Bayesian Mixed Frequency VAR's

Bjørn Eraker**
Ching Wai (Jeremy) Chiu*
Andrew Foerster*
Tae Bong Kim*
Hernan Seoane *

September 1, 2008

Abstract

Economic data can be collected at a variety of frequencies. Typically, estimation is done using the frequency of the coarsest data. This paper discusses how to combine data of different frequencies in estimating Vector Autoregressions (VAR's). The method is based on Bayesian Gibbs sampling using a missing data formulation for coarsely observed data. Our approach has the primary advantage that it increases the accuracy of parameter estimates relative to estimates obtained from coarse data. We demonstrate this through an example where we estimate a model for economic growth based on quarterly observations of GDP, monthly industrial production, and yield curve data. Estimates of the posterior standard deviations are uniformly lower for the BMF estimator. Experiments with artificially simulated data further documents the efficiency gains.

JEL classification: C11, C22, G10, E27

*Duke University, Department of Economics, ** Wisconsin School of Business, Department of Finance. Corresponding author: beraker@bus.wisc.edu. Comments welcome.

1 Introduction

Economic data can rarely be collected at the same instances in time. Data from liquid markets can be collected almost continuously, while aggregate macro data in many cases are collected at quarterly or annual frequencies. These mixed sampling frequencies represent a significant challenge to time-series econometricians: how do we combine data that can be sampled at different frequencies? A typical approach to dealing with mixed frequency observations is to simply "throw away" the high frequency data, using the coarsest sampling frequency as the common denominator for econometric modeling. Since this procedure implies that high frequency data are discarded, it is reasonable to ask if there is a way to avoid discarding the data would lead to increased efficiency in parameter estimates and econometric forecasts.

This paper explores Bayesian estimation of mixed frequency Vector Autoregressive models (VAR's) - arguably the basic workhorse of Macro time-series econometrics. We derive a simple, yet very powerful method for Markov-Chain-Monte-Carlo sampling from the posterior distributions of the parameters in VAR models. The posterior is conditioned on data observed at mixed frequencies. The method is based on the assumption that the econometrician simply does not observe the high frequency realizations of the low frequency data. We accordingly treat these data as missing values. Consistent with the standard utilization of missing values in Bayesian econometrics, we construct a Gibbs sampler that produces alternate draws from the missing data and the unknown parameters in the model, respectively. Under assumptions about normally distributed exogenous shocks, the linear structure of the VAR model produces a Gibbs sampler that requires draws from Gaussian distributions for estimating the missing data, as well as Gaussian and inverse Wishart conditional posterior distributions for the parameters in the model. Our Gibbs sampler thus requires only simulation from known densities which makes it extremely simple to implement.

Despite the obvious problem of mixed arrival frequencies of economic data, surprisingly little work has been done to address the resulting econometric challenges. One exception is the approach suggested in Miller and Chin (1996) which uses monthly data to improve quarterly variable forecasts. Their method is a iterative procedure. In a first step they use a quarterly model to forecast variables at quarterly frequency, in this model they include all the relevant variables at a quarterly frequency. Next, they forecast the relevant monthly variables and design a monthly model to forecast the quarterly data. Finally, they combine the two different forecasts using efficient weights (estimated using available data) to produce a final forecast. Our method, in contrast, uses all the relevant information to make multi-frequency forecasting for each variable in the VAR. This implies that our method is exploiting all the available information to forecast any variable in the model.

A notable contribution to the issue of estimating mixed frequency models is the recent body of work on MIDAS (Mixed frequency Data Analysis) described in Ghysels, Santa-Clara, and Valankov (2004), Andreou, Ghysels, and Kourtellos (2007), Ghysels, Sinko, and Valankov (2007), and references therein. The MIDAS method allows regressions of some low frequency variable onto high frequency variables. For example, Ghysels, Santa-Clara, and Valankov (2004) study the predictability of stock returns over relative low frequencies (monthly or quarterly) from high frequency volatility estimates. To our knowledge MIDAS represents the first systematic attempt to incorporate data sampled at different frequencies for econometric analysis, there are many methods that deal with missing data in econometrics. For example, Harvey and Pierse (1984) discuss the use of Kalman filtering for linear VAR models and note that missing observations can be incorporated by simply skipping a term from the updating equation whenever an observation is missing. In our context, the missing observations will occur at a regular frequency, but not for all variables in the VAR. The fact that the model is gaussian and linear makes it possible to formulate a state-space form. The resulting likelihood function, however, is non-linear and non-gaussian over a potentially very large parameter space. Analyzing such likelihood functions is difficult both from frequentist and Bayesian viewpoints. By contrast, the simulation based approach suggested in this paper solves the problem by just sampling from standard densities.

A second advantage of the Bayesian simulation approach is that it easily generalizes to more complicated models. For example, both financial market and macro variables are known to exhibit time-varying volatility. It is simple to incorporate stochastic volatility into our setting using methods such as described in Jacquier, Polson, and Rossi (1994). This require an additional step in the Gibbs sampler for updating the unobserved volatility path. In a similar vein, it is easy to incorporate heavy tailed error distributions in the form of mixture of normalsor t distributions, Markov mixtures as in Albert and Chib (1993) or jumps (Eraker, Johannes, and Polson (2003)).

In putting our method to work with real data, we consider a model of macro-economic output and interest rates. Specifically, we look at GDP data recorded quarterly, monthly records of industrial production, and high frequency interest rate data in the form of nominal short rates and a slope-of-the-yield-curve variable. We compare our method to the standard approach of sampling at the coarser frequency which we coin the "throw away estimator" (TAE). TAE can be used to identify parameters in the VAR even if the econometrician assumes that the true VAR evolve at some higher frequency than what is used for estimation because Gaussian VAR's are closed under temporal aggregation. We thus use TAE as a basic benchmark to our Bayesian Mixed Frequency (BMF) estimator. We find that BMF outperforms TAE using our real data example in that the posterior standard deviations are uniformly smaller for BMF than TAE. The posterior standard deviations decrease as we increase the sampling frequency for BMF from monthly to bi-weekly. It is typically the case that the posterior standard deviations decrease the most for parameters relating to the dynamics in the high-frequency observed variables (i.e., interest rates in our case). It is also the case, perhaps surprisingly, that posterior standard deviation decrease for parameters the low frequency variables (i.e., GDP). Thus, the increased accuracy that we experience by incorporating additional high frequency data is not contained locally to the high frequency variables but imply increased accuracy for the low frequency variables as well.

The paper also reports extensive numerical simulation results for a range of parameter constellations. The experiments generate data sets of different length using different, pseudo-true parameter values. We run the TAE and BMF estimators on the artificial data and record the posterior means from each method. In interpreting these as frequentist point estimates we compare the performance of the two estimators in mean square deviation from the true values. We find that BMF uniformly dominates TAE for all the parameter constellations.

We also investigate the difference between the throw away estimator and our Bayesian approach in estimating impulse response functions. We find that the decrease in parameter uncertainty associated with BMF is typically reflected in tighter confidence bands for the impulse response functions. Among other things, this implies that our method allows for sharper conclusions about impact of economic policies.

The remainder of the paper is organized as follows: In section two we discuss the construction of a Gibbs sampler for the model. Section three presents empirical evidence using real and artificial data. Section four concludes.

2 Econometric Methodology

This section discussed the main algorithm of data augmentation and estimation procedure in the presence of missing data. While our method applies easily to VAR's with multiple lags, we present the case of a first order VAR here. Thus the model is

$$y_t = A + By_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma)$$

where $\dim(y_t) = N$. Denote $y_t = (x_t, z_t)$ where $\dim(x_t) = N_x$ and $\dim(z_t) = N_z$ such that $N_z + N_x = N$ and suppose x_t is a fully observed variable and z_t is a variable with missing data. We assume in the following that z and x are recorded at two frequencies[1]. For example, x_t is observed monthly but z_t is observed quarterly. In this case the missing data are $\{\hat{z}_1, \hat{z}_2, \hat{z}_4, \hat{z}_5, \hat{z}_7, ...\}$. Let $z_{\backslash t}$ denote all elements of z except the t^{th} ones. Let $\hat{Y}^(i)$ denote the collection observed and augmented data at iteration i.

Given prior distributions and initial values of the parameters, the ith iteration of our MCMC algorithm reads

- step 1 : for $t = 1, .., T$, draw missing data $\hat{z}_t^{(i)} \mid x, \hat{z}_{\backslash t}^{(i-1)}, A^{(i)}, B^{(i)}, \Sigma^{(i)}$.
- step 2 : draw $A^{(i)}, B^{(i)} \mid \hat{Y}^{(i)}, \Sigma^{(i-1)}$
- step 3 : draw $\Sigma^{(i)} \mid \hat{Y}^{(i)}, A^{(i)}, B^{(i)}$

where $\hat{z}_{\backslash t}^{(i-1)}$ is the vector of most recently updated missing values. For example if we update in a consecutive order, we have $\hat{z}_{\backslash t}^{(i-1)} = (z_1^i, z_2^i, .., z_{t-1}^i, z_{t+1}^{(i-1)}, .., z_T^{(i-1)})$.

Note that except for the first step the procedure can be viewed as a standard normal linear model which yields normal and wishart posterior distributions. Drawing from these distributions is straightforward, as illustrated next.

[1] In general, our method can be extended to a multi-frequency dataset, which may contain weekly, monthly and quarterly data.

2.1 Drawing parameters

2.1.1 Updating the parameter matrices A and B

For the sake of drawing parameters in each iteration, we need to derive a posterior density that is composed of the likelihood function and the prior. Let us first examine the likelihood function of the parameter matrices, A and B. For notational simplicity, variables are defined as follows.

$$
\begin{aligned}
\underset{N\times(T-1)}{Y} &\equiv \begin{bmatrix} x_2 & x_3 & \cdots & x_T \\ z_2 & z_3 & \cdots & z_T \end{bmatrix} \\
\underset{N\times(N+1)}{\beta} &\equiv \begin{bmatrix} A & B \end{bmatrix} \equiv \begin{bmatrix} A_x & B_{xx} & B_{xz} \\ A_z & B_{zx} & B_{zz} \end{bmatrix} \\
\underset{(N+1)\times(T-1)}{X} &\equiv \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_{T-1} \\ z_1 & z_2 & \cdots & z_{T-1} \end{bmatrix}
\end{aligned}
$$

So the likelihood can be written as

$$
\begin{aligned}
L\left(\beta|y_T,...,y_1,\Sigma\right) &\propto \prod_{t=0}^{T} \exp\left\{-\frac{1}{2}\left(y_t - A - By_{t-1}\right)'\Sigma^{-1}\left(y_t - A - By_{t-1}\right)\right\} \\
&= \exp\left[-\frac{1}{2} tr\left(\left(Y-\beta X\right)'\Sigma^{-1}\left(Y-\beta X\right)\right)\right] \\
&\propto \exp\left[-\frac{1}{2} tr\left(\left(XX'\right)\left(\beta - YX'\left(XX'\right)^{-1}\right)'\Sigma^{-1}\left(\beta - YX'\left(XX'\right)^{-1}\right)\right)\right]
\end{aligned}
$$

Proportionality in the last step is obtained by removing all the constant terms and using the trace properties[2].

The conditional posterior for β is

[2] $tr\left(\underset{n\times n}{ABC} - \underset{n\times n}{DEF}\right) = tr(\underset{n\times m}{A}\underset{m\times k}{B}\underset{k\times n}{C}) - tr(\underset{n\times m}{D}\underset{m\times k}{E}\underset{k\times n}{F}) = tr(\underset{k\times k}{CAB}) - tr\left(\underset{k\times k}{FDE}\right) = tr\left(\underset{k\times k}{CAB} - \underset{k\times k}{FDE}\right)$

$$p\left(\beta|y_T,...,y_1,\Sigma\right)\sim MN\left(YX'\left(XX'\right)^{-1},\left(XX'\right)^{-1},\Sigma\right),$$

a matrix normal density. The vectorized form of β thus has density

$$p\left(vec\left(\beta\right)|y_T,...,y_1,\Sigma\right)\sim N\left(vec\left(YX'\left(XX'\right)^{-1}\right),\left(XX'\right)^{-1}\otimes\Sigma\right)\equiv N\left(\mu_\lambda,\Sigma_\lambda\right)$$

Assuming a conjugate prior[3]

$$\pi\left(vec\left(\beta\right)\right)\sim N\left(\mu_\Omega,\Sigma_\Omega\right)$$

We derive our posterior density for $vec\left(\beta\right)$ as follows:

$$p\left(vec\left(\beta\right)|y_T,...,y_1,\Sigma\right)\sim N\left(\left(\Sigma_\lambda^{-1}+\Sigma_\Omega^{-1}\right)^{-1}\left(\Sigma_\Omega^{-1}\mu_\Omega+\Sigma_\lambda^{-1}\mu_\lambda\right),\left(\Sigma_\lambda^{-1}+\Sigma_\Omega^{-1}\right)^{-1}\right)\quad(1)$$

where

$$\mu_\lambda=vec\left(YX'\left(XX'\right)^{-1}\right)\ ,\ \Sigma_\lambda=\left(XX'\right)^{-1}\otimes\Sigma.\quad(2)$$

Sampling β is obviously straightforward.

2.1.2 Updating the covariance matrix Σ

The conditional posterior for Σ is of the conjugate inverse Wishart from and follows straightforwardly.

Let

$$\epsilon_t^{(i)}=Y_t^{(i)}-A^{(i)}-B^{(i)}Y_{t-1}^{(i)}$$

denote the estimated error term at iteration i of the Gibbs sampler. The conditional posterior distribution of Σ is

$$p(\Sigma\mid Y^{(i)},A^{(i)},B^{(i)})\propto\Sigma^{-n/2}\exp(-\sum_t\epsilon_t^{(i)}\Sigma^{-1}\epsilon_t^{(i)\prime})p(\Sigma)\quad(3)$$

[3]In our baseline estimation, we have a zero vector for the prior mean and ten times the identity matrix for the prior variance. This setup constitutes an uninformative prior.

where $p(\Sigma)$ denotes the prior. By choosing an inverse Wishart prior with prior mean Ψ and degrees of freedom m the conditional posterior is

$$\Sigma \mid Y^{(i)}, A^{(i)}, B^{(i)} \sim IW(\sum_{t=1}^{n} \epsilon_t^{(i)} \epsilon_t^{(i)'} + \Phi, n+m), \tag{4}$$

another inverse Wishart. Algorithms for sampling from IW distributions are readily available.

2.2 Sampling the latent data

Our method requires drawing the latent data from its conditional posterior distribution. We proceed by drawing a single tth element in one operation. Thus, we wish to draw $\hat{z}_t^{(i)} \mid x, \hat{z}_{\setminus t}^{(i-1)}, A^{(i)}, B^{(i)}, \Sigma^{(i)}$. Write

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} A_x \\ A_z \end{bmatrix} + \begin{bmatrix} B_{xx} & B_{xz} \\ B_{zx} & B_{zz} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

where

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}\right).$$

WE show in appendix that the conditional density for z_t is the multivariate normal

$$\begin{aligned}
\hat{z}_t \mid \hat{z}_{\setminus t}, x, \Theta &\sim N(M, W_1) \\
W_1 &= \Sigma^{zz} + B_{zz}'\Sigma^{zz}B_{zz} + B_{xz}'\Sigma^{xx}B_{xz} + B_{zz}'\Sigma^{zx}B_{xz} + B_{xz}'\Sigma^{xz}B_{zz} \\
W_2 &= -\Sigma^{zx}\left(x_t - A_x - B_{xx}x_{t-1} - B_{xz}z_{t-1}\right) \\
&\quad +\Sigma^{zz}\left(A_z + B_{zx}x_{t-1} + B_{zz}z_{t-1}\right) \\
&\quad +B_{xz}'\Sigma^{xx}\left(x_{t+1} - A_x - B_{xx}x_t\right) \\
&\quad +B_{zz}'\Sigma^{zx}\left(x_{t+1} - A_x - B_{xx}x_t\right) \\
&\quad +B_{xz}'\Sigma^{xz}\left(z_{t+1} - A_z - B_{zx}x_t\right) \\
&\quad +B_{zz}'\Sigma^{zz}\left(z_{t+1} - A_z - B_{zx}x_t\right)
\end{aligned}$$

where

$$M \equiv W_1^{-1} W_2.$$

It is now straightforward to construct Gibbs sampling to draw z_t. One possibility is to draw the elements in a consecutive order. Another approach is to draw odd and even element of z alternately, which can easily be implemented in a vectorized programming environment.

2.3 The throw-away-estimator

The standard approach to mixed data frequency estimation is to delete the high frequency data such that the VAR is estimated at whichever frequency is jointly available. Thus, in estimating a model with, say, interest rates and GDP growth, one would sample both variables at the quarterly frequency since GDP is released quarterly. Of course, interest rates can be collected at almost any frequency. In choosing a quarterly sampling frequency for interest rates therefore, one throws away information contained in high frequency (hourly, daily, weekly) interest rate data. We coin this approach the throw-away-estimator (TAE).

The throw away estimator is not an unreasonable estimator. In particular, it is true that the estimator can be used to estimate the true values of the parameters in a VAR even if the true VAR is assumed to evolve at a higher frequency than that used for estimation. This is so because the model $Y_{t+1} = A + BY_t + \epsilon_t$ is closed under temporal aggregation, so that $Y_{t+n} = A_n + B_n Y_t + \epsilon_{t_n}$ with $\epsilon_{t_n} \sim N(0, \Sigma_n)$ where the coefficients A_n, B_n and Σ_n are given by

$$A_n = \sum_{s=1}^{n} B^{s-1} A \qquad (5)$$

$$B_n = B^n \qquad (6)$$

$$\Sigma_n = \sum_{s=1}^{n} B^{s-1} \Sigma B^{s-1\prime} \qquad (7)$$

This means, for example, that we can recover the implied quarterly dynamics from monthly estimates by applying (5) to (7) directly. We can also invert these equations to find, say, monthly estimates implied by quarterly estimates. If for example B_q is an estimate of B obtained using the TAE at quarterly sampled data, we can convert this into a monthly equivalent by inverting (6). The implied monthly estimate is thus $(n = 3)$ $B_q^{\frac{1}{3}}$. Equations

(5)-(7) thus allow us to compare estimates obtained through BMF with estimates from TAE. Such comparisons are best done by transforming the posterior simulations and then computing the quantity of interest. For example, in order to compare posterior standard deviations from quarterly TAE to monthly BMF, we compute the sample standard deviation $\frac{1}{G}\sum_i^G(B^{(i)})^{\frac{1}{3}}$ for G iterations of the Gibbs sampler. This now gives an estimate of the monthly implied standard deviation from quarterly TAE estimates.

3 Empirical Results

In the following we discuss the basic results of a VAR(1) model which makes use of financial and macro data observed at biweekly, monthly, and quarterly frequencies. We are primarily interesting in formulating a model for forecasting GDP growth. In doing so, we want to use higher frequency variables including the monthly releases of industrial production and interest rate variables. Specifically, we include short term interest rates in the form of one year constant maturity zero coupon yields. This short term interest rate variable should be able to reveal information about the Federal Reserves' expectations about growth if they use a standard Taylor rule in determining short term interest rates. The second variable in the VAR is the slope of the yield curve, defined as the difference between the seven and one year zero coupons. We include this variable as it has been frequently noted that an inverted yield curve (negative slope) tend to precede recessions. Both interest variables are from the data-set by Gurkaynak, Sack and Wright (2006). The third variable in our VAR is the industrial production (IP) index which is released monthly. Since IP is highly correlated with GDP, we include this variable to be able to predict quarterly GDP growth from the higher frequency IP data. Both industrial production and GDP are included in the form of logarithmic twelve month growth rates. Table 1 presents the summary statistics for sample data.

The data used in both TAE and BMF estimation are similar in nature, although the lower frequency variables change slightly depending upon the estimation method used. The relevant yield curve data are drawn from the last Friday of each bi-weekly period, month, or quarter. The industrial production (IP) index is in monthly frequency, and it is converted to the 12-month growth rate. Notice that the the relevant IP series for TAE estimation is the growth rate computed from the last month of each quarter. Finally, the real GDP growth rate is computed over a four-quarter span. We estimate the VAR over the

Table 1: Descriptive Statistics

The table reports the basic statistics for the one-year interest rate, slope of the yield curve, quarterly GDP growth and monthly growth in industrial production.

	Mean	Std.Dev.	Autocorrelation
1-year	6.10	2.79	0.998
Slope	0.75	1.00	0.991
ln(ΔGDP)	3.3	2.1	0.86
ln(ΔIP)	3.3	3.7	0.95

Table 2: Parameter Estimates, A

The table reports posterior means and standard deviations (in parenthesis) of the VAR constant terms, A. The variables are, in order of appearance, the one year zero coupon bond yield, the slope of the yield curve, log growth in industrial production, and log growth of GDP. All parameters are at the monthly frequency.

	BMF:Bi-Weekly	BMF:Monthly	TAE:Quarterly
A_1	-0.0	0.01	0.06
	(0.07)	(0.08)	(0.13)
A_2	0.11	0.08	0.07
	(0.04)	(0.05)	(0.07)
A_3	-0.3	-0.2	-0.2
	(0.18)	(0.18)	(0.23)
A_4	0.15	0.15	0.18
	(0.11)	(0.11)	(0.14)

time period of July 1962 and June 2007, with a total of 1087 bi-weekly observations, 544 monthly observations, and 182 quarterly observations. Summary statistics for the variables are presented in table 1.

Tables 2-4 display the VAR estimates for the BMF estimator using bi-weekly and monthly observations, and the TAE using quarterly observations. In all cases the ordering of the variables is as follows: yield curve intercept, yield curve slope, the growth rate of IP, and the growth rate of real GDP. Table 2 contains the estimates of the constant terms in the VAR. The table does show that the three estimators produce somewhat different posterior means for the constant terms. Noticeably however, the posterior standard deviations are uniformly smaller as we go from the quarterly TAE to the monthly BMF to the bi-weekly BMF estimators.

Table 3: Parameter Estimates, B

The table reports posterior means and standard deviations of the VAR coefficients, B, using data from the United States on GDP, Industrial Production (IP), one year zero coupon yield, and the slope of the yield curve. The sample is 1972 through 2007. The table reports posterior means and standard deviations using the TAE and BMF estimators. All parameters are at the monthly frequency.

	BMF:Bi-Weekly	BMF:Monthly	TAE:Quarterly
B_{11}	0.991	0.988	0.984
	(0.008)	(0.009)	(0.012)
B_{12}	0.019	0.018	0.033
	(0.023)	(0.025)	(0.037)
B_{13}	0.007	0.012	0.021
	(0.010)	(0.011)	(0.018)
B_{14}	0.011	0.001	-0.01
	(0.018)	(0.019)	(0.031)
B_{21}	-0.00	0.000	0.001
	(0.005)	(0.005)	(0.006)
B_{22}	0.952	0.949	0.939
	(0.014)	(0.015)	(0.018)
B_{23}	-0.00	-0.00	-0.01
	(0.006)	(0.007)	(0.009)
B_{24}	-0.01	-0.00	0.001
	(0.011)	(0.012)	(0.015)
B_{31}	-0.00	-0.00	-0.01
	(0.020)	(0.019)	(0.020)
B_{32}	0.107	0.108	0.141
	(0.054)	(0.054)	(0.060)
B_{33}	0.858	0.865	0.900
	(0.024)	(0.024)	(0.030)
B_{34}	0.206	0.190	0.124
	(0.042)	(0.042)	(0.051)
B_{41}	-0.00	-0.00	-0.00
	(0.012)	(0.012)	(0.012)
B_{42}	0.130	0.126	0.111
	(0.035)	(0.035)	(0.038)
B_{43}	0.068	0.067	0.036
	(0.017)	(0.017)	(0.018)
B_{44}	0.853	0.854	0.899
	(0.030)	(0.030)	(0.032)

Table 4: Parameter Estimates, Σ

The table reports posterior means and standard deviations of Σ using data from the United States on GDP, Industrial Production (IP), one year zero coupon yield, and the slope of the yield curve. The sample is 1972 through 2007. The table reports posterior means and standard deviations using the TAE and BMF from Monthly data. All parameters are converted to quarterly frequency.

	TAE		
1.016	-0.43	0.402	0.246
(0.112)	(0.054)	(0.139)	(0.083)
-0.43	0.306	-0.15	-0.05
(0.054)	(0.033)	(0.074)	(0.044)
0.402	-0.15	2.988	0.992
(0.139)	(0.074)	(0.326)	(0.159)
0.246	-0.05	0.992	1.085
(0.083)	(0.044)	(0.159)	(0.119)
	Monthly BMF		
0.715	-0.31	0.195	0.184
(0.046)	(0.025)	(0.068)	(0.054)
-0.31	0.252	-0.09	-0.03
(0.025)	(0.017)	(0.041)	(0.033)
0.195	-0.09	2.781	0.654
(0.068)	(0.041)	(0.192)	(0.103)
0.184	-0.03	0.654	1.038
(0.054)	(0.033)	(0.103)	(0.107)

Table 3 shows the estimates of the B coefficients. Again it is the case that our BMF estimator dominates the TAE. For example, for $B_{1,1}$ which measures the persistence in the one year zero coupon yield, the three methods produce posterior means that are similar while the standard deviations decrease from 0.012 for TAE to 0.008 for the bi-weekly BMF estimator. Figure 1 displays the kernel density estimates of the marginal posterior densities for the implied monthly VAR coefficients using the high frequency BFM estimators and the quarterly TAE. The pictures echoes the results from table 3: posterior densities obtained through BMF have visibly lower variance.

Table 4 reports estimates of the residual covariance matrix Σ from monthly BMF and quarterly TAE. Here we see a very substantial reduction in the posterior standard deviations for almost all the coefficients. For example, the standard deviation of $\Sigma_{1,1}$ is more than halved from 0.112 to 0.046. For other coefficients the standard deviations are typically lower by a factor of two thirds. The smallest improvement is for $\Sigma_{4,4}$, the variance of shocks to GDP growth which goes from 0.119 to 0.107.

3.1 Impulse Response

One of the primary advantages of VAR's is the ability to study the causal relationships between variables through impulse response functions. Figure 2 displays a comparison of impulse response functions when the BMF bi-weekly estimate is used versus the TAE quarterly estimate. The figure shows the 5,50 and 95% pointwise posterior percentiles for both estimators. Despite the similarities in the shapes of the impulse responses between the two estimates, the accompanied confidence bands suggest that the impulse responses produced using BMF are more precise than those produced with TAE. The most marked increase in precision occurs in identifying the effects of the various shocks on the interest rate level and slope. Since these are the most frequently observed data, the ability to include these more frequent data in the estimation using BMF produces uniformly tighter confidence bands. For the shocks to the interest rate level and slope, the decrease in the size of the confidence bands is relatively small, which reflects the fact that since those data are observed at the same frequency, adding more frequent observations doesn't have as much of an impact.

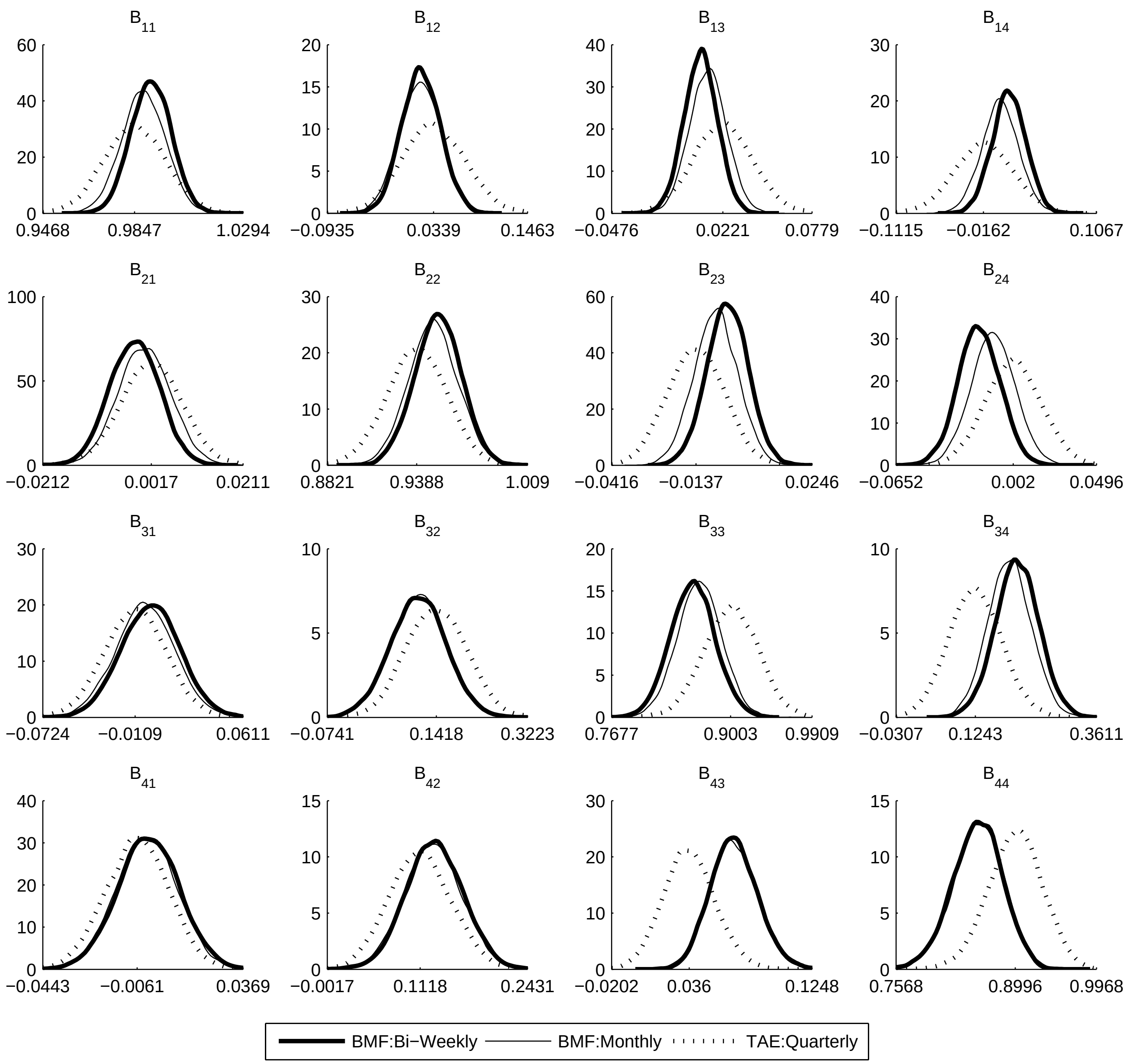

Figure 1: Posterior densities for TAE (quarterly), BMF (monthly) and BMF (bi-weekly).

The impact of a shock to industrial production on interest rate (labeled "IP on Level") is interesting. It is "insignificant" for the TAE in that its lower 5 percentile is well below zero. For BMF the lower 5 percentile is greater than zero at 20-40 quarters. A similar pattern is found from the impact of GDP shocks on the level of interest rates. These plots thus suggest that while the TAE estimator will be inconclusive about the impact of macro

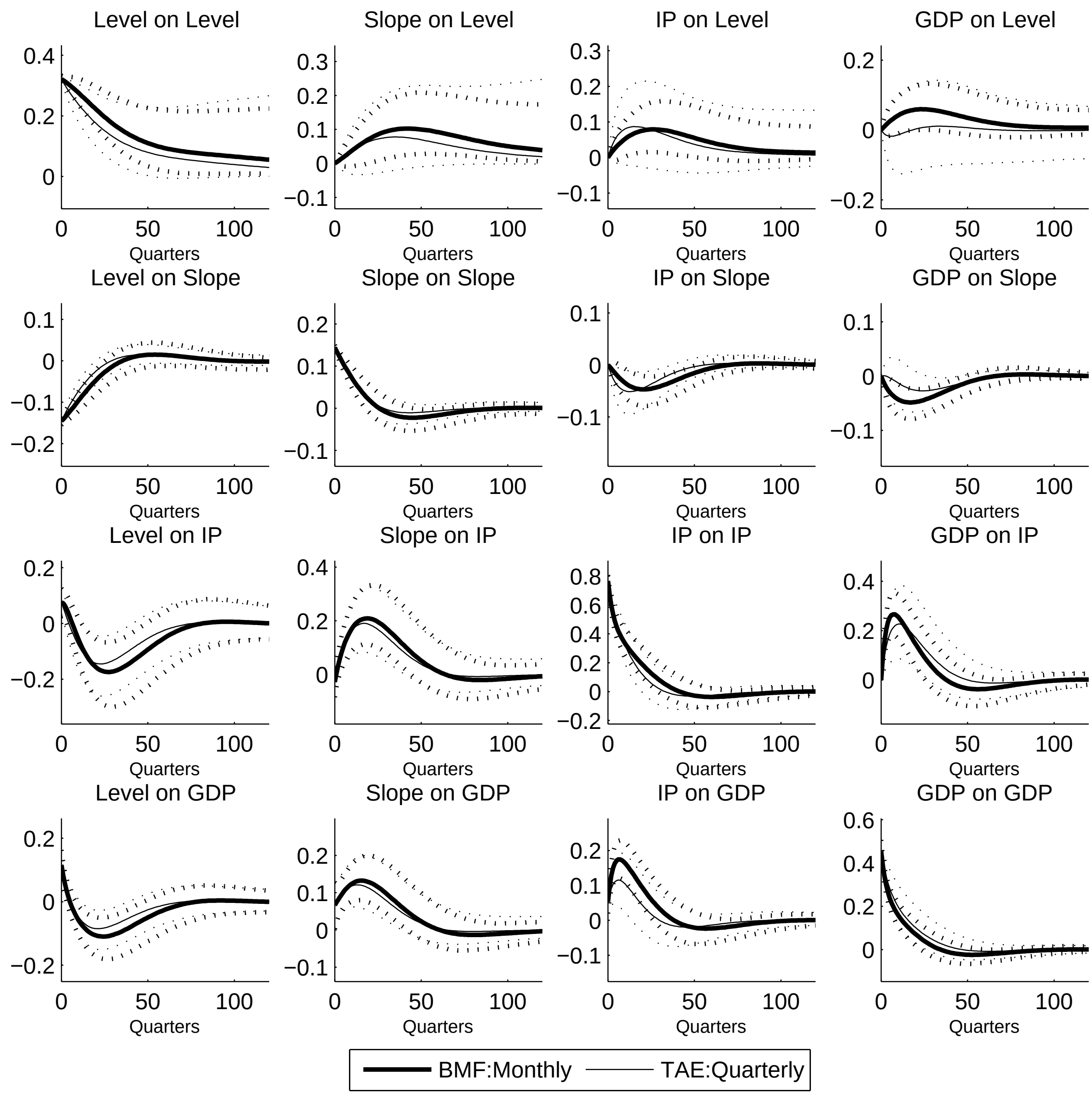

Figure 2: Impulse Response functions and 95% coverage.

shocks on interest rates, our BMF estimator is able to pinpoint a statistically significant impact of macro shocks on subsequent interest rates.

3.2 Small Sample Evidence

In the following we examine the performance of our estimator using simulated data. The purpose is to see how our method fares relative to alternatives when the objective is to recover parameter estimates, say the posterior mean, that are as close as possible in a least square sense to the truth. This is very much a frequentist way of thinking. The exercise should accordingly be interpreted as a small sample study of the posterior mean as a frequentist parameter estimate.

Table 5 reports the root mean square error of the parameters estimated using BMF and TAE. As can be seen, BMF attains smaller root mean square error than TAE for all the parameter estimates regardless sample size, correlation and diagonal elements of matrix B. Several additional features should be mentioned.

First, BMF improvements are higher for small sample than for larger sample estimation. Second, if we are performing a small sample estimation the smaller the correlation of x and z and the smaller the diagonal elements of B, the better performance of BMF is achieved. As can be seen in the first block of the table 5, a sample size of 20 and autocorrelation of zero-, the improvement of BMF over TAE goes from 17.5% to 72.8%. However, for the same sample size and B with a correlation of 95% , the improvement of BMF goes from 9.43% to 26.4%. We can see similar results when the diagonal elements of B are increased from .90 to .99. BMF is still an improvement to TAE but the relative gains in root mean square error are smaller, running 4.18% to 33% for the case of no correlation and from 2.16 % to 10.1% for a correlation of 95%. Finally, with high diagonal elements of B and bigger sample size, BMF has a better performance than TAE but relatively smaller than in the previous cases. However, it is still important in many cases. Summarizing, Table 1 shows that BMF is always an improvement to TAE. The improvement is higher for the smaller sample size and the smaller correlation between the variables of the system.

4 Concluding Remarks

This paper considers estimation of first order VAR's using data sampled at mixed frequencies. We propose a new method based on Gibbs sampling the unobserved data at the high frequency. We find that both real data and simulation experiments demonstrate that our

Table 5: Accuracy of BMF vs TAE

The table reports root mean squared of parameters estimated using BMF and TAE relative to true values over different sample length and pseudo true parameter values. In each case, $A = (0, 0)$ and $\text{Corr}(x, z) = \rho$. T is sample size. The row labeled '%' gives the percentage difference in in BMF relative to TAE.

			$B = [.9 - .04; .04 .9]; \rho = 0$			
			$T = 20$			
TAE	0.000478	0.000766	0.0231	0.0322	0.0701	0.133
BFM	0.000395	0.000414	0.0188	0.0184	0.0332	0.0362
%	-17.5	-45.9	-18.9	-43	-52.7	-72.8
			$T = 80$ (these numbers are wrong)			
TAE	0.00066	0.000748	0.0295	0.0435	0.0935	0.155
BFM	0.000502	0.000418	0.0247	0.021	0.0426	0.0387
%	-23.9	-44.1	-16.2	-51.7	-54.4	-75.1
			$B = [.9 - .04; .04 .9]; \rho = 0.95$			
			$T = 20$			
TAE	0.00117	0.00103	0.0645	0.0661	0.0481	0.0499
BFM	0.000862	0.000859	0.0552	0.0549	0.044	0.0452
%	-26.4	-16.4	-14.4	-16.9	-8.5	-9.43
			$T = 80$			
TAE	0.000946	0.00084	0.052	0.0517	0.0392	0.0384
BFM	0.000717	0.000724	0.0467	0.0466	0.0374	0.0369
%	-24.2	-13.9	-10.1	-9.84	-4.51	-3.98
			$B = [.99 - .04; .04 .99], \rho = 0$			
			$T = 20$			
TAE	0.0029	0.00238	0.0354	0.035	0.0249	0.0258
BFM	0.00194	0.00198	0.0324	0.0316	0.0239	0.0245
%	-33.2	-16.6	-8.3	-9.83	-4.18	-4.85
			$T = 80$			
TAE	0.00223	0.00176	0.0259	0.0258	0.0188	0.019
BFM	0.00151	0.00152	0.0255	0.0248	0.0189	0.0189
%	-32	-13.4	-1.65	-3.89	0.611	-0.349
			$B = [.99 .004; .004 .99]; \rho = 0.95$			
			$T = 20$			
TAE	0.0029	0.00299	0.0589	0.0554	0.057	0.0576
BFM	0.00263	0.00269	0.0576	0.0542	0.0554	0.0561
%	-9.23	-10.1	-2.16	-2.15	-2.82	-2.69
			$T = 80$			
TAE	0.00204	0.00194	0.0418	0.0402	0.0406	0.0416
BFM	0.00189	0.0018	0.0406	0.0388	0.0396	0.0405
%	-7.47	-7.43	-2.96	-3.43	-2.39	-2.48

approach produces more accurate estimates of model parameters than the basic approach of sub-sampling at the coarse data frequency.

Improved accuracy is not the only advantage of our mixed frequency estimator. Another strong point of our method is the ability to update forecasts of a coarsely observed variable in response to new arrival of data measured at high frequencies. One example is to update forecasts of next quarter GDP in response to daily or weekly measurements of financial market data. Our framework allows for a natural approach to incorporate high frequency observations to the low frequency forecasts. This again avoids the use of ad-hoc forecast revisions.

Our approach is implemented here using a first order vector auto-regression. It is straightforward to generalize our approach in various ways. First, higher order VAR's can be implemented straightforwardly as the conditional distribution for the unobserved data is still conditionally normal. The moments of these distributions involve some tedious algebra. Other extensions, such as for example time-varying volatility can be straightforwardly implemented and would typically involve only one additional step in the Gibbs sampler to draw the unobserved volatility path.

It is important to see how our approach could conceivably be used in connection with linear state-space models because state-space models comprise such an important class of time-series processes. So consider the case where some observation equation $Y_{t+1} = C + DX_{t+1}$, and state-equation $X_{t+1} = A + BX_t + \epsilon_{t+1}$. Assume a mixed sampling frequency for Y. We can nest this model within our VAR(1) framework by writing $Y^* = (Y, X)$ such that $Y^*_{t+1} = A^* + B^* Y^*_t + \epsilon^*_{t+1}$ and $A^* = (C, A)$,

$$B^* = \begin{pmatrix} 0 & D \\ 0 & B \end{pmatrix}.$$

To proceed to use our method for estimating this model we simulate, as before, the sparsely observed elements of Y but in addition, we now treat X as an unobserved variable. We may also interpret X as a variable observed with *zero frequency*. Importantly, our algorithm for drawing the missing data applies directly in this setting. To proceed to the second step of the Gibbs sampler which involves drawing the parameters, our algorithm needs slight modifications to impose the zero-constraints on B^*. Note that the estimation of VARMA models can be implemented using this approach. While our algorithm applies in general, identification considerations must be investigated on a case by case basis as is the usual case with such models. We will investigate generalizations along these lines in future work.

5 Appendix

We illustrate how $p\left(z_t|x_t, x_{t-1}, x_{t+1}, z_{t+1}, z_{t-1}, \Theta^{(i-1)}\right)$ can be derived analytically. First note that the conditional distribution is proportional to the product of the densities of $p\left(x_{t+1}, z_{t+1}|x_t, z_t, \Theta\right)$ and $p\left(x_t, z_t|x_{t-1}, z_{t-1}, \Theta\right)$:

$$p\left(z_t|x_t, x_{t-1}, x_{t+1}, z_{t+1}, z_{t-1}, \Theta\right) \propto p\left(x_{t+1}, z_{t+1}|x_t, z_t, \Theta\right) p\left(x_t, z_t|x_{t-1}, z_{t-1}, \Theta\right)$$

Recall

$$\begin{aligned} p\left(z_t, x_t|x_{t-1}, z_{t-1}, \Theta\right) &\propto l\left(y_t|y_{t-1}, \Theta\right) \\ &= \exp\left\{-\frac{1}{2}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}' \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}^{-1} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}\right\} \\ \text{where } \Sigma^{-1} &\equiv \begin{bmatrix} \Sigma^{xx} & \Sigma^{xz} \\ \Sigma^{zx} & \Sigma^{zz} \end{bmatrix} \\ v_1 &\equiv x_t - A_x - B_{xx}x_{t-1} - B_{xz}z_{t-1}, v_2 \equiv z_t - A_z - B_{zx}x_{t-1} - B_{zz}z_{t-1} \end{aligned}$$

Similarly,

$$\begin{aligned} p\left(x_{t+1}, z_{t+1}|x_t, z_t, \Theta\right) &\propto \exp\left\{-\frac{1}{2}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}' \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}^{-1} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right\} \\ \text{where } w_1 &= x_{t+1} - A_x - B_{xx}x_t - B_{xz}z_t, w_2 = z_{t+1} - A_z - B_{zx}x_t - B_{zz}z_t \end{aligned}$$

This implies

$$p\left(z_t|x_t, x_{t-1}, x_{t+1}, z_{t+1}, z_{t-1}, \Theta\right) \propto \exp\left\{-\frac{1}{2}\left(z_t - M\right)' W_1 \left(z_t - M\right)\right\}$$

where M,W_1, and W_2 have been defined in the previous subsection.

References

Albert, James H., and Siddharta Chib, 1993, Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts, *Journal of the American Statistical Association* 11, 1–15.

Andreou, Elena, Eric Ghysels, and Andros Kourtellos, 2007, Regression Models with Mixed Sampling Frequencies, *Working paper, UNC.*

Eraker, B., M. J. Johannes, and N. G. Polson, 2003, The Impact of Jumps in Returns and Volatility, *Journal of Finance* 53, 1269–1300.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valankov, 2004, There is a Risk-Return Tradeoff After All, *Journal of Financial Economics* forthcoming.

Ghysels, Eric, Arthur Sinko, and Rossen Valankov, 2007, MIDAS Regressions: Further Results and New Directions, *working paper, UNC.*

Gurkaynak, R. S., B. Sack, and J. Wright, 2006, The U.S. Treasury Yield Curve: 1961 to the Present, *working paper, Fed. Reserve Board.*

Harvey, A. C., and R. G. Pierse, 1984, Missing Observations in Economic Time Series, *Journal of the American Statistical Association* 79, 1–131.

Jacquier, E., N. G. Polson, and P. Rossi, 1994, Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* 12, 371–389.