ELSEVIER Journal of Econometrics 68 (1995) 339 360

JOURNAL OF Econometrics

Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models

Siddhartha Chib*[,a], Edward Greenberg[b]

[a] John M. Olin School of Business, Washington University, St. Louis, WA 63130, USA
[b] Department of Economics, Washington University, St. Louis, WA 63130, USA

(Received May 1993; final version received March 1994)

Abstract

We consider the use of Markov chain Monte Carlo methods to analyze hierarchical versions of Zellner's SUR model. In this context, the questions of Bayes estimation and model adequacy checking are considered. The approach is extended to SUR model with vector autoregressive and vector moving average errors of the first order. Finally, an efficient algorithm is developed to estimate a Markov time-varying parameter SUR model. The ideas are applied to both simulated and real data.

Key words: Gibbs sampling; Metropolis algorithm; Data augmentation; Markov chain Monte Carlo; Bayes factor; Hierarchical model; Vector autoregressive process; Vector moving average process; Time-varying parameter model; State space model
JEL classification: C11; C15; C21; C22

1. Introduction

Arnold Zellner's classic 1962 paper, 'An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests of Aggregation of Bias', introduced the seemingly unrelated regression (SUR) model. This landmark paper has stimulated extensive theoretical work and countless empirical application in

* Corresponding author.

We would like to thank Arnold Zellner, the Associate Editor, and two anonymous referees for helpful comments.

SSDI 0304407694016S3 H

econometrics and other areas (see, for example, Zellner, 1971; Box and Tiao, 1973; Srivastava and Giles, 1987; Spanos, 1986; or Goldberger, 1991).

Recently, Percy (1992) has shown that the Monte Carlo tool of Gibbs sampling is eminently suitable for estimating the parameters β and Ω of the Gaussian SUR model, where β is the regression vector and Ω is the covariance matrix of the errors. By iterating on draws from the posterior conditional distributions of β given Σ, and of Ω given β, an exact Bayesian result becomes possible. This approach also provides a solution to the problems of predicting out-of-sample observations and in-sample missing data.

Due to the importance of the SUR model in econometrics, this paper continues the study of the SUR model by simulation-based methods. In particular, we are interested in two key extensions of the basic SUR model. First, we study the estimation of the model under the assumption that the errors are serially correlated. In particular, we allow the errors to follow a vector autoregressive (VAR) process of order 1, and then a vector moving average (VMA) process of order 1. In the latter case, it is shown that the sampling of the moving average parameter matrix requires the Metropolis–Hastings algorithm; the former case (following the approach of Chib, 1993) requires only minor modifications to the basic Gibbs sampler strategy. Second, we investigate the estimation of time-varying parameter (TVP) models of the type recently considered by Min and Zellner (1993) and Gamerman and Migon (1993). This model only becomes amenable to the Gibbs sampler if the time-varying parameters are treated as additional unknown parameters. This data augmentation, however, leads to a major proliferation in the number of unknown parameters. To deal with this problem, a very important result is derived that allows for the simulation of all time-varying parameters from their joint distribution given the data and the remaining parameters.

Both these extensions are considered in the context of a hierarchical SUR model. We work within the hierarchical setup because it is straightforward to do so in the simulation context and also because it provides a great deal of flexibility in modeling. Specifically, the hierarchical model can be used to incorporate cross-equation restrictions, to assess the adequacy of reduced models, and to obtain the TVP model described above.

Finally, this paper proposes (and illustrates) the use of partial Bayes factors (Dawid, 1984) in the present context. Computing the Bayes factor in the Gibbs sampling setting is possible (see Carlin and Chib, 1993, and references therein) but the partial Bayes factor is more readily obtained provided one is willing to use a certain number of data points as a 'training sample' to update the prior distribution. The output of the Gibbs sampler applied to the training sample then yields the partial Bayes factors.

In this paper, the following notation and definitions will be employed. The p-variate normal distribution and the inverted gamma distribution will be signified by $\mathcal{N}_p(\mu, \Sigma)$ and $\mathcal{IG}(v, \delta)$, respectively. A matrix variable W: $p \times q$ that follows the matrix normal distribution will be denoted by $\text{MATN}(W \mid \mu, A \otimes B)$, which means that $\text{vecr}(W) \sim \mathcal{N}_{pq}(\mu, A \otimes B)$, where A is $p \times p, B$ is $q \times q$, and $\text{vecr}(\cdot)$ is the row-wise vectorization of a matrix (see Dawid, 1981). Its density function is

$$(2\pi)^{-pq/2}|A|^{-q/2}|B|^{-p/2}\exp\{-\tfrac{1}{2}\text{tr}[A^{-1}(W-\mu)B^{-1}(W-\mu)']\}.$$

For a symmetric positive definite matrix G which follows the p-dimensional Wishart distribution with v degrees of freedom and scale matrix A, we use the notation $\mathcal{W}_p(v, A)$. Its density is

$$k\frac{|G|^{(v-p-1)/2}}{|A|^{v/2}}\exp\{-\tfrac{1}{2}\text{tr}[A^{-1}G]\}, \qquad |G| > 0,$$

where k is a normalizing constant and A is a hyperparameter matrix (Press, 1982).

The rest of the paper is organized as follows. In Section 2 we describe the basic SUR model, the prior distributions, and the issue of model checking. In Section 3 we analyze the correlated error model, and in Section 4 we consider the SUR model with time-varying parameters. Section 5 applies the ideas to simulated and real data sets. In the real data example we fit SUR models to explain the growth rate of GNP for five OECD countries using 28 annual observation on eight variables. One of the models fit to this data set is a TVP-SUR model containing eight time-varying parameters. This example clearly demonstrates the usefulness and significance of the methods developed in the paper. The last section is a brief summary of the paper.

2. The hierarchical SUR model

We are interested in a p-vector of observations at time t, y_t, that is generated by the following hierarchical structure:

$$y_t = X_t\beta + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}_p(0, \Omega), \tag{1}$$

$$\beta = A_0\beta_0 + u, \qquad u \sim \mathcal{N}_k(0, B_0), \tag{2}$$

$$\beta_0 = A_1\mu + \eta, \qquad \eta \sim \mathcal{N}_m(0, B_1), \tag{3}$$

where the regression vector β: $k \times 1$ is related to a parameter β_0: $m \times 1$, which in turn is modeled in terms of a parameter μ: $r \times 1$. If $r \leqslant m \leqslant k$, then that would imply that the parameters are being successively projected onto lower-dimensional subspaces. We refer to (1), the observation equation, as the *first stage*, of the hierarchy, and to (2) and (3) as the *second* and *third stages* of the hierarchy.

Throughout this paper, we make the assumptions that (i) X_t, A_0, and A_1 are known; (ii) the variance matrices Ω, B_0, and B_1 are unknown; and (iii) the errors are mutually independent, which implies, for example, that $\{y_t\}$ is conditionally independent of (β_0, B_0, μ, B_1) given (β, Ω).

Note that this model gives rise to Zellner's SUR model in hierarchical form if X_t and β take the special forms $\mathrm{diag}(x'_{1t}, \ldots, x'_{pt})$ and $(\beta_1, \ldots, \beta_p)'$: $k \times 1$, respectively, where x_{it}: $k_i \times 1$ is the covariate vector and $k = \sum_i k_i$. The observation equation then reduces to the familiar form $y_{it} = x'_{it}\beta_i + \varepsilon_{it}$, with $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{pt})'$ distributed as $\mathcal{N}_p(0, \Omega)$.

In later sections, we will consider important variations on this model by letting ε_t be serially correlated and β be time-varying. But first, we present the Gibbs sampling implementation for the model in (1)–(3).[1]

2.1. Gibbs sampler for basic model

Suppose that prior information about the parameter vector μ and the unknown variance matrices is given by the forms $\mu \sim \mathcal{N}_r(\mu_0, M_0)$, $\Omega^{-1} \sim \mathcal{W}_p(v_0, R_0)$, $B_0^{-1} \sim \mathcal{W}_k(\rho_0, D_0)$, and $B_1^{-1} \sim \mathcal{W}_m(\rho_1, D_1)$, where the hyperparameters are assumed known. Then, the posterior density of the parameter $\psi = (\beta, \Omega, \beta_0, B_0, \mu, B_1)$ is given by

$$\pi(\psi \mid Y_n) \propto f(Y_n \mid \beta, \Omega)\pi(\psi), \tag{4}$$

where $f(Y_n \mid \beta, \Omega) \propto |\Omega|^{-n/2} \exp[-\frac{1}{2}\sum_{t=1}^{n}(y_t - X_t\beta)'\Omega^{-1}(y_t - X_t\beta)]$, $Y_n = (y_1, \ldots, y_n)$ is the sample data, and $\pi(\psi)$ is constructed from stages 2 and 3 of the hierarchy and the prior distribution on the parameters.

For this posterior density, the Gibbs sample can be shown to be defined through the following distributions, referred to as the full conditional distributions.[2]

(i) $\beta \mid Y_n, \psi_{-\beta} \sim \mathcal{N}_k(\hat{\beta}, V_n)$ and $\Omega^{-1} \mid Y_n, \psi_{-\Omega} \sim \mathcal{W}_p(v_0 + n, R_n)$, where $\hat{\beta} = V_n(B_0^{-1}A_0\beta_0 + \sum_{t=1}^{n} X_t\Omega^{-1}y_t)$, $V_n = (B_0^{-1} + \sum_{t=1}^{n} X_t'\Omega^{-1}X_t)^{-1}$, and $R_n^{-1} = R_0^{-1} + \sum_{t=1}^{n}(y_t - X_t\beta)(y_t - X_t\beta)'$. (These are obtained from the first and second stages of the hierarchy since the conditioning on μ and B_1 is irrelevant.)

(ii) $\beta_0 \mid Y_n, \psi_{-\beta_0} \sim \mathcal{N}_m(\Delta_1(B_1^{-1}A_1\mu + A_0'B_0^{-1}\beta), \Delta_1)$ and $B_0^{-1} \mid Y_n, \psi_{-B_0} \sim \mathcal{W}_k(\rho_0 + 1, [D_0^{-1} + (\beta - A_0\beta_0)(\beta - A_0\beta_0)']^{-1})$, where $\Delta_1 = (B_1^{-1} +$

[1] See Tanner and Wong (1987) and Gelfand and Smith (1990) for the original work on the Gibbs sampling algorithm. For a recent summary, with econometric applications, see Chib and Greenberg (1993b).

[2] The full conditional distribution of β, for example, is the distribution $\beta \mid Y_n, \psi_{-\beta}$, where $\psi_{-\beta} = \{\Omega, \beta_0, B_0, \mu, B_1\}$. Each full conditional distribution is proportional to the posterior distribution of ψ.

$A_0'B_0^{-1}A_0)^{-1}$. (These are obtained from the second and third stages of the hierarchy since the conditioning on Y_n, β, Ω is irrelevant.)

(iii) The distributions $\mu \mid Y_n, \psi_{-\mu} \sim \mathcal{N}_r(\Delta_2(M_0^{-1}\mu_0 + A_1'B_1^{-1}\beta_0), \Delta_2)$ and $B_1^{-1} \mid Y_n, \psi_{-B_1} \sim \mathcal{W}_m(\rho_1 + 1, [D_1^{-1} + (\beta_0 - A_1\mu)(\beta_0 - A_1\mu)']^{-1})$, where $\Delta_2 = (M_0^{-1} + A_1'B_1^{-1}A_1)^{-1}$. (These are obtained from the third and fourth stages of the hierarchy.)

All these conditional distributions are in standard form and therefore the Gibbs sampling algorithm is easy to apply. In addition, the theoretical convergence of the Gibbs sampler is assured because the sufficient conditions of Roberts and Smith (1993) are easily verified.[3] Hence, given the draw at the jth iteration, the next iteration is completed by simulating

$$
\begin{array}{lll}
\Omega^{-1(j+1)} & \text{from} & \Omega^{-1} \mid Y_n, \beta^{(j)}, \\
\beta^{(j+1)} & \text{from} & \beta \mid Y_n, \Omega^{-1(j+1)}, \beta_0^{(j)}, B_0^{(j)}, \\
\beta_0^{(j+1)} & \text{from} & \beta_0 \mid B_0^{(j)}, B_1^{(j)}, \mu^{(j)}, \beta^{(j+1)}, \\
B_0^{-1(j+1)} & \text{from} & B_0^{-1} \mid \beta^{(j+1)}), \beta_0^{(j+1)}, \\
\mu^{(j+1)} & \text{from} & \mu \mid \beta_0^{(j+1)}, B_1^{(j)}, \\
B_1^{-1(j+1)} & \text{from} & B_1^{-1} \mid \beta_0^{(j+1)}, \mu^{(j+1)}.
\end{array}
$$

This process can be iterated a large number of times, and after discarding the 'initial transient' all subsequent draws can be used for inferential purposes. Information on convergence can be obtained by monitoring the serial correlation in the draws, the numerical standard errors of estimates based on the output of the sampler, and through the diagnostics of Ritter and Tanner (1992), Gelman and Rubin (1992), and Zellner and Min (1993).

2.2. Model checking and Bayes factors

Estimation of the hierarchical SUR model is thus seen to be rather straightforward. A problem of dimensionality, however, can occur if k is large. Then, it might be of interest to shrink (or pool) the β_i towards a common value. This can, of course, be readily achieved in the hierarchical setting by defining A_0 to consist of identity matrices in stacked form.

[3] The conditions are (i) $\pi(\psi \mid Y_n) > 0$ implies that there exists an open neighbourhood N_ψ containing ψ and $\varepsilon > 0$ such that, for all $\psi' \in N_\psi$, $\pi(\psi' \mid Y_n) \geqslant \varepsilon > 0$; (ii) $\int \pi(\psi \mid Y_n)\,d\psi_i$ is locally bounded for all i, where ψ_i is the ith block of parameters; and (iii) the support of the posterior is arc-connected.

In this section, we consider the complementary question of model checking. Suppose it is of interest to assess the evidence in favor of a given reduced model of the kind $\beta = A_0\beta_0$, where for example β_0 may be the common value of β_i. The adequacy of this reduced model can be examined in the following way: Let $B_0 = \tau^2 I_k$, with τ^2 following an *a priori* inverse gamma $\mathcal{IG}(v_{00}/2, \delta_{00}/2)$ distribution, and include the distribution

$$\mathcal{IG}\left(\frac{v_{00}+k}{2}, \frac{\delta_{00}+(\beta - A_0\beta_0)'(\beta - A_0\beta_0)}{2}\right) \tag{5}$$

into the Gibbs sampler in place of the distribution of B_0. Then evidence about the reduced model can be obtained from the posterior draws of τ^2. If the draws are concentrated on small values of τ^2 *relative* to its prior, then that provides evidence in favor of the restrictions $\beta = A_0\beta_0$ (Albert and Chib, 1993a, b). Formally, the ratio

$$\frac{\Pr(\tau^2 \leqslant \varepsilon \mid y)\Pr(\tau^2 > \varepsilon \mid y)}{\Pr(\tau^2 \leqslant \varepsilon)\Pr(\tau^2 > \varepsilon)} \tag{6}$$

can be calculated for various values of ε, where the numerator is computed from the posterior relative frequencies and the denominator from the inverse gamma prior distribution. This approach is illustrated in Section 5.2.

Now consider the calculation of Bayes factors as a tool to discriminate between models. The Bayes factor for two models M_1 and M_2 is computable as $B_{12} = p(Y_n \mid M_1)/p(Y_n \mid M_2)$, the ratio of observed marginal densities for the two models. This quantity cannot be calculated directly in the present context and so recourse must be taken to simulation. Several valuable methods to do this have appeared recently (see George and McCulloch, 1993; Albert and Chib, 1993b; Carlin and Chib, 1993; and the references therein).

Alternatively, one can compute the partial Bayes factor (Dawid, 1984; O'Hagan, 1991) which is defined as follows. Let Y_{n_0} denote a set of n_0 observations that we may use as the training sample. Now suppose these observations are used to update a (possibly diffuse) prior distribution, yielding a proper posterior for all the models that are to be compared. Then, the *partial marginal density* for model M_k is defined as $m(Y_{n_1} \mid Y_{n_0}, M_k) = \int f(Y_{n_1} \mid \psi, M_k)\,\mathrm{d}\pi \times (\psi \mid Y_{n_0}, M_k)$, where Y_{n_1} is the remaining data and the integration is performed over the posterior distribution based on Y_{n_0}.

In the simulation setting, the estimation of this partial marginal density is straightforward and a simulation consistent estimate is available as

$$N^{-1}\sum_{i=1}^{N} f(Y_{n_1} \mid \psi^{(i)}, M_k), \tag{7}$$

where $\{\psi^{(j)}\}_{j=1}^{N}$ are drawn obtained from $\pi(\psi \mid Y_{n_0}, M_k)$. This calculation can be repeated for each model under consideration to form the partial Bayes factors.

The choice of Y_{n_0} in this calculation is somewhat arbitrary although it is necessary to ensure that the same Y_{n_0} is used for all models. In our time series application in Section 4, the first n_0 observations comprise Y_{n_0}.

3. SUR with correlated serial errors

With this background, we turn to the analysis of the hierarchical SUR model with serially correlated observation errors. We first let the observations errors follow a stationary VAR (1) process, and then consider the case of an invertible VMA(1) process. More general processes could be analyzed in the same way, but would entail additional notational complexity and a more detailed prior structure than what we have specified below. For the univariate ARMA(p,q) context, however, full Gibbs sampling frameworks are developed by Albert and Chib (1993c), Chib (1993), Chib and Greenberg (1994), and Marriott et al. (1995).

3.1. VAR(1) errors

Consider first the case in which the observation errors follow the stationary, first-order vector autoregression,

$$\varepsilon_t = \Phi \varepsilon_{t-1} + u_t, \qquad u_t \overset{\text{iid}}{\sim} \mathcal{N}_p(0, \Omega), \tag{8}$$

where Φ: $p \times p$ is a matrix with characteristic roots inside the unit circle.[4] The problem of interest is to simulate the posterior distribution of (ψ, Φ). We suppose that little is known a priori about Φ, and let $\Phi \propto I_S(\Phi)$, where $I_S(\Phi)$ is an indicator function that takes the value 1 if all roots of Φ lie in the unit circle and 0 otherwise. Thus, we assume that Φ be distributed uniformly on the stationary region (the distribution of the roots of Φ is not uniform but, if required, this distribution can be examined by simulation). Also suppose that the prior on ψ is of the form specific earlier.

Now to derive the relevant inputs for the Gibbs sampler, we first observe that the full conditional distributions of (β_0, B_0, μ, B_1) are unaffected by the serial correlation in the errors. We are left to determine the full conditional

[4] Non-Bayesian estimators of (β, Φ, Ω) for the nonhierarchical model setting of (1) and (8) can be obtained in several different ways (Parks, 1967; Guilkey and Schmidt, 1973; Judge et al., 1985, pp. 483–496).

distribution of Φ and the revised distributions of (β, Ω^{-1}). These three distributions can be derived along the lines of Chib (1993) provided we are willing to condition the analysis on (y_1, X_1). For $t \geqslant 2$, let

$$y_t^* = y_t - \Phi y_{t-1} \quad \text{and} \quad X_t^* = X_t - \Phi X_{t-1}.$$

Then, instead of (i) in Section 2 we now have

(i′) $\beta \mid Y_n, \psi_{-\beta}, \Phi \sim \mathcal{N}_k(\tilde{\beta}, \tilde{V}_n)$ and $\Omega^{-1} \mid Y_n, \psi_{-\Omega}, \Phi \sim \mathcal{W}_p(v_0 + n - 1, \tilde{R}_n)$, where $\tilde{\beta} = \tilde{V}_n(B_0^{-1}A_0\theta + \sum_{t=1}^n X_t^* \Omega^{-1} y_t^*)$, $\tilde{V}_n = (B_0^{-1} + \sum_{t=2}^n X_t^{*\prime}\Omega^{-1}X_t^*)^{-1}$, and $\tilde{R}_n^{-1} = R_0^{-1} + \sum_{t=2}^n (y_t^* - X_t^*\beta)(y_t^* - X_t^*\beta)'$.

In the case of Φ, the full conditional distribution is defined in terms of the 'realized errors' $e_t = y_t - X_t\beta$, $t \geqslant 1$. A simple calculation shows that:

(iv) $\Phi' \mid Y_n, \psi \propto \text{MATN}(\bar{\Phi}', (E_n'E_n)^{-1} \otimes \Omega) \times I_S(\Phi)$, where $\bar{\Phi}' = (E_n'E_n)^{-1} \times (\sum_{t=2}^n e_{t-1}e_t')$ and $E_n = (e_1, \ldots, e_{n-1})'$: $(n-1) \times p$.

To simulate the distribution in (iv), we let $E \equiv \bar{\Phi}' + PTQ'$, where T: $p \times p$ consists of independent standard normal variables, and P and Q satisfy the Choleski factorizations $PP' = (E_n'E_n)^{-1}$ and $QQ' = \Omega$, respectively. If all the roots of E are less than unity (the stationarity condition) then E is accepted; otherwise another T is drawn and the process repeated.

Thus, the estimation of the hierarchical SUR model with VAR(1) errors via the Gibbs sampler proceeds as before with an additional step involving the distribution of Φ.

3.2. *VMA(1) errors*

Next suppose that in place of (8) the errors follow an invertible first-order VMA process,

$$\varepsilon_t = u_t + \Theta u_{t-1}, \qquad u_t \overset{\text{iid}}{\sim} \mathcal{N}_p(0, \Omega), \tag{9}$$

where all characteristic roots of Θ: $p \times p$ are inside the unit circle. We begin with the full conditional distributions of β and Ω^{-1}. On the condition that $u_0 = 0$ and under the specification (1)–(3) and (9), the joint density of the observations is

$$f(Y_n \mid \beta, \Omega^{-1}, \Theta) \propto |\Omega^{-1}|^{n/2} \exp\left[-\tfrac{1}{2}\sum_{t=1}^n u_t'\Omega^{-1}u_t\right] \tag{10}$$

$$\propto |\Omega^{-1}|^{n/2} \exp\left[-\tfrac{1}{2}\sum_{t=1}^n (y_t^* - X_t^*\beta)'\Omega^{-1}(y_t^* - X_t^*\beta)\right], \tag{11}$$

where the star variables are defined as

$$y_t^* = \begin{cases} y_t & \text{if} \quad t = 1, \\ y_t - \Theta y_{t-1}^* & \text{if} \quad t \geqslant 2, \end{cases}$$

and

$$X_t^* = \begin{cases} X_t & \text{if} \quad t = 1, \\ X_t - \Theta X_{t-1}^* & \text{if} \quad t \geqslant 2. \end{cases}$$

The definitions of y_t^* and X_t^* thus differ from those in the previous section, and the range of the summation in (11) is from $t = 1$ to n. The full conditional distributions of β and Ω^{-1} are now seen to have the same form as those in (i′) in the previous section.

Simulation of Θ: This is a bit more challenging because the posterior distribution of Θ, which is proportional to the exponential term of (10), cannot be simulated directly. Nevertheless, Θ can be simulated by taking recourse to the Metropolis–Hastings algorithm which, like the Gibbs sampler, is a Markov chain Monte Carlo procedure. To apply this algorithm it is necessary to define a suitable candidate generating density to generate trial draws (see Tierney, 1993). The choice of a candidate generating density is important since it has a bearing on the rejection rate (the percentage of trial draws that are rejected).

A natural way to proceed is to obtain trial draws from a density that approximates the target density. Suppose we let $\bar{\Theta}$ denote the nonlinear least squares estimate of Θ, and then take a first-order Taylor's expansion of u_t around $\bar{\theta} = \text{vec}(\bar{\Theta})$. This gives

$$u_t(\theta) \approx u_t(\bar{\theta}) - W_t'(\theta - \bar{\theta}) \equiv u_t^* - W_t'\theta,$$

where $u_t(\bar{\theta})$ and W_t' can be obtained via the recursions $u_t(\bar{\theta}) = y_t - X_t\beta - \bar{\Theta}u_{t-1}(\bar{\theta})$ and $W_t' = (I_p \otimes u_{t-1}') + \bar{\Theta}'W_{t-1}'$, with $u_0 = 0$ and $W_0 = 0$. Now suppose this is substituted into (10) and the quadratic form in the exponent simplified. Then, we get a density for θ of the type

$$q(\theta \mid Y_n, \Omega) \equiv q(\theta) \propto |V|^{-1/2} \exp[-\tfrac{1}{2}(\theta - \hat{\theta})'V^{-1}(\theta - \hat{\theta})] \times I_V(\Theta), \quad (12)$$

where $\theta = \text{vec}(\Theta')$: $p^2 \times 1$, $\hat{\theta} = V(\sum_{t=1}^n W_t\Omega^{-1}u_t^*)$, $V = (\sum_{t=1}^n W_t\Omega^{-1}W_t')^{-1}$, $W_t = -\partial u_t'/\partial\theta$: $p^2 \times p$ is evaluated at $\bar{\Theta}$, and $I_V(\Theta)$ is the indicator function for the region of invertibility. The density in (12) can be used to generate candidate draws.

The Metropolis–Hastings algorithm is now applied as follows.

Suppose that the draw at the start of the $(i+1)$st iteration is $\theta^{(i)}$. To obtain $\theta^{(i+1)}$ we draw θ' from $q(\cdot)$ and compute

$$\alpha = \min\{w(\theta')/w(\theta^{(i)}), 1\},$$

where $w(\theta) = \pi(\theta \mid Y_n, \beta, \Omega)/q(\theta)$ is the ratio of the target density to the candidate generating density. Then, with probability α we accept the candidate, i.e., we set $\theta^{(i+1)} = \theta'$; otherwise we set $\theta^{(i+1)} = \theta^{(i)}$.

Note that the normalizing constants of (10) and (12) are not required as they cancel in the ratio $w(\theta')/w(\theta^{(i)})$. We have found that this formulation of the Metropolis-within-Gibbs algorithm leads to a well-behaved Markov chain with low rejection rates. In addition, it can be shown, following Chib and Greenberg (1994), that the output of this hybrid algorithm is distributed according to the target posterior density as the number of iterations of the sampler goes to infinity.

4. SUR with time-varying parameters

We now consider another important variant of the SUR model given in (1)–(3). We assume that the regression parameters are time-varying. In particular, the parameter vector β at time t is denoted by $\beta_t = (\beta'_{1t}, \ldots, \beta'_{pt})$, and the basic model is re-specified as

$$y_t = X_t \beta_t + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}_p(0, \Omega).$$

The evolution of the parameters is governed by

$$\beta_t = A_0 \theta_t + u_t, \qquad u_t \sim \mathcal{N}_k(0, B_0), \tag{13}$$

$$\theta_t = \theta_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}_m(0, B_1). \tag{14}$$

Compared to the model in Section 2, we have let $\beta_0 = \theta_t, \mu = \theta_{t-1}$, and $A_1 = I$. The parameter θ_t drifts according to a random walk. Min and Zellner (1993) have recently studied this model under the restrictions that each equation in the SUR model contains the same number of variables, i.e., $k_i = \bar{k}$ for all i, and that

$$\Omega = \sigma^2 I_p; \qquad A_0 = (I_{\bar{k}}, \ldots, I_{\bar{k}})';$$

$$B_0 = (I_p \otimes \Sigma), \quad \Sigma\colon \bar{k} \times \bar{k}; \qquad B_1 = \phi I_m, \tag{15}$$

where Σ is a diagonal matrix and is known. Gamerman and Migon (1993) study a version of this model and focus on the issues of deriving filtering and smoothing recursions under the assumption that the matrices Ω, B_0, and B_1 are known.

To deal with this model, we note that conditioned on $\{\theta_t\}_{t=0}^{n}$ the full conditional distributions of the remaining parameters can be derived rather easily. For example, if θ_t were known, then $\beta_t \sim \mathcal{N}_k(A_0\theta_t, B_0)$ is the prior distribution of β_t and its full conditional distribution is obtained by combining this prior with $f(y_t \mid \beta_t, \Omega) \propto \exp\{-\frac{1}{2}(y_t - X_t\beta_t)'\Omega^{-1}(y_t - X_t\beta_t)\}$. Proceeding in this manner, we obtain the following full conditional distributions for the parameters other than $\{\theta_t\}$:

$$\beta_t \mid Y_n, \Omega^{-1}, \{\theta_t\}, B_0 \sim \mathcal{N}_k(\hat{\beta}_t, V_t), \qquad t \leqslant n,$$

$$\Omega^{-1} \mid Y_n, \{\beta_t\}, \{\theta_t\}, B_0 \sim \mathcal{W}_p(v_0 + n, R_n),$$

$$B_0^{-1} \mid \{\beta_t\}, \{\theta_t\} \sim \mathcal{W}_k\left(\rho_0 + n, \left[D_0^{-1} + \sum_{t=1}^{n} (\beta_t - A_0\theta_t)(\beta_t - A_0\theta_t)'\right]^{-1}\right),$$

$$B_1^{-1} \mid Y_n, \{\theta_t\} \sim \mathcal{W}_m\left(\rho_1 + n, \left[D_1^{-1} + \sum_{t=1}^{n} (\theta_t - \theta_{t-1})(\theta_t - \theta_{t-1})'\right]^{-1}\right),$$

where $\hat{\beta}_t = V_t(B_0^{-1}A_0\theta_t + X_t\Omega^{-1}y_t)$, $R_n = [R_0^{-1} + \sum_{t=1}^{n}(y_t - X_t\beta_t) \times (y_t - X_t\beta_t)']^{-1}$, and $V_t = (B_0^{-1} + X_t'\Omega^{-1}X_t)^{-1}$. Note that if $\Omega^{-1} = \sigma^{-2}I_p$ and $B_0 = (I_p \otimes \Sigma)$, as in (15), then the distribution of σ^{-2} reduces to a Gamma distribution and that of Σ^{-1} (under a diffuse prior) to

$$\mathcal{W}_k\left(n \times p, \left[\sum_{t=1}^{n}\sum_{i=1}^{p} (\beta_{it} - \theta_t)(\beta_{it} - \theta_t)'\right]^{-1}\right).$$

Simulation of $\{\theta_t\}$: This now leaves us with the simulation of $\{\theta_t\}$ given $Y_n, \{\beta_t\}$, and values of the remaining parameters $\psi = (\Omega, B_0, B_1)$. We first note that conditioned on $(\{\beta_t\}, \psi), \{\theta_t\}$ is independent of Y_n. In fact, given $\{\beta_t\}$, we have a state space model for θ_t as can be seen from (13) and (14). Dealing with this state space model requires considerable care as the obvious strategy of simulating θ_t from its full conditional distribution $\theta_t \mid \{\beta_t\}, \psi_{-\theta_t}$ produces very slow convergence to the target distribution. This problem is magnified as the dimension of θ_t increases.

To circumvent this problem, we develop an extremely important result (see also Chib, 1992; Carter and Kohn, 1994) that allows us to simulate $\{\theta_t\}$ from the joint distribution

$$\theta_0, \theta_1, \ldots, \theta_n \mid \{\beta_t\}, \psi. \tag{16}$$

Intuitively, sampling the joint distribution is superior to sampling the collection $\{\theta_t \mid \{\beta_t\}, \psi_{-\theta_t}\}$ because in the former case only one additional block is introduced into the Gibbs sampler instead of the $n+1$ blocks in the latter case.

Despite what may be expected, sampling the joint distribution in (16) is not difficult. To begin with, we write the joint density of the $\{\theta_t\}$ in reverse time order as

$$p(\theta_n \mid \{\beta_t\}, \psi) \times p(\theta_{n-1} \mid \{\beta_t\}, \theta_n, \psi) \times \cdots \times p(\theta_0 \mid \{\beta_t\}, \theta_1, \ldots, \theta_n, \psi), \tag{17}$$

which shows that to obtain a draw from the joint distribution, first draw $\tilde{\theta}_n$ from $\theta_n \mid \{\beta_t\}, \psi$, then draw $\tilde{\theta}_{n-1}$ from $\theta_{n-1} \mid \{\beta_t\}, \tilde{\theta}_n, \psi$, and so on, until $\tilde{\theta}_0$ is drawn from $\theta_0 \mid \{\beta_t\}, \tilde{\theta}_1, \ldots, \tilde{\theta}_n, \psi$. Hence, all that remains to be done is to derive the density of the typical term in (17) which is

$$p(\theta_t \mid \{\beta_t\}, \theta_{t+1}, \ldots, \theta_n, \psi). \tag{18}$$

By way of notation, let $\theta^s = (\theta_s, \ldots, \theta_n)$ and $\beta^s = (\beta_s, \ldots, \beta_n)$ for $s \leqslant n$, and let $\beta_1^s = (\beta_1, \ldots, \beta_s)$. Then, by the Bayes theorem we can write

$$\begin{aligned} p(\theta_t \mid \{\beta_t\}, \theta^{t+1}, \psi) &\propto p(\theta_t \mid \beta_1^t, \psi) p(\theta_{t+1} \mid \beta_1^t, \theta_t, \psi) f(\beta^{t+1}, \theta^{t+2} \mid \beta_1^t, \theta_t, \theta_{t+1}, \psi) \\ &\propto p(\theta_t \mid \beta_1^t, \psi) p(\theta_{t+1} \mid \theta_t, \psi), \end{aligned} \tag{19}$$

from (13) and the observation that $(\beta^{t+1}, \theta^{t+2})$ is independent of θ_t given $(\beta_1^t, \theta_{t+1}, \psi)$.

These two terms can now be simplified by using a well-known result in Kalman filtering (see West and Harrison, 1989) that $(\theta_t \mid \beta_1^t, \psi) \sim \mathcal{N}_k(\hat{\theta}_{t|t}, R_{t|t})$, where $\hat{\theta}_{t|s} \equiv \mathrm{E}(\theta_t \mid \beta_1^s, \psi)$ and $R_{t|s} = \mathrm{cov}(\theta_t \mid \beta_1^s, \theta)$ for $s \leqslant t \leqslant n$ are obtained according to the recursions $\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + K_t(\beta_t - A_0 \hat{\theta}_{t|t-1})$ and $R_{t|t} = (I - K_t A_0) R_{t|t-1}$, with $\hat{\theta}_{t|t-1} = \hat{\theta}_{t-1|t-1}$, $F_{t|t-1} = A_0 R_{t|t-1} A_0' + B_0$, $R_{t|t-1} = R_{t-1|t-1} + B_1$, and $K_t = R_{t|t-1} A_0' F_{t|t-1}^{-1}$.

After the required substitutions in (19) and some manipulations we find that

$$\theta_t \mid \{\beta_t\}, \theta_{t+1}, \ldots, \theta_n, \psi \sim \mathcal{N}_k(\hat{\theta}_t, R_t), \qquad t \leqslant n-1,$$

where $\hat{\theta}_t = \hat{\theta}_{t|t} + M_t(\theta_{t+1} - \hat{\theta}_{t|t})$, $R_t = R_{t|t} - M_t R_{t+1|t} M_t'$, and $M_t = R_{t|t} R_{t+1|t}^{-1}$. At $t = n$, $[\theta_n \mid \{\beta_t\}, \psi]$ is available from the Kalman recursions. Thus, to obtain a sample $\{\theta_t\}$ from the *joint posterior distribution* of the states, given ψ we proceed as follows:

- First, the Kalman filter is run and its output $\{\hat{\theta}_{t|t}, R_t, M_t\}$ saved.

- Second, θ_n is simulated from $\mathcal{N}_k(\hat{\theta}_{n|n}, R_{n|n})$, then θ_{n-1} is simulated from $\mathcal{N}_k(\hat{\theta}_{n-1}, R_{n-1})$, and so on, until θ_0 is sampled from $\mathcal{N}_k(\hat{\theta}_0, R_0)$.

Note that this algorithm can be used to estimate state space models, as has been discussed elsewhere (Chib, 1992; Carter and Kohn, 1994). Also we mention that without this algorithm it would be difficult, if not impossible, to find Bayes estimates in the sort of high-dimensional models that arise in applications. Its value is illustrated in Section 5 where we successfully estimate a TVP-SUR model for five countries ($p = 5$), eight variables ($k = 40$), and $m = 8$ time-varying parameters. As far as we know, there are not other examples of so complex a TVP model being estimated by full Bayesian methods.

5. Examples

Our first example utilizes simulated data for a time series VMA(1) model with three variables. The second and third examples consider an OECD data set.

5.1. A simulated VMA(1) model

We estimate a vector moving average model with $p = 3$ variables. The data are generated by the process

$$y_t = \varepsilon_t + \Theta\varepsilon_{t-1}, \qquad t = 1, \ldots, 100,$$

where

$$\Theta = \begin{pmatrix} 0.6 & 0.2 & 0 \\ 0.1 & 0.6 & 0 \\ 0.2 & 0.2 & 0.6 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} 7 & -3 & 1 \\ -3 & 6 & 0 \\ 1 & 0 & 7 \end{pmatrix}.$$

No x's were specified because the novelty of the example is in the inclusion of a VMA(1) process. In the prior distribution of Ω^{-1}, ν_0 was set equal to 3 and R_0 to I_3. The prior of Θ is diffuse over the invertibility region. The MCMC algorithm generates 4000 values from the posterior distribution. The Metropolis acceptance rate was about 50%, and the serial correlation of the draws was negligible after about lag 5. We show in Fig. 1 the posterior distributions of Ω and Θ. The box plots summarize every fifth value from the run. By comparing the true values with the posterior distribution, it can seen that the parameters are rather tightly concentrated in the correct region of the parameter space.

5.2. Pooled and upooled models for OECD and GNP data

We next apply our methods to a data set that has recently been the subject of considerable analysis. The data are taken from the University of Chicago IMF International Financial Statistics data base; they consist of GNP data for 18 OECD countries. See Garcia-Ferrer, Highfield, Palm, and Zellner (1987) for further details. The SUR model we fit in this illustration, following Min and Zellner (1993), is given by

$$y_{it} = x_{it}\beta_i + \varepsilon_{it}, \qquad i = 1, \ldots, 5, \qquad t = 1960, \ldots, 1987. \tag{20}$$

The five countries in our study are Australia, Canada, Germany, Japan, and the United States. In this model, y_{it} is annual output growth rate for the ith country in the tth year and $x_{it} = (1, y_{it-1}, y_{it-2}, y_{it-3}, SR_{it-1}, SR_{it-2}, GM_{it-1}, MSR_{t-1})'$, where SR_{it} = rate of growth of real stock prices, GM_{it} = rate of growth of real money, and MSR_t = median of SR_{it} in year t. The dimension of β_i is 8 and that of β is 40. Min and Zellner (1993) and others have fit pooled model to this data set. In our framework the pooled model corresponds to a choice of $A_0 = (I_8, \ldots, I_8)'$: 40×8 in the second stage of the hierarchy in (2).

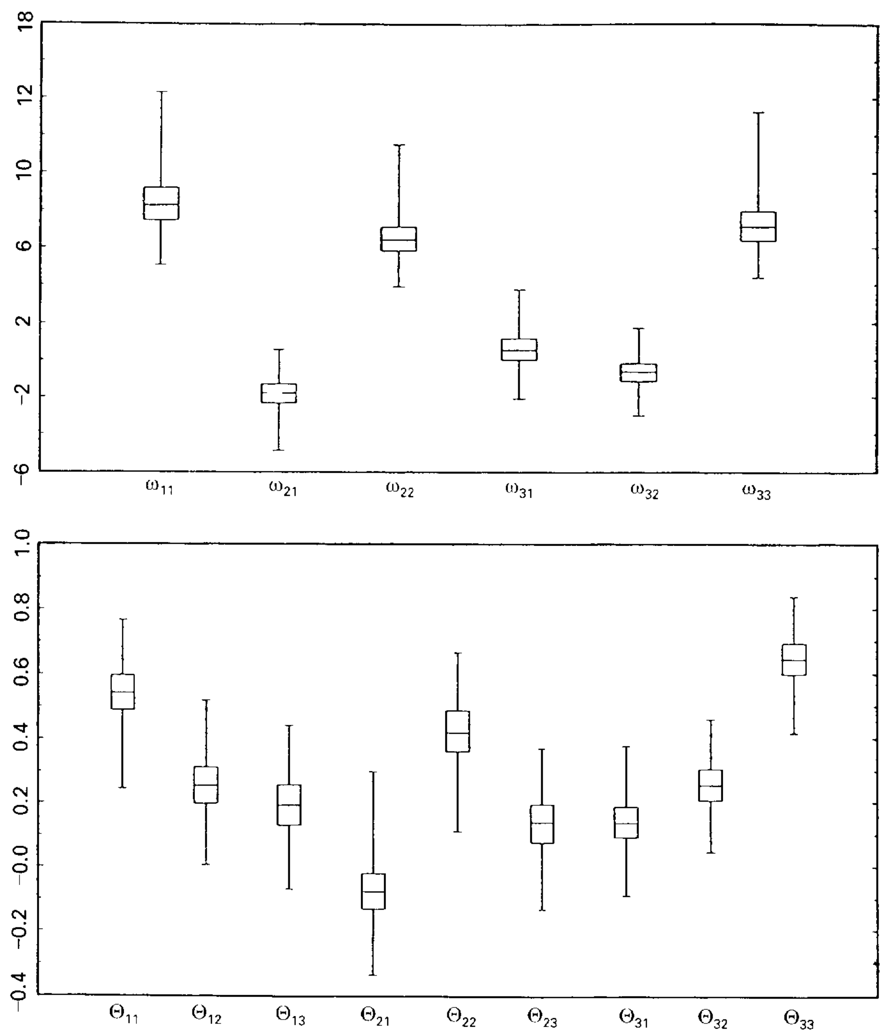

Fig. 1. Posterior pdf's in the simulated trivarite VMA(1) model; top panel: posterior of Ω, bottom panel: posterior of Θ.

As discussed in Section 2.3, to gauge the adequacy of this model we can let $B_0 = \tau^2 I_{40}$, with the prior of τ^2 given by (5). We further specified μ to be the eight-dimensional 0 vector, $B_1 = 10^6 \times I_8, R_0 = 0.2 \times I_5, \nu_{00} = 6$, and $\delta_{00} = 1$. These choices imply that the prior information on Ω is weak, that the fourth stage is degenerate, and that the prior mean and standard deviation or τ^2 are 0.25 and 0.25, respectively. The Gibbs sampler was run to obtain 6000 draws form the posterior distribution after discarding the first 1000 draws. Serial

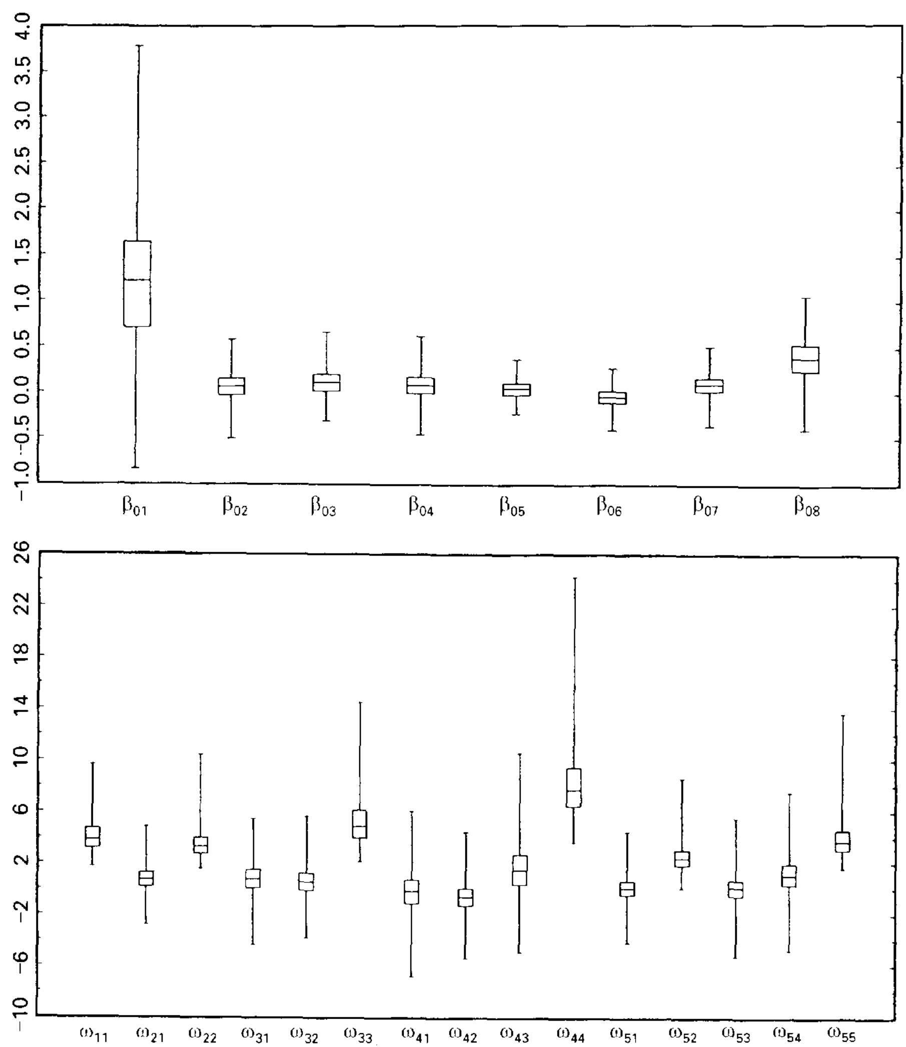

Fig. 2a. Selected posterior pdf's in the pooled SUR model for OECD GNP data; top panel posteiror of β_0, bottom panel: posterior of Ω.

correlation in the output was not significant, and the Gelman and Rubin (1992) convergence checks based on three (shorter) runs of the sampler indicate that this sampling design was adequate.[5]

[5] Additional confirmation was obtained with the Zellner and Min (1983) approach applied to three points in the parameter space of (β, Ω, τ^2).

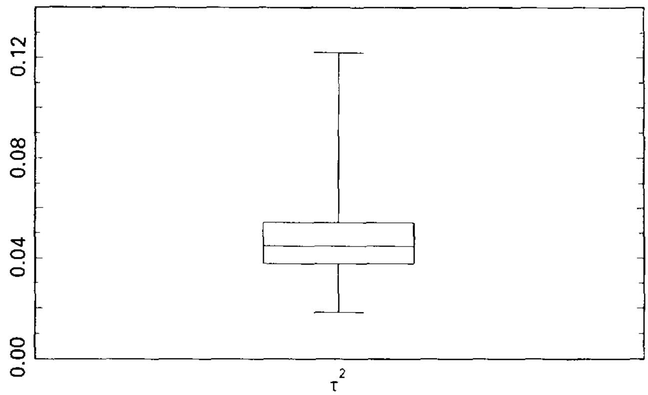

Fig. 2b. Posterior pdf of τ^2 in the pooled SUR model for OECD GNP data.

In Fig. 2 we present posterior distributions of β_0 (the pooled parameter), Ω, and τ^2. As in the previous example, these posteriors are based on every fifth draw from the Monte Carlo output. It should be noted that the posterior of τ^2 is concentrated on small values relative to its prior. In fact, the ratio in (6) for $\varepsilon = 0.05$ is 666.33 and for $\varepsilon = 0.08$ it is 692.12. These values are strong evidence in favor of the pooled model.

Partial Bayes factors for the full and pooled models can be computed by simulation as mentioned in Section 2.3. The first fourteen observations are used to update the prior. The log of the partial Bayes factor in favor of the pooled model is approximately 40, which also provides strong evidence for the pooled model.

5.3. Time-varying SUR with OECD GNP data

Min and Zellner (1993) estimate the TVP-SUR model under the restrictions embodied in (15). They select values of the parameters (Σ, ϕ) to minimize the root-mean-squared error of the one-step-ahead forecasts. In this section we use our full Bayesian approach to fit the pooled model to the specification given in (20) and the restrictions of (15), but we do not impose the restriction that Σ is diagonal.

Since there are a large number of parameters to be simulated, the MCMC sampling algorihtm is run for 10,000 iterations. The iterations are started by specifying values of $\sigma^2 = 1, \phi = 0.5$, and $\Sigma = I_8$. Each of the β_t's is initialized at the posterior mean estimate obtained from the pooled model of the previous section. With these starting values (all priors are diffuse), the simulation process begins by generating $\{\theta_t\}_{t=0}^n$, an eight-dimensional vector for each t. The simulation is continued by generating $\{\beta_t\}_{t=1}^n, \sigma^2, \Sigma$, and ϕ. The first 1000 values from the simulation are discarded.

In Fig. 3 we present the posterior densities for σ^2, ϕ, and the diagonal elements of Σ, and in Fig. 4 we present the posterior distribution of $\theta_5, \theta_{10}, \theta_{15}, \theta_{20}$ (corresponding to the years 1964, 1969, 1974, and 1979). These

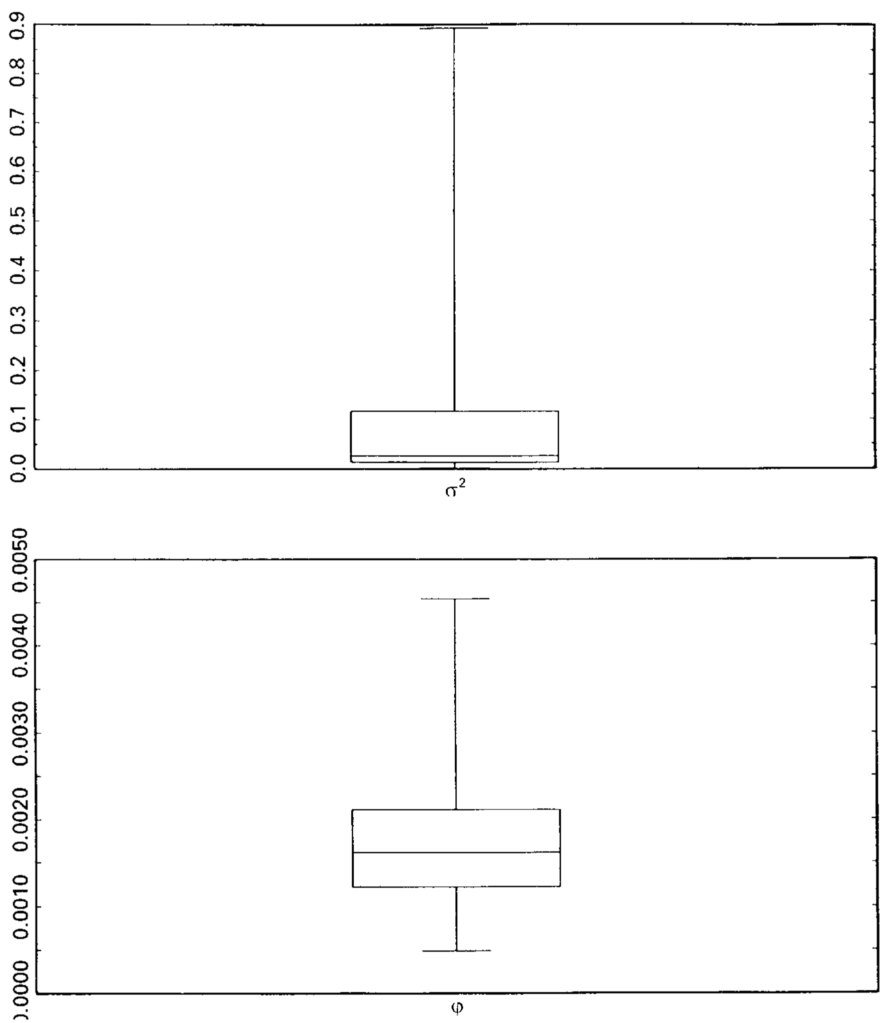

Fig. 3a. Selected posterior pdf's in SUR-TVP model for OECD GNP data; top panel: posterior of σ^2, bottom panel: posterior of ϕ.

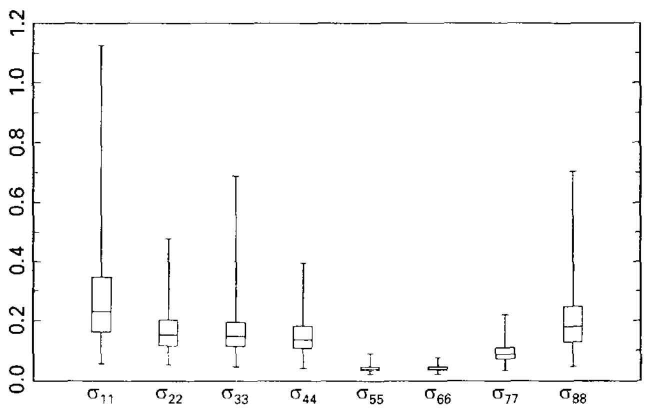

Fig. 3b. Diagonal elements of Σ in SUR-TVP model for OECD GNP data.

posterior densities show the evolution of the θ's and their components over time.

From the results on ϕ (the posterior is concentrated on values close to zero) and the posterior distributions of θ_t (which indicate constancy across time) we conclude that the case for time-varying parameters is weak. This is consistent with the results of Min and Zellner (1993, Table 4a). From the posterior distribution of the elements of Σ we find considerable support for the conclusion that the matrix is diagonal.

6. Conclusions

We have presented methods for a full Bayesian analysis of the SUR model over a more general set of specifications than has previously been examined. In particular, we have analyzed SUR models under hierarchical priors, correlated errors, and time-varying parameters. We have shown that Markov chain Monte Carlo methods provide a unified framework for dealing with these extensions of the SUR model. Since the Bayes estimators are obtained by simulation from standard distributions, the methods can be readily duplicated.

In addition we have discussed alternative Bayesian approaches to the problem of model specification. In the simulation framework it is not difficult to use the hierarchical prior specification to check the adequacy of nested models of interest. Such calculations are illustrated in the paper. We also discuss and illustrate the computation of partial Bayes factors.

All the ideas are illustrated with real and simulated data. The real data example, in particular, illustrates the ease with which high dimensional models can be estimated in the Markov chain Monte Carlo framework.

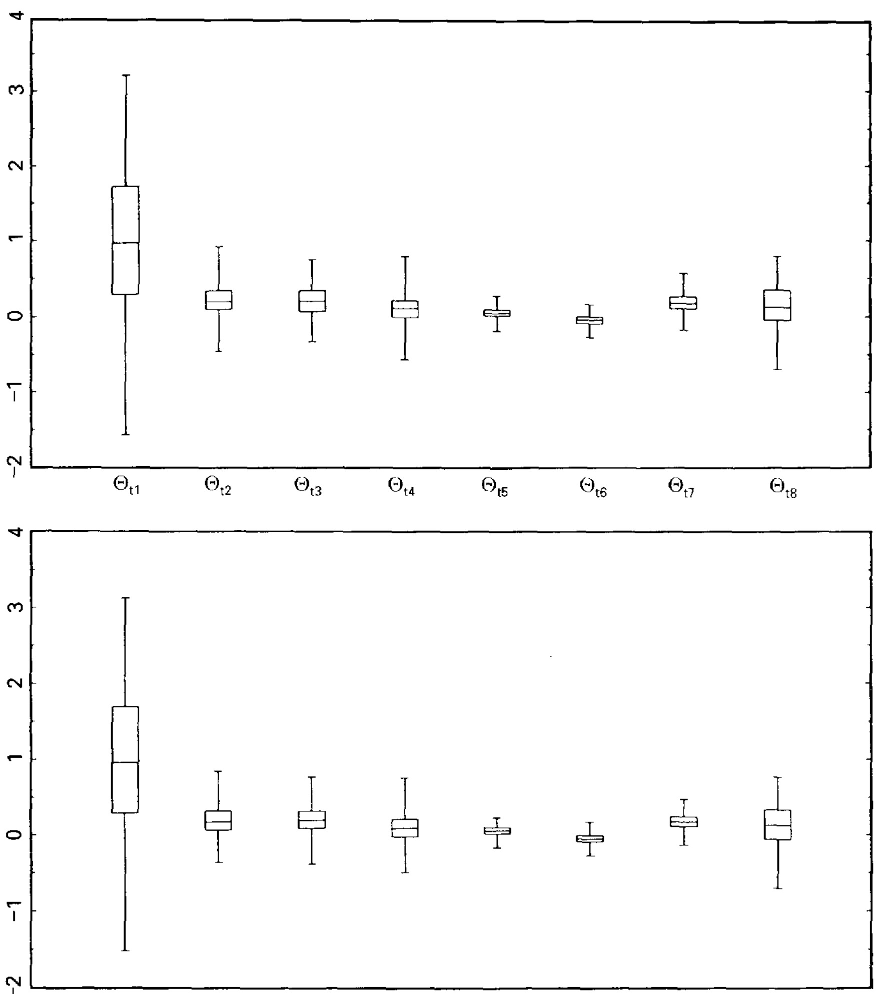

Fig. 4a. Posterior pdf's of θ_t in SUR-TVP model for OECD GNP data; top panel: $t = 1964$, bottom panel: $t = 1969$.

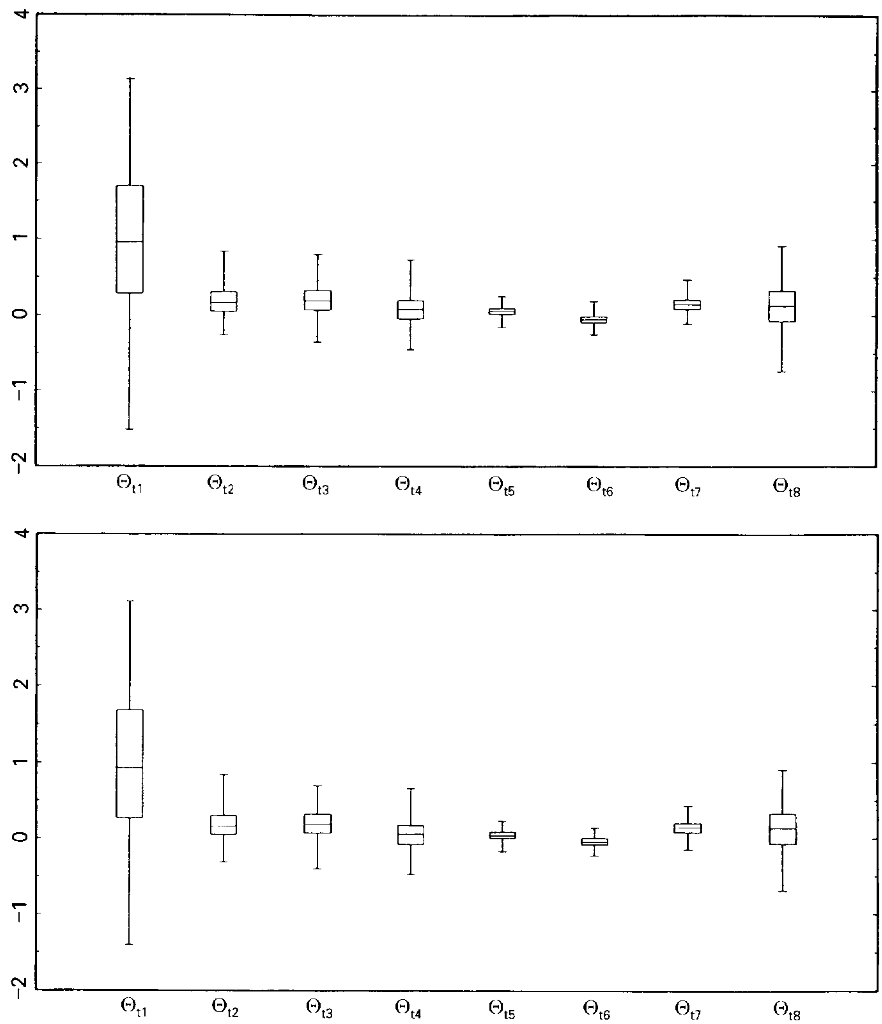

Fig. 4b. Posterior pdf's of θ_t in SUR-TVP model for OECD GNP data; top panel: t = 1974, bottom panel: t = 1979.

References

Albert, J. and S. Chib, 1993a, Bayesian analysis of binary and polychotomous response data, Journal of the American Statistical Association 88, 657–667.

Albert, J. and S. Chib, 1993b, Bayesian model checking for binary and categorical response data, Manuscript.

Albert, J. and S. Chib, 1993c, Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, Journal of Business and Economic Statistics 11, 1–15.

Box, G.E.P. and G.C. Tiao, 1973, Bayesian inference in statistical analysis (Addison-Wesley, Reading, MA).

Carlin, B.P. and S. Chib, 1993, Bayesian model choice via Markov chain Monte Carlo, Journal of the Royal Statistical Society B, forthcoming.

Carter, C. and R. Kohn, 1994, On Gibbs sampling for state space models, Biometrika 81, 541–553.

Chib, S., 1992, An accelerated Gibbs sampler for state space models and other results, Manuscript.

Chib, S., 1993, Bayes regression with autocorrelated errors: A Gibbs sampling approach, Journal of Econometrics 58, 275–294.

Chib, S. and E. Greenberg, 1993, Markov chain Monte Carlo simulation methods in econometrics, Manuscript.

Chib, S. and E. Greenberg, 1994, Bayes inference for regression model with ARMA (p, q) errors, Journal of Econometrics 64, 183–206.

Dawid, A.P., 1981, Some matrix-variate distribution theory: Notational considerations and a Bayesian application, Biometrika 68, 265–274.

Dawid, A.P., 1984, Statistical theory, the prequential approach, Journal of the Royal Statistical Society A 147, 278–292.

Gamerman, D. and H.S. Migon, 1993, Dynamic hierarchical models, Journal of the Royal Statistical Society B 55, 629–642.

Garcia-Ferrer, R., A. Highfield, F.C. Palm, and A. Zellner, 1987, Macroeconomic forecasting using pooled international data, Journal of Business and Economic Statistics 5, 53–67.

Gelfand, A.E. and A.F.M. Smith, 1990, Sampling-based approaches to calculating marginal densities, Journal of the American Statistical Association 85, 398–409.

Gelman, A. and D.B. Rubin, 1992, Inference from iterative simulation using multiple sequences, Statistical Science 4, 457–472.

George, E.I. and R.E. McCulloch, 1993, Variable selection via Gibbs sampling, Journal of the American Statistical Association 88, 881–889.

Goldberger, A.S., 1991, A course in econometrics (Harvard University Press, Cambridge, MA).

Guilkey, D. and P. Schmidt, 1973, Estimation of seemingly unrelated regressions with vector autoregressive errors, Journal of the American Statistical Association 68, 642–647.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lükepohl, and T.-C. Lee, 1985, The theory and practice of econometrics, 2nd ed. (Wiley, New York, NY).

Lindley, D. and A.F.M. Smith, 1972, Bayes estimates for the linear model, Journal of the Royal Statistical Society B 34, 1–41.

Marriott, J., N. Ravishanker, A. Gelfand, and J. Pai, 1995, Bayesian analysis of ARMA processes: Complete sampling based inference under full likelihoods, in: D. Berry, K. Chaloner, and J. Geweke, eds., Bayesian statistics and econometrics: Essays in honor of Arnold Zellner, forthcoming.

Min, C. and A. Zellner, 1993, Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates, Journal of Econometrics 56, 89–118.

O'Hagan, A., 1991, Discussion of 'posterior Bayes factors', Journal of the Royal Statistical Society B 53, 111–142.

Parks, R., 1967, Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated, Journal of the American Statistical Association 62, 500–509.

Percy, D.F., 1992, Prediction for seemingly unrelated regressions, Journal of the Royal Statistical Society B, 243–252.

Press, S.J., 1982, Applied multivariate analysis: Using Bayesian and frequentist methods of inference (Krieger, Malabar, FL).

Ritter, C. and M.A. Tanner, 1992, The Gibbs stopper and the griddy Gibbs sampler, Journal of the American Statistical Association 87, 861–868.

Roberts, G.O. and A.F.M. Smith, 1993, Some convergence theory for Markov chain Monte Carlo, Manuscript.

Spanos, A., 1986, Statistical foundations of econometric modelling (Cambridge University Press, Cambridge).

Srivastava, V.K. and D.E.A. Giles, 1987, Seemingly unrelated regression equations models: Estimation and inference (Marcel Dekker, New York, NY).

Tanner, M.A. and W.H. Wong, 1987, The calculation of posterior distributions by data augmentation, Journal of the American Statistical Association 82, 528–549.

Tierney, L., 1993, Markov chains for exploring posterior distributions, Annals of Statistics, forthcoming.

West, M. and P. Harrison, 1989, Bayesian forecasting and dynamic models (Springer-Verlag, New York, NY).

Zellner, A., 1962, An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias, Journal of the American Statistical Association 57, 500–509.

Zellner, A., 1971, Introduction to Bayesian inference in econometrics (Wiley, New York, NY).

Zellner, A. and C. Min, 1993, Gibbs sampler convergence criteria (GSC^2), Manuscript.