ISSN 1518-3548

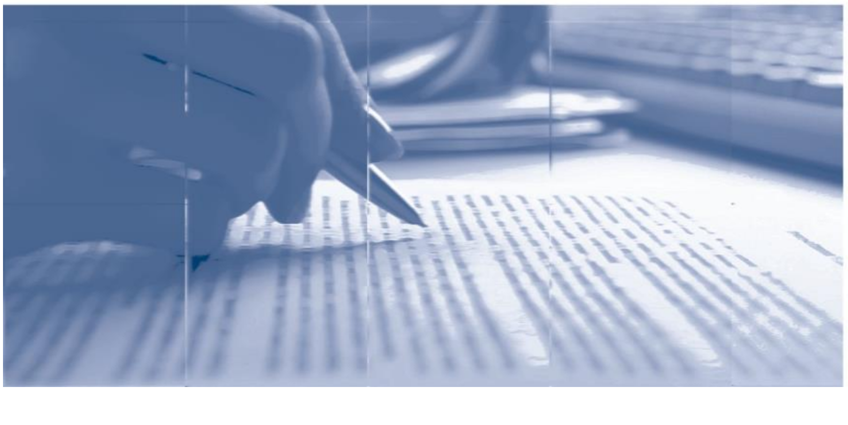
BANCO CENTRAL DO BRASIL

Not Just Another Mixed Frequency Paper

Sergio Afonso Lago Alves and Angelo Marsiglia Fasolo

September, 2015

Working Papers



400

ISSN 1518-3548
CGC 00.038.166/0001-05

Working Paper Series	Brasília	n. 400	September	2015	p. 1-34

Working Paper Series

Edited by Research Department (Depep) – E-mail: workingpaper@bcb.gov.br

Editor: Francisco Marcos Rodrigues Figueiredo – E-mail: francisco-marcos.figueiredo@bcb.gov.br
Editorial Assistant: Jane Sofia Moita – E-mail: jane.sofia@bcb.gov.br
Head of Research Department: Eduardo José Araújo Lima – E-mail: eduardo.lima@bcb.gov.br

The Banco Central do Brasil Working Papers are all evaluated in double blind referee process.

Authorized by Luiz Awazu Pereira da Silva, Deputy Governor for Economic Policy.

General Control of Publications

Banco Central do Brasil

Comun/Dipiv/Coivi

SBS – Quadra 3 – Bloco B – Edifício-Sede – 14º andar

Caixa Postal 8.670

70074-900 Brasília – DF – Brazil

Phones: +55 (61) 3414-3710 and 3414-3565

Fax: +55 (61) 3414-1898

E-mail: editor@bcb.gov.br

Citizen Service Division

Banco Central do Brasil

Deati/Diate

SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo

70074-900 Brasília – DF – Brazil

Toll Free: 0800 9792345

Fax: +55 (61) 3414-2553

Internet: <http//www.bcb.gov.br/?CONTACTUS>

Not Just Another Mixed Frequency Paper*

Sergio Afonso Lago Alves†

Angelo Marsiglia Fasolo‡

Abstract

This paper presents a new algorithm, based on a two-part Gibbs sampler with FFBS method, to recover the joint distribution of missing observations in a mixed-frequency dataset. The new algorithm relaxes most of the constraints usually presented in the literature, namely: (i) it does not require at least one time series to be observed every period; (ii) it provides an easy way to add linear restrictions based on the state space representation of the VAR; (iii) it does not require regularly-spaced time series at lower frequencies; and, (iv) it avoids degeneration problems arising when states, or linear combination of states, are actually observed. In addition, the algorithm is well suited for embedding high-frequency real-time information for improving nowcasts and forecasts of lower frequency time series. We evaluate the properties of the algorithm using simulated data. Moreover, as empirical applications, we simulate monthly Brazilian GDP, comparing our results to the Brazilian IBC-BR, and recover what would historical PNAD-C unemployment rates look like prior to 2012.

Keywords: Gibbs Sampling; Forward Filtering Backward Sampling (FFBS); Mixed Frequency Data; Bayesian methods; Real-time data; Macroeconomic forecasting; Vector autoregressions; PNAD.

JEL Classification: C11, C32, C53, E27

*The authors would like to thank Hernán D. Seoane and Andrew Foerster, for early comments on the main ideas of this paper, and the participants of seminars at the Central Bank of Brazil. All errors and omissions, however, are our own fault.

†Research Department, Banco Central do Brasil. Email: sergio.lago@bcb.gov.br

‡Research Department, Banco Central do Brasil. Email: angelo.fasolo@bcb.gov.br

1 Introduction

Many economic time series of developing and underdeveloped countries suffer from issues preventing their information content to be fully explored. These common issues usually arise from: (*i*) *short length*; (*ii*) measurement only available in *low frequencies*, such as quarterly or annual; (*iii*) *discontinuity* due to changes in measurement methodology into another; (*iv*) measurement or technical issues leading to *randomly spaced missing observations* in time series; and (*v*) *frequency switch* when the frequency of observations changes along the sample, e.g. series starts as annual, then change to quarterly after a couple of years.

Short length is a common issue that might affect developed countries as well, but is more often present in developing countries. In a nutshell, the problem arises when a time series either ceases to be measured, or has only recently started to be measured. Understanding the inference limitations caused by low frequency time series is straightforward due to the reduced number of observations available. The discontinuity problem arises when a time series has its measurement method changed while not updating the method for past observations. In this case, data properties might be significantly affected due to the new measurement method, in such a way that observations obtained from the old and new method must actually be considered as two different time series.

Randomly spaced missing observations may arise by either measurement flaws when collecting and treating raw data, or by deliberately excluding observations that are considered outliers or not consistent with the effect of hypotheses one wants to measure and test. Finally, series with frequency switches may be seen as a special case of series with missing observations, except that spacing between observations is set from a discretionary choice. For simplicity, we treat this case just like we treat randomly spaced missing observations.[1] In the context of multivariate analysis, a problematic situation occurs when a combination of the issues listed above leads to what is known as *unequally spaced mixed-frequency* data sets, where the extreme case shows periods with no available observations at all, i.e. all elements from the time series might be missing in given time.

Under mixed-frequency, multivariate framework, usual choices might compromise the statistical inference, leading to wrong conclusions. For instance, the analyst might either

[1] See Prado and West (2010, Section 4.3.3) for a better discussion on series with discretionary spaced missing observations.

fill the gaps in time series with ad-hoc values, i.e. mechanical approaches without formal statistical support, such as interpolations or trimming out observations in periods in which one or more time series have missing observations. Alternatively, one might consider reducing the time series frequency of the whole dataset in order to avoid dealing with mixed-frequency issues. Both trimming out periods or reducing frequencies might lead to the loss of relevant information, though. Hence, a problem arises in finding ways to simultaneously explore the full information content available in the mixed-frequency dataset, while avoiding the use of ad-hoc methods to fill information gaps.

In economics, the literature has evolved into the development of Bayesian Gibbs samplers to recover the entire joint distribution of the missing observations, rather than pin-pointing single values to fill the gaps. In this paper, we contribute to the literature by bringing an efficient and intuitive way to recover this joint distribution using a two-part Gibbs sampler, built mostly upon recent important contributions to the literature on mixed-frequency filtering by Schorfheide and Song (2014) and Eraker et al. (2014), and, from the statistics literature, on the efficient algorithm described in West (1996, 1997) and West and Harrison (1997, Section 15.3.2).

We highlight the point that the literature in statistics has considered the problem of recovering the entire joint distribution of missing observations way before the literature in economics, with efficient algorithms using Gibbs samplers, as the one developed by West and detailed further on, developed during the 1990's. Therefore, the main point of this paper is to adapt and bring some of the early developments from the statistics literature to applications in economics. That been said, our view is that the method proposed by West (1996, 1997) is more efficient, simpler and more intuitive than the ones proposed by Schorfheide and Song (2014) and Eraker et al. (2014), even though generating about the same results.

As in Schorfheide and Song (2014) and Eraker et al. (2014), we assume that a VAR with unknown parameters is able to describe the dynamics of the multivariate time series, if their missing values were all observed. The task of the Gibbs sampler is to recover the joint distribution of the missing values and the unknown VAR parameters. As in West (1996, 1997) and West and Harrison (1997, Section 15.3.2), our strategy for sampling time series is using a data augmentation approach based on Carter and Kohn (1994) and Fruhwirth-Schnatter (1994) — Fruhwirth-Schnatter (1994) calls this approach *forward* *filtering, backward sampling* (FFBS) algorithm, also used in Schorfheide and Song (2014). This is one step of the Gibbs sampler, which takes sampled parameters and covariance matrix from a previous step of the Gibbs sampler, based on a Bayesian VAR with a modified Minnesota-type prior (see e.g. Doan et al. (1984) and Litterman (1986) for the original Minnesota prior).

In its general form, the FFBS algorithm is straightforward to implement. An issue arises, however, in models with AR components, or any other model in which consecutive state vectors contain common components. In this case, linear combinations of elements from current state are actually observed, and the basic FFBS algorithm degenerates. From this point, we depart from Schorfheide and Song (2014) by proposing an algorithm intended to perform FFBS for a VAR case, avoiding degeneration issues. The algorithm is based on West's method on efficiently sampling state vectors in AR models, adapted to our VAR case. We also depart from Schorfheide and Song (2014) by allowing for irregularities in the frequency of observations, generalizing the shape of the selection matrix introduced by the authors for regularly-spaced mixed-frequency time series.

Eraker et al. (2014) also work on unevenly-spaced mixed-frequency time-series, but do not benefit from using a state-space specification, which allows for additional linear restrictions on the relation between observed time series and latent variables. Normally, dealing with missing values in a state space framework, as the one used in our Gibbs step for sampling missing observations conditional on the parameter set, is relatively easy (see e.g. Prado and West (2010, Section 4.3.3) and Durbin and Koopman (2012, Section 4.10)). The approach is flexible enough for not requiring the existence of at least one time series to be observed every period. Indeed, it works even when there are no observations at particular periods. Embedding this approach, we also depart from Schorfheide and Song (2014) and Eraker et al. (2014), for they require at least one completely observed time series. Therefore, our approach benefits from the best features of both Schorfheide and Song (2014) and Eraker et al. (2014) procedures.

The flexibility of not requiring the existence of at least one time series to be observed every period comes extremely in handy when performing forecasts and backcasts, as the algorithm only has to consider these exercises as an extended sample for which there is no observation at all. In other words, the FFBS algorithm samples missing values in forecast and backcast exercises. Therefore, our approach is efficient, intuitive and well suited for dealing with all four common issues previously described, i.e. short length, low frequencies, discontinuity, randomly spaced missing observations, and multivariate time series with periods in which no observations are available.

The algorithm is also well suited for dealing with high-frequency real-time information, in order to improve nowcasts and forecasts of lower frequency time series. An empirical exercise in Section 5 shows the improvement in quarterly Brazilian GDP's nowcasts as real-time high-frequency information arrives, and compares the results to the Brazilian IBC-BR. The exercise also highlights the role of a proper setup of the state space representation of the model, in order to establish the relationship between monthly and quarterly observations. In another exercise, we use the algorithm to recover the joint distribution of historical PNAD-C Brazilian unemployment rates prior to 2012, using information from other surveys such as Annual PNAD and PME. The importance of the latter exercise is due to the fact that PNAD-C is a new monthly time series, whose values are available only from 2012 on.

The remainder of the paper is organized as follows. The model is described in Section 2. The Gibbs sampler, and details on how to recover the full joint distribution of the missing observations are described in Section 3. Details on the efficient FFBS algorithm and Bayesian VAR are described in sections 3.1 and 3.2. We test the performance of our approach with simulated data sets in Section 4. Two empirical exercises are shown in Section 5, based on information of Brazilian GDP and unemployment. The empirical exercises explore both the issue of regularity of missing information and use of linear restrictions in the model. Section 6 summarizes the paper's conclusions.

2 The model

We assume the existence of a sample $\mathbf{Y}^{ob} \equiv \left\{y_1^{ob}, ..., y_T^{ob}\right\}$ of size T. At each period, y_t^{ob} is a $\left(m_t^{ob} \times 1\right)$ vector of observed endogenous variables, whose dimension m_t^{ob} changes over time due to irregularly-spaced mixed-frequency. This definition of m_t^{ob} does not exclude even the possibility that, at certain periods, no observations are available, i.e. $m_t^{ob} = 0$.

In this context, define first y_t as a $(m \times 1)$ vector of endogenous variables of interest, whose dimension $m \geq m_t^{ob}$ ($m > m_t^{ob}$, if $m_t^{ob} = 0$) is time-invariant. Notice that it is possible that the set of endogenous variables of interest might not be directly observed at all in the sample. For instance, y_t^{ob} might contain moving averages of y_t, which are affine transformations of y_t and its lagged values. In this case, y_t^{ob} is not even a subset of y_t. The relation between y_t^{ob} and y_t is formally addressed in Section 3.1.

Our objective is then to infer the joint distribution of the $(T \times m)$ sequence $\mathbf{Y} \equiv [y_1, ..., y_T]'$ of endogenous variables of interest. To fulfill this task, assume that the dynamics of y_t can be represented by a $VAR_m(q)$:

$$y_t = \Phi_1 y_{t-1} + ... + \Phi_q y_{t-q} + \bar{\Phi}_c \bar{x}_{ct} + e_t \quad , \; e_t \overset{iid}{\sim} N(0, \Sigma_e) \tag{1}$$

where $\bar{x}_{ct}$ is a $(m_c \times 1)$ matrix of deterministic variables, such as the constant, trend and seasonal dummies, Φ_ℓ is a $(m \times m)$ matrix of coefficients for $\ell \in \{1, ..., q\}$, $\bar{\Phi}_c$ is a $(m \times m_c)$ matrix of coefficients, and Σ_e is a $(m \times m)$ positive definite covariance matrix. Each equation has $k \equiv mq + m_c$ regressors, the VAR has mk coefficients, and the whole system has $mk + m^2$ parameters.

$$\begin{aligned}
y_t &= \begin{bmatrix} y_{1t} & \cdots & y_{mt} \end{bmatrix}' &\qquad \Phi_\ell &= \begin{bmatrix} \Phi_{\ell;11} & \cdots & \Phi_{\ell;m1} \\ \vdots & & \vdots \\ \Phi_{\ell;1m} & \cdots & \Phi_{\ell;mm} \end{bmatrix} \\
e_t &= \begin{bmatrix} e_{1t} & \cdots & e_{mt} \end{bmatrix}' & & \\
\bar{x}_{ct} &= \begin{bmatrix} \bar{x}_{c;1t} & \cdots & \bar{x}_{c;m_c t} \end{bmatrix}' &\qquad \bar{\Phi}_c &= \begin{bmatrix} \bar{\Phi}_{c;11} & \cdots & \bar{\Phi}_{c;m1} \\ \vdots & & \vdots \\ \bar{\Phi}_{c;1m_c} & \cdots & \bar{\Phi}_{c;mm_c} \end{bmatrix}
\end{aligned}$$

For analytical convenience, each part of the Gibbs sampler, described in Section 3, considers different companion forms when dealing with the linear system. More details are provided in Sections 3.1 and 3.2.

3 Gibbs sampling

This section describes the steps of a Gibbs algorithm to draw a sample of size S, given the information set described by $\mathbf{Y}^{ob}$, from the joint distribution of $\left(\mathbf{Y}, \mathbf{\Phi} | \mathbf{Y}^{ob}\right)$. Each step of the sampler consists of two parts: (i) sample draws of $\mathbf{Y}$ from the conditional distribution of $\left(\mathbf{Y} | \mathbf{Y}^{ob}, \mathbf{\Phi}\right)$, using Kalman filtering and a modified form of the *forward filtering, backward sampling* (FFBS) algorithm, as described in Section 3.1; and (ii) sample draws of $\mathbf{\Phi}$ from the conditional distribution of $\left(\mathbf{\Phi}|\mathbf{Y}^{ob},\mathbf{Y}\right)$, using a Bayesian VAR, as described in Section 3.2.

The sampler is initialized, in part (i) of step 0, by setting $\mathbf{\Phi}_{init}$ to imply a random walk dynamics to $\mathbf{Y}$, i.e. $\Phi_\ell = \frac{1}{q}\mathbf{I}_{(m\times m)}$, for $\ell \in \{1,...,q\}$, and $\bar{\Phi}_c = 0$. The FFBS algorithm of part (i) generates a smoothed path of $\mathbf{Y}_0$ from $\left(\mathbf{Y}|\mathbf{Y}^{ob},\mathbf{\Phi}_{init}\right)$ to be used in part (ii), where a Bayesian VAR generates the initial parameter set $\mathbf{\Phi}_0$ from $\left(\mathbf{\Phi}|\mathbf{Y}^{ob},\mathbf{Y}_0\right)$, still at step 0. From steps $s=1$ to $s=S$, the Gibbs algorithm evolves as usual: (i) draw $\mathbf{Y}_s$ from $\left(\mathbf{Y}|\mathbf{Y}^{ob},\mathbf{\Phi}_{s-1}\right)$; and (ii) draw $\mathbf{\Phi}_s$ from $\left(\mathbf{\Phi}|\mathbf{Y}^{ob},\mathbf{Y}_s\right)$.

3.1 The dynamic linear model

In what follows, we characterize the $VAR_m\left(q\right)$ system (1) as a dynamic linear model (DLM), as in Schorfheide and Song (2014), and describe the sequential updating, i.e. Kalman (1960) filtering, equations.[2] A companion form of the $VAR_m\left(q\right)$ system is:

$$\begin{array}{llll} y_t^{ob} & \overset{\substack{Observation\\Equation}}{=} F_t z_t & ; \; F_t & \equiv M_t\Lambda_z \\ z_t & \overset{\substack{Transition\\Equation}}{=} G_1 z_{t-1}+\omega_t & ; \; \omega_t & \overset{ind}{\sim} N\left(G_2\bar{x}_{ct},\Omega_e\right) \end{array} \tag{2}$$

where y_t^{ob} is a $\left(m_t^{ob}\times 1\right)$ vector of observed endogenous variables, whose dimension m_t^{ob} changes due to mixed-frequency or irregularly-spaced observations, M_t and Λ_z are generalizations of what was done in in Schorfheide and Song (2014), i.e. M_t is a $\left(m_t^{ob}\times m\right)$ *selection matrix* and Λ_z is a $(m\times m\bar{q})$ *transformation matrix*,[3] useful when the observed variable is a known affine transformation of the states (e.g. 3-month moving averages, or any other linear combinations), $z_t\equiv\left[y_t',...,y_{t-\bar{q}+1}'\right]'$ is a $(m\bar{q}\times 1)$ vector of states, G_1 is a $(m\bar{q}\times m\bar{q})$ matrix of coefficients for endogenous variables, G_2 is a $(m\bar{q}\times m_c)$ matrix of coefficients for exogenous variables, and Ω_e is a $(m\bar{q}\times m\bar{q})$ positive semi-definite

[2] Great references on dynamic linear models (DLM) and inference using the Kalman filter are Hamilton (1994), Prado and West (2010) and West and Harrison (1997). For DLMs applications in macroeconomic models, see Basdevant (2003).

[3] In this regard, Λ_z should be designed with care in order to avoid *aliasing* issues when we only observe averages, or any other linear combination, of latent variables of interest. For instance, suppose that we observe three quarterly time series $y_t = [y_{1t},y_{2t},y_{3t}]'$, which actually aggregates or averages latent monthly variables during each quarter, such as sectoral GDP's. In this case, there are an infinite set of possible monthly sectoral GDP time series that are consistent with the observed variables. Those sets include monthly time series with unreasonable extremely large variances. In this case, the algorithm will not converge to a unique stationary distribution.

covariance matrix:

$$G_1 = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{\bar{q}-1} & \Phi_{\bar{q}} \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \quad G_2 = \begin{bmatrix} \bar{\Phi}_c \\ 0_{m \times m_c} \\ 0_{m \times m_c} \\ \vdots \\ 0_{m \times m_c} \end{bmatrix} \quad \Omega_e = \begin{bmatrix} \Sigma_e & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Note that $\bar{q}$ does not necessarily matches q, the number of lags in the $VAR_m(q)$ describing the dynamics of y_t. For instance, suppose that y_t^{ob} represents quarterly averages ($\bar{q} = 3$) of monthly non-observed values of y_t, whose dynamics is described by a VAR with four lags ($q = 4$). Conversely, if the dynamics is better described by a $VAR_m(1)$, a similar issue arises. Those cases are easily addressed as follows: (i) if $\bar{q} < q$, set $\bar{q} = q$ and extend Λ_z with zero matrices, i.e. $\Lambda_z = \left[\Lambda_z, 0_{m \times m(q-\bar{q})}\right]$; (ii) if $\bar{q} > q$, consider additional zero matrices $\Phi_{\bar{\ell}} = 0_{m \times m}$, for $\bar{\ell} \in \{q+1, ..., \bar{q}\}$.

In this context, M_t is a time-varying selection matrix where the number of rows is adjusted to match the number of absent observations in each period. From this perspective, starting from an identity matrix M with size $(m \times m)$, M_t is set in each period by eliminating $m - m_t^{ob}$ rows of M, corresponding to non-observed variables of vector y_t^{ob}.

Let $\mathcal{D}_t \equiv \left\{\mathcal{D}_{t-1}, y_t^{ob}\right\}$ denote the information set at each period t. If there are no observed values at period t, i.e. y_t^{ob} is an empty vector, then M_t is an empty matrix and $\mathcal{D}_t = \mathcal{D}_{t-1}$.[4]

Consider the following initial prior Gaussian density for z_0 is $(z_0|\mathcal{D}_0) \sim N(\mu_0, C_0)$. Therefore, the Kalman filtering (KF) equations are:

a) Prior density for z_t with information up to period $(t-1)$:

$$(z_t|\mathcal{D}_{t-1}) \sim N(a_t, R_t)$$
$$a_t \equiv G_1\mu_{t-1} + G_2\bar{x}_{ct} \quad ; \; R_t \equiv G_1 C_{t-1} G_1' + \Omega_e$$

[4] A better discution on models with missing and unequally spaced data can be found in e.g. Prado and West (2010, Section 4.3.3) and Durbin and Koopman (2012, Section 4.10).

b) One-step forecast of y_t^{ob} with information up to period $(t-1)$:

$$\left(y_t^{ob}|\mathcal{D}_{t-1}\right) \sim N\left(f_t, Q_t\right)$$
$$f_t \equiv F_t a_t \quad ; \; Q_t \equiv F_t R_t F_t'$$

c) Posterior density for z_t with information up to period t:

$$\left(z_t|\mathcal{D}_t\right) \sim N\left(\mu_t, C_t\right)$$

If there are observations at period t:

$\mu_t \equiv a_t + A_t v_t \quad ; \; C_t \equiv R_t - A_t Q_t A_t' \quad ; \; v_t \equiv y_t^{ob} - f_t \quad ; \; A_t \equiv R_t F_t' Q_t^{-1}$

Otherwise:

$\mathcal{D}_t = \mathcal{D}_{t-1} \quad ; \; \left(z_t|\mathcal{D}_t\right) \equiv \left(z_t|\mathcal{D}_{t-1}\right) \quad ; \; \mu_t \equiv a_t \quad ; \; C_t \equiv R_t$

Our strategy for completing the series with mixed-frequency or irregularly-spaced observations is to use a data augmentation approach based on Carter and Kohn (1994) and Fruhwirth-Schnatter (1994) to sample complete sequences of state variables. We also follow Fruhwirth-Schnatter (1994) by calling this approach as *forward filtering, backward sampling* (FFBS) algorithm. Let $Z_T \equiv [z_T, z_{T-1}, ..., z_1]$ denote the whole sequence of state vectors. The nature of the FFBS algorithm comes from exploring the Markov structure of the evolution equation of any DLM, which allows us to write

$$\Pr\left(Z_T|\mathcal{D}_T\right) = \Pr\left(z_1|z_2, \mathcal{D}_1\right) ... \Pr\left(z_t|z_{t+1}, \mathcal{D}_t\right) ... \Pr\left(z_{T-1}|z_T, \mathcal{D}_{T-1}\right) \Pr\left(z_T|\mathcal{D}_T\right)$$

The proof is shown in the Appendix.

Therefore, the FFBS algorithm consists of sampling z_T from $\Pr\left(z_T|\mathcal{D}_T\right)$, and sequentially sampling z_t from $\Pr\left(z_t|z_{t+1}, \mathcal{D}_t\right)$ using equations directly derived from the Kalman filter.

In a recent paper, similar to ours, Schorfheide and Song (2014) also use a data augmentation approach based on Carter and Kohn (1994). We depart from Schorfheide and Song (2014) by allowing for irregularities in the frequency of observations, instead of simple, regularly-spaced mixed-frequency time-series. This extension, also carried out in Eraker et al. (2014), is obtained by generalizing the shape of M_t matrix. Eraker et al. (2014), on the other hand, do not benefit from using a state-apace specification, which allows for additional linear restrictions on the relation between observables and states, captured by matrix Λ_z. Therefore, our approach benefits from the best features of both procedures.

We also depart from Schorfheide and Song (2014) and Eraker et al. (2014) by not requiring the existence of at least one series whose values are observed in all periods. In this regard, our approach is also well suited for the following cases: (i) data sets from countries in which series are relatively short; (ii) new series, or series whose measuring method changes without applying to previous periods.

In its general form, the FFBS algorithm is very easy to implement and requires inverting matrix R_t, i.e. the covariance matrix of $(z_t|\mathcal{D}_{t-1})$.[5] An issue arises, however, in models with AR components, or any other model in which consecutive state vectors contain common components. In this case, linear combinations of elements from current state are actually observed. This fact implies that there will be many instances in which R_t is singular, and the basic FFBS algorithm degenerates.

Therefore, we also depart from those authors by proposing an algorithm intended to perform FFBS for a VAR case, avoiding degeneration issues. It is based on West (1996, 1997) and West and Harrison (1997, Section 15.3.2) on efficiently sampling state vectors in AR models. We adapt West's approach to our VAR case, detailed below.

3.1.1 Adapted FFBS Algorithm

The idea behind West's approach comes from the structure depicted in eq. (3). Note that, given a previously sampled vector z_{t+1}, the backward sampling procedure implies that there is only one element of vector z_t yet to be sampled: the $(m \times 1)$ vector y_{t-q+1}. All we have to do is to sample values from the conditional distribution of $(y_{t-\bar{q}+1}|z_{t+1}, \mathcal{D}_t)$ and complete vector z_t using the relevant elements from vector z_{t+1}. The case in which $\bar{q} = 1$, which does not fit the depicted structure, is easily addressed by imposing $\bar{q} = 2$ and defining $\Phi_2 = 0_{(m\times m)}$.

$$
\begin{array}{rcccccl}
z'_{t+1} = [\ y'_{t+1} & , y'_t & , y'_{t-1} & , \cdots & , y'_{t-\bar{q}+2} & &] \\
 & \updownarrow & \updownarrow & \updownarrow & \updownarrow & & \\
z'_t = & [\ y'_t & , y'_{t-1} & , \cdots & , y'_{t-\bar{q}+2} & , y'_{t-\bar{q}+1} &]
\end{array}
\tag{3}
$$

[5] Indeed, for each $t \in \{(T-1), (T-2), ..., 1\}$ and having already sampled z_{t+1}, we would obtain the conditional distribution $(z_t|z_{t+1}, \mathcal{D}_t) \sim N(h_t, H_t)$, where $h_t = \mu_t + B_t(z_{t+1} - a_{t+1})$, $H_t = C_t - B_t R_{t+1} B'_t$ and $B_t \equiv C_t G'_1 R^{-1}_{t+1}$.

Adapting West's modification of the FFBS algorithm for sampling z_t in our case is straightforward. First, let $\tilde{z}_t \equiv \left[y_t', ..., y_{t-\bar{q}+2}'\right]'$ denote the already sampled $(m(\bar{q}-1) \times 1)$ vector of common elements of z_{t+1} and z_t, i.e. we partition both vectors as $z_{t+1} = \left[y_{t+1}', \tilde{z}_t'\right]'$ and $z_t = \left[\tilde{z}_t', y_{t-\bar{q}+1}'\right]'$. In this context, let $\tilde{z}_{j,t}$ denote each singular element of $\tilde{z}_t$ for $j \in \{1, ..., m(\bar{q}-1)\}$.

For analytical simplicity, consider the sub-vector $\mathfrak{z}_{j,t} \equiv \left[\tilde{z}_{j:m(\bar{q}-1),t}', y_{t-\bar{q}+1}'\right]'$ where $\tilde{z}_{j_1:j_2,t} \equiv [\tilde{z}_{j_1,t}, ...\tilde{z}_{j_2,t}]'$ for $j_2 \geq j_1$ and $\tilde{z}_{j_1:j_2,t} \equiv empty$ for $j_2 < j_1$,[6] and the following partitions, for $j \leq i$:

$$\mathfrak{z}_{j,t} = \left[\begin{array}{c} \tilde{z}_{j,t} \\ \hline \mathfrak{z}_{j+1,t} \end{array}\right] \quad, \ \mu_{j,j,t} = \left[\begin{array}{c} \tilde{\mu}_{j,t} \\ \hline \mu_{j,j+1,t} \end{array}\right] \quad, \ C_{j,j,t} = \left[\begin{array}{c|c} \tilde{\sigma}_{j,t}^2 & \Sigma_{j,j+1,t}' \\ \hline \Sigma_{j,j+1,t} & C_{j,j+1,t} \end{array}\right]$$

where $\mu_{j,i,t} \equiv E\left(\mathfrak{z}_{i,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t\right)$, $C_{j,i,t} \equiv Var\left(\mathfrak{z}_{i,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t\right)$, $\tilde{\mu}_{j,t} \equiv E\left(\tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t\right)$, $\tilde{\sigma}_{j,t}^2 \equiv Var\left(\tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t\right)$ and $\Sigma_{j,i,t} \equiv Cov\left(\mathfrak{z}_{i,t}, \tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t\right)$. The partitions imply that $\tilde{z}_{j,t}$, $\tilde{\mu}_{j,t}$ and $\tilde{\sigma}_{j,t}^2$ are scalars.

The following steps describe the sampling strategy:

1. Sample a value of z_T from $(z_T | \mathcal{D}_T) \sim N(\mu_T, C_T)$;

2. For each $t \in \{(T-1), (T-2), ..., 1\}$, sequentially sample values from the conditional distribution of $(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t)$ and complete vector z_t using the relevant elements from vector z_{t+1}. This is achieved by obtaining the distributions of $(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t)$ and $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$. Also note, as shown in the Appendix, that $\Pr(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \Pr(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$. Since $z_t = \left[\tilde{z}_t', y_{t-\bar{q}+1}'\right]'$, the first choice to retrieve $\Pr(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ would be using the KF distribution $(z_t | \mathcal{D}_t) \sim N(\mu_t, C_t)$ to directly obtain the conditional distribution of $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$. This method, however, involves inverting $Var(\tilde{z}_t | \mathcal{D}_t)$ which is often singular.[7] Therefore, based on West's method, we efficiently obtain the conditional distribution $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ by sequentially conditioning on each scalar element $\tilde{z}_{j,t}$ of $\tilde{z}_t$, reducing the dimension of the distribution by 1 at each stage. The steps are described below:

   (a) Since $\mathfrak{z}_{1,t} = z_t$, note that $\mu_{1,1,t} = \mu_t$ and $C_{1,1,t} = C_t$.

[6] Note that $\mathfrak{z}_{1,t} = z_t$ and $\mathfrak{z}_{(m(\bar{q}-1)+1),t} = y_{t-\bar{q}+1}$.

[7] The same reasons why R_t is eventually singular might cause $Var(\tilde{z}_t | \mathcal{D}_t)$ to be singular as well.

(b) For $j \in \{1, ..., m(\bar{q}-1)\}$, compute the distribution of $\left(\mathfrak{z}_{(j+1),t} | \tilde{z}_{1:j,t}, \mathcal{D}_t\right)$:[8]

$$\left(\mathfrak{z}_{(j+1),t} | \tilde{z}_{1:j,t}, \mathcal{D}_t\right) \sim N\left(\mu_{(j+1),(j+1),t}, C_{(j+1),(j+1),t}\right)$$

$$\mu_{(j+1),(j+1),t} = \begin{cases} \mu_{j,j+1,t} + \frac{1}{\tilde{\sigma}^2_{j,t}} \Sigma_{j,j+1,t} \left(\tilde{z}_{j,t} - \tilde{\mu}_{j,t}\right) & \text{, if } \tilde{\sigma}^2_{j,t} > 0 \\ \mu_{j,j+1,t} & \text{, if } \tilde{\sigma}^2_{j,t} = 0 \end{cases}$$

$$C_{(j+1),(j+1),t} = \begin{cases} C_{j,j+1,t} - \frac{1}{\tilde{\sigma}^2_{j,t}} \Sigma_{j,j+1,t} \Sigma'_{j,j+1,t} & \text{, if } \tilde{\sigma}^2_{j,t} > 0 \\ C_{j,j+1,t} & \text{, if } \tilde{\sigma}^2_{j,t} = 0 \end{cases}$$

(c) Since $\mathfrak{z}_{(m(\bar{q}-1)+1),t} = y_{t-\bar{q}+1}$, $\left(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t\right)$ is distributed as $\left(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t\right) \sim N\left(\tilde{\mu}_{t-\bar{q}+1}, \tilde{C}_{t-\bar{q}+1}\right)$, where

$$\tilde{\mu}_{t-\bar{q}+1} = \mu_{(m(\bar{q}-1)+1),(m(\bar{q}-1)+1),t} \quad , \quad \tilde{C}_{t-\bar{q}+1} = C_{(m(\bar{q}-1)+1),(m(\bar{q}-1)+1),t}$$

3. Based on the already sampled vector $\tilde{z}_{t+1}$ and exogenous vector $\bar{x}_{ct+1}$, compute the partial residual

$$\tilde{e}_{t+1} \equiv y_{t+1} - \left(\Phi_1 y_t + ... + \Phi_{\bar{q}-1} y_{t-\bar{q}+2} + \bar{\Phi}_c \bar{x}_{ct+1}\right)$$

Note that $\Pr\left(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t\right) = \Pr\left(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t\right)$, whose conditional distribution is $\left(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t\right) \sim N\left(\Phi_{\bar{q}} y_{t-\bar{q}+1}, \Sigma_e\right)$.

4. Note that $\Pr\left(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t\right) \propto \Pr\left(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t\right) \Pr\left(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t\right)$, as the Appendix shows. If $\tilde{C}_{t-\bar{q}+1}$ is not singular, all we have to do is to sample y_{t-q+1} from

$$\left(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t\right) \sim N\left(\mu^y_{t-q+1}, C^y_{t-q+1}\right) \tag{4}$$

where

$$\mu^y_{t-q+1} \equiv \left(\Phi'_{\bar{q}} \Sigma_e^{-1} \Phi_{\bar{q}} + \left(\tilde{C}_{t-\bar{q}+1}\right)^{-1}\right)^{-1} \left(\Phi'_{\bar{q}} \Sigma_e^{-1} \tilde{e}_{t+1} + \left(\tilde{C}_{t-\bar{q}+1}\right)^{-1} \tilde{\mu}_{t-\bar{q}+1}\right)$$

$$C^y_{t-q+1} \equiv \left(\Phi'_{\bar{q}} \Sigma_e^{-1} \Phi_{\bar{q}} + \left(\tilde{C}_{t-\bar{q}+1}\right)^{-1}\right)^{-1}$$

[8] If $\tilde{\sigma}^2_{j,t} = 0$, then $\Sigma_{j,j+1,t} = 0$ and $\tilde{z}_{j,t} = \tilde{\mu}_{j,t}$. In this case, the limiting value of $\frac{1}{\tilde{\sigma}^2_{j,t}} \Sigma_{j,j+1,t} \left(\tilde{z}_{j,t} - \tilde{\mu}_{j,t}\right)$ and $\frac{1}{\tilde{\sigma}^2_{j,t}} \Sigma_{j,j+1,t} \Sigma'_{j,j+1,t}$ are both zero. For coding purposes, there still might be some numerical instability in cases where $\tilde{\sigma}^2_{j,t}$ is positive, but very small. In this case, the restriction $\tilde{\sigma}^2_{j,t} > 0$ might be translated into something as, for instance, $\tilde{\sigma}^2_{j,t} > Toler \cdot \tilde{\mu}_{j,t}$, where $Toler$ is a relative tolerance parameter.

Up to this point, we closely followed West's suggestions. There might be, however, an issue when sampling $(y_{t-\bar{q}+1}|z_{t+1},\mathcal{D}_t)$ based on the last result, i.e. there might be cases in which $\tilde{C}_{t-\bar{q}+1}$ is singular. In this case, we propose a procedure similar to what was done in (2.b.). Our suggestion is to sequentially sample each scalar element $y_{j,t-q+1}$ of y_{t-q+1}, for $j \in \{1,...,m\}$, reducing the dimension of the distribution $(y_{t-\bar{q}+1}|\tilde{z}_t,\mathcal{D}_t)$ by 1 at each stage. For that, consider the sub-vectors $\check{y}_{j,t} \equiv y_{j:m,t-q+1}$ and $\hat{y}_{j,t} \equiv y_{1:j,t-q+1}$ the conditional distribution $\Pr\left(\tilde{y}_{j,t}|\hat{y}_{(j-1)},z_{t+1},\mathcal{D}_t\right) \propto \Pr\left(\tilde{e}_{t+1}|\tilde{z}_t,y_{t-\bar{q}+1},\mathcal{D}_t\right)\Pr\left(\tilde{y}_{j,t}|\hat{y}_{(j-1)},\tilde{z}_t,\mathcal{D}_t\right)$, where $\tilde{y}_{j,t} \equiv y_{j,t-q+1}$ is a simplifying notation, and the following partitions

$$\check{y}_{j,t} = \left[\begin{array}{c} y_{j,t-q+1} \\ \hline \check{y}_{(j+1),t} \end{array}\right] \quad , \ \check{\mu}_{j,j,t} = \left[\begin{array}{c} \mu_{j,t-q+1} \\ \hline \check{\mu}_{j,j+1,t} \end{array}\right] \quad , \ \check{C}_{j,j,t} = \left[\begin{array}{c|c} \check{\sigma}^2_{j,t} & \check{\Sigma}'_{j,j+1,t} \\ \hline \check{\Sigma}_{j,j+1,t} & \check{C}_{j,j+1,t} \end{array}\right]$$

where the means are $\check{\mu}_{j,i,t} \equiv E\left(\check{y}_{i,t}|\hat{y}_{(j-1)},\tilde{z}_t,\mathcal{D}_t\right)$ and $\mu_{j,t-q+1} \equiv E\left(\tilde{y}_{j,t}|\hat{y}_{(j-1)},\tilde{z}_t,\mathcal{D}_t\right)$, while the variances are $\check{\sigma}^2_{j,t} \equiv Var\left(\tilde{y}_{j,t}|\hat{y}_{(j-1)},\tilde{z}_t,\mathcal{D}_t\right)$, $\check{C}_{j,i,t} \equiv Var\left(\check{y}_{i,t}|\hat{y}_{(j-1)},\tilde{z}_t,\mathcal{D}_t\right)$ and $\check{\Sigma}_{j,i,t} \equiv Cov\left(\check{y}_{i,t},\tilde{y}_{j,t}|\hat{y}_{(j-1)},\tilde{z}_t,\mathcal{D}_t\right)$. The partitions imply that $y_{j,t-q+1}$, $\mu_{j,t-q+1}$ and $\check{\sigma}^2_{j,t}$ are scalars. Therefore, take the steps described below:

(a) Since $\check{y}_{1,t} = y_{t-q+1}$, note that $\check{\mu}_{1,1,t} = \tilde{\mu}_{t-\bar{q}+1}$ and $\check{C}_{1,1,t} = \tilde{C}_{t-\bar{q}+1}$.

(b) For each $j \in \{1,...,m\}$, consider the jth $(m \times 1)$ vector $\Phi_{j,\bar{q}}$ in the partition $\Phi_{\bar{q}} \equiv [\Phi_{1,\bar{q}},...,\Phi_{j,\bar{q}},...,\Phi_{m,\bar{q}}]$, and sample $y_{j,t-\bar{q}+1}$ from:[9]

$$\left(y_{j,t-\bar{q}+1}|\hat{y}_{(j-1)},z_{t+1},\mathcal{D}_t\right) \sim N\left(\check{\mu}^y_{j,t-q+1},\left(\check{\sigma}^y_{j,t}\right)^2\right)$$

$$\check{\mu}^y_{j,t-q+1} = \begin{cases} \left(\dfrac{\Phi'_{j,\bar{q}}\Sigma_e^{-1}\tilde{e}_{t+1}+\left(\check{\sigma}^2_{j,t}\right)^{-1}\mu_{j,t-q+1}}{\Phi'_{j,\bar{q}}\Sigma_e^{-1}\Phi_{j,\bar{q}}+\left(\check{\sigma}^2_{j,t}\right)^{-1}}\right) & , \text{if } \check{\sigma}^2_{j,t} > 0 \\ \mu_{j,t-q+1} & , \text{if } \check{\sigma}^2_{j,t} = 0 \end{cases}$$

$$\left(\check{\sigma}^y_{j,t}\right)^2 = \begin{cases} \dfrac{1}{\Phi'_{j,\bar{q}}\Sigma_e^{-1}\Phi_{j,\bar{q}}+\left(\check{\sigma}^2_{j,t}\right)^{-1}} & , \text{if } \check{\sigma}^2_{j,t} > 0 \\ 0 & , \text{if } \check{\sigma}^2_{j,t} = 0 \end{cases}$$

[9] For coding purposes, there still might be some numerical instability in cases where $\check{\sigma}^2_{j,t}$ is positive, but very small. In this case, the restriction $\check{\sigma}^2_{j,t} > 0$ might be translated into something as, for instance, $\check{\sigma}^2_{j,t} > Toler \cdot \mu_{j,t-q+1}$, where $Toler$ is a relative tolerance parameter.

where

$$\check{\mu}_{(j+1),(j+1),t} = \begin{cases} \check{\mu}_{j,j+1,t} + \frac{1}{\check{\sigma}^2_{j,t}} \check{\Sigma}_{j,j+1,t} \left(y_{j,t-\bar{q}+1} - \mu_{j,t-q+1} \right) & \text{, if } \check{\sigma}^2_{j,t} > 0 \\ \check{\mu}_{j,j+1,t} & \text{, if } \check{\sigma}^2_{j,t} = 0 \end{cases}$$

$$\check{C}_{(j+1),(j+1),t} = \begin{cases} \check{C}_{j,j+1,t} - \frac{1}{\check{\sigma}^2_{j,t}} \check{\Sigma}_{j,j+1,t} \check{\Sigma}'_{j,j+1,t} & \text{, if } \check{\sigma}^2_{j,t} > 0 \\ \check{C}_{j,j+1,t} & \text{, if } \check{\sigma}^2_{j,t} = 0 \end{cases}$$

5. Sampled vector z_t is obtained by concatenating

$$z_t = \left[\tilde{z}'_t, y'_{t-\bar{q}+1} \right]' \tag{5}$$

Smoothing As for the smoothed distribution, retrieving it is also eventually subject to inverting the covariance matrix R_t when using the standard method. However, we propose modifying the FFBS algorithm to obtain the distribution, avoiding the issue. In the spirit of the algorithm, all the way from $t = (T - 1)$ to $t = 1$, we keep the means and variances, instead of sampling the states, and adjust vectors and matrices accordingly. By the end, we have retrieved the smoothed distribution. Note that the procedure is not the same as keeping the means and variances at $t = \tau$ when performing the FFBS algorithm for sampling the states at $t = \tau + 1$, for the auxiliary matrices and vectors will not be the same.

Forecasts, backcasts and nowcasts When performing forecasts and backcasts, all the algorithm has to do is to consider the extended sample for which there is no obervation at all. Therefore, the FFBS algorithm will also sample missing values in forecast and backcast exercises. The only requirement is that the VAR step, described in Section 3.2, must only consider the actual available sample period when sampling parameters conditional on observables and missing values.

No additional care is needed in nowcast exercises, for the definition of nowcasts requires at least one observed value to be available. Therefore, nowcasts are carried out on the actual available sample period.

3.2 Bayesian VAR

Following Canova (2007, chap. 10), the linear system can be represented in two companion forms. The first one is defined as follows:

$$\mathbf{Y} = \mathbf{X}\mathbf{\Phi} + \mathcal{E} \tag{6}$$

where $\mathbf{Y} \equiv [y_1, ..., y_T]'$ is a $(T \times m)$ matrix of variables, $\mathbf{\Phi} \equiv \left[\Phi_1, ..., \Phi_q, \bar{\Phi}_c\right]'$ denotes a $(k \times m)$ matrix of coefficients, $\mathbf{X} \equiv [x_1, ..., x_T]'$ is a $(T \times k)$ matrix of regressors, $x_t \equiv \left[y'_{t-1}, ..., y'_{t-q}, \bar{x}'_{ct}\right]'$ is a $(k \times 1)$ vector of regressors, and $\mathcal{E} \equiv [e_1, ..., e_T]'$ is a $(T \times m)$ matrix of error terms[10]. Note that each column of $\mathbf{Y}$, $\mathbf{\Phi}$ and $\mathcal{E}$ corresponds to equations describing the dynamics of a unique variable sequence $\{y_{it}\}$, for $i \in \{1, ..., m\}$:

$$\mathbf{X} = \left[\begin{array}{ccccccc|ccc} y_{1,1-1} & \cdots & y_{m,1-1} & \cdots & y_{1,1-q} & \cdots & y_{m,1-q} & \bar{x}_{c,11} & \cdots & \bar{x}_{c,m_c1} \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ y_{1,t-1} & \cdots & y_{m,t-1} & \cdots & y_{1,t-q} & \cdots & y_{m,t-q} & \bar{x}_{c,1t} & \cdots & \bar{x}_{c,m_ct} \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ y_{1,T-1} & \cdots & y_{m,T-1} & \cdots & y_{1,T-q} & \cdots & y_{m,T-q} & \bar{x}_{c,1T} & \cdots & \bar{x}_{c,m_cT} \end{array}\right]$$

$$\mathbf{Y} = \left[\begin{array}{ccccc} y_{1,1} & \cdots & \boxed{y_{i,1}} & \cdots & y_{m,1} \\ \vdots & & \vdots & & \vdots \\ y_{1,T} & \cdots & \boxed{y_{i,T}} & \cdots & y_{m,T} \end{array}\right] \qquad \mathcal{E} = \left[\begin{array}{ccccc} e_{11} & \cdots & \boxed{e_{i1}} & \cdots & e_{m1} \\ \vdots & & \vdots & & \vdots \\ e_{1T} & \cdots & \boxed{e_{iT}} & \cdots & e_{mT} \end{array}\right]$$

$$\mathbf{\Phi} = \left[\begin{array}{ccccc} \Phi_{1;11} & \cdots & \Phi_{1;i1} & \cdots & \Phi_{1;m1} \\ \vdots & & \vdots & & \vdots \\ \Phi_{1;1m} & \cdots & \Phi_{1;im} & \cdots & \Phi_{1;mm} \\ \vdots & & \vdots & & \vdots \\ \Phi_{q;11} & \cdots & \Phi_{q;i1} & \cdots & \Phi_{q;m1} \\ \vdots & & \vdots & & \vdots \\ \Phi_{q;1m} & \cdots & \Phi_{q;im} & \cdots & \Phi_{q;mm} \\ \hline \bar{\Phi}_{c;11} & \cdots & \bar{\Phi}_{c;i1} & \cdots & \bar{\Phi}_{c;m1} \\ \vdots & & \vdots & & \vdots \\ \bar{\Phi}_{c;1m_c} & \cdots & \bar{\Phi}_{c;im_c} & \cdots & \bar{\Phi}_{c;mm_c} \end{array}\right]$$

Applying the $vec(\cdot)$ operator to both sides of equation (6), we obtain the second companion form:[11]

$$\mathbf{y} = \mathbf{x}\boldsymbol{\phi} + \varepsilon \tag{7}$$

where $\mathbf{y} \equiv vec(\mathbf{Y})$ is a $(Tm \times 1)$ vector of endogenous variables, $\mathbf{x} \equiv (I_m \otimes \mathbf{X})$ is a $(Tm \times km)$ vector of regressors, $\boldsymbol{\phi} \equiv vec(\mathbf{\Phi})$ is a $(km \times 1)$ vector of coefficients, and

[10] Note that the VAR equation for a period t is $y_t = \mathbf{\Phi}' x_t + e_t$.

[11] See the Appendix for some results on matrix algebra and the $vec(\cdot)$ operator.

$\varepsilon \equiv vec(\mathcal{E})$ is a $(Tm \times 1)$ matrix of error terms.

Given the model and a proper mapping between the set of observable variables $(\mathbf{Y}^{ob})$ and endogenous variables $(\mathbf{Y})$ provided above, the next sections describe the procedure to obtain the joint distribution of $\left(\mathbf{Y}, \mathbf{\Phi}|\mathbf{Y}^{ob}\right)$ by means of a Monte Carlo simulation.

3.2.1 Computing the likelihood function

Assuming that $\varepsilon \sim N_{Tm}(0, \Sigma_\varepsilon)$, for $\Sigma_e \equiv (\Sigma_e \otimes I_T)$, the likelihood function is

$$p(\mathbf{y}|\boldsymbol{\phi},\Sigma_\varepsilon) = |2\pi\Sigma_\varepsilon|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi})' \Sigma_\varepsilon^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi})\right]$$

Let $\boldsymbol{\phi}_{\mathbf{ml}}$ and $\Sigma_{\mathbf{ml}}$ denote the maximum likelihood estimator of $\boldsymbol{\phi}$ and its covariance matrix:

$$\boldsymbol{\phi}_{\mathbf{ml}} \equiv \left(I_m \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{y} \quad ; \ \Sigma_{\mathbf{ml}} \equiv \left(\Sigma_e \otimes (\mathbf{X}'\mathbf{X})^{-1}\right)$$

Using the definitions of $\mathbf{x}$ and Σ_ε, note that the likelihood function can be written[12] as follows:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\phi},\Sigma_e) &= |2\pi|^{-\frac{Tm}{2}} |\Sigma_e|^{-\frac{(T-k-m-1)+m+1}{2}} \exp\left(-\frac{1}{2}tr\left(\bar{\Lambda}_s\Sigma_e^{-1}\right)\right) \\ &\quad \times |\Sigma_e|^{-\frac{k}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}})' \Sigma_{\mathbf{ml}}^{-1}(\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}})\right) \end{aligned} \tag{8}$$

where

$$\mathbf{\Phi}_{\mathbf{ml}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad ; \ \bar{\Lambda}_s \equiv (\mathbf{Y}-\mathbf{X}\mathbf{\Phi}_{\mathbf{ml}})'(\mathbf{Y}-\mathbf{X}\mathbf{\Phi}_{\mathbf{ml}}) \tag{9}$$

3.2.2 Joint prior and posterior

Given the form of the likelihood function, we consider that a natural conjugate joint prior distribution for $(\boldsymbol{\phi},\Sigma_e)$ is a generalization of what Gelman et al. (2003) call the Normal-Inverse-Wishart distribution. Thus the joint prior distribution is specified according to the following hierarchical structure:

$$\Sigma_e \sim W_m^{-1}\left(\delta_0, \Lambda_0^{-1}\right) \quad ; \ (\boldsymbol{\phi}|\Sigma_e) \sim N_{km}(\boldsymbol{\phi}_0, \Sigma_0) \quad ; \ \Sigma_0 \equiv \tfrac{1}{\sigma_0}\left(\Sigma_e \otimes \tilde{I}_k\right)$$

[12] See the Appendix for more details.

where $\delta_0 > (m+1)$ is the degrees of freedom of the Inverse-Wishart distribution (which can be understood as the prior sample size for Σ_e), $E\Sigma_e = \Lambda_0 / (\delta_0 - m - 1)$ is the prior estimate for Σ_e, σ_0 is a scale parameter for the conditional Normal distribution (which can be understood as the prior sample size for $\boldsymbol{\phi}$), $\boldsymbol{\phi}_0$ is the prior estimate for $\boldsymbol{\phi}$, and $\tilde{I}_k$ is a $(k \times k)$ diagonal matrix whose diagonal terms satisfy $\tilde{I}_{\mathfrak{K}\mathfrak{K}} \in (0,1]$, for $\mathfrak{K} \in \{1, ..., k\}$.

Note that the identity matrix I_k is a particular case of $\tilde{I}_k$. Moreover, this definition allows for using a Minnesota-type prior (see e.g. Doan et al. (1984) and Litterman (1986) for the original Minnesota prior).

It implies that the joint prior pdf $p(\boldsymbol{\phi},\Sigma_e) = p(\Sigma_e)\, p(\boldsymbol{\phi}|\Sigma_e)$ is proportional to:

$$\begin{aligned} p(\boldsymbol{\phi},\Sigma_e) \;\; &\propto \;\; |\Sigma_e|^{-\frac{(\delta_0+m+1)}{2}} \exp\left(-\frac{1}{2} tr\left(\Lambda_0 \Sigma_e^{-1}\right)\right) \\ &\quad \times |2\pi\Sigma_0|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)' \Sigma_0^{-1} (\boldsymbol{\phi}-\boldsymbol{\phi}_0)\right] \end{aligned} \tag{10}$$

Therefore, the joint posterior pdf $p(\boldsymbol{\phi},\Sigma_e|\mathbf{y}) \propto p(\boldsymbol{\phi},\Sigma_e)\, p(\mathbf{y}|\boldsymbol{\phi},\Sigma_e)$ is proportional to:[13]

$$\begin{aligned} p(\boldsymbol{\phi},\Sigma_e|\mathbf{y}) \;\; &\propto \;\; |\Sigma_e|^{-\frac{(\delta_0+T+m+1)}{2}} \exp\left\{-\frac{1}{2} tr\left[\left(\Lambda_0 + \bar{\Lambda}_s + \Lambda_T\right)\Sigma_e^{-1}\right]\right\} \\ &\quad \times |2\pi\Sigma_T|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\phi}_T)' \Sigma_T^{-1} (\boldsymbol{\phi}-\boldsymbol{\phi}_T)\right] \end{aligned} \tag{11}$$

where

$$\begin{aligned} \Sigma_T^{-1} &\equiv \left(\Sigma_0^{-1} + \Sigma_{\mathbf{ml}}^{-1}\right) = \left(\Sigma_e^{-1} \otimes \left(\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}\right)\right) \\ \boldsymbol{\phi}_T &\equiv \Sigma_T \left(\Sigma_0^{-1}\boldsymbol{\phi}_0 + \Sigma_{\mathbf{ml}}^{-1}\boldsymbol{\phi}_{\mathbf{ml}}\right) \\ \mathbf{\Phi}_T &\equiv vec^{-1}(\boldsymbol{\phi}_T) \\ &= \left(\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1} \left(\sigma_0 \tilde{I}_k^{-1}\right) \mathbf{\Phi}_0 + \left(\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1} (\mathbf{X}'\mathbf{X})\, \mathbf{\Phi}_{\mathbf{ml}} \\ \Lambda_T &\equiv \mathbf{\Phi}_0' \left(\sigma_0 \tilde{I}_k^{-1}\right) \mathbf{\Phi}_0 + \mathbf{\Phi}_{\mathbf{ml}}' (\mathbf{X}'\mathbf{X})\, \mathbf{\Phi}_{\mathbf{ml}} - \mathbf{\Phi}_T' \left(\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}\right) \mathbf{\Phi}_T \end{aligned}$$

This means that the joint posterior distribution of $(\boldsymbol{\phi},\Sigma_e|\mathbf{y})$ is specified according to the following hierarchical structure:

$$(\Sigma_e|\mathbf{y}) \sim W_m^{-1}\left(\delta_0 + T, \left(\Lambda_0 + \bar{\Lambda}_s + \Lambda_T\right)^{-1}\right) \quad ; \quad (\boldsymbol{\phi}|\Sigma_e, \mathbf{y}) \sim N_{km}(\boldsymbol{\phi}_T, \Sigma_T)$$

[13] Details are shown in the Appendix.

3.2.3 Specifying the joint prior distribution

As for the joint prior distribution, we propose the following variation of the Minnesota prior:

$$
\begin{aligned}
\Phi_{0\ell;ij} &= \begin{cases} 1 & \text{if } \ell = 1 \text{ and } i = j \\ 0 & \text{otherwise} \end{cases} \\
\bar{\Phi}_{0c} &= 0_{m\times m_c} \\
\tilde{I}_{\mathfrak{K}\mathfrak{K}} &= \begin{cases} \begin{array}{l} \tilde{\mathfrak{K}}^{-\alpha}\,(\textit{harmonic decay}) \text{ or} \\ \alpha^{1-\tilde{\mathfrak{K}}}\,(\textit{geometric decay}) \end{array} & \text{if } 1 \leq \mathfrak{K} \leq (k - m_c) \\ \\ \qquad\qquad 1 & \text{otherwise} \end{cases}
\end{aligned}
\tag{12}
$$

where $\tilde{\mathfrak{K}} \equiv int\left(\frac{\mathfrak{K}-1}{m}\right)+1$, $\alpha > 0$, $\mathfrak{K} \in \{1,...,k\}$, $\ell \in \{1,...,q\}$, $i \in \{1,...,m\}$, $j \in \{1,...,m\}$, and $\bar{m}_c \in \{1,...,m_c\}$, and again $k \equiv mq + m_c$.

For consistency, we assume that $E\Sigma_e = \bar{\Sigma}_e$, where $E\Sigma_e = \Lambda_0/\left(\delta_0 - m - 1\right)$ and $\bar{\Sigma}_e \equiv \bar{\Lambda}_s/\left(T-1\right)$. This assumption implies that

$$
\Lambda_0 = \frac{(\delta_0 - m - 1)}{(T-1)}\bar{\Lambda}_s \tag{13}
$$

A diffuse prior, if chosen, requires low levels for $\delta_0 > (m+1)$ and $\sigma_0 > 0$. In the limiting case $\delta_0 \to (m+1)$ and $\sigma_0 \to 0$, the posterior distribution implies the maximum likelihood estimators:[14]

$$
(\Sigma_e|\mathbf{y}) \sim W_m^{-1}\left(m+1+T, \left(\bar{\Lambda}_s\right)^{-1}\right) \quad ; \quad (\boldsymbol{\phi}|\Sigma_e, \mathbf{y}) \sim N_{km}\left(\boldsymbol{\phi}_{\mathbf{ml}}, \Sigma_{\mathbf{ml}}^{-1}\right)
$$

for which $\Sigma_{\mathbf{ml}}^{-1} = E\Sigma_e = \bar{\Lambda}_s/T$, $\Sigma_{\mathbf{ml}} \equiv \left(\Sigma_e \otimes \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$, $\boldsymbol{\phi}_{\mathbf{ml}} \equiv \left(I_m \otimes \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\mathbf{y}$, and $\bar{\Lambda}_s \equiv \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}}\right)'\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}}\right)$.

4 Properties of mixed-frequency VAR: simulations

The most common framework for mixed frequency estimation is when the researcher has available an information set with regularly spaced information in a complete panel. As an example, one might be interested in estimating a model with inflation and interest

[14]If $\sigma_0 = 0$ and $\delta_0 = (m+1)$, we conclude that $\boldsymbol{\Phi}_T = \boldsymbol{\Phi}_{\mathbf{ml}}$ (or $\boldsymbol{\phi}_T = \boldsymbol{\phi}_{\mathbf{ml}}$), $\Lambda_T = \Lambda_0 = \Sigma_0^{-1} = 0$, and $\Sigma_T = \Sigma_{\mathbf{ml}}$.

rates (usually measured at monthly frequency) and output (from the National Accounts at quarterly frequency). Instead of aggregating information on inflation and interest rates to quarterly frequency, in order to match output data, our framework allows a joint estimation of model.

For the first exercise, assume that the dataset has four time series covering a sample T of 200 periods at monthly frequency. In this exercise, $N < 4$ artificial series will have equally distributed gaps along the time series (N will change with simulations). The VAR model will obtain estimates of the missing observations and compare with the true value of artificial dataset initially simulated. Time series for the exercise are simulated from the following VAR at monthly frequency:

$$Y_t = AY_{t-1} + Se_t$$

where

$$A = \begin{bmatrix} 0.900 & 0.010 & -0.020 & 0.050 \\ 0.000 & 0.900 & -0.113 & -0.010 \\ 0.000 & 0.195 & 0.800 & 0.000 \\ -0.269 & 0.000 & 0.000 & 0.700 \end{bmatrix} \qquad S = \begin{bmatrix} 0.01 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.01 \end{bmatrix}$$

Values on matrix A were selected in order to ensure the same long run variance for each observed variable in vector Y_t, given the lack of contemporaneous conditional correlation given by matrix S. Alternative values for matrix S were tested, imposing some correlation across observed variables. Results were qualitatively similar.

Table 1 shows the median of the RMSE across 1000 simulated time series, measured as a proportion of the unconditional standard deviation of the series in the true VAR. A total of six simulations are shown in each line. First, each group of two columns compute the RMSE using the mean, the median and the mode of the simulations in the Gibbs sampling as the forecast for the missing observation. For each group of columns, two simulations try to disentangle the uncertainty from parameter estimation in the VAR when simulating the missing observations: the columns labeled "True Values" report the RMSE of the estimated VAR against the true values of the missing observations; the columns labeled "True VAR" use the true values of parameters describing the VAR to simulate the missing

Table 1: Median RMSE and Simulation of Missing Values

	RMSE – Mean		RMSE – Median		RMSE – Mode	
	True Values	True VAR	True Values	True VAR	True Values	True VAR
			$T = 200; \quad N = 1$			
$F = Q$	0.4181	0.1074	0.4182	0.1068	0.4531	0.2073
$F = S$	0.6592	0.4186	0.6572	0.4137	0.7966	0.6155
$F = Y$	0.9520	0.7655	0.9461	0.7592	1.1432	0.9899
			$T = 200; \quad N = 2$			
$F = Q$	0.4143	0.0976	0.4143	0.0972	0.4414	0.1813
$F = S$	0.6706	0.4357	0.6634	0.4264	0.8381	0.6622
$F = Y$	1.1254	0.9445	1.0264	0.8336	1.8651	1.7677
			$T = 200; \quad N = 3$			
$F = Q$	0.4301	0.0912	0.4293	0.0904	0.4555	0.1718
$F = S$	0.6677	0.3866	0.6630	0.3788	0.8505	0.6517
$F = Y$	1.6257	1.4690	1.1518	0.9336	6.6475	6.6008

observation. Table 1 also split the analysis in terms of number of observed time series with missing values in the VAR (N) and the frequency of missing observations (Q, S and Y for information in each quarter, semester and year, respectively), while keeping the sample size, $T = 200$, fixed.

Table 1 shows three interesting results. First, as expected, the RMSE of simulated time series of the mixed-frequency VAR does depend on the frequency of missing observations. For all exercises shown in the table, changing the frequency of missing informations from quarterly to yearly data increased the RMSE of the simulations, irrespective of the statistic summarizing the forecast of missing values. Despite an huge increase in the RMSE when using the mode of simulations in the Gibbs sampling with yearly information in three time series of the VAR, all simulations showed similar variations in the RMSE when changing the frequency of missing information in the model.

Second, also expected, the RMSE of simulated time series does dependent on the number of time series with missing observations in the VAR. While this result does not look significant when information is available at quarterly frequency, simulations with semi-annual and, especially, annual data shows that increasing the number of time series with missing observations increases the RMSE of simulations, hurting the performance of the algorithm.

Finally, there is a significant gain in knowing the true values of parameters in the

Table 2: Median RMSE and Simulation of Missing Values – T=400

	RMSE – Mean		RMSE – Median		RMSE – Mode	
	True Values	True VAR	True Values	True VAR	True Values	True VAR
	$T = 400;\quad N = 1$					
$F = Q$	0.4119	0.0817	0.4120	0.0819	0.4366	0.1616
$F = S$	0.6107	0.3431	0.6100	0.3424	0.6915	0.4753
$F = Y$	0.8119	0.5677	0.8081	0.5619	0.9963	0.8011
	$T = 400;\quad N = 2$					
$F = Q$	0.4103	0.0750	0.4103	0.0750	0.4245	0.1343
$F = S$	0.6083	0.3340	0.6067	0.3324	0.6802	0.4543
$F = Y$	0.8692	0.6334	0.8516	0.6090	1.1470	0.9759
	$T = 400;\quad N = 3$					
$F = Q$	0.4185	0.0616	0.4184	0.0615	0.4310	0.1231
$F = S$	0.6144	0.3049	0.6140	0.3049	0.6870	0.4327
$F = Y$	1.0238	0.7834	0.9577	0.6968	1.7854	1.6559

system, meaning that estimation problems should be carefully addressed in empirical applications. Comparing results from columns labeled "True Values" with those using the true set of parameters of the VAR, labeled "True VAR", the gains in terms of RMSE in knowing exactly the parameters of the model are significant, irrespective of the statistic summarizing the simulations in the Gibbs sampler and the frequency of missing information.

Table 2 considers the case where the sample available, T, at monthly frequency is twice the size of the benchmark exercise. Results show that the performance of the VAR with missing data at quarterly frequency does not show significant improvement, unless all parameters of the system of equations are known. The ratio between the RMSE and the volatility of the simulated time series remains almost the same in both exercises with missing values at quarterly frequency. On the other hand, simulations with observable variables in each semester or in each year show a significant improvement, even considering the uncertainty related to parameter estimation of the VAR.

5 Empirical Analysis

In this section, two exercises show the ability of MF-VAR to simulate time series of economic indicators that might be useful for economic analysis. The two exercises explore some of the properties discussed in the previous section, while also showing the flexibility of the procedure when describing the datasets. The first exercise provides estimates of time series of real GDP for Brazil at monthly frequency, using a large set of observable variables to build an indicator of economic activity at higher frequency. In this exercise, estimates show the consequences of properly specifying the transformation matrix Λ_z in empirical applications.

The second exercise tries to retrieve information from a large set of observable variables to build a time series for monthly unemployment rates based on the latest release of the unemployment survey for Brazil – the so-called "Pesquisa Nacional por Amostra de Domicílios Contínua" (PNAD-C, in Portuguese). The first results of this survey cover the period between March-2012 and March-2015. Using a set of MF-VAR models with different observable variables, including information from an annual version of PNAD, the section explores the ability of the models to generate consistent estimates of unemployment rate if missing data is irregularly spaced.

5.1 Estimating real GDP at monthly frequency

One interesting application of mixed-frequency VARs is to build high-frequency estimations of real gross domestic product using indicators of economic activity. Here, we use a group of economic indicators to build 5-variable MF-VARs and analyze the properties of the combination of the forecasts of these MF-VARs. These indicators include total industrial production, industrial production of capital goods, retail sales, total exports, total imports, energy consumption, energy consumption of manufacturing industry, the ratio of the trade balance over the trade flow, oil production and steel production. All these indicators are available at monthly frequency. The MF-VARs include seasonal dummies and change the assumption about deterministic trends between no trend, linear trend or a quadratic trend.

In empirical applications of the MF-VAR, it is important to properly characterize the setup of matrix Λ_z in equation 2, relating observable variables at high- and low-frequency. Figure 1 below shows the simulated path from one of the MF-VARs[15] for the quarterly

[15] The MF-VAR has three lags, and includes, besides the log of GDP at quarterly frequency, seasonal factors and a linear trend, the log of production of capital goods, retail sales, exports and energy consumption.

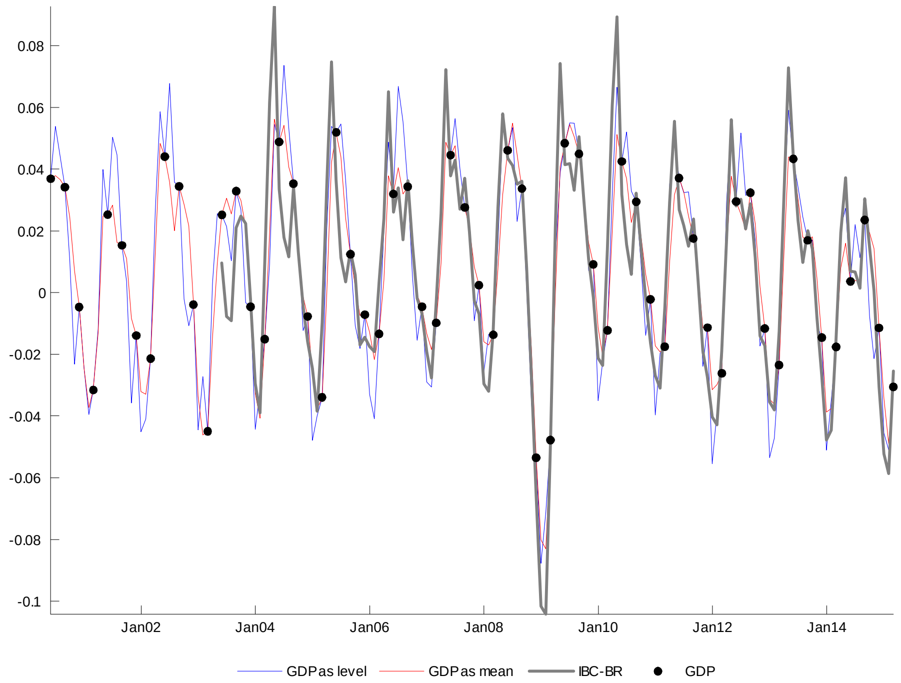


Figure 1: Quarterly GDP Change in High-Frequency – June/2000 to March/2015

growth rates of high-frequency estimates of (log-) real GDP[16] under two assumptions for matrix Λ_z: the blue line assumes a simple setup, where observed real GDP is directly correlated with high-frequency observation in the last month of the quarter; the red line sets a proper framework, with observed real GDP as the sum of simulated real GDP in the three months of the quarter. The gray line shows an indicator of high-frequency estimates of real GDP calculated by the Central Bank of Brazil – the so-called IBC-BR. It is clear from the figure that a proper setup of matrix Λ_z allows a closer match between the simulated path of the MF-VAR and the estimates provided by IBC-BR, as the real GDP in levels tends to systematically overestimate seasonal movements early in every year, while also missing the location of peaks in the middle of each year. The setup with real GDP as a quarterly mean provides a better match with IBC-BR at both these points.

For the exercise, combined forecast of MF-VARs are compared with the forecasts of a simple AR(1) process for the growth rate of (log-) real GDP with seasonal dummies and a set of Bayesian VARs at quarterly frequency with the same observables used in MF-VARs. The main objective of using Bayesian VARs to compare with MF-VARs is

[16]Data on real GDP used in this exercise was published by IBGE in the end of May 2015, with first vintage of data for the first quarter of 2015. This exercise does not handle with "real time forecasting", as a methodological change in GDP computation prevents a clear comparison between recent and previous vintages of data.

Table 3: Forecasting GDP with MF-VAR: RMSE – Dec/10 to Mar/15

		RMSE			Relative RMSE	
	Obs	MF-VAR	VAR	AR(1)	VAR	AR(1)
t = 0	18	0.0083	0.0200	0.0109	2.40	1.31
t = 1	17	0.0170	0.0298	0.0157	1.76	0.92
t = 2	16	0.0256	0.0278	0.0189	1.08	0.74
t = 3	15	0.0298	0.0291	0.0256	0.98	0.86
t = 4	14	0.0341	0.0415	0.0277	1.22	0.81

to measure the effect of using high-frequency information to build forecasts, instead of simply aggregating information at lower frequencies. The comparison with an AR(1) process provides a benchmark in terms of forecasting ability of the MF-VAR. The MF-VARs are arbitrarily set with three lags, in order to properly characterize the dynamics between monthly and quarterly frequency data. Bayesian VARs at quarterly frequency, on the other hand, due to the small sample size, are estimated with only one lag.

Table 3 compares the outcome of forecasts in the three models. In the table, $t = 0$ shows the properties of the so-called "nowcasting" in the MF-VAR, which is the forecast for real GDP at a given time t, given that all information for other variables, except GDP, is available for forecasting. It is worth noting that the additional information building the "nowcast" affects only forecasts from the MF-VARs and the Bayesian VARs; for the AR(1) process, the "nowcasting" is equivalent to a one-step-ahead forecast.

Table 3 shows the gains of using high-frequency information, instead of aggregating data to lower frequencies. The performance of the MF-VAR is, in a worst case scenario, similar to the Bayesian VAR, when forecasting three quarters ahead. In all other horizons considered, the performance of the MF-VAR clearly dominates the Bayesian VAR. However, despite providing a good performance in terms of "nowcasting", both the MF-VAR and the Bayesian VAR are not better than an univariate system estimated at quarterly frequency. The main advantage of using the MF-VAR, thus, seems to be associated with the ability of using high-frequency information to make inference about the current period.

5.2 Estimating unemployment rate with irregularly spaced data

One of the possible uses of MF-VARs is in the simulation of missing information in the time series of a variable when there are other proxy variables to approximate the dynamics

Table 4: Unemployment Surveys in Brazil

	PME (IBGE)	PED (DIEESE)	PNAD (IBGE)	PNAD-C (IBGE)
Frequency	Monthly	Monthly	Annual	Monthly*
Coverage	6 metro areas	4 metro areas	Brazil	Brazil
Pop. Ratio (2012)	0.25	0.15	1.00	1.00
Time Span	1999M05 2015M03	1999M05 2015M03	1999 2013	2012M03 2015M03
Observations	Major review in 2001	Restricted to same metro areas	Missing censitary years (2000 and 2010)	Information about last quarter

(*) Information refers to the moving average of the quarter finished at a given month.

of the variable with missing information. Here, we present a set of MF-VARs designed to estimate the missing values of unemployment rate in Brazil measured by a recent survey of IBGE – "Pesquisa Nacional por Amostra de Domicílios Contínua" (PNAD-C) in Portuguese. This survey shows estimates of unemployment rate in the previous quarter at monthly frequency, covering almost four times the total population surveyed in the old survey conducted by IBGE ("Pesquisa Mensal de Emprego" – PME). There is also an annual survey from IBGE, called "Pesquisa Nacional por Amostra de Domicílios" (PNAD), where information on employment is also collected, with similar coverage compared to PNAD-C. Information on employment status in Brazil is very irregular. However, given the role of labor markets for economic analysis, generating a time series of unemployment based on PNAD-C constitutes a valuable piece of information. Table 4 summarizes the main characteristics of four surveys on employment for Brazil.

In order to simulate time series for unemployment in Brazil based on PNAD-C, we first assume that unemployment rates from PNAD-C and PNAD differ only by a multiplicative factor,[17] which is retrieved by considering the only month for which we have observations from both series, i.e. Sept. 2012.[18] After that, we adjust the level of PNAD annual observations and merge them with PNAD-C monthly observations. Given a linear relation between the overlapping information between PNAD and PNAD-C, we set at least one point of data for the period between 1999 and 2011 – the period before PNAD-C results were published. After that, a set of four-variable MF-VARs were estimated

[17]Results do not change much when we assume an additive factor instead.

[18]Figures from PNAD evaluate employment situation on the last day of September in each year.

using information from: PME and PED surveys, with measures of unemployment and participation rates, CAGED data of formal employment and CNI data on employment in manufacturing sector. MF-VARs included seasonal dummies and were estimated with and without deterministic linear and quadratic trends. Estimations including a larger number of observable variables were not performed due to the small sample size combined with larger gaps in the information from PNAD-C. Also, in terms of restriction to variable selection in the VARs, at least one of the unemployment time series was included in the estimation, together with PNAD-C data. A total of 36 models were used to simulate the time series of unemployment.

Final estimates based on the mean and the median for the time series of PNAD-C are shown in figure 2, together with confidence bands based on the distribution of simulations across models. As expected, confidence bands are larger when the gap between information available at PNAD-C is also larger. Also as expected, confidence bands increase during turning points of the simulated time series. Of course, as the analysis in Section 4 suggest, as more data for PNAD-C become available, it should be possible to obtain better estimates of the dynamics of the time series during these turning points. As a consequence, we expect these confidence intervals to become smaller with the increase in the amount of information from PNAD-C. We have also run an in-sample forecasting exercise over the last year of the sample, in order to set weights across different models based on the mean square error of the models. The final time series for unemployment generated from this exercise, not shown in figure 2, is very close to the median of the simulations across models.

The simulated time series follows a close path when compared to the actual time series of PME and PED surveys. The main difference between these two surveys and the simulated path of PNAD-C presented here is related to the gap between the minimum and the maximum of each time series: while PME and PED, on average, showed unemployment rates above 12% in the period between 1999 and 2004 – even reaching 19% in one of the measures provided in PED –, the average of PNAD-C in the sample is set at 10.5%. On the other hand, for the late part of the sample, while PME reached historical lows in December of 2013 and December of 2014 of 4.3%, the minimum of PNAD-C – which, in the sample, is an observable, not a simulation – was found in December, 2013, at 6.2%. The main reason for this discrepancy is the use of annual information from PNAD for the

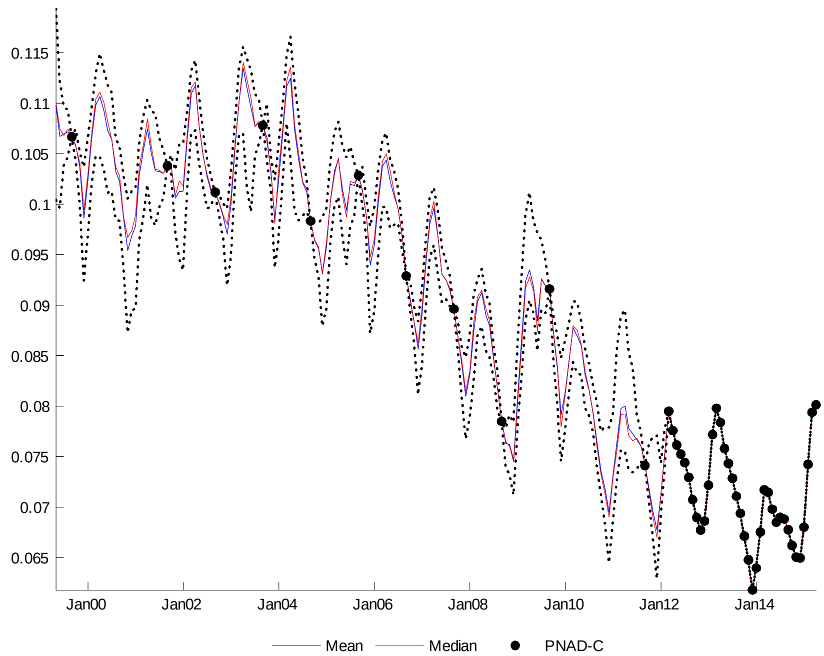


Figure 2: Simulated Time Series: PNAD-C – May/1999 to April/2015

period between 1999 and 2011: without the support of such information, simulated paths of PNAD-C would be closer to the mean between PME and PED for the early part of the sample.

6 Conclusion

This paper presented a new efficient algorithm to draw inference from the joint distribution of a dataset with missing information. The algorithm presented here is flexible enough to accommodate most of the issues faced when dealing with irregularly spaced datasets. It also allows for an easy setup of linear restrictions in observable variables, irrespective of the presence of missing information in those time series. From this perspective, the implementation of the model uses most of the findings in early literature on the subject, generalizing some results and applications for the framework of mixed-frequency estimation of VAR models.

Both the analysis with simulated and real datasets show the flexibility of the framework to properly handle datasets with an usual size observed in real world's applications. Despite results with simulated data suggest that some careful analysis of convergence of the Gibbs sampler is necessary to ensure that simulated time series are being drawn from appropriate parameter values, the small volatility of simulations show that the algorithm is a very useful tool to handle problems associated with missing information in large datasets.

In terms of the empirical applications with real data, the main results show the importance of using high-frequency information to make inference on macroeconomic dataset with missing values. The applications show the advantage of using high-frequency information in two different contexts: first, in terms of forecasting, MF-VARs using the framework proposed here had a better performance than traditional Bayesian VARs estimated using data at lower frequencies; second, the algorithm proposed here estimated past trajectories of variables with missing information combining a large set of information, generating simulations with relatively small dispersion around its mean.

References

Basdevant, O. (2003). On applications of state-space modelling in macroeconomics. Reserve Bank of New Zealand Discussion Paper Series DP2003/02, Reserve Bank of New Zealand. 00027.

Canova, F. (2007). *Methods for applied macroeconomic research.* Princeton: Princeton University Press.

Carter, C. K. and R. Kohn (1994, August). On Gibbs Sampling for State Space Models. *Biometrika 81*(3), 541–553.

Doan, T., R. B. Litterman, and C. A. Sims (1984, January). Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews 3*(1), 1–100.

Durbin, J. and S. J. Koopman (2012, August). *Time Series Analysis by State Space Methods* (2nd ed.). Number 38 in Oxford Statistical Science. Oxford University Press.

Eraker, B., C. W. J. Chiu, A. T. Foerster, T. B. Kim, and H. D. Seoane (2014). Bayesian Mixed Frequency VARs. *Journal of Financial Econometrics In Press.*

Fruhwirth-Schnatter, S. (1994, March). Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis 15*(2), 183–202.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003, July). *Bayesian Data Analysis, Second Edition* (2 edition ed.). Boca Raton, Fla: Chapman and Hall/CRC.

Hamilton, J. D. (1994). *Time series analysis.* Princeton, N.J.: Princeton University press. 15916.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering 82*(1), 35–45.

Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions - Five Years of Experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

Prado, R. and M. West (2010). *Time series.* London: Chapman & Hall.

Schorfheide, F. and D. Song (2014). Real-Time Forecasting with a Mixed-Frequency VAR. *Journal of Business & Economic Statistics In Press.*

West, M. (1996). Bayesian Time Series: Models and Computations for the Analysis of Time Series in the Physical Sciences. In K. M. Hanson and R. N. Silver (Eds.), *Maximum Entropy and Bayesian Methods*, Number 79 in Fundamental Theories of Physics, pp. 23–34. Springer Netherlands.

West, M. (1997, January). Time series decomposition. *Biometrika 84*(2), 489–494.

West, M. and J. Harrison (1997). *Bayesian forecasting and dynamic models.* New York: Springer.

A Some results

Matrix Algebra: Recall that, for any matrices $A_{k\times\ell}$, $B_{\ell\times m}$ and $C_{m\times n}$, the $vec(\cdot)$ operator satisfies the following properties:

$$\begin{array}{ll} tr(A'B) = (vec(B))'vec(A) = (vec(A))'vec(B) & vec(ABC) = (C'\otimes\mathbf{A})vec(B) \\ vec(ABC) = vec(AB\cdot C\cdot I_n) = (I_n\otimes AB)vec(C) & vec(ABC) = vec(I_k\cdot A\cdot BC) = (C'B'\otimes I_k)vec(A) \\ vec(AB){=}vec(A\cdot B\cdot I_m){=}(I_m\otimes\mathbf{A})vec(B) & vec(AB) = vec(I_k\cdot A\cdot B) = (B'\otimes I_k)vec(A) \end{array}$$

Recall that, for any matrices A, B, C and D, the Kronecker product satisfies (when it makes sense) the following properties:

$$\begin{array}{lll} (A\otimes B)(C\otimes D) = (AC\otimes BD) & (A\otimes B)^{-1} = \left(A^{-1}\otimes B^{-1}\right) & tr(A\otimes B) = tr(A)\cdot tr(B) \\ (A\otimes B)' = (A'\otimes B') & |A\otimes B| = |A|^{\ell}|B|^{k} & \text{if } A \text{ is } (k\times k)\text{, and } B \text{ is } (\ell\times\ell) \end{array}$$

Maximum Likelihood Estimator - Multiv Gaussian Distribution: The estimator comes from $0 = -\left(\Sigma_{\varepsilon}^{-\frac{1}{2}}\mathbf{x}\right)'\left[\Sigma_{\varepsilon}^{-\frac{1}{2}}\left(\mathbf{y}-\mathbf{x}\boldsymbol{\phi}_{\mathbf{ml}}\right)\right]$, which implies:

$$\begin{aligned} \boldsymbol{\phi}_{\mathbf{ml}} &= \left[\left(\Sigma_{\varepsilon}^{-\frac{1}{2}}\mathbf{x}\right)'\left(\boldsymbol{\Sigma}_{\varepsilon}^{-\frac{1}{2}}\mathbf{x}\right)\right]^{-1}\left(\Sigma_{\varepsilon}^{-\frac{1}{2}}\mathbf{x}\right)'\left(\Sigma_{\varepsilon}^{-\frac{1}{2}}\mathbf{y}\right) \\ &= \left[\left((\Sigma_e\otimes I_T)^{-\frac{1}{2}}(I_m\otimes\mathbf{X})\right)'\left((\Sigma_e\otimes I_T)^{-\frac{1}{2}}(I_m\otimes\mathbf{X})\right)\right]^{-1}\left((\Sigma_e\otimes I_T)^{-\frac{1}{2}}(I_m\otimes\mathbf{X})\right)'\left((\Sigma_e\otimes I_T)^{-\frac{1}{2}}\mathbf{y}\right) \\ &= \left[\left(\left(\Sigma_e^{-\frac{1}{2}}\otimes I_T\right)(I_m\otimes\mathbf{X})\right)'\left(\left(\Sigma_e^{-\frac{1}{2}}\otimes I_T\right)(I_m\otimes\mathbf{X})\right)\right]^{-1}\left(\left(\Sigma_e^{-\frac{1}{2}}\otimes I_T\right)(I_m\otimes\mathbf{X})\right)'\left(\left(\Sigma_e^{-\frac{1}{2}}\otimes I_T\right)\mathbf{y}\right) \\ &= \left[\left(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\right)'\left(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\right)\right]^{-1}\left(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\right)'\left(\left(\Sigma_e^{-\frac{1}{2}}\otimes I_T\right)\mathbf{y}\right) \\ &= \left[\left(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}'\right)\left(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\right)\right]^{-1}\left(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}'\right)\left(\Sigma_e^{-\frac{1}{2}}\otimes I_T\right)\mathbf{y} \\ &= \left(\Sigma_e^{-1}\otimes\mathbf{X}'\mathbf{X}\right)^{-1}\left(\Sigma_e^{-1}\otimes\mathbf{X}'\right)\mathbf{y}{=}\left(\Sigma_e\otimes(\mathbf{X}'\mathbf{X})^{-1}\right)\left(\Sigma_e^{-1}\otimes\mathbf{X}'\right)\mathbf{y} = \left(I_m\otimes(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{y} \end{aligned}$$

Rewriting the Likelihhod Function: Note that:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\phi},\Sigma_e) &= |2\pi\Sigma_\varepsilon|^{-\frac{1}{2}}\exp\Big[-\tfrac{1}{2}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi})'\Sigma_\varepsilon^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi})\Big] \\ &= |2\pi\Sigma_\varepsilon|^{-\frac{1}{2}}\exp\Big\{-\tfrac{1}{2}\Big[\Sigma_\varepsilon^{-\frac{1}{2}}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi})\Big]'\Big[\Sigma_\varepsilon^{-\frac{1}{2}}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi})\Big]\Big\} \end{aligned}$$

And hence, replacing $(\boldsymbol{\phi}_{\mathbf{ml}}+\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}})$ for $\boldsymbol{\phi}$, we obtain:

$$\begin{aligned} \Big[\Sigma_\varepsilon^{-\frac{1}{2}}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi})\Big]'\Big[\Sigma_\varepsilon^{-\frac{1}{2}}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi})\Big] &= \Big[\Big(\Sigma_e^{-\frac{1}{2}}\otimes I_T\Big)(\mathbf{y}-(I_m\otimes\mathbf{X})\boldsymbol{\phi})\Big]'\Big[\Big(\Sigma_e^{-\frac{1}{2}}\otimes I_T\Big)(\mathbf{y}-(I_m\otimes\mathbf{X})\boldsymbol{\phi})\Big] \\ &= \Big[\Big(\Sigma_e^{-\frac{1}{2}}\otimes I_T\Big)\mathbf{y}-\Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)\boldsymbol{\phi}\Big]'\Big[\Big(\Sigma_e^{-\frac{1}{2}}\otimes I_T\Big)\mathbf{y}-\Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)\boldsymbol{\phi}\Big] \\ &= \Big[\Big(\Sigma_e^{-\frac{1}{2}}\otimes I_T\Big)\mathbf{y}-\Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)\boldsymbol{\phi}_{\mathbf{ml}}\Big]'\Big[\Big(\Sigma_e^{-\frac{1}{2}}\otimes I_T\Big)\mathbf{y}-\Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)\boldsymbol{\phi}_{\mathbf{ml}}\Big] \\ &\quad + (\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}})'\Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)'\Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)(\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}}) - 2(\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}})'\mathcal{X}_e \end{aligned}$$

where

$$\begin{aligned} \mathcal{X}_e &\equiv \Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)'\Big[\Big(\Sigma_e^{-\frac{1}{2}}\otimes I_T\Big)\mathbf{y}-\Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)\boldsymbol{\phi}_{\mathbf{ml}}\Big] \\ &= \Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)'\Big[\Big(\Sigma_e^{-\frac{1}{2}}\otimes I_T\Big)-\Big(\Sigma_e^{-\frac{1}{2}}\otimes\mathbf{X}\Big)\Big(I_m\otimes(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)\Big]\mathbf{y} \\ &= \Big[\Big(\Sigma_e^{-1}\otimes\mathbf{X}'\Big)-\Big(\Sigma_e^{-1}\otimes\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)\Big]\mathbf{y} = \Big[\Big(\Sigma_e^{-1}\otimes\mathbf{X}'\Big)-\Big(\Sigma_e^{-1}\otimes\mathbf{X}'\Big)\Big]\mathbf{y} = \mathbf{0} \end{aligned}$$

Therefore:

$$p(\mathbf{y}|\boldsymbol{\phi},\Sigma_e) = |2\pi\Sigma_\varepsilon|^{-\frac{1}{2}}\exp\Big[-\tfrac{1}{2}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi}_{\mathbf{ml}})'(\Sigma_e\otimes I_T)^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi}_{\mathbf{ml}}) - \tfrac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}})'\Big(\Sigma_e^{-1}\otimes\mathbf{X}'\mathbf{X}\Big)(\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}})\Big]$$

where

$$\begin{aligned} (\mathbf{y}-\mathbf{x}\boldsymbol{\phi}_{\mathbf{ml}})'(\Sigma_e\otimes I_T)^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\phi}_{\mathbf{ml}}) &= (vec(\mathbf{Y}-\mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}}))'\Big(\Sigma_e^{-1}\otimes I_T\Big)vec(\mathbf{Y}-\mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}}) \\ &= (vec(\mathbf{Y}-\mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}}))'vec\Big(I_T(\mathbf{Y}-\mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}})\Sigma_e^{-1}\Big) = tr\Big[(\mathbf{Y}-\mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}})\Sigma_e^{-1}\Big] \end{aligned}$$

Note also that

$$\begin{aligned} \mathbf{vec}(\boldsymbol{\Phi}_{\mathbf{ml}}) &= \boldsymbol{\phi}_{\mathbf{ml}} = \Big(I_m\otimes(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)\mathbf{y} = \Big(I_m\otimes(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)vec(\mathbf{Y}) = vec\Big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\Big) \\ \mathbf{vec}(\mathbf{X}\boldsymbol{\Phi}_{\mathbf{ml}}) &= \mathbf{x}\boldsymbol{\phi}_{\mathbf{ml}} = (I_m\otimes\mathbf{X})\Big(I_m\otimes(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)\mathbf{y} = \Big(I_m\otimes\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)\mathbf{y} \\ &= \Big(I_m\otimes\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)vec(\mathbf{Y}) = vec\Big(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\Big) \end{aligned}$$

Joint posterior pdf of the Bayesian VAR: We need to compute the following summation terms:

$$\begin{aligned} S_1 &= (\boldsymbol{\phi}-\boldsymbol{\phi}_0)'\Sigma_0^{-1}(\boldsymbol{\phi}-\boldsymbol{\phi}_0) + (\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}})'\Sigma_{\mathbf{ml}}^{-1}(\boldsymbol{\phi}-\boldsymbol{\phi}_{\mathbf{ml}}) \\ &= (\boldsymbol{\phi}-\boldsymbol{\phi}_T)'\Sigma_{\mathbf{T}}^{-1}(\boldsymbol{\phi}-\boldsymbol{\phi}_T) + \boldsymbol{\phi}_0'\Sigma_0^{-1}\boldsymbol{\phi}_0 + \boldsymbol{\phi}_{\mathbf{ml}}'\Sigma_{\mathbf{ml}}^{-1}\boldsymbol{\phi}_{\mathbf{ml}} - \boldsymbol{\phi}_T'\Sigma_T^{-1}\boldsymbol{\phi}_T \end{aligned}$$

$$\begin{aligned} S_2 &= \boldsymbol{\phi}_0'\Sigma_0^{-1}\boldsymbol{\phi}_0 + \boldsymbol{\phi}_{\mathbf{ml}}'\Sigma_{\mathbf{ml}}^{-1}\boldsymbol{\phi}_{\mathbf{ml}} - \boldsymbol{\phi}_T'\Sigma_T^{-1}\boldsymbol{\phi}_T \\ &= vec(\boldsymbol{\Phi}_0)'\Big(\Sigma_e^{-1}\otimes\sigma_0\tilde{I}_k\Big)vec(\boldsymbol{\Phi}_0) + vec(\boldsymbol{\Phi}_{\mathbf{ml}})'\Big(\Sigma_e^{-1}\otimes(\mathbf{X}'\mathbf{X})\Big)vec(\boldsymbol{\Phi}_{\mathbf{ml}}) - vec(\boldsymbol{\Phi}_T)'\Big(\Sigma_e^{-1}\otimes\Big(\sigma_0\tilde{I}_k+\mathbf{X}'\mathbf{X}\Big)\Big)vec(\boldsymbol{\Phi}_T) \\ &= vec(\boldsymbol{\Phi}_0)'vec\Big(\Big(\sigma_0\tilde{I}_k\Big)\boldsymbol{\Phi}_0\Sigma_e^{-1}\Big) + vec(\boldsymbol{\Phi}_{\mathbf{ml}})'vec\Big((\mathbf{X}'\mathbf{X})\boldsymbol{\Phi}_{\mathbf{ml}}\Sigma_e^{-1}\Big) - vec(\boldsymbol{\Phi}_T)'vec\Big(\Big(\sigma_0\tilde{I}_k+\mathbf{X}'\mathbf{X}\Big)\boldsymbol{\Phi}_T\Sigma_e^{-1}\Big) \\ &= tr\Big[\Big(\boldsymbol{\Phi}_0'\Big(\sigma_0\tilde{I}_k\Big)\boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_{\mathbf{ml}}'(\mathbf{X}'\mathbf{X})\boldsymbol{\Phi}_{\mathbf{ml}} - \boldsymbol{\Phi}_T'\Big(\sigma_0\tilde{I}_k+\mathbf{X}'\mathbf{X}\Big)\boldsymbol{\Phi}_T\Big)\Sigma_e^{-1}\Big] = tr\Big[\Lambda_T\Sigma_e^{-1}\Big] \end{aligned}$$

where

$$\Sigma_T^{-1} \equiv \left(\Sigma_0{}^{-1}+\Sigma_{\mathbf{ml}}^{-1}\right) = \Sigma_e^{-1}\otimes\left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)$$
$$\phi_T \equiv \Sigma_T\left(\Sigma_0^{-1}\phi_0 + \Sigma_{\mathbf{ml}}^{-1}\phi_{\mathbf{ml}}\right) = \left(I_m\otimes\left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)^{-1}\left(\sigma_0\tilde{I}_k\right)\right)\phi_0+ \left(I_m\otimes\left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)^{-1}(\mathbf{X'X})\right)\phi_{\mathbf{ml}}$$
$$\mathbf{\Phi}_T \equiv vec^{-1}(\phi_T) = \left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)^{-1}\left(\sigma_0\tilde{I}_k\right)\mathbf{\Phi}_0 + \left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)^{-1}(\mathbf{X'X})\mathbf{\Phi}_{\mathbf{ml}}$$
$$\Lambda_T \equiv \mathbf{\Phi}_0'\left(\sigma_0\tilde{I}_k\right)\mathbf{\Phi}_0 + \mathbf{\Phi}_{\mathbf{ml}}'(\mathbf{X'X})\mathbf{\Phi}_{\mathbf{ml}} - \mathbf{\Phi}_T'\left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)\mathbf{\Phi}_T$$

Deriving the expression for $\mathbf{\Phi}_T$ goes as follows:

$$\begin{aligned} vec(\mathbf{\Phi}_T) &\equiv \left(I_m\otimes\left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)^{-1}\left(\sigma_0\tilde{I}_k\right)\right)vec(\mathbf{\Phi}_0) + \left(I_m\otimes\left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)^{-1}(\mathbf{X'X})\right)vec(\mathbf{\Phi}_{\mathbf{ml}}) \\ &= vec\left(\left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)^{-1}\left(\sigma_0\tilde{I}_k\right)\mathbf{\Phi}_0 + \left(\sigma_0\tilde{I}_k+\mathbf{X'X}\right)^{-1}(\mathbf{X'X})\mathbf{\Phi}_{\mathbf{ml}}\right) \end{aligned}$$

Note also that

$$\begin{aligned} \phi_T &\equiv \Sigma_T\left(\Sigma_0^{-1}\phi_0+\Sigma_{\mathbf{ml}}^{-1}\phi_{\mathbf{ml}}\right) = \left(\Sigma_0^{-1}+\Sigma_{\mathbf{ml}}^{-1}\right)^{-1}\left(\Sigma_0^{-1}\phi_0+\Sigma_{\mathbf{ml}}^{-1}\phi_{\mathbf{ml}}\right) \\ &= \left(\sigma_0\left(\Sigma_e^{-1}\otimes\tilde{I}_k^{-1}\right)+\left(\Sigma_e^{-1}\otimes(\mathbf{X'X})\right)\right)^{-1}\left(\sigma_0\left(\Sigma_e^{-1}\otimes\tilde{I}_k^{-1}\right)\phi_0+\left(\Sigma_e^{-1}\otimes(\mathbf{X'X})\right)\phi_{\mathbf{ml}}\right) \\ &= \left(\Sigma_e^{-1}\otimes\left(\sigma_0\tilde{I}_k^{-1}+(\mathbf{X'X})\right)\right)^{-1}\left(\left(\Sigma_e^{-1}\otimes\sigma_0\tilde{I}_k^{-1}\right)\phi_0+\left(\Sigma_e^{-1}\otimes(\mathbf{X'X})\right)\phi_{\mathbf{ml}}\right) \\ &= \left(\Sigma_e\otimes\left(\sigma_0\tilde{I}_k^{-1}+(\mathbf{X'X})\right)^{-1}\right)\left(\left(\Sigma_e^{-1}\otimes\sigma_0\tilde{I}_k^{-1}\right)\phi_0+\left(\Sigma_e^{-1}\otimes(\mathbf{X'X})\right)\phi_{\mathbf{ml}}\right) \\ &= \left(I_m\otimes\left(\sigma_0\tilde{I}_k^{-1}+\mathbf{X'X}\right)^{-1}\left(\sigma_0\tilde{I}_k^{-1}\right)\right)\phi_0+\left(I_m\otimes\left(\sigma_0\tilde{I}_k^{-1}+\mathbf{X'X}\right)^{-1}(\mathbf{X'X})\right)\phi_{\mathbf{ml}} \end{aligned}$$

B Proofs

Proposition 1 *The Markov structure of the evolution equation of any DLM implies*

$$\Pr\left(Z_T|\mathcal{D}_T\right) = \Pr\left(z_T|\mathcal{D}_T\right)\Pr\left(z_{T-1}|z_T,\mathcal{D}_{T-1}\right)...\Pr\left(z_1|z_2,\mathcal{D}_1\right)$$

Proof. Note that $\Pr\left(Z_T|\mathcal{D}_T\right)$ can be expanded as follows:

$$\begin{aligned} \Pr\left(Z_T|\mathcal{D}_T\right) &= \Pr\left(z_T,Z_{T-1}|\mathcal{D}_T\right) = \Pr\left(z_T|\mathcal{D}_T\right)\Pr\left(Z_{T-1}|z_T,\mathcal{D}_T\right) \\ &= \Pr\left(z_T|\mathcal{D}_T\right)\Pr\left(z_{T-1},Z_{T-2}|z_T,\mathcal{D}_T\right) \\ &= \Pr\left(z_T|\mathcal{D}_T\right)\Pr\left(z_{T-1}|z_T,\mathcal{D}_T\right)\Pr\left(Z_{T-2}|z_{T-1},z_T,\mathcal{D}_T\right) \\ &= \Pr\left(z_T|\mathcal{D}_T\right)\Pr\left(z_{T-1}|z_T,\mathcal{D}_T\right)\Pr\left(z_{T-2},Z_{T-3}|z_{T-1},z_T,\mathcal{D}_T\right) \\ &= \Pr\left(z_T|\mathcal{D}_T\right)...\underbrace{\Pr\left(z_{T-2}|z_{T-1},z_T,\mathcal{D}_T\right)}_{\Pr\left(z_{T-2}|z_{t>(T-2)},\mathcal{D}_T\right)}\Pr\left(Z_{T-3}|z_{T-2},z_{T-1},z_T,\mathcal{D}_T\right) \\ &\;\;\vdots \\ &= \Pr\left(z_T|\mathcal{D}_T\right)\Pr\left(z_{T-1}|z_{t>(T-1)},\mathcal{D}_T\right)\Pr\left(z_{T-2}|z_{t>(T-2)},\mathcal{D}_T\right)...\Pr\left(z_1|z_{t>1},\mathcal{D}_T\right) \end{aligned}$$

Let now $y^{ob}_{t>\bar{\tau}} \equiv \left[y^{ob}_{\bar{\tau}+1}, y^{ob}_T\right]$ and $z_{t>\bar{\tau}} \equiv \left[z_{\bar{\tau}+1}, z_T\right]$ denote the sets of all observable and state variables for periods between $\bar{\tau}+1$ up to T. Recall that the observation equation implies that, conditional on z_t, y^{ob}_t does not depend on past values of the state variable. As shown below, the Markovian property copped with this fact implies let us show that

$\Pr\left(z_{T-j}|z_{t>T-j}, \mathcal{D}_T\right) = \Pr\left(z_{T-j}|z_{T-j+1}, \mathcal{D}_{T-j}\right)$. Indeed:

$$
\begin{array}{rl}
\Pr\left(z_{T-j}|z_{t>T-j}\ ,\ \mathcal{D}_T\right) & = \Pr\left(z_{T-j}|z_{t>T-j}\ ,\ y^{ob}_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right) \\
& = \frac{\Pr\left(z_{T-j}\ ,\ y^{ob}_{t>T-j}|z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right)}{\Pr\left(\mathbf{y}^{ob}_{t>T-j}|z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right)} \\
& = \frac{\Pr\left(y^{ob}_{t>T-j}|z_{T-j}\ ,\ z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right)\Pr\left(z_{T-j}|z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right)}{\Pr\left(y^{ob}_{t>T-j}|z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right)} \\
\textit{Due to Observ Equation} & = \frac{\Pr\left(y^{ob}_{t>T-j}|z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right)\Pr\left(z_{T-j}|z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right)}{\Pr\left(y^{ob}_{t>T-j}|z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right)} \\
& = \Pr\left(z_{T-j}|z_{t>T-j}\ ,\ \mathcal{D}_{T-j}\right) \\
& = \Pr\left(z_{T-j}|z_{T-j+1}\ ,\ z_{t>T-j+1}\ ,\ \mathcal{D}_{T-j}\right) \\
& = \frac{\Pr\left(z_{T-j}\ ,\ z_{t>T-j+1}|z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)}{\Pr\left(z_{t>T-j+1}|z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)} \\
& = \frac{\Pr\left(z_{t>T-j+1}|z_{T-j}\ ,\ z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)\Pr\left(z_{T-j}|z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)}{\Pr\left(z_{t>T-j+1}|z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)} \\
\textit{Markovian Property} & = \frac{\Pr\left(z_{t>T-j+1}|z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)\Pr\left(z_{T-j}|z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)}{\Pr\left(z_{t>T-j+1}|z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)} \\
& = \Pr\left(z_{T-j}|z_{T-j+1}\ ,\ \mathcal{D}_{T-j}\right)
\end{array}
$$

Using our previous result, we conclude that

$$\Pr\left(Z_T|\mathcal{D}_T\right) = \Pr\left(z_T|\mathcal{D}_T\right)\Pr\left(z_{T-1}|z_T, \mathcal{D}_{T-1}\right)...\Pr\left(z_1|z_2, \mathcal{D}_1\right)$$

■

Proposition 2 *The conditional density of* $(y_{t-q+1}|z_{t+1}, \mathcal{D}_t)$ *is proportional to*

$$\Pr\left(y_{t-q+1}|z_{t+1}, \mathcal{D}_t\right) \propto \Pr\left(\tilde{e}_{t+1}|\tilde{z}_{t+1}, y_{t-q+1}, \mathcal{D}_t\right)\Pr\left(y_{t-q+1}|\tilde{z}_{t+1}, \mathcal{D}_t\right)$$

Proof.

$$
\begin{array}{rl}
\Pr\left(y_{t-q+1}|z_{t+1}, \mathcal{D}_t\right) & = \Pr\left(y_{t-q+1}|y_{t+1}, \tilde{z}_t, \mathcal{D}_t\right) \\
& = \frac{\Pr(y_{t+1},y_{t-q+1}|\tilde{z}_t,\mathcal{D}_t)}{\Pr(y_{t+1}|\tilde{z}_t,\mathcal{D}_t)} = \frac{\Pr(y_{t+1}|\tilde{z}_t,y_{t-q+1},\mathcal{D}_t)\Pr(y_{t-q+1}|\tilde{z}_t,\mathcal{D}_t)}{\Pr(y_{t+1}|\tilde{z}_t,\mathcal{D}_t)} \\
\text{conditioning first term also on } \tilde{e}_{t+1} & = \frac{\Pr(y_{t+1},\tilde{e}_{t+1}|\tilde{z}_t,y_{t-q+1},\mathcal{D}_t)}{\Pr(\tilde{e}_{t+1}|y_{t+1},\tilde{z}_t,y_{t-q+1},\mathcal{D}_t)}\frac{\Pr(y_{t-q+1}|\tilde{z}_t,\mathcal{D}_t)}{\Pr(y_{t+1}|\tilde{z}_t,\mathcal{D}_t)} \\
& = \frac{\Pr(y_{t+1}|\tilde{z}_t,y_{t-q+1},\tilde{e}_{t+1},\mathcal{D}_t)}{\Pr(\tilde{e}_{t+1}|y_{t+1},\tilde{z}_t,y_{t-q+1},\mathcal{D}_t)}\frac{\Pr(\tilde{e}_{t+1}|\tilde{z}_t,y_{t-q+1},\mathcal{D}_t)\Pr(y_{t-q+1}|\tilde{z}_t,\mathcal{D}_t)}{\Pr(y_{t+1}|\tilde{z}_t,\mathcal{D}_t)} \\
\text{conditioned on } \tilde{e}_{t+1}\text{, first term does not depend on } y_{t-q+1} & = \frac{\Pr(y_{t+1}|\tilde{z}_t,\tilde{e}_{t+1},\mathcal{D}_t)}{\Pr(\tilde{e}_{t+1}|y_{t+1},\tilde{z}_t,\mathcal{D}_t)}\frac{\Pr(\tilde{e}_{t+1}|\tilde{z}_t,y_{t-q+1},\mathcal{D}_t)\Pr(y_{t-q+1}|\tilde{z}_t,\mathcal{D}_t)}{\Pr(y_{t+1}|\tilde{z}_t,\mathcal{D}_t)}
\end{array}
$$

Since terms $\Pr\left(y_{t+1}|\tilde{z}_t, \tilde{e}_{t+1}, \mathcal{D}_t\right)$, $\Pr\left(\tilde{e}_{t+1}|y_{t+1}, \tilde{z}_t, \mathcal{D}_t\right)$ and $\Pr\left(y_{t+1}|\tilde{z}_t, \mathcal{D}_t\right)$ are not functions of y_{t-q+1}, they are constantans with respect to $\Pr\left(y_{t-q+1}|z_{t+1}, \mathcal{D}_t\right)$. Therefore, we obtain:

$$\Pr\left(y_{t-q+1}|z_{t+1}, \mathcal{D}_t\right) \propto \Pr\left(\tilde{e}_{t+1}|\tilde{z}_{t+1}, y_{t-q+1}, \mathcal{D}_t\right)\Pr\left(y_{t-q+1}|\tilde{z}_{t+1}, \mathcal{D}_t\right)$$

■