

# **Bayesian Inference in the Seemingly Unrelated Regressions Model**

**William E Griffiths\***

*Economics Department*

*University of Melbourne*

*Vic. 3010*

*Australia*

April 18, 2001

Chapter prepared for *Computer-Aided Econometrics*, edited by David Giles, to be published by Marcel Dekker.

\* I am grateful to Michael Chua for research assistance, and to Denzil Fiebig and Chris O'Donnell for comments on an earlier draft.

## I. INTRODUCTION

Zellner's idea of combining several equations into one model to improve estimation efficiency (Zellner 1962) ranks as one of the most successful and lasting innovations in the history of econometrics. The resulting seemingly unrelated regressions (SUR) model has generated a wealth of both theoretical and empirical contributions. Reviews of work on or involving the SUR model can be found in Srivastava and Dwivedi (1979), Judge et al (1985), Srivastava and Giles (1987) and Fiebig (2001). It was also Zellner (in Zellner 1971) who popularised Bayesian inference in econometrics generally and described the SUR model within the context of Bayesian inference. However, at that time, convenient methods for deriving or estimating marginal posterior density functions and moments for individual SUR coefficients were not generally available. Subsequently, analytical results were derived for some special cases (Drèze and Morales 1976, Richard and Tompa 1980, Richard and Steel 1988, Steel 1992) and importance sampling was suggested as a means for estimating marginal posterior density functions and their moments (Kloek and van Dijk 1978). More recently, the application of Markov Chain Monte Carlo (MCMC) methodology to Bayesian inference has made available a new range of numerical methods that make Bayesian estimation of the SUR model more convenient and accessible. The literature on MCMC is extensive; for a general appreciation of its scope and purpose, see Tierney (1994), Albert and Chib (1996), Chen et al (2000), Chib and Greenberg (1996), Gilks et al (1996), Tanner (1996), and the chapter by Geweke et al (this volume). For application of MCMC to the SUR model, see, for example, Percy (1992, 1996), Chib and Greenberg (1995), Griffiths and Chotikapanich (1997) and Griffiths et al (2000).

The objective of this chapter is to provide a practical guide to computer-aided Bayesian inference for a variety of problems that arise in applications of the SUR model. We describe examples of problems, models and algorithms that have been placed within a general framework in the chapter by Geweke et al (this volume); our chapter can be viewed as complimentary to that chapter. The model is described in Section II; the joint, conditional and marginal posterior density functions that result from a noninformative prior are derived. In Section III we describe how to use sample draws of parameters from their posterior densities to estimate posterior quantities of interest; two Gibbs sampling algorithms and a Metropolis-Hastings algorithm are given. Modifications necessary for nonlinear equations, equality restrictions and inequality restrictions are presented in Sections IV, V and VI, respectively. Three applications are described in Section VII. Section VIII contains methodology for forecasting. Some extensions are briefly mentioned in Section IX and a few concluding remarks are given in Section X.

## II. MODEL SPECIFICATION AND POSTERIORS FROM A NONINFORMATIVE PRIOR

Consider  $M$  equations written as

$$y_i = X_i \beta_i + e_i \quad i = 1, 2, \dots, M \quad (1)$$

where  $y_i$  is a  $T$ -dimensional vector of observations on a dependent variable,  $X_i$  is a  $(T \times K_i)$  matrix of observations on  $K_i$  nonstochastic explanatory variables, possibly including a constant term,  $\beta_i$  is a  $K_i$ -dimensional vector of unknown coefficients that we wish to estimate, and  $e_i$  is a  $T$ -dimensional unobserved random vector. The  $M$  equations can be combined into one big model written as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix} \quad (2)$$

that we then write compactly as

$$y = X\beta + e \quad (3)$$

where  $y$  is of dimension  $(TM \times 1)$ ,  $X$  is of dimension  $(TM \times K)$ , with  $K = \sum_{i=1}^M K_i$ ,  $\beta$  is  $(K \times 1)$  and  $e$  is  $(TM \times 1)$ . We assume the distribution for  $e$  is given by

$$e \sim N(0, \Sigma \otimes I_T) \quad (4)$$

Thus, the errors in each equation are homoskedastic and not autocorrelated. There is, however, contemporaneous correlation between corresponding errors in different equations. The variance of the error of the  $i$ -th equation we denote by  $\sigma_{ii}$ , the  $i$ -th diagonal element of  $\Sigma$ . The covariance between two corresponding errors in different equations (say  $i$  and  $j$ ), we write as  $\sigma_{ij}$ , an off-diagonal element of  $\Sigma$ .

Using  $f(\cdot)$  as generic notation for a probability density function (pdf), the likelihood function for  $\beta$  and  $\Sigma$  can be written as

$$f(y | \beta, \Sigma) = (2\pi)^{-MT/2} |\Sigma|^{-T/2} \exp\{-\frac{1}{2}(y - X\beta)'(\Sigma^{-1} \otimes I_T)(y - X\beta)\} \quad (5)$$

This pdf can also be written as

$$f(y | \beta, \Sigma) = (2\pi)^{-MT/2} |\Sigma|^{-T/2} \exp\{-\frac{1}{2} \text{tr}(A\Sigma^{-1})\} \quad (6)$$

where  $A$  is an  $(M \times M)$  matrix with  $(i, j)$ -th element given by

$$[A]_{ij} = (y_i - X_i\beta_i)'(y_j - X_j\beta_j) \quad (7)$$

Note that  $A$  can also be written as

$$A = (Y - X^*B)'(Y - X^*B) \quad (8)$$

where  $Y$  is the  $(T \times M)$  matrix  $Y = (y_1, y_2, \dots, y_M)$ ,  $X^*$  is the  $(T \times K)$  matrix  $X^* = (X_1, X_2, \dots, X_M)$ , and  $B$  is the  $(K \times M)$  matrix

$$B = \begin{bmatrix} \beta_1 & & & \\ & \beta_2 & & \\ & & \ddots & \\ & & & \beta_M \end{bmatrix} \quad (9)$$

Result (9) on page 42 of Lütkepohl (1996) can be used to establish the equivalence of equations (5) and (6). Specifically,

$$\begin{aligned} \text{tr}[(Y - X^*B)'(Y - X^*B)\Sigma^{-1}] &= [\text{vec}(Y - X^*B)]'[\Sigma^{-1} \otimes I_T] \text{vec}(Y - X^*B) \\ &= (y - X\beta)'(\Sigma^{-1} \otimes I_T)(y - X\beta) \end{aligned} \quad (10)$$

Two prior pdfs will be considered in this chapter; they are the conventional noninformative prior (see, for example, Zellner 1971, ch.8)

$$f(\beta, \Sigma) = f(\beta)f(\Sigma) \propto |\Sigma|^{-(M+1)/2} \quad (11)$$

and another prior that imposes inequality restrictions on  $\beta$ , but is otherwise noninformative. The inequality prior and its consequences will be considered later in the chapter. The noninformative prior in (11) is chosen to provide objectivity in reporting, not because we believe total ignorance is prevalent. Geweke et al (chapter in this volume) discuss how to modify results to accommodate the prior of a specific client.

#### **A. Joint Posterior pdf for $(\beta, \Sigma)$**

Applying Bayes' theorem to the prior pdf in (11) and the likelihood function in (5) and (6) yields the joint posterior pdf for  $\beta$  and  $\Sigma$

$$\begin{aligned}
f(\boldsymbol{\beta}, \boldsymbol{\Sigma} | y) &\propto f(y | \boldsymbol{\beta}, \boldsymbol{\Sigma}) f(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \\
&\propto |\boldsymbol{\Sigma}|^{-(T+M+1)/2} \exp\left\{-\frac{1}{2}(y - X\boldsymbol{\beta})'(\boldsymbol{\Sigma}^{-1} \otimes I_T)(y - X\boldsymbol{\beta})\right\} \\
&= |\boldsymbol{\Sigma}|^{-(T+M+1)/2} \exp\left\{-\frac{1}{2}\text{tr}(A\boldsymbol{\Sigma}^{-1})\right\}
\end{aligned} \tag{12}$$

In the remainder of this section we describe a number of marginal and conditional posterior pdf's that are derived from equation (12). These pdf's will prove useful in later sections when we discuss methods for estimating quantities of interest. We will assume that interest centers on individual coefficients, say the  $k$ -th coefficient in the  $i$ -th equation  $\beta_{ik}$ , and, more generally, on some functions of the coefficients, say  $g(\boldsymbol{\beta})$ . Forecasting future values  $y^*$  will also be considered. The relevant pdf's that express our uncertain post-sample knowledge about these quantities are the marginal pdf's  $f(\beta_{ik} | y)$ ,  $f(g(\boldsymbol{\beta}) | y)$  and  $f(y^* | y)$ , respectively. Typically, we report results by graphing these pdf's, and tabulating their means, standard deviations and probabilities for regions of interest. Describing the tools for doing so is the major focus of this chapter.

## B. Conditional Posterior pdf for $(\boldsymbol{\beta} | \boldsymbol{\Sigma})$

The term in the exponent of equation (12) can be written as

$$\begin{aligned}
(y - X\boldsymbol{\beta})'(\boldsymbol{\Sigma}^{-1} \otimes I_T)(y - X\boldsymbol{\beta}) &= (y - X\hat{\boldsymbol{\beta}})'(\boldsymbol{\Sigma}^{-1} \otimes I_T)(y - X\hat{\boldsymbol{\beta}}) \\
&\quad + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X'(\boldsymbol{\Sigma}^{-1} \otimes I_T) X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})
\end{aligned} \tag{13}$$

where  $\hat{\boldsymbol{\beta}} = [X'(\boldsymbol{\Sigma}^{-1} \otimes I_T)X]^{-1} X'(\boldsymbol{\Sigma}^{-1} \otimes I_T)y$ . It follows that the conditional posterior pdf for  $\boldsymbol{\beta}$  given  $\boldsymbol{\Sigma}$  is the multivariate normal pdf

$$f(\boldsymbol{\beta} | \boldsymbol{\Sigma}, y) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X'(\boldsymbol{\Sigma}^{-1} \otimes I_T) X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \tag{14}$$

with posterior mean equal to the generalised least squares estimator

$$E(\beta | y, \Sigma) = \hat{\beta} = [X'(\Sigma^{-1} \otimes I_T)X]^{-1} X'(\Sigma^{-1} \otimes I_T)y \quad (15)$$

and posterior covariance matrix equal to

$$V(\beta | y, \Sigma) = [X'(\Sigma^{-1} \otimes I_T)X]^{-1} \quad (16)$$

The last two expressions are familiar ones in sampling theory inference for the SUR model. They show that the traditional SUR estimator, written as

$$\hat{\beta} = [X'(\hat{\Sigma}^{-1} \otimes I_T)X]^{-1} X'(\hat{\Sigma}^{-1} \otimes I_T)y \quad (17)$$

where  $\hat{\Sigma}$  is a 2-step estimator or a maximum likelihood estimator, can be viewed as the mean of the conditional posterior pdf for  $\beta$  given  $\hat{\Sigma}$ . The traditional covariance matrix estimator  $[X'(\hat{\Sigma}^{-1} \otimes I_T)X]^{-1}$  can be viewed as the conditional covariance matrix from the same pdf. Since this pdf does not take into account uncertainty from not knowing  $\Sigma$  (the fact that  $\hat{\Sigma}$  is an estimate is not recognised), it overstates the reliability of our information about  $\beta$ . This dilemma was noted by Fiebig and Kim (2000) in the context of an increasing number of equations.

### C. Marginal Posterior pdf for $\beta$

The more appropriate representation of our uncertainty about  $\beta$  is the marginal posterior pdf  $f(\beta | y)$ . It can be shown that this pdf is given by

$$\begin{aligned} f(\beta | y) &= \int f(\beta, \Sigma | y) d\Sigma \\ &\propto |A|^{-T/2} \end{aligned} \quad (18)$$

The integral in (18) is performed by using properties of the inverted Wishart distribution (see, for example, Zellner 1971, p.395). For  $f(\beta | y) \propto |A|^{-T/2}$  to be

proper, we require  $T \geq M + \text{rank}(X^*)$  (Griffiths et al 2001). Also, this pdf is not of a standard recognisable form. Except for special cases, analytical expressions for its normalising constant and moments are not available. Estimating these moments, and marginal pdf's for individual coefficients  $\beta_{ik}$ , is considered in the next section; first, we describe some more pdf's that will prove to be useful.

#### D. Conditional Posterior pdf for $(\beta_1 | \beta_2, \dots, \beta_M)$

It is possible to show that the posterior pdf for the coefficient vector from one equation, conditional on those from other equations, is a multivariate  $t$ -distribution. To derive this result, we will consider the posterior pdf for  $\beta_1$ , conditional on  $(\beta_2, \beta_3, \dots, \beta_M)$ . We write a partition of  $(Y - X^*B)$  into its first and remaining  $(M - 1)$  columns as

$$Y - X^*B = \begin{pmatrix} y_1 - X_1\beta_1 & E_{(1)} \end{pmatrix}$$

The corresponding partition of  $A$  is

$$A = \begin{bmatrix} (y_1 - X_1\beta_1)'(y_1 - X_1\beta_1) & (y_1 - X_1\beta_1)'E_{(1)} \\ E_{(1)}'(y_1 - X_1\beta_1) & E_{(1)}'E_{(1)} \end{bmatrix}$$

Using a result on the determinant of a partitioned matrix, we have

$$|A| = |E_{(1)}'E_{(1)}| \left( (y_1 - X_1\beta_1)'(y_1 - X_1\beta_1) - (y_1 - X_1\beta_1)'E_{(1)}(E_{(1)}'E_{(1)})^{-1}E_{(1)}'(y_1 - X_1\beta_1) \right)$$

Defining  $Q_{(1)} = I_T - E_{(1)}(E_{(1)}'E_{(1)})^{-1}E_{(1)}'$ , and  $\tilde{\beta}_1 = (X_1'Q_{(1)}X_1)^{-1}X_1'Q_{(1)}y_1$  the second term in the above equation can be written as

$$(y_1 - X_1\beta_1)'Q_{(1)}(y_1 - X_1\beta_1) = (y_1 - X_1\tilde{\beta}_1)'Q_{(1)}(y_1 - X_1\tilde{\beta}_1) + (\beta_1 - \tilde{\beta}_1)'X_1'Q_{(1)}X_1(\beta_1 - \tilde{\beta}_1)$$



Collecting all these results, substituting into equation (18), and letting  $|E_{(1)}'E_{(1)}|$  be absorbed into the proportionality constant, we can write

$$f(\beta_1 | y, \beta_2, \beta_3, \dots, \beta_M) \propto \left[ v_1 + \frac{(\beta_1 - \tilde{\beta}_1)' X_1' Q_{(1)} X_1 (\beta_1 - \tilde{\beta}_1)}{\tilde{s}_1^2} \right]^{-(K_1 + v_1)/2} \quad (19)$$

where  $v_1 = T - K_1$  and  $\tilde{s}_1^2 = (y_1 - X_1 \tilde{\beta}_1)' Q_{(1)} (y_1 - X_1 \tilde{\beta}_1) / v_1$ . Equation (19) is in the form of a multivariate  $t$ -distribution with degrees of freedom  $v_1$ , mean  $\tilde{\beta}_1$ , and covariance matrix  $(v_1 / (v_1 - 2)) \tilde{s}_1^2 (X_1' Q_{(1)} X_1)^{-1}$ . See, for example, Zellner (1971, p.383). The conditional posterior pdf's for other  $\beta_i$  are similarly defined.

### E. Conditional Posterior pdf for $(\Sigma | \beta)$

Viewing the joint posterior pdf in equation (12) as a function of only  $\Sigma$  yields the conditional posterior pdf for  $\Sigma$  given  $\beta$ . It is the inverted Wishart pdf (see, for example Zellner 1971, p.395)

$$f(\Sigma | \beta, y) \propto |\Sigma|^{-(T+M+1)/2} \exp\{-\frac{1}{2} \text{tr}(A\Sigma^{-1})\} \quad (20)$$

It has  $T$  degrees of freedom, and parameter matrix  $A$ .

### F. Marginal Posterior pdf for $\Sigma$

The marginal pdf for  $\Sigma$ , obtained by using the result in (13), and then using properties of the multivariate normal pdf to integrate out  $\beta$ , is given by

$$\begin{aligned} f(\Sigma | y) &= \int f(\beta, \Sigma | y) d\beta \\ &\propto \left| X'(\Sigma^{-1} \otimes I_T) X \right|^{-1/2} |\Sigma|^{-(T+M+1)/2} \exp\{-\frac{1}{2} (y - X\hat{\beta})' (\Sigma^{-1} \otimes I_T) (y - X\hat{\beta})\} \\ &= \left| X'(\Sigma^{-1} \otimes I_T) X \right|^{-1/2} |\Sigma|^{-(T+M+1)/2} \exp\{-\frac{1}{2} \text{tr}(\hat{A}\Sigma^{-1})\} \end{aligned} \quad (21)$$

where  $\hat{A}$  is an  $(M \times M)$  matrix with  $(i, j)$ -th element given by

$$[\hat{A}]_{ij} = (y_i - X_i \hat{\beta}_i)' (y_j - X_j \hat{\beta}_j)$$

The posterior pdf in (21) is not an analytically tractable one whose moments are known. However, as we will see, we can draw observations from it using the Gibb's sampler.

### III. ESTIMATING POSTERIOR QUANTITIES

Given the intractability of the posterior pdf  $f(\beta | y) \propto |A|^{-T/2}$ , methods for estimating marginal posterior pdf's for individual coefficients  $\beta_{ik}$ , their moments, and probabilities of interest, are required. Suppose that we have draws  $\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(N)}$  taken from  $f(\beta | y)$  and, possibly, draws  $\Sigma^{(1)}, \Sigma^{(2)}, \dots, \Sigma^{(N)}$  taken from  $f(\Sigma | y)$ . We will describe a number of ways one can proceed to estimate the desired quantities; then, we discuss how the required posterior draws can be obtained.

#### A. Estimating Posterior pdf's

A simple way to estimate the marginal posterior pdf of  $\beta_{ik}$ , say, is to construct a histogram of draws of that parameter. Joining the mid points of the histogram classes provides a continuous representation of the pdf, but, typically, it will be a jagged one unless some kind of smoothing procedure is employed. Alternatively, one can obtain a smooth pdf, and a more efficient estimate, by averaging conditional posterior pdf's for the quantity of interest. In this case, for conditional posterior pdf's one can use the  $t$ -distributions defined by (19), or, if draws on both  $\beta$  and  $\Sigma$  are available, the normal distributions defined by (14).

Considering the  $t$ -distribution first, an estimate of  $f(\beta_{ik} | y)$  is given by

$$\begin{aligned}
\hat{f}(\beta_{ik} | y) &= \frac{1}{N} \sum_{\ell=1}^N f(\beta_{ik} | y, \beta_1^{(\ell)}, \dots, \beta_{i-1}^{(\ell)}, \beta_{i+1}^{(\ell)}, \dots, \beta_M^{(\ell)}) \\
&= c \frac{1}{N} \sum_{\ell=1}^N \frac{1}{\sqrt{\tilde{s}_i^{2(\ell)} q_{(i)kk}^{(\ell)}}} \left[ v_i + \frac{(\beta_{ik} - \tilde{\beta}_{ik}^{(\ell)})^2}{\tilde{s}_i^{2(\ell)} q_{(i)kk}^{(\ell)}} \right]^{-(v_i+1)/2} \quad (22)
\end{aligned}$$

The univariate  $t$ -distribution that is being averaged in equation (22) is the conditional pdf for a single coefficient from  $\beta_i$ , obtained from the multivariate  $t$ -distribution in (19), after generalising from  $\beta_1$  to  $\beta_i$ . The previously undefined terms in (22) are the constant

$$c = \frac{\Gamma[(v+1)/2] v^{v/2}}{\Gamma(v/2) \sqrt{\pi}}$$

where  $\Gamma(\cdot)$  is the gamma function, the conditional posterior mean  $\tilde{\beta}_{ik}$  which is the  $k$ -th element in  $\tilde{\beta}_i$ , and  $q_{(i)kk}$  that is the  $k$ -th diagonal element of  $(X_i' Q_{(i)} X_i)^{-1}$ . To plot the pdf in (22), we choose a grid of values for  $\beta_{ik}$  (50-100 is usually adequate), and for each value of  $\beta_{ik}$ , we compute the average in (22). These averages are plotted against the  $\beta_{ik}$ .

Alternatively, the conditional normal distributions in (14) can be averaged over  $\Sigma$ . In this case an estimate of the marginal posterior pdf for  $\beta_{ik}$  is given by

$$\begin{aligned}
\hat{f}(\beta_{ik} | y) &= \frac{1}{N} \sum_{\ell=1}^N f(\beta_{ik} | y, \Sigma^{(\ell)}) \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{N} \sum_{\ell=1}^N \frac{1}{\sqrt{h_{(i)kk}^{(\ell)}}} \exp \left\{ -\frac{1}{2h_{(i)kk}^{(\ell)}} (\beta_{ik} - \hat{\beta}_{ik}^{(\ell)})^2 \right\} \quad (23)
\end{aligned}$$

where  $\hat{\beta}_{ik}$  is the  $k$ -element in the  $i$ -th vector component of  $\hat{\beta}$  (see equation(15)), and  $h_{(i)kk}$  is the  $k$ -diagonal element in the  $i$ -th diagonal block of  $[X'(\Sigma^{-1} \otimes I_T)X]^{-1}$  (see

equation (16)). Like in (22), the average in (23) is computed for, and plotted against, a grid of values for  $\beta_{ik}$ .

## B. Estimating Posterior Means and Standard Deviations

Corresponding to the three ways given for estimating posterior pdf's, there are three ways of estimating their posterior means and variances. The first way is to use the sample mean and covariance matrix of the draws. That is,

$$\hat{E}(\beta | y) = \frac{1}{N} \sum_{\ell=1}^N \beta^{(\ell)} = \bar{\beta} \quad (24)$$

and

$$\hat{V}(\beta | y) = \frac{1}{N-1} \sum_{\ell=1}^N (\beta^{(\ell)} - \bar{\beta})(\beta^{(\ell)} - \bar{\beta})' \quad (25)$$

The second and third approaches use the results (1) an unconditional mean is equal to the mean of the conditional means, and (2) the unconditional variance is equal to the mean of the conditional variances plus the variance of the conditional means. Applying these two results to the conditional posterior pdf in (19) yields

$$\hat{E}(\beta_i | y) = \frac{1}{N} \sum_{\ell=1}^N \tilde{\beta}_i^{(\ell)} = \frac{1}{N} \sum_{\ell=1}^N (X_i' Q_{(i)}^{(\ell)} X_i)^{-1} X_i' Q_{(i)}^{(\ell)} y_i = \bar{\tilde{\beta}}_i \quad (26)$$

and

$$\hat{V}(\beta_i | y) = \left( \frac{v_i}{v_i - 2} \right) \frac{1}{N} \sum_{\ell=1}^N \tilde{s}_i^{2(\ell)} (X_i' Q_{(i)}^{(\ell)} X_i)^{-1} + \frac{1}{N-1} \sum_{\ell=1}^N (\tilde{\beta}_i^{(\ell)} - \bar{\tilde{\beta}}_i)(\tilde{\beta}_i^{(\ell)} - \bar{\tilde{\beta}}_i)' \quad (27)$$

Applying the two results to the normal conditional posterior pdf's in (14) yields

$$\hat{E}(\beta | y) = \frac{1}{N} \sum_{\ell=1}^N \hat{\beta}^{(\ell)} = \frac{1}{N} \sum_{\ell=1}^N [X'(\Sigma^{-1(\ell)} \otimes I_T)X]^{-1} X'(\Sigma^{-1(\ell)} \otimes I_T)y = \bar{\beta} \quad (28)$$

and

$$\hat{V}(\beta | y) = \frac{1}{N} \sum_{\ell=1}^N [X'(\Sigma^{-1(\ell)} \otimes I_T)X]^{-1} + \frac{1}{N-1} \sum_{\ell=1}^N (\hat{\beta}^{(\ell)} - \bar{\beta})(\hat{\beta}^{(\ell)} - \bar{\beta})' \quad (29)$$

Clearly, using the sample means and standard deviations from equations (24) and (25) is much easier than using the conditional quantities in equations (26) through (29). However, averaging conditional moments generally leads to more efficient estimates.

### C. Estimating Probabilities

Often, we are interested in reporting the probability that  $\beta_{ik}$  lies with a particular interval or finding an interval with a pre-specified probability content. In sampling theory inference intervals with 95% probability content are popular. An estimate of the probability that  $\beta_{ik}$  lies in a particular interval is given by the proportion of draws that lie within that interval. Alternatively, one can find conditional probabilities and average them, along the lines that the conditional means are averaged in equations (26) and (28). Using the conditional normal distribution as an example, we can estimate the probability that  $\beta_{ik}$  lies in the interval  $(a, b)$  as

$$\hat{P}(a < \beta_{ik} < b) = \frac{1}{N} \sum_{\ell=1}^N P(a < (\beta_{ik} | \Sigma^{(\ell)}) < b) \quad (30)$$

Order statistics can be used to obtain an interval with a prespecified probability content. For example, for a 95% probability interval for  $\beta_{ik}$ , we can take the 0.025 and 0.975 empirical quantiles of the draws of the  $\beta_{ik}$ .

#### D. Functions of $\beta$

How do we proceed if we are interested in some functions of  $\beta$ , say  $g(\beta)$ ? Examples of such functions considered later in this chapter are monotonicity and curvature conditions in a cost system, and the relative magnitudes of equivalence scales in household expenditure functions. Examples outside the context of SUR models are the evaluation of consumer surplus (Griffiths 1999) and the stationary region in a time series model (Geweke 1988).

If it is possible to derive, analytically, conditional distributions of the form  $f(g(\beta_1) | y, \beta_2, \dots, \beta_M)$  or  $f(g(\beta) | y, \Sigma)$ , then one can work with these conditional distributions along the lines described above. However, the ability to proceed analytically is rare, given that  $g(\beta)$  is frequently nonlinear and of lower dimension than  $\beta$ . Instead, we can compute values  $g(\beta^{(\ell)})$ ,  $\ell = 1, 2, \dots, N$ , from the draws of  $\beta$ . These values can be placed in a histogram to estimate the pdf of  $g(\beta)$ . Their sample mean and variance can be used to estimate the corresponding posterior mean and variance. Probabilities can be estimated using the proportion of values in a given region and order statistics can be used to find an interval with a given probability content.

#### E. Gibbs Sampling with $\beta$ and $\Sigma$

We now turn to the question of how to obtain draws  $\beta$  and  $\Sigma$  from their respective marginal posterior pdf's. One possible way is to use an MCMC algorithm known as Gibbs sampling. In this procedure draws are made iteratively from the conditional posterior pdf's. Specifically, given a particular starting value for  $\Sigma$ , say  $\Sigma^{(0)}$ , the  $\ell$ -th draw from the Gibbs sampler  $(\beta^{(\ell)}, \Sigma^{(\ell)})$  is obtained using the following two steps:

1. Draw  $\beta^{(\ell)}$  from  $f(\beta | \Sigma^{(\ell-1)}, y)$ .
2. Draw  $\Sigma^{(\ell)}$  from  $f(\Sigma | \beta^{(\ell)}, y)$ .

Making these draws is straightforward, given that the two conditional posterior pdf's are normal and inverted Wishart, respectively. (See the appendix for details.) MCMC theory suggests that, after a sufficiently large number of draws, the Markov Chain created by the draws will converge. After convergence, the subsequent draws can be viewed as draws from the marginal posterior pdf's  $f(\beta | y)$  and  $f(\Sigma | y)$ . It is these draws that can be used to present results in the desired fashion. Draws taken prior to the point at which convergence is assumed to have taken place are sometimes called the "burn in"; they are discarded. A large number of diagnostics have been suggested for assessing whether convergence has taken place. See, for example, Cowles and Carlin (1996). Assessing whether convergence has taken place is similar to assessing whether a time series is stationary. Thus, visual inspection of a graph of the sequence of draws, testing whether the mean and variance are the same at the beginning of the chain as at the end of the chain, and testing whether two or more separately-run chains have the same mean and variance, are ways of checking for convergence.

Since we are using a sample of draws of  $\beta$  and  $\Sigma$  to estimate posterior means and standard deviations and other relevant population quantities, the accuracy of the estimates is a concern. Estimation accuracy is assessed using numerical standard errors. Methods for computing such standard errors are described in the chapter by Geweke et al. Because the draws produced by MCMC algorithms are correlated, time series methods are used to compute the standard errors; also, larger samples are required to achieve a given level of accuracy relative to a situation involving independent draws.

Although the above remarks on convergence and numerical standard errors were made in the context of the Gibbs sampler for  $\beta$  and  $\Sigma$ , they also apply to other MCMC algorithms including the Gibbs sampler for  $\beta$  and the Metropolis Hastings algorithm described below.

## F. Gibbs Sampling with $\beta$

If the number of equations is large, making  $\Sigma$  of high dimension, then it may be preferable to use a Gibbs sampler based on the conditional posterior pdfs for the  $\beta_i$  from each equation. Note, however, that this alternative is not feasible if cross-equation restrictions on the  $\beta_i$ , as discussed in Sections V and VII, are present.

To proceed with this Gibbs sampler, we begin with starting values for all coefficients except the first, say  $(\beta_2^{(0)}, \beta_3^{(0)}, \dots, \beta_M^{(0)})$  and then sample iteratively using the following steps for the  $\ell$ -th draw:

1. Draw  $\beta_1^{(\ell)}$  from  $f(\beta_1 | \beta_2^{(\ell-1)}, \dots, \beta_M^{(\ell-1)})$ .
2. Draw  $\beta_2^{(\ell)}$  from  $f(\beta_2 | \beta_1^{(\ell)}, \beta_3^{(\ell-1)}, \dots, \beta_M^{(\ell-1)})$ .
- $\vdots$
- i. Draw  $\beta_i^{(\ell)}$  from  $f(\beta_i | \beta_1^{(\ell)}, \dots, \beta_{i-1}^{(\ell)}, \beta_{i+1}^{(\ell-1)}, \dots, \beta_M^{(\ell-1)})$ .
- $\vdots$
- M. Draw  $\beta_M^{(\ell)}$  from  $f(\beta_M | \beta_1^{(\ell)}, \dots, \beta_{M-1}^{(\ell)})$ .

The conditional posterior pdfs are multivariate  $t$ -distributions from which we can readily draw values (see the Appendix). Ordinary least squares estimates are adequate for starting values.



## G. A Metropolis-Hastings Algorithm

An alternative to Gibbs sampling is a Metropolis-Hastings algorithm that draws observations from the marginal posterior pdf  $f(\beta | y)$ . As we will see, this algorithm is particularly useful for an inequality-restricted prior, or if the equations are nonlinear. The algorithm we describe is a random-walk algorithm; it is just one of many possibilities. For others see, for example, Chen et al (2000).

The Metropolis–Hastings algorithm generates a candidate value  $\beta^*$  that is accepted or rejected as a draw from the posterior pdf  $f(\beta | y)$ . When it is rejected, the previously accepted draw is repeated as a draw. Thus, rules are needed for generating the candidate value  $\beta^*$  and for accepting it. Let  $V$  be the covariance matrix for the distribution used to generate a candidate value. The maximum likelihood covariance matrix is usually suitable. For the linear SUR model this matrix is  $[X'(\hat{\Sigma}^{-1} \otimes I_T)X]^{-1}$ . Choose a feasible starting value  $\beta^{(0)}$ . The following steps can be used to draw the  $(\ell + 1)$ –th observation in a random walk Metropolis–Hastings chain.

1. Draw a candidate value  $\beta^*$  from a  $N(\beta^{(\ell)}, cV)$  distribution where  $c$  is a scalar set such that  $\beta^*$  is accepted approximately 40-50% of the time.
2. Compute the ratio of the posterior pdf evaluated at the candidate draw to the posterior pdf evaluated at the previously accepted draw.

$$r = \frac{f(\beta^* | y)}{f(\beta^{(\ell)} | y)}$$

Note that this ratio can be computed without knowledge of the normalising constant for  $f(\beta | y)$ . Also, if any of the elements of  $\beta^*$  fall outside a feasible parameter region defined by an inequality-restricted prior (see

Section V), then  $f(\beta^* | y) = 0$ . When  $r > 1$ ,  $\beta^*$  is a more likely value than  $\beta^{(\ell)}$  in the sense that it is closer to the mode of the distribution. When  $r < 1$ ,  $\beta^*$  is further into the tails of the distribution. If  $r > 1$ ,  $\beta^*$  is accepted; if  $r < 1$ ,  $\beta^*$  is accepted with probability  $r$ . Thus, more draws occur in regions of high probability and fewer draws occur in regions of low probability. Details of the acceptance-rejection procedure follow in step 3.

3. Draw a value  $u$  for a uniform random variable on the interval (0,1).

If  $u \leq r$ , set  $\beta^{(\ell+1)} = \beta^*$ .

If  $u > r$ , set  $\beta^{(\ell+1)} = \beta^{(\ell)}$ .

Return to step 1 with  $\ell$  set to  $\ell + 1$ .

Let  $q(\beta^* | \beta^{(\ell)})$  be the distribution used to generate the candidate value  $\beta^*$  in step 1. In our case it is a normal distribution. In more general Metropolis-Hastings algorithms, where our choice of distribution is not necessarily utilized,  $r$  is defined as

$$r = \frac{f(\beta^* | y) \cdot q(\beta^{(\ell)} | \beta^*)}{f(\beta^{(\ell)} | y) \cdot q(\beta^* | \beta^{(\ell)})}$$

In our case  $q(\beta^{(\ell)} | \beta^*) = q(\beta^* | \beta^{(\ell)})$ . Various alternatives for  $q(\cdot)$  have been suggested in the literature.

#### IV. NONLINEAR SUR

Many economic models are intrinsically nonlinear, or a nonlinear model may result from substituting nonlinear restrictions on  $\beta$  into a linear model. The Gibbs sampling

algorithms that we described are no longer applicable for a nonlinear SUR model. However, we can still proceed with the Metropolis-Hastings algorithm.

Suppose that the nonlinear SUR model is given by

$$y_i = h_i(X, \beta) + e_i \quad i = 1, 2, \dots, M \quad (31)$$

where the  $h_i$  are the nonlinear functions. The dimensions of  $y_i$ ,  $h_i$  and  $e_i$  are  $(T \times 1)$ .

In this context  $X$  represents a set of explanatory variables and  $\beta$  is the vector of all unknown coefficients. The omission of an  $i$ -subscript on  $X$  and  $\beta$  is deliberate; the same coefficients and the same explanatory variables can occur in different equations. The earlier assumptions about the  $e_i$  are retained.

With a nonlinear model,  $f(\beta) \propto \text{constant}$  may no longer be suitable as a noninformative prior; consideration needs to be given to the type of nonlinear function and to whether particular values for some parameters need to be excluded. Thus, we give results for a general prior on  $\beta$ , denoted by  $f(\beta)$ . We retain the noninformative prior  $f(\Sigma) \propto |\Sigma|^{-(M+1)/2}$ , and assume a priori independence of  $\beta$  and  $\Sigma$ . Thus, the prior pdf is given by

$$f(\beta, \Sigma) \propto |\Sigma|^{-(M+1)/2} f(\beta) \quad (32)$$

The likelihood function can be written as

$$\begin{aligned} f(y | \beta, \Sigma) &= (2\pi)^{-MT/2} |\Sigma|^{-T/2} \exp\{-\frac{1}{2}(y - h(X, \beta))'(\Sigma^{-1} \otimes I_T)(y - h(X, \beta))\} \\ &= (2\pi)^{-MT/2} |\Sigma|^{-T/2} \exp\{-\frac{1}{2}\text{tr}(A\Sigma^{-1})\} \end{aligned} \quad (33)$$

where  $y = (y'_1, y'_2, \dots, y'_M)'$ ,  $h = (h'_1, h'_2, \dots, h'_M)'$ , and, now,  $A$  is an  $(M \times M)$  matrix with  $(i, j)$ -th element given by

$$[A]_{ij} = [y_i - h_i(X, \beta)]' [y_j - h_j(X, \beta)] \quad (34)$$

The joint posterior pdf for  $(\beta, \Sigma)$  is

$$f(\beta, \Sigma | y) \propto f(\beta) |\Sigma|^{-(T+M+1)/2} \exp\{-\frac{1}{2} \text{tr}(A\Sigma^{-1})\} \quad (35)$$

and, integrating out  $\Sigma$ , the marginal posterior pdf for  $\beta$  is

$$f(\beta | y) \propto f(\beta) |A|^{-T/2} \quad (36)$$

Thus, the posterior for  $\beta$  in the nonlinear SUR model involves the same determinant of sums of squares and cross products of residuals as it does in the linear model. A more general prior has been added. (Of course, it also could have been included in the linear model.)

The Metropolis-Hastings algorithm described in Section III can be readily applied to the posterior pdf in equation (36). Because the earlier results on conditional posterior pdfs for  $\beta$  and the  $\beta_i$  no longer hold, the draws need to be used directly to estimate posterior pdfs and their moments.

## V. IMPOSING LINEAR EQUALITY RESTRICTIONS

Economic applications of SUR models frequently involve linear restrictions on the coefficients. For example, the same coefficient may appear in more than one equation, the Slutsky symmetry conditions in demand models lead to cross-equation restrictions, or one might want to hypothesize that all equations have the same coefficient vector. Under the existence of cross-equation linear restrictions, the Gibbs sampler using  $\beta$  and  $\Sigma$ , and the Metropolis-Hastings algorithm, can still be used. However, the Gibbs sampler involving only  $\beta$  is no longer applicable. If the restrictions are all within equation restrictions, all three algorithms are possible.

Suppose a set of  $J$  linear restrictions is written as

$$R\beta = (R_1 \ R_2) \begin{pmatrix} \eta \\ \gamma \end{pmatrix} = r \quad (37)$$

where  $R_1$  is  $(J \times J)$  and nonsingular,  $R_2$  is  $(J \times (K - J))$ , and  $\eta$  and  $\gamma$  are  $J$  and  $(K - J)$  dimensional sub-vectors of  $\beta$ , respectively. To make this partition, it may be necessary to reorder the elements in  $\beta$ . Correspondingly, we can reorder the columns of  $X$  and partition it so that the linear SUR model can be written as

$$y = X\beta + e = (X_1 \ X_2) \begin{pmatrix} \eta \\ \gamma \end{pmatrix} + e \quad (38)$$

This reordering may destroy the block-diagonal properties of  $X$ . From (37), we can solve for  $\eta$  as

$$\eta = R_1^{-1}(r - R_2\gamma) \quad (39)$$

Substituting (39) into (38) and rearranging yields

$$y - X_1 R_1^{-1} r = (X_2 - X_1 R_1^{-1} R_2) \gamma + e$$

or

$$z = Z\gamma + e \quad (40)$$

where  $z = y - X_1 R_1^{-1} r$ , and  $Z = X_2 - X_1 R_1^{-1} R_2$  represent new sets of ‘‘observations’’.

In general,  $Z$  and  $\gamma$  can no longer be partitioned unambiguously into  $M$  separate equations. However, the stochastic properties of  $e$  remain the same. Thus, all the results in Sections II and III that did not rely on a partitioning of  $X$  and  $\beta$  can still be applied to the model in (40). In particular, a Gibbs sampler can be used to draw  $\gamma$  and

$\Sigma$ , and the Metropolis-Hastings algorithm can be used to draw  $\gamma$ , from their respective posterior pdf's.

## VI. IMPOSING INEQUALITY RESTRICTIONS

Possible inequality restrictions on the coefficients range from simple ones such as a sign restriction on a single coefficient to more complex ones such as enforcing the eigenvalues of a matrix of coefficients to be nonpositive. Letting the feasible region defined by the inequality constraints be denoted by  $S$ , and defining the indicator function

$$I_S(\beta) = \begin{cases} 1 & \text{for } \beta \in S \\ 0 & \text{for } \beta \notin S \end{cases} \quad (41)$$

the inequality restrictions can be accommodated by setting up the otherwise noninformative prior pdf

$$f(\beta, \Sigma) \propto |\Sigma|^{-(M+1)/2} I_S(\beta) \quad (42)$$

Using Bayes' Theorem to combine this prior with the likelihood function in equation (5), we obtain the joint posterior pdf.

$$\begin{aligned} f(\beta, \Sigma | y) &\propto f(y | \beta, \Sigma) f(\beta, \Sigma) \\ &\propto |\Sigma|^{-(T+M+1)/2} \exp\{-\frac{1}{2}(y - X\beta)'(\Sigma^{-1} \otimes I_T)(y - X\beta)\} I_S(\beta) \\ &= |\Sigma|^{-(T+M+1)/2} \exp\{-\frac{1}{2}\text{tr}(A\Sigma^{-1})\} I_S(\beta) \end{aligned} \quad (43)$$

From this result we can derive the following conditional and marginal posterior pdf's.

The conditional posterior pdf for  $(\beta | \Sigma)$  is the truncated multivariate normal distribution

$$f(\beta | \Sigma, y) \propto \exp\{-\frac{1}{2}(\beta - \hat{\beta})' X'(\Sigma^{-1} \otimes I_T) X(\beta - \hat{\beta})\} I_S(\beta) \quad (44)$$

The conditional posterior pdf for  $(\Sigma | y)$  is the same inverted-Wishart distribution as was given in equation (20). The marginal posterior pdf for  $\beta$  is

$$f(\beta | y) \propto |A|^{-T/2} I_S(\beta) \quad (45)$$

The posterior pdf for  $\beta_1$  conditional on the remaining  $\beta_i$  is the truncated multivariate  $t$ -distribution

$$f(\beta_1 | y, \beta_2, \beta_3, \dots, \beta_M) \propto \left[ v_1 + \frac{(\beta_1 - \tilde{\beta}_1)' X_1' Q_{(1)} X_1 (\beta_1 - \tilde{\beta}_1)}{\tilde{s}_1^2} \right]^{-(K_1 + v_1)/2} I_S(\beta) \quad (46)$$

Of interest is how to best use these pdfs to draw observations on  $\beta$ , and possibly  $\Sigma$ , from their respective posterior pdf's. The conditional posterior pdf's for  $(\beta | \Sigma)$  and  $(\Sigma | \beta)$  can be used within a Gibbs sampler providing the inequality restrictions are sufficiently mild for a simple acceptance-rejection algorithm to be practical when sampling from the truncated multivariate normal distribution. By a “simple acceptance–rejection algorithm”, we mean that a draw is made from a nontruncated multivariate normal distribution and, if it lies outside the feasible region, it is discarded and replaced by another draw. This procedure will not be practical if the probability of obtaining a draw within the feasible region is small, which will almost always be the case if the number of inequality restrictions is moderate to large. Thus, we are using the term “mild” inequality restrictions to describe a situation where the maximum number of draws necessary before a feasible draw is obtained is not excessive.

If the inequality restrictions are not mild, then a Metropolis-Hastings algorithm can be employed. In the steps we described in Section III, if a candidate

value  $\beta^*$  is infeasible, then  $r = 0$ , and the retained draw is automatically the last accepted feasible draw. That is,  $\beta^{(\ell+1)} = \beta^{(\ell)}$ .

If the inequality restrictions are not mild, but are linear, then using a Gibbs sampler on subcomponents of  $\beta$  might prove successful. For example, using the truncated multivariate  $t$ -distributions for each of the  $\beta_i$ , as specified in equation (46), could be useful. Also within different contexts, sampling from truncated multivariate  $t$  and multivariate normal distributions has been broken down into sampling from univariate conditional distributions by Geweke (1991) and Hajivassiliou and McFadden (1990). Also, see the Appendix.

## VII. THREE APPLICATIONS

### A. Wheat Yield

In Griffiths et al (2001) the following model was used for predicting wheat yield in five Western Australian shires.

$$Y_t = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 t^3 + \beta_5 G_t + \beta_6 G_t^2 + \beta_7 D_t + \beta_8 D_t^2 + \beta_9 F_t + \beta_{10} F_t^2 + e_t \quad (47)$$

Yield ( $Y_t$ ) depends on a cubic time trend to capture technological change and on quadratic functions of rainfall during the germination period ( $G_t$ ), the development period ( $D_t$ ), and the flowering period ( $F_t$ ). The rainfalls are measured relative to their sample means. Inequality restrictions are imposed to ensure that the response of yield to rainfall, at average rainfall, is positive. That is, for germination rainfall, for example,  $\partial Y / \partial G = \beta_5 + 2\beta_6 > 0$ . Thus, the feasible region for this example is

$$S(\beta) = \{\beta \mid \beta_5 + 2\beta_6 > 0, \beta_7 + 2\beta_8 > 0, \beta_9 + 2\beta_{10} > 0\} \quad (48)$$



Although Griffiths et al (2001) used separate single equation estimation for the five shires and focussed on several forecasting issues, investigation within a five-equation SUR model has started. Given that the inequality restrictions within each equation are relatively mild, but in total they are not, a Gibbs sampler using the truncated  $t$  densities in equation (46) seems a profitable direction to follow. Also, some preliminary work involving the Metropolis-Hastings algorithm on the complete  $\beta$  vector has proved effective.

## B. Cost and Share Equations

In a second application, a translog cost function (constant returns to scale) and cost-share equations for merino woolgrowers (310 observations over 23 years) was estimated by Griffiths et al (2000) using, as inputs, land, capital, livestock and other. In the equations that follow  $c$  is cost,  $q$  is output, the  $w_i$  are input prices and the  $S_i$  are input shares.

$$\log\left(\frac{c}{q}\right) = \beta_0 + \beta_T T + \sum_{i=1}^4 \beta_i \log(w_i) + 0.5 \sum_{i=1}^4 \sum_{j=1}^4 \beta_{ij} \log(w_i) \log(w_j) + e_1$$

$$S_i = \beta_i + \sum_{j=1}^4 \beta_{ij} \log(w_j) + e_i \quad i = 2, 3, 4$$

This SUR model has the following characteristics.

1. The equations are linear.
2. There are a number of linear equality restrictions that need to be imposed.

Specifically, the  $\beta_{ij}$ s in the cost function are equal to the  $\beta_{ij}$ s in the share equations, and, furthermore, to satisfy homogeneity and symmetry, we require

$$\sum_{i=1}^4 \beta_i = 1 \quad \sum_{j=1}^4 \beta_{ij} = 0 \quad \beta_{ij} = \beta_{ji}$$

3. Inequality restrictions are required for the functions to satisfy concavity and monotonicity. These restrictions are

- Monotonicity  $0 < S_i < 1$
- Concavity  $B - S + ss'$  is negative semidefinite where

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \\ \beta_{31} & \beta_{32} & \beta_{33} & \beta_{34} \\ \beta_{41} & \beta_{42} & \beta_{43} & \beta_{44} \end{bmatrix}$$

$$S = \begin{bmatrix} S_1 & & & \\ & S_2 & & \\ & & S_3 & \\ & & & S_4 \end{bmatrix} \quad s = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{bmatrix}$$

Note that  $B - S + ss'$  is negative semidefinite if and only if its largest eigenvalue is nonpositive.

Since  $S_i$  depends on the input prices, a decision concerning the input prices at which  $S_i$  is evaluated, and the inequality restrictions imposed, needs to be made. The inequality restrictions were imposed at average input prices for each of the 23 years.

4. Given the severe inequality restrictions that were imposed, the Metropolis-Hastings algorithm was used.
5. The quantities of interest are nonlinear functions of the parameters. They are the elasticities of substitution

$$\sigma_{ij} = \frac{\beta_{ij}}{S_i S_j} + 1 \quad i \neq j$$

and input demand elasticities

$$\eta_{ii} = \frac{\beta_{ii}}{S_i} + S_i - 1$$

$$\eta_{ij} = \frac{\beta_{ij}}{S_i} + S_j$$

### C. Expenditure Functions

Our third example involves two expenditure functions estimated from a sample of 1,834 Bangkok households, and deflated by an “equivalence scale” measure of household size (Griffiths and Chotikapanich 1997). For the  $t$ -th observation, the functions are

$$w_{1t} = \frac{\alpha_1 m_{1t}}{x_t} + \frac{\beta_1 m_{1t} (x_t - \alpha_1 m_{1t} - \alpha_2 m_{2t} - \alpha_3 m_{3t})}{x_t (\beta_1 m_{1t} + \beta_2 m_{2t} + \beta_3 m_{3t})} + e_{1t}$$

$$w_{2t} = \frac{\alpha_2 m_{2t}}{x_t} + \frac{\beta_2 m_{2t} (x_t - \alpha_1 m_{1t} - \alpha_2 m_{2t} - \alpha_3 m_{3t})}{x_t (\beta_1 m_{1t} + \beta_2 m_{2t} + \beta_3 m_{3t})} + e_{2t}$$

$$m_{1t} = 1 + \delta_{21} n_{2t} + \delta_{31} n_{3t}$$

$$m_{2t} = 1 + \delta_{22} n_{2t} + \delta_{32} n_{3t}$$

$$m_{3t} = 1 + \delta_{23} n_{2t} + \delta_{33} n_{3t}$$

where

$w_{jt}$  = expenditure proportion for commodity  $j$ ,

$x_t$  = total expenditure,

$m_{jt}$  = equivalence scale for commodity  $j$ ,

$n_{2t}$  = number of extra adults (each household has at least one adult),

$n_{3t}$  = number of children.

The unknown parameters are

$$(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \delta_{21}, \delta_{31}, \delta_{22}, \delta_{32}, \delta_{23}, \delta_{33})$$

This SUR model has the following characteristics:

1. The equations are nonlinear in the parameters.
2. A number of inequality restrictions were imposed, namely,

$0 < \beta_1, \beta_2 < 1$  Additional expenditure from a one-unit increase in supernumerary income must lie between zero and one.

$0 \leq \delta_{3j} \leq \delta_{2j} \leq 1$  Expenditure requirements for extra adults are less than those for the first adult but greater than those for children.

$$\alpha_i < \min_t \left\{ \frac{e_{it}}{1 + \delta_{2i}n_{2t} + \delta_{3i}n_{3t}} \right\} \quad i = 1, 2$$

The smallest level of consumption in the sample must be greater than subsistence expenditure, a constraint from the utility function.

3. Given the nonlinear equations and the inequality constraints, the Metropolis-Hastings algorithm was used.
4. Two nonlinear functions of the parameters are of interest. They are the general scale or "household size":

$$m_0 = \frac{x \sum_k \beta_k m_k}{x - \sum_k \alpha_k m_k + \sum_k \alpha_k \sum_k \beta_k m_k}$$

and the elasticities. Expressions for the latter can be found in Griffiths and Chotikapanich (1997).

## VIII. FORECASTING

Suppose that we are interested in forecasting dependent variable values in the next period. The shire-level wheat yield application in the previous section is an example of where such a forecast would be of interest. In that case the objective is to forecast yield for each of the five shires. Since the yields are correlated via the stochastic assumptions of the SUR model, a joint forecast is appropriate. We can write next period's observation as

$$y_* = X_* \beta + e_* \quad (49)$$

where  $y_*$  is an  $M$ -dimensional vector,  $X_*$  is an  $(M \times K)$  block diagonal matrix with the  $i$ -th block being a  $(1 \times K_i)$  row vector containing next period's explanatory variables for the  $i$ -th equation,

$$X_* = \begin{bmatrix} x_{1*} & & & \\ & x_{2*} & & \\ & & \ddots & \\ & & & x_{3*} \end{bmatrix} \quad (50)$$

and  $e_* \sim N(0, \Sigma)$  is next periods  $(M \times 1)$  random error vector. The conventional Bayesian forecasting tool is the predictive pdf  $f(y_* | y)$ . Graphing marginal predictive pdf's from this density function, and computing its means, standard deviations and probabilities of interest are the standard ways of reporting results.

The procedure for deriving the predictive pdf is to begin with the joint pdf  $f(y_*, \beta, \Sigma | y)$  and to then integrate out  $\Sigma$  and  $\beta$ , either analytically or via a numerical sampling algorithm. Now,

$$\begin{aligned} f(y_* | \beta, \Sigma) &= (2\pi)^{-M/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2}(y_* - X_*\beta)' \Sigma^{-1} (y_* - X_*\beta)\} \\ &\propto |\Sigma|^{-1/2} \exp\{-\frac{1}{2} \text{tr}(A_* \Sigma^{-1})\} \end{aligned} \quad (51)$$

where  $A_* = [y - X_*\beta][y - X_*\beta]'$ . Thus, using the posterior pdf in equation (12) (no inequality restrictions), we have

$$\begin{aligned} f(y_*, \beta, \Sigma | y) &= f(y_* | \beta, \Sigma) f(\beta, \Sigma | y) \\ &\propto |\Sigma|^{-(T+M+2)/2} \exp\{-\frac{1}{2} \text{tr}[(A + A_*) \Sigma^{-1}]\} \end{aligned}$$

Using properties of the inverted Wishart distribution to integrate out  $\Sigma$  yields

$$\begin{aligned} f(y_*, \beta | y) &= \int f(y_*, \beta, \Sigma | y) d\Sigma \\ &\propto |A + A_*|^{-(T+1)/2} \end{aligned} \quad (52)$$

Because analytical integration of  $\beta$  out of equation (52) is not possible, we consider the conditional predictive pdf  $f(y_* | \beta, y)$ . It turns out that this pdf is a multivariate student  $t$ . Thus,  $f(y_* | y)$  and its moments can be estimated by averaging quantities from  $f(y_* | \beta, y)$  over draws of  $\beta$  obtained using one of the MCMC algorithms described earlier.

To establish that  $f(y_* | \beta, y)$  is a multivariate  $t$ -distribution, we first note that (see, for example, Dhrymes 1978, p. 458)

$$|A + A_*| = |A| (1 + (y_* - X_*\beta)' A^{-1} (y_* - X_*\beta)) \quad (53)$$

Thus,

$$\begin{aligned}
f(y_* | \beta, y) &\propto \left[ 1 + (y_* - X_*\beta)' A^{-1} (y_* - X_*\beta) \right]^{-(T+1)/2} \\
&\propto \left[ v_* + (y_* - X_*\beta)' \left( \frac{A}{v_*} \right)^{-1} (y_* - X_*\beta) \right]^{-(M+v_*)/2}
\end{aligned} \tag{54}$$

where  $v_* = T - M + 1$ . Equation (54) is a multivariate  $t$ -distribution with mean

$$E(y_* | \beta, y) = X_*\beta \tag{55}$$

covariance matrix

$$V(y_* | \beta, y) = \frac{A}{v_* - 2} \tag{56}$$

and degrees of freedom  $v_*$ . Given draws  $\beta^{(\ell)}$ ,  $\ell = 1, 2, \dots, N$  from an MCMC algorithm, one can average the quantities in equations (54) to (56) over these draws to estimate the required marginal predictive pdf's and their moments. Marginal univariate  $t$  distributions from (54) are averaged and the formulas are analogous to those in equations (22), (26) and (27) except, of course, that our random variable of interest is now an element of  $y_*$ , say  $y_{*i}$ , not  $\beta_{ik}$ .

Percy (1992) describes an alternative Gibbs sampling approach where  $y_*$ ,  $\beta$  and  $\Sigma$  are recursively generated from their respective conditional pdf's. With our approach, it is not necessary to generate draws on  $y_*$ . Also, because we have derived the predictive pdf conditional on  $\beta$ , the introduction of inequality restrictions on  $\beta$  does not change the analysis. The range of values of  $\beta$  over which averaging takes place is restricted, but that is accommodated by the way in which  $\beta$  is drawn, and the result in (54) still holds.

An interesting extension, and one that is of concern to Griffiths et al. (2001), is capturing the extra uncertainty created by not knowing the value of one or more

regressors in  $X_*$ . We have this problem if a wheat yield forecast is made prior to all rainfalls having been observed. The effect can be captured by modelling rainfall and averaging the predictive pdf for yield conditional on rainfall over rainfalls draws made from its predictive pdf.

## IX. SOME EXTENSIONS

Consider estimating  $\beta$  in the SUR model when there are missing observations on one or more of the dependent variables. This problem was considered in the context of expenditure functions by Supat (1996). For the moment, assume the observations are truly missing and that they are missing at random; they are not zeros created by negative values of an unobserved latent variable, as in the case with the Tobit model. Writing  $y^O$  to denote observed components and  $y^U$  to denote unobserved components, estimation can proceed within a Gibbs sampling framework using the conditional posterior pdfs  $f(\beta | \Sigma, y^O, y^U)$ ,  $f(\Sigma | \beta, y^O, y^U)$  and  $f(y^U | \beta, \Sigma, y^O)$ . The conditional posterior pdfs for  $\beta$  and  $\Sigma$  are the normal and inverted Wishart pdf's given in equations (14) and (20). To investigate how to draw observations from  $f(y^U | \beta, \Sigma, y^O)$ , we write the  $(M \times 1)$   $t$ -th observation  $y_{(t)}$  as

$$y_{(t)} = X_{(t)}\beta + e_{(t)} \quad (57)$$

The subscript  $t$  has been placed in the parentheses to distinguish the  $(M \times 1)$   $t$ -th observation all equations  $y_{(t)}$  from the  $(T \times 1)$  observations on the  $i$ -th equation  $y_i$ . The structure of  $X_{(t)}$  is similar to that of  $X_*$  defined in equation (50). We wish to consider equation (57) for all values of  $t$  where  $y_{(t)}$  has one or more unobserved components. Reordering the elements if necessary, we can partition equation (57) as



$$\begin{pmatrix} y_{(t)}^U \\ y_{(t)}^O \end{pmatrix} = \begin{pmatrix} X_{(t)}^U \\ X_{(t)}^O \end{pmatrix} \beta + \begin{pmatrix} e_{(t)}^U \\ e_{(t)}^O \end{pmatrix} \quad (58)$$

where we write

$$E \begin{bmatrix} e_{(t)} & e'_{(t)} \end{bmatrix} = \begin{bmatrix} \Sigma^{UU} & \Sigma^{UO} \\ \Sigma^{OU} & \Sigma^{OO} \end{bmatrix} \quad (59)$$

The conditional posterior pdf  $f(y_{(t)}^U | \beta, \Sigma, y_{(t)}^O)$  is a multivariate normal distribution with mean

$$E(y_{(t)}^U | \beta, \Sigma, y_{(t)}^O) = X_{(t)}^O \beta + \Sigma^{UO} \Sigma^{OO^{-1}} (y_{(t)}^O - X_{(t)}^O \beta) \quad (60)$$

and covariance matrix

$$V(y_{(t)}^U | \beta, \Sigma, y_{(t)}^O) = \Sigma^{UU} - \Sigma^{UO} \Sigma^{OO^{-1}} \Sigma^{OU} \quad (61)$$

Furthermore,  $(y_{(t)}^U | \beta, \Sigma, y_{(t)}^O)$ ,  $t = 1, 2, \dots, T$  are independent. Thus, for generating  $y_{(t)}^U$  within the Gibbs sampler, we use the conditional normal distributions defined by equations (60) and (61) for all observations where an unobserved component is present.

Suppose, now, that the unobserved components represent negative values of a Tobit-type latent variable. In this case we have the additional posterior information that the elements of  $y_{(t)}^U$  are negative. The conditional posterior pdf for  $(y_{(t)}^U | \beta, \Sigma, y_{(t)}^O)$  becomes a truncated (multivariate) normal distribution with a truncation that forces  $y_{(t)}^U$  to be negative. Its location vector and scale matrix (no longer the mean and covariance matrix) are given in equations (60) and (61). A

convenient algorithm for drawing from this truncated normal distribution is described in the Appendix.

For extensions into Probit models, see Geweke et al. (1997) and references therein. The literature on simultaneous equation models with Tobit and Probit variables can be accessed through Li (1998). Sets of SUR expenditure functions with a common parameter and with unobserved expenditures that result from infrequency of purchase are considered by Griffiths and Valenzuela (1998). Smith and Kohn (2000) study Bayesian estimation of nonparametric SURs.

## **X. CONCLUDING REMARKS**

With the recent explosion of literature on MCMC techniques, Bayesian inference in the SUR model has become a practical reality. However, it is the author's view that, prior to the writing of this chapter, the relevant results have not been collected and summarised in a form convenient for applied researchers to implement. It is my hope this chapter will facilitate and motivate many more applications of Bayesian inference in the SUR model.

## **XI. APPENDIX – DRAWING RANDOM VARIABLES AND VECTORS**

### **A. Multivariate Normal Distribution**

To draw a vector  $y$  from a  $N(\mu, \Sigma)$  distribution:

1. Compute the Cholesky decomposition  $H$  such that  $HH' = \Sigma$ .
2. Generate  $z$  from  $N(0, I)$ .
3. Calculate  $y = \mu + Hz$ .

## B. Multivariate $t$ Distribution

Consider the multivariate  $k$ -dimensional  $t$ -distribution with pdf

$$f(x|\mu, V) \propto \left[ v + (x - \mu)'V^{-1}(x - \mu) \right]^{-(k+v)/2}$$

It has  $v$  degrees of freedom, mean  $\mu$  and covariance matrix  $(v/(v-2))V$ . (Assume  $v > 2$ .) To draw a vector  $x$  from this pdf:

1. Compute the Cholesky decomposition  $H$  such that  $HH' = V$ .
2. Generate the  $(k \times 1)$  vector  $z_1$  from  $N(0, I_k)$ .
3. Generate the  $(v \times 1)$  vector  $z_2$  from  $N(0, I_v)$ .
4. Calculate  $x = \mu + Hz_1 / \sqrt{z_2'z_2/v}$ .

## C. Inverted Wishart Distribution

Let  $\Sigma$  have an  $m$ -dimensional inverted Wishart distribution with parameter matrix  $S$  and degrees of freedom  $v$ . It has pdf

$$f(\Sigma|S) \propto |\Sigma|^{-(v+m+1)/2} \exp\{-\frac{1}{2} \text{tr}(S\Sigma^{-1})\}$$

To draw observations on  $\Sigma$ :

1. Compute the Cholesky decomposition  $H$  such that  $HH' = S^{-1}$ .
2. Draw independent  $(m \times 1)$  normal random vectors  $z_1, z_2, \dots, z_v$  from  $N(0, I_m)$ .
3. Calculate  $\Sigma = \left( H \sum_{i=1}^v z_i z_i' H' \right)^{-1}$ .

#### D. Univariate Truncated Normal Distribution

Suppose that  $x$  is a truncated normal random variable with location  $\mu$ , scale  $\sigma$  and truncation  $a < x < b$ . To draw  $x$ :

1. Draw a uniform (0,1) random variable  $U$ .
2. Calculate

$$x = \mu + \sigma \Phi^{-1} \left[ \Phi \left( \frac{a - \mu}{\sigma} \right) + U \left( \Phi \left( \frac{b - \mu}{\sigma} \right) - \Phi \left( \frac{a - \mu}{\sigma} \right) \right) \right] \quad (\text{A.1})$$

where  $\Phi$  is the standard normal cumulative distribution function.

#### E. Multivariate Truncated Normal Distribution

Suppose that  $x$  is an  $m$ -dimensional multivariate truncated normal distribution such that  $a_1 < x_1 < b_1$ ,  $a_2 < x_2 < b_2, \dots, a_m < x_m < b_m$ .

1. Use (A.1) to draw  $x_1$ .
2. Find the location and scale parameters for the truncated conditional normal distribution  $(x_2 | x_1)$  conditional on  $x_1$  drawn in step 1.
3. Apply (A.1) to the distribution  $(x_2 | x_1)$ .
4. Find location and scale parameters for the distribution of  $(x_3 | x_2, x_1)$  conditional on the draws made in steps 1 and 3.
5. Apply (A.1) to the distribution  $(x_3 | x_2, x_1)$ .
6. And so on.

**REFERENCES**

- Albert, J.H., Chib, S., 1996. Computation in Bayesian econometrics: an introduction to Markov chain Monte Carlo. In: Hill, R.C. (Ed.), *Advances in Econometrics Volume 11A: Computational Methods and Applications*. JAI Press, Greenwich, pp.3-24.
- Chen, M.-H., Shao, Q.-M., Ibrahim, J.G., 2000. *Monte Carlo Methods in Bayesian Computation*. Springer, New York.
- Chib, S., 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., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12, 409-431.
- Cowles, M.K., Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 91, 883-904.
- Dhrymes, P.J., 1978. *Introductory Econometrics*. Springer, New York.
- Drèze, J.H., Morales, J.A., 1976. Bayesian full information analysis of simultaneous equations. *Journal of the American Statistical Association* 71, 919-923.
- Fiebig, D., 2001. Seemingly unrelated regression. In: Baltagi, B. (Ed.), *Companion in Econometrics*, Basil Blackwell, London (Chapter 5).
- Fiebig, D.G., Kim J.H., 2000. Estimation and inference in SUR models when the number of equations is large. *Econometric Reviews* 19, 105-130.

- Geweke, J., 1988. The secular and cyclical behavior of real GDP in nineteen OECD countries. *Journal of Business and Economic Statistics* 6, 479-486.
- Geweke, J., 1991. Efficient simulation from the multivariate normal and student t-distributions subject to linear constraints. In: Keramidas E.M. (ed.), *Computing Science and Statistics: Proceedings of the Twenty-third Symposium on the Interface*, Interface Foundation of North America, Fairfax, pp.571-578.
- Geweke, J., Keane, M.P., Runkle, D.E., 1997. Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*. 80, 127-166.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., (Eds.), 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- Griffiths, W.E., 1999. Estimating consumer surplus: comment on using simulation methods for Bayesian econometric models: inference development and communication. *Econometric Reviews* 18, 75-88.
- Griffiths, W.E., Chotikapanich, D., 1997. Bayesian methodology for imposing inequality constraints on a linear expenditure function with demographic factors. *Australian Economic Papers* 36, 321-341.
- Griffiths, W.E., O'Donnell, C.J., Tan-Cruz, A., 2000. Imposing regularity conditions on a system of cost and factor share equations. *Australian Journal of Agricultural and Resource Economics* 44, 107-127.
- Griffiths, W.E., Newton, L.S., O'Donnell, C.J., 2001, Predictive densities for shire level wheat yield in Western Australia. Paper contributed to the Australian Agricultural and Resource Economics Society Conference in Adelaide.
- Griffiths, W.E., Skeels, C.J., Chotikapanich, D., 2001. Sample size requirements for estimation in SUR models. In: Ullah, A., Wan, A., Chaturvedi, A., (Eds.),

Handbook of Applied Econometrics and Statistical Inference, Marcel Dekker, New York, forthcoming.

Griffiths, W.E., Valenzuela, M.R., 1998. Missing data from infrequency of purchase: Bayesian estimation of a linear expenditure system. In: Fomby, T.B., Hill, R.C. (Eds.), *Advances in Econometrics, Volume 13, Messy Data, Missing Observations, Outliers and Mixed-Frequency Data*, JAI Press, Greenwich, pp.75-102.

Hajivassiliou, V.A., McFadden, D.L., 1990, The method of simulated scores for the estimation of LDV models with an application to external debt crises. Yale Cowles Foundation Discussion Paper No. 967.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., Lee, T.-C., 1985. *The Theory and Practice of Econometrics*, second ed. John Wiley and Sons, New York.

Kloek, T., van Dijk, H.K., 1978. Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. *Econometrica* 46, 1-19.

Li, K., 1998. Bayesian inference in a simultaneous equation model with limited dependent variables. *Journal of Econometrics* 85, 387-400.

Lütkepohl, H., 1996. *Handbook of Matrices*. John Wiley and Sons, Chichester.

Percy, D.F., 1992. Prediction for seemingly unrelated regressions. *Journal of the Royal Statistical Society B* 54, 243-252.

Percy, D.F., 1996. Zellner's influence on multivariate linear models. In: Berry, D.A., Chaloner, K.M., Geweke, J.K. (Eds.), *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, John Wiley and Sons, New York (Chapter 17).

- Richard, J.-F., Tompa, H., 1980. On the evaluation of poly- $t$  density functions. *Journal of Econometrics* 12, 335-351.
- Richard, J.-F., Steel, M.F.J., 1988. Bayesian analysis of systems of seemingly unrelated regression equations under a recursive extended natural conjugate prior density. *Journal of Econometrics* 38, 7-37.
- Smith, M., Kohn, R., 2000. Nonparametric seemingly unrelated regression. *Journal of Econometrics* 98, 257-282.
- Srivastava, V.K., Dwivedi, T.D., 1979. Estimation of seemingly unrelated regression equations: a brief survey. *Journal of Econometrics* 10, 15-32.
- Srivastava, V.K., Giles, D.E.A., 1987. *Seemingly Unrelated Regression Equations Models: Estimation and Inference*. Marcel Dekker, New York.
- Steel, M.F.J., 1992. Posterior analysis of restricted seemingly unrelated regression equation models: a recursive analytical approach. *Econometric Reviews* 11, 129-142.
- Supat, K., 1996. *Seemingly Unrelated Regression with Missing Observations on Dependent Variables*. B.Ec. Honours dissertation, University of New England, Armidale.
- Tanner, M.A., 1996. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, third ed. Springer, New York.
- Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion and rejoinder). *Annals of Statistics* 22, 1701-1762.



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

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