Journal of Econometrics 193 (2016) 433–437

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A computationally efficient method for vector autoregression with mixed frequency data

Hang Qian *

Iowa State University, USA

ARTICLE INFO

Article history:
Available online 3 May 2016

Keywords:
VAR
Kalman filter
Bayesian

ABSTRACT

A linear transformation method is proposed to handle the vector autoregression with mixed frequency time series data. Temporally aggregated observations impose linear constraints on the distribution of latent variables, which are converted such that each observation replaces a latent variable. Full-sample transformation yields a closed-form simulation smoother, while partial-sample transformation leads to a computationally efficient sampler suitable for parallel computing.

1. Introduction

A standard vector autoregression (VAR) model assumes that variables are sampled at the same frequency. However, economic data may be observed at mixed frequencies. For example, GDP data are quarterly, while many financial series are daily or more frequent. In addition, recent observations may have a higher frequency while historical data are coarsely sampled. For instance, quarterly GDP data are not available until 1947.

Mixed frequency regressions can be handled by the Kalman filter. Early studies of Jones (1980), Ansley and Kohn (1983) and Harvey and Pierse (1984) show the state space representations of the ARMA process subject to missing data and temporal aggregation. The idea can be extended to the VAR and dynamic factor models, which have been studied by Zadrozny (1988), Brockwell et al. (1991), Mariano and Murasawa (2003, 2010), Foroni et al. (2013), etc. Recent years have also seen Bayesian estimation of the mixed frequency VAR as in Viefers (2011) and Schorfheide and Song (2015), in which forward filtering and backward sampling is utilized for simulating the smoothed states.

Chiu et al. (2011) and Eraker et al. (2015) propose an innovative approach that does not rely on the Kalman filter. In their Gibbs sampler, latent variables are generated one at a time conditional on the neighboring states. Refer to Carter and Kohn (1994) for a comparison of the single-move sampler and the simulation smoother based on the Kalman filter.

Most mixed frequency VAR models have latent autoregressive processes. Ghysels (forthcoming) proposes a new VAR formulated exclusively in terms of observed mixed frequency data. Stacking or blocking (also see Filler, 2013) makes the observation matrix time-invariant. This model is also closely related to the MIDAS regression introduced by Ghysels et al. (2006) and Ghysels et al. (2007).

The main contribution of this paper is a linear transformation method that addresses frequency mismatch. Mixed frequency observations impose linear constraints on the distribution of the latent variables. The constraints can be absorbed by linear transformation such that each observation replaces a latent variable. Full-sample transformation yields the explicit form of the likelihood function and the simulation smoother. Partial-sample transformation leads to a computationally efficient sampler, whose required floating-point operations (FLOPS) are less than 1/10 of those for the Kalman filter. Moreover, the Kalman filter is inherently sequential, but the new method is suitable for parallel computing. A simulation study shows that our estimator is as accurate as the Kalman filter, but has a substantial speed advantage. In addition, our method differs from that of Eraker et al. (2015) in that it is not a single-move sampler and states of disjoint periods are generated simultaneously after linear transformation.

Our analysis is from the Bayesian perspective, but the analytic, non-recursive likelihood function can also be obtained by the proposed method, which may be useful for frequentist estimation. Parameter estimation is not further discussed in the reminder of the paper, because conditional on the augmented data (i.e., the smoothed states of the latent VAR series), the model is reduced to a single frequency VAR. Bayesian inference of a standard VAR model is well developed, usually under a conjugate normal-Wishart prior with coefficient shrinkage. Refer to Litterman (1986), Kadiyala and Karlsson (1997) and Banbura et al. (2010). This paper focuses on the smoothing problem: the posterior distribution of the latent VAR series conditional on the model parameters and observations.

* Correspondence to: 55 Centre Street, Natick, MA 01760, USA.
E-mail address: matlabist@gmail.com.

http://dx.doi.org/10.1016/j.jeconom.2016.04.016
0304-4076/

For brevity, conditioning on model parameters is implicit when we mention 'the posterior distribution' throughout the paper.

The reminder of the paper is structured as follows. Section 2 shows the analytic smoother obtained from the linear transformation approach. Section 3 refines the approach by partial-sample transformation, which yields a computationally efficient posterior sampler. Section 4 compares our approach with the Kalman filter and Section 5 concludes the paper.

2. Full-sample transformation

For notational convenience, we present the linear transformation approach using a bivariate VAR(1) model. Extension to multivariate VAR(p) models with irregular frequencies is feasible. Our software implements the VAR(p) cases and the simulation results are presented in Section 4.

Consider a stationary VAR(1) process:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \Phi \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + P \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \tag{1}$$

where $\varepsilon_{1,t}, \varepsilon_{2,t}$ are i.i.d. standard normal disturbances, whose loading matrix P is lower triangular (the Cholesky factor of the covariance matrix). Assume that the second series $\{y_t\}$ is fully observed, while the first series $\{x_t\}$ is observed every other period as the sum of two consecutive values:

$$\bar{x}_t = x_{t-1} + x_t, \quad t = 2, 4, \ldots, T. \tag{2}$$

By stationarity, $(x_t, y_t)' \sim N(0, \Sigma), \forall t$, where Σ satisfies the Lyapunov equation $\Sigma = \Phi \Sigma \Phi' + PP'$. It follows that $(x_1, y_1, \ldots, x_T, y_T)' \sim N(0, \Omega)$, where Ω is a $2T \times 2T$ matrix with 2×2 blocks taking the form $\Phi^i \Sigma$, $i = 0, 1, 2, \ldots$. For example, if $T = 4$, we have

$$\Omega = \begin{pmatrix} \Sigma & (\Phi\Sigma)' & \left(\Phi^2\Sigma\right)' & \left(\Phi^3\Sigma\right)' \\ \Phi\Sigma & \Sigma & (\Phi\Sigma)' & \left(\Phi^2\Sigma\right)' \\ \Phi^2\Sigma & \Phi\Sigma & \Sigma & (\Phi\Sigma)' \\ \Phi^3\Sigma & \Phi^2\Sigma & \Phi\Sigma & \Sigma \end{pmatrix}. \tag{3}$$

We want to replace the even-numbered latent variables $x_2, x_4, \ldots, x_T$ by the low frequency observations $\bar{x}_2, \bar{x}_4, \ldots \bar{x}_T$, and thus construct a transformation matrix

$$\widetilde{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{4}$$

and a block diagonal matrix $A = diag\left(\widetilde{A}, \ldots, \widetilde{A}\right)$ in which $\widetilde{A}$ repeats itself $\frac{T}{2}$ times. Linear transformation refers to

$$v_2 = A \cdot v_1, \tag{5}$$

where

$$v_1 = (x_1, y_1, x_2, y_2, \ldots, x_{T-1}, y_{T-1}, x_T, y_T)',$$
$$v_2 = (x_1, y_1, \bar{x}_2, y_2, \ldots, x_{T-1}, y_{T-1}, \bar{x}_T, y_T)'.$$

The vector v_2 contains three types of variables: (1) high frequency observations $Y \equiv (y_1, \ldots, y_T)'$, (2) low frequency observations $\overline{X} \equiv (\bar{x}_2, \bar{x}_4, \ldots, \bar{x}_T)'$, and (3) high frequency latent variables $x_1, x_3, \ldots, x_{T-1}$. The goal of smoothing is to compute the posterior distribution of the latent variables conditional on Y and $\overline{X}$. To this end, we may construct a $\frac{T}{2} \times 2T$ selection matrix E_0 obtained from the row $1, 5, 9, \ldots, 2T-3$ of a $2T \times 2T$ identity matrix, and a $\frac{3T}{2} \times 2T$ selection matrix E_1 consisting of the remaining rows of that identity matrix. Since $v_2 \sim N\left(0, A\Omega A'\right)$, by the formula of the conditional normal distribution, we obtain the analytic smoother:

$$(x_1, x_3, \ldots, x_{T-1})' \left| \left(Y', \overline{X}'\right)' \sim N\left(\overline{\mu}, \overline{\Omega}\right), \right. \tag{6}$$

where

$$\overline{\mu} = \left(E_0 A\Omega A' E_1'\right) \cdot \left(E_1 A\Omega A' E_1'\right)^{-1} \cdot (y_1, \bar{x}_2, y_2, \ldots, y_{T-1}, \bar{x}_T, y_T)', \tag{7}$$

$$\overline{\Omega} = E_0 A\Omega A' E_0' - \left(E_0 A\Omega A' E_1'\right) \cdot \left(E_1 A\Omega A' E_1'\right)^{-1} \cdot \left(E_1 A\Omega A' E_0'\right). \tag{8}$$

The $\frac{3T}{2} \times \frac{3T}{2}$ matrix $E_1 A\Omega A' E_1'$ is invertible because the rank of the matrix E_1, A, Ω is $\frac{3T}{2}, 2T, 2T$, respectively.

Note that latent variables in the model also include $x_2, x_4, \ldots, x_T$, which have been removed during the linear transformation. However, their posterior distributions are degenerate conditional on $x_1, x_3, \ldots, x_{T-1}$ and their posterior draws can be restored by the identity $x_2 = \bar{x}_2 - x_1, \ldots, x_T = \bar{x}_T - x_{T-1}$.

An alternative interpretation of the degenerate sampler is joint posterior sampling for $x_1, \ldots, x_T$ subject to $T/2$ linear constraints. That is, we generate draws from a T dimensional multivariate normal distribution whose covariance matrix has the rank $T/2$.

Full sample transformation leads to the closed-form simulation smoother. In addition, we have obtained the explicit likelihood function (i.e., the joint distribution of Y and $\overline{X}$), which is simply the density of $N\left(0, E_1 A\Omega A' E_1'\right)$. This is in contrast with the Kalman filter that generates the likelihood function and the smoother in a recursive manner.

3. Partial-sample transformation

Full-sample transformation sheds light on the analytic properties of the mixed frequency VAR models, but it is computationally intensive. However, with slight modification the linear transformation approach can be computationally efficient. Instead of generating posterior draws for $x_1, x_3, \ldots, x_{T-1}$ simultaneously, we divide them into two blocks: $X_A \equiv (x_1, x_5, x_9, \ldots, x_{T-3})'$ and $X_B \equiv (x_3, x_7, x_{11}, \ldots, x_{T-1})'$. We alternately generate draws for X_A conditional on $X_B, Y, \overline{X}$, and then draw X_B conditional on $X_A, Y, \overline{X}$.

On the surface, we need a multivariate random number generator for X_A (and X_B as well), but it is valid to sample the component variables individually from univariate normal distributions, because they are mutually independent conditional on $X_B, Y, \overline{X}$. Formally, we have

Proposition 1.

$$p\left(X_A \left| X_B, Y, \overline{X}\right.\right) = p\left(x_1 \left| y_1, \bar{x}_2, y_2, x_3, y_3\right.\right) \cdot \prod_{t=\{5,9,\ldots,T-3\}} p\left(x_t \left| x_{t-1}, y_{t-1}, y_t, \bar{x}_{t+1}, y_{t+1}, x_{t+2}, y_{t+2}\right.\right), \tag{9}$$

$$p\left(X_B \left| X_A, Y, \overline{X}\right.\right) = p\left(x_{T-1} \left| x_{T-2}, y_{T-2}, y_{T-1}, \bar{x}_T, y_T\right.\right) \cdot \prod_{t=\{3,7,\ldots,T-5\}} p\left(x_t \left| x_{t-1}, y_{t-1}, y_t, \bar{x}_{t+1}, y_{t+1}, x_{t+2}, y_{t+2}\right.\right). \tag{10}$$

A proof is provided in the Appendix. The conditioning variables for each term on the right hand side of Eq. (9) can be equivalently written in the verbose form: $X_B, Y, \overline{X}$, hence conditional independence. When we sample for x_t, conditioning on x_{t+2}, y_{t+2} cannot be dropped as the univariate sampler can be interpreted as joint sampling for x_t, x_{t+1} subject to a linear constraint.

Conditional independence facilitates parallel computing, as multiple computational threads can simultaneously sample elements in X_A (and X_B). In contrast, the Kalman filter is inherently

sequential and the sampling order is forward filtering (from period 1 to T) and then backward smoothing/sampling (from period T to 1). Under such recursion, we cannot sample states of disjoint periods, say $x_1, x_5, x_9, \ldots$, in parallel.

We may use a partial-sample transformation method to generate univariate draws for $p\left(x_t \left| x_{t-1}, y_{t-1}, y_t, \bar{x}_{t+1}, y_{t+1}, x_{t+2}, y_{t+2}\right.\right)$ in Proposition 1.

First, construct a transformation matrix

$$A = diag\left(I_2, \widetilde{A}, I_2\right), \tag{11}$$

where I_2 is a 2×2 identity matrix and $\widetilde{A}$ is defined in Eq. (4). The purpose is to replace x_{t+1} by $\bar{x}_{t+1}$. Similar to Eq. (5), we have $v_{2,t} = A \cdot v_{1,t}$, where

$$\begin{aligned} v_{1,t} &= (x_{t-1}, y_{t-1}, x_t, y_t, x_{t+1}, y_{t+1}, x_{t+2}, y_{t+2})', \\ v_{2,t} &= (x_{t-1}, y_{t-1}, x_t, y_t, \bar{x}_{t+1}, y_{t+1}, x_{t+2}, y_{t+2})'. \end{aligned} \tag{12}$$

Second, in order to separate the latent and observed variables in $v_{2,t}$, we construct two selection matrices:

$$E_0 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{13}$$

and E_1 is obtained from a 8×8 identity matrix with its third row deleted. Since $v_{2t} \sim N\left(0, A\Omega A'\right)$, where Ω is given by Eq. (3), we have

$$x_t \left| x_{t-1}, y_{t-1}, y_t, \bar{x}_{t+1}, y_{t+1}, x_{t+2}, y_{t+2} \right. \sim N\left(\overline{\mu}_t, \overline{\Omega}\right), \tag{14}$$

where $\overline{\Omega}$ takes the same form as that in Eq. (8), and

$$\begin{aligned} \overline{\mu}_t = &\left(E_0 A\Omega A' E_1'\right) \cdot \left(E_1 A\Omega A' E_1'\right)^{-1} \\ &\cdot (x_{t-1}, y_{t-1}, y_t, \bar{x}_{t+1}, y_{t+1}, x_{t+2}, y_{t+2})'. \end{aligned} \tag{15}$$

The degenerate x_{t+1} can be restored by the identity $x_{t+1} = \bar{x}_{t+1} - x_t$.

In summary, the posterior series $x_1, \ldots, x_T$ are simulated in this order: (1) draw $x_1, x_5, \ldots, x_{T-3}$ individually, possibly with parallel computing; (2) restore $x_2, x_6, \ldots, x_{T-2}$; (3) draw $x_3, x_7, \ldots, x_{T-1}$; and (4) restore $x_4, x_8, \ldots, x_T$. Fig. 1 illustrates the sampling method by blocks.

Note that the matrices $\Omega, A, E_0, E_1, \overline{\Omega}$ are time-invariant; only the posterior mean $\overline{\mu}_t$ changes over time. Our method requires less computation compared with the Kalman filter. To analyze computational complexity, we depart from the bivariate case and consider a n dimensional VAR model with T observations. Since the complexity of matrix–matrix multiplication, inversion and the Cholesky decomposition are $O(n^3)$, the Kalman simulation smoother requires $O(n^3T)$ operations to draw a path of $x_1, \ldots, x_T$. For the partial-sample transformation algorithm, computing $\Omega, A, E_0, E_1, \overline{\Omega}$ has a one-time cost of $O(n^3)$ operations, while computing $\overline{\mu}_t$ for $t = 1, \ldots, T$ only involves matrix–vector multiplication, hence $O(n^2T)$ operations. Our total cost $O(n^3) + O(n^2T)$ is much smaller than $O(n^3T)$, and Section 4 provides numerical evidence on that assertion.

The Kalman filter iterations may converge to periodic values (as the observation equation is time-varying), and one may take advantage of the periodic steady-state Kalman gain. We did not take that shortcut for two reasons. First, convergence is an asymptotic property, and the solution is approximated if the iterations are truncated. In contrast, the covariance matrix $\overline{\Omega}$ in Eq. (14) is time-invariant except for the first and last period (due to missing presample and postsample observations). Our result is not an approximation. Second, if there are changes in sampling frequency or other sources of missing data, convergence of the Kalman filter is lost and the complexity is necessarily $O(n^3T)$. However, such occasional irregularity only has a local impact on our sampler, as we modify Eq. (14) only for the irregular periods. Therefore, the computational complexity remains $O(n^3) + O(n^2T)$.

One may argue that the proposed sampler requires extra computation due to burn-in draws, while the Kalman filter does not. This is not true because in practice we do not know the VAR coefficients. Even if we use the Kalman simulation smoother, we will alternately sample the posterior states and the VAR coefficients, so the burn-in draws cannot be avoided.

4. Simulation study

This section is devoted to numerical comparison of the Kalman filter (KF hereafter) and the linear transformation (LT) method for estimating mixed frequency VAR models. We consider four aspects of numerical performance: estimator accuracy, computational complexity, computing time and sampler quality. Estimator accuracy is evaluated by the discrepancy between the estimated VAR coefficients and the 'true values', and between KF and LT with the same dataset. Computational complexity is measured by FLOPS. Both a running count by MATLAB 5 and an estimation of FLOPS by parsing codes are reported in Table 1. Computing time is a direct measure of speed, though it depends on the hardware and software platforms. Sampler quality refers to the degree of mixing measured by the effective sample size (ESS). See Kass et al. (1998).

$$ESS = \frac{M}{1 + 2\sum_{j=1}^{M}\left(1 - \frac{j}{M}\right)\rho_j},$$

where M is the number of draws and ρ_j is the autocorrelation of lag j.

Conditional on the VAR coefficients, KF and LT are two simulation smoothing algorithms that generate latent variables. Conditional on the latent variables, we use the standard Bayesian VAR techniques to sample the VAR coefficients.

We first consider the bivariate VAR(1) model specified by Eqs. (1) and (2) with

$$\Phi = \begin{pmatrix} 0.5 & 0.4 \\ 0.3 & 0.6 \end{pmatrix}, \qquad P = \begin{pmatrix} 0.9 & 0 \\ 0.8 & 0.7 \end{pmatrix}, \qquad T = 1000.$$

We generate latent variables by Eq. (1), and then aggregate the first series by Eq. (2). Using the mixed frequency data, we employ both KF and LT to estimate the VAR model. Both samplers take 5000 draws with the first half burned in. For stability of results, the experiment is repeated for 100 times; each time with a new set of mixed frequency data. The posterior mean is treated as the Bayesian estimator with the standard deviation in parenthesis.

The last two columns of Table 1 report the KF and LT estimators. Both are reasonably close to the true values with discrepancies less than 11%. Such discrepancies are induced by sampling variations instead of algorithms, because the discrepancies between KF and LT are less than 1.3%. Since we have generated 100 different datasets, the average estimators should remove sampling variations and converge to the true values. The middle columns of Table 1 show that both KF and LT can recover the true values within 2.5% tolerance. The differences between KF and LT are close to zero. Based on the Monte Carlo evidence, we claim that LT is as accurate as KF.

Given the same accuracy, KF is overshadowed by LT in terms of computational complexity and time. LT only costs 1/10 of the FLOPS compared with KF and the computing time is roughly proportional to the algorithm complexity, that is, LT is 10 times faster. As for the sampler quality, there is no evidence that LT draws are inferior in terms of autocorrelation.

We also consider three scenarios of multivariate VAR(p) models, in which the first series is aggregated such that $\bar{x}_t = x_t + x_{t-1} + x_{t-2}$, $t = 3, 6, \ldots, T$. The three scenarios are 3-variable VAR(2), 6-variable VAR(3) and 12-variable VAR(3). We simulate

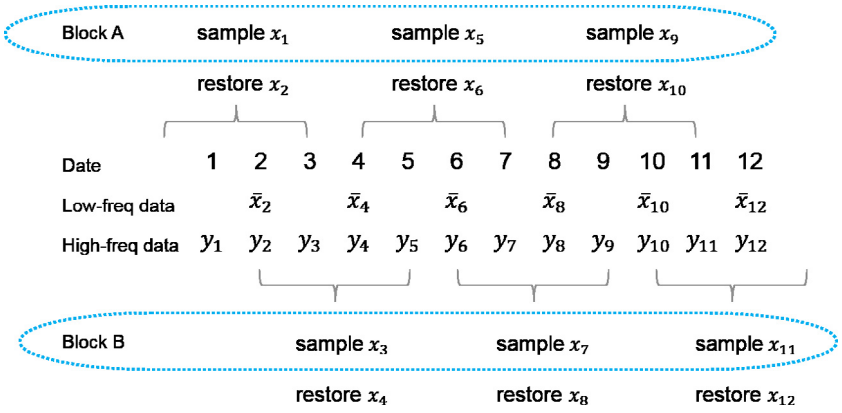

Fig. 1. Illustration of partial-sample transformation sampler. The figure illustrates a bivariate VAR(1) model in which the first series is observed every other period as the sum of two consecutive values. Mixed frequency observations of the first 12 periods are depicted in the rows 'High-freq data' and 'Low-freq data'. We alternately sample variables in Blocks A and B. Block A consists of $x_1, x_5, x_9, \ldots$, and we sample them individually from univariate normal distributions, possibly by parallel computing. Then the companion series $x_2, x_6, x_{10}, \ldots$ can be restored, using $x_{t+1} = \bar{x}_{t+1} - x_t$. Similarly, variables in Block B, namely $x_3, x_7, x_{11}, \ldots$, are drawn conditional on Block A, and then $x_4, x_8, x_{12}, \ldots$ can be restored.

Table 1
Bivariate mixed frequency VAR(1) estimation.

		100 Experiments avg		1 Experiment	
	True	KF	LT	KF	LT
ϕ_{11}	0.5	0.488	0.489	0.511	0.512
		(0.060)	(0.059)	(0.060)	(0.058)
ϕ_{12}	0.4	0.410	0.410	0.370	0.369
		(0.056)	(0.056)	(0.055)	(0.053)
ϕ_{21}	0.3	0.307	0.307	0.266	0.263
		(0.059)	(0.059)	(0.062)	(0.064)
ϕ_{22}	0.6	0.594	0.594	0.630	0.633
		(0.055)	(0.055)	(0.056)	(0.058)
p_{11}	0.9	0.897	0.897	0.876	0.878
		(0.029)	(0.029)	(0.028)	(0.028)
p_{21}	0.8	0.788	0.788	0.801	0.804
		(0.037)	(0.037)	(0.035)	(0.035)
p_{22}	0.7	0.708	0.708	0.662	0.661
		(0.026)	(0.026)	(0.026)	(0.025)
Runtime FLOPS		4.3 e9	3.7 e8	4.3 e9	3.7 e8
Parsed FLOPS		4.2 e9	4.1 e8	4.2 e9	4.1 e8
CPU time (s)		334	31	334	30
Effective sample size		669	689	560	825

The mixed frequency model is specified by Eqs. (1) and (2). The third and fourth columns report the Kalman filter and the linear transformation estimators averaged by 100 datasets, each of which has 1000 observations. The last two columns correspond to the results under a single dataset. Runtime FLOPS are obtained from MATLAB 5, and parsed FLOPS are estimated by our program.

Table 2
Multivariate mixed frequency VAR(p) estimation.

	3D VAR(2)		6D VAR(3)		12D VAR(3)	
	KF	LT	KF	LT	KF	LT
$\lvert\Phi_{True} - \widehat{\Phi}_{KF/LT}\rvert$	0.049	0.051	0.024	0.024	0.018	0.018
$\lvert\widehat{\Phi}_{KF} - \widehat{\Phi}_{LT}\rvert$	0.005	0.005	0.002	0.002	0.001	0.001
CPU time (s)	143	18	892	60	7057	870
Effective sample size	438	452	863	898	1422	1439

The columns show the Kalman filter and the linear transformation estimators in three scenarios: 3-variable VAR(2), 6-variable VAR(3) and 12-variable VAR(3), respectively. The first variable is temporally aggregated. The third row reports the discrepancies between the Kalman filter (linear transformation) estimators and the true VAR coefficients. The fourth row shows the discrepancy between the Kalman filter and linear transformation estimators. Note that the VAR coefficient Φ contains many elements. We compute the absolute value of the discrepancy for each element and then take the average. CPU time is measured on a personal computer with optimized MATLAB codes with mex files and vectorization techniques. The optimized codes are faster than the codes for the bivariate VAR(1).

3000 mixed frequency observations and then use KF and LT to estimate the model with 10,000 posterior draws. The experiment is repeated for ten times with new datasets, and the average results are presented in Table 2. To save space, we only report the average discrepancies between (1) KF estimators and the true VAR coefficients, (2) LT estimators and the true values, and (3) KF and LT estimators. Codes are optimized for maximum speed with mex files and vectorization. The optimized codes cannot run on MATLAB 5 and FLOPS are not available.

As we can see in Table 2, the discrepancies from the true values mostly stem from sampling variations. With the same dataset, the KF and LT estimators are nearly identical. However, the computing time of KF is 7.9, 14.9 and 8.1 times longer than that of LT in the three scenarios.

5. Conclusion

The Kalman filter provides a convenient solution to the mixed frequency VAR model. However, it is not necessarily the best approach for two reasons. First, the filter/smoother covariance matrix recursion is computationally intensive. Second, the Kalman filter uses observations one at a time and is inherently sequential. The proposed partial-sample linear transformation approach overcomes those two drawbacks. The covariance matrix of the smoothing distribution in Eq. (14) is time-invariant if there are no other sources of missing observations. Even if such missing data exist, they only have a local impact on the sampler. Furthermore, the sampling order is flexible and parallel computing within a Gibbs sampler is feasible. Monte Carlo experiments confirm the superior performance of the new sampler.

Acknowledgments

We would like to thank Eric Ghysels, Brent Kreider, Lucrezia Reichlin, Gray Calhoun, and anonymous referees for helpful comments on this paper.

Appendix. Proof of Proposition 1

Let $z_t = (x_t, y_t)'$, and $\{z_t\}$ follow a bivariate VAR(1) process.

Lemma 1. *For $s < t < u - 1$, we have*

$$p\left(z_t, z_{t+1} \,|z_s, \ldots, z_{t-1}, z_{t+2}, \ldots, z_u\right) \tag{A.1}$$

$$= p\left(z_t, z_{t+1} \,|z_{t-1}, z_{t+2}\right). \tag{A.2}$$

Proof.

(A.1)

$$\propto p(z_s)\prod_{\tau=s+1}^{u} p(z_\tau\,|z_{\tau-1}) \tag{A.3}$$

$$\propto \prod_{\tau=t}^{t+2} p(z_\tau\,|z_{\tau-1}) \tag{A.4}$$

$\propto$ (A.2).

Since both (A.1) and (A.2) integrate to one w.r.t. z_t, z_{t+1}, proportionality indicates equality. ■

Lemma 2. *Let $\alpha, \beta, \gamma, \delta$ be random vectors. If $p(\alpha\,|\beta,\gamma,\delta) = p(\alpha\,|\beta)$, then $p(\alpha\,|\beta,\delta) = p(\alpha\,|\beta)$.*

Proof. $p(\alpha\,|\beta,\delta) = \frac{\int p(\alpha,\beta,\gamma,\delta)d\gamma}{p(\beta,\delta)} = \frac{\int p(\beta,\gamma,\delta)p(\alpha|\beta)d\gamma}{p(\beta,\delta)} = p(\alpha\,|\beta)$. ■

Lemma 1 indicates that z_{t-1}, z_{t+2} in (A.1) are the essential conditioning variables, while $z_s, \ldots, z_{t-2}, z_{t+3}, \ldots, z_u$ are redundant. Lemma 2 suggests that any subset of $z_s, \ldots, z_{t-2}, z_{t+3}, \ldots, z_u$ are redundant as well.

Lemma 3.

$$p\left(z_1, z_2, z_5, z_{6,}\ldots, z_{T-3}, z_{T-2}\,\middle|z_3, z_4, z_7, z_{8,}\ldots, z_{T-1}, z_T\right) \tag{A.5}$$

$$= p(z_1, z_2\,|z_3)\cdot\prod_{t=\{5,9,\ldots,T-3\}} p(z_t, z_{t+1}\,|z_{t-1}, z_{t+2}). \tag{A.6}$$

Proof. We first decompose (A.5) as the product of the marginal density and conditional densities, and then use Lemmas 1 and 2 to simplify expressions. Let $w = \left(z_3, z_4, z_7, z_{8,}\ldots, z_{T-1}, z_T\right)'$.

(A.5)

$$\propto p(z_1, z_2\,|w)\cdot\prod_{t=\{5,9,\ldots,T-3\}} p(z_t, z_{t+1}\,|z_{t-4}, z_{t-3}, \ldots, z_1, z_2, w) \tag{A.7}$$

$=$ (A.6). ■

Proposition 1.

$$p\left(X_A\,\middle|X_B, Y, \overline{X}\right) \tag{A.8}$$

$$= p(x_1\,|y_1, \bar{x}_2, y_2, x_3, y_3)\times\prod_{t=\{5,9,\ldots,T-3\}} p(x_t\,|x_{t-1}, y_{t-1}, y_t, \bar{x}_{t+1}, y_{t+1}, x_{t+2}, y_{t+2}). \tag{A.9}$$

Proof. Recall Eq. (4) where the transformation matrix $\widetilde{A}$ is constructed so that the aggregated observations and latent variables are connected by Eq. (5). We use the change of variable technique to compute the densities:

(A.8)

$$\propto p\left(X_A, X_B, Y, \overline{X}\right) \tag{A.10}$$

$$= \left\{\det\left[diag\left(\widetilde{A}, \ldots \widetilde{A}\right)\right]\right\}^{-1}\cdot p(z_1, \ldots, z_T) \tag{A.11}$$

$$\propto \left[\det\left(\widetilde{A}\right)\right]^{-T/2}\cdot \text{(A.5)} \tag{A.12}$$

$$= \left[\det\left(\widetilde{A}\right)\right]^{-T/2}\cdot \text{(A.6)} \tag{A.13}$$

$$\propto p(x_1, y_1, \bar{x}_2, y_2\,|z_3)\cdot\prod_{t=\{5,9,\ldots,T-3\}} p(x_t, y_t, \bar{x}_{t+1}, y_{t+1}\,|z_{t-1}, z_{t+2}) \tag{A.14}$$

$\propto$ (A.9).

Since both (A.8) and (A.9) integrate to one w.r.t. X_A, proportionality indicates equality. ■

References

Ansley, C.F., Kohn, R., 1983. Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. Biometrika 70 (1), 275–278.

Banbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector auto regressions. J. Appl. Econometrics 25 (1), 71–92.

Brockwell, P.J., Davis, R.A., Salehi, H., 1991. A state-space approach to transfer-function modeling. In: Statistical Inference in Stochastic Processes. Marcel Dekker, New York.

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

Chiu, C., Eraker, B., Foerster, A., Kim, T.B., Seoane, H., 2011. Estimating VAR sampled at mixed or irregular spaced frequencies: A Bayesian approach (manuscript).

Eraker, B., Chiu, C.W. J., Foerster, A.T., Kim, T.B., Seoane, H.D., 2015. Bayesian mixed frequency VARs. J. Financ. Econom. 13 (3), 698–721.

Filler, A., 2013. Generalized Dynamic Factor Models: Structure Theory and Estimation for Single and Mixed Frequency Data (Ph.D. thesis), Vienna University of Technology, Austria.

Foroni, C., Ghysels, E., Marcellino, M., 2013. Mixed-frequency vector autoregressive models. VAR Models in Macroeconomics—New Developments and Applications: Essays in Honor of Christopher A. Sims 13, 247–272.

Ghysels, E., 2012. Macroeconomics and the reality of mixed frequency data. J. Econometrics (forthcoming).

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. J. Econometrics 131 (1–2), 59–95.

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

Harvey, A.C., Pierse, R.G., 1984. Estimating missing observations in economic time series. J. Amer. Statist. Assoc. 79 (385), 125–131.

Jones, R.H., 1980. Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics 22 (3), 389–395.

Kadiyala, K.R., Karlsson, S., 1997. Numerical methods for estimation and inference in Bayesian VAR-models. J. Appl. Econometrics 12 (2), 99–132.

Kass, R.E., Carlin, B.P., Gelman, A., Neal, R.M., 1998. Markov chain Monte Carlo in practice: A roundtable discussion. Amer. Statist. 52 (2), 93–100.

Litterman, R.B., 1986. Forecasting with Bayesian vector autoregressions-five years of experience. J. Bus. Econom. Statist. 4 (1), 25–38.

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

Mariano, R.S., Murasawa, Y., 2010. A coincident index, common factors, and monthly real GDP. Oxf. Bull. Econ. Stat. 72 (1), 27–46.

Schorfheide, F., Song, D., 2015. Real-time forecasting with a mixed-frequency VAR. J. Bus. Econom. Statist. 33 (3), 366–380.

Viefers, P., 2011. Bayesian inference for the mixed-frequency VAR model. Discussion Papers of DIW Berlin 1172.

Zadrozny, P., 1988. Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. Econometric Theory 4 (01), 108–124.