Efficient Simulation and Integrated Likelihood Estimation in State Space Models

Joshua C.C. Chan*
University of Queensland

Ivan Jeliazkov†
University of California, Irvine

November 1, 2009

Abstract

We consider the problem of implementing simple and efficient Markov chain Monte Carlo (MCMC) estimation algorithms for state space models. A conceptually transparent derivation of the posterior distribution of the states is discussed, which also leads to an efficient simulation algorithm that is modular, scalable, and widely applicable. We also discuss a simple approach for evaluating the integrated likelihood, defined as the density of the data given the parameters but marginal of the state vector. We show that this high-dimensional integral can be easily evaluated with minimal computational and conceptual difficulty. Two empirical applications in macroeconomics demonstrate that the methods are versatile and computationally undemanding. In one application, involving a time-varying parameter model, we show that the methods allow for efficient handling of large state vectors. In our second application, involving a dynamic factor model, we introduce a new blocking strategy which results in improved MCMC mixing at little cost. The results demonstrate that the framework is simple, flexible, and efficient.

Keywords: Banded matrix; Bayesian estimation; Collapsed sampler; Markov chain Monte Carlo (MCMC); Kalman filter; State smoothing; Dynamic factor model; Time-varying parameter model.

Biographical Notes: Joshua Chan is a Ph.D. student in the Department of Mathematics, University of Queensland, Australia. His research interests lie in the field of Monte Carlo simulation, particularly the methodological and computational issues of Markov chain Monte Carlo methods.

Ivan Jeliazkov is Assistant Professor of Economics at the University of California, Irvine, whose main research interests are in Bayesian econometrics, Markov chain Monte Carlo (MCMC) estimation, latent variable models in discrete data and time series analysis, and model comparison.

1 Introduction

A number of important econometric models can be cast in state space form, including models with dynamic factors, time-varying parameters, autoregressive moving average disturbances, and

*Department of Mathematics, University of Queensland, Brisbane, QLD 4072, Australia. The author's research was supported by the Australian Research Council (Discovery Grant DP0985177). E-mail: c.chan5@uq.edu.au.

†Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100. E-mail: ivan@uci.edu. We would like to thank Bill McCausland for stimulating discussions on this subject.

stochastic volatility, among others. Extensive reviews of state space models and their applications in time series analysis are given in Kim and Nelson (1999) and Durbin and Koopman (2001). In this paper we consider the linear Gaussian state space form in which, for $t = 1, \ldots, T$, the $n \times 1$ vector of observations y_t is assumed to depend on the $q \times 1$ latent state vector η_t according to the hidden Markov structure

$$y_t = X_t\beta + G_t\eta_t + \varepsilon_t, \tag{1}$$

$$\eta_t = Z_t\gamma + F_t\eta_{t-1} + \nu_t, \tag{2}$$

where (2) is initialized with $\eta_1 \sim N(Z_1\gamma, D)$, and

$$\begin{pmatrix} \varepsilon_t \\ \nu_t \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{pmatrix}\right).$$

Equation (1) is often referred to as the measurement or observation equation, while (2) is called the transition or evolution equation. Define $y = (y_1', \ldots, y_T')'$ and $\eta = (\eta_1', \ldots, \eta_T')'$, and let θ represent the model parameters (i.e. β, γ, $\{G_t\}$, $\{F_t\}$, and the unique elements of Ω_{11}, Ω_{22}, and D). The covariates $\{X_t\}$ and $\{Z_t\}$ are treated as given and will be suppressed in all conditioning sets below.

Despite the availability of tractable conditional posterior distributions for the model unknowns in (1) and (2), estimation of state space models by Markov chain Monte Carlo (MCMC) methods can in many instances be rather challenging. The usual complications due to high dimensionality imply that estimation algorithms should be scalable in the key dimensions of the model (n, q, or T). Moreover, because the model unknowns enter equations (1) and (2) additively or multiplicatively, MCMC algorithms using full conditional sampling may suffer from slow convergence and poor mixing. This problem can be further exacerbated by the intertemporal correlation in η due to its dynamic behavior. To improve MCMC performance, special care is needed in grouping the parameters in blocks that can be sampled jointly, rather than conditionally on each other.

One important advance in this regard involves the sampling of the latent states. Carter and Kohn (1994) and Früwirth-Schnatter (1994) demonstrate that joint sampling of the entire vector $\eta \sim \pi(\eta|y, \theta)$ in one step results in important efficiency improvements relative to one-at-a-time Gibbs

sampling from $\pi(\eta_t|y,\theta,\{\eta_j\}_{j\neq t})$. Their techniques build upon a traditional two-pass procedure. In the forward filtering pass, conducted for $t = 1,\ldots,T$, the Kalman filter iterations produce the predictive mean and variance of η_t given only the observations up to period t. At the conclusion of the forward filtering pass, the states are drawn in reverse time order $\eta_t \sim \pi(\eta_t|y,\theta,\{\eta_j\}_{j>t})$ for $t = T,\ldots,1$, which gives draws from the joint distribution $\pi(\eta|y,\theta)$. A similar approach was also developed by Chib (1996, 1998) for hidden Markov models. The sampling techniques have been applied in other settings such as seemingly unrelated regression (SUR) models with serial correlation and time varying parameters (Chib and Greenberg, 1995) and stochastic volatility models (Jacquier et al., 1994; Shephard and Pitt, 1997; Kim et al., 1998). The methods are carefully reviewed in Kim and Nelson (1999) and Durbin and Koopman (2001), and the basic recursions are briefly summarized in Appendix A. Important further refinements of the simulation approach are offered in de Jong and Shephard (1995), using the distribution of the errors in the model, and in Durbin and Koopman (2002), using auxiliary data samples. Another sampling improvement was recently introduced by Chib et al. (2006), who suggested that because G_t and η_t enter multiplicatively in (1), they should be sampled jointly. Their approach builds upon a Metropolis-Hastings (MH) algorithm in which, given the remaining unknowns in the model, the parameters in G_t are sampled marginally of the states, which is followed by sampling of the latent states conditioned on G_t. The additional computational costs of that algorithm pay off through a dramatic improvement in MCMC performance relative to full conditional sampling.

Motivated by the aforementioned considerations, we pursue several main goals in this paper. First, we discuss a simple derivation of the joint density of the states, $\pi(\eta|y,\theta)$, conditional on the data and the parameters. A key feature of this derivation is its conceptual clarity, which significantly facilitates subsequent aspects of the analysis. In comparison, the reliance of existing methods on multiple loops through time, though effective in many instances, may limit their transparency. Second, we show how density evaluation and simulation of draws $\eta \sim \pi(\eta|y,\theta)$ can be achieved efficiently using block-banded and sparse matrix algorithms working with the precision matrix of

that distribution. The techniques are scalable in the dimension of the model and do not involve the Kalman filtering and smoothing recursions. The approach builds upon techniques that have been heavily used in nonparametric regression (e.g. Silverman, 1985; Fahrmeir and Lang, 2001; Chib and Jeliazkov, 2006; Chib et al., 2009), spatial models (Rue, 2001; Knorr-Held and Rue, 2002), and smooth coefficient models (Koop and Tobias, 2006), and although the applicability of these methods to state space models has been recognized (Fahrmeir and Kaufmann, 1991; Asif and Moura, 2002, 2005; Knorr-Held and Rue, 2002), they have not yet gained prominence in time-series analysis despite their versatility. Third, we show that the integrated likelihood $f(y|\theta)$, which gives the density of the data conditional on the parameters but integrated over the state vector η, can be obtained very easily. This is useful in estimation, maximization, and Bayesian and frequentist model comparison. Work on approximate state space models (Chin, 2001) and integrated likelihoods $f(y|\theta)$ (Bell, 2000) has also relied on the computational advantages of sparse and banded matrices, but in contrast to those approaches, our methods are exact because bandedness emerges as a feature of the precision matrix of $\pi(\eta|y,\theta)$ and is not just an approximation. Finally, we use the preceding developments to propose and implement a new blocking scheme for analysis of state space models that relies on collapsed MCMC sampling (Liu, 1994; Liu et al., 1994) and allows not only G_t, but also β, to be sampled marginally of the states η. An application involving a dynamic factor model of the U.S. economy shows that our approach leads to important improvements in estimation.

The remainder of the paper is organized as follows. In Section 2, we discuss several aspects of the inferential framework. Specifically, Section 2.1 offers a simple and conceptually clear derivation of the joint distribution $\pi(\eta|y,\theta)$. We build upon that derivation in Section 2.2 and discuss methods that can be used to sample the states jointly using efficient algorithms for banded matrices. Section 2.3 shows that the integrated likelihood can be evaluated easily without involving high-dimensional integration over η. Section 3 provides two empirical applications in macroeconomics which demonstrate the performance of the algorithms and the benefits conveyed by the new MCMC blocking scheme for fitting factor models. The paper concludes with a summary in Section 4.

2 Methodological Framework

2.1 A Simple Derivation

We begin by writing the joint sampling density $f(y|\theta,\eta)$ implied by (1). In particular, stacking (1) over the T time periods, we have

$$y = X\beta + G\eta + \varepsilon,$$

where

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_T \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & & \\ & \ddots & \\ & & G_T \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix},$$

with $\varepsilon \sim \mathcal{N}(0, I_T \otimes \Omega_{11})$, which implies that $f(y|\theta,\eta)$ is Gaussian with mean $X\beta + G\eta$ and covariance matrix $(I_T \otimes \Omega_{11})$, i.e.

$$f(y|\theta,\eta) = f_{\mathcal{N}}(y|X\beta + G\eta, I_T \otimes \Omega_{11}). \tag{3}$$

Turning attention to the distribution of η, we note that the directed conditional structure for $\pi(\eta_t|\theta,\eta_{t-1})$ in (2) implies a joint density for η, which can be obtained, following Fahrmeir and Kaufmann (1991), by defining

$$H = \begin{pmatrix} I_q & & & & \\ -F_2 & I_q & & & \\ & -F_3 & I_q & & \\ & & \ddots & \ddots & \\ & & & -F_T & I_q \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} D & & & & \\ & \Omega_{22} & & & \\ & & \Omega_{22} & & \\ & & & \ddots & \\ & & & & \Omega_{22} \end{pmatrix},$$

so that (2) can be written as $H\eta = Z\gamma + \nu$, where

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_T \end{bmatrix} \quad \text{and} \quad \nu = \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_T \end{bmatrix} \sim \mathcal{N}(0, S).$$

A simple change of variable from ν to η implies that

$$[\eta|\theta] \sim \mathcal{N}\left(\tilde{\eta}, K^{-1}\right), \tag{4}$$

where $\tilde{\eta} = H^{-1}Z\gamma$ (which can equivalently, and more efficiently, be obtained by iterating (2) in expectation) and the $Tq \times Tq$ precision matrix K is given by $K = H'S^{-1}H$, i.e.,

$$K = \begin{pmatrix} F_2'\Omega_{22}^{-1}F_2 + D^{-1} & -F_2'\Omega_{22}^{-1} & & & \\ -\Omega_{22}^{-1}F_2 & F_3'\Omega_{22}^{-1}F_3 + \Omega_{22}^{-1} & -F_3'\Omega_{22}^{-1} & & \\ & \ddots & \ddots & \ddots & \\ & & -\Omega_{22}^{-1}F_{T-1} & F_T'\Omega_{22}^{-1}F_T + S_{22}^{-1} & -F_T'\Omega_{22}^{-1} \\ & & & -\Omega_{22}^{-1}F_T & \Omega_{22}^{-1} \end{pmatrix}. \quad (5)$$

In the preceding derivation, of foremost significance is the fact that the precision matrix K is block-banded and contains only a small number of non-zero elements concentrated in a narrow band around the main diagonal. This observation, together with the fact that K is symmetric, implies that storage costs are low and that the computational benefits of working with banded or sparse matrix algorithms can be exploited in this setting. This will be discussed in Section 2.2. In addition, this approach to deriving the joint distribution of η makes it easy to accommodate (by simply changing H and S) versions of the model where (2) is specified so that η_t follows a second or higher order dynamic process. This is important for avoiding the degeneracies that result when a higher order dynamic process for η_t is written as a first order process by enforcing a set of identities in (2). Another point is that because of their modularity, the methods discussed here can be adapted to settings where G_t and F_t depend on a small number of parameters through a hierarchical structure; of course, the methods trivially accommodate the case where those parameter matrices are time-invariant, which is typically the case in applications.

Given the densities $f(y|\theta,\eta)$ and $f(\eta|\theta)$ in (1) and (2), a simple application of Bayes' theorem implies that the conditional distribution $f(\eta|y,\theta) \propto f(y|\eta,\theta)\, f(\eta|\theta)$ is immediately available and the derivations can be found in standard references such as Zellner (1971), Poirier (1995), Koop (2003), Gelman et al. (2003), or Greenberg (2008). In particular, upon combining the Gaussian likelihood in (3) with the Gaussian prior in (4), the conditional posterior is also Gaussian

$$[\eta|y,\theta] \sim \mathcal{N}\left(\hat{\eta}, P^{-1}\right), \quad (6)$$

where the precision P and mean $\hat{\eta}$ are given by

$$P = K + G'\left(I \otimes \Omega_{11}^{-1}\right) G, \tag{7}$$

$$\hat{\eta} = P^{-1}\left(K\tilde{\eta} + G'\left(I \otimes \Omega_{11}^{-1}\right)(y - X\beta)\right). \tag{8}$$

Since $G'\left(I \otimes \Omega_{11}^{-1}\right) G$ is banded, it follows that P is also banded, thus affording the aforementioned computational benefits that will be reviewed shortly. We mention that when D is finite, the distribution of η is proper. If D is diffuse, then K is of reduced rank. This is easily handled by redefining H (simply by excluding the first q rows) and S (excluding the first q rows and columns). As a result, the method works with both proper and improper prior distributions.

2.2 Efficient Estimation of the Latent States

We begin by discussing an efficient simulation method for the density $\pi(\eta|y,\theta)$ given in (6)–(8). Since the precision matrix P is banded, one should note that sampling of $[\eta|y,\theta]$ need not include an inversion to obtain P^{-1} and $\hat{\eta}$ in (8). Instead, the mean $\hat{\eta}$ can be found by computing the (banded) Cholesky factor C of P such that $C'C = P$ and solving

$$C'C\hat{\eta} = K\tilde{\eta} + G'\left(I \otimes \Omega_{11}^{-1}\right)(y - X\beta), \tag{9}$$

which is done in $\mathcal{O}(Tq^2)$ operations by first solving (9) for $C\hat{\eta}$ by forward substitution and then using the result to solve for $\hat{\eta}$ by back substitution. Similarly, to obtain a random draw from $\mathcal{N}(\hat{\eta}, P^{-1})$ efficiently, sample $u \sim \mathcal{N}(0, I)$, and solve $Cx = u$ for x by back-substitution. It follows that $x \sim \mathcal{N}\left(0, P^{-1}\right)$. Adding the mean $\hat{\eta}$ to x, one obtains a draw $\eta \sim \mathcal{N}(\hat{\eta}, P^{-1})$. It is worth mentioning that because K and P are not only banded (which eliminates large numbers of zeros from consideration), but they are also block-banded (i.e. even within the bands, there are blocks of zeros), the Cholesky factors will also have zero entries in the same location. For this reason additional computational savings can be generated in operations such as multiplication, forward and back-substitution, and Cholesky factorization by using block-banded or sparse matrix algorithms that take those zeros into account. We summarize the above procedures as follows.

Algorithm 1 ***Efficient State Smoothing and Simulation***

1. *Compute* P *in* (7) *and obtain its Cholesky factor* C *such that* $C'C = P$.

2. `Smoothing:` *Solve* (9) *by forward- and back-substitution to obtain* $\hat{\eta}$.

3. `Simulation:` *Sample* $u \sim \mathcal{N}(0, I)$, *and solve* $Cx = u$ *for* x *by back-substitution and take* $\eta = \hat{\eta} + x$, *so that* $\eta \sim \mathcal{N}(\hat{\eta}, P^{-1})$.

Several remarks about Algorithm 1 can be made. First, this method is very efficient and has been widely used in nonparametric regression (see, for example, Silverman, 1985; Fahrmeir and Lang, 2001; Chib and Jeliazkov, 2006). Computational efficiency aspects of banded matrix operations are discussed in Golub and van Loan (1983), Knorr-Held and Rue (2002), Rue and Held (2005) and McCausland et al. (2009). In the examples in Section 3, this algorithm generated 20-40% faster run times than the Kalman filter and smoother recursions. Second, compared with existing algorithms, Algorithm 1 is transparent and requires trivial effort to program. Its applicability is further facilitated by the inclusion of banded and sparse matrix algorithms in standard statistical programming packages such as Gauss (Aptech Systems, Inc., Black Diamond, WA) and Matlab (The MathWorks, Inc., Natick, MA). Third, because the method does not require any auxiliary simulation (c.f. Durbin and Koopman, 2002), it is scalable as the size of y_t grows. Finally, computation of the density ordinate $\pi(\eta|y, \theta)$ is also very efficient because banded matrix multiplications and determinant evaluations are inexpensive. This observation, together with the integrated likelihood approach that is discussed next, allows for more efficient blocking in the construction of MCMC algorithms such as those considered in Section 3.

2.3 Integrated Likelihood Evaluation

The integrated likelihood $f(y|\theta)$, defined as

$$f(y|\theta) = \int f(y|\theta, \eta)\pi(\eta|\theta)d\eta, \tag{10}$$

where $f(y|\theta,\eta)$ is the likelihood and $\pi(\eta|\theta)$ represents the prior, is often used in likelihood evaluation for classical and Bayesian problems involving optimization and model comparison. Evaluating the integrated likelihood via the multivariate integration in (10) is computationally intensive and often impractical. The derivation of the full conditional density $[\eta|y,\theta]$ discussed in Section 2.1, however, provides a simple way to evaluate $f(y|\theta)$ without invoking (10) and instead using the approach in Chib (1995). By Bayes' theorem, we can write the integrated likelihood as

$$f(y|\theta) = \frac{f(y|\theta,\eta)\pi(\eta|\theta)}{\pi(\eta|y,\theta)}, \tag{11}$$

where $\pi(\eta|y,\theta)$ denotes the full conditional posterior density of η given (y,θ). Due to the banded nature of prior and posterior precision matrices (K and P, respectively), evaluation of this quantity can be done very efficiently by simply evaluating the terms in (11) at a single point such as the posterior mean or mode as suggested by Chib (1995). Note, in particular, that the choice $\eta = \hat{\eta}$ eliminates the need to compute the exponential part of the pdf in the denominator density. Also note that this approach to evaluating the integrated likelihood is exact and does not require approximations (such as the Laplace approximation whose applicability is studied, for example, in DiCiccio et al., 1997, and Bell, 2000). In conjunction with the results in Section 2.2, the method provides a platform for Bayesian model comparison (Chib, 1995; Chib and Jeliazkov, 2001, 2005) and allows efficient MCMC blocking as discussed in our examples.

3 Empirical Applications

In this section we apply the techniques discussed thus far to estimate two models in empirical macroeconomics. The data set consists of $T = 229$ quarterly observations (1948:Q1 to 2005:Q1) on $n = 4$ macroeconomic series: output (GDP) growth, unemployment rate, interest rate, and inflation. A popular model for empirical analysis in macroeconomics is the vector autoregression (VAR) model, which has traditionally been employed to capture the evolution and interdependence among the time series. In our examples we extend the traditional VAR specification in two ways. In Section 3.1 we consider a time-varying parameter vector autoregressive (TVP-VAR) model which

allows us to study the issue of structural instability, possibly due to evolving technology, expectations, or institutional settings. The regression parameters in the model follow a random walk, which permits the economy to evolve differently over time, while maintaining that such changes in dynamic behavior should occur smoothly. Section 3.2 considers a dynamic factor VAR (DF-VAR) model that is useful in studying macroeconomic shocks and business cycles. The model specification allows the vector of macroeconomic variables to depend on its past realizations as well as on a systematic unobserved factor that evolves dynamically over time. The proposed MCMC sampler for this model involves a new collapsed blocking scheme, made possible by the techniques of Section 2, that significantly improves the convergence and mixing of the Markov chain.

In both applications, we report the inefficiency factors for the sampled parameters. The inefficiency factors approximate the ratio of the numerical variance of the posterior mean from the MCMC chain relative to that from an independent sample of equal size, and hence provide a useful measure of the mixing of the Markov chain. Inefficiency factors are related to an alternative metric, known as the effective sample size, which gives the size of an independent sample giving the same numerical variance as the MCMC sample (Robert and Casella, 2004).

3.1 Time-varying parameter model

In this section we consider the analysis of a time-varying parameter VAR (TVP-VAR) model in order to study the extent to which structural instability may be a feature in this type of macroeconomic model. Specifically, for $t = 1, \ldots, T$, we consider the first-order TVP-VAR model

$$y_t = \mu_t + \Gamma_t y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \Omega_{11}), \tag{12}$$

where y_t is a vector containing measurements on the aforementioned $n = 4$ economic variables (output growth, unemployment, income, and inflation), μ_t and Γ_t contain time-varying parameters, and Ω_{11} is an unknown $n \times n$ positive definite covariance matrix. The analysis is performed conditionally on the initial data point and consequently our sample consists of $T = 228$ observations.

For estimation purposes, equation (12) will be written in SUR form (Zellner, 1962) as

$$y_t = X_t \beta_t + \varepsilon_t, \tag{13}$$

where X_t is given by

$$\begin{pmatrix} (1, y'_{t-1}) & & & \\ & (1, y'_{t-1}) & & \\ & & \ddots & \\ & & & (1, y'_{t-1}) \end{pmatrix} \tag{14}$$

and β_t is a vector containing the corresponding parameters from μ_t and Γ_t ordered equation by equation, i.e., $\beta_t \equiv \text{vec}\,([\mu_t \,:\, \Gamma_t]')$, where the vec$(\cdot)$ operator stacks the columns of a matrix into a vector. The evolution of the vector of parameters β_t for $t = 2, \ldots, T$ is governed by

$$\beta_t = \beta_{t-1} + \nu_t, \tag{15}$$

with $\beta_1 \sim \mathcal{N}(0, D)$, where $\nu_t \sim \mathcal{N}(0, \Omega_{22})$, and $\Omega_{22} = \text{diag}(\sigma_1^2, \ldots, \sigma_q^2)$ is an unknown $q \times q$ diagonal matrix with $q = n(n+1)$. The transition equation (15) allows the elements of β_t to evolve gradually over time by penalizing large changes between successive values, thus *a priori* favoring the simple VAR model with time-invariant parameters. The joint distribution of the entire vector $\beta = (\beta'_1, \ldots, \beta'_T)'$ can be determined in a straightforward fashion following the derivations in Section 2.1. Chib and Greenberg (1995) have considered a TVP-SUR model using a hierarchical setup in which the evolution of β_t is related to a random walk in a lower-dimensional latent factor, and estimation proceeds by forward-filtering backward-sampling methods. In contrast, using the block-banded matrix methods discussed in Section 2.2, here we sample the entire sequence $\beta = (\beta'_1, \ldots, \beta'_T)'$ consisting of 228 20-dimensional vectors β_t in a single step. Specifically, stacking the observations in (13) over t, and using the derivations in Section 2.1, we have

$$y = X\beta + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_T \otimes \Omega_{11})$$

where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_T \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix},$$

and for $\beta = (\beta_1', \ldots, \beta_T')'$, we have

$$[\beta|\Omega_{22}] \sim \mathcal{N}\left(0, K^{-1}\right),$$

where K is a simplification of (5) with $\{F_t\}$ being replaced by identity matrices. Once written in this form, the model becomes conceptually simple and estimation can proceed as in Section 2.2.

Our analysis of the U.S. macroeconomic data is carried out under the following prior distributions: $\Omega_{11} \sim \mathcal{IW}(\nu_1^0, S_1^0)$ and $\sigma_i^2 \sim \mathcal{IG}(\nu_{i2}^0/2, S_{i2}^0/2)$, for $i = 1, \ldots, q$, with $\nu_1^0 = n + 3$, $S_1^0 = I$, $\nu_{12}^0 = \cdots = \nu_{q2}^0 = 6$, $S_{12}^0 = \cdots = S_{q2}^0 = 0.01$, and $D = 5 \times I$, where $\mathcal{IW}(\cdot, \cdot)$ and $\mathcal{IG}(\cdot, \cdot)$ denote the inverse-Wishart and inverse-gamma distributions, respectively. Posterior inference is based on an MCMC sampler that sequentially draws from $\pi(\beta|y, \Omega_{11}, \Omega_{22})$, $\pi(\Omega_{11}|y, \beta, \Omega_{22})$ and $\pi(\Omega_{22}|y, \beta, \Omega_{11})$. The techniques we present are also applicable to similar models such as the smooth coefficient models of Koop and Tobias (2006). We summarize the MCMC approach as follows.

Algorithm 2 ***MCMC Estimation of the TVP-VAR Model***

1. *Sample* $[\beta|y, \Omega_{11}, \Omega_{22}] \sim \mathcal{N}(\hat{\beta}, P^{-1})$, *with precision* $P = K + X'\left(I \otimes \Omega_{11}^{-1}\right)X$ *and mean* $\hat{\beta} = P^{-1}X'\left(I \otimes \Omega_{11}^{-1}\right)y$, *which is done efficiently via Algorithm 1.*

2. *Sample* $[\Omega_{11}|y, \beta] \sim \mathcal{IW}\left(\nu_1^0 + T, S_1^0 + \sum_{t=1}^{T}(y_t - X_t\beta_t)(y_t - X_t\beta_t)'\right)$.

3. *Sample* $[\Omega_{22}|y, \beta]$ *from* q *independent inverse-gamma distributions of the form*

$$[\sigma_i^2|y, \beta] \sim \mathcal{IG}\left(\frac{\nu_{i2}^0 + T - 1}{2}, \frac{S_{i2}^0 + \sum_{t=2}^{T}(\beta_{t,i} - \beta_{t-1,i})^2}{2}\right), \quad i = 1, \ldots, q,$$

where $\beta_{t,i}$ *is the* i*th element of* β_t.

Estimation of the TVP-VAR model for the post-war U.S. data was based on a sample of 20000 MCMC draws following a burn-in of 1000 draws. The results were verified by a forward-filtering backward-sampling algorithm, but the run time of Algorithm 2 was approximately one third less than that of the alternative approach.

The results for the main parameters of interest β_t are presented in Figure 1. The figure reveals that all effects in the GDP growth equation appear to have drifted towards zero in the latter part of the sample. This supports the contention that economic growth has become more stable over time due to lower dependence on past outcomes. Indeed, there has been much debate among economists about the existence, sources, and ramifications of a "Great Moderation" in U.S. output growth volatility based on comparisons of data before and after the mid-1980s (Stock and Watson, 2003). The unemployment equation parameters appear to be more stable over time relative to the parameters in the other three panels. The interest and inflation rate equations also exhibit some strong instability in the early part of the sample and the late 70s/early 80s; however, the broader trends that are especially visible in those equations concern the behavior of the time-varying intercepts, which have risen in recent years relative to the levels in the aftermath of World War II.

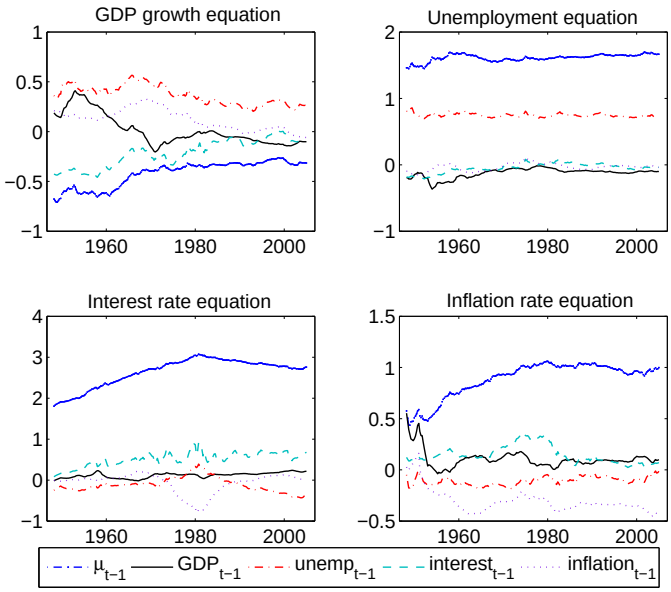

Figure 1: Evolution of the parameters β_t in the TVP-VAR model.

As an illustration of the mixing properties of the MCMC sampler in this application, Figure 2 presents boxplots of the inefficiency factors. The inefficiency factors show that despite the size of the latent state vector, the MCMC sampler performs very well. We note that, as in most latent variable models, parameters that depend on observed data are better estimated than parameters

that depend on latent outcomes, and for this reason the inefficiency factors for the elements of Ω_{11} are lower than those for Ω_{22}.

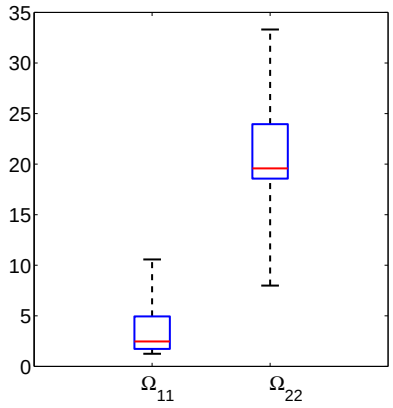

Figure 2: Inefficiency factors in the estimation of the TVP-VAR model of the U.S. economy.

3.2 Dynamic factor model

In our second application, we consider a dynamic factor VAR (DF-VAR) model given by

$$y_t = \mu + \Gamma y_{t-1} + A f_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \Omega), \tag{16}$$

where $\Omega = \text{diag}(\omega_{11}, \ldots, \omega_{nn})$, and

$$f_t = \gamma f_{t-1} + \nu_t, \qquad \nu_t \sim \mathcal{N}(0, \sigma^2), \tag{17}$$

is a dynamic economy-wide factor which reflects unobserved sources of macroeconomic volatility. Equation (17) is initialized with f_1 distributed according to the steady state distribution $f_1 \sim N\left(0, \sigma^2/\left(1-\gamma^2\right)\right)$. Inclusion of the factor $f = (f_1, \ldots, f_T)'$ in empirical models in economics and finance is often motivated by theoretical considerations, but it also serves as a parsimonious way of modeling the co-movement in the series. The joint distribution $\pi\left(f|\gamma, \sigma^2\right)$ can be easily derived following Section 2.1, whereby

$$[f|\gamma, \sigma^2] \sim \mathcal{N}\left(0, \sigma^2 F_0^{-1}\right), \tag{18}$$

and the precision involves

$$F_0 = \begin{pmatrix} 1 & -\gamma & & & \\ -\gamma & 1+\gamma^2 & -\gamma & & \\ & \ddots & \ddots & \ddots & \\ & & -\gamma & 1+\gamma^2 & -\gamma \\ & & & -\gamma & 1 \end{pmatrix}.$$

Since neither the vector of loadings A nor the factor f is known, there are two identification problems that relate to the scale and sign of A and f because $Af_t = (cA)(f_t/c)$ for any $c \neq 0$. Both the scale and sign problems are resolved by fixing the first element of A at 1, i.e. $A = (1, a')'$, and only the vector of unrestricted loadings a needs to be estimated. Although estimation of a DF-VAR model with more than one factor poses no conceptual or computational difficulties (see Sections 2 and 3.1) and all factors can still be sampled jointly in one step, identification in those models requires additional constraints. In particular, to prevent the possibility for simultaneous permutation of the factors and the columns of A, the latter matrix is typically restricted so that $A[i,j] = 0$ for $j > i$ and $A[i,i] = 1$. For the purposes of estimation the model is written in the form

$$y_t = X_t\beta + Af_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{19}$$

where X_t is given in (14) and $\beta = \text{vec}([\mu : \Gamma]')$. Stacking the equations in (19) over t, we have

$$y = X\beta + (I_T \otimes A)f + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_T \otimes \Omega), \tag{20}$$

where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ \vdots \\ X_T \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix}.$$

Given the distribution $[f|\gamma, \sigma^2] \sim \mathcal{N}(0, \sigma^2 F_0^{-1})$ in (18) and the priors $\beta \sim \mathcal{N}(\beta_0, B_0)$, $a \sim \mathcal{N}(a_0, A_0)$, $\omega_{ii} \sim \mathcal{IG}(\nu_{i0}/2, r_{i0}/2)$ for $i = 1, \ldots, n$, $\sigma^2 \sim \mathcal{IG}(s_0/2, \delta_0/2)$, and $\gamma \sim \mathcal{TN}_{(-1,1)}(\gamma_0, G_0)$, where $\mathcal{TN}_{(a,b)}(\cdot)$ represents a truncated normal distribution on the interval (a, b), all full conditional posterior distributions are tractable and in principle estimation could proceed by simulation from those densities. This gives rise to MCMC Scheme 1, which uses the following parameter blocking (conditioning on parameters that do not enter the full-conditional distributions is suppressed).

Scheme 1 ***Full-Conditional MCMC Blocking***

1. $[\beta|y, A, f, \Omega]$
2. $[a|y, \beta, f, \Omega]$, *where* $A = (1, a')'$
3. $[f|y, \beta, A, \Omega, \gamma, \sigma^2]$
4. $[\Omega|y, \beta, A, f]$
5. $[\gamma|f, \sigma^2]$
6. $[\sigma^2|f, \gamma]$

The main benefit of Scheme 1 is that the full-conditional distributions are straightforward and tractable, and hence MCMC simulation is very simple. For this reason, samplers based on such blocking have appeared in various contexts (e.g. Jacquier et al., 1994; Geweke and Zhou, 1996; Aguilar and West, 2000; Kose et al., 2003; Hogan and Tchernis, 2004). Unfortunately, the MCMC output produced under Scheme 1 often suffers from slow convergence and poor mixing. For this reason, Chib et al. (2006) suggest an alternative blocking scheme where the loadings A and the factors f are sampled in a single step by simulating A marginally of f, followed by drawing f from its full-conditional distribution. This reduces the inefficiency factors of the sampler and justifies the additional costs of sampling $[a|y, \beta, \Omega]$ by an MH step. This scheme is presented next.

Scheme 2 ***Sampling A and f in One Block***

1. $[\beta|y, A, f, \Omega]$
2. $[a, f|y, \beta, \Omega, \gamma, \sigma^2]$, *which is done using*
   (a) $[a|y, \beta, \Omega, \gamma, \sigma^2]$, *which does not depend on* f
   (b) $[f|y, \beta, A, \Omega, \gamma, \sigma^2]$
3. $[\Omega|y, \beta, A, f]$
4. $[\gamma|f, \sigma^2]$
5. $[\sigma^2|f, \gamma]$

Two key considerations motivate our proposed extensions of these techniques. First, the model in (19) involves two components, $X_t\beta$ and Af_t, both of which involve unknown quantities and enter

additively to determine the conditional mean of y_t. This feature has typically caused slow MCMC mixing in other settings, such as panel data models (Chib and Carlin, 1999), and has required some care in the construction of Markov chains. While it is not practical to integrate both A and f in the sampling of β, we show that a collapsed MCMC algorithm (Liu, 1994; Liu et al., 1994) that simulates both β and A marginally of f can be constructed easily, sampled efficiently, and can lead to substantial improvements in performance by reducing the conditional dependence in the Markov chain. Second, while the approach to marginalizing f suggested in Chib et al. (2006) works well in static factor models, extensions to models with dynamic factors are more difficult. However, we show that the methods in Section 2.3 allow efficient evaluation of the density of a marginally of f in both dynamic and static settings. We present this sampling scheme next.

Scheme 3 ***Collapsed MCMC Sampling of β and A***

1. $[\beta|y, A, \Omega, \gamma, \sigma^2]$, *which does not depend on* f
2. $[a, f|y, \beta, \Omega, \gamma, \sigma^2]$, *which is done using*
   (a) $[a|y, \beta, \Omega, \gamma, \sigma^2]$, *which does not depend on* f
   (b) $[f|y, \beta, A, \Omega, \gamma, \sigma^2]$
3. $[\Omega|y, \beta, A, f]$
4. $[\gamma|f, \sigma^2]$
5. $[\sigma^2|f, \gamma]$

The main difference between Schemes 2 and 3 is that in Scheme 3, β is sampled marginally of f. In anticipation of the summary in Algorithm 3, we present the computational details in this step. In order to sample β marginally of f, we begin by integrating f out of (20) using the distribution in (18) to obtain

$$y = X\beta + \nu, \qquad \nu \sim \mathcal{N}(0, \Sigma),$$

where $\Sigma = [(I_T \otimes \Omega) + (I_T \otimes A)\sigma^2 F_0^{-1}(I_T \otimes A)']$ is a dense $Tn \times Tn$ covariance matrix. This representation is conceptually convenient because the conditional posterior is of known form

$$[\beta|y, A, \Omega, \gamma, \sigma^2] \sim \mathcal{N}(\hat{\beta}, B),$$

with mean and variance given by

$$B = (B_0^{-1} + X'\Sigma^{-1}X)^{-1} \quad \text{and} \quad \hat{\beta} = B(B_0^{-1}\beta_0 + X'\Sigma^{-1}y). \tag{21}$$

However, in order for the approach to be practical, we must find a way to obtain $X'\Sigma^{-1}X$ and $X'\Sigma^{-1}y$ that does not involve brute-force inversion of Σ. Fortunately, Woodbury's formula yields

$$\begin{aligned}
\Sigma^{-1} &= \left[(I_T \otimes \Omega) + (I_T \otimes A)\,\sigma^2 F_0^{-1}\,(I_T \otimes A)'\right]^{-1} \\
&= \left(I_T \otimes \Omega^{-1}\right) - \left(I_T \otimes \Omega^{-1}\right)(I_T \otimes A) \times \\
&\quad \times \left[\sigma^{-2}F_0 + (I_T \otimes A)'\left(I_T \otimes \Omega^{-1}\right)(I_T \otimes A)\right]^{-1}(I_T \otimes A)'\left(I_T \otimes \Omega^{-1}\right) \\
&= \left(I_T \otimes \Omega^{-1}\right) - \left(I_T \otimes \Omega^{-1}A\right)\left[\sigma^{-2}F_0 + I_T\left(A'\Omega^{-1}A\right)\right]^{-1}\left(I_T \otimes A'\Omega^{-1}\right).
\end{aligned}$$

Note that the matrix $P = \left[\sigma^{-2}F_0 + I_T\left(A'\Omega^{-1}A\right)\right]$ in the last line above is banded since F_0 is banded and $A'\Omega^{-1}A$ is a scalar. Letting $\tilde{y}_t = A'\Omega^{-1}y_t$, $\tilde{X}_t = A'\Omega^{-1}X_t$, $\tilde{y} = (\tilde{y}_1', \ldots, \tilde{y}_T')'$, and $\tilde{X} = (\tilde{X}_1', \ldots, \tilde{X}_T')'$, the mean and variance in (21) can be written as

$$B = \left(B_0^{-1} + \sum_{t=1}^{T} X_t'\Omega^{-1}X_t - \tilde{X}'P^{-1}\tilde{X}\right)^{-1}, \quad \hat{\beta} = B\left(B_0^{-1}\beta_0 + \sum_{t=1}^{T} X_t'\Omega^{-1}y_t - \tilde{X}'P^{-1}\tilde{y}\right), \tag{22}$$

which can be easily evaluated because P, as defined in the previous sentence, is banded.

We use the results in Section 2.3 to extend the approach of Chib et al. (2006) to the case of a dynamic factor and efficiently sample $[a|y, \beta, A, \Omega, \gamma, \sigma^2]$ marginally of f. Similarly to (11), which gives the density of y marginally of f, we can find the conditional density of a marginally of f:

$$\pi(a|y, \beta, A, \Omega, \gamma, \sigma^2) \propto \frac{\pi(a|y, \beta, f, \Omega, \gamma, \sigma^2)}{\pi(f|y, \beta, A, \Omega, \gamma, \sigma^2)}, \tag{23}$$

which holds for any value of f and is easy to evaluate because the full-conditional densities in the model are known. Note, moreover, that typically we will not need to include $\pi(f|y, \beta, \Omega, \gamma, \sigma^2)$ because it does not involve A, and can be absorbed in the constant of proportionality – the kernel of the density is all that is needed in optimization and MH steps (because any constants cancel in the probability of acceptance ratio). With these derivations, we are now ready to proceed with a formal statement of the MCMC estimation algorithm.

Algorithm 3 ***MCMC Sampler for the DF-VAR model***

1. *Sample* $[\beta|y, A, \Omega, \gamma, \sigma^2] \sim \mathcal{N}\left(\hat{\beta}, B\right)$ *where* $\hat{\beta}$ *and* B *are given in* (22).

2. *Sample* $[a, f|y, \beta, \Omega, \gamma, \sigma^2]$ *in one block as follows*

   (a) *Sample* $[a|y, \beta, \Omega, \gamma, \sigma^2]$ *marginally of* f *by MH with tailored proposal* $a^\dagger \sim q(\hat{a}, V)$*, and accept the proposed draw* $a^\dagger$ *with probability*

$$\alpha_{MH}(a, a^\dagger) = \min\left\{1, \frac{\pi(a^\dagger|y, \beta, A, \Omega, \gamma, \sigma^2)q(a|\hat{a}, V)}{\pi(a|y, \beta, A, \Omega, \gamma, \sigma^2)q(a^\dagger|\hat{a}, V)}\right\}$$

   (b) *Sample* $[f|y, \beta, A, \Omega, \gamma, \sigma^2] \sim \mathcal{N}(\hat{f}, F)$*, where* $\hat{f} = F\left(I_T \otimes A\right)'\left(I_T \otimes \Omega^{-1}\right)(y - X\beta)$ *and* $F = \left(F_0/\sigma^2 + (I_T \otimes A)'(I_T \otimes \Omega^{-1})(I_T \otimes A)\right)^{-1}$*, which is done efficiently via Algorithm 1.*

3. *Sample* $[\Omega|y, \beta, A, f]$ *by drawing* $\omega_{ii} \sim \mathcal{IG}((\nu_{i0} + T)/2, (r_{i0} + e_i'e_i)/2)$ *for* $i = 1, \ldots, n$*, where* e_i *is a* T*-vector of residuals from the* i*th observation equation.*

4. *Sample* $[\gamma|f, \sigma^2]$ *by MH with proposal* $\gamma^\dagger \sim \mathcal{N}(\hat{\gamma}, G)$*, where* $G = (G_0^{-1} + f_{1:T-1}'f_{1:T-1}/\sigma^2)^{-1}$ *and* $\hat{\gamma} = G(G_0^{-1}\gamma_0 + f_{1:T-1}'f_{2:T}/\sigma^2)$*, and the proposed value* $\gamma^\dagger$ *is accepted with probability*

$$\alpha_{MH}(\gamma, \gamma^\dagger) = \min\left\{1, \frac{f_{\mathcal{N}}(f_1|0, \sigma^2/(1 - \gamma^{\dagger 2}))}{f_{\mathcal{N}}(f_1|0, \sigma^2/(1 - \gamma^2))}\right\}.$$

5. *Sample* $[\sigma^2|f, \gamma] \sim \mathcal{IG}\left(\frac{s_0+T}{2}, \frac{\delta_0+(f^*-\hat{f})'(f^*-\hat{f})}{2}\right)$*, where* $f^* = (f_1\sqrt{1-\gamma^2}, f_2, \ldots, f_T)'$ *and* $\hat{f} = (0, \gamma f_2, \ldots, \gamma f_{T-1})'$.

Several clarifications about Algorithm 3 are in order. First, the tailored proposal density for sampling a is typically taken to be a multivariate Student's t density with low degrees of freedom to ensure heavy tails, and $\hat{a}$ and V are the mode and inverse of the negative Hessian at the mode of $\pi(a|y, \beta, \Omega, \gamma, \sigma^2)$ as in Chib et al. (2006). The ordinate $\pi(a|y, \beta, \Omega, \gamma, \sigma^2)$, when required, is efficiently determined by (23) as discussed earlier. Second, the MH algorithm for sampling γ is adapted from Chib and Greenberg (1994), and as a result the generalization studied in that paper

(e.g. longer lags or ARMA dynamics) can also be applied in our context. Third, applications of Algorithm 3 can also be extended to similar models estimated by Bayesian MCMC methods such as those presented in Belviso and Milani (2006).

We applied Algorithm 3 to study the post-war macroeconomic data for the United States. The estimated dynamic macroeconomic factor is shown in Figure 3, together with the timing of officially announced U.S. recessions. Figure 3 reveals that the factor captures the timing of these recessions quite well, but also gives an idea of the relative severity and dynamic evolution of each recession. Moreover, while the last two officially announced recessions (during the "Great Moderation" following the mid-1980s) were shorter, shallower, and further apart than recessions in earlier parts of the sample, the estimated dynamic factor does suggest that their end might have been announced a bit prematurely. The ability of the model to capture these features is notable because what followed the officially announced end of the 1990-1991 recession was a period of jobless growth when unemployment actually increased and remained elevated for several years. Recovery from the 2001 recession was stymied by the immediate aftermath of the 9/11 attacks, which affected consumer behavior, investor sentiment, and demand in key industries. Data limitations, however, make it difficult to draw definitive conclusions about recent macroeconomic trends. For instance, because macroeconomic data are revised for several years after their initial release, our sample does not include the latest U.S. recession which started in December 2007.

To demonstrate the efficiency gains of MCMC sampling by Scheme 3, in Figure 4 we present boxplots of the inefficiency factors of algorithms constructed under the three sampling schemes discussed earlier. The inefficiency factors were obtained from samples of length 100000 MCMC draws for each scheme, following a burn-in of 10000 draws. The figure reveals that the proposed collapsed MCMC sampler (S3) implemented by Algorithm 3 significantly improves MCMC mixing for several key parameters, most notably for β. Improvement is also visible in the draws for σ^2 and γ, while the savings are less pronounced in the case of Ω and A.

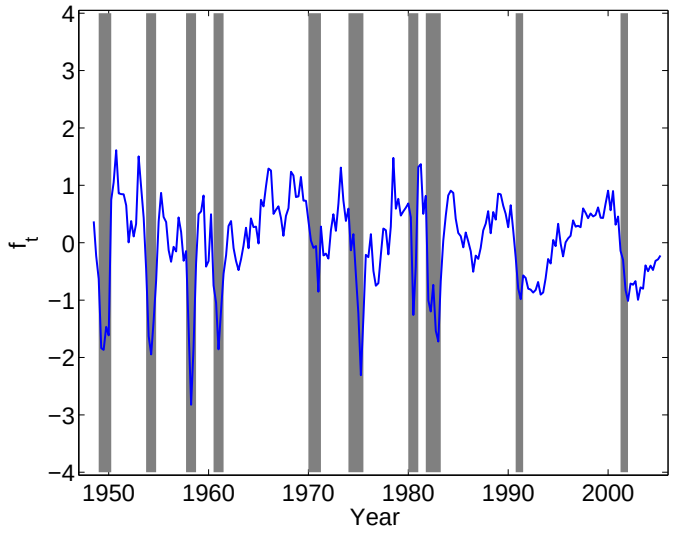

Figure 3: Estimated factor in the dynamic factor VAR model of the U.S. economy together with the timing of officially announced U.S. recessions (shaded regions).

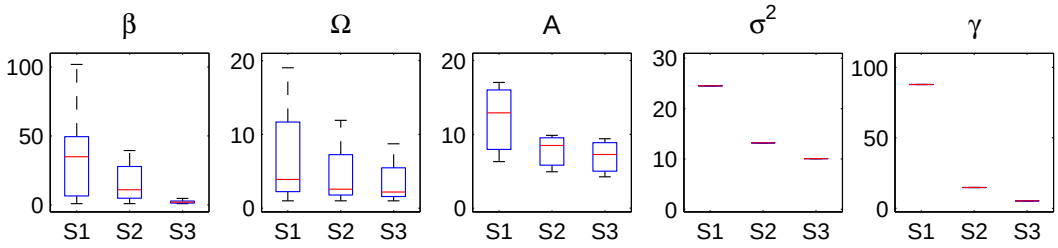

Figure 4: Inefficiency factors for the three sampling schemes (S1, S2, S3) in the DF-VAR model.

4 Concluding Remarks

The paper has examined the problem of estimating state space models by MCMC techniques. We have presented a straightforward derivation of the posterior distribution of the states, which is appealing not only for its conceptual simplicity, but also because it allows for efficient simulation of the states and efficient blocking in this class of models. We have also discussed a simple approach for evaluating the integrated likelihood which can be obtained easily and does not rely on approximations. The applicability of the proposed techniques is examined in two applications in economics, where a time-varying parameter and a dynamic factor model are used to analyze macroeconomic data for the post-war United States. The first of these applications shows that the

methods are viable in high-dimensional problems. In the second application, these methods allow us to provide a new MCMC blocking scheme that can significantly improve MCMC performance. While our discussion in this paper has focused on time-series analysis, the techniques in this paper are modular and can easily be applied in cross-sectional and spatial analysis. We intend to study their usefulness in those settings in future work.

Appendix: The Basic Kalman Filter and State Smoother

The forward filtering pass, conducted for $t = 1, \ldots, T$, involves a prediction and an updating step for each period. The prediction step computes conditional forecasts using the predictive distribution for η_t given information up to time $t-1$:

$$\begin{array}{ll} \eta_{t|t-1} = Z_t\gamma + F_t\eta_{t-1|t-1}, & P_{t|t-1} = F_tP_{t-1|t-1}F_t' + \Omega_{22}, \\ \varepsilon_{t|t-1} = y_t - X_t\beta - G_t\eta_{t|t-1}, & f_{t|t-1} = G_tP_{t|t-1}G_t' + \Omega_{11}. \end{array}$$

The recursions are initialized with the moments of the distribution of η_1. The quantities computed in this step are then updated to incorporate information up to time t:

$$K_t = P_{t|t-1}G_t'f_{t|t-1}^{-1}, \quad \eta_{t|t} = \eta_{t|t-1} + K_t\varepsilon_{t|t-1}, \quad P_{t|t} = P_{t|t-1} - K_tG_tP_{t|t-1}.$$

Upon completion of the forward filtering pass, a smoothing recursion is initiated in reverse time order to produce moments based on information from the entire sample. Starting with $\eta_{T|T}$ and $P_{T|T}$ from the last updating step, for $t = T-1, \ldots, 1$ one computes

$$\eta_{t|T} = \eta_{t|t} + S_t\left(\eta_{t+1|T} - F_{t+1}\eta_{t|t} - Z_t\gamma\right), \quad P_{t|T} = P_{t|t} - S_t\left(P_{t+1|T} - P_{t+1|t}\right)S_t',$$

where $S_t = P_{t|t}F_{t+1}'P_{t+1|t}^{-1}$. The two pass procedure underlies many of the forward-filtering backward-sampling methods referenced in the paper. In those approaches, η_T is sampled given $\eta_{T|T}$ and $P_{T|T}$ computed at the conclusion of the forward filtering pass, while the remaining states are simulated in reverse time order by replacing the conditional mean $\eta_{t+1|T}$ in each smoothing recursion with the value η_{t+1} that has already been simulated.

References

Aguilar, O. and West, M. (2000), "Bayesian Dynamic Factor Models and Portfolio Allocation," *Journal of Business & Economic Statistics*, 18, 338–357.

Asif, A. and Moura, J. M. F. (2002), "Fast Inversion of L-Block Banded Matrices and Their Inverses," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1369–1372.

Asif, A. and Moura, J. M. F. (2005), "Block Matrices with L-Block-Banded Inverse: Inversion Algorithms," *IEEE Transactions on Signal Processing*, 53, 630–642.

Bell, B. M. (2000), "The Marginal Likelihood for Parameters in a Discrete Gauss-Markov Process," *IEEE Transactions on Signal Processing*, 48, 870–873.

Belviso, F. and Milani, F. (2006), "Structural Factor-Augmented VARs (SFAVARs) and the Effects of Monetary Policy," *Topics in Macroeconomics*, 6, Iss. 3, Article 2.

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

Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S. (1996), "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–97.

Chib, S. (1998), "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 86, 221–241.

Chib, S. and Carlin, B. P. (1999), "On MCMC Sampling in Hierarchical Longitudinal Models," *Statistics and Computing*, 9, 17–26.

Chib, S. and Greenberg, E. (1994), "Bayes Inference in Regression Models With ARMA(p,q) Errors," *Journal of Econometrics*, 64, 183–206.

Chib, S. and Greenberg, E. (1995), "Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-Varying Parameter Models," *Journal of Econometrics*, 68, 339–360.

Chib, S. and Jeliazkov, I. (2001), "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.

Chib, S. and Jeliazkov, I. (2005), "Accept-Reject Metropolis-Hastings Sampling and Marginal Likelihood Estimation," *Statistica Neerlandica*, 59, 30–44.

Chib, S. and Jeliazkov, I. (2006), "Inference in Semiparametric Dynamic Models for Binary Longitudinal Data," *Journal of the American Statistical Association*, 101, 685–700.

Chib, S., Nardari, F., and Shephard, N. (2006), "Analysis of High Dimensional Multivariate Stochastic Volatility Models," *Journal of Econometrics*, 134, 341–371.

Chib, S., Greenberg, E., and Jeliazkov, I. (2009), "Estimation of Semiparametric Models in the Presence of Endogeneity and Sample Selection," *Journal of Computational and Graphical Statistics*, 18, 321–348.

Chin, T. M. (2001), "On Kalman Filter Solution of Space-Time Interpolation," *IEEE Transactions on Image Processing*, 10, 663–666.

de Jong, P. and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339–350.

DiCiccio, T. J., Kass, R. E., Raftery, A. E., and Wasserman, L. (1997), "Computing Bayes Factors by Combining Simulation and Asymptotic Approximations," *Journal of the American Statistical Association*, 92, 903–915.

Durbin, J. and Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

Durbin, J. and Koopman, S. J. (2002), "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89, 603–615.

Fahrmeir, L. and Kaufmann, H. (1991), "On Kalman Filtering, Posterior Mode Estimation and Fisher Scoring in Dynamic Exponential Family Regression," *Metrika*, 38, 37–60.

Fahrmeir, L. and Lang, S. (2001), "Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors," *Journal of the Royal Statistical Society – Series C*, 50, 201–220.

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

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003), *Bayesian Data Analysis*, Chapman & Hall, New York, 2 edn.

Geweke, J. and Zhou, G. (1996), "Measuring the Pricing Error of the Arbitrage Pricing Theory," *The Review of Financial Studies*, 9, 557–587.

Golub, G. H. and van Loan, C. F. (1983), *Matrix Computations*, Johns Hopkins University Press, Baltimore.

Greenberg, E. (2008), *Introduction to Bayesian Econometrics*, Cambridge University Press, New York.

Hogan, J. W. and Tchernis, R. (2004), "Bayesian Factor Analysis for Spatially Correlated Data, With Application to Summarizing Area-Level Material Deprivation From Census Data," *Journal of the American Statistical Association*, 99, 314–324.

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

Kim, C.-J. and Nelson, C. R. (1999), *State-Space Models with Regime Switching*, MIT Press, Cambridge.

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

Knorr-Held, L. and Rue, H. (2002), "On Block Updating in Markov Random Field Models for Disease Mapping," *Scandinavian Journal of Statistics*, 29, 597–614.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley & Sons, Inc., New York.

Koop, G. and Tobias, J. L. (2006), "Semiparametric Bayesian Inference in Smooth Coefficient Models," *Journal of Econometrics*, 134, 283–315.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003), "International Business Cycles: World, Region and Country Specific Factors," *American Economic Review*, 93, 1216–1239.

Liu, J. S. (1994), "The Collapsed Gibbs Sampler in Bayesian Computations With Applications to a Gene Regulation Problem," *Journal of the American Statistical Association*, 89, 958–966.

Liu, J. S., Wong, W. H., and Kong, A. (1994), "Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes," *Biometrika*, 81, 27–40.

McCausland, W., Miller, S., and Pelletier, D. (2009), "Precision-Based Methods for State Smoothing," `University of Montreal working paper`.

Poirier, D. J. (1995), *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Cambridge, MA.

Robert, C. P. and Casella, G. (2004), *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

Rue, H. (2001), "Fast Sampling of Gaussian Markov Random Fields," *Journal of the Royal Statistical Society – Series B*, 63, 325–338.

Rue, H. and Held, L. (2005), *Gaussian Markov Random Fields*, Chapman & Hall/CRC, Boca Raton, FL.

Shephard, N. and Pitt, M. K. (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, 84, 653–667.

Silverman, B. (1985), "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting," *Journal of the Royal Statistical Society – Series B*, 47, 1–52.

Stock, J. H. and Watson, M. W. (2003), "Has the Business Cycles Changed? Evidence and Explanations," in *Monetary Policy and Uncertainty: Adapting to a Changing Economy*, pp. 9–56, Federal Reserve Bank of Kansas City.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57, 348–368.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, New York.