J. C. C. Chan, and J. L. Cross (2020): "Stochastic volatility models with ARMA innovations: An application to G7 inflation forecasts," *International Journal of Forecasting*, 36(4), 1318–1328.