
Munich Personal RePEc Archive

Vector Autoregression with Mixed Frequency Data

Qian, Hang

June 2013

Online at https://mpra.ub.uni-muenchen.de/47856/
MPRA Paper No. 47856, posted 27 Jun 2013 04:29 UTC

Vector Autoregression with Mixed Frequency Data

Hang Qian[1]

The MathWorks, Inc.
First Draft: 10/2011
This Draft: 06/2013

Abstract

Three new approaches are proposed to handle mixed frequency Vector Autoregression. The first is an explicit solution to the likelihood and posterior distribution. The second is a parsimonious, time-invariant and invertible state space form. The third is a parallel Gibbs sampler without forward filtering and backward sampling. The three methods are unified since all of them explore the fact that the mixed frequency observations impose linear constraints on the distribution of high frequency latent variables. By a simulation study, different approaches are compared and the parallel Gibbs sampler outperforms others. A financial application on the yield curve forecast is conducted using mixed frequency macro-finance data.

Keywords: VAR, Temporal aggregation, State space, Parallel Gibbs sampler

[☆]Previously titled: Vector Autoregression with Varied Frequency Data

[1]We would like to thank Eric Ghysels, Brent Kreider, Lucrezia Reichlin, Gray Calhoun, Timothy Fuerst for helpful comments on this paper. Corresponding author: 3 Apple Hill Drive, the MathWorks, Inc., Natick, MA 01760. Email: matlabist@gmail.com

1. Introduction

A standard Vector Autoregression (VAR) model assumes that data are sampled at the same frequency since variables at date t are regressed on variables dated at $t-1, t-2$, etc. However, economic and financial data may be sampled at varied frequencies. For example, GDP data are quarterly, while many financial variables might be daily or more frequent. In addition, for a given variable, recent data can be observed at a higher frequency while historical data are coarsely sampled. For instance, quarterly GDP data are not available until 1947.

In the presence of mixed frequency data, a VAR practitioner usually aligns variables either downward by aggregating the data to a lower frequency or upward by interpolating the high frequency data with heuristic rules such as polynomial fillings. Downward alignment discards valuable information in the high frequency data. Furthermore, temporal aggregation can change the lag order of ARMA models (Amemiya and Wu, 1972), reduce efficiency in parameter estimation and forecast (Tiao and Wei, 1976), affect Granger-causality and cointegration among component variables (Marcellino, 1999), induce spurious instantaneous causality (Breitung and Swanson, 2002), and so on. Silvestrini and Veredas (2008) provide a comprehensive review on the theory of temporal aggregation. On the other hand, upward alignment on the basis of *ad hoc* mathematical procedures is also problematic. Pavia-Miralles (2010) surveys various methods of interpolating and extrapolating time series. The problem is that by using a VAR model we assume high frequency data are generated by that model. However, interpolation is not based on the multivariate model that generates the data, but on other heuristic rules,

which inevitably introduce noises, if not distortion, to the data.

This paper focuses on VAR models that explicitly include heterogeneous frequency data. In the literature, there are several directions to the mixed frequency VAR modeling: state space solution, observed-data VAR and an alternative Gibbs sampler without the Kalman filter. We review these methods and describe the contribution of this paper.

First, the state space model (SSM) can bridge frequency mismatch and the Kalman filter yields the likelihood function and the posterior state distribution in a recursive form. The seminal paper of Harvey and Pierse (1984) outlines an SSM of the ARMA process subject to temporal aggregation. High frequency variables of adjacent periods are stacked in the state vector, whose linear combinations constitute observed mixed frequency data. This idea can be easily extended to the VAR and dynamic factor models, which have been explored by Zadrozny (1988), Mittnik and Zadrozny (2004), Mariano and Murasawa (2003, 2010), Hyung and Granger (2008). With the aid of the Kalman filter, a mixed frequency model can be estimated either by numerical maximum likelihood or the expectation-maximization algorithm. Recent years have also seen Bayesian estimation of the mixed frequency VAR as in Viefers (2011) and Schorfheide and Song (2012), who apply the forward filtering and backward sampling (FFBS) for data augmentation. Refer to Carter and Kohn (1994) for the original FFBS and Durbin and Koopman (2002), Chan and Jeliazkov (2009), among others, for improved FFBS algorithms.

It is tempting to think that an SSM is employed because direct likelihood evaluation of the mixed frequency VAR model is not obvious and the Kalman filter offers a recursive solution. We show that the mixed frequency

VAR model has analytic likelihood and posterior distribution of states if we interpret mixed frequency observations as linear constraints on the distribution of high frequency latent variables, but it is the computation that makes the state space solution numerically attractive. However, our explicit form sheds light on a high-performance parallel Gibbs sampler that explores the same idea but works on smaller observation chunks.

Second, a VAR system can be built on observed mixed frequency data, without resorting to latent variables. Proposed by Ghysels (2012), this new VAR is closely related to the Mi(xed) Da(ta) S(ampling), or MIDAS, regression introduced by Ghysels et al. (2006), Ghysels et al. (2007). The VAR system includes both a MIDAS regression (projecting high frequency data onto low frequency data with tightly parameterized weights) and autoregressions of observed high frequency variables plus low-frequency regressors. The coexistence of mixed frequency data in a VAR is achieved by rewriting a regular VAR model as a stacked and skip-sampled form.

Inspired by this new VAR representation, we propose a stick-and-skip SSM within the latent variable VAR framework. It is more parsimonious than the existing SSMs in that skip-sampling effectively shortens the recursion periods of the Kalman filter. More importantly, our SSM observation equation is time-invariant, non-cyclical, and not padded with artificial data points. It is as standard as a usual ARMA state space form, and thus readily applicable on any statistical software that supports SSM. Another advantage of our SSM is that the predicted state covariance matrix is invertible, which is not true for the existing SSMs for mixed frequency regression. Therefore, our SSM is suitable for the classical state smoother or simulation smoother

that requires inverting the state covariance matrix.

Third, given the latent variable framework, the state space form is not the unique solution to data augmentation. Chiu et al. (2011) propose a non-FFBS Gibbs sampler in which a single-period (say, period t) latent variables are drawn conditional on all other latent values. In a VAR(1) model, two neighbors (that is, values in periods $t-1$ and $t+1$) are relevant. Though this is an innovative approach handling mixed frequency data, it works under an assumption that low frequency data are the result of sampling every m periods from the high frequency variable. This might be appropriate for some types of stock variables. For example, the daily S&P 500 index can be thought as the closing price. However, some other stock variables such as monthly CPI might be more reasonably viewed as an average of the latent "weekly CPI" in a month, or there is an ambiguity on whether it reflects the price of the first or last week of a month. Similarly, flow variables such as quarterly GDP are the sum of the latent "monthly GDP" in a quarter.

Our parallel Gibbs sampler overcomes that limitation and accommodates the sampler of Chiu et al. (2011) as a special case. Low frequency observations are linear combinations of high frequency latent variables. If an observation binds several high frequency data as their sum or average, it is a temporal aggregation problem. If it binds only one high frequency data point, it reduces to the algorithm of Chiu et al. (2011).

Note that the FFBS draws of all latent variables as a whole by decomposing the joint distribution into cumulative conditionals, while any non-FFBS sampler increases the chain length of the Markov Chain Monte Carlo (MCMC). However, it does not necessarily imply that a non-FFBS sampler

is inferior. The FFBS is inherently sequential because the Kalman filter and smoother are computed recursively. However, our Gibbs sampler as well as that of Chiu et al. (2011) have an attractive feature that parallel computation can be performed within a single MCMC chain, since it satisfies the blocking property. For reasons that will be explained in Section 5, the parallel Gibbs sampler only moderately increases the correlation of draws, but substantially accelerates the sampler even on a personal computer. Perhaps a good way to describe the speed of our parallel sampler is that it takes longer time to sample the posterior VAR coefficients than the latent high frequency data.

In presenting our approaches, we do not label ourselves as frequentists or Bayesians, for our explicit solution and stick-and-skip SSM can be applied to both maximum likelihood and Bayesian inference. Also, for the sake of exposition, we first describe a bivariate evenly mixed frequency VAR(1), and then extend the method to a higher-order, higher-dimension VAR subject to arbitrary temporal aggregation.

The rest of the paper is organized as follows. Section 2 discusses the explicit likelihood and posterior states estimation of the mixed frequency VAR model. Section 3 introduces the stick-and-skip SSM in contrast to the standard state space solution. Section 4 shows the connections between the explicit solution and the recursive Kalman filter. Section 5 explains how the idea of the explicit solution can be adapted to a parallel Gibbs sampler. Section 6 conducts a simulation exercise to compare the sampling speed and efficiency of different mixed frequency VAR approaches. Section 7 applies the parallel Gibbs sampler to the yield curve forecast with mixed frequency macro-finance data. Section 8 concludes the paper.

2. The Explicit Solution

To fix ideas, first consider a bivariate stationary VAR(1) model

$$\begin{pmatrix} x^*_{1,t} \\ x^*_{2,t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} x^*_{1,t-1} \\ x^*_{2,t-1} \end{pmatrix} + \begin{pmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad (1)$$

where $\varepsilon_{1,t}, \varepsilon_{2,t}$ are independent standard normal noises.

To introduce mixed frequency data, assume the first series is fully observed, while the second series is temporally aggregated every other period. In other words, $x_{1,t} = x^*_{1,t}$, for $t = 1, \ldots, T$, while

$$x_{2,t} = \begin{cases} NaN & t = 1, 3, \ldots, T-1 \\ x^*_{2,t-1} + x^*_{2,t} & t = 2, 4, \ldots, T \end{cases}.$$

If the average, instead of sum, of the high frequency data are observed, rescale $x_{2,t}$ by a constant.

We are interested in the likelihood function for classical inference as well as the posterior distribution of latent high frequency data for Bayesian inference. For conciseness, throughout the paper, conditioning on model parameters is implicit when we mention "posterior distribution". We only discuss posterior latent variables, since posterior model parameters conditional on augmented data follow well-developed Bayesian VAR approaches. Refer to Litterman (1986); Kadiyala and Karlsson (1997); Banbura et al. (2010), among others.

The explicit solution can be obtained by the following three steps.

First, with the obvious vector (matrix) notation, rewrite Eq (1) in the matrix form

$$x^*_t = \Phi x^*_{t-1} + P\varepsilon_t.$$

Suppose x_1^* comes from the stationary distribution $N(0,\Omega)$, where Ω satisfies the Lyapunov Equation $\Omega = \Phi\Omega\Phi' + PP'$. The stacked variable $x^* \equiv (x_1^{*\prime}, \ldots, x_T^{*\prime})'$ will follow $N(0,\Gamma)$, where Γ is a symmetric matrix with $T \times T$ blocks and the $(i,j), i \geq j$ block equals $\Phi^{i-j}\Omega$.

Next, construct a transformation matrix

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

and a block diagonal matrix $A \equiv diag\left(\widetilde{A}, \ldots, \widetilde{A}\right)$ in which $\widetilde{A}$ repeats itself $\frac{T}{2}$ times. Then we have $Ax^* \sim N(0, A\Gamma A')$. Essentially Ax^* is a linear transformation of x^*. For each $t = 2, 4, \ldots, T$, the $2t-3, 2t-1, 2t$ elements of Ax^* contain the mixed-frequency observations $x_{1,t-1}, x_{1,t}, x_{2,t}$, while its $2t-2$ element corresponds to the unobserved variable $x_{2,t-1}^*$. Let e be a $2T \times 1$ logical vector whose $2t-3, 2t-1, 2t$ elements are ones (logical true), which serves as an indexing array to select entries of Ax^* and $A\Gamma A'$. Denote the corresponding subvectors (submatrices) by $x_{(0)}, x_{(1)}$, $\Gamma_{00}, \Gamma_{01}, \Gamma_{11}, \Gamma_{10}$. For example, Γ_{01} is a submatrix of $A\Gamma A$ with rows selected by $1-e$ and columns selected by e. Basically, the subscript 0 stands for unobserved variables, while 1 for observations.

Third, the likelihood function is given by the joint distribution of $x_{(1)}$, that is, the density of $N(0,\Gamma_{11})$. The posterior distribution $x_{(0)} \left| x_{(1)} \right.$ follows

$$N\left(\Gamma_{01}\Gamma_{11}^{-1}x_{(1)}, \Gamma_{00} - \Gamma_{01}\Gamma_{11}^{-1}\Gamma_{10}\right).$$

Note that $x_{(0)}$ only contains a fraction of high frequency series, namely

$x_{2,1}, x_{2,3}, \ldots, x_{2,T-1}$. However, the rest high frequency variables are degenerated conditional on $x_{(0)}, x_{(1)}$ since $x^*_{2,t} = x_{2,t} - x^*_{2,t-1}$.

This method can be extended to a general VAR(p) model with irregularly mixed frequency data. Assume that the k dimensional latent series $\{x^*_t\}_{t=1}^T$ follow a stationary VAR(p) process:

$$x^*_t = \sum_{j=1}^{p} \Phi_j x^*_{t-j} + P\varepsilon_t, \tag{2}$$

The reference time unit is t, which indexes the highest frequency data in the VAR system. Let $\left\{x^*_{i,t}\right\}_{t=1}^T$ be the i^{th} component series. Suppose in some time interval $[a, b]$, $1 \leq a \leq b \leq T$, latent values $x^*_{i,a}, \ldots, x^*_{i,b}$ are aggregated. This interval is called an *aggregation cycle*. We then construct the data series $\{x_{i,t}\}_{t=1}^T$ such that $x_{i,a} = \ldots = x_{i,b-1} = NaN$ and $x_{i,b} = \sum_{j=0}^{b-a} x^*_{1,a+j}$. As a special case, $a = b$ implies that the highest frequency data is observed. The data series $\{x_{i,t}\}_{t=1}^T$ contains both observations and aggregation structure, since by counting a run of NaN entries preceding a data point reveals an aggregation cycle. Define a kT-by-1 logical vector e whose $(i-1)T + t$ element equals zero if $x_{i,t}$ is NaN, and equals one otherwise.

The explicit likelihood and posteriors can be found by three steps. First, let x^* be the kT-by-1 stacked latent series and let its joint distribution be $N(0, \Gamma)$. Essentially Γ consists of the auto-covariances of the VAR(p) series, which can be obtained from its companion form (i.e., a giant VAR(1)). See Hamilton (1994, p.265-266) for the auto-covariance formulae. Second, link the aggregated and disaggregated data by a kT-by-kT transformation matrix A such that for a m-period ($m \geq 1$) temporal aggregation, the first $m-1$ variates are retained, while the last one is replaced by the sum of the

variates in the aggregation cycle. Then we have the transformed series $Ax^* \sim N(0, A\Gamma A')$. Third, the likelihood function and posteriors are given by $x_{(1)} \sim N(0, \Gamma_{11})$ and $x_{(0)} \left| x_{(1)} \right. \sim N\left(\Gamma_{01}\Gamma_{11}^{-1}x_{(1)}, \Gamma_{00} - \Gamma_{01}\Gamma_{11}^{-1}\Gamma_{10}\right)$, where $x_{(0)}, x_{(1)}$, $\Gamma_{00}, \Gamma_{01}, \Gamma_{11}, \Gamma_{10}$ are defined similarly as in the bivariate VAR(1) example.

3. The Recursive Solution

The state space representation is the most popular solution to the mixed frequency VAR. The existing SSMs are similar: the state equation is the companion form of VAR(p), possibly with more lags if the length of an aggregation cycle is larger than p. The observation equation extracts observed components or takes linear combinations of the state vector. For example, the state space form of Eq (1) with $x_{2,t}^*$ being aggregated every other period is given by

$$\begin{pmatrix} x_t^* \\ x_{t-1}^* \end{pmatrix} = \begin{pmatrix} \Phi & 0 \\ I & 0 \end{pmatrix} \begin{pmatrix} x_{t-1}^* \\ x_{t-2}^* \end{pmatrix} + \begin{pmatrix} P \\ 0 \end{pmatrix} \varepsilon_t,$$

$$x_{1,t} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_t^* \\ x_{t-1}^* \end{pmatrix}, t = 1, 3, \ldots, T-1,$$

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_t^* \\ x_{t-1}^* \end{pmatrix}, t = 2, 4, \ldots, T.$$

There are two problems of this state space form. First, mixed frequency data imply both time-varying dimensions of the observation vector and time-varying coefficients in the observation equation. Time-varying dimensions can be circumvented by filling in pseudo realizations (say zeros) of some exogenous process (say standard normal) that does not depend on model

parameters, as proposed by Mariano and Murasawa (2003). However, the coefficient matrix in the observation equation remains cyclical. It does not pose a problem in theory, but a time-varying model is inconvenient for both programmers and users. In addition, padding observations with artificial data slows down the Kalman filter. Second, in the Kalman filter recursion, states are predicted and updated conditional on past observations. In this case, the covariance matrix of the predicted states is not invertible, because $x^*_{1,t-1}$, the third component of the state vector, is known conditional on all information up to period $t-1$. Therefore, old smoothing algorithms that require inverting that matrix cannot be applied directly, unless one carefully squeezes non-random components out of the matrix before taking inversion.

We propose a parsimonious, time-invariant and invertible state space representation (stick-and-skip form) for the mixed frequency VAR. The idea is to stick two periods together so we only count a half periods $t = 2, 4, ..., T$ such that

$$\begin{pmatrix} x^*_t \\ x^*_{t-1} \end{pmatrix} = \begin{pmatrix} \Phi^2 & 0 \\ \Phi & 0 \end{pmatrix} \begin{pmatrix} x^*_{t-2} \\ x^*_{t-3} \end{pmatrix} + \begin{pmatrix} P & \Phi P \\ 0 & P \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix},$$

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{1,t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x^*_t \\ x^*_{t-1} \end{pmatrix}.$$

In the stick-and-skip form, the state vectors are non-overlapping, but the state equation represents the same evolution as a VAR(1). As a result, the covariance matrix of the predicted states is of full rank. As for the observation equation, since period t and $t-1$ stick together, the coefficient matrix is time-invariant and non-cyclical. In practice, after reshaping the mixed-frequency

data of T periods as pooled data of $T/2$ periods, the model can readily work on any statistical software that supports state space modeling.

Extension to a general VAR(p) model is straightforward. The state space model is still time-invariant for balanced temporal aggregation. Suppose the least common multiple of the aggregation cycles of the mixed frequency data is m. Put $r = ceil\left(\frac{p}{m}\right) \cdot m$, where $ceil\left(\cdot\right)$ rounds a number towards positive infinity. Let $\xi_t = \begin{pmatrix} x_t^{*\prime} & x_{t-1}^{*\prime} & \cdots & x_{t-r+1}^{*\prime} \end{pmatrix}'$, and rewrite Eq (1) in its companion form: $\xi_t = \widetilde{\Phi}\xi_{t-1} + \widetilde{P}\varepsilon_t$, where $\widetilde{\Phi}, \widetilde{P}$ are largely sparse with $\Phi_1, ..., \Phi_p, P$ sitting in the northwest corner. By iteration we obtain the state equation that glues variables of r periods [2]

$$\xi_t = \widetilde{\Phi}^r \xi_{t-r} + \begin{pmatrix} \widetilde{P} & \widetilde{\Phi}\widetilde{P} & \cdots & \widetilde{\Phi}^{r-1}\widetilde{P} \end{pmatrix} \begin{pmatrix} \varepsilon_t' & \varepsilon_{t-1}' & \cdots & \varepsilon_{t-r+1}' \end{pmatrix}'.$$

The observation equation pools the mixed frequency data of period $t, t-1, \ldots, t-r+1$, which can be extracted from the state vector. With pooled data, the Kalman filter jumps through period $r, 2r, \ldots, T$.

4. Comparison Between the Explicit and Recursive Form

Now we show the connections between the explicit and recursive solution. Since the underlying data generating process is the same, different approaches must lead to the same likelihood function and posterior states estimation. Our finding is that the explicit solution is an unconscious recursive formula implemented automatically by a computer.

[2] Computationally, it is more favorable to directly iterate the VAR(p) to obtain the coefficients of the state equation, though iteration based on the companion form appears more straightforward.

As already explained, $x_{(1)} \sim N(0, \Gamma_{11})$, and the explicit form of the likelihood function is a multivariate normal density

$$(2\pi)^{-\frac{n}{2}} |\Gamma_{11}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} x'_{(1)} \Gamma_{11}^{-1} x_{(1)}\right),$$

where n is the total number of observed data points.

Evaluating this expression on a computer, we seldom directly compute $|\Gamma_{11}|$ and Γ_{11}^{-1} due to concerns on numerical stability. Instead, the Cholesky decomposition or its non-square-root version LDL decomposition is routinely performed. Let $\Gamma_{11} = LDL'$, where L is lower triangular with unitary diagonals and $D \equiv diag(d_1, \ldots, d_n)$ is a diagonal matrix. It follows that

$$|\Gamma_{11}| = \prod_{t=1}^{n} d_t,$$
$$x'_{(1)} \Gamma_{11}^{-1} x_{(1)} = \sum_{t=1}^{n} \frac{v_t^2}{d_t},$$

where $(v_1, \ldots, v_n)' \equiv L^{-1} x_{(1)}$, which can be computed easily by backward substitution. Therefore, the multivariate density is effectively evaluated by the product of univariate normal densities

$$\prod_{t=1}^{n} (2\pi d_t)^{-\frac{1}{2}} \exp\left(-\frac{v_t^2}{2d_t}\right).$$

The LDL decomposition itself is a numerical device, but Fact 1 gives it a statistical interpretation. Proofs of Facts in this paper are provided in the appendix.

Fact 1. *Let* $x_{(1)} \equiv (y_1, \ldots, y_n)'$, *then* $v_t = y_t - E(y_t | y_{t-1}, \ldots, y_1)$, $d_t = Var(v_t)$.

Recall that $x_{(1)}$ represents all mixed frequency observations. Fact 1 shows that direct likelihood evaluation by LDL algorithm is equivalent to rewriting the likelihood in the prediction error decomposition form, which can also be achieved by the Kalman filter. We leave the details on how to obtain v_t, d_t by the Kalman forward recursion in the Appendix. In spite of result equivalence, the Kalman filter is a conscious recursion, since it critically explores the AR(1) state transition which enables the iterative one-step prediction of the states and observations. The computational complexity of the Kalman filter is $O(n)$. In contrast, the LDL algorithm is mechanical (say, by the rank-one update scheme) and can decompose any covariance matrix. Though it unconsciously obtains the same likelihood function in the prediction error decomposition form, its computational complexity is roughly $\frac{1}{3}n^3$.

Similar arguments apply to the simulation smoother. We focus on the posterior mean, for posterior draws can be obtained solely from the mean-adjusted smoother without computing its variance, as proposed by Durbin and Koopman (2002). We show that the explicit form of the posterior mean is indeed computed by an unconscious recursion, which ensembles the smoothing algorithm of de Jong (1989).

We have shown $E\left(x_{(0)} \left| x_{(1)} \right.\right) = \Gamma_{01}\Gamma_{11}^{-1}x_{(1)}$. In realistic computation, we seldom invert Γ_{11} and then times this quantity by $x_{(1)}$. Instead, we treat it as a linear equation and adopt LU, QR or Cholesky decomposition. Since Γ_{11} is symmetric positive definite, the most appropriate numerical method is Cholesky decomposition or its variant LDL decomposition. Let $\Gamma_{11} = LDL'$,

then

$$\begin{aligned}\Gamma_{01}\Gamma_{11}^{-1}x_{(1)} &= E\left[x_{(0)}x_{(1)}'\right]\left(LDL'\right)^{-1}x_{(1)} \\ &= E\left[x_{(0)}\left(L^{-1}x_{(1)}\right)'\right]D^{-1}\left(L^{-1}x_{(1)}\right) \\ &= \sum_{t=1}^{n} E\left[x_{(0)}v_t\right]d_t^{-1}v_t\end{aligned}$$

In the appendix we show this expression can be obtained from a backward recursion similar to the smoother of de Jong (1989). The computational complexity of the backward recursion is $O(n)$, while direct computation by the LDL decomposition requires a complexity of $O(n^3)$.

Note that the variable n stands for the number of periods times number of observations in a period. Clearly, n must be large in a typical empirical study and the difference between $O(n)$ and $O(n^3)$ is substantial. This, however, does not imply the explicit solution is worthless. Note that $O(n)$ and $O(n^3)$ imply nothing but asymptotics. When n is small, $O(n^3)$ is not necessarily larger than $O(n)$. If we slightly adapt the explicit-form posterior distribution and apply it on a small chunk of observations, we obtain a parallel Gibbs sampler that works much faster than the Kalman simulation smoother.

5. Gibbs Sampler with Blocks

The sampling procedure in Section 2 allows us to draw latent variables all at once. However, it is computationally unfavorable to transform the entire sample at one time, so we partition the sample according to the aggregation cycle and explore the (high order) Markov property of the VAR(p) process.

Fact 2. *Suppose* $\{x_t^*\}_{t=1}^T$ *follow a VAR(p) process as in Eq (2), then*

$$\begin{aligned} & p\left(x_s^*, \ldots, x_t^* \left| x_1^*, \ldots, x_{s-1}^*, x_{t+1}^*, \ldots, x_T^* \right.\right) \\ & = p\left(x_s^*, \ldots, x_t^* \left| x_{s-p}^*, \ldots, x_{s-1}^*, x_{t+1}^*, \ldots, x_{t+p}^* \right.\right), \end{aligned}$$

for $1 \leq s \leq t \leq T$*, with proper adjustments for the initial and terminal variables.*

To fix ideas, consider again Eq (1) with $\left\{x_{2,t}^*\right\}$ aggregated every other period. Treat variables in each aggregation cycle as a *chunk* (the word *block* is reserved for later use), so we have $\frac{T}{2}$ chunks. That is, $c_j \equiv \left(x_{2j-1}^{*\prime}, x_{2j}^{*\prime}\right)'$, $j = 1, \ldots, \frac{T}{2}$. Suppose we want to sample latent variables chunk by chunk. For each j, we sample $x_{2,2j-1}^*$ conditional on all other chunks and all observations. Once $x_{2,2j-1}^*$ is sampled, other latent variables in j^{th} chunk can be sampled trivially. By the Markov property we actually work on

$$x_{2,2j-1}^* \left| x_{1,2j-2}^*, x_{2,2j-2}^*, x_{1,2j-1}, x_{1,2j}, x_{2,2j}, x_{1,2j+1}^*, x_{2,2j+1}^* \right. ,$$

with proper adjustments for the initial and terminal variables. Similar to the method described in Section 2, the posterior conditional distribution of $x_{2,2j-1}^*$ can be obtained by three steps. First, obtain the joint distribution of $\widetilde{x}_j \equiv \left(x_{2j-2}^{*\prime}, x_{2j-1}^{*\prime}, x_{2j}^{*\prime}, x_{2j+1}^{*\prime}\right)'$. We have $\widetilde{x}_j \sim N\left(0, \Gamma\right)$, where Γ is a symmetric matrix with 4×4 chunks and the $(i, r), i \geq r$ chunk equals $\Phi^{i-r}\Omega$. Second, construct a 8×8 transformation matrix $A = I_8 + 1_{(6,4)}$, where I_8 is an identity matrix and $1_{(6,4)}$ is an 8×8 zero matrix except for its $(6, 4)$ element being one. $A\widetilde{x}_j$ transform $\widetilde{x}_j$ such that $x_{2,2j}^*$ is replaced by the low frequency data $x_{2,2j}$. Let e be an 8×1 logical vector in which all but the 4^{th} element equals to one. Third, $A\widetilde{x}_j \sim N\left(0, A\Gamma A'\right)$, and $x_{(0,j)} \left| x_{(1,j)} \right. \sim$

$N\left(\Gamma_{01}\Gamma_{11}^{-1}x_{(1,j)}, \Gamma_{00}-\Gamma_{01}\Gamma_{11}^{-1}\Gamma_{10}\right)$, where $x_{(0,j)}, x_{(1,j)}, \Gamma_{00}, \Gamma_{01}, \Gamma_{11}, \Gamma_{10}$ are sub-vectors (submatrices) of $A\widetilde{x}_j$ and $A\Gamma A'$ selected by e and/or $1-e$. In this case $x_{(0,j)} = x^*_{2,2j-1}$ and $x_{(1,j)} = \left(x^*_{1,2j-2}, x^*_{2,2j-2}, x_{1,2j-1}, x_{1,2j}, x_{2,2j}, x^*_{1,2j+1}, x^*_{2,2j+1}\right)'$. Note that the conditional variance $\Gamma_{00}-\Gamma_{01}\Gamma_{11}^{-1}\Gamma_{10}$ is equal for all but the initial and terminal blocks. This feature saves substantial amount of computation.

More importantly, the proposed sampler satisfies the blocking property. Consider a genetic setting where $\theta_0, \theta_1, \ldots \theta_q$ are latent variable chunks. The Gibbs sampler takes turns to sample from $p\left(\theta_i \left| \theta_{-i} \right.\right)$, where θ_{-i} denotes all but the i^{th} chunk. If $p\left(\theta_1, \ldots \theta_q \left| \theta_0 \right.\right) = \prod_{i=1}^{q} p\left(\theta_i \left| \theta_0 \right.\right)$, then we say $\theta_1, \ldots \theta_q$ satisfy the blocking property. Since $\theta_1, \ldots \theta_q$ are independent conditional on θ_0, the blocking property implies that $p\left(\theta_i \left| \theta_{-i} \right.\right) = p\left(\theta_i \left| \theta_0 \right.\right)$, $i = 1, \ldots, q$. Suppose we have q parallel processors and let the i^{th} processor take a draw from $p\left(\theta_i \left| \theta_{-i} \right.\right)$. Then we effectively obtain draws from $p\left(\theta_1, \ldots \theta_q \left| \theta_0 \right.\right)$. In this sense, $\theta_1, \ldots \theta_q$ are sampled as a block.

Figure 1 illustrate the chunks and blocks for the bivariate VAR(1) example. The $\frac{T}{2}$ chunks are represented by $c_1, \ldots, c_{\frac{T}{2}}$ and the two blocks are $b_1 \equiv \left(c_1, c_{3,} \ldots, c_{\frac{T}{2}-1}\right)$, $b_2 \equiv \left(c_2, c_{4,} \ldots, c_{\frac{T}{2}}\right)$. The Markov property of AR(1) implies the blocking property $p\left(c_1, c_{3,} \ldots, c_{\frac{T}{2}-1} \left| b_2 \right.\right) = \prod_{j=1,3,\ldots,\frac{T}{2}-1} p\left(c_j \left| b_2 \right.\right)$ as well as $p\left(c_2, c_{4,} \ldots, c_{\frac{T}{2}} \left| b_1 \right.\right) = \prod_{j=2,4,\ldots,\frac{T}{2}} p\left(c_j \left| b_1 \right.\right)$. Therefore, the Gibbs sampler cycles through two blocks, namely $p\left(b_1 \left| b_2 \right.\right)$ and $p\left(b_2 \left| b_1 \right.\right)$. Within each block, multiple processors collaborate to sample chunks simultaneously. Compared with FFBS by which all latent variables are sampled as one block, parallel sampler increases the length of the MCMC chain by one, but not $\frac{T}{2}$.

So we can expect the correlation of MCMC draws only slightly rises.

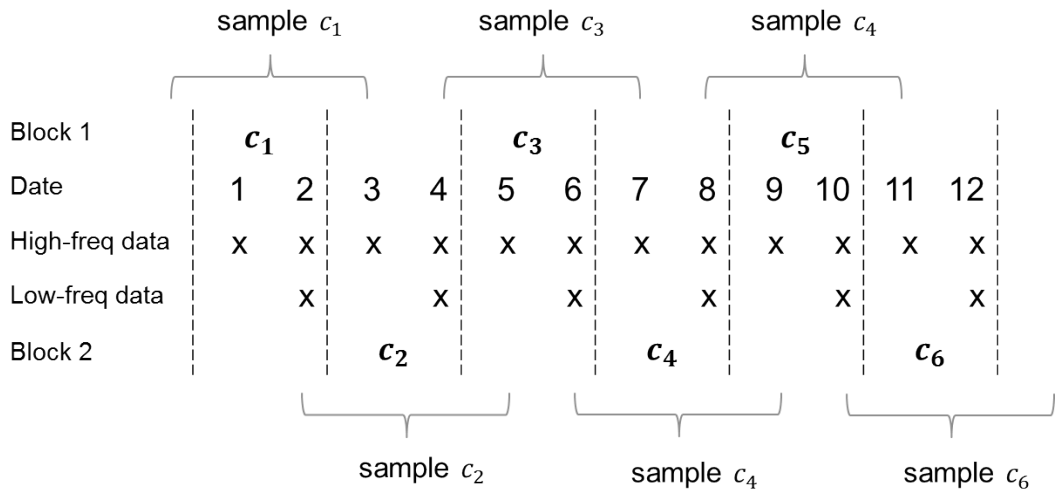
Implementation of parallel computation for this problem is not demanding. Consider sampling from $p\left(b_1 \left| b_2 \right.\right)$. We need to compute the conditional mean for each chunk in that block. However, the formula $\Gamma_{01}\Gamma_{11}^{-1}x_{(1,j)}$ implies that we may concatenate the vectors $x_{(1,j)}$ as a matrix, say $x_{(b_1)} \equiv \left[x_{(1,1)}, x_{(1,3)}, \ldots, x_{\left(1,\frac{T}{2}-1\right)}\right]$. Then all these conditional means can be computed in bulk through $\Gamma_{01}\Gamma_{11}^{-1}x_{(b_1)}$. In modern matrix-based computational platforms such as MATLAB, parallel computation is inherent in matrix multiplication and will be automatically invoked when the matrix size is large enough.

Also note that the method of designating chunks and blocks is not unique. It is not wrong to put variables in two aggregation cycles as a chunk, but a larger trunk intensifies the $O\left(n^3\right)$ complexity of the LDL decomposition. It is not wrong to group chunks in other manners to form a block, but more blocks translate to higher correlation of the MCMC draws.

In a general VAR(p)-based model, the blocking strategy still applies. Suppose the least common multiple of the aggregation cycles of the mixed frequency data is m. Then variables of m periods can be treated as a chunk. That is, $c_j \equiv \left(x_{mj-m+1}^{*\prime}, \ldots, x_{mj}^{*\prime}\right)^{\prime}$,$j = 1, \ldots, \frac{T}{m}$.According to the high order Markov property, $x_{mj-m-p+1}^{*}, \ldots, x_{mj-m+1}^{*}, \ldots, x_{mj}^{*}, \ldots, x_{mj+p}^{*}$ are relevant for sampling latent variables in c_j. The procedures are same as the bivariate VAR(1) example. The chunks can also be grouped into blocks. Let $s = ceil\left(\frac{p}{m}\right) + 1$, then the blocks are $b_i \equiv \left(c_i, c_{i+s}, \ldots, c_{i+\frac{T}{m}-s}\right)$, $i = 1, \ldots s$.

In summary, triple factors contribute to the acceleration of the parallel Gibbs sampler. First, it works with small trunks of observations and the

Figure 1: A Graphic Illustration of the Parallel Gibbs Sampler



Consider a bivariate VAR(1) model with the second series aggregated every other period. Mixed frequency observations of the first 12 periods are depicted in the rows "High-freq data"and "Low-freq data". Each aggregation cycle (i.e., two periods) constitutes a chunk, as illustrated by $c_1, \ldots, c_6$. To sample latent variables in a chunk, we need not only mixed frequency observations in that chunk, but also two neighboring latent variables outside the chunk. The parallel Gibbs sampler cycles through blocks rather than chunks. In this case, Block 1 consists of chunks c_1, c_3, c_5 and Block 2 contains chunks c_2, c_4, c_6. Within a block, multiple computational threads can collaborate to sample chunks regardless of the sampling order.

$O(n^3)$ complexity of LDL decomposition does not loom large. Second, for balanced aggregation, the same conditional variance applies to many chunks. Third, multiple processors can simultaneously sample chunks within a block. The potential of the third factor is unlimited, for it is the reality that a personal computer is equipped with increasingly larger multi-threads in the CPU and GPU. Even in the absence of multiple computational threads, the second factor alone makes the sampler faster than FFBS, which computes the entire sequence of the predicted and filtered state variances for the whole sample periods.[3]

6. Simulation Study

The mixed frequency VAR model under discussion is fully parametric and various approaches would generate the same results if we could take infinite amount of MCMC draws. However, their numerical performance may vary with the size of the model. We conduct a Bayesian simulation exercise to compare the sampling speed and the quality of the correlated MCMC draws. The four candidate methods are the explicit form of posterior states (EXPLICIT), the traditional state space form (SSM1), the stick-and-skip state space form (SSM2), and the parallel Gibbs sampler (PARGIBBS). The new VAR proposed by Ghysels (2012) and MIDAS family models are not experimented, for they assume a different data generating process. See

[3]For a stationary model, the Riccati equation will converge and state variances eventually level off under suitable regularity conditions. One may stop computing the variances halfway when the sequence stabilizes. However, it is just an approximation. To obtain the precise likelihood or smoother, one should compute the entire variance sequence.

Kuzin et al. (2011) for comparisons between the mixed frequency VAR and MIDAS.

The hardware platform is a personal computer with a 3.0G Hz four-core Xeon W3550 CPU and 12 GB RAM. The software platform is Win64 MATLAB 2013a without Parallel Computing Toolbox. Codes of the four candidate methods reflect the developer's best efforts and offer the same level of generality: a Bayesian multivariate VAR(p) model under Normal-Wishart priors (independent, dependent or diffuse) with Minnesota-style shrinkage; high and low mixed frequency data subject to balanced temporal aggregation. An unfair part is that the Kalman filter and the simulation smoother are implemented by a mex file based on compiled C codes, while others are completely MATLAB codes. The mex file accelerates the two state space methods by thirty to fifty times, which qualifies them to compete with the parallel Gibbs sampler on smaller models. Note that the only channel that the parallel Gibbs sampler gets access to multiple computational threads is matrix multiplication, in which parallel computation is automatic when the matrix size is large.

We consider three scenarios that differ in model sizes. In all scenarios the VAR coefficients have independent Normal-Wishart priors with shrinkage on regression coefficients. The Gibbs sampler are employed to cycle through posterior conditionals of regression coefficients, disturbance covariance matrix as well as the latent high frequency data. The four candidate methods only differ in sampling the latent variables. We compare their sampling speed measured in total computation time and the quality of the MCMC draws by

the relative numerical efficiency (RNE), which is defined as

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

where ρ_j is j^{th} autocorrelation of the draws sequence of length r.[4] It is expected that the first three methods should yield similar RNE, since all latent variables are drawn as a whole. However, the last method may exhibit lower RNE because latent variables are sampled for each block conditional on other blocks, which increases the length of the MCMC chain and the correlation of draws.

The first scenario is a small model in which one component variable in a bivariate VAR(1) is temporally aggregated every other period. The true VAR parameters are artificial and 500 observations are simulated. Then each of the four candidate samplers takes 5000 draws with the first half burned in. For credibility and robustness of results, the experiment of this scenario is repeated for 100 times; each time with new pseudo observations. The average sampling speed and RNE are reported with standard deviations in parenthesis. As seen in Table 1, The EXPLICIT method is significantly slower than others, due to manipulation of a 1000-by-1000 covariance matrix. In such a small model, there is no speed advantage of PARGIBBS over SSM1/SSM2, or vice versa. They all take 3 to 4 seconds for 5000 draws. On average, the RNE of PARGIBBS is reduced by less than 4% compared with the RNE of about 0.2 for other methods. Note the standard deviation of RNE is roughly

[4]It is not feasible to estimate autocorrelations of order close to the sample size. In practice, a cut-off lag is designated. We put 100 lags for the sample size of 5000. Lags increase with square root of the sample size.

0.04 for all methods, so the reduction of RNE of PARGIBBS is not significant in our experiment.

The second scenario is a median-sized model with six variables and three lags in the regression. The 1200 simulated data ensemble the monthly-quarterly aggregation. Other settings are same as the first scenario. The EXPLICIT method is not experimented, for it is already too slow even on a small model. As seen in Table 1, the stick-and-skip SSM shortens the sampling time by 37% compared to the traditional SSM that costs 185 seconds. However, both SSM1 and SSM2 are overshadowed by PARGIBBS, for it only takes 11 seconds with invisible RNE decrease. Also recall that it is an unfair competition between MATLAB and C codes in favor of SSM. The RNE is smaller for all methods compared with previous scenario. We are not aware of the exact reason, which might due to more VAR parameters and longer aggregation cycle. In view of that, we suggest practitioners ran mixed frequency VAR with thinned MCMC draws, which nevertheless is affordable since PARGIBBS is fast.

The third scenario is a larger model that contains 12 variables, 3 lags, 3000 observations and 10000 draws. SSM2 is still preferable to SSM1 in terms of 34% reduced sampling time, though the speed advantage of PARGIBBS renders state space solution little attraction. Its computation time is 122 seconds, while that of SSM1 and SSM2 are 6215, 4085 respectively. We want to clarify that a truly large Bayesian VAR model that contains a hundred variables, as discussed in Banbura et al. (2010), can hardly work with mixed frequency data. In their paper, dependent Normal-Wishart priors are employed so that the model has analytic posterior marginal distribution for

Table 1: Comparison of Sampling Speed and Efficiency of Mixed Frequency VAR Algorithms

	Small Model			
	EXPLICIT	SSM 1	SSM 2	PARGIBBS
Speed	294.215	4.055	3.720	3.632
	(4.334)	(0.109)	(0.073)	(0.137)
RNE	0.201	0.196	0.196	0.193
	(0.041)	(0.048)	(0.045)	(0.039)
	Median Model			
Speed		184.678	116.440	11.475
		(2.711)	(2.254)	(0.395)
RNE		0.024	0.024	0.024
		(0.003)	(0.003)	(0.003)
	Larger Model			
Speed		6215.359	4085.393	122.067
		(15.730)	(20.281)	(1.781)
RNE		0.017	0.016	0.016
		(0.002)	(0.002)	(0.001)

In a Bayesian simulation study, the explicit solution, traditional SSM, stick-and-skip SSM and parallel Gibbs sampler are compared on small, median and larger models. Sampling speed is measured by total computational time in seconds. Sampling efficiency is measured by relative numerical efficiency. Each experiment is repeated multiple times, average speed and efficiency are reported with standard deviations in parenthesis.

disturbance covariance matrix (inverse Wishart distribution) and regression coefficients (matric-variate t distribution). However, once mixed frequency data are added to the model, we rely on the computationally intensive Gibbs sampler. For very large models, if one has access to computer clusters, one may coordinate many computers to fulfill the parallel Gibbs sampler in a more efficient manner.

7. Yield Curve Forecast Application

In this section, we apply our mixed frequency approach to a dynamic factor model for the yield curve forecast. U.S. Treasury yields with maturities of 3, 6, 12, 24, 36, 60, 84, 120, 240, 360 are studied for the sample period 1990:01 - 2013:03. The model setup is the same as Diebold et al. (2006), in which the yields are determined by the Nelson and Siegel (1987) curve:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right),$$

where $y_t(\tau)$ is the period-t Treasury yield with the maturity τ. The parameter λ determines the exponential decay rate. The three dynamic factors $\beta_{1t}, \beta_{2t}, \beta_{3t}$ are interpreted by Diebold and Li (2006) as level, slope and curvature factors, which interact with macroeconomic variables in a VAR model.

To capture the basic macroeconomic dynamics, we include the four-quarter growth rate of real GDP, 12-month growth rate of price index for personal consumption expenditures as well as the effective federal funds rate. GDP is the single best measure of economic activity, but only available at quarterly frequency. Researchers often use monthly proxies such as capacity utilization, industrial production index or interpolated GDP from quarter

data. In this application, however, we assume that the latent 12-month growth rate of "monthly GDP" interacts with inflation rate, federal funds rates as well as three yield curve factors in a six-variate VAR(1) system:

$$x_t^* = \Phi x_{t-1}^* + P\varepsilon_t,$$

where $x_t^* = (g_t^*, \pi_t, i_t, \beta_{1t}, \beta_{2t}, \beta_{3t})$ and g_t, π_t, i_t are output growth, inflation and fed funds rate respectively. Put $\Sigma \equiv PP'$.

The observed four-quarter growth rate of GDP can be approximately viewed as the average of monthly GDP growth rate: [5]

$$g_t = \frac{1}{3}\left(g_t^* + g_{t-1}^* + g_{t-2}^*\right).$$

for $t = 3, 6, \ldots T$.

We propose the following Bayesian method, based on the parallel Gibbs sampler in Section 5, to estimate the model. In each step, we sample one parameter/variable block from their full posterior conditional distributions.

Step 1: sample regressive coefficients in the VAR. This is a standard Bayesian VAR procedure. We assume a multivariate normal prior indepen-

[5]The simple average holds precisely if the monthly GDP in that quarter of last year is a constant. Otherwise, larger weights should be assigned to months with higher GDP level. For practical purposes, assuming aggregation by simple average is less harmful, since monthly variation of (seasonally adjusted) series should be relatively small compared with the level of the series. A related issue is the aggregation under logarithmic data. Mariano and Murasawa (2003, 2010) document this nonlinear aggregation problem and suggest redefining the disaggregated data as the geometric mean (instead of the arithmetic mean) of the disaggregated data. Camacho and Perez-Quiros (2010) argue that the approximation error is almost negligible if monthly changes are small and the geometric averaging works well in practice.

dent to the prior of Σ, which allows us to treat own and foreign lags asymmetrically. Shrinkage to random walks in the spirit of Minnesota prior is applied (see Litterman, 1986; Kadiyala and Karlsson, 1997, for details). For each equation of the VAR system, the prior mean for the coefficient on the first own lag is set to one, while other coefficients have prior mean of zero. The prior variance is set to 0.05 for the own lag, 0.01 for foreign lags and 1e6 for the constant term. The posterior conditional distribution is multivariate normal.

Step 2: sample disturbance covariance matrix, which has inverse Wishart posterior distribution under the reference prior $p(\Sigma) \propto \Sigma^{-\frac{7}{2}}$ as in a standard Bayesian VAR model.

Step 3: sample latent monthly GDP growth rate. This follows the parallel Gibbs sampling procedure for mixed frequency VAR described in Section 5.

Step 4: sample three yield curve factors. Note that conditional on the latent monthly GDP series, the dynamic factor model is readily in the standard state space form in which the state vector consists of three factors and three macro variables. Then we apply forward filtering and backward sampling to obtain posterior state draws.

Step 5: sample the disturbance variances in the Nelson-Siegel curve. Though Nelson-Siegel curve has excellent fit of the yield data, adding a disturbance term is necessary. Otherwise, the three factors can be solved precisely from any of the three yields. We assume that the yield curve for each maturity has an uncorrelated disturbance variance σ_τ^2, with a reference prior $p(\sigma_\tau^2) \propto \sigma_\tau^{-2}$. The posterior conditional distribution is inverse gamma.

Step 6: sample the scalar parameter λ in the Nelson-Siegel model. Our

prior comes from Diebold and Li (2006)'s description that λ "determines the maturity at which the loading on the medium-term, or curvature, factor achieves maximum. Two- or three-year maturities are commonly used in that regard ". So we put a uniform prior that ranges from 0.037 to 1.793, which correspond to the maximizer of one- and four- year maturities. The posterior conditional distribution of λ is not of known form, and we may either insert a metropolis step or discretize the value of λ into grids as the maximizer of the maturity of each month. The two methods yield similar results, and the result of random walk metropolis (normal proposal density with adaptive variance) is reported. In fact, after some vibrations in the burn-in periods, λ draws are extremely stable. This provides support for Diebold and Li (2006)'s decision of a predetermined $\lambda = 0.0609$.

Note that state space forms, traditional or stick-and-skip, can also be applied to this model so that Step 3 and 4 can be merged, though the states should keep track of factors and macro variables of three periods. In that case, the length of the state vector would be 18. Since the parallel Gibbs sampler runs much faster than the FFBS, it is worthwhile to first impute latent GDP and then work on a smaller state space representation.

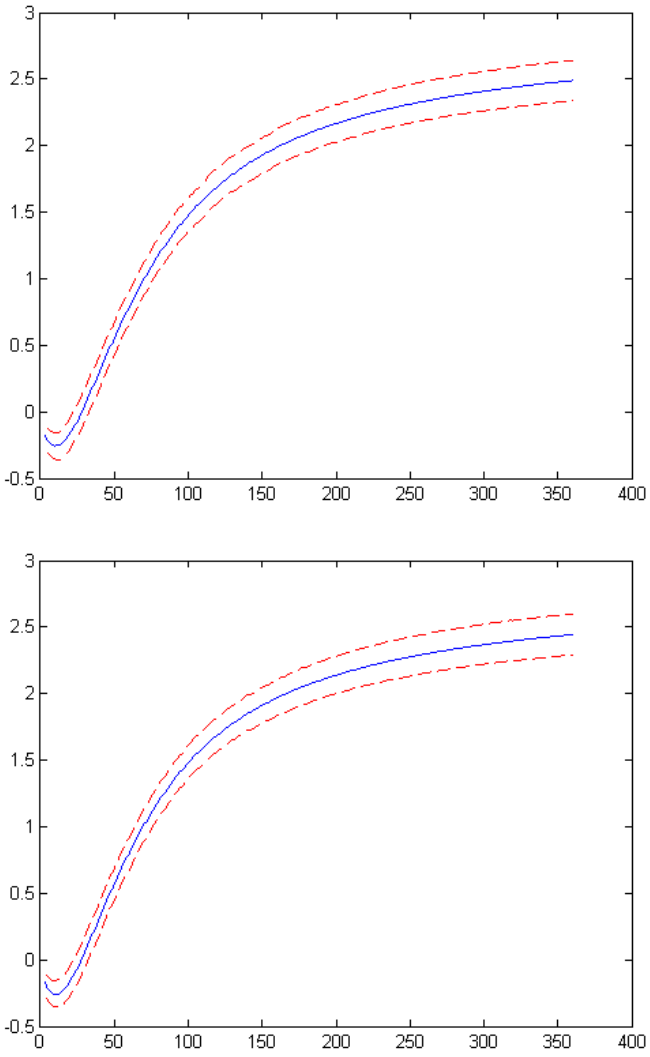
Cycling through these steps for 1,000,000 times with the first half of draws burned-in and the rest thinned every 100 draws, we obtain posterior draws of model parameters, latent monthly GDP growth series and three dynamic factor series. Diagnostic tests suggest the chain has converged and mixed well. The computation time is about an hour on the machine described in Section 6.

Consider the yield curve forecast in two scenarios. First, suppose the

current period is the end of a quarter (say, March). Conditional on all the monthly and quarterly observations up to the current period, we make a one-month ahead forecast. Second, suppose the current period is one month after the end of a quarter (say, April). Conditional on all available observations including the current month, we make a one-month ahead forecast. The two scenarios differ by the real-time monthly information. Our approach can handle both scenarios by treating the last-period GDP as missing data without temporal aggregation. This treatment is essentially the sampler proposed by Chiu et al. (2011). Once we obtain the posterior draws of the three factors and monthly GDP of the current period, we plug them back to the VAR and predict factors for the next month. Then we forecast the Nelson-Siegel yield curve for the next month. The upper and lower panels of Figure 2 depicts the one-month yield curve forecast using observations up to 2013:03 and 2013:04 respectively.

Lastly, consider a real-time forecast of another type. We use the yield quotes of the first trading day of a month, which must be released earlier than the macroeconomic data of that month. Suppose we observe Treasury yields up to 2013:04, while macro data has one month in lag. Then we could use this model to forecast GDP, inflation and fed funds rate for 2013:04. Heuristically, the most recent yield curve carries information on current-month factors, which interact with macro variables under the model assumption. This information helps forecast on the macro variables, in additional to historical observations. Our FFBS codes support missing observations, so the easiest way to conduct this forecast is to put macro observations in 2013:04 as missing values, so that the simulation smoother will generate the forecast

Figure 2: One-Month-Ahead Yield Curve Forecast



The upper panel uses data up to 2013:03 to estimate the model and obtain one-period ahead forecast. The lower panel adds new information of 2013:04 to predict the yield curve of the next month. The horizontal axis represents yield maturities in months, and the vertical axis is the yield level. The solid line depicts the posterior mean of the forecast curve, and the dotted lines bracket the 90 % credible interval with highest posterior density.

of these macro variables. The projected GDP growth, inflation and fed funds rate of 2013:04 are 0.0192, 0.0102, 0.0016 with standard deviations 0.0063, 0.0029, 0.0011 respectively. Note that the forecast on the fed funds rate has larger uncertainty. This is not surprising, for the exploratory analysis on the dataset shows that the fed funds rate plummets around 2008 and remains historically low. The sample means before and after 2008:12 are 0.043 and 0.001 respectively. For future work, it might be interesting to add a structural break at some unknown date or introduce Markov-switching regimes.

8. Conclusion

We considered the mixed frequency VAR model with latent variables. Under the assumed data generating process, the three proposed methods offer the optimal estimate by exploring the idea that lower frequency observations impose linear constraints on latent variables. Our simulation study suggests outstanding performance of the parallel Gibbs sampler. On the one hand, it is fast even on a personal computer, let alone its suitability for future computational environments. On the other hand, its sampling procedure can easily be integrated with the Gibbs sampler for other models. Essentially, it transforms mixed frequency observations to augmented data of homogeneous frequency, so that methods handling the complete-data model apply.

Appendix A. Proof of Fact 1

First, by definition, $v_1, \ldots, v_n$ are invertible linear transformation of $y_1, \ldots, y_n$, so $v_1, \ldots, v_n$ are also multivariate normal.

$E\left[(v_1,\ldots,v_n)'(v_1,\ldots,v_n)\right] = L^{-1}(LDL')L^{-1\prime} = D$, which implies $v_1,\ldots,v_n$ are independent to each other and $Var(v_t) = d_t$.

Let L_{tj} be (t,j) element of L, which is an invertible lower triangular matrix with unitary diagonals. $(y_1,\ldots,y_n)' = L\cdot(v_1,\ldots,v_n)'$ implies

$y_t = v_t + \sum_{j=1}^{t-1} L_{tj}v_j$, or $v_t = y_t - \sum_{j=1}^{t-1} L_{tj}v_j$.

Since v_t is independent to $v_1,\ldots,v_{t-1}$, by the property of linear projection, we have $v_t = y_t - E(y_t|v_1,\ldots,v_{t-1})$.

Also, invertible lower triangular L also implies $v_1,\ldots,v_{t-1}$ are invertible linear transformation of $y_1,\ldots,y_{t-1}$,

It follows that $E(y_t|v_1,\ldots,v_{t-1}) = E(y_t|y_{t-1},\ldots,y_1)$ and $v_t = y_t - E(y_t|y_{t-1},\ldots,y_1)$.

Appendix B. Proof of Fact 2

For notational convenience, define $\mathbf{Y}_s^t = \{\mathbf{Y}_s^*,...,\mathbf{Y}_t^*\}$. Let

$$\mathbf{Y}_t^* = \mathbf{c} + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{Y}_{t-i}^* + \varepsilon_t.$$

This data generating process suggests

$$p\left(\mathbf{Y}_t^* \middle| \mathbf{Y}_{t-p}^{t-1}\right) = p\left(\mathbf{Y}_t^* \middle| \mathbf{Y}_{t-p-1}^{t-1}\right) = ... = p\left(\mathbf{Y}_t^* \middle| \mathbf{Y}_1^{t-1}\right),$$

since their distributions are all equal to the distribution of ε_t with the mean shifted by $\mathbf{c} + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{Y}_{t-i}^*$. Then we have

$$\begin{aligned} p\left(\mathbf{Y}_s^t \middle| \mathbf{Y}_1^{s-1}, \mathbf{Y}_{t+1}^T\right) &\propto p\left(\mathbf{Y}_1^T\right) \\ &= p\left(\mathbf{Y}_1^{s-1}\right) \cdot \prod_{j=s}^{t+p} p\left(\mathbf{Y}_j^* \middle| \mathbf{Y}_{j-p}^{j-1}\right) \cdot p\left(\mathbf{Y}_{t+p+1}^T \middle| \mathbf{Y}_{t+1}^{t+p}\right) \\ &\propto \prod_{j=s}^{t+p} p\left(\mathbf{Y}_j^* \middle| \mathbf{Y}_{j-p}^{j-1}\right). \end{aligned}$$

Similarly,

$$\begin{aligned} p\left(\mathbf{Y}_s^t \left| \mathbf{Y}_{s-p}^{s-1}, \mathbf{Y}_{t+1}^{t+p}\right.\right) &\propto p\left(\mathbf{Y}_{s-p}^{t+p}\right) \\ &= p\left(\mathbf{Y}_{s-p}^{s-1}\right) \cdot \prod_{j=s}^{t+p} p\left(\mathbf{Y}_j^* \left| \mathbf{Y}_{j-p}^{j-1}\right.\right) \\ &\propto \prod_{j=s}^{t+p} p\left(\mathbf{Y}_j^* \left| \mathbf{Y}_{j-p}^{j-1}\right.\right). \end{aligned}$$

Both $p\left(\mathbf{Y}_s^t \left| \mathbf{Y}_1^{s-1}, \mathbf{Y}_{t+1}^T\right.\right)$ and $p\left(\mathbf{Y}_s^t \left| \mathbf{Y}_{s-p}^{s-1}, \mathbf{Y}_{t+1}^{t+p}\right.\right)$ are proper. If they are proportional to the same expression, they must be equal.

Appendix C. Kalman Filter and Smoother

We outline the Kalman filtering algorithm and highlight how v_t, d_t obtained from LDL decomposition can also be derived in forward recursion and how $\sum_{t=1}^{n} E\left[x_{(0)} v_t\right] d_t^{-1} v_t$ is computed by backward recursion. Consider a state space model

$$\begin{aligned} x_t^* &= A_t x_{t-1}^* + B_t \varepsilon_t, \\ y_t &= C_t x_t^* + D_t u_t. \end{aligned}$$

where $\{\varepsilon_t\}_{t=1}^T, \{u_t\}_{t=1}^T$ are independent Gaussian white noise series. Note that we assume time-varying coefficients so that we can treat $\{y_t\}_{t=1}^T$ as a scalar observation series, which is essentially the univariate treatment of multivariate series.

Let $Y_1^t = (y_1, \ldots, y_t)$, $\widehat{x}_{s|t} = E\left(x_s^* | Y_1^t\right)$, $P_{s|t} = Var\left(x_s^* | Y_1^t\right)$. Given the initial state distribution $N\left(\widehat{x}_{0|0}, P_{0|0}\right)$, the forward recursion in period $t = 1, \ldots, T$ consists of the prediction and update steps. First, predict

states: $x_t^* \left| Y_1^{t-1} \right. \sim N\left(\widehat{x}_{t|t-1}, P_{t|t-1}\right)$, where $\widehat{x}_{t|t-1} = A_t \widehat{x}_{t-1|t-1}$, $P_{t|t-1} = A_t P_{t-1|t-1} A_t' + B_t B_t'$. Second, predict observations: $y_t \left| Y_1^{t-1} \right. \sim N\left(\widehat{y}_{t|t-1}, d_t\right)$, where $\widehat{y}_{t|t-1} = E\left(y_t \left| Y_1^{t-1} \right.\right) = C_t \widehat{x}_{t|t-1}$, $d_t = Var\left(y_t \left| Y_1^{t-1} \right.\right) = C_t P_{t|t-1} C_t' + D_t D_t'$. Equivalently, the prediction error form is $v_t \sim N(0, d_t)$, where $v_t = y_t - \widehat{y}_{t|t-1}$. The definition of v_t, d_t here is conformable to that in the LDL decomposition, as established by Fact 1. Third, update states: $x_t^* \left| Y_1^t \right. \sim N\left(\widehat{x}_{t|t}, P_{t|t}\right)$, where $\widehat{x}_{t|t} = \widehat{x}_{t|t-1} + P_{t|t-1} C_t \left(d_t\right)^{-1} v_t$, $P_{t|t} = P_{t|t-1} - P_{t|t-1} C_t \left(d_t\right)^{-1} C_t' P_{t|t-1}'$. This completes a recursion cycle and the filter proceeds to the next period.

The smoother $\widehat{x}_{t|T}$ is the posterior mean of states conditional on all observations.

$$\begin{aligned}
\widehat{x}_{t|T} &= E\left(x_t^* \left| Y_1^{t-1}, v_t, \ldots, v_T \right.\right) \\
&= \widehat{x}_{t|t-1} + \sum_{s=t}^{T} E\left(x_t^* v_s\right) d_s^{-1} v_s \\
&= \widehat{x}_{t|t-1} + P_{t|t-1} \sum_{s=t}^{T} \left(\prod_{j=t}^{s-1} J_j' \right) C_s' d_s^{-1} v_s
\end{aligned}$$

where $J_t = A_{t+1} - A_{t+1} P_{t|t-1} C_t' \left(d_t\right)^{-1} C_t$. Put $r_t = \sum_{s=t}^{T} \left(\prod_{j=t}^{s-1} J_j' \right) C_s' d_s^{-1} v_s$, which can be efficiently computed by a backward recursion such that $r_t = J_t' r_{t+1} + C_t' d_t^{-1} v_t$, $r_{T+1} = 0$.

Recall that the explicit solution by the LDL decomposition works on $\sum_{t=1}^{n} E\left[x_{(0)} v_t\right] d_t^{-1} v_t$, where $x_{(0)}$ contains latent variables of all periods and are estimated in bulk. However, the Kalman smoother works on a similar expression $\sum_{s=t}^{T} E\left(x_t^* v_s\right) d_s^{-1} v_s$, where x_t^* contains latent variables of a single period. It is computed efficiently by the backward recursion of r_t. Furthermore, it utilizes the forward recursion result $\widehat{x}_{t|t-1}$ so that the backward

recursion length is kept minimum.

Amemiya, T., Wu, R. Y., 1972. The effect of aggregation on prediction in the autoregressive model. Journal of the American Statistical Association 67 (339), 628–632.

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

Breitung, J., Swanson, N., 2002. Temporal aggregation and spurious instantaneous causality in multiple time series models. Journal of Time Series Analysis 23 (6), 651–665.

Camacho, M., Perez-Quiros, G., 2010. Introducing the euro-sting: Short-term indicator of euro area growth. Journal of Applied Econometrics 25 (4), 663–694.

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

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

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).

de Jong, P., 1989. Smoothing and interpolation with the state-space model. Journal of the American Statistical Association 84 (408), 1085–1088.

Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130 (2), 337–364.

Diebold, F. X., Rudebusch, G. D., Borag[caron]an Aruoba, S., 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. Journal of Econometrics 131 (1-2), 309–338.

Durbin, J., Koopman, S. J., 2002. A simple and efficient simulation smoother for state space time series analysis. Biometrika 89 (3), 603–615.

Ghysels, E., 2012. Macroeconomics and the reality of mixed frequency data. Tech. rep., SSRN No. 2069998.

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

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

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press: Princeton.

Harvey, A. C., Pierse, R. G., 1984. Estimating missing observations in economic time series. Journal of the American Statistical Association 79 (385), 125–131.

Hyung, N., Granger, C. W., 2008. Linking series generated at different frequencies. Journal of Forecasting 27 (2), 95–108.

Kadiyala, K. R., Karlsson, S., 1997. Numerical methods for estimation and inference in bayesian var-models. Journal of Applied Econometrics 12 (2), 99–132.

Kuzin, V., Marcellino, M., Schumacher, C., 2011. Midas vs. mixed-frequency var: Nowcasting gdp in the euro area. International Journal of Forecasting 27 (2), 529–542.

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

Marcellino, M., 1999. Some consequences of temporal aggregation in empirical analysis. Journal of Business & Economic Statistics 17 (1), 129–136.

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

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

Mittnik, S., Zadrozny, P. A., 2004. Forecasting quarterly german gdp at monthly intervals using monthly ifo business conditions data. CESifo Working Paper Series.

Nelson, C. R., Siegel, A. F., 1987. Parsimonious modeling of yield curves. The Journal of Business 60 (4), 473–89.

Pavia-Miralles, J., 2010. A survey of methods to interpolate, distribute and extrapolate time series. Journal of Service Science and Management 3, 449–463.

Schorfheide, F., Song, D., 2012. Real-time forecasting with a mixed-frequency var. Tech. Rep. 701, Federal Reserve Bank of Minneapolis.

Silvestrini, A., Veredas, D., 2008. Temporal aggregation of univariate and multivariate time series models: A survey. Journal of Economic Surveys 22 (3), 458–497.

Tiao, G. C., Wei, W. S., 1976. Effect of temporal aggregation on the dynamic relationship of two time series variables. Biometrika 63 (3), 513–523.

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.