ISSN 1936-5330

THE FEDERAL RESERVE BANK *of* KANSAS CITY
RESEARCH WORKING PAPERS

Estimating VAR's Sampled at Mixed or Irregular Spaced Frequencies: A Bayesian Approach

Ching Wai (Jeremy) Chiu, Bjørn Eraker, Andrew T. Foerster, Tae Bong Kim and Hernán D. Seoane

December 2011; Revised December 2012
RWP 11-11

RESEARCH WORKING PAPERS

Estimating VAR's Sampled at Mixed or Irregular Spaced Frequencies : A Bayesian Approach*

Ching Wai (Jeremy) Chiu† Bjørn Eraker‡ Andrew T. Foerster§ Tae Bong Kim¶

Hernán D. Seoane‖

December 10, 2012

Abstract

Economic data are collected at various frequencies but econometric estimation typically uses the coarsest frequency. This paper develops a Gibbs sampler for estimating VAR models with mixed and irregularly sampled data. The approach allows efficient likelihood inference even with irregular and mixed frequency data. The Gibbs sampler uses simple conjugate posteriors even in high dimensional parameter spaces, avoiding a non-Gaussian likelihood surface even when the Kalman filter applies. Two applications illustrate the methodology and demonstrate efficiency gains from the mixed frequency estimator: one constructs quarterly GDP estimates from monthly data, the second uses weekly financial data to inform monthly output.

Keywords: Gibbs Sampling, Mixed Frequency Data, VAR, Bayesian Estimation

JEL classification: C11, C22, G10, E27

---

*Previously titled "Bayesian Mixed Frequency VAR's." The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

†Duke University

‡Corresponding author: 5374A Grainger Hall, University of Wisconsin School of Business, Madison, WI 53706, beraker@bus.wisc.edu

§Federal Reserve Bank of Kansas City

¶Korea Development Institute

‖Universidad Carlos III de Madrid

1 Introduction

Economic data are rarely collected at the same instances in time. Data from liquid markets are available almost continuously, while aggregate macro data in many cases are available only at monthly, quarterly, or annual frequencies. Mixed and irregular sampling frequencies represent a significant challenge to time-series econometricians.

This paper develops Bayesian estimation of mixed frequency Vector Autoregressions (VAR's). The method is a simple, yet very powerful algorithm for Markov-Chain-Monte-Carlo sampling from the posterior distributions of the VAR parameters. The algorithm works in the presence of mixed frequency or irregularly spaced observations. The posterior is conditioned on data observed at mixed frequencies rather than simply data observed at the coarsest frequency. The method follows from the assumption that the econometrician simply does not observe the high frequency realizations of the low frequency data, and can accordingly treat these data as missing values. Consequently, and consistent with the standard utilization of missing values in Bayesian econometrics, the Bayesian Mixed Frequency (BMF) algorithm developed is a Gibbs sampler that produces alternate draws from the missing data and the unknown parameters in the model. Under typical assumptions about normally distributed exogenous shocks, the VAR's linear structure allows for draws from Gaussian conditional distributions for estimating the missing data, along with draws from Gaussian and inverse Wishart conditional posterior distributions for the parameters in the model. Since this Gibbs sampler requires only simulation from known densities, it is extremely simple to implement.

There has been much work addressing the issue of mixed frequency data from a variety of different approaches. An early contribution is the Kalman filtering approach introduced by Harvey & Pierse (1984), which notes that for linear VAR models, missing observations can be incorporated by simply skipping a term from the updating equation whenever an observation is missing. The VAR's linear and Gaussian form makes it straightforward to formulate a state-space form. However, the Kalman filter approach is potentially cumbersome when the missing data occur at irregular frequencies, especially if there are multiple series with missing data at differing frequencies. In addition, the Kalman filter yields a likelihood function that is non-linear and non-Gaussian over a potentially very large parameter space; analyzing such likelihood functions often proves difficult both from frequentist and Bayesian viewpoints. The BMF approach, by contrast, handles irregular and multiple missing series with ease, and the Gibbs sampling from standard densities makes the analysis of the resulting posterior densities very tractable.

Another approach, suggested by Miller & Chin (1996), uses monthly data to improve quarterly variable forecasts. The method is an iterative procedure that first uses quarterly observed variables to construct quarterly forecasts, then uses monthly observed variables to construct quarterly forecasts, and finally combines the two forecasts using estimated weights. Corrado & Greene (1988) show that adding monthly information via a monthly pooling procedure can improve quarterly forecasts. The BMF method, in comparison, uses all the relevant information to make multi-frequency forecasting for each variable in the VAR, so in the context of monthly and quarterly data, considers forecasts of monthly variables as well. This difference implies that BMF exploits all the available information to forecast any variable in the model, which offers the advantage of producing additional forecasts but also allows for inference based upon the effects of quarterly variables on monthly ones. Other papers that use bridging type models include Baffigi *et al.* (2004), and those that use bridging with factors, such as Giannone *et al.* (2008) and Angelini *et al.* (2008).

A growing body of work considering the estimation of mixed frequency models is the work on MIDAS (MIxed DAta Sampling) described in Ghysels *et al.* (2004), Andreou *et al.* (2010), Ghysels *et al.* (2007), among others. The MIDAS method allows regressions of a low frequency variable onto high frequency variables. For example, Ghysels *et al.* (2004) study the predictability of stock returns over relative low frequencies (monthly or quarterly) from high frequency volatility estimates, Andreou *et al.* (2009) consider the importance of daily data for forecasting monthly or quarterly real data, and Bai *et al.* (2010) expand MIDAS to deal with state-space models. While the MIDAS approach differs substantially from the Kalman filter approach of Harvey & Pierse (1984), it potentially suffers from the same pitfalls: handling observations that are irregularly spaced requires altering the estimated equations as in Francis *et al.* (2011), and larger systems may lead to significant numerical burdens.

In contrast to these methods, the approach taken in this paper is from a Bayesian perspective, and will consequently treat lower frequency data as missing. The missing data approach to higher frequency data has a history from both a Bayesian and frequentist perspective. Chow & Lin (1971) discuss how to interpolate time series using related series. Sims & Zha (2006b) and Leeper & Zha (2003), for example, use quarterly

GDP interpolated to monthly intervals in monthly VARs. Other mixed-frequency VAR approaches use stock-flow relationships for interpolation, such as Zadrozny (1988), Mittnik & Zadrozny (2005), or Mariano & Murasawa (2010). The BMF approach, on the other hand, follows the Bayesian approach to missing data, similar to, for example, Kim *et al.* (1998).

The traditional approach for dealing with mixed frequency data is to discard high frequency data and simply perform estimation on the coarsest frequency data. This estimation strategy potentially discards information contained in the higher frequency data, yet is used often in macro time series econometrics, especially within the context of VAR estimation, making it a useful benchmark. Indeed, a number of Bayesian and frequentist applications, including studying the effects of monetary policy, oil, or uncertainty shocks, include VAR's estimated at a monthly frequency despite the availability of higher-frequency data. The coarse estimation can be used to identify parameters in the VAR even if the econometrician assumes that the true VAR evolves at some higher frequency than that used for estimation because Gaussian VAR's are closed under temporal aggregation.

In addition to developing the methodology, this paper demonstrates the advantages of the BMF estimation method using numerical simulations and actual data. For numerical simulations over a range of parameter constellations, BMF uniformly dominates estimation using coarse sampling from the frequentist perspective of mean squared deviations from the true values. After considering simulated data, two applications highlight the advantages of BMF using actual data. The first involves a monthly and quarterly set of data on the real economy, and the second involves combining monthly real economic variables with high-frequency financial variables. In both contexts, BMF outperforms the coarsely sampled estimator in that the posterior standard deviations are smaller when using BMF. Which posterior standard deviations decrease the most depend on the application, it can either help accuracy for the low or the high frequency variables. The BMF approach also improves the estimation of impulse response functions, as the decrease in parameter uncertainty associated with BMF is typically reflected in tighter confidence bands for the impulse response functions. Among other things, this result implies that BMF can allow for sharper conclusions about the impact of economic policies or the effects of shocks.

The remainder of the paper is organized as follows: Section 2 discusses the construction of a Gibbs sampler for the model. Section 3 presents simulation based evidence for the advantages of using the BMF approach. Sections 4 and 5 present two examples of applications of mixed frequency estimation; the first example uses a monthly and quarterly model of the macroeconomy, and the second example uses weekly financial data along with monthly data on output. Finally, Section 6 concludes.

2 Econometric Methodology

This section discusses the main algorithm of data augmentation and estimation in the presence of missing data. The model is

$$y_t = A + \sum_{l=1}^{k} B_l y_{t-l} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma) \tag{1}$$

where $\dim(y_t) = N$. Denote the set of parameters $\Theta = (A, B, \Sigma)$, $y_t = (x_t, z_t)$ 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.

For simplicity, focus on the case when $k = 1$ – Appendix A extends the following discussion to the case when $k > 1$ – and assume that z and x are recorded at two frequencies, but note that the method applies to a multi-frequency dataset. In the example application in Section 4, x_t is observed monthly and z_t is observed quarterly; Section 5's example has x_t observed weekly and z_t observed monthly. In the case of monthly and quarterly observations, the missing data are $\{\hat{z}_1, \hat{z}_2, \hat{z}_4, \hat{z}_5, \hat{z}_7, ...\}$, where $\hat{z}_t$ denotes a sampled observation at time t. Let $\hat{z}$ denote the vector of observed and sampled data, $\hat{z}_{\backslash t}$ denote all elements of $\hat{z}$ except the t-th ones, and let $\hat{Y}^{(i)}$ denote the full collection of observed and sampled data at iteration i.

The BMF estimator is an application of Bayesian Gibbs sampling, which requires iterating over objects of interest, sampling those objects from known distributions conditional upon the remaining objects. In the current setup, the objects of interest are the missing observations, the matrices A and B, and the covariance matrix Σ. Given prior distributions and initial values of the parameters, the i-th iteration of the MCMC algorithm reads

- Step 1 : for $t = 1, .., T$, draw missing data $\hat{z}_t^{(i)} \mid x, \hat{z}_{\backslash t}^{(i-1)}, A^{(i-1)}, B^{(i-1)}, \Sigma^{(i-1)}$, where $\hat{z}_{\backslash t}^{(i-1)}$ is the vector of most recently updated missing values and $A^{(i-1)}, B^{(i-1)}, \Sigma^{(i-1)}$ are the latest draws of A, B, Σ, respectively. For example, in the case of consecutive updating, $\hat{z}_{\setminus t}^{(i-1)} = (\hat{z}_1^{(i)}, \hat{z}_2^{(i)}, .., \hat{z}_{t-1}^{(i)}, \hat{z}_{t+1}^{(i-1)}, .., \hat{z}_T^{(i-1)})$.[1]

- Step 2 : draw $A^{(i)}, B^{(i)} \mid \hat{Y}^{(i)}, \Sigma^{(i-1)}$

- Step 3 : draw $\Sigma^{(i)} \mid \hat{Y}^{(i)}, A^{(i)}, B^{(i)}$

The new step in the procedure is Step 1, which is drawing missing data given the parameters in the model and the fully observed data. Except for this first step, the procedure is a standard Normal linear model which, under conjugate priors, yield Normal and inverse Wishart posterior distributions. Since drawing from the relevant posteriors in Steps 2 and 3 is a well-known procedure, the following focuses on Step 1, sampling the missing data given a set of parameters.

2.1 Step 1: Sampling the Latent Data

Step 1 of the Gibbs sampler requires drawing the latent data from its conditional posterior distribution. It is convenient to draw a single t-th element in one operation, so the goal is to draw $\hat{z}_t^{(i)} \mid x, \hat{z}_{\setminus t}^{(i-1)}, A^{(i-1)}, B^{(i-1)}, \Sigma^{(i-1)}$. Separating the VAR into its components:

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} A_x \\ A_z \end{bmatrix} + \begin{bmatrix} B_{xx} & B_{xz} \\ B_{zx} & B_{zz} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix} \tag{2}$$

where

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}\right). \tag{3}$$

Appendix A shows that the conditional density for z_t is the multivariate normal

$$\hat{z}_t \mid \hat{z}_{\setminus t}, x, \Theta \sim N(M, W)$$

It is now straightforward to construct Gibbs sampling to draw $\hat{z}_t$, since it is also conditionally normal. One

---

[1] For notational simplicity, this notation assumes that all z_t are missing; implicitly the updating equation for non-missing data is just $\hat{z}_t^{(i)} = \hat{z}_t^{(i-1)}$.
In addition, the exact timing of updating is flexible; a possible alternative is to use the entire vector of missing values from the previous iteration, so $\hat{z}_{\setminus t}^{(i-1)} = (\hat{z}_1^{(i-1)}, \hat{z}_2^{(i-1)}, .., \hat{z}_{t-1}^{(i-1)}, \hat{z}_{t+1}^{(i-1)}, .., \hat{z}_T^{(i-1)})$. However, this timing tends to be less efficient.

possibility is to draw the elements in a consecutive order. Another approach is to draw odd and even elements of z alternately, which can easily be implemented in a vectorized programming environment.

2.2 Coarse Sampling Estimation

The standard approach to mixed 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, in the case of the example in Section 4, monthly and quarterly data, one would sample both variables at the quarterly frequency. In choosing a quarterly sampling frequency for the monthly data, one throws away information contained in the higher frequency data.

It should be noted that, in the context of many macroeconomic applications that use mixed frequency data, many simply perform estimation at the lowest frequency. Within the literature on monetary policy shocks, papers such as Christiano *et al.* (1996), Christiano *et al.* (1999), Sims & Zha (2006a), Sims & Zha (2006b), and Banbura *et al.* (2010), ignore the high frequency movements in interest rates and financial variables and estimate monthly or quarterly VARs. In studying the effects of oil price shocks, Kilian & Park (2009), Kilian *et al.* (2009), and Kilian & Vigfusson (2011) discard information in high-frequency price movements, again estimating using a monthly frequency. Monthly estimation of VARs is also used by Bloom (2009) to study the effects of uncertainty shocks, even though some of the relevant asset pricing data are available at much higher frequencies. In each of these applications, discarding data at high frequencies and estimating using the lowest sampling interval is standard procedure.

The estimator based solely on coarse data 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 evolves at a higher frequency than that used for estimation. This fact follows because the model $Y_t = A + BY_{t-1} + \epsilon_t$ is closed under temporal aggregation, so that $Y_{t+n} = A_n + B_n Y_t + \epsilon_{t_n}$ where the new coefficients A_n and B_n are given by

$$A_n = \sum_{s=1}^{n} B^{s-1} A \tag{4}$$

$$B_n = B^n \tag{5}$$

and the covariance matrix of the error ϵ_{t_n} is

$$\Sigma_n = \sum_{s=1}^{n} B^{s-1} \Sigma B^{s-1\prime}. \tag{6}$$

Estimation of the lower frequency VAR produces an estimate of B_n, denoted $\hat{B}_n$, which can then produce an estimate $\widehat{B}$ by computing $\hat{B}_n^{1/n}$. Then, using (4) and (6) produces estimates of A and Σ. If, for example, B_q is an estimate of B obtained using quarterly sampled data, inverting (5) gives the implied monthly estimate $B_q^{\frac{1}{3}}$. Equations (4)-(6) thus allow the comparison of estimates obtained through BMF with those produced by coarse estimation. Such comparisons are best done by transforming the posterior simulations and then computing the quantity of interest. For example, to compare posterior standard deviations from coarse estimation at a quarterly frequency to BMF at a monthly one, the sample standard deviation should be computed based upon the converted draws of the Gibbs sampler, so computing $\frac{1}{G}\sum_{i=1}^{G}\left(B^{(i)}\right)^{\frac{1}{3}}$ for G iterations. This conversion gives an estimate of the monthly implied standard deviation from quarterly coarse estimates for easy comparisons across the two methodologies.

3 Simulation Results

Having presented the methodology, this section examines BMF using simulated data. The purpose is to analyze how BMF fares relative to estimation at the coarsest frequency when the objective is to recover parameter estimates, say the posterior mean, that are as close as possible in some sense to the truth. This is very much a frequentist way of thinking, and so the exercise should accordingly be interpreted as a small sample study of the posterior mean as a frequentist parameter estimate.

Table 1 reports the root mean squared error of the parameters estimated using BMF versus estimation after discarding the high frequency data for four different parameter constellations, each with two sample sizes. The parameterizations use two different coefficient matrices and two different covariances among the errors. The data are generated by a monthly bivariate VAR, where the variable x is observed every period, and the variable z is observed every third period $t = 1, 4, 7, ..., 3T$, so T is the number of quarters in the sample, meaning there are $3T$ months. In the shorter sample, there are $T = 20$ quarters, and in the longer sample there are $T = 80$.

As can be seen, BMF tends to attain smaller root mean squared errors and lower absolute biases for the parameter estimates. Two additional features should be mentioned.

Comparing the results for Simulation 1 with Simulation 2, and Simulation 3 with Simulation 4, the matrices A and B are identical, and the the difference in the parameterization is the covariance matrix Σ. Simulations 1 and 3 have diagonal covariance matrices, and Simulations 2 and 4 have correlated shocks. In the case of correlated shocks, BMF uniformly performs better than coarse estimation. Incorporating the additional high frequency data allows better inference on the parameters because it helps distinguish between the sources of varation. In other words, BMF exhibits better performance in the case of correlated shocks, because information contained in the high frequency observations becomes important as the contemporaneous correlation between the unobserved and observed data increases.

Second, for each simulation parameterization, the improvement provided by BMF is comparable across the different dataset lengths. This suggests that incorporating mixed frequency data can boost accuracy for both small and longer datasets. Since the performance of coarse estimation improves with sample size, a relatively constant relative improvement of BMF suggests that mixed frequency data will produce a larger gain with small sample sizes.

Hence, this section shows that BMF has significant gains compared to the usual estimation strategy of using the coarsest frequency data in terms of root mean squared errors and absolute bias. Having provided a comparison on simulated data, the following two sections illustrate two experiments widely studied in the economic literature: first, estimating GDP at monthly frequency from monthly data, and second, exploring the information contained in financial variables for real variables.

4 Application I: Monthly and Quarterly Data

4.1 Data Description

This section provides an example implementation of the algorithm to data collected at monthly and quarterly frequencies. The primary objective is to formulate a model that allows analysis of GDP at a frequency higher than the quarterly data readily available. Since GDP is a widely cited indicator of economic performance, using higher frequency data may help to provide more timely snapshots of the economy than quarterly data allows.

This application is of considerable interest in the mixed frequency estimation literature. Kuzin *et al.* (2010) use quarterly GDP and a set of monthly indicators to compute monthly estimates of quarterly GDP in the Euro area following the MIDAS estimation strategy, Mittnik & Zadrozny (2005) pursue a similar objective using the Kalman Filter. Aruoba *et al.* (2009) use a dynamic factor model to estimate economic activity. Barhoumi *et al.* (2008) compare several estimation strategies that generate monthly GDP estimates using quarterly GDP and monthly indicators, Diron (2008) assesses the ability of real-time monthly data to help forecast quarterly GDP, and Marcellino & Musso (2010) study the estimation of real-time output gaps.

To keep the exercise simple and transparent, this application combines monthly data on industrial production, inflation, and the unemployment rate with quarterly real GDP data for the US. The data are the twelve-month change in industrial production and inflation, the four-quarter change in real GDP, and the unemployment rate, all expressed as percentage points. For the mixed frequency data, the timing assumption is that every month, the monthly data are observed, but the quarterly data are observed only during the last month of each quarter. Since the timing assumes that a period equals a month, the analysis below converts the quarterly results to a monthly frequency using the method described in Section 2.2. The data run from Jan-1948 to Jun-2011, for a total of 762 months or 254 quarters. Summary statistics for the variables are presented in Table 2.

4.2 Estimation Results

Tables 3, 4, and 5 display the estimates of the constant A, the coefficient matrix B, and the variance-covariance matrix Σ, respectively, for the BMF estimator using monthly observations versus the coarse estimation strategy of discarding the first two months of each quarter and therefore using only quarterly observations. In all cases, the ordering of the variables follows that in Table 2: industrial production growth, inflation, unemployment rate, and real GDP growth.

Table 3 shows the estimates of the constant terms in the VAR, converted to a monthly frequency. The two estimates produce somewhat different posterior means for the constant terms, but they are all within one standard deviation across the methods. Noticeably, however, the posterior standard deviations are uniformly smaller for the monthly BMF than for the quarterly estimates, so BMF produces more precise estimates.

Table 4 shows the estimates of the coefficient matrices B, again with the quarterly estimates converted to their monthly counterpart. Similar to the estimates for the constant in Table 3, the posterior means across methods are close with the two methods. Again, the standard deviations generated by BMF are uniformly smaller than those produced by quarterly estimation, which reflects the fact that including monthly data provides more information about the persistence of the process.

Figure 1 displays the kernel density estimates of the marginal posterior densities for the implied monthly VAR coefficients using the monthly BMF and versus quarterly estimation. As noted in Table 4, the point estimates do not vary greatly across methods, the bigger difference is in the posterior standard deviations. The posteriors using the monthly BMF have sharper peaks, reflecting the lower standard deviations. The most notable differences come in the fourth row and fourth column of plots, which are the estimates associated with the infrequently observed quarterly real GDP. So BMF improves the precision of the estimates for all variables, but has the biggest difference for the coarsely observed quarterly GDP.

Table 5 reports estimates of the residual covariance matrix Σ from monthly BMF and from quarterly estimation. In this case, there is a very substantial reduction in the posterior standard deviations for each of the coefficients. In general, the decrease in posterior standard deviations is the most noticeable for the monthly variables, showing that including the higher frequency data tends to produce more precise estimates of the covariances to the shocks.

4.3 Impulse Responses

The previous subsection discussed the improvement in estimates from incorporating the higher frequency data, but the question remains whether the lower standard deviation translates to changes in dynamics implied by the VAR. Figure 2 displays a comparison of impulse response functions between the monthly BMF estimate and the quarterly estimate converted to its monthly counterpart. The plots are very similar for the most part, suggesting that both estimates yield similar responses of the economy to the exogenous shocks. In most cases, the responses are almost identical, with both methodologies yielding humped-shaped responses for certain variables and monotonic responses from others. The cases that differ the most, which are the effects of GDP shocks, have generally the same shape, and the responses are within each others' confidence bands.

Despite the similarities in the shapes of the impulse responses between the two estimates, the accompanied confidence bands suggest that the responses produced using BMF are often more precise than those produced with just quarterly data. The most marked increase in precision occurs in identifying the effects of the various shocks on the monthly variables. 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 monthly variables, 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. The more significant reduction in the width of the confidence bands occurs for the effects of GDP on the monthly variables. With the coarsely sampled estimation method, since the data are observed quarterly, comovements of the variables, and hence the effects of shocks, are relatively difficult to distinguish. On the other hand, with BMF the shocks are more easily traced out since the comovements of variables are observed at a monthly frequency. Hence it is not surprising that the improvement of BMF is strongest among the variables for which BMF allows more observations.

The use of the BMF estimates, therefore, serves as a significant improvement when analyzing impulse response functions. The actual response functions differ only marginally, but the confidence bands for the BMF estimates suggest stronger estimates of the effects of exogenous shocks. The inclusion of the additional, more frequent observations that the BMF estimate makes it much easier to identify the responses of those more frequent observations.

5 Application II: Weekly and Monthly Data

5.1 Data Description

The application in the previous section showed how to use BMF to combine monthly and quarterly data, this application turns to using higher frequency financial data to inform inference about aspects of the real economy. Financial data are often available on an almost continual basis: asset prices and interest rates change even by the minute. This example examines how high-frequency data may help investigate the impact of financial variables or asset prices on output. As noted above, industrial production is available at a monthly frequency, and measures output in a set of subsectors in the economy. Since these production sectors may be especially influenced by changes in interest rates or oil prices, the high frequency data are measures of the level and slope of the yield curve, as well as spot oil prices.

Interest rate conditions may affect production decisions, and oil is often an essential input to production, so it is natural to consider these variables along with IP. A number of papers such as Kilian & Park (2009), Kilian *et al.* (2009) or Kilian & Vigfusson (2011), to name a few, estimate monthly VARs in order to study the affects of oil shocks, even though the spot oil price changes much more frequently. So the analysis uses the spot real price of West Texas Intermediate crude oil, and a measure of the intercept and slope of the yield curve of interest rates. The slope of the yield curve is defined as the difference in yields between the seven- and one-year zero coupons. The slope gives expectations about future interest rates and because it has been frequently noted that an inverted yield curve (negative slope) tend to precede recessions. All constant maturity zero-coupon yield data are from the dataset by Gurkaynak *et al.* (2006). While these variables are available at extremely high frequencies, the analysis below focuses on weekly data. The data run from the first week of Jan-1986 to the last week of Jul-2011, for a total of 1336 weekly observations and 307 monthly observations. Summary statistics for the variables are presented in Table 6.

In addition to being able to address the impact of interest rates and oil on industrial production, the choice of weekly intervals presents an interesting challenge for mixed-frequency data. The assumption of timing is the following: the last business day of each week (usually Friday but occasionally Thursday), the yield curve and oil spot prices are observed, and the last Friday (or Thursday) of each month, the twelve-month growth rate of industrial production is observed. The challenge is that most months will have four weekly observations per month, but there will be some months that have five weeks associated with them. While BMF can handle this irregularly observed data with ease, using a method such as the Kalman filter or MIDAS would require either ignoring the fifth week in these months or changing the structure of the estimated equations in these months. Since the base period considered is a week, the analysis below converts the monthly estimates to their weekly counterparts following the method described in Section 2.2.[2]

5.2 Estimation

Tables 7 and 8 display the estimates of the constant A and the coefficient matrix B, respectively, for the BMF estimator using weekly observations and versus the discarding all but the last week of the month and therefore using only monthly observations. In all cases, the ordering of the variables follows that in Table 6: yield curve intercept, yield curve slope, oil price, and industrial production.

Table 7 contains the estimates of the constant terms in the VAR, converted to a weekly frequency. Three of the posterior means are similar across methods, the third variable, oil, shows a marked change in posterior estimate and reduction in posterior standard deviation, but even this change is small relative to the standard deviations.

Table 8 shows the estimates of the B coefficients, again with the monthly estimates converted to their weekly counterpart. As with the previous application, most of the posterior means are similar across methodologies, and BMF tends to have smaller posterior standard deviations. The notable exception to the similar posterior means are the estimates associated with oil, the third variable in the VAR. In the previous example, BMF provided the biggest reduction in standard errors for the coarsely observed GDP, but here the biggest reduction is associated with the finely observed oil price. The reason for this result is apparent from comparing the descriptive statistics in Table 6, which show that the oil variable has by far the biggest standard deviation and the lowest autocorrelation by a significant margin. Consequently, when oil is included at a high frequency, the inclusion adds a lot more information about the dynamics of the VAR

[2] In practice, the conversion of irregularly spaced data can be cumbersome, since, in this case, there are not a fixed number of weeks per month, and therefore equations (4), (5), and (6) cannot be directly applied. So to convert the monthly estimates to weekly estimates, the first step is to convert the monthly to their "whole-sample" counterparts (307 months per sample) and then to their weekly counterparts (1 sample per 1336 weeks). Because of the difficulty of inverting (6) in this case, the reported results here focus on the estimates of A and B.

at a weekly interval.

Figure 3 displays the kernel density estimates of the marginal posterior densities for the implied weekly VAR coefficients using the weekly BMF versus when only monthly data are used. The results seen in Table 8 are striking in these plots: the posterior estimates tend to be fairly similar, but there is significant gain in the precision of estimates associated with the third variable, which is the spot price of oil. Moving from monthly data to weekly has a large gain in precision of the posterior estimates associated with oil, as reflected by the high peaks of the marginal posterior in the third column and third row of subplots.

5.3 Impulse Responses

After noting the gains in accuracy from the parameter estimates, especially for the oil variable, Figure 4 shows the effects of incorporating weekly data on the impulse response functions. As with the previous application, the confidence bands are smaller for the BMF estimator. And reflecting the smaller standard errors associated with the parameters corresponding to the effects of oil, the narrowing of confidence bands is most pronounced for the oil variables. The shapes of the impulse responses are nearly identical across the two methodologies, but slights shifts in impulse responses from the weekly data may lead to different conclusions. For example, the "Oil on Oil" response is not very persistent for either case, but the relative persistence is greater for BMF. In addition, the BMF confidence intervals for oil cover zero in the case of "Oil on Level" and "Oil on Slope" meaning that the significance of the shocks may be reversed with the inclusion of higher frequency data.

6 Concluding Remarks

This paper considers estimation of first order VAR's using data sampled at mixed frequencies. The methodology uses Gibbs sampling the unobserved data at the high frequency to generate estimates with generally smaller standard errors. The simulation experiments demonstrate that BMF produces more accurate estimates of model parameters than the basic approach of sub-sampling at the coarse data frequency, and the two example applications show that using higher frequency data may produce sizable gains.

Improved accuracy is not the only advantage of the BMF estimator. Another benefit is the ability to update forecasts of a coarsely observed variable in response to new arrival of data measured at high frequencies. Along the lines of the applications presented above, examples include updating forecasts of next quarter GDP in response to monthly measurements of data or using weekly or even daily financial data to forecast aspects of the real economy. The BMF framework allows for a natural approach to incorporate high frequency observations to the low frequency forecasts, which would avoid the use of ad-hoc forecast revisions.

The approach is implemented here using a first order vector auto-regression for simplicity, but there are many possible extensions that generalize the this approach.

One potential 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. Following Jacquier *et al.* (1994), incorporating stochastic volatility into a VAR setting simply requires adding an additional Gibbs sampling draw for the unobserved volatility path. In a similar vein, the Bayesian approach can easily incorporate heavy tailed error distributions in the form of mixture of normals or t distributions, Markov mixtures as in Albert & Chib (1993) or jumps as in Eraker *et al.* (2003).

It is useful to consider how BMF compares to a Kalman filtering approach. The linearity of the VAR coupled with Gaussian errors implies that it is possible to write the unobserved high frequency data in a state space form, enabling the use of a Kalman filter approach to a basic model. One advantage of the Kalman filtering approach is that it does not require posterior simulation of unobserved data. On the other hand, the resulting likelihood function needs to be analyzed numerically. Frequentist analysis through maximum likelihood is possible, although this requires a numerical search in a typically high dimensional parameter space. For Bayesian inference, BMF avoids using Metropolis Hastings or other methods that would be required in to deal with the non-Gaussian nature of the posterior computed through Kalman filtering.

The application to weekly and monthly data in Section 5 highlights a second advantage of BMF over the Kalman filter and MIDAS. In the case of irregularly spaced data, where there are an unequal number of observations per unit of coarsely observed data, the specification of the Kalman filter and MIDAS equations can be awkward and tedious, whereas BMF requires no such changes. For the Kalman filter approach, the state equation is the VAR equation

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} A_x \\ A_z \end{bmatrix} + \begin{bmatrix} B_{xx} & B_{xz} \\ B_{zx} & B_{zz} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix} \tag{7}$$

and the observation equation changes from

$$Y_t^{obs} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x_t \\ z_t \end{bmatrix} \tag{8}$$

if z_t is observed, to

$$Y_t^{obs} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ z_t \end{bmatrix} \tag{9}$$

if z_t is not observed. In the case of mixed frequency data where the missing data occur at regular frequencies, such as the monthly and quarterly application, the observation equation switches between (9) and (8) systematically. With irregularly spaced data, such as the weekly and monthly application, the switching observation equation changes at different intervals, making implementation more difficult. With even more irregular observations, such as if z_t has variables that are observed not at the same frequency, the complexity in changing the observation equation can grow substantially. The BMF approach does not require this constant changing of observation equations, since it simply draws all the missing data conditional on the parameters regardless of the observation interval.

A potential disadvantage of the VAR approach is that, especially as the number of lags grows, the number of parameters to estimate grows considerably. The MIDAS approach, by comparison, typically focuses on one equation to forecast one coarsely observed variable with more frequently observed variables. Considering this setup for the application of weekly and monthly data from Section 5, the MIDAS approach forecasts monthly industrial production using the weekly interest rate and oil data without putting restrictions on how industrial production affects the weekly data. However, BMF can incorporate this flexibility by including certain variables as exogenous rather than in the endogenous vector in the VAR. Then, for example, interest rates and IP could be affected by the exogenous evolution of oil, thereby ignoring other feedback effects on oil.

Of course, whether to incorporate variables as endogenous or exogenous depends upon the application and research question, but the important point is that BMF handles these restrictions with minimal modification to the VAR model considered here.

Another useful generalization of the VAR approach is to consider the use of BMF in connection with linear state-space models. Consider the case with an observation equation $Y_{t+1} = C + DX_{t+1}$, and state-equation $X_{t+1} = A + BX_t + \epsilon_{t+1}$, where there is a mixed sampling frequency for Y_t. The VAR(1) framework discussed in this paper nests the state-space model by writing $Y^* = (Y, X)$, and that $Y^*_{t+1} = A^* + B^*Y^*_t + \epsilon^*_{t+1}$ and $A^* = (C, A)$,

$$B^* = \begin{pmatrix} 0 & D \\ 0 & B \end{pmatrix}. \tag{10}$$

BMF then proceeds to estimate this model by simulating, as before, the sparsely observed elements of Y, but in addition treats X as an unobserved variable – a variable observed with zero frequency. Importantly, the 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, the algorithm needs only slight modifications to impose the zero-constraints on B^*. Note that the estimation of VARMA models can be implemented using this approach.

This paper has also not considered the out-of-sample forecasting ability of BMF estimators. Of course, given the applications, forecasting is a natural extension given mixed frequency data. As with any VAR-based method, forecasting given BMF estimates involves iterated forecasting rather than direct forecasting. Marcellino *et al.* (2006) and Chevillon & Hendry (2005) discuss the advantages of both types of forecasting, and De Mol *et al.* (2008) show Bayesian VARs can have good forecasting performance.

Finally, while the BMF algorithm applies in general, identification considerations must, as usual, be investigated on a case by case basis depending upon the application. Consequently, the BMF framework developed in this paper represents an interesting starting point for a number of different extensions.

References

Albert, J.H., & Chib, S. 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, E., Ghysels, E., & Kourtellos, A. 2009. Should Macroeconomic Forecasters Use Daily Financial Data and How? *UNC Working Paper*.

Andreou, E., Ghysels, E., & Kourtellos, A. 2010. Regression Models with Mixed Sampling Frequencies. *Journal of Econometrics*, **158**, 246–261.

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., & Runstler, G. 2008. Short-Term Forecasts of Euro Area GDP Growth. *ECB Working Paper*, **949**.

Aruoba, S.B., Diebold, F.X., & Scotti, C. 2009. Real-Time Measurement of Business Conditions. *Journal of Business and Economic Statistics*, **27**(4), 417–427.

Baffigi, A., Golinelli, R., & Parigi, G. 2004. Bridge Models to Forecast the Euro Area GDP. *International Journal of Forecasting*, 447–460.

Bai, J., Ghysels, E., & Wright, J.H. 2010. State Space Models and MIDAS Regressions. *UNC Working Paper*.

Banbura, M., Giannone, D., & Reichlin, L. 2010. Large Bayesian VARs. *Journal of Applied Econometrics*, 71–92.

Barhoumi, K., Benk, S., Cristadoro, R., Den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Runstler, G., Ruth, K., & Nieuwenhuyze, C.V. 2008. Short-term Forecasting of GDP Using Large Monthly Datasets - A Pseudo Real-time Forecast Evaluation Exercise. *National Bank of Belgium Working Paper*.

Bloom, N. 2009. The Impact of Uncertainty Shocks. *Econometrica*, **77**(3), 623–685.

Chevillon, G., & Hendry, D. F. 2005. Non-Parametric Direct Multi-Step Estimation for Forecasting Economic Processes. *International Journal of Forecasting*, **21**, 201–218.

Chow, G.C., & Lin, A. 1971. Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics*, **53**(4), 372–375.

Christiano, L., Eichenbaum, M., & Evans, C. 1996. The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds. *The Review of Economics and Statistics*, 16–34.

Christiano, L., Eichenbaum, M., & Evans, C. 1999. Monetary Policy Shocks: What Have We Learned, and To What End. *Handbook of Monetary Economics*.

Corrado, C., & Greene, M. 1988. Reducing Uncertainty in Short-Term Projections: Linkage of Monthly and Quarterly Models. *Journal of Forecasting*, **7**, 77–102.

De Mol, C., Giannone, D., & Reichlin, L. 2008. Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components? *Journal of Econometrics*, **146**, 318–328.

Diron, M. 2008. Short-Term Forecasts of Euro Area Real GDP Growth: An Assessment of Real-Time Performance Based on Vintage Data. *Journal of Forecasting*, **27**(5), 371–390.

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

Francis, N.R., Ghysels, E., & Owyang, M.T. 2011. The Low-Frequency Impact of Daily Monetary Policy Shocks. *Federal Reserve Bank of St. Louis Working Paper*, **2011-009B**.

Ghysels, E., Santa-Clara, P., & Valankov, R. 2004. There is a Risk-Return Tradeoff After All. *Journal of Financial Economics*.

Ghysels, E., Sinko, A., & Valankov, R. 2007. MIDAS Regressions: Further Results and New Directions. *UNC Working Paper*.

Giannone, D., Reichlin, L., & Small, D. 2008. Nowcasting: The Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, **55**(4), 665–676.

Gurkaynak, R.S., Sack, B., & Wright, J. 2006. The U.S. Treasury Yield Curve: 1961 to the Present. *Federal Reserve Board Working Paper*.

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

Jacquier, E., Polson, N. G., & Rossi, P. 1994. Bayesian Analysis of Stochastic Volatility Models. *Journal of Business and Economic Statistics*, **12**, 371–389.

Kilian, L., & Park, C. 2009. The Impact of Oil Price Shocks on the US Stock Market. *International Economic Review*, **50**(4), 1267–1287.

Kilian, L., & Vigfusson, R.J. 2011. Nonlinearities in the Oil Price-Output Relationship. *Macroeconomic Dynamics*.

Kilian, L., Rebucci, A., & Spatafora, N. 2009. Oil Shocks and External Balances. *Journal of International Economics*, **77**(2), 181–194.

Kim, S., Shephard, N., & Chib, S. 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, **65**(3), 361–393.

Kuzin, V., Marcellino, M., & Schumacher, C. 2010. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area. *International Journal of Forecasting*.

Leeper, E., & Zha, T. 2003. Modest Policy Interventions. *Journal of Monetary Economics*, **50**(8), 1673–1700.

Marcellino, M., & Musso, A. 2010. Real Time Estimates of the Euro Area Output Gap: Reliability and Forecasting Performance. *ECB Working Paper*, **1157**.

Marcellino, M., Stock, J. H., & Watson, M. W. 2006. A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. *Journal of Econometrics*, **135**, 499–526.

Mariano, R., & Murasawa, Y. 2010. A Coincident Index, Common Factors, and Monthly Real GDP. *Oxford Bulletin of Economics and Statistics*, **72**, 27–46.

Miller, P., & Chin, D. 1996. Using Monthly Data to Improve Quarterly Model Forecasts. *Federal Reserve Bank of Minneapolis Quarterly Review*, 16–33.

Mittnik, S., & Zadrozny, P. 2005. Forecasting Quarterly German GDP at Monthly Intervals Using Monthly Ifo Business Conditions Data. *Ifo Survey Data in Business Cycle and Monetary Policy Analysis*, 19–48.

Sims, C.A., & Zha, T. 2006a. Does Monetary Policy Generate Recessions? *Macroeconomic Dynamics*, **10**(2), 231.

Sims, C.A., & Zha, T. 2006b. Were There Regime Switches in US Monetary Policy? *American Economic Review*, **96**(1), 54–81.

Zadrozny, P. 1988. Gaussian Likelihood of Continuous-Time Armax Models When Data are Stocks and Flows at Different Frequencies. *Econometric Theory*, 108–124.

7 Appendix A: Derivations

This appendix shows the derivations for the analytical conditional distribution of the latent data. The structure is a VAR(k) model:

$$y_t = A + B_1 y_{t-1} + B_2 y_{t-2} + \cdots + B_k y_{t-k} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma) \tag{11}$$

The partition between the fully observed variable x_t and the variable with missing data z_t is given by

$$\begin{aligned} \begin{bmatrix} x_t \\ z_t \end{bmatrix} &= \begin{bmatrix} A_x \\ A_z \end{bmatrix} + \begin{bmatrix} B_{1,xx} & B_{1,xz} \\ B_{1,zx} & B_{1,zz} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \cdots \\ &\quad + \begin{bmatrix} B_{k,xx} & B_{k,xz} \\ B_{k,zx} & B_{k,zz} \end{bmatrix} \begin{bmatrix} x_{t-k} \\ z_{t-k} \end{bmatrix} + \begin{bmatrix} \varepsilon_{x,t} \\ \varepsilon_{z,t} \end{bmatrix}. \end{aligned} \tag{12}$$

where

$$\begin{bmatrix} \varepsilon_{x,t} \\ \varepsilon_{z,t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix} \right).$$

By Bayes rule, the conditional probability, for $l = 0, ..., k$, is given by

$$p\left(x_{t+l}, z_{t+l} | x_{t-k:t+l-1}, z_{t-k:t+l-1}, \Theta\right) = \frac{p\left(x_{t-k:t+l}, z_{t-k:t+l}, \Theta\right)}{p\left(x_{t-k:t+l-1}, z_{t-k:t+l-1}, \Theta\right)} \tag{13}$$

where $x_{t-k:t}$ denote the sequence of observations of x from $t-k$ to t. The conditional probability for z_t is related to the previous expression by

$$p\left(z_t | x_{t-k:t+k}, z_{t-k:t-1}, z_{t+1:t+k}, \Theta\right) \propto \prod_{l=0}^{k} p\left(x_{t+l}, z_{t+l} | x_{t-k:t+l-1}, z_{t-k:t+l-1}, \Theta\right) \tag{14}$$

Since each joint distribution is conditionally normal, each of these has the form

$$\begin{aligned} & p\left(x_{t+l}, z_{t+l} | x_{t-k:t+l-1}, z_{t-k:t+l-1}, \Theta\right) \\ \propto \quad & \exp\left\{-\frac{1}{2}\begin{bmatrix} \varepsilon_{x,t+l} \\ \varepsilon_{z,t+l} \end{bmatrix}' \begin{bmatrix} \Sigma^{xx} & \Sigma^{xz} \\ \Sigma^{zx} & \Sigma^{zz} \end{bmatrix} \begin{bmatrix} \varepsilon_{x,t+l} \\ \varepsilon_{z,t+l} \end{bmatrix}\right\} \\ \propto \quad & \exp\left\{-\frac{1}{2}\left(\varepsilon'_{x,t+l}\Sigma^{xx}\varepsilon_{x,t+l} + \varepsilon'_{x,t+l}\Sigma^{xz}\varepsilon_{z,t+l} + \varepsilon'_{z,t+l}\Sigma^{zx}\varepsilon_{x,t+l} + \varepsilon'_{z,t+l}\Sigma^{zz}\varepsilon_{z,t+l}\right)\right\} \end{aligned} \tag{15}$$

where

$$\begin{bmatrix} \Sigma^{xx} & \Sigma^{xz} \\ \Sigma^{zx} & \Sigma^{zz} \end{bmatrix} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}^{-1}$$

Define, for $l = 0, 1, ..., k$, the portion not explained by z_t:

$$u_{x,t+l} = \varepsilon_{x,t+l} + B_{l,xz} z_t, \text{ and } u_{z,t+l} = \varepsilon_{z,t+l} + B_{l,zz} z_t \tag{16}$$

or

$$\begin{aligned} \begin{bmatrix} u_{x,t+l} \\ u_{z,t+l} \end{bmatrix} \quad = \quad & \begin{bmatrix} x_{t+l} \\ z_{t+l} \end{bmatrix} - \begin{bmatrix} A_x \\ A_z \end{bmatrix} - \begin{bmatrix} B_{1,xx} & B_{1,xz} \\ B_{1,zx} & B_{1,zz} \end{bmatrix} \begin{bmatrix} x_{t+l-1} \\ z_{t+l-1} \end{bmatrix} - \cdots \\ & - \begin{bmatrix} B_{k,xx} & B_{k,xz} \\ B_{k,zx} & B_{k,zz} \end{bmatrix} \begin{bmatrix} x_{t+l-k} \\ z_{t+l-k} \end{bmatrix} + \begin{bmatrix} B_{l,xz} \\ B_{l,zz} \end{bmatrix} z_t \end{aligned} \tag{17}$$

where $B_{0,xz} = 0$ and $B_{0,zz} = -I$. Using these definitions to substitute out $\varepsilon_{x,t+l}$ and $\varepsilon_{z,t+l}$:

$$\begin{aligned}
& p\left(x_{t+l}, z_{t+l} | x_{t-k:t+l-1}, z_{t-k:t+l-1}, \Theta\right) \\
\propto \quad & \exp\left\{-\frac{1}{2}\begin{pmatrix} \left(u_{x,t+l} - B_{l,xz} z_t\right)' \Sigma^{xx} \left(u_{x,t+l} - B_{l,xz} z_t\right) \\ + \left(u_{x,t+l} - B_{l,xz} z_t\right)' \Sigma^{xz} \left(u_{z,t+l} - B_{l,zz} z_t\right) \\ + \left(u_{z,t+l} - B_{l,zz} z_t\right)' \Sigma^{zx} \left(u_{x,t+l} - B_{l,xz} z_t\right) \\ + \left(u_{z,t+l} - B_{l,zz} z_t\right)' \Sigma^{zz} \left(u_{z,t+l} - B_{l,zz} z_t\right) \end{pmatrix}\right\} \\
\propto \quad & \exp\left\{-\frac{1}{2}\begin{pmatrix} +z_t' \left[B_{l,xz}' \Sigma^{xx} B_{l,xz} + B_{l,xz}' \Sigma^{xz} B_{l,zz} + B_{l,zz}' \Sigma^{zx} B_{l,xz} + B_{l,zz}' \Sigma^{zz} B_{l,zz}\right] z_t \\ -z_t' \left[B_{l,xz}' \Sigma^{xx} u_{x,t+l} + B_{l,xz}' \Sigma^{xz} u_{z,t+l} + B_{l,zz}' \Sigma^{zx} u_{x,t+l} + B_{l,zz}' \Sigma^{zz} u_{z,t+l}\right] \\ - \left[u_{x,t+l}' \Sigma^{xx} B_{l,xz} + u_{x,t+l}' \Sigma^{xz} B_{l,zz} + u_{z,t+l}' \Sigma^{zx} B_{l,xz} + u_{z,t+l}' \Sigma^{zz} B_{l,zz}\right] z_t \\ + \left[u_{x,t+l}' \Sigma^{xx} u_{x,t+l} + u_{x,t+l}' \Sigma^{xz} u_{z,t+l} + u_{z,t+l}' \Sigma^{zx} u_{x,t+l} + u_{z,t+l}' \Sigma^{zz} u_{z,t+l}\right] \end{pmatrix}\right\}
\end{aligned} \tag{18}$$

These imply that the conditional distribution is normal

$$\begin{aligned}
& p\left(x_{t+l}, z_{t+l} | x_{t-k:t+l-1}, z_{t-k:t+l-1}, \Theta\right) \\
\propto \quad & \exp\left\{-\frac{1}{2}\left(z_t - M_l\right)' W_l^{-1} \left(z_t - M_l\right)\right\} \\
\propto \quad & \exp\left\{-\frac{1}{2}\left(z_t' W_l^{-1} z_t - z_t' W_l^{-1} M_l - M_l' W_l^{-1} z_t + M_l' W_l^{-1} M_l\right)\right\}
\end{aligned} \tag{19}$$

Matching coefficients produces

$$W_l^{-1} = B_{l,xz}' \Sigma^{xx} B_{l,xz} + B_{l,xz}' \Sigma^{xz} B_{l,zz} + B_{l,zz}' \Sigma^{zx} B_{l,xz} + B_{l,zz}' \Sigma^{zz} B_{l,zz} \tag{20}$$

$$M_l = W_l \left[B_{l,xz}' \Sigma^{xx} u_{x,t+l} + B_{l,xz}' \Sigma^{xz} u_{z,t+l} + B_{l,zz}' \Sigma^{zx} u_{x,t+l} + B_{l,zz}' \Sigma^{zz} u_{z,t+l}\right] \tag{21}$$

The original distribution was

$$
\begin{aligned}
& p\left(z_t | x_{t-k:t+k}, z_{t+1:t+k}, z_{t-k:t-1}, \Theta\right) \\
\propto \quad & \prod_{l=0}^{k} p\left(x_{t+l}, z_{t+l} | x_{t-k:t+l-1}, z_{t-k:t+l-1}, \Theta\right) \\
\propto \quad & \prod_{l=0}^{k} \exp\left\{-\frac{1}{2}\left(z_t' W_l^{-1} z_t - z_t' W_l^{-1} M_l - M_l' W_l^{-1} z_t + M_l' W_l^{-1} M_l\right)\right\} \\
\propto \quad & \exp\left\{-\frac{1}{2}\sum_{l=0}^{k}\left(z_t' W_l^{-1} z_t - z_t' W_l^{-1} M_l - M_l' W_l^{-1} z_t + M_l' W_l^{-1} M_l\right)\right\}
\end{aligned}
\tag{22}
$$

Assuming conditional normality of z_t implies

$$
\begin{aligned}
& p\left(z_t | x_{t-k:t+k}, z_{t+1:t+k}, z_{t-k:t-1}, \Theta\right) \\
\propto \quad & \exp\left\{-\frac{1}{2}\left(z_t - M\right)' W^{-1}\left(z_t - M\right)\right\} \\
\propto \quad & \exp\left\{-\frac{1}{2}\left(z_t' W^{-1} z_t - z_t' W^{-1} M - M' W^{-1} z_t + M' W^{-1} M\right)\right\}
\end{aligned}
\tag{23}
$$

Again matching coefficients

$$W^{-1} = \sum_{l=0}^{k} W_l^{-1} \tag{24}$$

$$M = W \sum_{l=0}^{k} W_l^{-1} M_l \tag{25}$$

Therefore,

$$p\left(z_t | x_{t-k:t+k}, z_{t+1:t+k}, z_{t-k:t-1}, \Theta\right) \sim N\left(M, W\right). \tag{26}$$

8 Appendix B: Tables and Figures

Table 1: Accuracy of BMF Using Simulated Data

Absolute bias and root mean squared errors of parameters estimated using BMF and coarse data relative to true values over different sample length and pseudo true parameter values. The simulated model is $y_t = A + By_{t-1} + \varepsilon_t$, where $y_t = [x_t \ \ z_t]'$, $A = [A_x \ \ A_z]'$, $B = [B_{xx} \ \ B_{xz};\ B_{zx} \ \ B_{zz}]$, $\varepsilon_t = [\varepsilon_t^x \ \ \varepsilon_t^z]'$, $\varepsilon_t \sim N(\mathbf{0}, \Sigma)$. T is sample size in quarters. The row labeled Abs Bias gives the percentage difference in the mean absolute bias for BMF relative to coarse sampling; negative numbers indicate BMF achieves a lower value. Similarly, the row labeled RMSE gives the percentage difference in root mean squared error for BMF relative to coarse sampling; negative numbers indicate BMF achieves a lower value. All values are the result of 1000 simulated datasets.

Simulation 1	$A = [0; 0]; B = [0.9 \ \ 0.9;\ 0 \ \ 0.9]; \Sigma = [0.01 \ \ 0;\ 0 \ \ 0.05]$					
			$T = 20$			
	$\widehat{A}_x$	$\widehat{A}_z$	$\widehat{B}_{xx}$	$\widehat{B}_{xz}$	$\widehat{B}_{zx}$	$\widehat{B}_{zz}$
Abs Bias	-77.643	-21.94	-34.55	-8.9621	-33.066	11.633
RMSE	-76.145	-22.059	-35.189	-8.066	-22.859	7.6958
			$T = 80$			
Abs Bias	-74.388	-18.858	-35.298	-7.6159	-47.802	-11.826
RMSE	-75.764	-21.155	-35.144	-7.6314	-47.423	-13.02
Simulation 2	$A = [0; 0]; B = [0.9 \ \ 0.9;\ 0 \ \ 0.9]; \Sigma = [0.01 \ \ 0.02;\ 0.02 \ \ 0.05]$					
			$T = 20$			
	$\widehat{A}_x$	$\widehat{A}_z$	$\widehat{B}_{xx}$	$\widehat{B}_{xz}$	$\widehat{B}_{zx}$	$\widehat{B}_{zz}$
Abs Bias	-61.839	-28.245	-34.249	-7.907	-31.841	-19.146
RMSE	-61.54	-33.555	-33.013	-7.3243	-24.727	-23.358
			$T = 80$			
Abs Bias	-64.286	-18.643	-34.002	-9.9857	-43.643	-20.311
RMSE	-64.337	-19.64	-33.806	-9.0189	-44.587	-22.497
Simulation 3	$A = [0; 0]; B = [0.95 \ \ 0.8;\ -0.01 \ \ 0.99]; \Sigma = [0.01 \ \ 0;\ 0 \ \ 0.05]$					
			$T = 20$			
	$\widehat{A}_x$	$\widehat{A}_z$	$\widehat{B}_{xx}$	$\widehat{B}_{xz}$	$\widehat{B}_{zx}$	$\widehat{B}_{zz}$
Abs Bias	-74.399	-11.03	-18.896	0.043434	-29.003	12.071
RMSE	-69.858	-16.463	-8.5897	4.1556	-19.055	8.2203
			$T = 80$			
Abs Bias	-69.573	-4.0152	-18.69	-0.38386	-28.168	0.48215
RMSE	-70.276	-4.8683	-19.855	-1.0017	-29.027	-0.34357
Simulation 4	$A = [0; 0]; B = [0.95 \ \ 0.8;\ -0.01 \ \ 0.99]; \Sigma = [0.01 \ \ 0.02;\ 0.02 \ \ 0.05]$					
			$T = 20$			
	$\widehat{A}_x$	$\widehat{A}_z$	$\widehat{B}_{xx}$	$\widehat{B}_{xz}$	$\widehat{B}_{zx}$	$\widehat{B}_{zz}$
Abs Bias	-57.378	-15.214	-20.013	-1.8474	-12.5	-3.1774
RMSE	-56.188	-21.155	-17.871	-3.4858	-6.8278	-9.611
			$T = 80$			
Abs Bias	-56.736	-4.2886	-23.924	-1.504	-27.321	-5.8757
RMSE	-57.394	-4.888	-23.717	-1.304	-27.335	-7.5488

Table 2: Descriptive Statistics, Monthly and Quarterly Application

Basic statistics for the change in industrial production, the inflation rate, the unemployment rate, and the change in real GDP.

	Mean	Std.Dev.	Autocorrelation
Δln(IP)	3.207	5.961	0.965
Inflation	3.719	3.005	0.988
Unempl	5.744	1.641	0.991
Δln(GDP)	3.249	2.755	0.845

Table 3: Parameter Estimates A, Monthly and Quarterly Application

Posterior means and standard deviations of the VAR constant terms, A. The variables are, in order of appearance, the change in industrial production, the inflation rate, the unemployment rate, and the change in GDP. All parameters are at the monthly frequency.

		BMF:Monthly	Quarterly
A_1	(Δln(IP))	-0.2291	-0.4058
		(0.2807)	(0.3329)
A_2	(Inflation)	-0.04862	0.02807
		(0.08379)	(0.1001)
A_3	(Unempl)	0.184	0.1861
		(0.0383)	(0.03681)
A_4	(Δln(GDP))	0.2096	0.01081
		(0.1659)	(0.148)

Table 4: Parameter Estimates B, Monthly and Quarterly Application

Posterior means and standard deviations of the VAR coefficients, B. The variables are, in order of appearance, the change in industrial production, the inflation rate, the unemployment rate, and the change in GDP. All parameters are at the monthly frequency.

		BMF:Monthly	Quarterly
B_{11}	(IP on IP)	1.041	0.9757
		(0.02434)	(0.0313)
B_{12}	(Inflation on IP)	-0.09648	-0.1123
		(0.01882)	(0.02415)
B_{13}	(Unempl on IP)	0.1654	0.181
		(0.03799)	(0.04578)
B_{14}	(GDP on IP)	-0.1526	-0.0423
		(0.05446)	(0.06724)
B_{21}	(IP on Inflation)	0.02811	0.04139
		(0.007012)	(0.009049)
B_{22}	(Inflation on Inflation)	0.989	0.9877
		(0.005656)	(0.006982)
B_{23}	(Unempl on Inflation)	0.009435	0.005014
		(0.01139)	(0.01369)
B_{24}	(GDP on Inflation)	-0.01963	-0.04567
		(0.01564)	(0.01958)
B_{31}	(IP on Unempl)	-0.0128	-0.007334
		(0.003248)	(0.003351)
B_{32}	(Inflation on Unempl)	0.008099	0.008291
		(0.002533)	(0.002611)
B_{33}	(Unempl on Unempl)	0.9721	0.974
		(0.005134)	(0.005036)
B_{34}	(GDP on Unempl)	-0.001592	-0.01146
		(0.007313)	(0.007225)
B_{41}	(IP on GDP)	0.1154	0.07235
		(0.01632)	(0.01371)
B_{42}	(Inflation on GDP)	-0.02403	-0.03121
		(0.01045)	(0.01053)
B_{43}	(Unempl on GDP)	0.06072	0.07473
		(0.02141)	(0.02027)
B_{44}	(GDP on GDP)	0.7425	0.8287
		(0.03703)	(0.02958)

Table 5: Parameter Estimates, Σ, Monthly and Quarterly Application

Posterior means and standard deviations of the covariance matrix Σ. The variables are, in order of appearance, the change in industrial production, the inflation rate, the unemployment rate, and the change in GDP. All parameters are at the monthly frequency.

	BMF: Monthly		
2.409	0.05786	-0.1332	0.6603
(0.2484)	(0.05267)	(0.02562)	(0.1224)
0.05786	0.216	-0.007187	-0.01001
(0.05267)	(0.02235)	(0.007209)	(0.03266)
-0.1332	-0.007187	0.04339	-0.06021
(0.02562)	(0.007209)	(0.004534)	(0.01905)
0.6603	-0.01001	-0.06021	0.6975
(0.1224)	(0.03266)	(0.01905)	(0.1109)
	Quarterly		
3.672	0.04347	-0.261	1.061
(0.3555)	(0.07152)	(0.03213)	(0.1288)
0.04347	0.3049	-0.007526	0.00222
(0.07152)	(0.02886)	(0.007617)	(0.03076)
-0.261	-0.007526	0.04261	-0.09586
(0.03213)	(0.007617)	(0.004016)	(0.01322)
1.061	0.00222	-0.09586	0.6863
(0.1288)	(0.03076)	(0.01322)	(0.06618)

Table 6: Descriptive Statistics, Weekly and Monthly Application

Basic statistics for the one-year interest rate, slope of the yield curve, monthly growth rate in the real price of oil, and industrial production.

	Mean	Std.Dev.	Autocorrelation
Level	4.467	2.365	0.999
Slope	5.699	1.862	0.997
$\Delta\ln(\text{Oil})$	0.232	10.039	0.706
$\Delta\ln(\text{IP})$	2.165	4.145	0.997

Table 7: Parameter Estimates A, Weekly and Monthly Application

Posterior means and standard deviations of the VAR constant terms, A. The variables are, in order of appearance, the yield curve level, yield curve slope, change in real price of oil, and industrial production growth. All parameters are at the weekly frequency.

		BMF:Weekly	Monthly
A_1	Level	-0.01066	-0.01016
		(0.01306)	(0.01408)
A_2	Slope	0.01728	0.01479
		(0.01567)	(0.01573)
A_3	$\Delta\ln(\text{Oil})$	0.7337	0.2396
		(0.7558)	(1.043)
A_4	$\Delta\ln(\text{IP})$	-0.1356	-0.1103
		(0.05597)	(0.04648)

Table 8: Parameter Estimates B, Weekly and Monthly Application

Posterior means and standard deviations of the VAR coefficients, B. The variables are, in order of appearance, the yield curve level, slope of the yield curve, change in real price of oil, and industrial production growth. All parameters are at the weekly frequency.

		BMF:Weekly	Monthly
B_{11}	(Level on Level)	0.9947	0.9949
		(0.003616)	(0.003858)
B_{12}	(Slope on Level)	0.004002	0.003834
		(0.004459)	(0.004786)
B_{13}	(Oil on Level)	0.0004077	0.00182
		(0.0003276)	(0.000925)
B_{14}	(IP on Level)	0.002752	0.002669
		(0.000874)	(0.0009284)
B_{21}	(Level on Slope)	0.004129	0.004154
		(0.004342)	(0.004319)
B_{22}	(Slope on Slope)	0.993	0.9936
		(0.005353)	(0.005354)
B_{23}	(Oil on Slope)	0.0005737	0.002077
		(0.0003957)	(0.001039)
B_{24}	(IP on Slope)	-0.0005504	-0.0006479
		(0.001048)	(0.00104)
B_{31}	(Level on Oil)	0.1136	-0.03534
		(0.2123)	(0.3238)
B_{32}	(Slope on Oil)	-0.2126	-0.02515
		(0.2593)	(0.372)
B_{33}	(Oil on Oil)	0.7067	0.4553
		(0.01958)	(0.1214)
B_{34}	(IP on Oil)	0.01025	0.04596
		(0.05193)	(0.08615)
B_{41}	(Level on IP)	-0.04592	-0.04722
		(0.01541)	(0.01277)
B_{42}	(Slope on IP)	0.06202	0.05808
		(0.01905)	(0.01585)
B_{43}	(Oil on IP)	0.003134	0.007225
		(0.001689)	(0.003171)
B_{44}	(IP on IP)	0.9939	0.9965
		(0.003735)	(0.003108)

Figure 1: Posterior Densities for Monthly and Quarterly Application

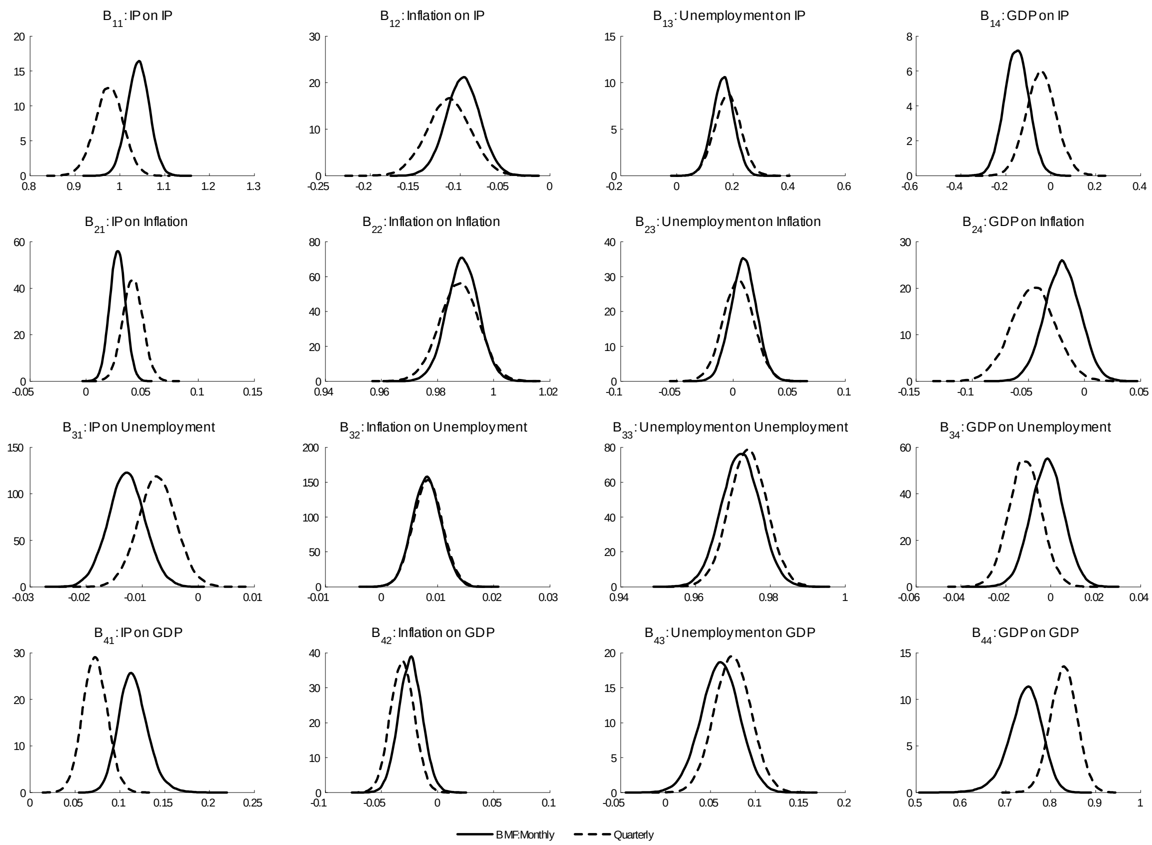

Figure 2: Impulse Responses for Monthly and Quarterly Application

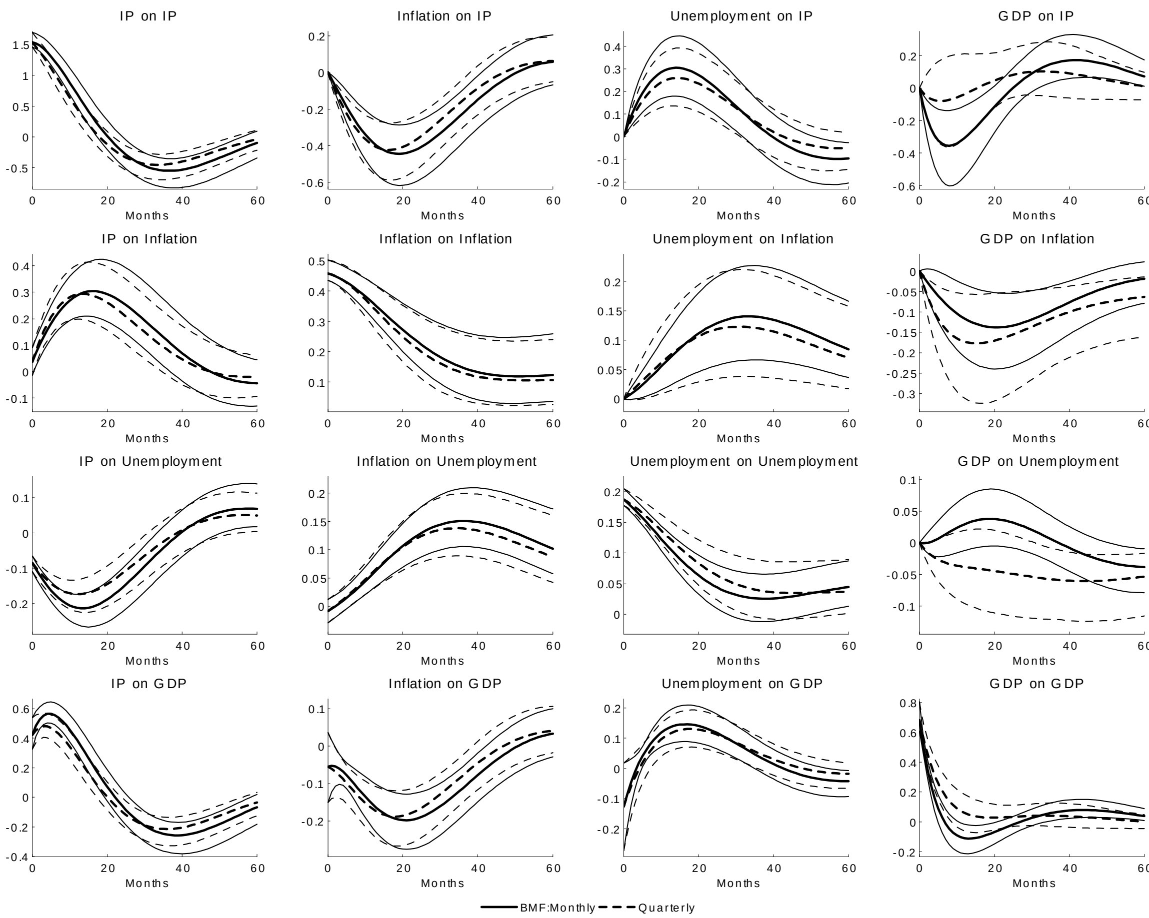

Figure 3: Posterior Densities for Weekly and Monthly Application

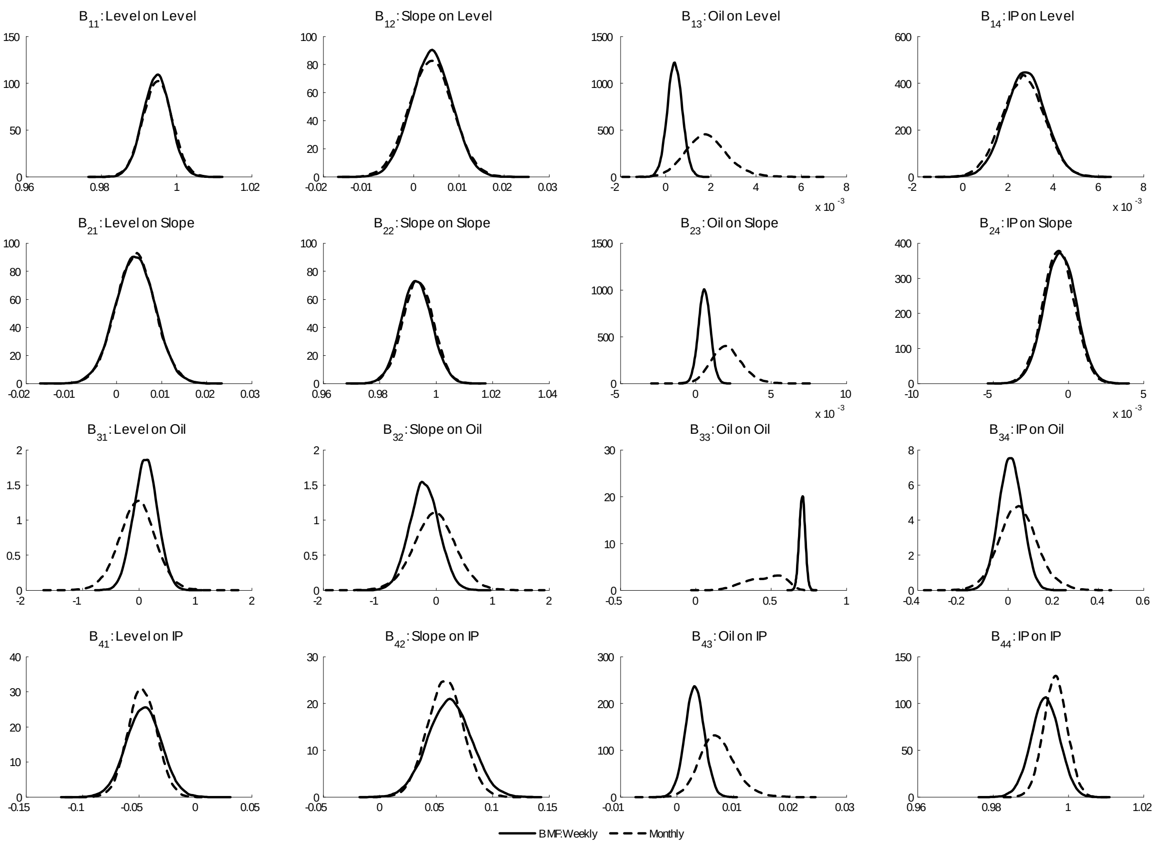

Figure 4: Impulse Responses for Weekly and Monthly Application