DEUTSCHE
BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 11/2021

Precision-based sampling with missing observations: A factor model application

Philipp Hauber
(Kiel Institute for the World Economy)

Christian Schumacher
(Deutsche Bundesbank)

Discussion Papers represent the authors' personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

Editorial Board: Daniel Foos
Stephan Jank
Thomas Kick
Malte Knüppel
Vivien Lewis
Christoph Memmel
Panagiota Tzamourani

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

Reproduction permitted only if source is stated.

ISBN 978–3–95729–819–5 (Internetversion)

Non-technical summary

Research Question

In many central banks, models with unobserved components or time-varying parameters are now routinely applied for policy analysis and forecasting. When the models are estimated in a Bayesian framework with iterative algorithms, recurrent sampling of the unobserved components and time-varying parameters is necessary.

Contribution

The paper proposes a conditional sampler for unobserved states in state-space models when observations are missing in the data. To illustrate the method, the approach is applied to a large dynamic factor model with many variables. The approach extends the existing literature on precision-based sampling, where typically complete-data applications are considered. The sampler not only provides estimates of the factors, but also of the missing values in the data. The paper compares the estimation accuracy and the computing time of different precision-based samplers. As an empirical application, we estimate international factors in GDP growth and compare estimation results based on balanced data and results on larger, unbalanced data.

Results

The estimation accuracy of the different precision-based samplers turns out to be very similar. In terms of computational speed, a 2-step approach performs best, which sequentially samples factors and missing observations. In the empirical application we find that global factors and the common components for GDP growth of the G7 countries are very similar when estimated on balanced and unbalanced data. However, African and Asian factors are more precisely estimated when using the larger, unbalanced data.

Nichttechnische Zusammenfassung

Fragestellung

In vielen Zentralbanken werden Modelle mit unbeobachteten Komponenten oder zeitvariierenden Parametern, sogenannte Zustandsraummodelle, regelmäßig für die Analyse wirtschaftspolitischer Maßnahmen und Prognosen verwendet. Bei der bayesianischen Schätzung solcher Modelle kommen häufig iterative Algorithmen zum Einsatz, in denen wiederholt aus der stochastischen Verteilung der unbeobachteten Komponenten oder zeitvariierenden Parameter gezogen wird.

Beitrag

Dieses Papier schlägt ein Verfahren zum Ziehen von unbeobachteten Komponenten in Zustandsraummodellen vor, wenn der verwendete Datensatz zum Teil fehlende Beobachtungen aufweist. Die Methode wird anhand eines Faktormodells mit einem großen Datensatz veranschaulicht. Der Ansatz ergänzt die bestehende Literatur zur Ziehung auf Basis von Präzisionsmatrizen, die typischerweise von fehlenden Beobachtungen abstrahiert. Das Verfahren liefert dabei nicht nur eine Schätzung der unbeobachteten Faktoren des Modells, sondern auch der fehlenden Beobachtungen. Das Papier vergleicht die Schätzgenauigkeit und die Rechenzeit verschiedener präzisionsbasierter Ansätze in Simulationen. In der empirischen Anwendung werden internationale Faktoren aus den Veränderungsraten des BIP für verschiedene Länder geschätzt. Wir vergleichen Ergebnisse auf Grundlage von balancierten Daten, bei denen Zeitreihen mit fehlenden Beobachtungen aus den Daten entfernt wurden, mit Ergebnissen auf Basis des umfangreicheren Datensatzes, in dem Datenlücken bestehen.

Ergebnisse

Die Schätzgenauigkeit der verschiedenen präzisionsbasierten Ansätze unterscheidet sich nur unwesentlich. In Bezug auf die Rechenzeit schneidet ein iterativer 2-Schritt-Ansatz am besten ab, bei dem erst die Faktoren und dann die fehlenden Beobachtungen geschätzt werden. In der empirischen Anwendung zeigen sich geringe Unterschiede beim globalen Faktor und den gemeinsamen Komponenten der BIP-Veränderungsraten der G7-Länder, wenn man Ergebnisse auf Basis balancierter und unbalancierter Daten vergleicht. Faktoren für afrikanische und asiatische Länder werden jedoch genauer geschätzt, wenn der umfangreiche Datensatz mit teilweise fehlenden Beobachtungen verwendet wird.

Deutsche Bundesbank Discussion Paper No 11/2021

Precision-based sampling with missing observations: A factor model application*

Philipp Hauber
Kiel Institute for the World Economy

Christian Schumacher
Deutsche Bundesbank

Abstract

We propose a new approach to sample unobserved states conditional on available data in (conditionally) linear unobserved component models when some of the observations are missing. The approach is based on the precision matrix of the states and model variables, which is sparse and banded in many economic applications and allows for efficient sampling. The existing literature on precision-based sampling is focused on complete-data applications, whereas the proposed samplers in this paper provide draws for states and missing observations by using permutations of the precision matrix. The approaches can be easily integrated into Bayesian estimation procedures like the Gibbs sampler. By allowing for incomplete data sets, the proposed sampler expands the range of potential applications for precision-based samplers in practice. We derive the sampler for a factor model, although it can be applied to a wider range of empirical macroeconomic models. In an empirical application, we estimate international factors in GDP growth in a large unbalanced data set of about 180 countries.

Keywords: Precision-based sampling, Bayesian estimation, state-space models, missing observations, factor models, banded matrices

JEL classification: C32, C38, C63, C55.

*Contact address: Christian Schumacher, Deutsche Bundesbank, Research Centre, Wilhelm-Epstein-Str. 14, 60431 Frankfurt, Germany. E-Mail: christian.schumacher@bundesbank.de. Helpful comments were received by William McCausland, Elmar Mertens and seminar participants at the Deutsche Bundesbank. We thank Max Schröder and Jiachun Zhang for research assistance. The computer codes for this paper were implemented in Matlab 2018a and are available upon request. Bundesbank Discussion Papers represent the authors' personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

1 Introduction

In the recent literature, conditional samplers for unobserved states given data and parameters based on the precision matrix have received considerable attention. First applications of precision-based samplers to address economic questions have been provided by Chan and Jeliazkov (2009) and McCausland (2012), building on seminal work by Rue (2001) and Rue and Held (2005) on Gaussian Markov random fields. There is now a huge number of applications of precision-based samplers in the empirical macroeconomic literature. Recent examples of state-space models with unobserved components such as output gaps or inflation trends and time-varying parameters are Chan, Koop, and Potter (2013, 2016), Grant and Chan (2017), and Chan, Clark, and Koop (2018) or time-varying parameter vector autoregressive (VAR) models with a vast number of applications such as Chan and Eisenstat (2018), Chan (2020), Chan, Eisenstat, and Strachan (2020) and references cited therein. Factor model applications are provided by Chan and Jeliazkov (2009), McCausland (2015), and Kaufmann and Schumacher (2017, 2019). These application typically employ precision-based sampling of states given data as part of a Bayesian estimation procedures like the Gibbs sampler, whose aim is to draw from the posterior density $p(\theta, \eta|x)$, where η are the unobserved states, x denotes data, and θ are model parameters. A standard Gibbs sampler iterates between drawing from the conditional posteriors $p(\eta|x, \theta)$ and $p(\theta|x, \eta)$. Drawing from $p(\eta|x, \theta)$ can be carried out efficiently by precision-based samplers, as the underlying precision matrix of states and variables is banded in many economic models and allows for the application of fast sparse matrix techniques (Rue, 2001).

The literature cited above applies precision-based samplers to complete data sets. In practice, however, observations in multivariate data sets can often be missing. In this case, an analyst has the choice of removing all time series with missing observation and using balanced data only. This, however, implies a loss of information. The alternative is to use the larger unbalanced data, but this raises the need for estimation methods that can tackle missing observations.

In this paper, we propose a precision-based sampler for unobserved states in the presence of partly missing observations. In line with the literature cited above, the state-space model is assumed to be (conditionally) linear and the disturbances follow normal distributions. If the data are completely available, the states can be sampled from a conditional distribution of a multivariate normal using fast band-matrix computation as in Rue (2001) and Chan and Jeliazkov (2009). Important alternative samplers from the literature based on the Kalman filter are provided in the seminal papers Carter and Kohn (1994), Frühwirth-Schnatter (1994), and Durbin and Koopman (2002). The main contribution of this paper is the extension of the literature on precision-based samplers by considering missing observations. We do so by implementing an efficient reordering of states, observed and unobserved variables that facilitates fast band-matrix computation. The sampler provides draws from the conditional posterior distribution $p(\eta, x_m|x_o, \theta)$ for the states and the missing observations x_m conditional on observed data x_o and parameters. Thus it can easily be integrated into Gibbs samplers to tackle missing data as proposed by Little and Rubin (2002).

To illustrate the sampling method, we use a factor model with vector autoregressive (VAR) dynamics for the factors and autoregressive (AR) idiosyncratic components in a

Bayesian framework (McCausland, 2015; Kaufmann and Schumacher, 2019). Alternative missing-data approaches for factor models in the literature are Angelini, Henry, and Marcellino (2006) and Marcellino (2007) for backdating and interpolation in a principal-components framework, see also Bai and Ng (2019) for a recent contribution. In a Bayesian framework, Otrok and Pourpourides (2017) interpolate data in panel data and Müller, Stock, and Watson (2019) interpolate international long-run growth data. Further approaches based on the Kalman filter have been proposed by Jungbacker, Koopman, and van der Wel (2011) and Banbura and Modugno (2014), amongst others.

For the factor model, we derive alternative precision-based samplers for the factors and missing values in the data, which differ with respect to the permutations of η, x_m and x_o in the precision matrix. We compare the accuracy and computational efficiency of the precision-based samplers in simulations. As an empirical application, we estimate international factors in GDP growth along the lines of the literature on international business cycles with Bayesian techniques (Kose, Otrok, and Whiteman, 2003, 2008; Francis, Owyang, and Savascin, 2017; Müller et al., 2019). We compare estimation results based on balanced data for about 50 country-GDP time series using the standard precision-based sampler and results on a larger, unbalanced data set consisting of more than 180 GDP time series. We check whether results based on balanced data are robust when using the larger information set.

The paper proceeds as follows. In Section 2, we introduce the factor model, whereas Section 3 describes its estimation using Bayesian methods given complete data with a focus on precision-based sampling of the unobserved factors. Section 4 provides alternative precision-based samplers for partly missing observations. A simulation exercise to compare the alternative precision-based samplers is provided in Section 5. In Section 6, we discuss the results of the empirical application. Section 7 briefly discusses the calculation of the marginal likelihood and extensions to other models such as time-varying parameter (TVP) Bayesian VAR models. Section 8 concludes.

2 The factor model

The factor model explains the $(N \times 1)-$dimensional vector of variables $x_t = (x_{1,t}, x_{2,t}, \ldots, x_{N,t})^\mathsf{T}$ in time period t according to

$$x_t = \lambda \eta_t + \epsilon_t, \tag{1}$$

$$\eta_t = \phi \eta_{t-1} + u_{\eta,t}, \quad \epsilon_t = \psi \epsilon_{t-1} + u_{\epsilon,t}. \tag{2}$$

The $(r \times 1)-$dimensional vector of factors is denoted as η_t, and λ is the $(N \times r)-$dimensional matrix of factor loadings. The factor representation (1) holds for $t = 1, \ldots, T$. The factors follow a VAR(1) process with the $(r \times r)-$dimensional lag parameter matrix ϕ. The factor VAR disturbances are distributed as $u_{\eta,t} \sim \mathcal{N}(0_{r \times 1}, \omega_\eta)$. The idiosyncratic components collected in the $(N \times 1)-$dimensional vector $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{N,t})^\mathsf{T}$ each follow AR(1) processes such that the $(N \times N)-$dimensional coefficient matrix ψ is diagonal containing the AR(1) lag parameters ψ_i for $i = 1, \ldots, N$ on the main diagonal. The one-lag specification only serves to illustrate the methods. The empirical applications later in the paper will consider a factor VAR(p) with AR(q) idiosyncratic components where $p, q > 1$. The idiosyncratic disturbances are distributed as $u_{\epsilon,t} \sim \mathcal{N}(0_{N \times 1}, \omega_\epsilon)$, where ω_ϵ

is also diagonal with diagonal elements $\omega_{\epsilon,i}$ for $i = 1, \ldots, N$. We assume that $u_{\eta,t}$ and $u_{\epsilon,t}$ are mutually independent and that the VAR and AR processes in (2) are stationary. In addition, the equations in (2) are defined for time periods $t = 2, \ldots, T$. For $t = 1$, let $\eta_1 \sim \mathcal{N}(\eta_{1|0}, \vartheta_{\eta,1|0})$ and $\epsilon_1 \sim \mathcal{N}(\epsilon_{1|0}, \vartheta_{\epsilon,1|0})$, respectively. The means of the distributions are set equal to their unconditional mean, which is zero in our case, so $\eta_{1|0} = 0_{r\times 1}$ and $\epsilon_{1|0} = 0_{N\times 1}$, respectively. As the processes underlying (2) are stationary, $\vartheta_{\eta,1|0}$ and $\vartheta_{\epsilon,1|0}$ are set equal to the unconditional covariances implied by the model equations: For the states, we define $\vartheta_{\eta,1|0} = \vartheta_\eta$, where ϑ_η is equal to the solution of the vector equation $\vartheta_\eta = \phi\vartheta_\eta\phi^\intercal + \omega_\eta$, and $\vartheta_{\epsilon,1|0}$ is a diagonal matrix with $\omega_{\epsilon,i}/(1-\psi_i^2)$ on the main diagonal for $i = 1, \ldots, N$.

For compact notation, we stack all time periods for the variables x_t into one $(NT \times 1)-$dimensional vector according to $x = (x_1^\intercal, x_2^\intercal, \ldots, x_T^\intercal)^\intercal$. In the same way, define the $(Tr \times 1)-$dimensional stacked vector of factors $\eta = (\eta_1^\intercal, \eta_2^\intercal, \ldots, \eta_T^\intercal)^\intercal$ and the idiosyncratic components, $\epsilon = (\epsilon_1^\intercal, \epsilon_2^\intercal, \ldots, \epsilon_T^\intercal)^\intercal$. We collect all the model parameters in the vector θ.

3 Precision-based sampling with complete data

For Bayesian estimation of the factor model, we first assume the data are complete and later generalize to the case when some observations are missing. Complete or balanced data means that we have one observation $x_{i,t}^o$ available for each variable explained in the model $x_{i,t} = x_{i,t}^o$ for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$, or $x = x^o$ in brief.

Our aim is to sample from the posterior distribution

$$p(\eta, \theta|x) \propto L(x|\eta, \theta)p(\eta|\theta)p(\theta), \tag{3}$$

where the likelihood function $L(x|\eta, \theta)$ is implied by (1) and (2), and the priors are chosen closely in line with the existing factor model literature, see Appendix A for details. To obtain draws from the posterior distribution, we sample sequentially from the following conditional posterior distributions:

1. $p(\eta|x, \theta)$
2. $p(\theta|x, \eta)$

The precision-based sampler to draw from $p(\eta|x, \theta)$ is discussed in detail below, whereas details on the samplers for $p(\theta|x, \eta)$ and further model specifications are provided in Appendix A.

To prepare the application of conditional sampling from partitioned multivariate normals, we follow McCausland (2015) and define the joint vector

$$z = \begin{pmatrix} \eta \\ x \end{pmatrix} = \begin{pmatrix} I_{Tr} & 0_{Tr\times TN} \\ \Lambda & I_{TN} \end{pmatrix} \begin{pmatrix} \eta \\ \epsilon \end{pmatrix} \tag{4}$$

as a function of unobserved factors and idiosyncratic components. The matrix Λ is given by the Kronecker product $\Lambda = I_T \otimes \lambda$. According to (4), the joint vector $z = (\eta^\intercal, x^\intercal)^\intercal$ is an affine transformation of $(\eta^\intercal, \epsilon^\intercal)^\intercal$, which are Gaussian and mutually independent.

Hence, z also follows a multivariate normal distribution by

$$z|\theta \sim \mathcal{N}\left(0_{T(r+N)\times 1}, Q^{-1}\right), \tag{5}$$

where Q^{-1} is the $(T(r+N) \times T(r+N))-$dimensional covariance matrix and Q the corresponding precision matrix, which is conditional on model parameters. To facilitate efficient sampling, we have to find a tractable blocked expression for the precision matrix Q, which will allow us to apply general rules for sampling from partitioned Gaussian vectors.

We start by deriving the covariance matrix of z and write the factor VAR as $\Phi\eta = u_\eta$, where $u_\eta|\theta \sim \mathcal{N}(0_{Tr\times 1}, \Omega_\eta)$,

$$\Phi = \begin{pmatrix} I_r & & & \\ -\phi & I_r & & \\ & -\phi & \ddots & \\ & & \ddots & I_r \\ & & & -\phi \quad I_r \end{pmatrix}, \quad \text{and} \quad \Omega_\eta = \begin{pmatrix} \vartheta_{\eta,1|0} & & & \\ & \omega_\eta & & \\ & & \ddots & \\ & & & \omega_\eta \end{pmatrix}. \tag{6}$$

In stacked form, the vector of factors follows the distribution $\eta|\theta \sim \mathcal{N}(0_{Tr\times 1}, \Phi^{-1}\Omega_\eta\Phi^{-\mathsf{T}})$, where Φ has full rank and, hence, is invertible. Similarly, the stacked idiosyncratic components are defined as $\Psi\epsilon = u_\epsilon$ where $u_\epsilon|\theta \sim \mathcal{N}(0_{TN\times 1}, \Omega_\epsilon)$, and Ω_ϵ is a matrix containing the matrix $\vartheta_{\epsilon,1|0}$ on the first main diagonal block and the matrices ω_ϵ on the final $t = 2, \ldots, T$ main diagonal blocks. Ψ is constructed in a similar way as Φ above, but the main diagonal blocks consist of I_N matrices, and all the subdiagonal blocks are equal to $-\psi$. It follows that the vector of idiosyncratic components is distributed as $\epsilon|\theta \sim \mathcal{N}(0_{TN\times 1}, \Psi^{-1}\Omega_\epsilon\Psi^{-\mathsf{T}})$.

We obtain the covariance matrix Q^{-1} of the joint vector $z = (\eta^\mathsf{T}, x^\mathsf{T})^\mathsf{T}$ in (5) by

$$Q^{-1} = \begin{pmatrix} I & 0 \\ \Lambda & I \end{pmatrix} \begin{pmatrix} \Phi^{-1}\Omega_\eta\Phi^{-\mathsf{T}} & 0 \\ 0 & \Psi^{-1}\Omega_\epsilon\Psi^{-\mathsf{T}} \end{pmatrix} \begin{pmatrix} I & \Lambda^\mathsf{T} \\ 0 & I \end{pmatrix}, \tag{7}$$

and we can directly derive the partitioned precision matrix

$$Q = \begin{pmatrix} \Phi^\mathsf{T}\Omega_\eta^{-1}\Phi + \Lambda^\mathsf{T}(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi)\Lambda & -\Lambda^\mathsf{T}(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi) \\ -(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi)\Lambda & \Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi \end{pmatrix} = \begin{pmatrix} Q_{\eta\eta} & Q_{\eta x} \\ Q_{x\eta} & Q_{xx} \end{pmatrix}. \tag{8}$$

Given complete data, we can make use of the general rules for conditional sampling from a partitioned multivariate normal distribution as in Anderson (2003), Theorem 2.5.1. In terms of the partitioned precision matrix, Rue (2001), Section 3.1.1, provides the conditional distribution of η given $x = x^o$ defined as

$$p(\eta|x = x^o, \theta) \overset{\mathcal{D}}{=} \mathcal{N}\left(-Q_{\eta\eta}^{-1}Q_{\eta x}x^o, Q_{\eta\eta}^{-1}\right). \tag{9}$$

To efficiently draw a sample η^* from this distribution, Rue (2001) and Chan and Jeliazkov (2009) propose the application of fast band-matrix techniques. In the factor model, $Q_{\eta\eta} = \Phi^\mathsf{T}\Omega_\eta^{-1}\Phi + \Lambda^\mathsf{T}(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi)\Lambda$ is a block-banded matrix. Sampling proceeds as follows: Compute first the sparse Cholesky decomposition $Q_{\eta\eta} = LL^\mathsf{T}$, which implies a banded Cholesky factor L. Then, following Rue (2001), solve $Lw = -Q_{\eta x}x^o$ for w with

a matrix equation solver. Afterwards, solve $L^\mathsf{T}\mu = w$ for μ. Solve $L^\mathsf{T}v = v^*$ for v, where v^* is drawn from the standard normal distribution $\mathcal{N}(0_{Tr\times 1}, I_{Tr})$. Finally, a draw of the factors is provided by $\eta^* = \mu + v$.

To compute the Cholesky decomposition and solve for the factors as outlined above, matrix programming languages such as Matlab can exploit the sparse structure of $Q_{\eta\eta}$ efficiently, see Chan and Jeliazkov (2009) and McCausland, Miller, and Pelletier (2011) for details.

4 Precision-based sampling with missing observations

Now consider the case when we do not observe all values in $x_{i,t}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$. We assume a fraction κ - chosen such that κTN is an integer - of the observations is missing, and the missing observations can be distributed randomly across the indexes (i, t) as in Angelini et al. (2006) and Marcellino (2007). We do not model the process which generates the missing observations explicitly, and rather take the patterns of missing observations as given in the data. Following Rubin (1976), we thereby assume that the missing-data mechanism is ignorable, which is common when using macroeconomic data.

We define those model variables with missing observations as x_m, whereas variables with available observations are denoted as x_o. In the presence of missing observations, we modify the posterior sampler from Section 3 along the lines of Little and Rubin (2002). The general Gibbs sampler by Little and Rubin (2002) starts by sampling values for missing observations from their conditional posterior distribution. In subsequent steps, these samples are combined with observed data, and enter the conditional posterior distributions for sampling the remaining model parameters.

In our case of the factor model (1) and (2), we want to sample from the posterior distribution $p(\eta, x_m, \theta|x_o)$. Following Little and Rubin (2002), we do so by sequentially sampling factors, missing observations, and model parameters from their conditional posterior distributions:

1. $p(\eta|x_o, x_m, \theta)$
2. $p(x_m|x_o, \eta, \theta)$
3. $p(\theta|x_o, x_m, \eta)$

Alternatively, we provide a joint sampler for factors and missing observations conditional on parameters and observed data in a single step. In this case, the conditional distribution $p(\eta, x_m|x_o, \theta)$ replaces the first two in the sampler above.

Our general sampling strategy works as follows: Given complete data in Section 3, we have used the variable ordering $z = (\eta^\mathsf{T}, x^\mathsf{T})^\mathsf{T}$, where variables x were ordered last. In the presence of missing observations, the main idea is to permute the variables in z and obtain reordered or permuted z_{P_z} such that we can apply the same techniques for conditional sampling as in Section 3. In particular, we move those variables with available observations x_o to the bottom of z_{P_z} and apply the same rules for conditional sampling from a Gaussian as in (8) and (9).

The reordering of variables in the vector z can be implemented by using a properly defined permutation matrix P_z such that $z_{P_z} = P_z z$. In general, a permutation matrix P_z is defined as a square binary matrix that has exactly one entry of 1 in each row and each column and zeros elsewhere. Permutation matrices are orthogonal matrices such that $P_z^{-1} = P_z^\mathsf{T}$ and $P_z P_z^\mathsf{T} = P_z^\mathsf{T} P_z = I$. The permutation by P_z implies a linear transformation of z in (5), and the distribution of the transformed Gaussian z_{P_z} becomes

$$z_{P_z}|\theta \sim \mathcal{N}\left(0_{T(r+N)\times 1}, P_z Q^{-1} P_z^\mathsf{T}\right), \tag{10}$$

following standard rules for linear transformations of Gaussian vectors as in Anderson (2003), Theorem 2.4.1. Note that $P_z Q^{-1} P_z^\mathsf{T}$ is equal to the row- and column-permuted covariance matrix of z since

$$P_z Q^{-1} P_z^\mathsf{T} = ((P_z^\mathsf{T})^{-1} Q P_z^{-1})^{-1} = (P_z Q P_z^\mathsf{T})^{-1} = Q_z^{-1}. \tag{11}$$

Thus, the permuted covariance is equal to the inverse of the permuted precision matrix. A variable ordering and permutation in z is associated with a column- and row-permutation of the elements in the precision matrix. After permutation, we can partition the permuted precision matrix and apply a similar conditional sampling from a Gaussian as in (9).[1]

All the permutations we consider order the variables with observed data x_o last in z_{P_z}. The reordering of variables in x into variables with missing observations x_m above the variables with available observations x_o can be implemented by using the permutation matrix P_x defined as

$$x_{P_x} = \begin{pmatrix} x_m \\ x_o \end{pmatrix} = P_x x = \begin{pmatrix} P_{x_m} \\ P_{x_o} \end{pmatrix} x. \tag{12}$$

The matrix P_{x_o} has TN columns corresponding to the TN elements in $x = (x_1^\mathsf{T}, x_2^\mathsf{T}, \ldots, x_T^\mathsf{T})^\mathsf{T}$, and the number of rows is equal to $(1-\kappa)TN$, the number of observations available for estimation. If observations were available for all variable values, P_x would equal the identity matrix. To construct P_x in the presence of missing observations, we can set P_{x_o} equal to the identity matrix and remove all those rows for which the corresponding observations are missing in the empirical data set. The matrix P_{x_m} just consists of these removed rows. Note that the position of missing observations in the data set is the only necessary information to derive the permutation matrix P_x. In particular, the permutation does not depend on model parameters.

Apart from the position of x_o, different sampling schemes can be derived depending on how the factors and variables with missing values are ordered. In this paper, we discuss two alternative samplers in the subsequent sections of the text:

1. Sequential 2-step sampler in Section 4.1 in the spirit of Little and Rubin (2002):

   (a) $p(\eta|x_o, x_m, \theta)$: Sampling factors conditional on a sample of missing values,

[1] Note that we show matrix permutations in the paper only for expositional purposes. The Matlab computer codes underlying the quantitative results in the paper are based on equivalent, but more efficient index permutations (Golub and Van Loan, 2013). Let P be a permutation matrix of dimension $(K \times K)$ and p be a permutation vector defined as $p = (1, 2, \ldots, K) \times P^\mathsf{T}$. For a $(K \times K)$-dimensional matrix S, indexing by `S(p,:)` in Matlab is equivalent to row permutation PS, and `S(:,p)` is equivalent to column permutation SP^T. To reverse the original permutation, we can use the inverse of the permutation matrix P^T or the corresponding index `r` defined as `r(p)=1:K;`.

observed data, and parameters. No permutation is needed in this step.

(b) $p(x_m|x_o, \eta, \theta)$: Conditionally sampling of missing observations given factors, data, and parameters using the permutation $z_{P_{2s}} = (x_m^\mathsf{T}, \eta^\mathsf{T}, x_o^\mathsf{T})^\mathsf{T}$.

2. Joint sampling from $p(\eta, x_m|x_o, \theta)$ using the period-wise time permutation $z_{P_\tau} = (\eta_1^\mathsf{T}, x_{m,1}^\mathsf{T}, \eta_2^\mathsf{T}, x_{m,2}^\mathsf{T}, \ldots, \eta_T^\mathsf{T}, x_{m,T}^\mathsf{T}, x_o^\mathsf{T})^\mathsf{T}$, see Section 4.2.

Of course, both samplers aim at drawing from the same conditional posterior distribution $p(\eta, x_m|x_o, \theta)$. Differences between the samplers can arise with respect to a) convergence and mixing of the Markov chain, and b) computational efficiency. Concerning a), sequential samplers are in many cases easy to implement, because conditional distributions can generally be easier derived than joint distributions. On the other hand, sequentially sampling in two blocks using conditional distributions might lead to more correlated samples and slower convergence compared to sampling from the joint distribution in one block. Concerning b), computational efficiency, the joint sampler relies on the sparse Cholesky decomposition of one huge precision matrix, whereas the sequential sampler is based on decompositions of two smaller precision matrices for factors and missing observations, respectively. In addition, the alternative permutations of variables can influence the speed of the sparse Cholesky decomposition (McCausland et al., 2011). Furthermore, set-up costs to fill the precision matrices in each step of the Gibbs sampler vary between the samplers. Finally, there are differences between the samplers with respect to model evaluation using the marginal likelihood.

We will provide details on the samplers in the next subsections. A discussion of their differences is provided in Section 5 by using simulation experiments. Details on how the marginal likelihood can be derived using the joint sampler are provided in Section 7.

4.1 Sequential sampling of factors and missing observations

This precision-based sampler iterates between sampling factors conditional on interpolated missing values from $p(\eta|x_o, x_m, \theta)$ and, thereafter, values for missing observations conditional on factors from $p(x_m|x_o, \eta, \theta)$:

1. $p(\eta|x_o, x_m, \theta)$: Assume we have a draw for missing values $x_m = x^{m*}$. We can stack the interpolated missing data and the observed data in $x^*_{P_x} = ((x^{m*})^\mathsf{T}, (x^o)^\mathsf{T})^\mathsf{T}$. By reversing the data permutation from (12) according to $P_x^{-1} = P_x^\mathsf{T}$, we can move the interpolated values to the positions of the missing observations in the original data set using $x^* = P_x^\mathsf{T} x^*_{P_x}$. Given the partly interpolated data, we can use the complete-data sampler from (9) to draw factors conditional on the data and parameters from

$$p(\eta|x_o = x^o, x_m = x^{m*}, \theta) = p(\eta|x = x^*, \theta) \overset{\mathcal{D}}{=} \mathcal{N}\left(-Q_{\eta\eta}^{-1} Q_{\eta x} x^*, Q_{\eta\eta}^{-1}\right). \tag{13}$$

   Note that the moments in (13) differ from those in the complete-data case (9) only with respect to x^* in the mean.

2. $p(x_m|x_o, \eta, \theta)$: We draw values for missing observations conditional on a factor sample η^* from step 1 and observed data. We use the permutation $z_{P_{2s}} = (x_m^\mathsf{T}, \eta^\mathsf{T}, x_o^\mathsf{T})^\mathsf{T}$, where the variables corresponding to missing observations in the data are ordered

first, whereas factors and observed data are ordered last. We permute by $z_{P_{2s}} = P_{2s} z$ with permutation matrix

$$P_{2s} = \begin{pmatrix} 0_{\kappa TN \times Tr} & P_{x_m} \\ I_{Tr} & 0_{Tr \times TN} \\ 0_{(1-\kappa)TN \times Tr} & P_{x_o} \end{pmatrix}, \tag{14}$$

where the permutation matrices P_{x_m} and P_{x_o} decompose the model variables as in (12). As (10) holds for any permutation matrix, the distribution of the transformed Gaussian $z_{P_{2s}}$ is

$$z_{P_{2s}} | \theta \sim \mathcal{N}\left(0_{T(r+N)\times 1}, P_{2s} Q^{-1} P_{2s}^{\mathsf{T}}\right), \tag{15}$$

and the permuted covariance is equal to the inverse of the permuted precision matrix by $P_{2s} Q^{-1} P_{2s}^{\mathsf{T}} = (P_{2s} Q P_{2s}^{\mathsf{T}})^{-1} = Q_{2s}^{-1}$. The precision matrix can be derived as the inverse of a block matrix product by

$$\begin{aligned} Q_{2s} &= P_{2s} \begin{pmatrix} I & -\Lambda^{\mathsf{T}} \\ 0 & I \end{pmatrix} \begin{pmatrix} Q_\eta & 0 \\ 0 & Q_\epsilon \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Lambda & I \end{pmatrix} P_{2s}^{\mathsf{T}} \\ &= \begin{pmatrix} P_{x_m} Q_\epsilon P_{x_m}^{\mathsf{T}} & -P_{x_m} Q_\epsilon \Lambda & P_{x_m} Q_\epsilon P_{x_o}^{\mathsf{T}} \\ -\Lambda^{\mathsf{T}} Q_\epsilon P_{x_m}^{\mathsf{T}} & Q_\eta + \Lambda^{\mathsf{T}} Q_\epsilon \Lambda & -\Lambda^{\mathsf{T}} Q_\epsilon P_{x_o}^{\mathsf{T}} \\ P_{x_o} Q_\epsilon P_{x_m}^{\mathsf{T}} & -P_{x_o} Q_\epsilon \Lambda & P_{x_o} Q_\epsilon P_{x_o}^{\mathsf{T}} \end{pmatrix}, \end{aligned} \tag{16}$$

where we have defined $Q_\epsilon = (\Psi^{-1} \Omega_\epsilon \Psi^{-\mathsf{T}})^{-1} = \Psi^{\mathsf{T}} \Omega_\epsilon^{-1} \Psi$ and $Q_\eta = (\Phi^{-1} \Omega_\eta \Phi^{-\mathsf{T}})^{-1} = \Phi^{\mathsf{T}} \Omega_\eta^{-1} \Phi$ to simplify notation from (7). The variable ordering in $z_{P_{2s}} = (x_m^{\mathsf{T}}, \eta^{\mathsf{T}}, x_o^{\mathsf{T}})^{\mathsf{T}}$ is useful for conditional sampling, as the first block contains the variables with missing observations, whereas the rest contains the conditioning information, namely, factors and observed data. We thus partition the precision matrix by

$$Q_{2s} = \begin{pmatrix} Q_{x_m, x_m} & Q_{x_m, \eta x_o} \\ Q_{\eta x_o, x_m} & Q_{\eta x_o, \eta x_o} \end{pmatrix}, \tag{17}$$

where the top-left $(\kappa TN \times \kappa TN)$-dimensional block is defined as $Q_{x_m, x_m} = P_{x_m} Q_\epsilon P_{x_m}^{\mathsf{T}}$. Similarly to (9), we can derive the conditional distribution of missing observations conditional on factors and observed data

$$p(x_m | x_o = x^o, \eta = \eta^*, \theta) \overset{\mathcal{D}}{=} \mathcal{N}\left(-Q_{x_m, x_m}^{-1} Q_{x_m, \eta x_o} ((\eta^*)^{\mathsf{T}}, (x^o)^{\mathsf{T}})^{\mathsf{T}}, Q_{x_m, x_m}^{-1}\right). \tag{18}$$

Note that in the precision matrix Q_{x_m, x_m}, the matrix $Q_\epsilon = \Psi^{\mathsf{T}} \Omega_\epsilon^{-1} \Psi$ is the precision matrix of the idiosyncratic components' prior distribution. Since the idiosyncratic components follow a VAR(1) process, the precision matrix is block-banded (Chan and Jeliazkov, 2009). Permutation using P_{2s} just selects those idiosyncratic components corresponding to missing observations in the data and thus leaves the precision matrix $Q_{x_m, x_m} = P_{x_m} Q_\epsilon P_{x_m}^{\mathsf{T}}$ block-banded.

In the subsequent parts of the text, we call this method 'Sequential 2-step sampling'.

4.2 Joint sampling of factors and missing observations with time permutation

To efficiently sample from the joint distribution, we permute z such that factors and missing values are ordered together for each time period t, $(\eta_t^\mathsf{T}, x_{m,t}^\mathsf{T})^\mathsf{T}$. These vectors are stacked for $t = 1, \ldots, T$ and placed on top of the variables with observed data x_o. We obtain the permuted vector of variables $z_{P_\tau} = (\eta_1^\mathsf{T}, x_{m,1}^\mathsf{T}, \eta_2^\mathsf{T}, x_{m,2}^\mathsf{T}, \ldots, \eta_T^\mathsf{T}, x_{m,T}^\mathsf{T}, x_o^\mathsf{T})^\mathsf{T}$. Note that the variables in $z = (\eta^\mathsf{T}, x^\mathsf{T})^\mathsf{T}$ are already ordered period-wise within the blocks for factors $\eta = (\eta_1^\mathsf{T}, \eta_2^\mathsf{T}, \ldots, \eta_T^\mathsf{T})^\mathsf{T}$ and variables $x = (x_1^\mathsf{T}, x_2^\mathsf{T}, \ldots, x_T^\mathsf{T})^\mathsf{T}$ in (4). To reorder the variables, we permute by $z_{P_\tau} = P_\tau z$ with

$$P_\tau = \begin{pmatrix} P_{\eta,1} & 0_{r \times NT} \\ 0_{N_{m,1} \times rT} & P_{x_m,1} \\ P_{\eta,2} & 0_{r \times NT} \\ 0_{N_{m,2} \times rT} & P_{x_m,2} \\ \vdots & \vdots \\ P_{\eta,T} & 0_{r \times NT} \\ 0_{N_{m,T} \times rT} & P_{x_m,T} \\ 0_{(N - \sum_t N_{m,t}) \times rT} & P_{x_o} \end{pmatrix}, \tag{19}$$

where $N_{m,t}$ denotes the number of missing values in x at time t.

The matrices $P_{\eta,t}$ for $t = 1, \ldots, T$ are equal to the rows from the identity matrix $I_{rT} = I_T \otimes I_r$ matrix corresponding to period t such that

$$P_{\eta,t} = \begin{pmatrix} 0_{r \times r(t-1)} & I_r & 0_{r \times r(T-t)} \end{pmatrix}. \tag{20}$$

Note that the matrix $P_{\eta,t}$ can be considered as a block row vector having T blocks, each consisting of r columns. The period-t block is just the $(r \times r)$ identity matrix, because all r factor values are ordered first in (19) in each period.

The matrices $P_{x_m,t}$ for $t = 1, \ldots, T$ contain the block rows of the matrix P_{x_m} as defined in (12) that correspond to missing observations in x_t according to

$$P_{x_m} = \begin{pmatrix} P_{x_m,1} \\ P_{x_m,2} \\ \vdots \\ P_{x_m,T} \end{pmatrix}. \tag{21}$$

The number of rows of $P_{x_m,t}$, $N_{m,t}$, is equal to the number of missing values in x_t at time t, whereas the number of columns is equal to TN. Note that $P_{x_m,t}$ can also be considered as a block row vector having T column blocks, each consisting of N columns, according to

$$P_{x_m,t} = \begin{pmatrix} 0_{N_{m,t} \times (t-1)N} & P_{x_m,(t,t)} & 0_{N_{m,t} \times (T-t)N} \end{pmatrix}, \tag{22}$$

where the $(N_{m,t} \times N)$ matrix $P_{x_m,(t,t)}$ selects the missing observations in period t, as implicitly defined in (12), taking into account the stack of variables $x = (x_1^\mathsf{T}, x_2^\mathsf{T}, \ldots, x_T^\mathsf{T})^\mathsf{T}$ in (4).

Note that all vectors and matrices above follow the convention that a block matrix

with zero row or column size implies the matrix is empty, and the corresponding blocks in the matrices above represent empty placeholders. This applies to either time periods without any missing observations or time periods without data observations.

Given the permutation matrix P_τ, we can obtain the permuted vector of variables $z_{P_\tau} = P_\tau z$ and the permuted precision matrix $Q_\tau = P_\tau Q P_\tau^\intercal$. In Appendix B, we derive the precision matrix Q_τ by exploiting the period-wise block structure implied by z_{P_τ}.

Our aim is to jointly sample factors and values for missing observations from the conditional distribution $p(\eta, x_m | x_o = x^o, \theta)$. The final block in the vector z_{P_τ} is x_o and the corresponding observations serve as the conditioning set. Thus, the permuted precision matrix Q_τ can be partitioned in the following way

$$Q_\tau = \begin{pmatrix} Q_{\eta x_m, \eta x_m} & Q_{\eta x_m, x_o} \\ Q_{x_o, \eta x_m} & Q_{x_o, x_o} \end{pmatrix}, \tag{23}$$

where the upper-left block $Q_{\eta x_m, \eta x_m}$ has dimensions $(rT + \sum_{t=1}^{T} N_{m,t}) \times (rT + \sum_{t=1}^{T} N_{m,t})$, and we can directly derive the conditional distribution

$$p(\eta, x_m | x_o = x^o, \theta) \overset{\mathcal{D}}{=} \mathcal{N}(-Q^{-1}_{\eta x_m, \eta x_m} Q_{\eta x_m, x_o} x^o, Q^{-1}_{\eta x_m, \eta x_m}) \tag{24}$$

for factors and missing values conditional on observed data. In Appendix B, we show that $Q_{\eta x_m, \eta x_m}$ is a block-banded matrix. In particular, it has block bandwidth equal to one, thus, representing a block tridiagonal matrix (Golub and Van Loan, 2013). The reason is that the reduced form of the factor model can be written as a VAR process of order one in the factors and explained variables. Following Rue and Held (2005), precision matrices of AR processes have a bandwidth equal to the lag order of the AR process. In the appendix, we show that this result also holds in the factor model (1) and (2) with missing observations, where the precision matrix is block tridiagonal.

In the subsequent parts of the text, we call this method 'Joint sampling, time permutation'.

5 Comparing the precision-based samplers by simulations

Based on simulations, we compare the convergence properties, the computing time of the precision-based samplers as well as their equivalence in terms of mean-squared errors of simulated factors and missing observations.

5.1 Data-generating process and model estimation

The data-generating process (DGP) has a factor structure with $r = 2$ factors, which follow a VAR process with one lag. The idiosyncratic components each follow AR processes with

one lag. The variables $x_{i,t}$ for $i = 1, \ldots, N$ are simulated according to

$$x_{i,t} = \lambda_{i.}\eta_t + \epsilon_{i,t}, \tag{25}$$

$$\eta_t = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.8 \end{bmatrix} \eta_{t-1} + u_{\eta,t}, \;\; u_{\eta,t} \sim \mathcal{N}(0, I_2), \tag{26}$$

$$\epsilon_{i,t} = 0.4\epsilon_{i,t-1} + u_{\epsilon,i,t}, \;\; u_{\epsilon,i,t} \sim \mathcal{N}(0, \omega_\epsilon). \tag{27}$$

For the loading matrix λ, we specify a point-mass normal mixture distribution often used for variable selection as in George and McCulloch (1993, 1997), and Geweke (1996):

$$p(\lambda_{ij}) \overset{\mathcal{D}}{=} (1 - \rho_j)\delta_0(\lambda_{ij}) + \rho_j \mathcal{N}(m_j, M), \tag{28}$$

$$p(\rho_j) \overset{\mathcal{D}}{=} \mathcal{B}(r_0 s_0, r_0(1 - s_0)), \tag{29}$$

where $\delta_0(\cdot)$ represents the Dirac function with point mass at zero and ρ_j is a factor-specific probability of a non-zero loading. The inclusion probabilities ρ_j follow a beta distribution, $\mathcal{B}(r_0 s_0, r_0(1-s_0))$, with mean $s_0 = 0.5$ and precision $r_0 = 30$. The non-zero factor loadings are simulated out of the normal distributions $\mathcal{N}(m_j, M)$ with $m_1 = 0.60$, $m_2 = 0.40$, and $M = 0.01$. The variance $M = 0.01$ is relatively tight in order to clearly separate zero and non-zero loadings. For the variances of the idiosyncratic components, we assume an inverse gamma distribution according to $\omega_\epsilon \sim \mathcal{IG}(2, 0.5)$.

Concerning sample size, we consider $T = 100$ time-series observations and $N = 100$ variables. Given this specification, we sample factors η_{DGP} and data x_{DGP} from (25) and (27). To address missing observations, we assume that 20% of observations are missing, $\kappa = 0.20$. We randomly set κNT observations in the sample x_{DGP} to missing values, yielding the data set x^o used for Bayesian estimation of factors and missing observations. In the experiment, we sample $K = 100$ times from the DGP and estimate the factor model on each data set. We ran further experiments with different specifications for (T, N, κ) and alternative priors. As the results in the alternative experiments are very similar compared to the baseline case summarized below, we only report the baseline results.

In the simulation experiments, the factor model is estimated using the posterior sampler outlined in Appendix A for each data $x^{o,(k)}$ set sampled from the DGP for $k = 1, \ldots, K$. Each time, we draw $G = 10000$ times from the posterior. We obtain samples for factors, missing values, and parameters according to $\eta^{(k,g)}, x^{m,(k,g)}, \theta^{(k,g)} \sim p(\eta, x_m, \theta | x_o = x^{o,(k)})$ for $g = 1, \ldots, G$ and $k = 1, \ldots, K$. To address convergence and computing time depending on the number of draws, we consider different partitions of the raw posterior draws. In particular, we compare alternative numbers of burn-in draws $G^{\text{burn-in}}$ and numbers of posterior evaluation draws G^{eff}.

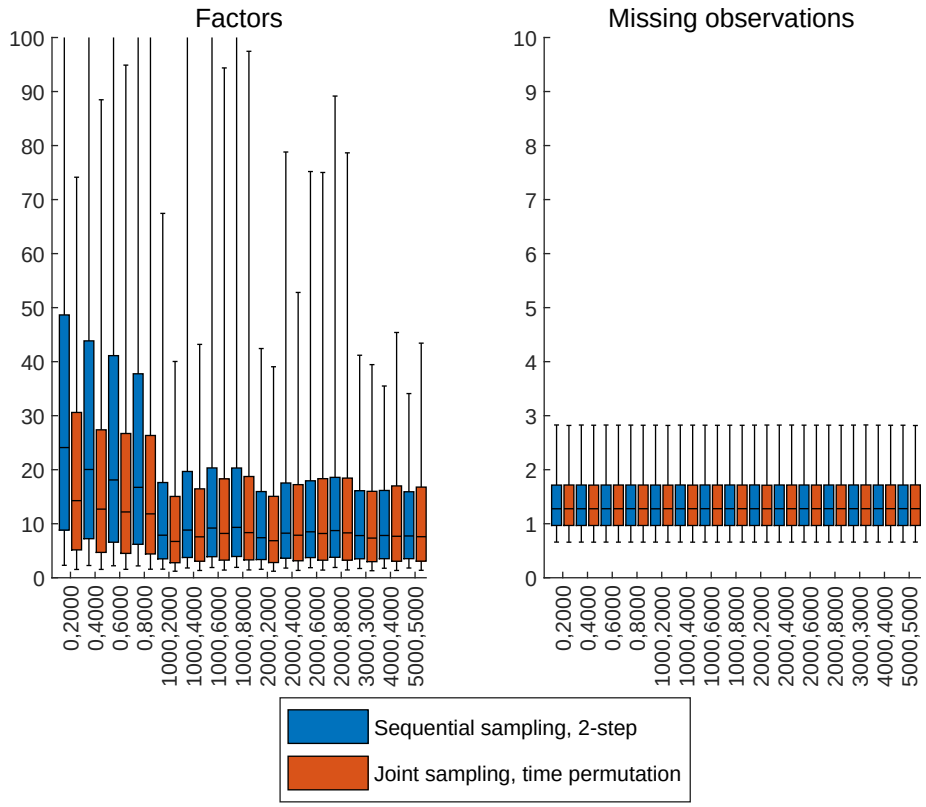
5.2 Comparing inefficiency factors

We compute inefficiency factors to discuss convergence and mixing of the Gibbs Sampler. The inefficiency factor can be defined as in Chib (2001) by

$$\text{IE}_{\eta,j,t,k} = 1 + 2\sum_{m=1}^{M}(1 - \frac{m}{M})\rho_{\eta,j,t,k}(m), \tag{30}$$

where $\rho_{\eta,j,t,k}(m)$ is the estimated autocorrelation at lag m of the posterior draws for model factors $\eta_{j,t}^{(k,g)}|x_o = x^{o,(k)}$ over the draws $g = G^{\text{burn-in}}+1,\ldots,G^{\text{burn-in}}+G^{\text{eff}}$. The maximum lag order is $M = 150$. Values of $\text{IE}_{\eta,j,t,k}$ greater than one indicate autocorrelation in the chain that might be due to poor mixing or lack of convergence. We compute inefficiency factors for the posterior samples of the model factors for $j = 1, 2$, time periods $t = 1,\ldots,T$, and for $k = 1,\ldots,K$ samples from the DGP. In the left panel of Figure 1, we show box plots for the whole distribution of inefficiency factors for model factors $\text{IE}_{\eta,j,t,k}$ across j, t, and k and for different numbers of burn-in and evaluation draws. In the right panel of Figure 1, we show box plots for the inefficiency factors of posterior draws of values for missing observations.

Figure 1: Inefficiency factors for posterior samples of model factors and missing observations.



Note: In the left figure, box plots of inefficiency factors are based on the posterior samples of model factors for different numbers of burn-in and posterior evaluation draws. The first number shown in the labels of the horizontal axis refers to the number of burn-in draws $G^{\text{burn-in}}$, whereas the second number refers to the number of draws used for posterior evaluation after burn-in G^{eff}. In the right figure, box plots show inefficiency factors for the estimated values of missing observations.

The results for model factors in the left panel of Figure 1 show that a number of burn-in draws greater or equal than $G^{\text{burn-in}} = 1000$ generally leads to a median of inefficiency factors around 10 for all three precision-based samplers. The upper bound of the interquartile ranges are in most cases below 20 for $G^{\text{burn-in}} \geq 1000$. Across different

posterior sample splits, there are differences with respect to the whiskers of the boxplots, which mark the 5^{th} and 95^{th} percentiles. However, with increasing numbers of burn-in and evaluation draws, the bands tend to become smaller, and we find no systematic differences between the two precision-based samplers. For sample splits with zero or 1000 burn-in draws, $G^{\text{burn-in}}$, the upper bounds of the interquartile ranges for the 2-step precision-based sampler are higher than for the joint time-permutation sampler. For $G^{\text{burn-in}} > 1000$, we see no substantial differences in terms of convergence in the model factors between the two precision-based samplers.

The results for missing observations in the right panel of Figure 1 show that convergence is very fast for all precision-based samplers. The median of the inefficiency factors is slightly greater than one, the upper bound of the interquartile range is about 1.3, and the bound of the upper whisker is slightly below 3. Note that the convergence for missing observations is substantially faster than for model factors. Convergence issues in factor models are often due to the lack of identification between factor loadings and factors, because $\lambda\eta_t = (\lambda H)(H^{-1}\eta_t)$ holds for any invertible H as documented in Lopes and West (2004); Ghosh and Dunson (2009); Bai and Wang (2014); Conti, Frühwirth-Schnatter, Heckman, and Piatek (2014); Kastner, Frühwirth-Schnatter, and Lopes (2017); Chan, Leon-Gonzalez, and Strachan (2018), amongst others. As discussed in parts of this literature, blocked sequential sampling of model factors conditional on loadings and subsequently loadings conditional on factors can sometimes lead to correlated draws and poor convergence, whereas joint sampling of model factors and loadings generally improves convergence, albeit making more complicated samplers necessary (Ghosh and Dunson, 2009; Chan and Jeliazkov, 2009; Conti et al., 2014; Kastner et al., 2017). This paper has a conceptually different focus on sampling missing values in the data and model factors given factor loadings. The key point is that the missing values in the data are a function of the common components, not of either the factors or the loadings alone. In Figure 2, we show the inefficiency factors for factor loadings and the common components to address this issue. The inefficiency factors of the loadings are comparable in magnitude to the inefficiency factors of the model factors, whereas the inefficiency factors of the common components are close to one, and thus comparable to the inefficiency factors of the estimated missing values.

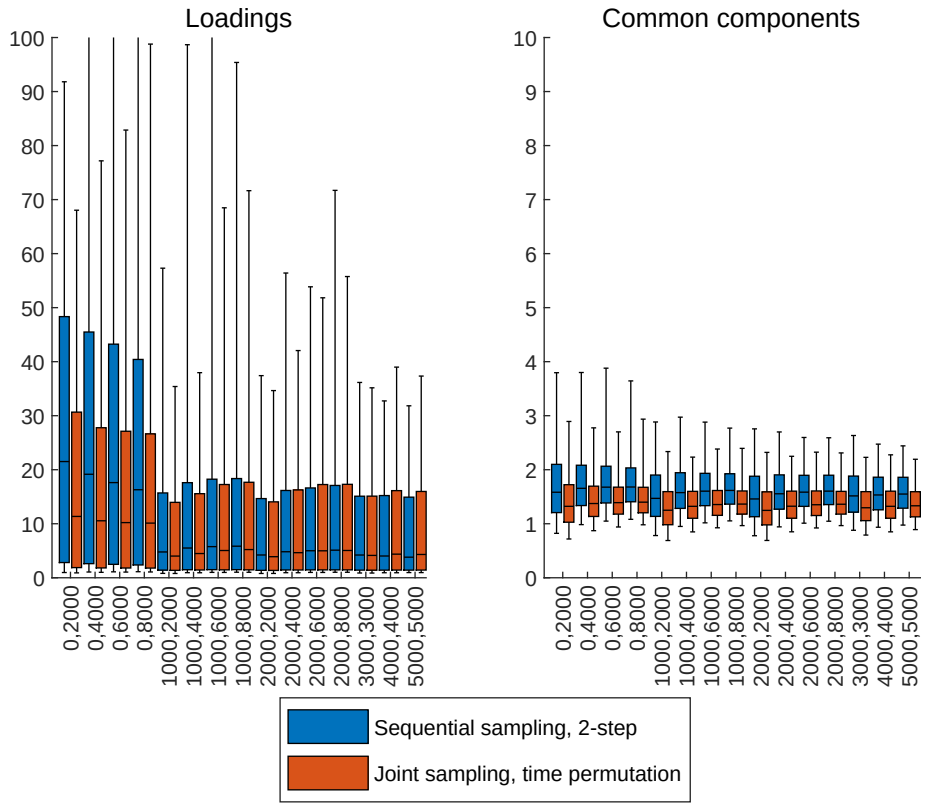
Overall, the simulation results indicate that convergence of estimates of missing values and common components do not to seem to be affected by any loadings-factor identification issue as mentioned above. Despite the fact that loadings and model factors show slower convergence, we see no major convergence issues for a reasonably chosen number of burn-in draws in general. The two precision-based samplers perform quite similar.

5.3 Comparing computing time

In Figure 3, the average computing time needed for posterior sampling is shown for different numbers of burn-in draws $G^{\text{burn-in}}$ and evaluation draws G^{eff} for the precision-based samplers. A box plot in the figure refers to the distribution of computing time across $k = 1, \ldots, K$ data sets sampled from the DGP.

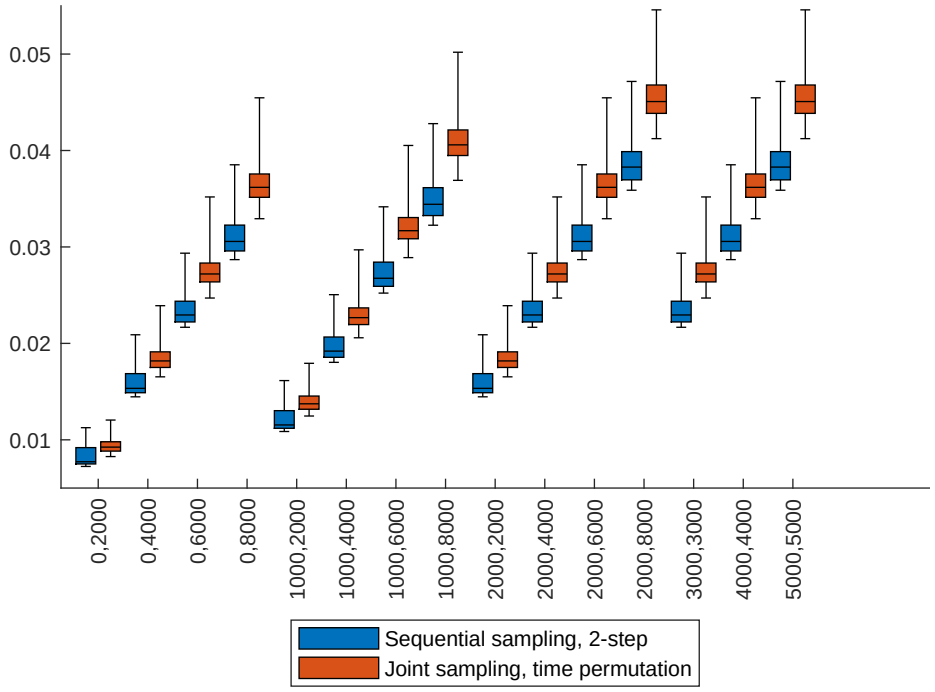
The results in Figure 3 show an overall better performance of the 2-step precision-based sampler than the time permutation sampler. The median and the bounds of the interquartile ranges of the 2-step sampler are in the majority of cases smaller than those

Figure 2: Inefficiency factors for posterior samples of model factor loadings and common components.



Note: In the left figure, box plots of inefficiency factors are based on the posterior samples of model factor loadings for different numbers of burn-in and posterior evaluation draws. The first number shown in the labels of the horizontal axis refers to the number of burn-in draws $G^{\text{burn-in}}$, whereas the second number refers to the number of draws used for posterior evaluation after burn-in G^{eff}. In the right figure, box plots show inefficiency factors for the common components.

Figure 3: Average computing time for posterior samples for different numbers of posterior draws.



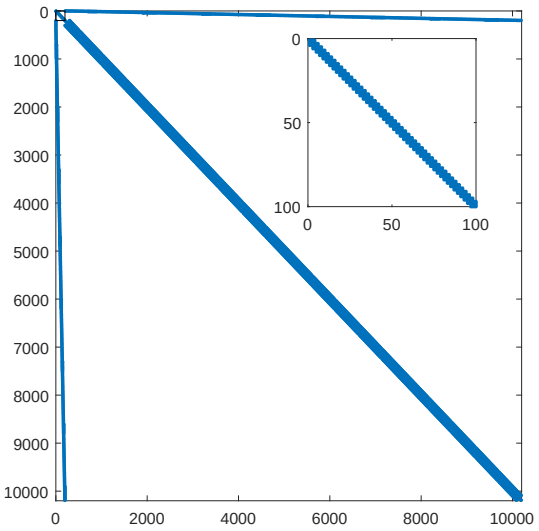
Note: Figure shows hours of computing time averaged over $K = 100$ estimation experiments. For each dataset from the DGP, we measure the elapsed time to draw from the posterior given alternative numbers of posterior draws. The first number shown in the labels of the horizontal axis refers to the number of burn-in draws $G^{\text{burn-in}}$, whereas the second number refers to the number of draws used for posterior evaluation after burn-in G^{eff}.

of the time permutation sampler. Despite some overlap of the 90% intervals, which are marked by the whiskers of the boxplots, the results overall indicate some computational advantages of the 2-step precision-based sampler.

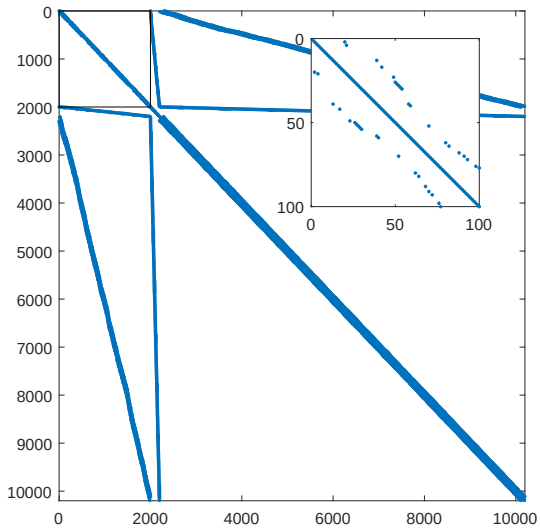
To get an understanding of the better performance of the 2-step approach, we have a closer look at the flops needed for precision-based sampling as in Rue (2001); McCausland et al. (2011). In general, the Cholesky factorization $Q = LL^\intercal$ of a banded, symmetric, and positive definite matrix Q with dimensions $(K \times K)$ and bandwidth b implies Kb^2 flops, where the bandwidth is the maximum number of off-diagonals, which have non-zero elements. For the forward and backward substitutions outlined at the end of Section 3, we need $4Kb$ flops, and we need $Kb^2 + 4Kb$ flops in total. The 2-step sequential approach of Section 4.1 requires two samples based on different precision matrices, whereas time-t permutation sampling requires one decomposition only. Both the matrix size and the bandwidth have a positive effect on the computing time according to the formulae for the overall flops. However, as the bandwidth also enters the formulae $Kb^2 + 4Kb$ squared, it has a comparatively huge impact on the overall computing time. The farther away non-zero elements are from the main diagonal, the more computing time is need for decomposing and solving. In Figure 4, we show the precision matrices for 2-step and time-t sampling. Each entry in the figures receives a blue sign, if the corresponding entry of the precision matrix is non-zero. Panel A and B show the two precision matrices for the sequential 2-step sampler. Panel C shows the precision matrix for the time-permutation sampler. The top-left blocks highlighted by a black rectangle refer to the submatrix, which will be decomposed by sparse Cholesky factorization. The dimensions of theses submatrices are $K_{2s,\eta} = 200$ and $K_{2s,x_m} = 2000$ for the 2-step sampler, and $K_\tau = 2200$ for the time-permutation sampler. Note that the position of the missing $\kappa NT = 2000$ observations is chosen randomly in the data and thereby affects the shape of the precision matrices and bandwidths. To highlight the bandwidth near the main diagonal, a subplot in the top-right of Figure 4 zooms the first top-left (100×100)-dimensional block of elements of the precision matrix. In the figure, the bandwidth of the precision matrix underlying the time-permutation sampler is larger than the bandwidths of the precision matrices used for 2-step sampling. In detail, the bandwidths are $b_{2s,\eta} = 3$ and $b_{2s,x_m} = 31$ for the 2-step sampler, and $b_\tau = 59$ for the time-permutation sampler. Thus, the comparatively large bandwidth of the time-permutation sampler contributes to the slower computational performance. Note, however, that the bandwidth of the precision matrix is only one source to explain the differences in computing time. There are also set-up costs to fill the precision matrix with model parameters every recursion of the sampler, which contribute to the computing time in the simulations. In the same way, permuting the moments of z at each draw is also costly. Note, however, that the permutation matrices depend only on the position of missing observations in the dataset, not the model parameters, and thus only have to be computed once for Bayesian estimation. The simulation results reflect all these different determinants of the overall computing time.

Figure 4: Precision matrices after permutation

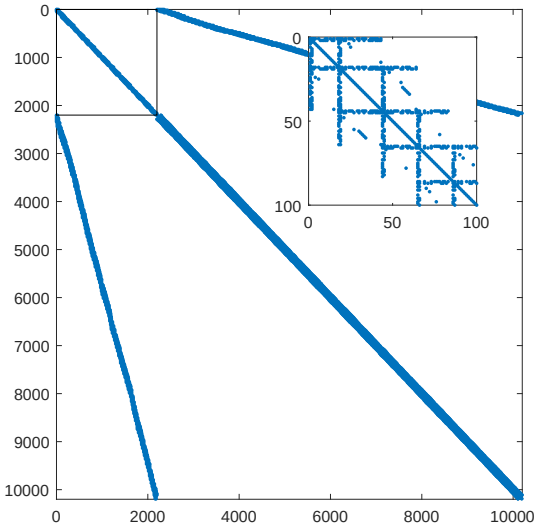
A. Sequential 2-step: Q to sample η



B. Sequential 2-step: $P_{2s}QP_{2s}^{\mathsf{T}}$ to sample x_m



C. Time permutation, $P_{\tau}QP_{\tau}^{\mathsf{T}}$



Note: The figure shows the non-zero elements in the precision matrices used in the alternative samplers. The top-left blocks highlighted by a black rectangle refer to the submatrices, which will be decomposed by Cholesky factorization. In Panel A, the highlighted part of the precision matrix is the block $Q_{\eta\eta}$ in (13). In Panel B, the highlighted part of the precision matrix is Q_{x_m,x_m} in (18). In Panel C, the highlighted part shows the block $Q_{\eta x_m,\eta x_m}$ under the joint time-permutation sampler in (24). To show the structure and bandwidth of these submatrices near the main diagonal in more detail, a subplot in the top-right of the figure zooms the first top-left (100×100)-dimensional block of elements of the precision matrix.

5.4 Comparing samples of factors and missing observations

As a final check of the accuracy of the precision-based samplers, we compare the mean-squared error $\text{MSE}_{\eta,t,j}$ for the factor posterior samples defined as

$$\text{MSE}_{\eta,j,t} = \frac{1}{G^{\text{eff}}} \sum_{g=G^{\text{burn-in}}+1}^{G^{\text{burn-in}}+G^{\text{eff}}} (\eta_{\text{DGP},j,t} - \eta_{j,t}^{(g)} | x_o = x^o)^2, \tag{31}$$

for each factor indexed by j for $j = 1, 2$ and time period $t = 1, \ldots, T$. In the MSE, the posterior samples of the factor $\eta_{j,t}^{(g)} | x_o = x^o$ are subtracted from the "true" factor values sampled from the DGP $\eta_{\text{DGP},j,t}$. As the samplers should provide draws from the same posterior distribution, we expect very similar values for large G^{eff}. Accordingly, we also compute the MSE for the missing observations by subtracting the sampled values drawn by the conditional samplers from the true values simulated from the DGP.

The results presented below are based on one data set and factors sampled from the DGP. We take $G = 10000$ raw draws from the posterior and discard the first $G^{\text{burn-in}} = 5000$ as burn-in. The MSEs are computed over the remaining $G^{\text{eff}} = 5000$ draws. The left panel of Figure 5 provides the MSEs for the first of the two factors obtained from the different samplers for each time period. As expected, the samplers yield very similar MSEs for all $t = 1, \ldots, T$. The results show no signs of any systematic differences between the samplers. We obtain very similar results when looking at the MSEs for the missing observations in the right panel of Figure 5, where the MSEs of the first 50 missing observations in the data are shown. The results indicate a very high similarity across the precision samplers.

To sum up the simulation results, we only find differences between the samplers with respect to the computing time. In this regard, the 2-step sampler tends to be faster than the time-permutation sampler. However, the precision-based samplers yield similar MSEs for factors and missing observations, and inefficiency factors indicate no major convergence issues for a reasonably chosen number of burn-in draws. Due to these similarities in the simulations, we only consider the 2-step sampler in the empirical application below.

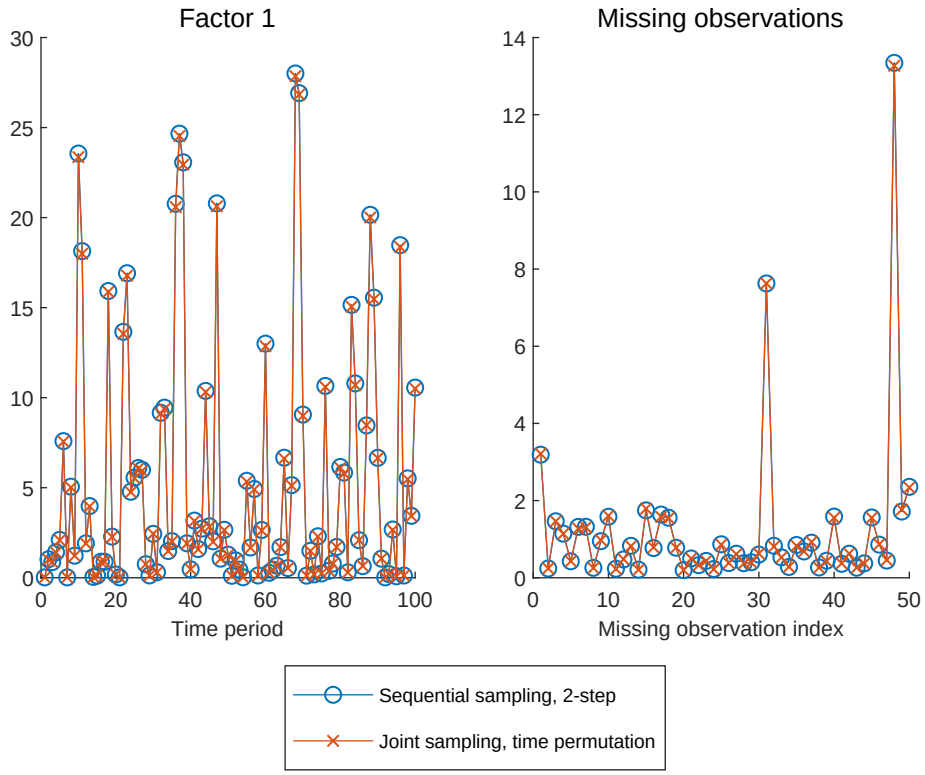
6 Empirical application: Bayesian estimation of international factors in GDP growth

To illustrate the precision-based sampler in the presence of missing observations in data, we estimate the factor model with Bayesian techniques on multi-country GDP growth data. The application follows the literature on international business cycles estimated in large factor models (Kose et al., 2003, 2008; Francis et al., 2017; Müller et al., 2019).

6.1 Data and motivation for the empirical exercise

Francis et al. (2017) estimate global and regional business cycles from a large set of country-specific annual GDP growth series. We follow these authors and choose the Penn World Tables (PWT) to construct the data set. We use PWT version 9.1 (Feenstra, Inklaar, and Timmer, 2015) and take annual real GDP (output concept) series for all

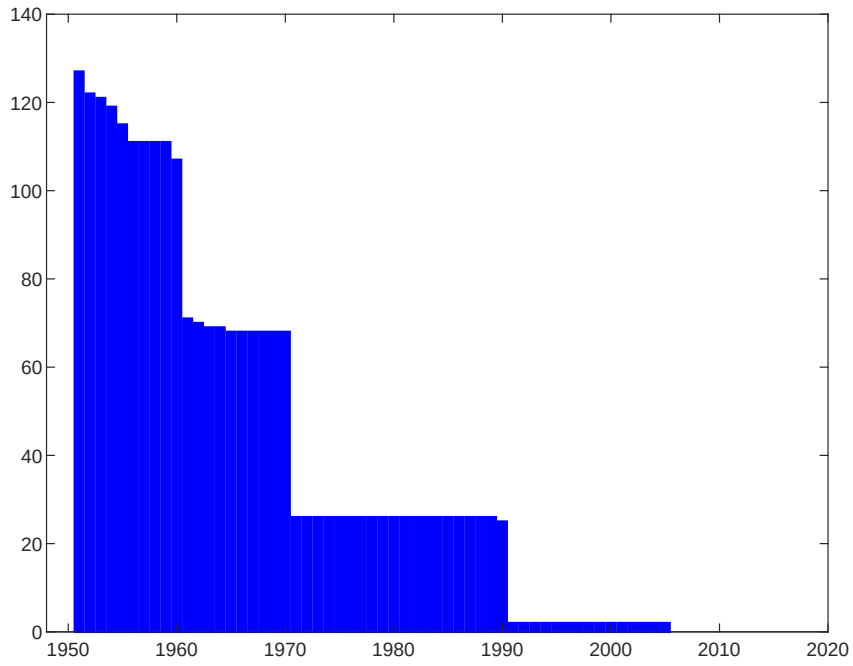
Figure 5: MSE comparison for one simulated dataset and posterior samples of first model factor and missing observations.



Note: In the left figure, the mean-squared errors (MSE) are computed by averaging the squared difference between $G^{\text{eff}} = 5000$ posterior draws for the first model factor and the corresponding factor sample from the DGP as defined in (31). The effective samples used are obtained by taking $G = 10000$ raw samples from the posterior and discarding the first $G^{\text{burn-in}} = 5000$ as burn-in. In the right figure, the MSE is shown for posterior draws of 50 randomly chosen missing values.

available countries. Growth rates are computed by taking first differences of the logarithm of the series in levels. We end up with $T = 67$ time series observations for the years 1951 to 2017 and $N = 182$ countries in the cross section. The data is unbalanced with respect to available observations per country. Overall, 2391 of the observations are missing, which is 19.6% in relation to $TN = 12194$ potential values in the data. The number of missing observations at each period in time are shown in Figure 6.

Figure 6: Missing observations in international GDP growth data



Note: The bars in the figure show the sum of missing observations in each time period. The maximum number of cross-section observations per period is equal to $N_{\text{unbal}} = 182$.

In the data we find that observations are mostly missing in earlier periods of the sample: In 1951 we have 127 observation missing out of 182, whereas no observations are missing at the end of the sample after 2005. The decline of the number of missing observations in the intermediate time periods can be roughly described as being step-wise across the covered decades. From 1952 to 1960, there are more than 100 observations missing per year, whereas the number of missing observations drops to about 70 until 1970. Between 1970 and 1990, slightly more than 20 time series observations are missing.

Compared to the existing literature by Kose et al. (2003) and Francis et al. (2017), the data set used here contains a larger number of variables with $N_{\text{unbal}} = 182$ and thus more cross-country information. If we remove all time series with any missing observations over the sample period 1951 and 2017, we end up with a balanced data set of $N_{\text{bal}} = 55$ time series, which is very close to the countries covered in Kose et al. (2003) and Francis et al. (2017). We estimate the factor model with these two different data sets and compare the results. In general, time-series data with partially missing observations encompasses balanced time-series data where all time series with any missing observations have been removed. Thus, estimation methods, which can tackle missing observations, can take

into account larger data sets and use more information to estimate models. Note that in general we cannot expect that larger data always improves factor estimation. Whether additional information is beneficial for factor estimation, depends - amongst other things - on the number of missing observations as well as the information content in the additional data. Boivin and Ng (2006) show that noisy data can even deteriorate accuracy of factor estimates.

The empirical comparison proceeds along several dimensions: We compare the factor estimates obtained from using the two different information sets. We also identify the relevant and irrelevant variables in both data sets along the lines of Kaufmann and Schumacher (2017). Irrelevant variables are defined as not being related to factors via the loading matrix and generally cannot contribute to estimating the factors in a state-space model (Koopman and Harvey, 2003). We also investigate how variables, which are part of both information sets, are explained by the different factor estimates. In particular, we estimate the common and idiosyncratic components and investigate how the variance contributions of the common components differ for the two different information sets.

6.2 Model specification and Bayesian estimation

In the model for the empirical application, the loading matrix has a group structure with $r = 6$ factors. Following the literature on international business cycles like Kose et al. (2003) and Francis et al. (2017), we define the first factor in the model as the global factor, such that all variables can load on this factor. Accordingly, there are no zero restrictions in the first column of the loading matrix. The other five factors are continental factors for Africa, Asia, Europe, North America, and South America. The continental group structure is imposed by zero restrictions: Each country GDP variable can load on only one of the continental factors in addition to the global factor. For each continental factor, those country GDP variables not belonging to this particular continent receive a zero loading element in the corresponding column of the loading matrix. Given these zero restrictions from the continental factor structure, the factor model is identified according to the Bekker criterion outlined in Bai and Wang (2014).

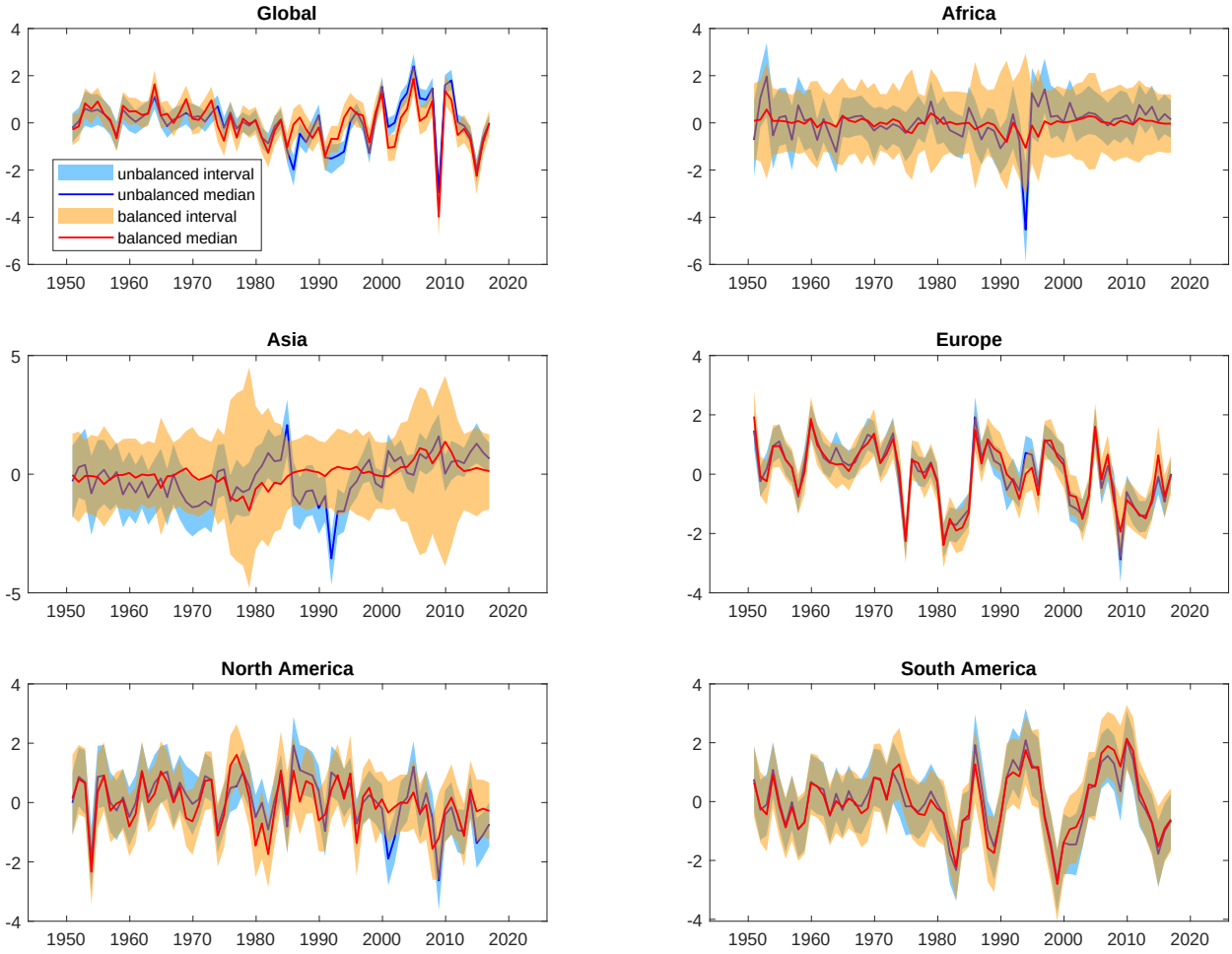
In the application, we use a more general model than specified in (1) and (2). We use $p = 2$ lags in the factor VAR and $q = 2$ lags in the AR equations for idiosyncratic components. Details on the extended model and the posterior sampler can be found in Appendix A. In the empirical application, we take 100000 draws from the posterior density, discard the first 50000 draws as burn-in, and take the remaining draws for final posterior evaluation. We estimate the factor model on balanced and unbalanced data and compare the results.

6.3 Results

The estimated factors using the balanced and unbalanced data set can be compared in Figure 7. We show the median of the posterior samples together with 90% posterior bands. The majority of the factors look similar when estimating the model on balanced and unbalanced data: The global factor, the European factor, and the two American factors do not change considerably with respect to the median and the posterior intervals. The African and Asian factors, however, are different when comparing balanced and

unbalanced data. Based on balanced data, the African and Asian factors are hardly different from zero in terms of 90% posterior intervals across time periods. Based on the larger unbalanced data, both factors are more precisely estimated and exhibit more pronounced cyclical swings.

Figure 7: Factors estimated on balanced and unbalanced data



Note: The straight lines refer to the median of the posterior distribution of the factors. The shaded ares cover the 90% posterior intervals.

Given the posterior draws of the loadings, we investigate the relevance or irrelevance of variables in the model. Following Kaufmann and Schumacher (2017), we define irrelevant variables as having only zero factor loadings, whereas relevant variables are related to at least one factor. Irrelevant variables are not explained by the factors and, vice versa, do not contain information for factor estimation. To distinguish irrelevant from relevant variables, we compute 90% highest posterior density (HPD) intervals for each element in the loading matrix to check whether the loading element is different from zero. We define relevant variables as having at least one non-zero factor loading in the corresponding row of the loading matrix. In Table 1, the number of relevant variables is shown for different continents. Within geographic regions, we observe some heterogeneity: Among

African GDP growth series in the balanced data set, we find only one out of nine (11%) relevant variables. In the unbalanced data set, we have 50 African GDP growth time series, and the proportion of relevant variables is 18% (9 of 50). In the balanced data, the proportion of relevant Asian GDP series in the model is 20% and thus a bit higher compared to African GDP series. Using unbalanced data implies an increase of the number of Asian GDP time series from 10 to 50. The proportion of relevant variables compared to balanced data increases to 38% (19 of 50). For Europe and North America, we observe a relatively large proportion of relevant variables compared to the other continents when using balanced data. However, there are very few relevant time series in the additional unbalanced data. For Europe, for example, the proportion of relevant variables decreases from 89% to 63% when adding unbalanced data. The results are similar for American GDP growth data. We summarize by looking at results for all variables in the two models: Overall we find 55% relevant variables in the balanced data, whereas we have only 37% relevant variables in the unbalanced data. This indicates that only some of the added unbalanced GDP growth time series provide additional information for factor estimation on top of the balanced data only.

Table 1: Relevant variables

Data	Unbalanced data	Balanced data
Method	Sequential 2-step	Standard precision-based
Africa	0.18 (9 of 50)	0.11 (1 of 9)
Asia	0.38 (19 of 50)	0.20 (2 of 10)
Europe	0.63 (25 of 40)	0.89 (16 of 18)
North America	0.23 (7 of 31)	0.50 (5 of 10)
South America	0.64 (7 of 11)	0.75 (6 of 8)
All variables	0.37 (67 of 182)	0.55 (30 of 55)

Note: Entries in the table are equal to the proportion of relevant variables in the geographical region. We compute highest posterior density (HPD) intervals for each element in the loading matrix to check whether the loading element is significantly different from the zero. We define relevant variables as having at least one significant non-zero factor loading in the corresponding row of the loading matrix.

We also look at the role of the common and idiosyncratic components for each time series in the data set. Figure 8 and Figure 9 show variance shares of common components, defined as the variance of the common component of a country GDP growth time series divided by the overall variance of the time series. The difference between one and the variance shares of common components provides the variance share of country-specific idiosyncratic components (Kose et al., 2003). Thus, the larger the variance share of common components, the more variance of GDP growth is explained by common factors and co-movements with other countries' GDP growth. The smaller the number, the more important are country-specific sources of GDP fluctuations. To summarize the posterior variability of variance shares of common components, we provide box plots. Green boxes refer to the results based on balanced data, whereas yellow boxes refer to the results based on unbalanced data. Figure 8 shows the variance share of the common components for those countries which are part of both the balanced an unbalanced data sets. Generally,

European countries (abbreviated by EU in the table) have a comparatively high share of common component variances, followed by South American (SA) countries. African (AF) and Asian (AS) countries' GDP growth have a comparatively low variance share of common components on average. When comparing estimates for balanced and unbalanced data, the box plots are often quite similar. For a large number of GDP growth series, there is no clear advantage from using larger unbalanced data with respect to the variance share of common components. This holds, for example, for the big developed countries USA, Canada, France, and Germany, amongst others. In Figure 9, we look at the variance share of the common components for those countries which are only part of the unbalanced data set. These are $N_{\text{unbal}} - N_{\text{bal}} = 182 - 55 = 127$ countries. We can see that there are some country GDP series with a considerable variance share of the common components, for example, Brunei, Georgia, and several Asian countries. The majority of GDP series, however, shows quite small numbers only up to 0.2, indicating a big role of idiosyncratic country-specific movements for many countries in the unbalanced data set.

To sum up the empirical results, many GDP series for Asian and African countries are not related to the factors. However, the proportion of Asian and African countries related to the factors is larger when using the unbalanced data rather than the balanced data. In the results based on unbalanced data, we generally find that the idiosyncratic components seem to dominate for the majority of countries, despite some country GDP growth series having a high variance share of the common components. The additional unbalanced data also has no strong effect on the commonality of the variables that are part of both the balanced and unbalanced data set. If an analyst were mostly interested in results for G7 countries, it might suffice to look at the posterior results from the smaller balanced data and using the simpler estimation methods. The use of the factor model based on the larger unbalanced data and the more demanding sampling methods to tackle missing observations may provide relevant insights, if the countries, which are exclusively part of the unbalanced data, are interesting in themselves to an analyst.

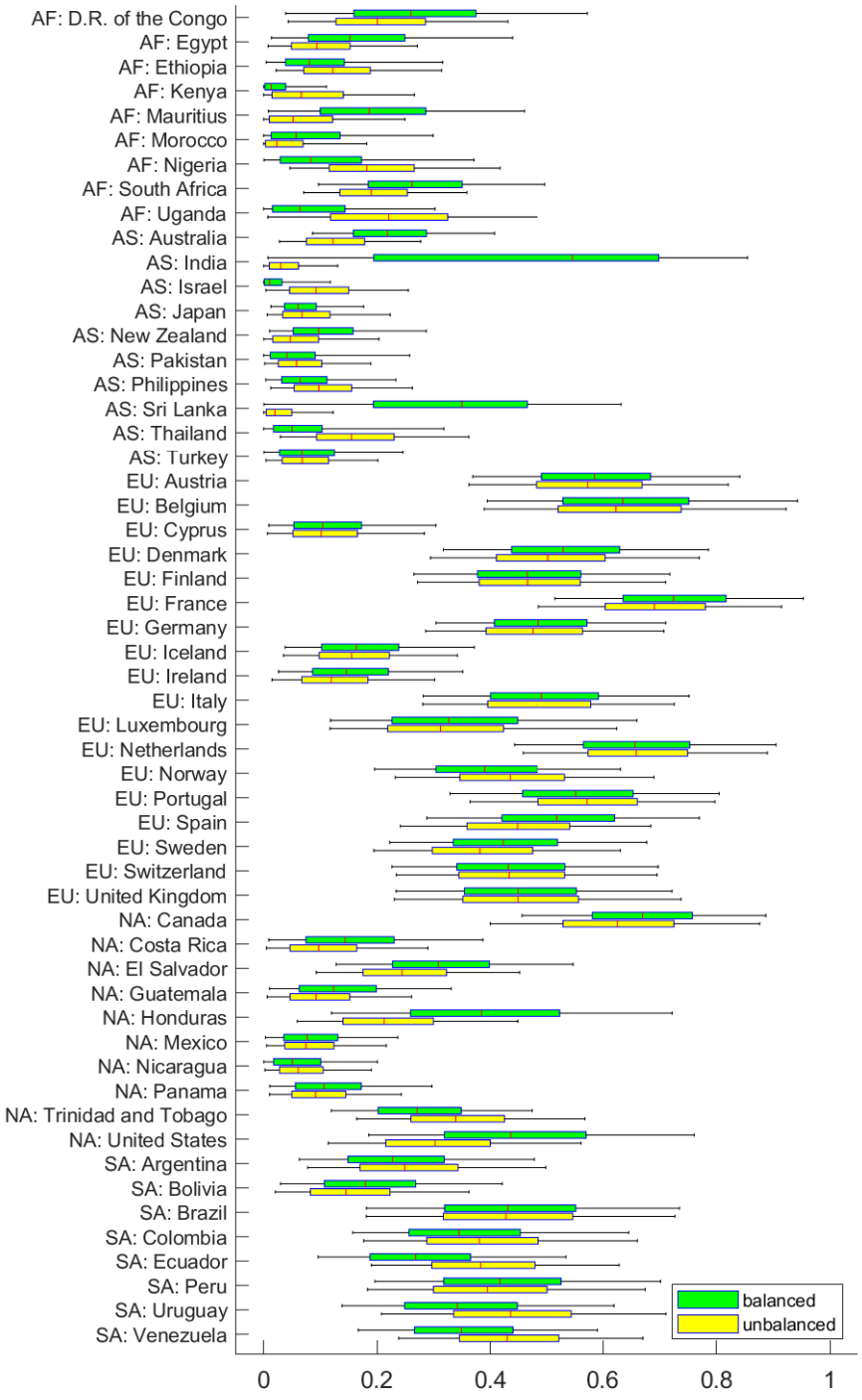
7 Extensions: Integrated likelihood, other state-space models

7.1 Integrated likelihood

The methods discussed in this paper can also be employed for model evaluation using the marginal likelihood. Following Chan and Grant (2016), the integrated likelihood function, obtained by integrating the unobserved states out of the conditional or complete likelihood function, $p(x|\theta) = \int p(x, \eta|\theta)d\eta = \int p(x|\eta, \theta)p(\eta|x, \theta)d\eta$, is the key quantity to derive the marginal likelihood. For factor models and given complete data for estimation, Chan and Grant (2016) and McCausland (2012) show that the integrated likelihood function has a normal distribution function. Given the quantities derived in this paper, we can show that the model (1) and (2) given missing observations also implies a normal density for the observed data.

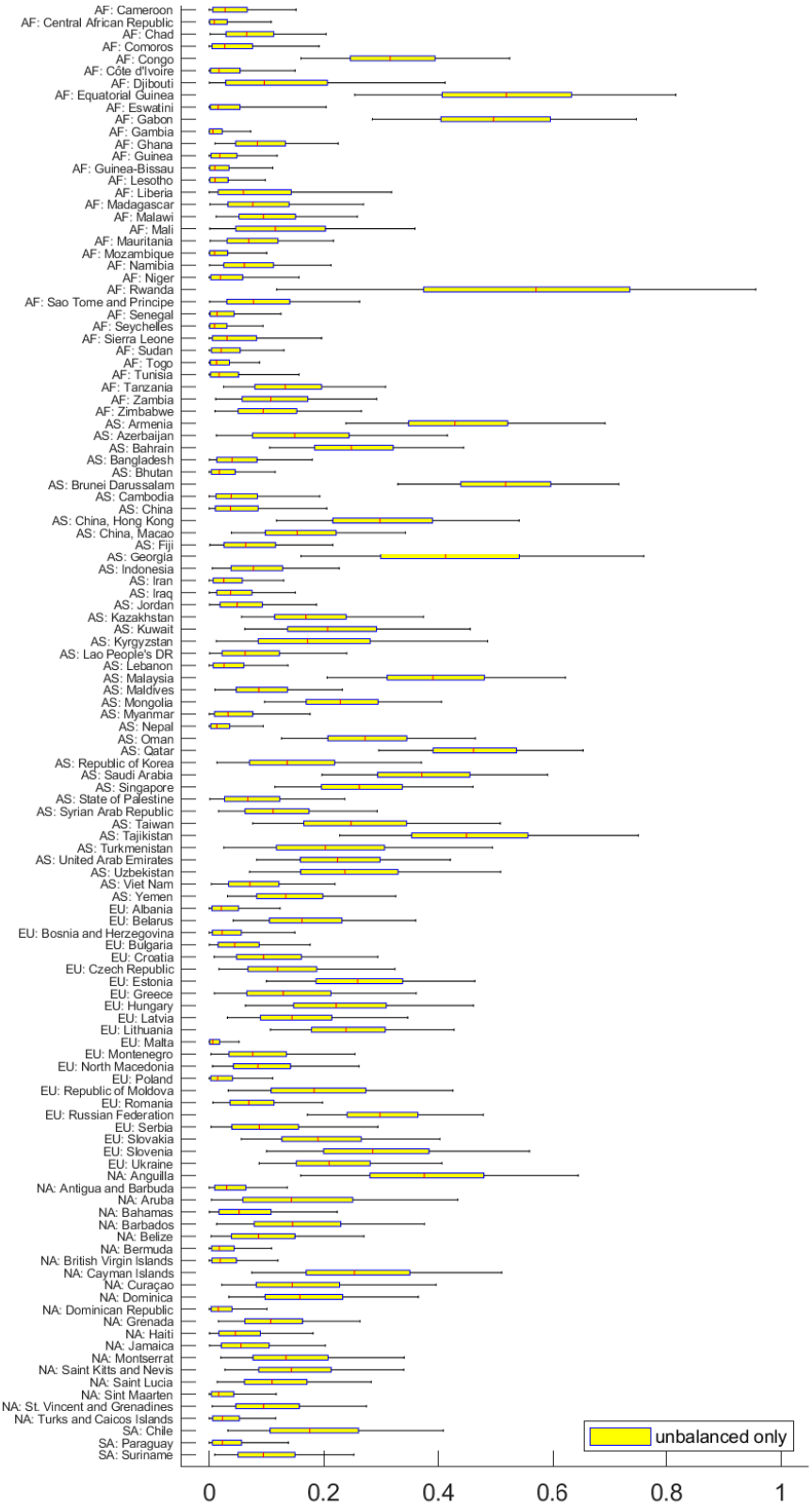
In Section 4, we have used joint normals for factors, variables with missing observations, and variables with observed data to derive conditional samplers. The permutations of variables $z_{P_{2s}} = (x_m^\intercal, \eta^\intercal, x_o^\intercal)^\intercal$ and $z_{P_\tau} = (\eta_1^\intercal, x_{m,1}^\intercal, \eta_2^\intercal, x_{m,2}^\intercal, \ldots, \eta_T^\intercal, x_{m,T}^\intercal, x_o^\intercal)^\intercal$ have all

Figure 8: Variance share of the common components, countries covered in both balanced and unbalanced data



Note: The figure shows box plots reflecting the posterior sample variability of the of the common components' variance divided by the overall variance of each GDP growth time series. The whiskers refer to 90% posterior interval bounds. Abbreviations for continents: AF Africa, AS Asia, EU Europe, NA North America, SA South America.

Figure 9: Variance share of the common components, countries only covered in unbalanced data



Note: The figure shows box plots reflecting the posterior sample variability of the of the common components' variance divided by the overall variance of each GDP growth time series. The whiskers refer to 90% posterior interval bounds. Abbreviations for continents: AF Africa, AS Asia, EU Europe, NA North America, SA South America.

in common that the variables with observed data x_o are ordered last.

According to (10), the permuted vector z_P is distributed as

$$z_{P_z}|\theta \sim \mathcal{N}\left(0_{T(r+N)\times 1}, Q_z^{-1}\right), \tag{32}$$

with $Q_z = P_z Q P_z^\intercal$ and the corresponding covariance matrix $\Omega_z = Q_z^{-1}$. For later use, we define the partitions

$$Q_z = \begin{pmatrix} Q_{\eta x_m, \eta x_m} & Q_{\eta x_m, x_o} \\ Q_{x_o, \eta x_m} & Q_{x_o, x_o} \end{pmatrix}, \quad \Omega_z = \begin{pmatrix} \Omega_{\eta x_m, \eta x_m} & \Omega_{\eta x_m, x_o} \\ \Omega_{x_o, \eta x_m} & \Omega_{x_o, x_o} \end{pmatrix}. \tag{33}$$

If variables jointly follow a normal distribution, the marginal distributions are also normal distributions, and the moments of the marginal distributions can be taken from the partitions of the joint moments (Anderson, 2003, Theorem 2.4.3). In our case, we obtain the marginal distribution of the variables with observed data $p(x_o|\theta)$ according to

$$p(x_o|\theta) \overset{\mathcal{D}}{=} \mathcal{N}\left(0_{(1-\kappa)TN\times 1}, \Omega_{x_o, x_o}\right). \tag{34}$$

It implies the log integrated likelihood function

$$\log p(x_o = x^o|\theta) = -\frac{(1-\kappa)TN}{2}\log(2\pi) - \frac{T}{2}\log\left|\Omega_{x_o,x_o}^{-1}\right| - \frac{1}{2}(x^o)^\intercal \Omega_{x_o,x_o}^{-1} x^o. \tag{35}$$

To evaluate the integrated likelihood function, we can make use of the time-permutation sampler from Section 4.2. From the definition of the covariance and precision matrix

$$\begin{pmatrix} \Omega_{\eta x_m, \eta x_m} & \Omega_{\eta x_m, x_o} \\ \Omega_{x_o, \eta x_m} & \Omega_{x_o, x_o} \end{pmatrix} \begin{pmatrix} Q_{\eta x_m, \eta x_m} & Q_{\eta x_m, x_o} \\ Q_{x_o, \eta x_m} & Q_{x_o, x_o} \end{pmatrix} = I \tag{36}$$

we obtain four matrix equations. From them, we can derive Ω_{x_o,x_o}^{-1} as an expression of the components of the precision matrix according to

$$\Omega_{x_o,x_o}^{-1} = Q_{x_o,x_o} - Q_{x_o,\eta x_m} Q_{\eta x_m,\eta x_m}^{-1} Q_{\eta x_m,x_o}, \tag{37}$$

and the log integrated likelihood function becomes

$$\begin{aligned} \log p(x_o = x^o|\theta) = &-\frac{(1-\kappa)TN}{2}\log(2\pi) - \frac{T}{2}\log\left|Q_{x_o,x_o} - Q_{x_o,\eta x_m} Q_{\eta x_m,\eta x_m}^{-1} Q_{\eta x_m,x_o}\right| \\ &- \frac{1}{2}(x^o)^\intercal \left(Q_{x_o,x_o} - Q_{x_o,\eta x_m} Q_{\eta x_m,\eta x_m}^{-1} Q_{\eta x_m,x_o}\right) x^o. \end{aligned} \tag{38}$$

Evaluating the quantities is straightforward using the results from the time-permutation sampler. In Appendix B, we show that $Q_{\eta x_m,\eta x_m}$ is block-banded, so the methods outline before can be directly applied in line with McCausland (2012) and Chan and Grant (2016).

Note that these results also allow to estimate a factor model with stochastic volatility as proposed in Chan and Eisenstat (2018).

7.2 Other state-space models

The methods in the paper were so far applied to a factor model. However, they can be easily adopted to other state-space models in applications with missing data.

If we set $\psi = 0_{N\times N}$ in (2), we obtain a general state-space model, where η_t represents the state vector. In this case, we have $\Psi = I_{TN}$, which simplifies the precision matrix in (8). We can also consider cases where the number of disturbances in the state equation is smaller to the number of states, for example, if $\eta_t = \phi\eta_{t-1} + R_t u_{\eta,t}$ with the number of columns in R_t and the length of $u_{\eta,t}$ smaller than r, as in the general state-space models discussed in Durbin and Koopman (2002).

In a similar way, we can generalize the TVP-BVAR models as in Chan and Eisenstat (2018) to applications with missing observations in the data. Consider the model

$$B_{0,t}y_t = \mu_t + B_{1,t}y_{t-1} + \cdots + B_{p,t}y_{t-p} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \Sigma_t), \tag{39}$$

where y_t is an N-dimensional vector of model variables, μ_t is a vector of time-varying intercepts, $B_{1,t}, \ldots, B_{p,t}$ are time-varying VAR coefficient matrices, $B_{0,t}$ is a lower triangular matrix with ones on the diagonal, and time-varying volatilities Σ_t for $t = 1, \ldots, T$. In the literature such as Chan and Eisenstat (2018); Chan (2020); Chan et al. (2020), estimation of the VAR coefficients using precision-based samplers given complete data proceeds by conditioning on lags of y_t on the right-hand side. If some of the observations are missing, this conditioning is not feasible. In this case, the methods developed in this paper can be adopted to obtain a Gibbs sampler for TVP-BVAR models in the presence of partly missing observations along the lines of Little and Rubin (2002). In particular, we can provide a sampling step for missing observations in y_t. Given augmented data, the samplers as proposed in the papers cited above can be applied for Bayesian estimation of the model parameters in (39).

To derive the conditional posterior distribution for missing observations, rewrite the model (39) in stacked form with $y = (y_1^\mathsf{T}, y_2^\mathsf{T}, \ldots, y_T^\mathsf{T})^\mathsf{T}$ and $\tilde{\mu}_y = (\mu_1^\mathsf{T}, \mu_2^\mathsf{T}, \ldots, \mu_T^\mathsf{T})^\mathsf{T}$ as

$$\Phi_y y = \tilde{\mu}_y + \epsilon_y \tag{40}$$

with

$$\Phi_y = \begin{pmatrix} B_{0,1} & & & & & & \\ -B_{1,1} & B_{0,2} & & & & & \\ -B_{2,1} & -B_{1,2} & B_{0,3} & & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ -B_{p,1} & & \ddots & \ddots & \ddots & & \\ & \ddots & & \ddots & \ddots & B_{0,T-1} & \\ & & -B_{p,T-p} & \cdots & -B_{2,T-2} & -B_{1,T-1} & B_{0,T} \end{pmatrix}, \tag{41}$$

$\epsilon_y \sim \mathcal{N}(0, \Sigma)$, and $\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_T)$.

Given model parameters θ_y, we have $y|\theta_y \sim \mathcal{N}(\mu_y, \Phi_y^{-1}\Sigma\Phi_y^{-\mathsf{T}})$ with $\mu_y = \Phi_y^{-1}\tilde{\mu}_y$ and precision matrix $Q_y = (\Phi_y^{-1}\Sigma\Phi_y^{-\mathsf{T}})^{-1} = \Phi_y^\mathsf{T}\Sigma^{-1}\Phi_y$. We can now reorder the variables in y into variables with missing observations y_m above the variables with available observations

y_o as in (12) by

$$y_{P_y} = \begin{pmatrix} y_m \\ y_o \end{pmatrix} = P_y y = \begin{pmatrix} P_{y_m} \\ P_{y_o} \end{pmatrix} y \tag{42}$$

such that the transformed vector y_{P_y} is distributed as

$$y_{P_y} | \theta_y \sim \mathcal{N}\left(P_y \mu_y, (P_y Q_y P_y^{\mathsf{T}})^{-1}\right). \tag{43}$$

We can now partition the mean and the precision matrix as in (42) and obtain

$$P_y \mu_y = \begin{pmatrix} \mu_m \\ \mu_o \end{pmatrix}, \qquad P_y Q_y P_y^{\mathsf{T}} = \begin{pmatrix} Q_{y_m,y_m} & Q_{y_m,y_o} \\ Q_{y_o,y_m} & Q_{y_o,y_o} \end{pmatrix}, \tag{44}$$

which provides us with the moments of the distribution for the missing observations y_m conditional on the observations $y_o = y^o$ and the model parameters. The conditional distribution of the missing values is given by

$$p(y_m | y_o = y^o, \theta_y) \overset{\mathcal{D}}{=} \mathcal{N}\left(\mu_m - Q_{y_m,y_m}^{-1} Q_{y_m,y_o}(y^o - \mu_o), Q_{y_m,y_m}^{-1}\right). \tag{45}$$

Note that it is not straightforward to jointly sample missing values and time-varying VAR parameters in the TVP-BVAR model. The reason is that products of missing values and time-varying VAR parameters are present on the right-hand side of the TVP-BVAR model, rendering the resulting state-space system non-linear.

8 Conclusions

We propose a simple and efficient precision-based sampler for unobserved states and missing observations conditional on available data. The approach extends the existing literature on precision-based sampling, which typically considers complete-data applications. By allowing the investigation of incomplete data sets, the sampler proposed here expands the range of potential applications for precision-based samplers in practice.

The approach can be applied to a wide range of state-space models such as time-varying parameter BVARs, as their corresponding precision matrix has a block-banded structure. In this paper we apply the sampler to a large dynamic factor model. To facilitate sampling in the presence of missing observations, we reorder the variables in the precision matrix by alternative permutations. Based on the permuted precision matrices with small bandwidth, we can employ fast band-matrix computing techniques to draw from the conditional distributions of factors and missing observations given available data. In the simulations, a 2-step precision-based sampler, which sequentially samples factors and missing observations, turns out to be computationally efficient. On the other hand, a joint sampler based on time permutation is slower, but facilitates computing the integrated likelihood for model comparison more easily. Both can be directly integrated into Bayesian estimation procedures like the Gibbs sampler.

Appendix

A Posterior sampler

A more general specification of the factor model than (1) and (2) has $p \geq 1$ lags in the factor VAR and $q \geq 1$ AR lags in the equations for idiosyncratic components. We obtain the model equations

$$x_t = \lambda \eta_t + \epsilon_t, \tag{46}$$
$$\phi(L)\eta_t = u_{\eta,t}, \quad \psi(L)\epsilon_t = u_{\epsilon,t}. \tag{47}$$

with polynomials $\phi(L) = I_r - \phi_1 L - \ldots - \phi_p L^p$ and $\psi(L) = I_N - \psi_1 L - \ldots - \psi_q L^q$ in the lag operator $Ly_t = y_{t-1}$. The polynomial matrices in $\psi(L)$ are assumed to be diagonal. The factor VAR disturbances are distributed as $u_{\eta,t} \sim \mathcal{N}(0_{r\times 1}, \omega_\eta)$, and we assume for simplicity $\omega_\eta = I_r$. With respect to the idiosyncratic components we assume $u_{\epsilon,t} \sim \mathcal{N}(0_{N\times 1}, \omega_\epsilon)$, where ω_ϵ is also diagonal with blocks $\omega_{\epsilon,i}$ for $i = 1, \ldots, N$ on the main diagonal.

Concerning the prior specifications of the model, we follow closely the existing factor model literature. We use sparse priors for the elements in the loading matrix as in George and McCulloch (1993, 1997); Geweke (1996); Carvalho, Chang, Lucas, Nevins, Wang, and West (2008); Kaufmann and Schumacher (2017, 2019). The hierarchical prior for the loadings is denoted as $p(\lambda|\theta_\lambda)$ with a prior for hyper-parameters $p(\theta_\lambda)$. The factors and missing observations follow multivariate normal priors given model parameters, $p(\eta|\theta_+)$ and $p(x_m|\theta_+)$, where further model parameters are collected in θ_+: It contains the factor VAR polynomial parameters $\phi(L)$ and the idiosyncratic components AR polynomial parameters $\psi(L)$, which follow normal distributions truncated to the stationary region (Litterman, 1986). θ_+ also contains the variances of the innovations in the idiosyncratic components AR models $\omega_{\epsilon,i}$ for $i = 1, \ldots, N$. We use an inverse gamma distribution as the prior. Details regarding the prior specifications can be found in the subsections below.

Given available data x_o, we want to obtain samples from the posterior distribution

$$p(\eta, x_m, \lambda, \theta_\lambda, \theta_+|x_o) \propto L(x_o|\eta, x_m, \lambda, \theta_\lambda, \theta_+)p(x_m|\theta_+)p(\lambda|\theta_\lambda)p(\eta|\lambda, \theta_+)p(\theta_+)p(\theta_\lambda). \tag{48}$$

The sampler closely follows the Gibbs sampler proposed by Little and Rubin (2002) to tackle missing data. In a first step, Little and Rubin (2002) sample missing observations for x_m, and combine these samples with observed data. In subsequent steps, complete-data conditional posterior distributions are used for sampling the remaining model parameters. The use of the complete-data conditional posterior distributions in these steps is valid because the product of $L(x_o|\eta, x_m, \lambda, \theta_\lambda, \theta_+)$ and $p(x_m|\theta_+)$ in (48) equals the complete-data likelihood $L(x_o, x_m|\eta, \lambda, \theta_\lambda, \theta_+)$, as implied by the model equations (46) and (47).

In our case, we expand on Little and Rubin (2002) by also estimating unobserved factors in the first block of the Gibbs sampler along with the missing observations. The sampler proceeds with the following three steps:

1. $p(\eta, x_m|x_o, \lambda, \theta_\lambda, \theta_+)$: To sample factors and missing observations, we have two options:

(a) When applying time permutation precision-based sampling, we provide a joint draw of $\eta, x_m|x_o, \lambda, \theta_\lambda, \theta_+$ in a single step.

(b) When applying the 2-step precision-based sampling, we sample sequentially from $p(\eta|x_o, x_m, \lambda, \theta_\lambda, \theta_+)$ and $p(x_m|x_o, \eta, \lambda, \theta_\lambda, \theta_+)$.

2. $p(\lambda|x_o, x_m, \eta, \theta_\lambda, \theta_+)$, $p(\theta_\lambda|x_o, x_m, \lambda)$: Sampling the loadings λ, and hyper-parameters related to the hierarchical prior for the loadings θ_λ given loadings λ.

3. $p(\theta_+|x_o, x_m, \eta, \lambda, \theta_\lambda)$: Sampling the rest of the model parameters.

A.1 Factors and missing observations: $p(\eta, x_m|x_o, \lambda, \theta_\lambda, \theta_+)$

With lag orders $p, q > 1$ in the factor model, the precision-based samplers as described in the main text for $p = q = 1$ have to be modified only in terms of the precision matrix Q. With respect to the stacked factors, we have $\eta|\phi(L), \omega_\eta, \vartheta_{\eta,1|0} \sim \mathcal{N}(0_{Tr\times 1}, \Phi^{-1}\Omega_\eta\Phi^{-\intercal})$, where

$$\Phi = \begin{pmatrix} I_r & & & & & & \\ -\phi_1 & I_r & & & & & \\ -\phi_2 & -\phi_1 & I_r & & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ -\phi_p & & \ddots & \ddots & \ddots & & \\ & \ddots & & \ddots & \ddots & I_r & \\ & & -\phi_p & \cdots & -\phi_2 & -\phi_1 & I_r \end{pmatrix}, \quad \Omega_\eta = \begin{pmatrix} \vartheta_{\eta,1|0} & & & \\ & \omega_\eta & & \\ & & \ddots & \\ & & & \omega_\eta \end{pmatrix}. \tag{49}$$

The matrix entries for the VAR lag polynomial $\phi(L)$ are taken from the previous Gibbs draw. We set $\omega_\eta = I_r$. As the initial covariance of factors $\vartheta_{\eta,1|0} = \vartheta_\eta$, we use the covariance matrix of factors and their lags implied by the VAR in companion form

$$\underbrace{\begin{pmatrix} \eta_t \\ \eta_{t-1} \\ \vdots \\ \eta_{t-p} \end{pmatrix}}_{\tilde{\eta}_t} = \underbrace{\begin{pmatrix} \phi_1 & \cdots & \cdots & \phi_p \\ I_r & & & 0 \\ & \ddots & & \vdots \\ & & I_r & 0 \end{pmatrix}}_{\tilde{\phi}} \begin{pmatrix} \eta_{t-1} \\ \eta_{t-2} \\ \vdots \\ \eta_{t-(p+1)} \end{pmatrix} + \underbrace{\begin{pmatrix} u_{\eta,1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\tilde{u}_{\eta,t}}. \tag{50}$$

The VAR innovations in the companion form $\tilde{u}_{\eta,t}$ have the $(pr \times pr)$ covariance matrix $\tilde{\omega}_\eta = \mathrm{Var}(\tilde{u}_{\eta,t})$, which contains ω_η in the top-left block and zeros elsewhere. The companion form implies the $(pr \times pr)$ second-moment matrix of stacked factors lags $\tilde{\vartheta}_\eta = \mathrm{Var}(\tilde{\eta}_t)$ from solving the equation $\mathrm{vec}(\tilde{\vartheta}_\eta) = [I_{(pr)^2} - \tilde{\phi} \otimes \tilde{\phi}^\intercal]^{-1} \times \mathrm{vec}(\tilde{\omega}_\eta)$. As the initial covariance of factors $\vartheta_{\eta,1|0} = \vartheta_\eta$, we use the first $(r \times r)$ block-diagonal element in $\tilde{\vartheta}_\eta$.

With respect to the stacked idiosyncratic components, we have $\epsilon|\psi(L), \omega_\epsilon, \vartheta_{\epsilon,1|0} \sim \mathcal{N}(0_{TN\times 1}, \Psi^{-1}\Omega_\epsilon\Psi^{-\intercal})$, where Ψ and Ω_ϵ are constructed along the lines of (49). The matrix entries for the lag polynomial $\phi(L)$ and the innovation variances ω_ϵ are taken from the previous Gibbs draw. To initialize the second moments of the idiosyncratic components by $\vartheta_{\epsilon,1|0}$, we use the second moments implied by the AR models for the idiosyncratic

components, which can be computed using a companion form in the same way as for the factors.

Given Φ, Ω_η, Ψ, and Ω_ϵ, we define the precision matrix Q in the same way as in (7) and (8) in the main text and can apply the alternative precision-based samplers accordingly.

A.2 Loadings: $p(\lambda|x_o, x_m, \eta, \theta_\lambda, \theta_+)$, $p(\theta_\lambda|x_o, x_m, \lambda)$

Following George and McCulloch (1993, 1997); Geweke (1996); Kaufmann and Schumacher (2019), we use a sparse point-mass normal mixture prior on the loadings with a common probability of non-zero loading on factor j across variables according to

$$p(\lambda_{ij}) \overset{\mathcal{D}}{=} (1-\rho_j)\delta_0(\lambda_{ij}) + \rho_j \, \mathcal{N}(0, \tau_j), \tag{51}$$

$$p(\rho_j) \overset{\mathcal{D}}{=} \mathcal{B}(r_0 s_0, r_0(1-s_0)), \tag{52}$$

where the Dirac delta function $\delta_0(\cdot)$ assigns all probability mass to zero. To capture potential factor-specific scaling of loadings, we specify an inverse gamma distribution for $\tau_j \sim \mathcal{IG}(g_0, G_0)$. The expected probability of non-zero factor loading, s_0, and precision r_0 are hyperparameters. In the empirical application, we specify $s_0 = 0.5$, $r_0 = 3.0$, $g_0 = 2$, and $G_0 = 0.5$. We define the hyper-parameters by $\theta_\lambda = \{\rho_j, \tau_j\}$ for $j = 1, \dots, r$.

Given complete data x, which combines samples of missing data and the observed data, we can sample from $p(\lambda|x, \eta, \theta, \theta_\lambda)$ in the following way. The posterior odds of a non-zero factor loading in (51) are given by

$$\frac{P(\lambda_{ij} \neq 0|x, \cdot)}{P(\lambda_{ij} = 0|x, \cdot)} = \frac{p(\lambda_{ij})|_{\lambda_{ij}=0}}{p(\lambda_{ij}|\cdot)|_{\lambda_{ij}=0}} \frac{\rho_j}{1-\rho_j} = \frac{\mathcal{N}(0; 0, \tau_j)}{\mathcal{N}(0; m_{ij}, M_{ij})} \frac{\rho_j}{1-\rho_j}, \tag{53}$$

where the moments m_{ij} and M_{ij} are

$$M_{ij} = \left(\frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} \left(\psi_i(L)\eta_{jt}\right)^2 + \frac{1}{\tau_j} \right)^{-1}, \quad m_{ij} = M_{ij} \left(\frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} \left(\psi_i(L)\eta_{jt}\right) x_{it}^* \right), \tag{54}$$

and x_{it}^* is a transform of the variables by

$$x_{it}^* = \psi_i(L)x_{it} - \sum_{l=1, l\neq j}^{k} \lambda_{il}\psi_i(L)\eta_{lt} = \lambda_{ij}\psi_i(L)\eta_{jt} + \epsilon_{it},$$

which isolates the effect of factor j on variable i. Note that the conditional sampler in the empirical application is applied only to those elements in λ, which are not fixed to zero in the continental group factors.

We choose $\lambda_{ij} \neq 0$ if $U \leq PO_{ij}/(1+PO_{ij})$, where U is a draw from the uniform distribution over $[0, 1]$. If we choose $\lambda_{ij} \neq 0$, λ_{ij} is drawn from $\mathcal{N}(m_{ij}, M_{ij})$, otherwise it is set equal to zero.

Given λ_{ij}, we can update the hyper-parameters θ_λ. The conditional posterior of ρ_j is $p(\rho_j|x, \cdot) \overset{\mathcal{D}}{=} \mathcal{B}(r_{1j}, r_{2j})$ with $r_{1j} = r_0 s_0 + S_j$, $r_{2j} = r_0(1-s_0) + N_j - S_j$, and N_j is the number of loading elements not fixed to zero a-priori in the continental factors for

$j = 2, \ldots, 6$, whereas for the global factor indexed by $j = 1$, we have $N_1 = N$. We also define $S_j = \sum_{i=1}^{N} I\{\lambda_{ij} \neq 0\}$. In the simulation experiments in Section 5, we do not set zero restrictions in the loadings. In that case, we have to use $r_{2j} = r_0(1 - s_0) + N - S_j$.

The conditional posterior of the τ_j is $p(\tau_j|x, \cdot) \overset{\mathcal{D}}{=} \mathcal{IG}(g_j, G_j)$ with $g_j = g_0 + 0.5N_j$ and $G_j = G_0 + 0.5\sum_{i=1}^{N} \lambda_{ij}^2$.

A.3 Further parameters: $p(\theta_+|x_o, x_m, \eta, \lambda, \theta_\lambda)$

In the VAR polynomial for the factors and the AR polynomial for the idiosyncratic components, we employ a standard Minnesota prior specification (Litterman, 1986). The prior variance for the first autoregressive lag is set to $\vartheta_0^2 = 0.09$, whereas the shrinkage factor for the prior variance on lags of other variables is equal to $\vartheta_1^2 = 0.03$. The normal prior moments for the VAR polynomial parameters are mean zero and variance $\text{Var}(\phi_{l,i,j}) = \vartheta_0^2/l^2$ for $i = l$ at lag l, and $\text{Var}(\phi_{l,i,j}) = \vartheta_1^2\vartheta_0^2/l^2$ for $i \neq l$ at lag l.

With respect to the variance of the innovations in the idiosyncratic components AR model, we choose $p(\omega_{\epsilon,i}) \overset{\mathcal{D}}{=} \mathcal{IG}(2.0, 1.0)$ for $i = 1, \ldots, N$. The conditional posteriors are standard in the literature and not reported here. More details can be found in Appendix A.3 in Kaufmann and Schumacher (2019).

B Precision matrix of joint time-permutation sampler

Given the permutation matrix P_τ, we obtain the permuted precision matrix $Q_\tau = P_\tau Q P_\tau^\mathsf{T}$ for sampling factors and missing observations. Below we show that Q_τ is a block-banded matrix. In particular, it is a block tridiagonal matrix with varying block dimensions, having blocks in the lower diagonal, main diagonal and upper diagonal, and all other blocks being zero matrices (Golub and Van Loan, 2013, Chapter 4.5).

To derive Q_τ, first note that the precision matrix Q from (8) is a function of the prior precision matrices Q_η and Q_ϵ according to

$$Q = \begin{pmatrix} \Phi^\mathsf{T}\Omega_\eta^{-1}\Phi + \Lambda^\mathsf{T}(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi)\Lambda & -\Lambda^\mathsf{T}(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi) \\ -(\Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi)\Lambda & \Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi \end{pmatrix} = \begin{pmatrix} Q_\eta + \Lambda^\mathsf{T}Q_\epsilon\Lambda & -\Lambda^\mathsf{T}Q_\epsilon \\ -Q_\epsilon\Lambda & Q_\epsilon \end{pmatrix}. \quad (55)$$

As factors and idiosyncratic components follow VAR(1) processes, both precision matrices Q_η and Q_ϵ are block tridiagonal (Chan and Jeliazkov, 2009). In particular, we have

$$\begin{aligned} Q_\eta &= \Phi^\mathsf{T}\Omega_\eta^{-1}\Phi \\ &= \begin{pmatrix} \vartheta_{\eta,1|0}^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi & -\phi^\mathsf{T}\omega_\eta^{-1} & & & \\ -\omega_\eta^{-1}\phi & \omega_\eta^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi & -\phi^\mathsf{T}\omega_\eta^{-1} & & \\ & \ddots & \ddots & \ddots & \\ & & -\omega_\eta^{-1}\phi & \omega_\eta^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi & -\phi^\mathsf{T}\omega_\eta^{-1} \\ & & & -\omega_\eta^{-1}\phi & \omega_\eta^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi \end{pmatrix}, \end{aligned} \quad (56)$$

and $Q_\epsilon = \Psi^\mathsf{T}\Omega_\epsilon^{-1}\Psi$ can be written accordingly.

Given Q_η and Q_ϵ, we can show that the four submatrices of Q in (55),

$$Q = \begin{pmatrix} Q_\eta + \Lambda^\mathsf{T} Q_\epsilon \Lambda & -\Lambda^\mathsf{T} Q_\epsilon \\ -Q_\epsilon \Lambda & Q_\epsilon \end{pmatrix} = \begin{pmatrix} Q_{\eta\eta} & Q_{\eta x} \\ Q_{x\eta} & Q_{xx} \end{pmatrix}, \tag{57}$$

are also block tridiagonal matrices each. Note that Λ is a block-diagonal matrix. Pre- and post-multiplying the block tridiagonal matrix Q_ϵ by Λ leads to a block tridiagonal matrix without altering the block bandwidth. The sum of two block tridiagonal matrices $Q_\eta + \Lambda^\mathsf{T} Q_\epsilon \Lambda$ is also a block tridiagonal matrix. In particular, we obtain

$$Q_{\eta\eta} = \begin{pmatrix} Q_{\eta\eta,(1,1)} & Q_{\eta\eta,(1,2)} & & & & & \\ Q_{\eta\eta,(2,1)} & Q_{\eta\eta,(2,2)} & Q_{\eta\eta,(2,3)} & & & & \\ & Q_{\eta\eta,(3,2)} & Q_{\eta\eta,(3,3)} & Q_{\eta\eta,(3,4)} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & Q_{\eta\eta,(T-1,T-2)} & Q_{\eta\eta,(T-1,T-1)} & Q_{\eta\eta,(T-1,T)} \\ & & & & Q_{\eta\eta,(T,T-1)} & Q_{\eta\eta,(T,T)} \end{pmatrix} \tag{58}$$

with

$$Q_{\eta\eta,(1,1)} = \vartheta_{\eta,1|0}^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi + \lambda^\mathsf{T}(\vartheta_{\epsilon,1|0}^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi)\lambda, \tag{59}$$

$$Q_{\eta\eta,(t,t)} = \omega_\eta^{-1} + \phi^\mathsf{T}\omega_\eta^{-1}\phi + \lambda^\mathsf{T}(\omega_\epsilon^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi)\lambda \quad \text{for} \quad t = 2, \ldots, T, \tag{60}$$

$$Q_{\eta\eta,(t,t+1)} = -\phi^\mathsf{T}\omega_\eta^{-1} - \lambda^\mathsf{T}(-\psi^\mathsf{T}\omega_\epsilon^{-1})\lambda \quad \text{for} \quad t = 1, \ldots, T-1, \tag{61}$$

$$Q_{\eta\eta,(t+1,t)} = -\omega_\eta^{-1}\phi - \lambda^\mathsf{T}(\omega_\epsilon^{-1}\psi)\lambda \quad \text{for} \quad t = 1, \ldots, T-1. \tag{62}$$

Accordingly, the non-zero submatrices of $Q_{\eta x} = \left(Q_{\eta x,(t_1,t_2)}\right)_{t_1,t_2=1,\ldots,T}$ are given by

$$Q_{\eta x,(1,1)} = -\lambda^\mathsf{T}(\vartheta_{\epsilon,1|0}^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi), \tag{63}$$

$$Q_{\eta x,(t,t)} = -\lambda^\mathsf{T}(\omega_\epsilon^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi) \quad \text{for} \quad t = 2, \ldots, T, \tag{64}$$

$$Q_{\eta x,(t,t+1)} = \lambda^\mathsf{T}(\psi^\mathsf{T}\omega_\epsilon^{-1}) \quad \text{for} \quad t = 1, \ldots, T-1, \tag{65}$$

$$Q_{\eta x,(t+1,t)} = \lambda^\mathsf{T}(\omega_\epsilon^{-1}\psi) \quad \text{for} \quad t = 1, \ldots, T-1, \tag{66}$$

and all other submatrices are zero blocks, $Q_{\eta x,(t_1,t_2)} = 0$ for $|t_1 - t_2| > 1$. Finally, the non-zero submatrices of $Q_{x\eta} = \left(Q_{x\eta,(t_1,t_2)}\right)_{t_1,t_2=1,\ldots,T}$ are given by

$$Q_{x\eta,(1,1)} = -(\vartheta_{\epsilon,1|0}^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi)\lambda, \tag{67}$$

$$Q_{x\eta,(t,t)} = -(\omega_\epsilon^{-1} + \psi^\mathsf{T}\omega_\epsilon^{-1}\psi)\lambda \quad \text{for} \quad t = 2, \ldots, T, \tag{68}$$

$$Q_{x\eta,(t,t+1)} = (\psi^\mathsf{T}\omega_\epsilon^{-1})\lambda \quad \text{for} \quad t = 1, \ldots, T-1, \tag{69}$$

$$Q_{x\eta,(t+1,t)} = (\omega_\epsilon^{-1}\psi)\lambda \quad \text{for} \quad t = 1, \ldots, T-1, \tag{70}$$

and $Q_{x\eta,(t_1,t_2)} = 0$ for $|t_1 - t_2| > 1$.

The precision matrix $Q_\tau = P_\tau Q P_\tau^\mathsf{T}$ can be obtained by block multiplication

$$Q_\tau = \begin{pmatrix} P_{\eta,1} & 0 \\ 0 & P_{x_m,1} \\ P_{\eta,2} & 0 \\ 0 & P_{x_m,2} \\ \vdots & \vdots \\ P_{\eta,T} & 0 \\ 0 & P_{x_m,T} \\ 0 & P_{x_o} \end{pmatrix} \begin{pmatrix} Q_{\eta\eta} & Q_{\eta x} \\ Q_{x\eta} & Q_{xx} \end{pmatrix} \begin{pmatrix} P_{\eta,1}^\mathsf{T} & 0 & P_{\eta,2}^\mathsf{T} & 0 & \cdots & P_{\eta,T}^\mathsf{T} & 0 & 0 \\ 0 & P_{x_m,1}^\mathsf{T} & 0 & P_{x_m,2}^\mathsf{T} & \cdots & 0 & P_{x_m,T}^\mathsf{T} & P_{x_o}^\mathsf{T} \end{pmatrix} \tag{71}$$

with $P_{\eta,t} = \begin{pmatrix} 0_{r\times r(t-1)} & I_r & 0_{r\times r(T-t)} \end{pmatrix}$ and $P_{x_m,t} = \begin{pmatrix} 0_{N_{m,t}\times(t-1)N} & P_{x_m,(t,t)} & 0_{N_{m,t}\times(T-t)N} \end{pmatrix}$. Note that the permuted variables z_{P_τ} permutation matrix P_τ can be considered as having $T+1$ row blocks, where T row blocks contain model factors η_t and variables corresponding to missing observations $x_{m,t}$ for each period t, and the $(T+1)$-st row block contains variables corresponding to observed data x_o. This blocking is implied by the variable ordering in $z_{P_\tau} = (\eta_1^\mathsf{T}, x_{m,1}^\mathsf{T}, \eta_2^\mathsf{T}, x_{m,2}^\mathsf{T}, \ldots, \eta_T^\mathsf{T}, x_{m,T}^\mathsf{T}, x_o^\mathsf{T})^\mathsf{T}$. Given the blocking in P_τ, the corresponding precision matrix can be partitioned in the same way. We obtain Q_τ as

$$Q_\tau = \begin{pmatrix} Q_{\tau,(1,1)} & Q_{\tau,(1,2)} & & & & & & Q_{\tau,(1,T+1)} \\ Q_{\tau,(2,1)} & Q_{\tau,(2,2)} & Q_{\tau,(2,3)} & & & & & Q_{\tau,(2,T+1)} \\ & \ddots & \ddots & \ddots & & & & \vdots \\ & & Q_{\tau,(T-2,T-3)} & Q_{\tau,(T-2,T-2)} & Q_{\tau,(T-2,T-1)} & & & Q_{\tau,(T-2,T+1)} \\ & & & Q_{\tau,(T-1,T-2)} & Q_{\tau,(T-1,T-1)} & Q_{\tau,(T-1,T)} & & Q_{\tau,(T-1,T+1)} \\ & & & & Q_{\tau,(T,T-1)} & Q_{\tau,(T,T)} & & Q_{\tau,(T,T+1)} \\ Q_{\tau,(T+1,1)} & Q_{\tau,(T+1,2)} & \cdots & & Q_{\tau,(T+1,T-1)} & Q_{\tau,(T+1,T)} & & Q_{\tau,(T+1,T+1)} \end{pmatrix}, \tag{72}$$

with $(T+1)\times(T+1)$ blocks. Note that the dimensions of the blocks in Q_τ are time-varying, depending on the number of missing observations in each period, $N_{m,t}$.

The submatrices of Q_τ are given by

$$Q_{\tau,(t,t)} = \begin{pmatrix} Q_{\eta\eta,(t,t)} & Q_{\eta x,(t,t)} P_{x_m,(t,t)}^\mathsf{T} \\ P_{x_m,(t,t)} Q_{x\eta,(t,t)} & P_{x_m,(t,t)} Q_{\epsilon,(t,t)} P_{x_m,(t,t)}^\mathsf{T} \end{pmatrix} \tag{73}$$

for $t = 1, 2, \ldots, T$ on the main diagonal, and

$$Q_{\tau,(t,t+1)} = \begin{pmatrix} Q_{\eta\eta,(t,t+1)} & Q_{\eta x,(t,t+1)} P_{x_m,(t+1,t+1)}^\mathsf{T} \\ P_{x_m,(t,t)} Q_{x\eta,(t,t+1)} & P_{x_m,(t,t)} Q_{\epsilon,(t,t+1)} P_{x_m,(t+1,t+1)}^\mathsf{T} \end{pmatrix}, \tag{74}$$

$$Q_{\tau,(t+1,t)} = \begin{pmatrix} Q_{\eta\eta,(t+1,t)} & Q_{\eta x,(t+1,t)} P_{x_m,(t,t)}^\mathsf{T} \\ P_{x_m,(t+1,t+1)} Q_{x\eta,(t+1,t)} & P_{x_m,(t+1,t+1)} Q_{\epsilon,(t+1,t)} P_{x_m,(t,t)}^\mathsf{T} \end{pmatrix}, \tag{75}$$

for $t = 1, \ldots, T-1$ on the upper and lower block diagonal, respectively. Furthermore, $Q_{\tau,(t_1,t_2)} = 0$ for $|t_1 - t_2| > 1$ and $t_1, t_2 = 1, \ldots, T$. Concerning the $(T+1)$-st row and

column, we obtain

$$Q_{\tau,(T+1,T+1)} = P_{x_o} Q_\epsilon P_{x_o}^\mathsf{T}, \tag{76}$$

$$Q_{\tau,(T+1,t)} = P_{x_o} \begin{pmatrix} Q_{x\eta} P_{\eta,t}^\mathsf{T} & Q_\epsilon P_{x_m,t}^\mathsf{T} \end{pmatrix} \quad \text{for} \quad t = 1, \ldots, T, \tag{77}$$

$$Q_{\tau,(t,T+1)} = \begin{pmatrix} P_{\eta,t} Q_{\eta x} \\ P_{x_m,t} Q_\epsilon \end{pmatrix} P_{x_o}^\mathsf{T} \quad \text{for} \quad t = 1, \ldots, T. \tag{78}$$

To sample from $p(\eta, x_m | x_o = x^o, \theta)$, the precision matrix Q_τ will be partitioned according to

$$Q_\tau = \begin{pmatrix} Q_{\eta x_m, \eta x_m} & Q_{\eta x_m, x_o} \\ Q_{x_o, \eta x_m} & Q_{x_o, x_o} \end{pmatrix}, \tag{79}$$

where the upper-left block $Q_{\eta x_m, \eta x_m}$ contains the first top-left $T \times T$ sub-blocks from (72) and has dimensions $(rT + \sum_{t=1}^T N_{m,t}) \times (rT + \sum_{t=1}^T N_{m,t})$, whereas the lower-right block Q_{x_o,x_o} has dimensions $(NT - \sum_{t=1}^T N_{m,t}) \times (NT - \sum_{t=1}^T N_{m,t})$. The conditional distribution becomes $p(\eta, x_m | x_o = x^o, \theta) \overset{\mathcal{D}}{=} \mathcal{N}(-Q_{\eta x_m, \eta x_m}^{-1} Q_{\eta x_m, x_o} x^o, Q_{\eta x_m, \eta x_m}^{-1})$. From (72), we can see that $Q_{\eta x_m, \eta x_m}$ is block tridiagonal and thus facilitates precision-based sampling efficiently.

References

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis* (third ed.). Wiley Series in Probability and Statistics. Wiley.

Angelini, E., J. Henry, and M. Marcellino (2006). Interpolation and backdating with a large information set. *Journal of Economic Dynamics and Control 30*(12), 2693 – 2724.

Bai, J. and S. Ng (2019). Matrix completion, counterfactuals, and factor analysis of missing data. Working paper, New York University.

Bai, J. and P. Wang (2014). Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics 178*(2), 794 – 804.

Banbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169 – 194.

Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Carvalho, C. M., J. Chang, J. E. Lucas, J. R. Nevins, Q. Wang, and M. West (2008). High-dimensional sparse factor modeling: applications in gene expression genomics. *Journal of the American Statistical Association 103*(484), 1438–1456.

Chan, J., R. Leon-Gonzalez, and R. W. Strachan (2018). Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association 113*(522), 819–828.

Chan, J. C., T. E. Clark, and G. Koop (2018). A new model of inflation, trend inflation, and long-run inflation expectations. *Journal of Money, Credit and Banking 50*(1), 5–53.

Chan, J. C., E. Eisenstat, and R. W. Strachan (2020). Reducing the state space dimension in a large TVP-VAR. *Journal of Econometrics 218*(1), 105 – 118.

Chan, J. C. and A. L. Grant (2016). Fast computation of the deviance information criterion for latent variable models. *Computational Statistics and Data Analysis 100*, 847 – 859.

Chan, J. C. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation 1*(1-2), 101–120.

Chan, J. C. C. (2020). Large bayesian VARs: A flexible kronecker error covariance structure. *Journal of Business & Economic Statistics 38*(1), 68–79.

Chan, J. C. C. and E. Eisenstat (2018). Bayesian model comparison for time-varying parameter vars with stochastic volatility. *Journal of Applied Econometrics 33*(4), 509–532.

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

Chan, J. C. C., G. Koop, and S. M. Potter (2016). A bounded model of time variation in trend inflation, NAIRU and the Phillips curve. *Journal of Applied Econometrics 31*(3), 551–565.

Chib, S. (2001). Chapter 57 - Markov Chain Monte Carlo Methods: Computation and Inference. Volume 5 of *Handbook of Econometrics*, pp. 3569 – 3649. Elsevier.

Conti, G., S. Frühwirth-Schnatter, J. J. Heckman, and R. Piatek (2014). Bayesian exploratory factor analysis. *Journal of Econometrics 183*(1), 31–57.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–616.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015, October). The next generation of the Penn World Table. *American Economic Review 105*(10), 3150–82.

Francis, N., M. T. Owyang, and O. Savascin (2017). An endogenously clustered factor approach to international business cycles. *Journal of Applied Econometrics 32*(7), 1261–1276.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*(2), 183–202.

George, E. and R. McCulloch (1997). Approaches for bayesian variable selection. *Statistica Sinica 7*(2), 339–373.

George, E. I. and R. E. McCulloch (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

Geweke, J. (1996). Variable selection and model comparison in regression. *Bayesian Statistics 5*, 339–373.

Ghosh, J. and D. Dunson (2009). Default prior distributions and efficient posterior computation in bayesian factor analysis. *Journal of Computational and Graphical Statistics 18*, 306 – 320.

Golub, G. and C. Van Loan (2013). *Matrix Computations* (fourth ed.). Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press.

Grant, A. L. and J. C. Chan (2017). Reconciling output gaps: Unobserved components model and Hodrick-Prescott filter. *Journal of Economic Dynamics and Control 75*, 114–121.

Jungbacker, B., S. J. Koopman, and M. van der Wel (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control 35*(8), 1358–1368.

Kastner, G., S. Frühwirth-Schnatter, and H. F. Lopes (2017). Efficient bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics 26*(4), 905–917.

Kaufmann, S. and C. Schumacher (2017). Identifying relevant and irrelevant variables in sparse factor models. *Journal of Applied Econometrics 32*(6), 1123–1144.

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.

Koopman, S. J. and A. Harvey (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control 27*(7), 1317–1333.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003, September). International business cycles: World, region, and country-specific factors. *American Economic Review 93*(4), 1216–1239.

Kose, M. A., C. Otrok, and C. H. Whiteman (2008). Understanding the evolution of world business cycles. *Journal of International Economics 75*(1), 110 – 130.

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

Little, R. J. and D. B. Rubin (2002). *Statistical analysis with missing data* (second ed.). John Wiley & Sons.

Lopes, H. F. and M. West (2004). Bayesian model assessment in factor analysis. *Statistica Sinica 14*(1), 41–67.

Marcellino, M. (2007). Pooling-based data interpolation and backdating. *Journal of Time Series Analysis 28*(1), 53–71.

McCausland, W. J. (2012). The HESSIAN method: Highly efficient simulation smoothing, in a nutshell. *Journal of Econometrics 168*(2), 189–206.

McCausland, W. J. (2015). Dynamic factor models with stochastic volatility. University of Montreal, Department of Economics, mimeo.

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

Müller, U. K., J. H. Stock, and M. W. Watson (2019). An econometric model of international long-run growth dynamics. Working Paper 26593, National Bureau of Economic Research.

Otrok, C. and P. M. Pourpourides (2017). On the cyclicality of real wages and wage differentials. *The B.E. Journal of Macroeconomics 19*(1), 1–18.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

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

Rue, H. and L. Held (2005). *Gaussian Markov random fields: theory and applications*. CRC press.