ISSN 0819-2642
ISBN 0 7340 2568 8

THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 912

AUGUST 2004

GIBBS SAMPLERS FOR A SET OF SEEMINGLY UNRELATED REGRESSIONS

by

W. E. Griffiths
&
Ma. Rebecca Valenzuela

Department of Economics
The University of Melbourne
Melbourne Victoria 3010
Australia.

Gibbs Samplers for A Set of Seemingly Unrelated Regressions

William E Griffiths

Economics Department

University of Melbourne

Ma. Rebecca Valenzuela

Economics Department

Monash University

August 19, 2004

Summary

Bayesian estimation of a collection of seemingly unrelated regressions, referred to as a 'set of seemingly unrelated regressions' is considered. The collection of seemingly unrelated regressions is linked by common coefficients and/or a common error covariance matrix. Gibbs samplers useful for estimating posterior quantities are described and applied to two examples – a set of linear expenditure functions and a cost function and share equations from production theory.

1. Introduction

The seemingly unrelated regressions (SUR) model was introduced by Zellner (1962) to accommodate and take advantage of contemporaneous correlation in the errors of linear models that might otherwise appear unrelated. Since then, it has been studied extensively (see surveys by Srivastava and Dwivedi (1979), Srivastava and Giles (1987), and Fiebig (2001)) and has become commonplace in economic applications involving joint estimation of a number of equations. Such applications include joint estimation of equations for expenditures on different commodity groups (with repeated observations over households), joint estimation of production or cost functions and their corresponding first order equations for profit maximization (with repeated observation over firms), and various panel-data applications where several cross-sectional units are observed in a number of time periods.

In this paper, we are concerned with estimating a model that contains within it several SUR models. We are thus adding a third dimension to the conventional SUR model that typically has two dimensions, a number of equations and repeated observations on the variables in these equations. We call the model that contains several SUR models "a set of seemingly unrelated regressions". In the application that motivated this study, one involving estimation of household expenditure functions, the several SUR models are linked by a common coefficient vector. The equations within each SUR model correspond to different expenditure categories, while the different SUR models correspond to households with different demographic compositions. When considering linear expenditure functions derived from a Klein-Rubin utility function with a common coefficient on supernumerary income, but with different subsistence levels for households with different demographic compositions, the demographic-specific coefficients vary over the SUR models whereas the demographic-invariant coefficients of income do not. A second application that we also consider in this paper relates to estimation of factor share equations for a production function, using pooled cross section and time series data, with the coefficients assumed constant over time except for the intercept and the trend coefficients which are time-varying.

Our approach is Bayesian. Given the intractability of the joint posterior density function for all unknown parameters, we are interested in deriving convenient

conditional posterior densities that can be used within a Gibbs sampler for sampling from the joint posterior density. Bayesian estimation of the SUR model was first considered by Zellner (1971) with analytical results for some special cases subsequently being derived by Dreze and Morales (1976), Richard and Tompa (1980), Richard and Steel (1988), and Steel (1992); 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 the SUR model, under various assumptions, has been considered by Percy (1992, 1996), Chib and Greenberg (1995), Griffiths and Chotikapanich (1997), and Smith and Kohn (2000, 2002). Previous work on a set of models seems to have been confined to sampling-theory estimation of models with error components to handle the different cross-section and time-series dimensions. See, for example, Baltagi (2001, Chapter 6) and references therein. In our work, each set is allowed to have a different unrestricted error covariance matrix.

The outline of the paper is as follows. In Section 2, we begin by considering the traditional SUR model with non-informative prior. In addition to reviewing the normal and inverted Wishart conditional posterior densities that are typically used for drawing posterior observations on the coefficients and error covariance matrix, respectively, we suggest an alternative Gibbs sampler that does not require drawing observations on the error covariance matrix. In the second part of Section 2, we describe Gibbs samplers that can be used for a set of SURs. The techniques are applied to estimation of equivalence scales from expenditure systems in Section 3 and to estimation of the parameters of a translog cost function in Section 4. Some concluding remarks are made in Section 5. Proofs of results are given in an appendix to the paper.

2. Models and Gibbs Samplers

2.1 The Traditional SUR Model

To introduce the various Gibbs samplers relevant for a set of SUR models under alternative assumptions, we begin with the traditional SUR model with M equations written as

$$y_i = X_i\beta_i + e_i \qquad i = 1, 2, \ldots, M \tag{1}$$

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 non-stochastic explanatory variables, possibly including a constant term, β_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

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

that we then write compactly as

$$y = X\beta + e \tag{3}$$

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

$$e \sim N(0, \Sigma \otimes I_T) \tag{4}$$

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 σ_{ii}, the *i*-th diagonal element of Σ. The covariance between two corresponding errors in different equations (say *i* and *j*), we write as σ_{ij}; the σ_{ij} appear as off-diagonal elements of Σ.

Using $f(.)$ as generic notation for a probability density function (pdf), the likelihood function for β and Σ can be written as

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

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

$$\left[A\right]_{ij}=\left(y_i-X_i\beta_i\right)'\left(y_j-X_j\beta_j\right) \tag{6}$$

The conventional non-informative prior for $\left(\beta,\Sigma\right)$, and the one we shall employ throughout the paper, is (see for example, Zellner 1971, Ch. 8),

$$f\left(\beta,\Sigma\right)=f\left(\beta\right)f\left(\Sigma\right)\propto\left|\Sigma\right|^{-(M+1)/2} \tag{7}$$

The marginal posterior pdf for β, obtained by combining the prior in (7) with the likelihood in (5), and integrating out Σ, is

$$f\left(\beta\,|\,y\right)\propto\left|A\right|^{-T/2} \tag{8}$$

For this posterior pdf to be proper, the sample size T must satisfy the condition $T\geq M+\operatorname{rank}\left(X^*\right)$, where $X^*=\left(X_1,X_2,\ldots,X_M\right)$. See Griffiths et al (2002) for a proof.

The posterior pdf $f\left(\beta\,|\,y\right)\propto\left|A\right|^{-T/2}$ is analytically intractable in the sense that marginal posterior pdf's for the individual elements in β, and their moments, cannot be readily derived. Thus, estimation of this pdf and its moments via MCMC is an attractive alternative. The most common MCMC algorithm is a Gibbs sampler used to draw from the conditional posterior pdfs for $\left(\beta\,|\,\Sigma\right)$ and $\left(\Sigma\,|\,\beta\right)$. It can be shown that these pdfs are the following normal and inverted Wishart pdfs, respectively

$$f\left(\beta\,|\,\Sigma,y\right)\propto\exp\left\{-\frac{1}{2}(\beta-\hat{\beta})'X'(\Sigma^{-1}\otimes I_T)X(\beta-\hat{\beta})\right\} \tag{9}$$

$$f\left(\Sigma \mid \beta, y\right) \propto \left|\Sigma\right|^{-(T+M+1)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(A\Sigma^{-1}\right)\right\} \tag{10}$$

where $\hat{\beta} = [X'(\Sigma^{-1} \otimes I_T)X]^{-1} X'(\Sigma^{-1} \otimes I_T)y$ is the (sampling-theory) generalized least squares estimator.

Another possible Gibbs sampler can be obtained by considering the conditional posterior pdf for the coefficients of each equation β_i, conditional on the coefficients in the other equations, $\beta_{-i} = (\beta_1', \ldots, \beta_{i-1}', \beta_{i+1}', \ldots, \beta_M')'$, but not conditional on the covariance matrix Σ. By partitioning the posterior pdf $f\left(\beta \mid y\right) \propto \left|A\right|^{-T/2}$ it can be shown that the conditional posterior pdf $f\left(\beta_i \mid \beta_{-i}, y\right)$ is the multivariate *t*-distribution

$$f\left(\beta_i \mid \beta_{-i}, y\right) \propto \left[v_i + \frac{(\beta_i - \tilde{\beta}_i)' X_i' Q_i X_i (\beta_i - \tilde{\beta}_i)}{\tilde{s}_i^2}\right]^{-(K_i + v_i)/2} \tag{11}$$

where Q_i is the idempotent matrix $Q_i = I_T - E_{-i}(E_{-i}'E_{-i})^{-1}E_{-i}'$, with the $[T \times (M-1)]$ matrix E_{-i} obtained by deleting the *i*-th column of

$$E = \left[(y_1 - X_1\beta_1), (y_2 - X_2\beta_2), \ldots, (y_M - X_M\beta_M)\right] \tag{12}$$

The degrees of freedom parameter is $v_i = T - K_i$, the conditional mean is given by $\tilde{\beta}_i = \left(X_i'Q_iX_i\right)^{-1} X_i'Q_iy$, and $\tilde{s}_i^2 = (y_i - X_i\tilde{\beta}_i)'Q_i(y_i - X_i\tilde{\beta}_i)/v_i$.

A proof of the result in equation (11) is given in the appendix.

The mean $\tilde{\beta}_i$ and inverse precision matrix $\tilde{s}_i^2\left(X_i'Q_iX_i\right)^{-1}$ can be found by regressing a "corrected" y_i on a "corrected" X_i. The corrections are obtained by "subtracting out" E_{-i}, the influence of the residuals in the other equations. The required steps can be summarized as follows.

1. Regress y_i and each of the columns of X_i on E_{-i}, and compute the corresponding predictions

$$y_i^* = E_{-i}\left(E_{-i}'E_{-i}\right)^{-1} E_{-i}'y_i$$
$$X_i^* = E_{-i}\left(E_{-i}'E_{-i}\right)^{-1} E_{-i}'X_i \tag{13}$$

2. Construct the variables

$$\tilde{y}_i = y_i - y_i^* = Q_i y_i$$
$$\tilde{X}_i = X_i - X_i^* = Q_i X_i \tag{14}$$

3. Regress $\tilde{y}_i$ on $\tilde{X}_i$ to obtain

$$\tilde{\beta}_i = \left(\tilde{X}_i'\tilde{X}_i\right)^{-1} \tilde{X}_i'\tilde{y}_i = \left(X_i'Q_iX_i\right)^{-1} X_i'Q_iy_i \tag{15}$$

and

$$\tilde{s}_i^2 = (\tilde{y}_i - \tilde{X}_i\tilde{\beta}_i)'(\tilde{y}_i - \tilde{X}_i\tilde{\beta}_i)/v_i$$
$$= (y_i - X_i\tilde{\beta}_i)'Q_i(y_i - X_i\tilde{\beta}_i)/v_i \tag{16}$$

In some models there are cross-equation restrictions on some of the coefficients. Such is the case when theoretical considerations suggest the same coefficients appear in different equations, as is the case in the second of our applications in this paper. In these circumstances the conditional posterior pdfs for $(\beta\,|\,\Sigma)$ and $(\Sigma\,|\,\beta)$ can still be used in a Gibbs sampler, after redefining *X* in a suitable manner. As an example of how *X* can be redefined, consider the following two-equation model with θ common to each of the equations

$$y_1 = X_{11}\alpha + X_{12}\theta + e_1$$
$$y_2 = X_{21}\delta + X_{22}\theta + e_2 \tag{17}$$

Writing these two equations jointly yields

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_{11} & 0 & X_{12} \\ 0 & X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} \alpha \\ \delta \\ \theta \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \tag{18}$$

The Gibbs sampler that uses $(\beta\,|\,\Sigma)$ and $(\Sigma\,|\,\beta)$, now with $\beta' = (\alpha', \delta', \theta')$, can proceed as before with *X* redefined as

$$X = \begin{pmatrix} X_{11} & 0 & X_{12} \\ 0 & X_{21} & X_{22} \end{pmatrix}$$

However, it is no longer straightforward to set up a Gibbs sampler by first integrating out Σ, and then considering the conditional posterior pdfs for $(\alpha \mid \delta, \theta)$, $(\delta \mid \alpha, \theta)$ and $(\theta \mid \alpha, \delta)$. The conditional posterior pdfs for $(\alpha \mid \delta, \theta)$ and $(\delta \mid \alpha, \theta)$ can readily be derived as multivariate *t* distributions by defining 'transformed dependent variables' $y_1^* = y_1 - X_{12}\theta$ and $y_2^* = y_2 - X_{22}\theta$ and proceeding as we did previously to derive pdf's analogous to the one in equation (11). The conditional posterior pdf for the common coefficients $f(\theta \mid \alpha, \delta, y)$ remains a problem, however. It is still of the form $|A|^{-T/2}$ described in equation (8). To set up a Gibbs sampler involving $(\alpha \mid \delta, \theta)$, $(\delta \mid \alpha, \theta)$ and $(\theta \mid \alpha, \delta)$, a Metropolis-Hastings step would be needed to draw from $f(\theta \mid \alpha, \delta, y)$.

2.2 A Set of SURs

To motivate the idea of a set of SURs consider the first of our applications. We are estimating expenditure functions designed to explain household expenditure on several commodity groups. There is an equation for each commodity group and the equations for all commodity groups constitute an SUR model. Now suppose, as expected, our sample of households contains households with varying numbers of adults and children. We use the numbers of adults and children in a household to define a particular *household type*. For example, in our application we consider 8 household types, 1 adult with 0, 1, 2 or 3 children and 2 adults with 0, 1, 2 or 3 children. We are concerned with specifying a different SUR model for each household type. The collection of such SUR models we call a set of SURs.

To specify such a set, we return to equations (3) and (4) and rewrite them as

$$y_{(h)} = X_{(h)}\beta_{(h)} + e_{(h)} \tag{20}$$

$$e_{(h)} \sim N\left(0, \Sigma_h \otimes I_{T_h}\right) \tag{21}$$

for $h = 1, 2, \ldots, H$ where H is the number of SUR models (number of household types) in a set. The numbers of observations can be different in each of the SUR models. We denote them by T_h, $h = 1, 2, \ldots, H$. The i-th equation in the h-th SUR model is written as

$$y_{(h)i} = X_{(h)i}\beta_{(h)i} + e_{(h)i} \tag{22}$$

where $y_{(h)i}$ is $(T_h \times 1)$, $X_{(h)i}$ is $(T_h \times K_i)$, $\beta_{(h)i}$ is $(K_i \times 1)$ and $e_{(h)i}$ is $(T_h \times 1)$. For the complete model in equation (20), $y_{(h)}$ is $(T_h M \times 1)$, $X_{(h)}$ is $(T_h M \times K)$, $\beta_{(h)}$ is $(K \times 1)$ and $e_{(h)}$ is $(T_h M \times 1)$.

The parameters $\beta_{(h)}$ and Σ_h in equations (20) and (21) are different in each of the SUR models, although we vary this assumption in the subsections to follow. If we specify independent non-informative priors for each model,

$$f\left(\beta_{(h)}, \Sigma_{(h)}\right) \propto \left|\Sigma_h\right|^{-(M+1)/2} \qquad h = 1, 2, \ldots, H \tag{23}$$

then, in terms of deriving posterior pdfs for the parameters, each of the SUR models can be treated separately. The Gibbs samplers described in Section 2.1 are relevant. If, however, a set of SUR models has parameters common to each SUR model, then a new set of conditional posterior pdfs is required for Gibbs sampling to proceed. We now consider two such situations. In the first some elements in $\beta_{(h)}$ are common; in the second we also assume the error covariance matrices Σ_h are identical for all h.

<u>A set of SURs with common coefficient vector</u>

The effect of household composition on expenditure patterns is often estimated by specifying a vector of intercepts that is different for each household type, while at the same time assuming a vector of income coefficients that is common to all household types. To accommodate common elements in $\beta_{(h)}$, we partition it as $(\theta'_{(h)}, \eta')'$, and partition $X_{(h)} = (Z_{(h)}, W_{(h)})$ correspondingly, so that the SUR model in equation (20) can be rewritten as

$$y_{(h)} = Z_{(h)}\theta_{(h)} + W_{(h)}\eta + e_{(h)} \tag{24}$$

The *i*-th equation in this SUR model we write as

$$y_{(h)i} = Z_{(h)i}\theta_{(h)i} + W_{(h)i}\eta_i + e_{(h)i} \tag{25}$$

Note that the vector η_i is indexed by *i* but not *h*. Thus, this coefficient vector is different for all equations within a given SUR model, but each SUR model has the same set of η_i, $i = 1, 2, ..., M$. For example, in the household expenditure model that follows, the different equations represent expenditure on different commodity groups and the different SUR models correspond to different household compositions. The coefficient of income is different for each of the commodity groups, but the vector of income coefficients is assumed to be the same for each household composition.

The following notation will be useful for describing the conditional posterior pdf's for this model.

$$\theta = \left\{\theta_{(h)} \mid h = 1, 2, \ldots, H\right\}$$

$$\theta_{-h} = \left\{\theta_{(k)} \mid k \neq h\right\}$$

$$Y = \left\{y_{(h)} \mid h = 1, 2, \ldots, H\right\}$$

$$\Sigma = \left\{\Sigma_h \mid h = 1, 2, \ldots, H\right\}$$

$$\Sigma_{-h} = \left\{\Sigma_k \mid k \neq h\right\}$$

Using the non-informative prior pdfs in equation (23), the following conditional posterior pdfs can be derived. Proofs are given in the appendix.

The conditional posterior pdf for Σ_h is the inverted Wishart distribution

$$f\left(\Sigma_h \mid \theta, \eta, \Sigma_h, Y\right) \propto \left|\Sigma_h\right|^{-(T_h+M+1)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(A_h\Sigma_h^{-1}\right)\right\} \tag{26}$$

$$h = 1, 2, \ldots, H$$

where A_h is an $(M \times M)$ matrix with $(i, j)^{\text{th}}$ element equal to

$$\left[A_h\right]_{ij} = \left(y_{(h)i} - Z_{(h)i}\theta_{(h)i} - W_{(h)i}\eta_i\right)'\left(y_{(h)j} - Z_{(h)j}\theta_{(h)j} - W_{(h)j}\eta_j\right) \tag{27}$$

The conditional posterior pdf for $\left(\theta_{(h)} \mid \theta_{-h}, \eta, \Sigma, Y\right)$ is multivariate normal

$$\left(\theta_{(h)} \mid \theta_{-h}, \eta, \Sigma, Y\right) \sim N\left[\hat{\theta}_{(h)}, \left[Z'_{(h)}\left(\Sigma_h^{-1} \otimes I_{T_h}\right)Z_{(h)}\right]^{-1}\right] \qquad h = 1, 2, \ldots, H \tag{28}$$

where

$$\hat{\theta}_{(h)} = \left[Z'_{(h)}\left(\Sigma_h^{-1} \otimes I_{T_h}\right)Z_{(h)}\right]^{-1} Z'_{(h)}\left(\Sigma_h^{-1} \otimes I_{T_h}\right)\left(y_{(h)} - W_{(h)}\eta\right) \tag{29}$$

In the special case where η is a vector of slope coefficients and $\theta_{(h)}$ are intercept coefficients, $Z_{(h)}$ is a matrix of dummy variables. It can be written as $Z_{(h)} = I_M \otimes \iota_{T_h}$, where ι_{T_h} is a T_h - dimensional vector of ones. Equations (28) and (29) simplify to

$$\left(\theta_{(h)} \mid \theta_{-h}, \eta, \Sigma, Y\right) \sim N\left[\hat{\theta}_{(h)}, T_h^{-1}\Sigma_h\right] \tag{30}$$

and

$$\hat{\theta}_{(h)} = \left(I_M \otimes T_h^{-1}\iota'_{T_h}\right)\left(y_{(h)} - W_{(h)}\eta\right) \tag{31}$$

respectively.

The remaining conditional posterior pdf is that for the common coefficients η. It is the following multivariate normal pdf

$$\left(\eta \mid \theta, \Sigma, Y\right) \sim N\left(\hat{\eta}, \left(\sum_{h=1}^{H} W'_{(h)}\left(\Sigma_h^{-1} \otimes I_{T_h}\right)W_{(h)}\right)^{-1}\right) \tag{32}$$

where

$$\hat{\eta} = \left[\sum_{h=1}^{H} W'_{(h)}\left(\Sigma_h^{-1} \otimes I_{T_h}\right)W_{(h)}\right]^{-1} \sum_{h=1}^{H} W'_{(h)}\left(\Sigma_h^{-1} \otimes I_{T_h}\right)\left(y_{(h)} - Z_{(h)}\theta_{(h)}\right) \tag{33}$$

The conditional posterior pdf's in equations (26) through (33) can be readily used in a Gibbs' sampling algorithm.

In Section 2.1 we showed that, for a single SUR model, it was possible to integrate out the error covariance matrix Σ, and then set up a Gibbs' sampler involving multivariate *t* conditional posterior pdfs for the coefficients β_i from each equation. In the Appendix we investigate the consequences of a similar strategy for the model defined in equation (24) where we have a set of SURs with a common coefficient. We find that the conditional posterior pdfs for the $\theta_{(h)i}$ are multivariate *t* pdfs and those for the η_i are poly-*t* pdfs.

<u>A set of SURs with common coefficient vector and common error covariance matrix</u>

Consider again the model in equation (24)

$$y_{(h)} = Z_{(h)}\theta_{(h)} + W_{(h)}\eta + e_{(h)} \tag{34}$$

but this time with a common error covariance matrix Σ. That is,

$$e_{(h)} \sim N\left(0, \Sigma \otimes I_{T_h}\right) \tag{35}$$

We use the non-informative prior pdf

$$f\left(\theta, \eta, \Sigma\right) \propto \left|\Sigma\right|^{-(M+1)/2} \tag{36}$$

Defining $T = \sum_{h=1}^{H} T_h$ and $S = \sum_{h=1}^{H} A_h$, the conditional posterior pdf for Σ is the inverted Wishart pdf,

$$f\left(\Sigma \mid \theta, \eta, Y\right) \propto \left|\Sigma\right|^{-(T+M+1)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(S\Sigma^{-1}\right)\right\} \tag{37}$$

The conditional posterior pdfs for $\theta_{(h)}$ and η are identical to those in equations (28) and (32) except that Σ_h is replaced by Σ. These results follow immediately from the earlier ones after noting that the joint posterior pdf is

$$f\left(\theta,\eta,\Sigma\,|\,Y\right)\propto\left|\Sigma\right|^{-(T+M+1)/2}\exp\left\{-\frac{1}{2}\sum_{h=1}^{H}\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\eta\right)'\left(\Sigma^{-1}\otimes I_{T_h}\right)\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\eta\right)\right\}$$

$$=\left|\Sigma\right|^{-(T+M+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(S\Sigma^{-1}\right)\right\} \tag{38}$$

Also, it is worth noting that, with these assumptions, this set of SURs could be treated as a single SUR model. We can accommodate the parameters that are not common $\theta_{(h)}$ by defining suitable dummy variables and then incorporating the products of the $Z_{(h)}$ and the dummy variables into the matrix $W_{(h)}$.

Furthermore, it follows that we can integrate Σ out from equation (38) and then set up a Gibbs' sampler involving multivariate *t* conditional posterior pdfs for $\theta_{(h)i}$ and η_i. These pdf's can be found from suitable partitionings of the posterior pdf $f\left(\theta,\eta\,|\,Y\right)\propto\left|S\right|^{-T/2}$. However, if some coefficients are common to more than one equation in each SUR model, a convenient partitioning of *S* that permits derivation of multivariate *t* student pdfs for all the coefficients cannot be found.

3. Estimation of Household Equivalence Scales

In the context of the notation introduced earlier, the first model we estimate can be written as a special case of equation (25). It is given by

$$y_{(h)i}=\theta_{(h)i}+\eta_i w_{(h)}+e_{(h)} \tag{39}$$

where $y_{(h)i}$ is a vector of expenditures on the ith commodity group by households of type *h* and $w_{(h)}$ is a vector of observations on income for households of type *h*. The intercept parameter $\theta_{(h)i}$ is assumed to vary over both commodity groups and household types. The slope coefficient η_i, describing the response of expenditure to changes in income, is assumed to be different for different commodity groups, but the same for different household types. Using Australian household expenditure survey data, Griffiths and Valenzuela (2004) obtained maximum likelihood estimates of the parameters of this model and of functions of the parameters called equivalence scales. The model is known as an extended linear expenditure system (Lluch 1973, Kakwani

1977); it is derived from the so-called Klein-Rubin utility function where consumption of each commodity must exceed a subsistence level to generate utility, and a micro-consumption function that relates total expenditure to income. To accommodate the fact that larger households (those with more adults and/or children) need to consume more of a commodity before a subsistence level is reached, commodity-specific equivalence scales are included in the utility function as deflators of quantity consumed. The ith commodity's equivalence scale for household type *h* can be viewed as the proportional increase in consumption of that commodity that is needed for subsistence, relative to the consumption needed by a reference household. Let this commodity equivalence scale be denoted by ϕ_{ih}. By definition, the equivalence scale for a reference household *r* is given by $\phi_{ir} = 1$. In our study a reference household is taken as one with two adults and no children. Thus, for households with one adult and no children we would expect $\phi_{ih} < 1$. For those with two adults and a positive number of children, we would expect $\phi_{ih} > 1$. Including equivalence scales in this way leads to intercepts that vary over household types, but income coefficients that do not, as represented in equation (39). In terms of the parameters of that equation, it can be shown that

$$\phi_{ih} = \frac{\theta_{(h)i}\left(1 - \sum_{j=1}^{M} \eta_j\right) + \eta_i \sum_{j=1}^{M} \theta_{(h)j}}{\theta_{(r)i}\left(1 - \sum_{j=1}^{M} \eta_j\right) + \eta_i \sum_{j=1}^{M} \theta_{(r)j}} \tag{40}$$

where *M* is the number of commodity groups. Details of the utility function and the derivation of equation (40) can be found in Griffiths and Valenzuela (2004).

In addition to the commodity-specific equivalence scales, interest often centres on a general equivalence scale for household type *h*, defined as the proportional increase in income necessary to make the utility of that household equal to the utility level of the reference household. Denoting this general scale by ϕ_h, it is possible to show that (Griffiths and Valenzuela 2004)

$$\phi_h = \left(\frac{\sum_{j=1}^{M} \theta_{(h)j}}{w_r \left(1 - \sum_{j=1}^{M} \eta_j \right)} \right) + \left(1 - \frac{\sum_{j=1}^{M} \theta_{(r)j}}{w_r \left(1 - \sum_{j=1}^{M} \eta_j \right)} \right) \prod_{i=1}^{M} \phi_{ih}^{b_i} \tag{41}$$

where $b_i = \eta_i / \sum_{j=1}^{M} \eta_j$ and w_r is the income level of the reference household. Equivalence scales are important devices for assessing the relative costs of living for households of different compositions. They are used by governments for assessing tax policies and welfare payments. For example, a comparison of equivalence scales for households with and without children is a popular means of representing the costs of raising children. Such costs are used for deciding the level of government support in income maintenance programs.

A sample of 5532 households was drawn from the 1988-89 Household Expenditure Survey conducted by the Australian Bureau of Statistics. Eight types of households were considered, those with 1 adult and 0, 1, 2 and 3 children and 2 adults and 0, 1, 2 and 3 children. Expenditure was divided into 11 different commodity groups. Thus, in terms of our earlier notation, there are $H = 8$ different SUR models, each with $M = 11$ equations. The 11 different expenditure categories are listed in the first column of Table 1. The numbers of observations on each household type, $T_1, T_2, ..., T_8$, are given in the first row in the body of Table 1. The Gibbs sampler given by the conditional posterior pdfs in equations (26), (30) and (32) was used to generate 18,000 observations on the parameters $\theta_{(h)i}$, η_i and Σ_h. No convergence problems were encountered. After discarding 3,000 observations for a burn in, the remaining 15,000 were used to estimate the posterior pdfs of the equivalence scales defined in equations (40) and (41). The general scales ϕ_h depend on a chosen level of income for the reference household, w_r. Rather than condition on one value of w_r, for the posterior pdfs for the general scales we used the empirical distribution of w_r to marginalise over this variable. Symbolically, the resulting posterior pdf for the general scale can be written as

$$f(\phi_h \mid Y) = \int_{w_r > a_r} f(\phi_h \mid w_r, Y) \, f(w_r) \, dw_r \tag{42}$$

In each Gibbs sampling iteration, a value for w_r was drawn from its empirical distribution and used along with the current drawings of the parameters to compute a value for ϕ_h. If a draw for w_r was such that income was less than subsistence expenditure, another draw for w_r was taken until the inequality $w_r > a_r$ was satisfied, where a_r is total subsistence expenditure given by

$$a_r = \frac{\sum_{j=1}^{M} \theta_{(r)j}}{1 - \sum_{j=1}^{M} \eta_j} \tag{43}$$

The estimated posterior means and standard deviations for the commodity-specific and general equivalence scales appear in Table 1, and the complete posterior pdfs for the scales for food, clothing and housing are graphed in Figures 1, 2 and respectively. The posterior pdfs for the general scales appear in Figure 4. The relative magnitudes of the various scales are generally as expected. For example, the posterior means for the first three commodity groups in Table 1 suggest that the arrival of a child in a household of two adults increases housing costs by 49%, fuel & power costs by 22% and food costs by 24%. The arrival of two children increases costs further, but by lesser amounts. For example, relative to a childless couple, the cost increases are 51%, 34% and 42% for housing, fuel & power and food, respectively. Housing costs for households with one adult are 82% of those for two adults while the corresponding figures for fuel & power and food are 67% and 53%, respectively. The estimated scales for alcohol and tobacco decline as the number of children in the household increases. It seems the presence of children in the household tends to influence expenses away from so-called ‘adult goods’.

The relationships between the scales for different household types and their precision of estimation are seen clearly in the posterior pdfs graphed in Figures 1-4. The food scales are the most precisely estimated with distinct pdfs that have little overlap except for that from the pdf for (1,3)-type households. In general, estimation of scales for (1,3)-type households is relatively imprecise because of the smaller number of observations in this category. At the other end of the spectrum, the posterior pdfs for scales for (1,0)-type households are very sharp, reflecting the relatively large number of observations in this category. The scales for housing and clothing are less precisely estimated than those for food, and there is considerable

overlap of the pdfs for different household types. Nevertheless, their location shifts to the right as expected when larger numbers of household members are present. Similar remarks can be made about the pdfs for the general scales that appear in Figure 4; in this case the increased cost of running a 2-adult household relative to a 1-adult household is quite distinct, irrespective of the number of children present in the household.

4. Estimating a Translog Cost Function

The second application is a modification of the translog cost function and share equations estimated by O'Donnell and Woodland (1995) and Griffiths, et.al. (2000) for the Australian merino wool growing section. These studies consider the constant returns to scale functional form

$$\ln\left(\frac{C}{q}\right) = \alpha_0 + \alpha_T t_T + \sum_{i=1}^{I} \alpha_i \ln w_i + 0.5 \sum_{i=1}^{I}\sum_{j=1}^{I} \alpha_{ij} \ln(w_i)\ \ln(w_j) \tag{44}$$

where C represents total costs, q denotes output, w_i represents the price of input i, t_T is a time trend used to capture the effects of exogenous technical change, and I is the number of inputs. The factor share equations obtained using Shephard's lemma are

$$s_i = \alpha_i + \sum_{j=1}^{I} \alpha_{ij} \ln w_j \qquad i = 1, 2, ..., I \tag{45}$$

where s_i represents the cost share of input i. It is clear from these equations that our assumed form of technical change is Hicks-neutral; factor shares are unaffected by technical change while unit cost decreases at a constant percentage rate.

Our sample consists of 310 time-series and cross section observations on Australian merino wool growers, over periods 1952-53 to 1962-63 ($t_T = 1, 2, ...11$) and 1964-65 to 1975-76 ($t_T = 13, 14, ..., 24$). Thus, data are available for a total of 23 years. Each observation in the original data set is a record of the average financial and physical characteristics for a group of firms. The number of groups of firms in each year varies, but it is equal to 12, 13 or 14. The data were used to construct observations on output (q), total cost (C), input prices (w) and input quantities. Inputs

were grouped into one of four broad categories: land, capital, livestock and other inputs (including labour, equipment, materials and services). A more complete description of the data can be found in O'Donnell and Woodland (1995).

The modification of the above model considered here is a replacement of the intercept and trend term in the cost function $(\alpha_0 + \alpha_T t_T)$ with a time-varying intercept $\theta_{(h)}$. The subscript h is in line with the notation used earlier. In this case, h refers to the hth year. Having a different intercept for each year continues the assumption of Hicks-neutral technical change, but no longer imposes the restriction that unit costs decrease at a constant percentage rate. In the context of the discussion in Section 2, the cost function and 3 share equations make up an SUR model $(M = 4)$, we have a different SUR model for each year $(H = 23)$, and the collection of SUR models for all years we call a set of SUR models. The number of observations available in each year (T_h in our earlier notation) varies between 12 and 14. The intercept in the cost function is different for each model, but the other coefficients, the α_i and α_{ij}, are assumed to be constant over time and hence the same for each SUR model. Since getting a reliable estimate of a (4 × 4) error covariance matrix with only 12 or 14 observations is somewhat ambitious, we assume identical error covariance matrices for each year, except for the following qualification. Each observation is based on an average from a group of firms. The number of firms used to create each average varies. To accommodate the fact that averages for large groups are more reliable than averages from small groups, we assume that the (4×4) error covariance matrix is given by Σ / N_{hj} where N_{hj} is the number of firms used to form the average for the jth observation $(j = 1, 2, ..., T_h)$ in the hth year $(h = 1, 2, ..., 23)$. After multiplying the $(h, j)^{\text{th}}$ observation on all variables by $\sqrt{N_{hj}}$, the covariance matrix for the 4-dimensional transformed error term is Σ.

Before making these assumptions explicit, some rearranging of the variables and the matrices defined earlier is necessary to impose homogeneity and symmetry restrictions and to allow for the fact that the same α_i and α_{ij} appear in more than one equation. Linear homogeneity and symmetry will be satisfied if

$$\sum_{j=1}^{I} \alpha_i = 1 \tag{46}$$

$$\sum_{j=1}^{I} \alpha_{ij} = 0 \qquad (i = 1,...,I) \tag{47}$$

$$\alpha_{ij} = \alpha_{ji} \qquad (i, j = 1,...,I)\,. \tag{48}$$

Introducing these restrictions, defining appropriately transformed variables and adding error terms leads to the following system of four equations (note the equation for the fourth share is redundant)

$$\begin{bmatrix} y \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \theta_h + \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \\ 1 & 0 & 0 & x_1 & 0 & 0 & x_2 & x_3 & 0 \\ 0 & 1 & 0 & 0 & x_2 & 0 & x_1 & 0 & x_3 \\ 0 & 0 & 1 & 0 & 0 & x_3 & 0 & x_1 & x_2 \end{bmatrix} \eta + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \tag{49}$$

where $\eta = (\alpha_1, \alpha_2, \alpha_3, \alpha_{11}, \alpha_{22}, \alpha_{33}, \alpha_{12}, \alpha_{13}, \alpha_{23},)'$, $x_1 = \ln\left(w_1 / w_4\right)$, $x_2 = \ln\left(w_2 / w_4\right)$, $x_3 = \ln\left(w_3 / w_4\right)$, $y = \ln\left(C / q\right) - \ln w_4$,

$$x_4 = .5(\ln w_1)^2 + .5(\ln w_4)^2 - (\ln w_1)(\ln w_4)$$

$$x_5 = .5(\ln w_2)^2 + .5(\ln w_4)^2 - (\ln w_2)(\ln w_4)$$

$$x_6 = .5(\ln w_3)^2 + .5(\ln w_4)^2 - (\ln w_3)(\ln w_4)$$

$$x_7 = (\ln w_1)(\ln w_2) - (\ln w_1)(\ln w_4) - (\ln w_2)(\ln w_4) + (\ln w_4)^2$$

$$x_8 = (\ln w_1)(\ln w_3) - (\ln w_1)(\ln w_4) - (\ln w_3)(\ln w_4) + (\ln w_4)^2$$

$$x_9 = (\ln w_2)(\ln w_3) - (\ln w_2)(\ln w_4) - (\ln w_3)\ln(\ln w_4) + (\ln w_4)^2$$

If the variables $(y_i, s_i, 1, x_i)$ in equation (49) are vectors of length T_h, containing all observations for the hth year, then this equation fits the structure of equation (24), and can be written as

$$y_{(h)} = Z_{(h)}\theta_{(h)} + W_{(h)}\eta + e_{(h)} \tag{50}$$

However, our assumption about the nature of the heteroskedasticity induced by the different numbers of firms used to construct each observation means that the covariance matrix for $e_{(h)}$ will be given by

$$E\left(e_{(h)}e'_{(h)}\right) = \Sigma \otimes N_{(h)} \tag{51}$$

where $N_{(h)}$ is the diagonal matrix $N_{(h)} = \text{diag}(N_{h1}, N_{h2}, ..., N_{hT_h})$. To overcome this problem we apply a Gibbs' sampler to the transformed model

$$y^*_{(h)} = Z^*_{(h)}\theta_{(h)} + W^*_{(h)}\eta + e^*_{(h)} \tag{52}$$

where $y^*_{(h)} = (I_{T_h} \otimes N^{-1/2}_{(h)})y_{(h)}$, $Z^*_{(h)} = (I_{T_h} \otimes N^{-1/2}_{(h)})Z_{(h)}$, $X^*_{(h)} = (I_{T_h} \otimes N^{-1/2}_{(h)})X_{(h)}$, $e^*_{(h)} = (I_{T_h} \otimes N^{-1/2}_{(h)})e_{(h)}$ and $N^{-1/2}_{(h)} = \text{diag}(N^{-1/2}_{h1}, N^{-1/2}_{h2}, ..., N^{-1/2}_{hT_h})$. Note that the covariance matrix for the transformed error term in (52) is $E\left(e^*_{(h)}e^{*\prime}_{(h)}\right) = \Sigma \otimes I_{T_h}$.

The Gibbs' sampler described in and below equation (37) was used to generate 23000 observations with 3000 discarded as a burn in. The posterior means and standard deviations for the θ_h, α_i and α_{ij} appear in Table 2. The posterior pdfs of some economic quantities of interest are graphed in Figures 5, 6 and 7. Figure 5 contains the posterior pdfs of four technical change coefficients that approximately equidistant in time, $\theta_2, \theta_9, \theta_{15}$ and θ_{22}. If the assumption of a constant percentage decline in unit costs is a reasonable one, the four posterior pdfs for the θ_h should be approximately equidistant. What we find is that the posterior pdfs for θ_9 and θ_{15} are approximately the same, and those for θ_2 and θ_{22} are approximately equidistant on each side of those for θ_9 and θ_{15}. The assumption has some merit, but may be too restrictive when all years are considered.

Figure 6 contains the posterior pdfs for the input demand elasticities for capital (η_2) and livestock (η_3) evaluated at the input means. These quantities are defined as

$$\eta_i = \frac{\alpha_{ii}}{s_i} + s_i - 1 \qquad i = 2,3$$

where the shares s_i are also evaluated at the input means. The capital elasticity is much larger (in absolute value) and more precisely estimated than that for labour. The posterior pdf for the labour elasticity suggests its value could be positive or negative, although the posterior probability of a (realistic) negative value is still high. Finally, in Figure 7 we plot two quite different posterior pdfs for two input substitution elasticities, that for land and capital (ε_{12}) and that for capital and livestock (ε_{23}). These elasticities are defined as

$$\varepsilon_{ij} = \frac{\alpha_{ij}}{s_i s_j} + 1 \qquad i \neq j$$

We find that ε_{12} is precisely estimated and relatively large whereas there is considerable uncertainty about the value of ε_{23}.

5. Concluding Remarks

Bayesian estimation of a set of seemingly unrelated regression models that are linked by common coefficients and/or a common covariance matrix has been considered. Building on alternative Gibbs' samplers that can be used for the case of a single SUR model, we explored various alternatives that can be employed when we have several models with common parameters. Samplers that involve the error covariance matrix as well as those obtained after integrating out this covariance matrix were considered. The results from two examples of applications were presented. The first was designed to obtain posterior pdfs for commodity and general household equivalence scales that are functions of the parameters in a linear expenditure system. In the second, posterior pdfs describing the changing impact of technical change in a translog cost function, and input demand and substitution elasticites were estimated.

Appendix

Derivation of the conditional posterior pdf for $f(\beta_i \mid \beta_{-i}, y)$ *given in equation (11)*

The starting point for derivation of the conditional posterior pdf $f(\beta_i \mid \beta_{-i}, y)$ is the marginal posterior pdf for the complete vector β

$$f\left(\beta \mid y\right) \propto \left|A\right|^{-T/2} = \left|E'E\right|^{-T/2} \tag{A1}$$

where

$$E = \left[(y_1 - X_1\beta_1),\ (y_2 - X_2\beta_2), \ldots, (y_M - X_M\beta_M)\right] \tag{A2}$$

For convenience, suppose the columns of E are reordered such that the i-th column comes first, followed by the remaining $(M-1)$ columns, such that we can partition E as

$$E = [(y_i - X_i\beta_i),\ E_{-i}] \tag{A3}$$

Then,

$$A = \begin{bmatrix} (y_i - X_i\beta_i)'(y_i - X_i\beta_i) & (y_i - X_i\beta_i)'E_{-i} \\ E'_{-i}(y_i - X_i\beta_i) & E'_{-i}E_{-i} \end{bmatrix} \tag{A4}$$

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

$$\left|A\right| = \left|E'_{-i}E_{-i}\right| \left((y_i - X_i\beta_i)'(y_i - X_i\beta_i) - (y_i - X_i\beta_i)'E_{-i}(E'_{-i}E_{-i})^{-1}E'_{-i}(y_i - X_i\beta_i)\right) \tag{A5}$$

Defining $Q_i = I_T - E_{-i}(E'_{-i}E_{-i})^{-1}E'_{-i}$ and $\tilde{\beta}_i = (X'_iQ_iX_i)^{-1}X'_iQ_iy_i$, the second term on the right side of the above equation can be written as

$$\begin{aligned} &(y_i - X_i\beta_i)'(y_i - X_i\beta_i) - (y_i - X_i\beta_i)'E_{-i}(E'_{-i}E_{-i})^{-1}E'_{-i}(y_i - X_i\beta_i) \\ &\qquad = (y_i - X_i\beta_i)'Q_i(y_i - X_i\beta_i) \\ &\qquad = (y_i - X_i\tilde{\beta}_i)'Q_i(y_i - X_i\tilde{\beta}_i) + (\beta_i - \tilde{\beta}_i)'X'_iQ_iX_i(\beta_i - \tilde{\beta}_i) \end{aligned} \tag{A6}$$

Substituting these results into the equation $f\left(\beta \mid y\right) \propto \left|A\right|^{-T/2} = \left|E'E\right|^{-T/2}$, and letting $\left|E'_{-i}E_{-i}\right|$ be absorbed into the proportionality constant, we can write

$$f\left(\beta_i \mid \beta_{i-1}, y\right) \propto [v_i \tilde{s}_i^2 + (\beta_i - \tilde{\beta}_i)' X_i' Q_i X_i (\beta_i - \tilde{\beta}_i)]^{-(K_i+v_i)/2}$$

$$\propto \left[v_i + \frac{(\beta_i - \tilde{\beta}_i)' X_i' Q_i X_i (\beta_i - \tilde{\beta}_i)}{\tilde{s}_i^2} \right]^{-(K_i+v_i)/2} \tag{A7}$$

where $v_i = T - K_i$ and $\tilde{s}_i^2 = (y_i - X_i\tilde{\beta}_i)' Q_i (y_i - X_i\tilde{\beta}_i)/v_i$. This pdf is in the form of a multivariate *t*-distribution with degrees of freedom v_i, location vector $\tilde{\beta}_i$, and inverse precision matrix $\tilde{s}_i^2 (X_i' Q_i X_i)^{-1}$.

Derivation of conditional posteriors for Σ_h, $\theta_{(h)}$ *and* η *in equations (26), (28), (32)*

The complete likelihood function for all SUR models in the set is given by

$$f\left(Y \mid \theta, \Sigma, \eta\right) \propto \prod_{h=1}^{H} \left[\left|\Sigma_h\right|^{-T_h/2} \exp\left\{ -\frac{1}{2}\left(y_{(h)} - Z_{(h)}\theta_{(h)} - W_{(h)}\eta\right)' \left(\Sigma_h^{-1} \otimes I_{T_h}\right)\left(y_{(h)} - Z_{(h)}\theta_{(h)} - W_{(h)}\eta\right) \right\} \right]$$

$$= \prod_{h=1}^{H} \left[\left|\Sigma_h\right|^{-T_h/2} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(A_h \Sigma_h^{-1}\right) \right\} \right] \tag{A8}$$

from which we obtain a joint posterior pdf given by

$$f\left(\theta, \eta, \Sigma \mid Y\right) \propto \prod_{h=1}^{H} \left[\left|\Sigma_h\right|^{-(T_h+M+1)/2} \exp\left\{ -\frac{1}{2}\left(y_{(h)} - Z_{(h)}\theta_{(h)} - W_{(h)}\eta\right)' \left(\Sigma_h^{-1} \otimes I_{T_h}\right)\left(y_{(h)} - Z_{(h)}\theta_{(h)} - W_{(h)}\eta\right) \right\} \right]$$

$$= \prod_{h=1}^{H} \left[\left|\Sigma_h\right|^{-(T_h+M+1)/2} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(A_h \Sigma_h^{-1}\right) \right\} \right] \tag{A9}$$

It is immediately clear that the conditional posterior pdf for Σ_h is given by

$$f\left(\Sigma_h \mid \theta, \eta, Y\right) \propto \left|\Sigma_h\right|^{-(T_h+M+1)/2} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(A_h \Sigma_h^{-1}\right) \right\} \tag{A10}$$

For the conditional posterior pdf for $\theta_{(h)}$, we note that

$$f\left(\theta_{(h)} \mid \theta_{-h}, \eta, \Sigma, Y\right) \propto \exp\left\{ -\frac{1}{2}\left(y_{(h)} - Z_{(h)}\theta_{(h)} - W_{(h)}\eta\right)' \left(\Sigma_h^{-1} \otimes I_{T_h}\right)\left(y_{(h)} - Z_{(h)}\theta_{(h)} - W_{(h)}\eta\right) \right\}$$

The term in the exponent of this pdf can be written as

$$\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\eta\right)'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\eta\right)$$
$$=(\theta_{(h)}-\hat{\theta}_{(h)})'Z_{(h)}'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)Z_{(h)}(\theta_{(h)}-\hat{\theta}_{(h)}) \tag{A11}$$
$$+(y_{(h)}-Z_{(h)}\hat{\theta}_{(h)}-W_{(h)}\eta)'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)(y_{(h)}-Z_{(h)}\hat{\theta}_{(h)}-W_{(h)}\eta)$$

where

$$\hat{\theta}_{(h)}=\left[Z_{(h)}'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)Z_{(h)}\right]^{-1}Z_{(h)}'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)\left(y_{(h)}-W_{(h)}\eta\right)$$

Absorbing the exponent of the second term in (A11) into the factor of proportionality, we have

$$f\left(\theta_{(h)}\,|\,\theta_{-h},\eta,\Sigma,Y\right)\propto\exp\left\{-\frac{1}{2}\left(\theta_{(h)}-\hat{\theta}_{(h)}\right)'Z_{(h)}'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)Z_{(h)}\left(\theta_{(h)}-\hat{\theta}_{(h)}\right)\right\} \tag{A12}$$

which is the normal distribution specified in equation (28).

To recognize the form of the conditional posterior pdf for η, we first note that, from (A9), this pdf can be written as

$$f\left(\eta\,|\,\theta,\Sigma,Y\right)\propto\exp\left\{-\frac{1}{2}\sum_{h=1}^{H}\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\eta\right)'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\eta\right)\right\} \tag{A13}$$

The term in the exponent can be written as

$$\sum_{h=1}^{H}\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\eta\right)'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\eta\right)$$
$$=\sum_{h=1}^{H}\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\hat{\eta}\right)'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)\left(y_{(h)}-Z_{(h)}\theta_{(h)}-W_{(h)}\hat{\eta}\right) \tag{A14}$$
$$+\left(\eta-\hat{\eta}\right)'\left[\sum_{h=1}^{H}W_{(h)}'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)W_{(h)}\right]\left(\eta-\hat{\eta}\right)$$

where

$$\hat{\eta}=\left[\sum_{h=1}^{H}W_{(h)}'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)W_{(h)}\right]^{-1}\sum_{h=1}^{H}W_{(h)}'\left(\Sigma_h^{-1}\otimes I_{T_h}\right)\left(y_{(h)}-Z_{(h)}\theta_{(h)}\right) \tag{A15}$$

Absorbing the exponent of the first term in (A14) into the proportionality constant in (A13) yields the multivariate normal distribution given in equation (32), namely

$$f\left(\eta \mid \theta, \Sigma, Y\right) \propto \exp\left\{-\frac{1}{2}\left(\eta-\hat{\eta}\right)'\left[\sum_{h=1}^{H} W_{(h)}'\left(\Sigma_h^{-1} \otimes I_{T_h}\right) W_{(h)}\right]\left(\eta-\hat{\eta}\right)\right\} \tag{A16}$$

Derivation of conditional posterior pdfs after integrating out the Σ_h

Integrating the Σ_h out of the joint posterior pdf in equation (A9) yields

$$f\left(\theta, \eta \mid Y\right) \propto \prod_{h=1}^{H} \left|A_h\right|^{-T_h/2} \tag{A17}$$

It is interesting to ask what happens if we use (A17) to derive conditional posterior pdfs for the $\theta_{(h)}$ and η.

If we write the model as

$$y_{(h)}^* = Z_{(h)}\theta_{(h)} + e_{(h)} \tag{A18}$$

where

$$y_{(h)}^* = y_{(h)} - W_{(h)}\eta \tag{A19}$$

then the derivation in equations (A1) to (A7) holds separately for each of the equations in each of the SUR models, providing we replace y_i by $y_{(h)i}^*$ and β_i by $\theta_{(h)i}$. Thus, the conditional posterior pdf for $\theta_{(h)i}$ is the multivariate *t* distribution

$$f\left(\theta_{(h)i} \mid \theta_{(h)-i}, \eta, Y\right) \propto \left[v_{hi}^* + \frac{\left(\theta_{(h)i} - \tilde{\theta}_{(h)i}\right)' Z_{(h)i}' Q_{(h)i} Z_{(h)i}\left(\theta_{(h)i} - \tilde{\theta}_{(h)i}\right)}{\tilde{s}_{(h)i}^2}\right]^{-(K_i^* + v_{hi}^*)/2} \tag{A20}$$

where, using obvious extensions of earlier notation,

$$Q_{(h)i} = I_{T_h} - E_{(h)-i}\left(E_{(h)-i}' E_{(h)-i}\right)^{-1} E_{(h)-i}' \tag{A21}$$

$$\tilde{\theta}_{(h)i} = \left(Z'_{(h)i}Q_{(h)i}Z_{(h)i}\right)^{-1} Z'_{(h)i}Q_{(h)i}\left(y_{(h)i} - W_{(h)i}\eta_i\right) \tag{A22}$$

$$\tilde{s}^2_{(h)i} = \left(y_{(h)i} - Z_{(h)i}\tilde{\theta}_{(h)i} - W_{(h)i}\eta_i\right)' Q_{(h)i}\left(y_{(h)i} - Z_{(h)i}\tilde{\theta}_{(h)i} - W_{(h)i}\eta_i\right)/v^*_{hi} \tag{A23}$$

and $v^*_{hi} = T_h - K^*_i$, where K^*_i is the dimension of $\theta_{(h)i}$. Equation (A20) holds for all equations in all SUR models and hence represents MH conditional posterior pdfs.

For the conditional posterior pdf for η_i (the vector of coefficients of $W_{(h)i}$, assumed different for each equation, but the same for the *i*-th equation in all SUR models), we write the *i*-th equation in the *h*-th model as

$$y^0_{(h)i} = W_{(h)i}\eta_i + e_i \tag{A24}$$

where

$$y^0_{(h)i} = y_{(h)i} - Z_{(h)i}\theta_{(h)i} \tag{A25}$$

The derivation in equations (A1) to (A7) can again be applied to each of the equations in each of the SUR models; this time with y_i replaced by $y^0_{(h)i}$ and β_i replaced by η_i. However, because η_i appears in each of the terms $|A_h|$ in equation (A17), the resulting pdf will be the product of *H* multivariate-*t* kernels. Specifically,

$$f\left(\eta_i \mid \eta_{-i}, \theta, Y\right) \propto \prod_{h=1}^{H}\left[v^0_{hi} + \frac{\left(\eta_i - \hat{\eta}_{(h)i}\right)' W'_{(h)i}Q_{(h)i}W_{(h)i}\left(\eta_i - \hat{\eta}_{(h)i}\right)}{\hat{s}^2_{(h)i}}\right]^{-\left(K^0_i + v^0_{hi}\right)/2} \tag{A26}$$

where

$$\hat{\eta}_{(h)i} = \left(W'_{(h)i}Q_{(h)i}W_{(h)i}\right)^{-1} W'_{(h)i}Q_{(h)i}\left(y_{(h)i} - Z_{(h)i}\theta_{(h)i}\right) \tag{A27}$$

$$\hat{s}^2_{(h)i} = \left(y_{(h)i} - Z_{(h)i}\theta_{(h)i} - W_{(h)i}\hat{\eta}_{(h)i}\right)' Q_{(h)i}\left(y_{(h)i} - Z_{(h)i}\theta_{(h)i} - W_{(h)i}\hat{\eta}_{(h)i}\right)/v^0_{hi} \tag{A28}$$

and $v^0_{hi} = T_h - K^0_i$ where K^0_i is the dimension of η_i. The pdf in equation (A26) belongs to a class of $H-0$ poly-*t* densities. See, for example, Drèze (1977), Richard and

Tompa (1980) and Bauwens and Richard (1985). Except in special cases, it is not possible to draw directly from such pdfs.

References

Baltagi, B. (2001), *Econometric Analysis of Panel Data, 2nd Edition*, John Wiley and Sons, New York.

Bauwens, L. and J.F. Richard (1985), "A 1-1 Poly-*t* Random Variable Generator with Application to Monte Carlo Integration", *Journal of Econometrics*, 29, 19-46.

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

Drèze, J.H. (1977), "Bayesian Regression Analysis Using Poly-*t* Density Functions", *Journal of Econometrics*, 6, 329-254.

Dreze, J.H. and J.A. Morales (1976), "Bayesian Full Information Analysis of Simultaneous Equations", *Journal of American Statistical Association*, 71, 919-923.

Fiebig, D. (2001), "Seemingly Unrelated Regressions" in Baltagi, B. (Ed.), A Companion to Theoretical Econometrics, Basil Blackwell, London (Chapter 5).

Griffiths, W.E. and D. Chotikapanich (1997), "Bayesian Methodology for Imposing Inequality Constraints on a Linear Expenditure Function with Demographic Factors", *Australian Economic Papers,* 36, 321-341.

Griffiths, W.E., C.J. O'Donnell and A. Tan-Cruz (2000), "Imposing Regularity Conditions on a System of Cost and Factor Share Equations", *The Australian Journal of Agricultural and Resource Economics*, 44, 107-127.

Griffiths, W.E., C.L. Skeels and D. Chotikapanich (2002), "Sample Size Requirements for Estimation in SUR Models", in A. Ullah, A. Wan and A. Chaturvedi, editors, *Handbook of Applied Econometrics and Statistical Inference*, Marcel Dekker, New York (Chapter 37).

Griffiths, W.E. and M.R. Valenzuela (2004), "Estimating Costs of Children from Micro-Unit Records: A New Procedure Applied to Australian Data", *Journal of Quantitative Economics, New Series* 2, 98-120.

Kakwani. N. (1977), "On the Estimation of Consumer Unit Scales", *Review of Economics and Statistics*, 59, 507-510.

Kloek, T. and H.K. van Dijk (1978), "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo", *Econometrica*, 46, 1-19.

Lluch, C. (1973), "The Extended Linear Expenditure System", *European Economic Review*, 4, 21-23.

O'Donnell, C.J. and A.D.Woodland (1995), "Estimation of Australian Wool and Land Production Technologies under Uncertainly: an Error-Components Approach", *American Journal of Agricultural Economics*, 77, 552-65.

Percy, D. (1992), "Predictions 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., K.M Chaloner,.and J.K. Geweke (Eds.), *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner,* John Wiley and Sons, New York (Chapter 17).

Richard, J.-F., and M.F.J. Steel, (1988), "Bayesian Analysis of Systems of Seemingly Unrelated Regression Equations under a Recursive Extended Natural Conjugate Prior Density", *Journal of Econometrics,* 38, 7-37.

Steel, M.F.J. (1992), "Posterior Analysis of Restricted Seemingly Unrelated Regression Equation Models: A Recursive Analytical Approach", *Econometric Reviews*, 11, 129-142.

Richard, J.F. and H. Tompa (1980), "On the Evaluation of Poly-*t* Density Functions", *Journal of Econometrics*, 12, 335-351.

Smith, M. and R. Kohn, (2000) "Nonparametric Seemingly Unrelated Regression," *Journal of Econometrics*, 98, 257-282.

Smith, M. and R. Kohn (2002), "Parsimonious Covariance Matrix Estimation for Longitudinal Data", *Journal of the American Statistical Association*, 97, 1141-1153.

Srivastava, V.K. and T.W. Dwivedi (1979), "Estimation of Seemingly Unrelated Regression Equations: A Brief Survey", *Journal of Econometrics*, 10, 15-32.

Srivastava, V.K. and D.E.A. Giles (1987), *Seemingly Unrelated Regression Equations: Estimation and Inference.* Marcel Dekker, New York.

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.

Table 1. Posterior Means and (Standard Deviations) of Equivalence Scales

	Household Type (no. of adults, no. of children)							
	(1,0)	*(1,1)*	*(1,2)*	*(1,3)*	*(2,0)*	*(2,1)*	*(2,2)*	*(2,3)*
Sample size	1372	132	103	42	2074	532	889	388
				Commodity Specific Scales				
Housing	**0.823**	**1.026**	**1.148**	**1.280**	**1.000**	**1.488**	**1.516**	**1.650**
	(0.040)	(0.084)	(0.096)	(0.225)	-	(0.092)	(0.079)	(0.100)
Fuel & Power	**0.673**	**0.920**	**1.063**	**1.110**	**1.000**	**1.216**	**1.343**	**1.439**
	(0.019)	(0.052)	(0.060)	(0.137)	-	(0.040)	(0.039)	(0.050)
Food	**0.530**	**0.727**	**0.944**	**1.059**	**1.000**	**1.236**	**1.423**	**1.575**
	(0.013)	(0.041)	(0.048)	(0.121)	-	(0.033)	(0.032)	(0.044)
Alcohol & Tobacco	**0.573**	**0.464**	**0.388**	**0.345**	**1.000**	**0.949**	**0.864**	**0.763**
	(0.034)	(0.053)	(0.060)	(0.079)	-	(0.066)	(0.063)	(0.054)
Clothing & Footwear	**0.534**	**0.910**	**0.921**	**1.410**	**1.000**	**1.280**	**1.402**	**1.647**
	(0.049)	(0.156)	(0.156)	(0.356)	-	(0.116)	(0.112)	(0.153)
Household Furnishings & Equipment	**0.548**	**0.664**	**0.773**	**0.814**	**1.000**	**1.451**	**1.152**	**1.325**
	(0.040)	(0.083)	(0.116)	(0.242)	-	(0.138)	(0.094)	(0.111)
Medical & Health Care	**0.538**	**0.468**	**0.676**	**0.507**	**1.000**	**1.261**	**1.280**	**1.313**
	(0.042)	(0.082)	(0.123)	(0.166)	-	(0.084)	(0.057)	(0.071)
Transport	**0.525**	**0.571**	**0.623**	**0.781**	**1.000**	**1.018**	**1.192**	**1.366**
	(0.039)	(0.076)	(0.095)	(0.281)	-	(0.077)	(0.077)	(0.130)
Recreation & Entertainment	**0.538**	**0.576**	**0.513**	**0.827**	**1.000**	**1.029**	**1.285**	**1.369**
	(0.041)	(0.099)	(0.076)	(0.249)	-	(0.105)	(0.115)	(0.146)
Personal Care	**0.542**	**0.781**	**0.972**	**0.734**	**1.000**	**1.194**	**1.291**	**1.176**
	(0.037)	(0.100)	(0.160)	(0.168)	-	(0.095)	(0.080)	(0.093)
Others	**0.567**	**1.024**	**0.890**	**0.809**	**1.000**	**1.397**	**1.794**	**2.083**
	(0.065)	(0.173)	(0.118)	(0.176)	-	(0.142)	(0.190)	(0.272)
General Scales	**0.582**	**0.723**	**0.786**	**0.899**	**1.000**	**1.234**	**1.334**	**1.470**
	(0.022)	(0.053)	(0.056)	(0.154)	-	(0.057)	(0.055)	(0.073)

Table 2. Posterior Means and (Standard Deviations) for θ_h, α_i and α_{ij}

Parameter	*Posterior means & std.devs*	*Parameter*	*Posterior means & std.devs*	*Parameter*	*Posterior means & std.devs*
θ_1	**-0.717** *(0.079)*	θ_{13}	**-0.869** *(0.075)*	α_1	**0.24148** *(0.00567)*
θ_2	**-0.611** *(0.081)*	θ_{14}	**-1.087** *(0.076)*	α_2	**0.67164** *(0.02181)*
θ_3	**-0.595** *(0.081)*	θ_{15}	**-0.919** *(0.075)*	α_3	**0.44929** *(0.01236)*
θ_4	**-0.782** *(0.080)*	θ_{16}	**-1.185** *(0.078)*	α_{11}	**0.02161** *(0.00068)*
θ_5	**-0.794** *(0.077)*	θ_{17}	**-1.169** *(0.080)*	α_{22}	**0.11583** *(0.00682)*
θ_6	**-0.576** *(0.072)*	θ_{18}	**-1.022** *(0.095)*	α_{33}	**0.07679** *(0.00222)*
θ_7	**-0.809** *(0.076)*	θ_{19}	**-1.328** *(0.099)*	α_{12}	**0.01673** *(0.00095)*
θ_8	**-0.918** *(0.076)*	θ_{20}	**-1.135** *(0.080)*	α_{13}	**-0.00555** *(0.00063)*
θ_9	**-0.941** *(0.078)*	θ_{21}	**-1.274** *(0.092)*	α_{23}	**-0.00737** (0.00268)
θ_{10}	**-0.965** *(0.078)*	θ_{22}	**-1.267** *(0.097)*		
θ_{11}	**-1.005** *(0.077)*	θ_{23}	**-1.347** *(0.098)*		
θ_{12}	**-0.953** *(0.075)*				

Figure 1: Posterior Distributions of Food Scales by Household Type

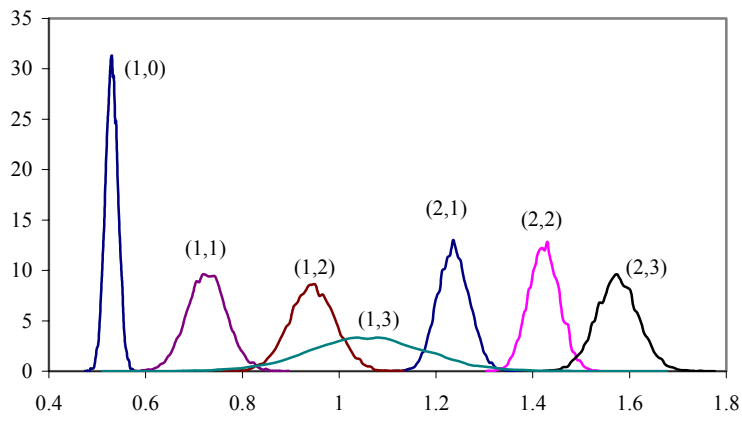


Figure 2: Posterior Distributions of Clothing Scales for each Household Type

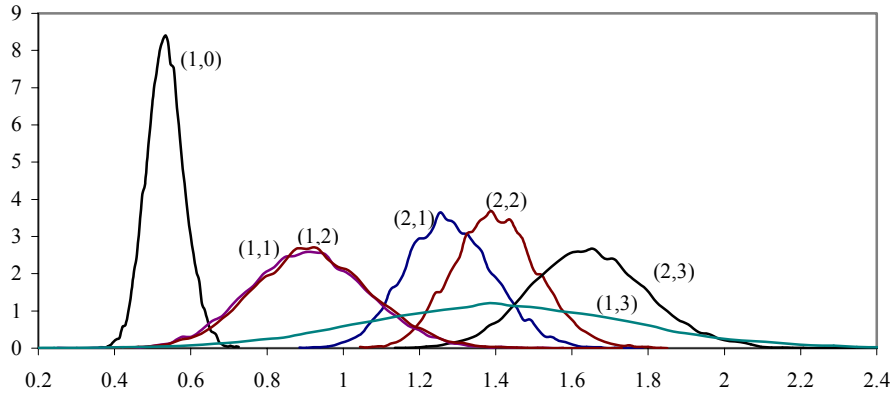


Figure 3: Posterior Distributions for Housing Scales for each Household Type

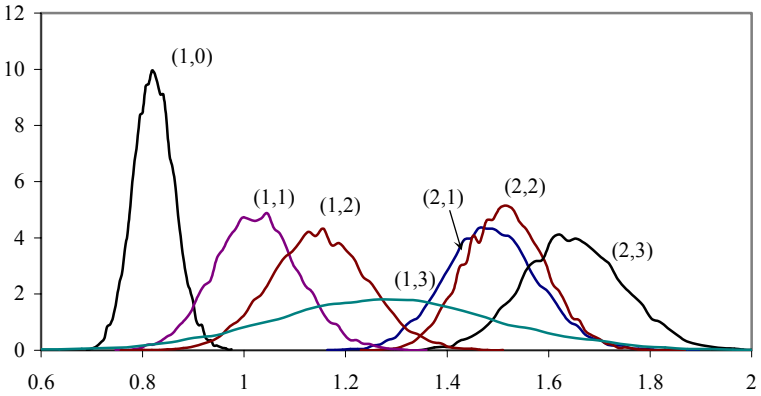


Figure 4: Posterior Distributions of General Scales by Household Type

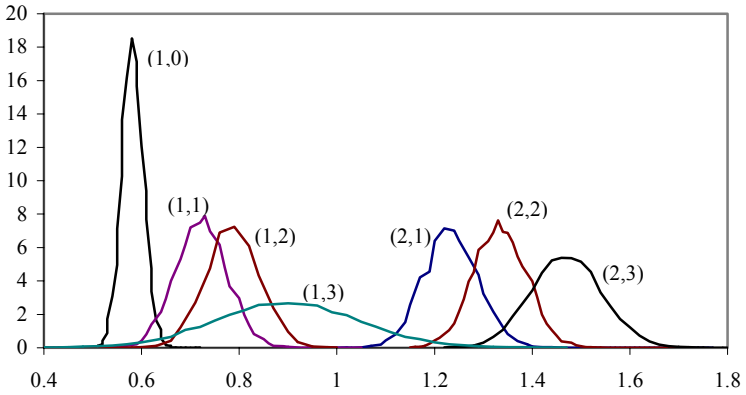


Figure 5. Posterior Distributions for Technical Change Coefficients

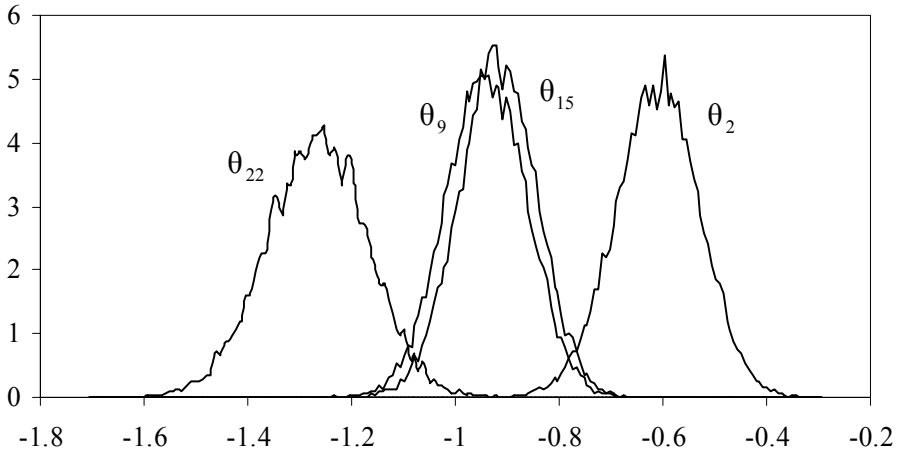


Figure 6. Posterior Distributions for Input Demand Elasticities

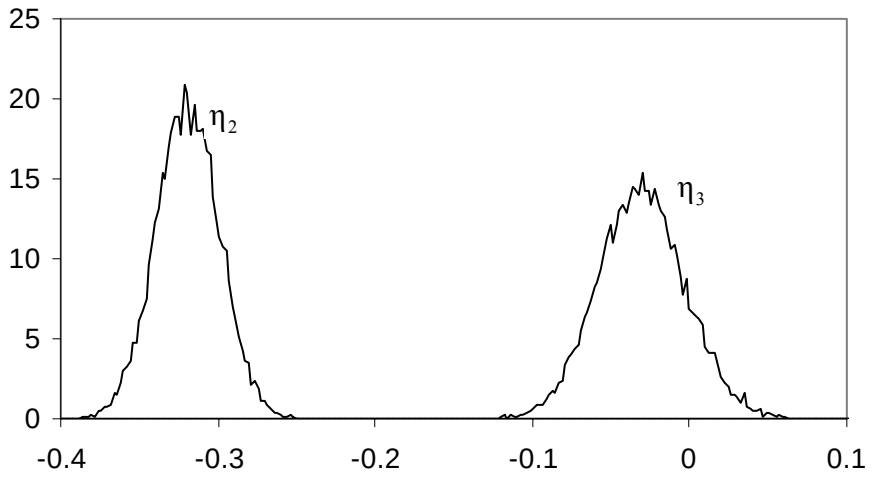
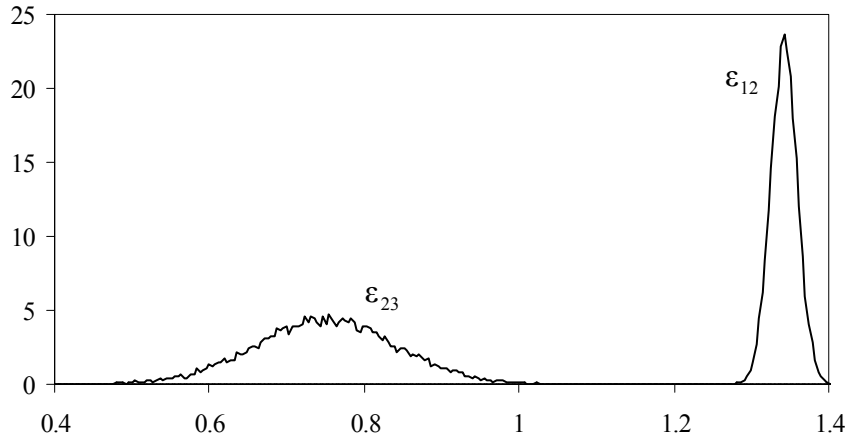


Figure 7. Posterior Distrbutions for Elasticities of Substitution



Author/s:
Griffiths, W. E.;Valenzuela, Ma. Rebecca

Title:
Gibbs samplers for a set of seemingly unrelated regressions

Date:
2004-08

Citation:
Griffiths, W. E. and Valenzuela, Ma. Rebecca (2004) Gibbs samplers for a set of seemingly unrelated regressions.

Persistent Link:
http://hdl.handle.net/11343/34381