Journal of Econometrics 38 (1988) 39–72. North-Holland

BAYESIAN SPECIFICATION ANALYSIS AND ESTIMATION OF SIMULTANEOUS EQUATION MODELS USING MONTE CARLO METHODS*

Arnold ZELLNER

University of Chicago, Chicago, IL 60638, USA

Luc BAUWENS

CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
GREQUE, Ecole des Hautes Etudes en Sciences Sociales, Marseille, France

Herman K. VAN DIJK

Erasmus University, 3000 DR Rotterdam, The Netherlands
CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Bayesian procedures for specification analysis or diagnostic checking of modeling assumptions for structural equations of econometric models are developed and applied using Monte Carlo numerical methods. Checks on the validity of identifying restrictions, exogeneity assumptions and other specifying assumptions are performed using posterior distributions for *discrepancy vectors and functions* representing departures from specifying assumptions. Several mappings or functions of reduced form coefficients are defined and their posterior distributions are computed. A restricted reduced form approach is used to compute posterior distributions for structural parameters. These procedures are applied in analyses of two econometric models.

1. Introduction

There have been many studies relating to limited information estimation of the parameters of the simultaneous equation model (SEM) from both the Bayesian and non-Bayesian points of view – see, e.g., Zellner (1971, 1979), Drèze (1976), Drèze and Richard (1983), Hausman (1983), Tsurumi (1985, 1987), and the references cited in these works. In non-Bayesian approaches, there is usually reliance on asymptotic approximations in making inferences.[1] Some previous Bayesian approaches also involve asymptotic approximations.

*The first and third authors received support from the National Science Foundation and from the H.G.B. Alexander Endowment Fund, Graduate School of Business, University of Chicago. The second author acknowledges support from Erasmus University. Comments by J. Drèze and J.F. Richard on an earlier draft were very helpful.

[1]A brief discussion of small sample results in non-Bayesian limited information estimation of the SEM is given by Anderson (1984, pp. 518–519). Tsurumi (1987) reports Monte Carlo experimental results.

A problem in previous exact Bayesian analyses is that posterior distributions of structural parameters are in most cases not analytically tractable[2] and thus must be integrated numerically to obtain their moments, marginal distributions, etc. As regards Monte Carlo numerical integration, usual posterior distributions of structural parameters do not have simple forms from which draws can be made easily. As a consequence, the success of Monte Carlo integration procedures depends importantly on an investigator's ability to find distribution functions that are good approximations to posterior distributions and from which pseudo-random drawings can be made easily. Also, past Bayesian analyses of the SEM have not devoted much attention to diagnostic checking of models' assumptions, that is to specification error analysis.

In the present paper, we start from the reduced form of the SEM and make a distinction between 'unrestricted reduced form analysis' (URFA) and 'restricted reduced form analysis' (RRFA). In our URFA, we define indirect least squares, generalized indirect least squares, two-stage least squares and limited information maximum likelihood mappings or functions of unrestricted reduced form coefficients which do not require that overidentifying restrictions hold exactly and obtain complete posterior distributions of these mappings or functions by a direct Monte Carlo simulation approach. Also *discrepancy vectors* and *discrepancy functions* are introduced which measure the extent to which overidentifying restrictions are in error and we indicate how to obtain their posterior distributions by a direct simulation approach. One may also use Bayesian realized error analysis [Zellner (1975)] to provide further diagnostic checks of the SEM.

In the case that exact identifying restrictions are imposed, we present a RRFA and discuss a method for computing posterior distributions of structural parameters which makes use of Monte Carlo integration in a relatively simple way, namely a direct simulation approach.

The plan of our paper is as follows. In section 2 we consider simple, canonical models to illustrate our approach and go on to specify a general system. Then various mappings of the URF coefficients are introduced and we indicate how to compute their posterior distributions, moments, etc. This is followed by an analysis of the RRF system to obtain posterior distributions of structural coefficients. Section 3 is devoted to further diagnostic checking procedures. In section 4, our methods are applied in illustrative analyses of several well known models using actual data. Section 5 provides some concluding remarks. An efficient algorithm for generating pseudo-random drawings from a matrix Student-t distribution is presented in the appendix.

[2]An exception is Drèze (1976) where the posterior density is in the poly-t family. Then one can, in some cases, compute moments of structural coefficients analytically. See also Bauwens and Richard (1985) and Tsurumi (1985, 1987).

2. Model specification, interpretation and analysis

In this section we first consider canonical models to illustrate features of our approach. Then we specify unrestricted reduced form (URF) systems and indicate how to compute posterior distributions for interesting functions or mappings of URF coefficients. These functions or mappings are related to *discrepancy vectors* which measure departures of the URF coefficients from satisfying usual overidentifying restrictions. Next, we impose identifying and normalizing restrictions, derive the posterior distribution of the parameters of a single structural equation using diffuse and informative prior distributions and discuss a Monte Carlo integration procedure for the computation of posterior moments and densities. Also, various conditional posterior distributions centered at OLS, 2SLS, LIML, and MELO point estimates and diagnostic checks of the validity of overidentifying restrictions are provided.

2.1. Canonical models

The first canonical model is a 'means model' for two endogenous variables, namely,

$$y_{1i} = \eta_i + v_{1i}, \qquad i = 1,2,\dots,n, \tag{2.1a}$$
$$y_{2i} = \xi_i + v_{2i}, \tag{2.1b}$$

where η_i and ξ_i are means of y_{1i} and y_{2i}, respectively, and the zero-mean disturbance terms, v_{1i} and v_{2i}, are assumed independently drawn from a bivariate normal distribution with 2×2 positive definite symmetric (pds) covariance matrix. For example, η_i and ξ_i can be interpreted as the ith individual's 'permanent' or 'anticipated' consumption and income, respectively, whereas y_{1i} and y_{2i} are their measured counterparts. Interest may center on various functions of the η_i's and ξ_i's, for example η_i/ξ_i, $i = 1,2,\dots,n$, the 'permanent consumption–income' ratios,

$$\bar{\eta} = \sum_{i=1}^{n} \eta_i/n, \quad \bar{\xi} = \sum_{i=1}^{n} \xi_i/n,$$

$$\sigma_{\xi\xi} = \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2/n,$$

$$\sigma_{\eta\eta} = \sum_{i=1}^{n} (\eta_i - \bar{\eta})^2/n,$$

$$\sigma_{\xi\eta} = \sum_{i=1}^{n} (\xi_i - \bar{\xi})(\eta_i - \bar{\eta})/n,$$

higher-order moments, skewness and kurtosis measures, etc. Further, weighted averages of the ratios η_i/ξ_i, e.g.,

$$\bar{\gamma}_1 = \bar{\eta}/\bar{\xi} = \sum_{i=1}^{n} (\eta_i/\xi_i)\xi_i \Big/ \sum_{i=1}^{n} \xi_i,$$

or

$$\bar{\gamma}_2 = \sum_{i=1}^{n} (\eta_i/\xi_i)\xi_i^2 \Big/ \sum_{i=1}^{n} \xi_i^2 = \sum_{i=1}^{n} \eta_i \xi_i \Big/ \sum_{i=1}^{n} \xi_i^2 = \boldsymbol{\xi}'\boldsymbol{\eta}/\boldsymbol{\xi}'\boldsymbol{\xi},$$

where

$$\boldsymbol{\eta}' = (\eta_1, \eta_2, \ldots, \eta_n) \quad \text{and} \quad \boldsymbol{\xi}' = (\xi_1, \xi_2, \ldots, \xi_n),$$

might be of interest. If we write

$$\boldsymbol{\eta} = \boldsymbol{\xi}\gamma + \boldsymbol{\Delta}_1, \tag{2.2}$$

where γ is a scalar parameter and $\boldsymbol{\Delta}_1$ is an $n \times 1$ *discrepancy vector*, which measures the extent to which the η_i/ξ_i depart from a common value γ, then $\bar{\gamma}_2 = \boldsymbol{\xi}'\boldsymbol{\eta}/\boldsymbol{\xi}'\boldsymbol{\xi}$ is the value of γ that minimizes $\boldsymbol{\Delta}_1'\boldsymbol{\Delta}_1 = (\boldsymbol{\eta} - \boldsymbol{\xi}\gamma)'(\boldsymbol{\eta} - \boldsymbol{\xi}\gamma)$, a *discrepancy function*. Also, the functions $\bar{\sigma}_1^2 = (\boldsymbol{\eta} - \boldsymbol{\xi}\bar{\gamma}_1)'(\boldsymbol{\eta} - \boldsymbol{\xi}\bar{\gamma}_1)/n$ and $\bar{\rho}_1^2 = 1 - n\bar{\sigma}_1^2/\boldsymbol{\eta}'\boldsymbol{\eta}$ are of interest and have obvious regression interpretations.

Given a posterior distribution for the $2n$ parameters, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, draws can be made from it and complete posterior distributions for η_i/ξ_i, $\bar{\eta}$, $\bar{\xi}$, $\sigma_{\xi\xi}$, $\sigma_{\eta\eta}$, $\sigma_{\xi\eta}$, $\bar{\gamma}$, $\bar{\sigma}_1^2$, $\bar{\rho}_1^2$, etc. can be obtained by a direct Monte Carlo approach, that is by repeated evaluation of these quantities using independent draws from the joint distribution. If it is the case that the distribution of $\bar{\sigma}_1^2$ is centered far from zero, there is little support for the assumption $\boldsymbol{\Delta}_1 = \mathbf{0}$ or $\boldsymbol{\eta} = \gamma\boldsymbol{\xi}$. On the other hand, if $\bar{\sigma}_1^2$'s distribution is centered close to zero, this provides some support for the assumption $\boldsymbol{\Delta}_1 = \mathbf{0}$ and, with this assumption, the model becomes a form of the usual 'errors-in-variables' model. While we do not pursue the matter now, it is also possible to compute posterior odds relating to the hypotheses $\boldsymbol{\Delta}_1 = \mathbf{0}$ and $\boldsymbol{\Delta}_1 \neq \mathbf{0}$.

If in addition to (2.1), we have proxies for η_i and ξ_i, namely,

$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\pi}_1, \tag{2.3a}$$

$$\xi_i = \boldsymbol{x}_i'\boldsymbol{\pi}_2, \tag{2.3b}$$

where $\boldsymbol{x}_i'$ is a $1 \times k$ vector of predetermined variables, a typical row of an $n \times k$ matrix X, assumed of full column rank, and $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ are $k \times 1$ coefficient vectors, the number of location parameters is reduced from $2n$ ξ_i's and η_i's to $2k$ π's. Using (2.3), we can express (2.1) in matrix form as follows:

$$y_1 = X\pi_1 + v_1, \tag{2.4a}$$

$$y_2 = X\pi_2 + v_2, \tag{2.4b}$$

where y_1, y_2, v_1, and v_2 are $n \times 1$ vectors with typical elements y_{1i}, y_{2i}, v_{1i} and v_{2i}, respectively.

In (2.4), we have two URF equations. Just as with (2.1), we may be interested in various functions or mappings of the URF coefficients, the analogues of those for η_i and ξ_i with $x_i'\pi_1$ and $x_i'\pi_2$ replacing η_i and ξ_i, respectively, in their definitions. Also, we can introduce

$$X\pi_1 = X\pi_2\gamma + \Delta_2, \tag{2.5}$$

where Δ_2 is an $n \times 1$ discrepancy vector. Then $\bar{\gamma}_2 = \pi_2'X'X\pi_1/\pi_2'X'X\pi_2$ is the value of γ that minimizes $\Delta_2'\Delta_2$. Further, $\bar{\sigma}_2^2 = (X\pi_1 - X\pi_2\bar{\gamma}_2)'(X\pi_1 - X\pi_2\bar{\gamma}_2)/n$ and $\bar{\rho}_2^2 = 1 - n\bar{\sigma}_2^2/\pi_1X'X\pi_1$ are regression-like mappings of the π's which are of interest. Also, if we consider

$$\pi_1 = \pi_2\gamma + \Delta_3, \tag{2.6}$$

where Δ_3 is a $k \times 1$ discrepancy vector, then the value of γ, say $\bar{\gamma}_3$, minimizing $\Delta_3'\Delta_3$, is just $\bar{\gamma}_3 = \pi_2'\pi_1/\pi_2'\pi_2$ and $\bar{\sigma}_3^2 = (\pi_1 - \pi_2\bar{\gamma}_3)'(\pi_1 - \pi_2\bar{\gamma}_3)/k$ and $\bar{\rho}_3^2 = 1 - k\bar{\sigma}_3^2/\pi_1'\pi_1$ are measures of the extent to which $\Delta_3 = \mathbf{0}$ holds.

Given a joint posterior pdf for π_1 and π_2 from which draws can be made, a direct Monte Carlo simulation approach can be employed to obtain the posterior distributions of $\bar{\gamma}_2$, $\bar{\gamma}_3$, $\bar{\sigma}_2^2$, $\bar{\sigma}_3^2$, $\bar{\rho}_2^2$, $\bar{\rho}_3^2$, etc., since these quantities are given functions or mappings of the unrestricted π's.

If $\Delta_2 = \mathbf{0}$ in (2.5) or $\Delta_3 = \mathbf{0}$ in (2.6), we have the case of *exact restrictions*. Then (2.4) can be written as

$$y_1 = X\pi_2\gamma + v_1, \tag{2.7a}$$

$$y_2 = X\pi_2 + v_2, \tag{2.7b}$$

or

$$y_1 = y_2\gamma + u_1, \tag{2.7c}$$

$$y_2 = X\pi_2 + u_2, \tag{2.7d}$$

where $u_1 = v_1 - v_2\gamma$ and $u_2 = v_2$. Eqs. (2.7a) and (2.7b) form the restricted reduced form (RRF) equation system which can also be expressed in structural form as shown in (2.7c) and (2.7d). On introducing a prior distribution for γ, $\boldsymbol{\pi}_2$ and the reduced form disturbance covariance matrix, we can obtain a posterior distribution for these parameters. Note that in working with (2.7a) and (2.7b), it is assumed that the overidentifying restrictions *hold exactly*, that is $\boldsymbol{\Delta}_2 = \mathbf{0}$ in (2.5) or $\boldsymbol{\Delta}_3 = \mathbf{0}$ in (2.6). The number of coefficients in (2.7) is $k+1$, usually a large reduction from the $2k$ coefficients in (2.4) for $k > 1$. When $k = 1$, the case of 'just-identification', the number of coefficients in the URF and RRF is the same. Also, relative to the $2n$ location parameters in (2.1), the reduction is much larger. This reduction, however, is dependent not only on the identifying restrictions holding exactly but also on the appropriateness of the proxy expressions in (2.3). Diagnostic checking procedures relating to these assumptions will be described in a subsequent section.

We now turn to provide results for general cases including mappings of reduced form coefficients in the unrestricted case and posterior distributions for structural parameters in the restricted reduced form case after introducing some needed notation. Let $Y_a = (y_1 \vdots Y_1 \vdots Y_0)$ denote an $n \times m'$ matrix of observations on m' endogenous variables with URF,

$$(y_1 \vdots Y_1 \vdots Y_0) = X(\boldsymbol{\pi}_1 \vdots \Pi_1 \vdots \Pi_0) + (\boldsymbol{v}_1 \vdots V_1 \vdots V_0), \tag{2.8}$$

where X is an $n \times k$ matrix of observations on k predetermined variables of rank k and the rows of the disturbance matrix have been independently drawn from a zero-mean multivariate normal distribution with a pds covariance matrix. A structural equation, say the first, with normalization imposed can be written as

$$(y_1 \vdots Y_1 \vdots Y_0)\begin{pmatrix} 1 \\ -\gamma_1 \\ \mathbf{0} \end{pmatrix} = (X_1 \vdots X_0)\begin{pmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{pmatrix} + \boldsymbol{u}_1, \tag{2.9a}$$

or

$$y_1 - Y_1\gamma_1 = X_1\boldsymbol{\beta}_1 + \boldsymbol{u}_1, \tag{2.9b}$$

where Y_0 and X_0 are observations on endogenous and predetermined variables excluded from the first equation and $X = (X_1 \vdots X_0)$. The $m_1 \times 1$ vector γ_1 and the $k_1 \times 1$ vector $\boldsymbol{\beta}_1$ are the structural coefficients and $\boldsymbol{u}_1$ is an $n \times 1$ vector of structural disturbance terms.

To obtain the well-known restrictions on the reduced form coefficients, we write (2.8) as

$$(y_1 \vdots Y_1 \vdots Y_0) = (X_1 \vdots X_0)\begin{pmatrix} \boldsymbol{\pi}_{11} \vdots \Pi_{11} \vdots \Pi_{01} \\ \boldsymbol{\pi}_{10} \vdots \Pi_{10} \vdots \Pi_{00} \end{pmatrix} + (\boldsymbol{v}_1 \vdots V_1 \vdots V_0), \tag{2.10}$$

and on multiplying both sides of (2.10) on the right by $(1 \vdots -\gamma_1' \vdots \mathbf{0}')$, the result is

$$y_1 - Y_1\gamma_1 = (X_1 \vdots X_0)\begin{pmatrix} \pi_{11} - \Pi_{11}\gamma_1 \\ \pi_{10} - \Pi_{10}\gamma_1 \end{pmatrix} + v_1 - V_1\gamma_1. \tag{2.11}$$

For compatibility with (2.9b), $u_1 = v_1 - V_1\gamma_1$ and

$$\pi_{11} - \Pi_{11}\gamma_1 = \beta_1, \tag{2.12a}$$

$$\pi_{10} - \Pi_{10}\gamma_1 = \mathbf{0}, \tag{2.12b}$$

which are restrictions on the reduced form coefficients with γ_1 and β_1 appearing in them, a generalization of (2.6) with $\Delta_3 = \mathbf{0}$. In (2.12b) Π_{10} is assumed to be of full column rank.

On substituting for $(\pi_{11}' \vdots \pi_{10}')'$ in (2.10) from (2.12), the RRF equations for y_1 and Y_1 are

$$y_1 = X\Pi_1\gamma_1 + X_1\beta_1 + v_1, \tag{2.13a}$$

$$Y_1 = X\Pi_1 + V_1, \tag{2.13b}$$

where $\Pi_1' = (\Pi_{11}' \vdots \Pi_{10}')$. It is seen that (2.13) is in the form of a multivariate non-linear regression model, a generalization of (2.7). The system in (2.13) will serve as the starting point for an analysis of the RRF system, whereas

$$(y_1 \vdots Y_1) = X(\pi_1 \vdots \Pi_1) + (v_1 \vdots V_1) \tag{2.14a}$$

will serve as the starting point for the URF analysis of the data $(y_1 \vdots Y_1)$.

2.2. Mappings of unrestricted reduced form (URF) coefficients

We shall obtain a posterior distribution for the parameters of (2.14a) and use it to obtain posterior distributions of interesting functions or mappings of the URF coefficients, $(\pi_1 \vdots \Pi_1)$. For convenience, we write $Y = (y_1 \vdots Y_1)$, $\Pi = (\pi_1 \vdots \Pi_1)$ and $V = (v_1 \vdots V_1)$ and thus (2.14a) becomes

$$\underset{n\times m}{Y} = \underset{n\times k}{X}\ \underset{k\times m}{\Pi} + \underset{n\times m}{V}. \tag{2.14b}$$

The n rows of V are assumed to be independently drawn from a multivariate normal distribution with zero mean vector and $m \times m$ pds covariance matrix Ω, i.e., MVN$(\mathbf{0}, \Omega)$. If X includes lagged endogenous variables, we assume that initial or starting values are given. Then the likelihood function for (2.14b) is

$$l(\Pi,\Omega|Y_1,X)\propto|\Omega|^{-n/2}\exp\{-\tfrac{1}{2}\mathrm{tr}(Y-X\Pi)'(Y-X\Pi)\Omega^{-1}\}$$

$$\propto|\Omega|^{-n/2}\exp\{-\tfrac{1}{2}\mathrm{tr}[S+(\Pi-\hat{\Pi})'\times X'X(\Pi-\hat{\Pi})]\Omega^{-1}\}, \tag{2.15}$$

where $\propto$ denotes 'is proportional to', and

$$\hat{\Pi}=(X'X)^{-1}X'Y, \tag{2.16a}$$

$$S=(Y-X\hat{\Pi})'(Y-X\hat{\Pi}). \tag{2.16b}$$

It is seen that the likelihood function in (2.15) is in the same form as that for a multivariate regression model – see, e.g., Zellner (1971, ch. 8) with $\hat{\Pi}$ and S sufficient statistics.

We shall employ the following standard diffuse prior distribution for Π and the distinct elements of Ω:[3]

$$p(\Pi,\Omega)\propto|\Omega|^{-(m+1+\nu_0)/2}, \tag{2.17}$$

where $\nu_0\geq 0$, that is the elements of Π and Ω are independent, with the former being uniformly distributed and the latter in the form of a degenerate, inverted Wishart distribution.

On multiplying (2.15) and (2.17) and using $\nu_0=0$, we obtain by Bayes' Theorem the joint posterior density of Π and Ω, namely,

$$p(\Pi,\Omega|D)\propto|\Omega|^{-(n+m+1)/2}\exp\{-\tfrac{1}{2}\mathrm{tr}[S+(\Pi-\hat{\Pi})'\times X'X(\Pi-\hat{\Pi})]\Omega^{-1}\}, \tag{2.18}$$

where D denotes the given sample information (Y, X) and prior information in (2.17). On integrating (2.18) with respect to Ω, we obtain the well-known marginal posterior density for Π,

$$p(\Pi|D)\propto|S+(\Pi-\hat{\Pi})'X'X(\Pi-\hat{\Pi})|^{-n/2}, \tag{2.19}$$

[3]The value $\nu_0=k$ in the exponent of (2.17) has been suggested by Drèze (1976) while Zellner (1971) employs $\nu_0=0$.

which is in the form of a matrix Student-t density – see, e.g., Dickey (1967), Box and Tiao (1973), Drèze and Richard (1983), Geisser (1965), and Zellner (1971) for properties of this distribution. As explained below, it is possible to make independent draws from (2.19) and to use them to determine the posterior distributions of interesting functions or mappings of the elements of Π. Some of these mappings are given below.

We first consider the case of 'just-identification' in which the matrix Π_{10} in (2.12b) is square and non-singular and the matrix $(\boldsymbol{\pi}_{10} \vdots - \Pi_{10})$ is not of full column rank. Then (2.12b) has a unique solution for γ_1 – see Graybill (1969, p. 140), and this solution can be substituted in (2.12a) to express $\boldsymbol{\beta}_1$ in terms of the RF coefficients. Explicitly, we have

$$\boldsymbol{\beta}_1 = \boldsymbol{\pi}_{11} - \Pi_{11}\Pi_{10}^{-1}\boldsymbol{\pi}_{10}, \tag{2.20a}$$

$$\gamma_1 = \Pi_{10}^{-1}\boldsymbol{\pi}_{10}, \tag{2.20b}$$

which we call the *Indirect Least Squares* (ILS) mapping since if least squares estimates of the Π's are inserted in (2.20), the result is the 'indirect least squares' estimate of non-Bayesian econometrics. In the Bayesian approach, with the posterior distribution for Π in (2.19), the least squares quantity $\hat{\Pi} = (X'X)^{-1}X'Y$ is the modal value and mean of (2.19) and the ILS estimate is the modal value of the posterior distribution of $\boldsymbol{\beta}_1$ and γ_1 in this case of 'exact identification' since (2.20) is a one-to-one transformation from the Π's to $\boldsymbol{\beta}_1$ and γ_1. Further, as explained below, we can make independent draws from the matrix Student-t posterior distribution for Π in (2.19) and evaluate $\boldsymbol{\beta}_1$ and γ_1 for each draw by use of (2.20) and thus obtain the complete posterior distributions for the elements of $\boldsymbol{\beta}_1$ and γ_1. Also, various measures associated with these distributions can be calculated, for example medians, inter-quartile ranges, means (if they exist), etc., as will be illustrated in computed examples below.[4]

In the case of overidentification, the matrix Π_{10} in (2.12b) has dimension $k_0 \times m_1$, where k_0 is the number of columns of X_0 or the number of predetermined variables left out of the first structural equation in (2.9b) and m_1 is the number of columns of Y_1 or the number of endogenous variables included in (2.9b) less one. The rank condition for identification of the structural coefficients γ_1 and $\boldsymbol{\beta}_1$ is that the rank of Π_0 is m_1 which requires $k_0 > m_1$, the order condition in the overidentified case. In the overidentified case, we cannot go from the URF coefficients, the elements of Π in (2.14b) and (2.19) to the elements of γ_1 and $\boldsymbol{\beta}_1$. For example in (2.6) with $\boldsymbol{\Delta}_3 = \mathbf{0}$, $\boldsymbol{\pi}_1 = \boldsymbol{\pi}_2\gamma$ and given that $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ are a.s. linearly independent in the URF, we cannot solve for γ in terms of the elements of the vectors of URF

[4] Drèze (1976, p. 1055) discusses conditions for existence of moments of structural coefficients.

coefficients, $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$. In fact, we can only find an *approximate* solution [Graybill (1969, p. 103ff.)] as follows. Just as in (2.6), we shall append a discrepancy vector $\boldsymbol{\Delta}_2$, to (2.12b). This yields

$$\boldsymbol{\pi}_{11} - \Pi_{11}\boldsymbol{\gamma}_1 = \boldsymbol{\beta}_1, \tag{2.21a}$$

$$\boldsymbol{\pi}_{10} - \Pi_{10}\boldsymbol{\gamma}_1 = \boldsymbol{\Delta}_2. \tag{2.21b}$$

We can now define discrepancy functions and obtain values of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\beta}_1$ which minimize them. One example of a discrepancy function is $\Delta'_2\Delta_2$ and the value of $\boldsymbol{\gamma}_1$ which minimizes this function, denoted by $\boldsymbol{\gamma}_1^*$ and the associated value of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_1^*$ are

$$\boldsymbol{\beta}_1^* = \boldsymbol{\pi}_{11} - \Pi_{11}\boldsymbol{\gamma}_1^*, \tag{2.22a}$$

$$\boldsymbol{\gamma}_1^* = (\Pi'_{10}\Pi_{10})^{-1}\Pi'_{10}\boldsymbol{\pi}_{10}. \tag{2.22b}$$

We shall call the mapping in (2.22) the *Generalized Indirect Least Squares* (GILS) mapping since when least squares estimates of the π's are inserted in (2.22), the result is the GILS estimate – see Khazzoom (1976). In our Bayesian approach, the posterior distribution of the elements of $\boldsymbol{\beta}_1^*$ and $\boldsymbol{\gamma}_1^*$ can be computed by direct Monte Carlo simulation based on draws from the matrix Student-*t* posterior distribution for Π in (2.19). Also posterior distributions for the discrepancy functions can be computed, for example

$$\tilde{\Delta}'_2\tilde{\Delta}_2/k_0 = (\boldsymbol{\pi}_{10} - \Pi_{10}\boldsymbol{\gamma}_1^*)'(\boldsymbol{\pi}_{10} - \Pi_{10}\boldsymbol{\gamma}_1^*)/k_0, \tag{2.23a}$$

and

$$\tilde{\rho}_2^2 = 1 - \tilde{\Delta}'_2\tilde{\Delta}_2/\boldsymbol{\pi}'_{10}\boldsymbol{\pi}_{10}. \tag{2.23b}$$

Also, the posterior distributions of the elements of $\tilde{\boldsymbol{\Delta}}_2 = \boldsymbol{\pi}_{10} - \Pi_{10}\boldsymbol{\gamma}_1^*$ can be computed by direct Monte Carlo simulation. The posterior distributions of $\tilde{\boldsymbol{\Delta}}_2$, $\tilde{\Delta}'_2\tilde{\Delta}_2/k_0$ and $\tilde{\rho}_2^2$ will provide information regarding the validity of the exact restrictions in (2.12) in the frequently encountered overidentified case.

We next turn to a mapping that involves the matrix of predetermined variables by multiplying both sides of (2.12) on the left by $X = (X_1 X_0)$ to obtain

$$X\boldsymbol{\pi}_1 = X\Pi_1\boldsymbol{\gamma}_1 + X_1\boldsymbol{\beta}_1 = \bar{Z}_1\boldsymbol{\delta}_1, \tag{2.24}$$

where $\boldsymbol{\pi}'_1 = (\boldsymbol{\pi}'_{11}\ \boldsymbol{\pi}'_{10})$, $\Pi'_1 = (\Pi'_{11}\ \Pi'_{10})$, $\bar{Z}_1 = (X\Pi_1\ \ X_1)$ and $\boldsymbol{\delta}'_1 = (\boldsymbol{\gamma}'_1\ \boldsymbol{\beta}'_1)$. To allow for possible errors in the exact restrictions in (2.24), we introduce a discrepancy vector, Δ_3, as follows:

$$X\pi_1 = \bar{Z}_1\delta_1 + \Delta_3. \tag{2.25}$$

Then, just as in the cases considered above, we can minimize the discrepancy function $\Delta_3'\Delta_3$ with respect to δ_1 to obtain

$$\delta_1^* = (\bar{Z}_1'\bar{Z}_1)^{-1}\bar{Z}_1'X\pi_1 \tag{2.26}$$

as the minimizing value which defines a mapping of the π's, which resembles that arising in 2SLS estimation.[5] Thus we call (2.26) the *2SLS Mapping*. Also from (2.25) and (2.26), we can define

$$\tilde{\Delta}_3 = X\pi_1 - \bar{Z}_1\delta_1^*, \tag{2.27a}$$

$$\tilde{\Delta}_3'\tilde{\Delta}_3/n = (X\pi_1 - \bar{Z}_1\delta_1^*)'(X\pi_1 - \bar{Z}_1\delta_1^*)/n, \tag{2.27b}$$

$$\tilde{\rho}_3^2 = 1 - \tilde{\Delta}_3'\tilde{\Delta}_3/\pi_1'X'X\pi_1. \tag{2.27c}$$

Posterior distributions of δ_1^*, $\tilde{\Delta}_3$, $\tilde{\Delta}_3'\tilde{\Delta}_3$, $\tilde{\rho}_3^2$ and other interesting functions of the URF coefficients can be calculated using a direct Monte Carlo simulation approach based on draws from the matrix Student-t distribution in (2.19).

Last, we define a *LIML mapping* as follows. Write the URF system for $Y = (y_1 \; Y_1)$ in (2.14) as

$$Y = X_1\Pi_{1\cdot} + X_0\Pi_{0\cdot} + V, \tag{2.28}$$

where $\Pi' = (\Pi_{1\cdot}' \; \Pi_{0\cdot}')$ and multiply both sides of (2.28) on the right by $\gamma_a = (1 \vdots \gamma_1')'$ to obtain

$$Y\gamma_a = X_1\Pi_{1\cdot}\gamma_a + X_0\Pi_{0\cdot}\gamma_a + V\gamma_a = X\Pi\gamma_a + V\gamma_a. \tag{2.29}$$

Note that $\Pi_{0\cdot}\gamma_a = \mathbf{0}$ if the restrictions in (2.12) hold and thus we introduce a 'variance ratio' discrepancy function,

$$\phi = \gamma_a'V_r'V_r\gamma_a/\gamma_a'V'V\gamma_a, \tag{2.30}$$

where $V = Y - X\Pi$ and $V_r = Y - X_1\Pi_1$. With $\bar{l}$ being the smallest root of $|V_r'V_r - lV'V| = 0$, the value of γ_a minimizing ϕ in (2.30) is obtained by solving the following set of equations, given Π, X and Y,

$$(V_r'V_r - \bar{l}V'V)\gamma_a = \mathbf{0}.$$

[5]An alternative procedure to compute δ_1^* is presented in section 3.

The solution is $\gamma_a^* = (1 \vdots -\gamma_1^{*\prime})'$ and we can then define $\beta_1^* = \pi_{11} - \Pi_{11}\gamma_1^*$ from the restrictions in (2.12). Thus $\delta_1^{*\prime} = (\gamma_1^{*\prime}\ \beta_1^{*\prime})$ is the *LIML mapping* which can be substituted in (2.30) to yield $\phi^* = \gamma_a^{*\prime}V_r'V_r\gamma_a^*/\gamma_a^{*\prime}V'V\gamma_a^*$. The posterior distributions of δ_1^*, ϕ^*, $\Pi_{0\cdot}\gamma_a^*$, etc. can be calculated by direct Monte Carlo simulation based on independent draws of Π from its posterior distribution in (2.19).

We have discussed various mappings that are useful in connection with URF analysis which do not involve assuming that identifying restrictions hold exactly. One may extend the GILS mapping and the 2SLS mapping to the case of a full system of equations [see van Dijk (1985)]. We shall not pursue this extension herein. We turn now to the derivation of posterior distributions for structural parameters in a RRF framework.

2.3. *Restricted reduced form analysis (RRFA)*

We now assume that the restrictions in (2.12) and in the line above (2.12), hold exactly and impose them to obtain the RRF system of the equations for y_1 and Y_1 as follows. Substitute the expression $v_1 = u_1 + V_1\gamma_1$ in (2.13a), use (2.13b) and (2.10), and re-express (2.13) as

$$(y_1\ Y_1)\begin{bmatrix} 1 & 0 \\ -\gamma_1 & I \end{bmatrix} = (X_1\ X_0)\begin{bmatrix} \beta_1 & \Pi_{11} \\ 0 & \Pi_{10} \end{bmatrix} + (u_1\ V_1). \tag{2.13'}$$

Assuming that the rows of $(u_1\ V_1)$ are independently drawn from a zero-mean normal distribution with PDS covariance matrix Ω^*, where

$$\Omega^* = \begin{bmatrix} \sigma_1^2 & \omega_1' \\ \omega_1 & \Omega_1 \end{bmatrix}, \tag{2.13''}$$

one can write the likelihood function

$$l(\delta_1, \Pi_1, \Omega^* | D) \propto |\Omega^*|^{-n/2} \exp\left\{-\tfrac{1}{2}\mathrm{tr}\left[(u_1\ V_1)'(u_1\ V_1)\Omega^{*-1}\right]\right\}, \tag{2.31}$$

where $\delta_1' = (\gamma_1'\ \beta_1')$, $D = (Y\ X)$ and $(u_1\ V_1)$ is restricted by eq. (2.13′). A well-known diffuse prior for the parameters of (2.31) is

$$p(\delta_1, \Pi_1, \Omega^*) \propto |\Omega^*|^{-(m_1+2+\nu_0)/2}, \tag{2.32}$$

where ν_0 (≥ 0) can be chosen in accordance with invariance considerations. More informative priors are discussed below. Multiplying (2.31) and (2.32) gives the posterior pdf as

$$p(\boldsymbol{\delta}_1, \Pi_1, \Omega^* | D) \propto |\Omega^*|^{-(n_* + m_1 + 2)/2} \times \exp\{-\tfrac{1}{2}\mathrm{tr}[(\boldsymbol{u}_1\ V_1)'(\boldsymbol{u}_1\ V_1)\Omega^{*-1}]\}, \tag{2.33}$$

where $n_* = n + \nu_0$. On integrating the posterior with respect to the elements of Ω^*, one obtains the marginal pdf for $\boldsymbol{\delta}_1$ and Π_1, given as

$$p(\boldsymbol{\delta}_1, \Pi_1 | D) \propto |(\boldsymbol{u}_1\ V_1)'(\boldsymbol{u}_1\ V_1)|^{-n_*/2}. \tag{2.34}$$

We now make use of

$$|(\boldsymbol{u}_1\ V_1)'(\boldsymbol{u}_1\ V_1)| = (\boldsymbol{u}_1' M_1 \boldsymbol{u}_1)|V_1'V_1|,$$

where

$$M_1 = I - V_1(V_1'V_1)^{-1}V_1',$$

and rewrite (2.34) as

$$p(\boldsymbol{\delta}_1, \Pi_1 | D) \propto |(y_1 - W_1\boldsymbol{\delta}_1)'M_1(y_1 - W_1\boldsymbol{\delta}_1)|^{-n_*/2} \times |(Y_1 - X\Pi_1)'(Y_1 - X\Pi_1)|^{-n_*/2}, \tag{2.35}$$

where $W_1 = (Y_1\ X_1)$. By making use of the definitions of the multivariate and matrix variate Student-t density functions [see Zellner (1971, app. B)] one can re-express (2.35) as

$$p(\boldsymbol{\delta}_1, \Pi_1 | D) = p_1(\boldsymbol{\delta}_1 | \Pi_1, D)\, p_2(\Pi_1 | D), \tag{2.36}$$

where

$$p_1(\boldsymbol{\delta}_1 | \Pi_1, D) = c_1 \left| \frac{W_1'M_1W_1}{s_1^2} \right|^{1/2} \{\nu_1 + (\boldsymbol{\delta}_1 - \tilde{\boldsymbol{\delta}}_1)'W_1'M_1W_1(\boldsymbol{\delta}_1 - \tilde{\boldsymbol{\delta}}_1)/s_1^2\}^{-(\nu_1 + l_1)/2}, \tag{2.36a}$$

and

$$p_2(\Pi_1 | D) = c_2 f(\Pi_1)\left\{c_3 |S_1 + (\Pi_1 - \hat{\Pi}_1)'X'X(\Pi_1 - \hat{\Pi}_1)|^{-n_*/2}\right\}, \tag{2.36b}$$

with $f(\Pi_1)$ given as

$$f(\Pi_1) = |W_1'M_1W_1|^{-1/2}(s_1^2)^{-\nu_1/2}. \tag{2.36c}$$

The normalizing constants c_1 and c_3 are well-known in terms of elementary functions [see our appendix and Zellner (1971, app. B)]. The parameters of (2.36a) are given as $\nu_1 = n_* - l_1$, $l_1 = m_1 + k_1$ and

$$\tilde{\delta}_1 = (W_1'M_1W_1)^{-1}W_1'M_1y_1, \qquad \nu_1 s_1^2 = (y_1 - W_1\tilde{\delta}_1)'M_1(y_1 - W_1\tilde{\delta}_1), \tag{2.37a}$$

with $|W_1'M_1W_1| > 0$, $\nu_1 s_1^2 > 0$. The parameters of (2.36b) are

$$\hat{\Pi}_1 = (X'X)^{-1}X'Y_1, \qquad S_1 = (Y_1 - X\hat{\Pi}_1)'(Y_1 - X\hat{\Pi}_1). \tag{2.37b}$$

Note that $p(\delta_1|\Pi_1, D)$ in (2.36a), the conditional posterior density for δ_1 given Π_1 and D, is in the form of a l_1-variate Student-t pdf with ν_1 degrees of freedom with mean $\tilde{\delta}_1$ and covariance matrix $(W_1'M_1W_1)^{-1}\nu_1 s_1^2/(\nu_1 - 2)$, both of which depend on Π_1. On integration over the elements of δ_1 in (2.36a), the marginal posterior density for Π_1 is given in (2.36b) which is written as $c_2 f(\Pi_1)$ times a normalized matrix Student-t factor with c_2 the normalizing constant that, to the best of our knowledge, is not known in terms of elementary functions.

To obtain the unconditional moments of the elements of δ_1, we make N draws $\Pi_1^{(i)}$, $i = 1, \ldots, N$, from the matrix Student-t factor in (2.36b) (see the algorithm described in the appendix) and use well-known formulas to compute marginal moments from conditional moments. For example, to compute the unconditional mean of δ_1, we have

$$\begin{aligned} \mathrm{E}(\delta_1|D) &= \int \tilde{\delta}_1 p_2(\Pi_1|D)\,\mathrm{d}\Pi_1 \\ &= \int \tilde{\delta}_1 f(\Pi_1) p_3(\Pi_1|D)\,\mathrm{d}\Pi_1 \Big/ \int f(\Pi_1) p_3(\Pi_1|D)\,\mathrm{d}\Pi_1, \end{aligned} \tag{2.38}$$

where

$$p_3(\Pi_1|D) = c_3\left|S_1 + (\Pi_1 - \hat{\Pi}_1)'X'X(\Pi_1 - \hat{\Pi}_1)\right|^{-n_*/2}.$$

To approximate the ratio of integrals in (2.38), we make N draws from $p_3(\Pi_1|D)$, evaluate $\tilde{\delta}_1 f(\Pi_1)$ and $f(\Pi_1)$ for each draw and then compute

$$\sum_{i=1}^{N} \tilde{\delta}^{(i)} f(\Pi_1^{(i)}) \Big/ \sum_{i=1}^{N} f(\Pi_1^{(i)}), \tag{2.39}$$

where $\tilde{\delta}_1^{(i)}$ is $\tilde{\delta}_1$ evaluated at $\Pi_1 = \Pi_1^{(i)}$. The marginal covariance matrix of δ_1 is defined as the sum of the expectation of the conditional variance and the variance of the conditional expectation, i.e.,

$$V(\delta_1|D) = \int \frac{\nu_1 s_1^2}{\nu_1 - 2} (W_1' M_1 W_1)^{-1} p_2(\Pi_1|D)\,\mathrm{d}\Pi_1$$

$$+ \int (\tilde{\delta}_1 - \mathrm{E}(\tilde{\delta}_1))(\tilde{\delta}_1 - \mathrm{E}(\tilde{\delta}_1))' p_2(\Pi_1|D)\,\mathrm{d}\Pi_1. \tag{2.40}$$

Each integral in the formula above can also be approximated by ratios of sums.

To compute the posterior density of an element of δ_1, say δ_{1i}, we integrate (2.36a) analytically to obtain the conditional posterior pdf for δ_{1i}, $p(\delta_{1i}|\Pi_1, D)$, which is in the form of a univariate Student-t pdf with ν_1 degrees of freedom. Then we consider

$$p(\delta_{1i}|D) = \int p(\delta_{1i}|\Pi_1, D) p_2(\Pi_1|D)\,\mathrm{d}\Pi_1, \tag{2.41}$$

with $p_2(\Pi_1|D)$ given in (2.36b). A Monte Carlo numerical integration procedure can be employed to evaluate the integral in (2.41). To approximate $p(\delta_{1i}|D)$ at a given value of δ_{1i}, say δ_{1i}^*, compute simply

$$\sum_{i=1}^{N} p(\delta_{1i}^*|\Pi_1^{(i)}, D) f(\Pi_1^{(i)}) \Big/ \sum_{i=1}^{N} f(\Pi_1^{(i)}).$$

In this way, complete marginal posterior pdfs for the elements of δ_1 can be calculated. Also joint posterior pdfs for δ_{1i} and δ_{1j} can be calculated in a similar manner since, from (2.36b), $p(\delta_{1i}, \delta_{1j}|\Pi_1, D)$ has a bivariate Student-t form and $\int p(\delta_{1i}, \delta_{1j}|\Pi_1, D) p_2(\Pi_1|D)\,\mathrm{d}\Pi_1$ can be evaluated using Monte Carlo integration procedures. Finally, we note that (2.33) can be integrated analytically with respect to the elements of δ_1, ω_1 and Ω_1 to obtain $p(\sigma_1^2|\Pi_1, D) p_2(\Pi_1|D)$ and numerical integration procedures can be utilized to obtain the marginal posterior pdf for σ_1^2, $p(\sigma_1^2|D)$.

Above, we have employed the diffuse prior assumptions in (2.32). As an alternative, we can use the following informative prior density:

$$p(\boldsymbol{\delta}_1, \Pi_1, \Omega^*) = p_1(\boldsymbol{\delta}_1, \Pi_1|\Omega^*)\, p_2(\Omega^*), \tag{2.42}$$

where $p_1(\boldsymbol{\delta}_1, \Pi_1|\Omega^*)$ is a multivariate normal density with mean $(\bar{\boldsymbol{\delta}}_1, \bar{\Pi}_1)$ and covariance matrix $\Omega^* \otimes C^{-1}$ and $p_2(\Omega^*)$ is an inverted Wishart form. With this prior, operations similar to those presented above in the case of a diffuse prior are easily performed given values of $\bar{\boldsymbol{\delta}}_1$, $\bar{\Pi}_1$, C and other prior parameters. It is also possible to use an informative prior for $\boldsymbol{\delta}_1$ given Ω^* and diffuse priors for the other parameters.

Various conditional posterior densities associated with (2.35) are now considered. If we condition on $X\Pi_1 = Y_1 - K\hat{V}_1$, where $K > 0$ is a given constant and $\hat{V}_1 = Y_1 - X\hat{\Pi}_1$, we have $X\Pi_1 = (1-K)Y_1 + KX\hat{\Pi}_1$ or $V_1 = K\hat{V}_1$, where $V_1 = Y_1 - X\Pi_1$. Then on defining $\hat{M}_1 = I - \hat{V}_1(\hat{V}_1'\hat{V}_1)^{-1}\hat{V}_1'$ and $\tilde{\boldsymbol{\delta}}_c' = (\tilde{\boldsymbol{\gamma}}_{1c}', \tilde{\boldsymbol{\beta}}_{1c}')$, the conditional posterior mean value, given by $\boldsymbol{\delta}_{1c} = (W_1'\hat{M}_1W_1)^{-1}W_1'\hat{M}_1\boldsymbol{y}_1$, is by direct evaluation

$$\tilde{\boldsymbol{\delta}}_{1c} = \begin{pmatrix} \tilde{\boldsymbol{\gamma}}_{1c} \\ \boldsymbol{\beta}_{1c} \end{pmatrix} = \begin{pmatrix} Y_1'Y_1 - K\hat{V}_1'\hat{V}_1 & X_1'Y_1 \\ Y_1'X_1 & X_1'X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_1' - K\hat{V}_1 \\ X_1' \end{pmatrix} \boldsymbol{y}_1. \tag{2.43}$$

With these conditioning assumptions, $\tilde{\boldsymbol{\delta}}_{1c}$, the conditional posterior mean of $\boldsymbol{\delta}_1$, is in the form of a K-class estimate. As is well known, for $K = 1$, $\tilde{\boldsymbol{\delta}}_{1c}$ is the 2SLS estimate, for $K = \lambda$, the smallest root of a determinantal equation encountered in maximum likelihood estimation, $\tilde{\boldsymbol{\delta}}_{1c}$ is the LIML estimate, and for $K = 1 - k/(\nu - 2)$, with $\nu = n - k - m_1 > 2$, $\tilde{\boldsymbol{\delta}}_{1c}$ is the MELO estimate; see Zellner (1986). Note that if $K = 0$, $\tilde{\boldsymbol{\delta}}_{1c}$ is the OLS estimate defined for $K = 0$. While the above conditional results are interesting, it is often the case that conditional means, etc. are not very good approximations to unconditional means, etc. in small or even moderate sized samples. This is illustrated in computed examples presented in section 4.

We end this section with two remarks. First, the model (2.13) or (2.13′) does not include a reduced form equation for Y_0, the endogenous variables excluded from the structural equation. This means that, in fact, our analysis in this section is conditional on the hypothesis that Y_0 is independent of y_1 and Y_1. This hypothesis can be suppressed easily and the Bayesian analysis of the RRF can be adapted to the more general case. We note that one may interpret the model (2.13′) as an incomplete simultaneous equation model [see Richard (1984)]. Second, we did not discuss conditions for the existence of the marginal posterior moments of $\boldsymbol{\delta}_1$. Given that our approach of computing posterior moments may be considered as an alternative to Drèze's (1976) approach, one may argue that Drèze's discussion of existence conditions [see also Drèze and Richard (1983)] is also applicable to our case. A more explicit discussion of conditions for existence of moments will be given in future work.

3. Some Bayesian diagnostics for the model specification

In this section we extend the computational procedures of the previous section in order to compute posterior moments and densities of parameters (or functions of parameters) that give diagnostic checks of the specification of the model (2.13) or, equivalently, (2.13′).

First, we discuss how to check the hypothesis of weak *exogeneity* [as defined by Engle et al. (1983)] of the included endogenous variables Y in eq. (2.13′).[6] In non-Bayesian econometrics this can be done by testing whether $\eta_1 = 0$ in the expanded first equation of (2.13′), which is written as

$$y_1 = Y_1\gamma_1 + X_1\beta_1 + \hat{V}_1\eta_1 + \varepsilon_1, \tag{3.1}$$

where $\hat{V}_1 = Y_1 - X\hat{\Pi}_1$ is the $n \times m_1$ matrix of ordinary least squares residuals of the set of reduced form equations for Y_1. [For details see, e.g., Hausman (1983), Holly (1982) and Engle (1984, ch. 9.3).] In our unrestricted reduced form (URF) approach one may proceed as follows:

(i) Use independent random drawings $\Pi_1^{(1)},\dots,\Pi_1^{(i)},\dots,\Pi_1^{(N)}$, that are generated from a matrix Student-t distribution with a density function proportional to (2.19) and compute the sequence $V_1^{(1)},\dots,V_1^{(i)},\dots,V_1^{(N)}$ where $V_1^{(i)} = Y_1 - X\Pi_1^{(i)}$, $i = 1,\dots,N$.

(ii) Run N ordinary least squares regressions on (3.1) with $V_1^{(i)}$ instead of $\hat{V}_1$. This yields the sequence $\hat{\eta}_1^{(1)},\dots,\hat{\eta}_1^{(i)},\dots,\hat{\eta}_1^{(N)}$ where $\hat{\eta}_1^{(i)}$ is the well-known OLS expression.

(iii) Compute the moments and densities of the elements of the vector $\hat{\eta}_1$ by standard sampling theory formulas. If the posterior density of $\hat{\eta}_1$ is located around zero, one has an indication that the variables Y_1 in eq. (2.13′) are weakly exogenous in the sense that the stochastic component V_1 of the variables Y_1 does not contribute much to the eq. (2.13′). The smaller the dispersion of $\hat{\eta}_1$ around zero the greater one's confidence in this indication.

The sequence $(\hat{\gamma}_1'^{(i)}, \hat{\beta}_1'^{(i)})$, $i = 1,\dots,N$, that is obtained in the OLS regression described in step (ii) above is equal to the sequence $\{\delta_1^{*(i)}\}$, $i = 1,\dots,N$, that is obtained by using the 2SLS mapping (2.26). This follows by direct verification. As a consequence, one expects that the sample mean $\hat{\eta}_1$ from the sequence $\{\hat{\eta}_1^{(i)}\}$, $i = 1,\dots,N$, contains an approximation error with respect to η_1 when the system (2.13′) is strongly overidentified since $\hat{V}_1 \neq V_1$ in general.

[6] For earlier Bayesian results on testing for exogeneity, see Reynolds (1980, 1982).

In order to deal with the overidentified case in an exact way, we consider again the RRF system (2.13) and (2.13′) and reformulate this model as follows. First, denote the ith row of $(\boldsymbol{u}_1\ V_1)$ by $(u_1\ \boldsymbol{v}_{1i}')$ and decompose the $(1+m_1)$-multivariate normal density of $(u_1\ \boldsymbol{v}_{1i}')$ as a conditional normal density of u_i given a value of $\boldsymbol{v}_{1i}'$ and a marginal multivariate normal density of $\boldsymbol{v}_{1i}$. This yields $(u_1|\boldsymbol{v}_{1i}') \sim \mathrm{N}(\boldsymbol{v}_{1i}'\boldsymbol{\eta}_1, \sigma_1^2 - \boldsymbol{\omega}_1'\Omega_1^{-1}\boldsymbol{\omega}_1)$ with $\boldsymbol{\eta}_1 = \Omega_1^{-1}\boldsymbol{\omega}_1$ and $\boldsymbol{v}_{1i} \sim \mathrm{N}(\mathbf{0}, \Omega_1)$. Next, perform the transformation of random variables from $(\boldsymbol{u}_1|V_1)$ to $(y_1|Y_1)$ and from V_1 to Y_1. This yields

$$(y_1|Y_1) \sim \mathrm{N}\big(Y_1\gamma_1 + X_1\boldsymbol{\beta}_1 + (Y_1 - X\Pi_1)\boldsymbol{\eta}_1, (\sigma_1^2 - \boldsymbol{\omega}_1'\Omega_1^{-1}\boldsymbol{\omega}_1)I\big), \tag{3.2}$$

$$Y_1 \sim \mathrm{N}(X\Pi_1, \Omega_1 \otimes I). \tag{3.3}$$

From (3.2) and (3.3) one can write the model

$$\begin{aligned} y_1 &= Y_1\gamma_1 + X_1\boldsymbol{\beta}_1 + V_1\boldsymbol{\eta}_1 + \boldsymbol{\varepsilon}, \\ Y_1 &= X\Pi_1 + V_1, \end{aligned} \tag{3.4}$$

where $(\varepsilon_i, \boldsymbol{v}_{1i}')$, $i = 1, \ldots, n$, are independent random drawings from a multivariate normal distribution with mean zero and covariance matrix

$$\begin{bmatrix} \sigma_\varepsilon^2 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 - \boldsymbol{\omega}_1'\Omega_1^{-1}\boldsymbol{\omega}_1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{bmatrix}. \tag{3.4′}$$

Note that $\mathrm{cov}(\varepsilon_i, \boldsymbol{v}_{1i}) = \mathbf{0}$ which follows from direct verification. Therefore, testing whether $\boldsymbol{\omega}_1 = \mathbf{0}$ in the model given in (2.13′) and (2.13″) is equivalent to testing whether $\boldsymbol{\eta}_1 = \mathbf{0}$ in the model given in (3.4) and (3.4′). Further, note that if $\boldsymbol{\eta}_1 = -\gamma_1$, one can substitute $X\Pi_1 = Y_1 - V_1$ in the first equation of (3.4). As a consequence, there are only predetermined variables on the right-hand side of eq. (3.4).

The likelihood function of the parameters $\boldsymbol{\delta}_1' = (\gamma_1', \boldsymbol{\beta}_1')$, $\boldsymbol{\eta}_1$, σ_ε^2 and Ω_1 is obtained from (3.4) and (3.4′) as

$$\begin{aligned} l(\boldsymbol{\delta}_1, \boldsymbol{\eta}_1, \sigma_\varepsilon^2, \Omega_1|D) \propto (\sigma_\varepsilon^2)^{-n/2}\exp\{-\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/2\sigma_\varepsilon^2\}|\Omega_1|^{-n/2} \\ \times\exp\{-\tfrac{1}{2}\mathrm{tr}[(V_1'V_1)^{-1}\Omega_1^{-1}]\}, \end{aligned} \tag{3.5}$$

where $\boldsymbol{\varepsilon}$ and V_1 are given by equations in (3.4). As a next step we have to transform the prior density on $(\boldsymbol{\delta}_1, \Pi_1, \sigma_1^2, \boldsymbol{\omega}_1, \Omega_1)$ [see (2.32)] to a prior density on the parameter set $(\boldsymbol{\delta}_1, \Pi_1, \sigma_\varepsilon^2, \boldsymbol{\eta}_1, \Omega_1)$. The relevant part is the transformation from $(\sigma_1^2, \boldsymbol{\omega}_1, \Omega_1)$ to $(\sigma_\varepsilon^2, \boldsymbol{\eta}_1, \Omega_1)$ which gives as Jacobian $|\Omega_1|$.

As a consequence the prior information specified in (2.32) is given in terms of $(\boldsymbol{\delta}_1, \Pi_1, \sigma_\varepsilon^2, \boldsymbol{\eta}_1, \Omega_1)$ as

$$p(\boldsymbol{\delta}_1, \Pi_1, \sigma_\varepsilon^2, \boldsymbol{\eta}_1, \Omega_1) \propto (\sigma_\varepsilon^2)^{-(m_1+v_0+2)/2} |\Omega_1|^{-(m_1+\nu_0)/2}. \tag{3.6}$$

The posterior density of the p-vector $\boldsymbol{\theta}' = (\boldsymbol{\delta}_1', \boldsymbol{\eta}_1')$, with $p = l_1 + m_1$, and $\Pi_1, \sigma_\varepsilon^2, \Omega_1$ is given by

$$p(\boldsymbol{\theta}, \Pi_1, \sigma_\varepsilon^2, \Omega_1|D) \propto (\sigma_\varepsilon^2)^{-(n_*+m_1+2)/2} \exp\{-\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/2\sigma_\varepsilon^2\} \times |\Omega_1|^{-(n_*+m_1)/2} \exp\{-\tfrac{1}{2}\mathrm{tr}[(V_1'V_1)\Omega_1^{-1}]\}. \tag{3.7}$$

Integrating (3.7) with respect to σ_ε^2 and Ω_1 yields the marginal posterior $p(\boldsymbol{\theta}, \Pi_1|D)$,

$$p(\boldsymbol{\theta}, \Pi_1|D) \propto (\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})^{-(n_*+m_1)/2} |V_1'V_1|^{-(n_*-1)/2}, \tag{3.8}$$

where $\boldsymbol{\varepsilon}$ and V_1 are given in (3.4). The density (3.8) may be compared with the pdf given in (2.34) and (2.35). In a similar way as done below (2.35) in subsection 2.3, we can rewrite (3.8) as

$$p(\boldsymbol{\theta}, \Pi_1|D) = p_1(\boldsymbol{\theta}|\Pi_1, D)\, p_2(\Pi_1|D), \tag{3.9}$$

where

$$p_1(\boldsymbol{\theta}|\Pi_1, D) = c_1 \left(\frac{W'W}{s_1^2}\right)^{1/2} \times \{\nu_1 + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})'W'W(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})/s_1^2\}^{-(\nu_1+p)/2}, \tag{3.9a}$$

and

$$p_2(\Pi_1|D) \propto c_2 h(\Pi_1) c_3 |S_1 + (\Pi_1 - \hat{\Pi}_1)'X'X(\Pi_1 - \hat{\Pi}_1)|^{-(n_*-1)/2}, \tag{3.9b}$$

and

$$h(\Pi_1) = |W'W|^{-1/2}(s_1^2)^{-\nu_1/2}. \tag{3.9c}$$

From the definition of $W = (W_1 \; V_1)$ it follows that $|W'W| = |W_1'M_1W_1||V_1'V_1|$, where M_1 is as given below (2.34). Therefore

$$h(\Pi_1) = f(\Pi_1)|V_1'V_1|^{-1/2}, \tag{3.10}$$

with $f(\Pi_1)$ given in (2.36c). It follows that the posterior density of Π_1 given in (2.36b) is equivalent to the posterior density given in (3.9b). The parameters of the conditional multivariate Student-t density of the p-vector $\boldsymbol{\theta}$ are given as $\nu_1 = n_* - l_1$, and

$$\tilde{\boldsymbol{\theta}} = (W'W)^{-1}W'y_1, \qquad \nu_1 s_1^2 = (y_1 - W\tilde{\boldsymbol{\theta}})'(y_1 - W\tilde{\boldsymbol{\theta}}). \tag{3.11}$$

The conditional density $p_1(\boldsymbol{\theta}|\Pi_1, D)$ in (3.9a) is in the form of a p-variate conditional Student-t pdf of $\boldsymbol{\theta}$ given Π_1 and D with ν_1 degrees of freedom, with mean $\tilde{\boldsymbol{\theta}}$ and covariance matrix $(W'W)^{-1}\nu_1 s_1^2/(\nu_1 - 2)$, both of which depend on Π_1. Similar remarks that were made with respect to (2.36a) apply to (3.9a) and are not repeated. We mention here only that if the marginal pdf of $\boldsymbol{\eta}_1$ is centered around zero, then one has an indication that Y_1 is weakly exogenous in the sense discussed before.

We note that one may use diffuse or informative priors other than (3.6). For instance an alternative type of diffuse prior is given by

$$p(\boldsymbol{\theta}, \Pi_1, \sigma_\varepsilon^2, \Omega_1) \propto (\sigma_\varepsilon^2)^{-(m_1+\nu_0+2)/2}|W'W|^{1/2}|\Omega_1|^{-(m_1+\nu_0)/2}. \tag{3.12}$$

This prior is equal to (3.6) times $|W'W|^{1/2}$, which is the root of the determinant of the information matrix of $\boldsymbol{\theta}$ given Π_1. As a result the factor $|W'W|^{-1/2}$ will not appear in (3.9c). Further, we note that conditional moments associated with (3.9) can be formulated in a similar way as was done in subsection 2.3. In particular, if we condition $p_1(\boldsymbol{\theta}|\Pi_1, D)$ on $\Pi_1 = \hat{\Pi}_1$ and integrate out $\boldsymbol{\delta}_1$, the posterior density $p_1(\boldsymbol{\eta}_1|\hat{\Pi}_1, D)$ is an m_1-variate Student pdf with mean $\hat{\boldsymbol{\eta}}_1$, the OLS estimate of $\boldsymbol{\eta}_1$ in (3.1). The non-Bayesian test procedure for the weak exogeneity of Y_1 using (3.1) is to reject the null hypothesis if a $(1-\alpha)\%$ confidence region centered at $\hat{\boldsymbol{\eta}}_1 = 0$ does not contain the point $\hat{\boldsymbol{\eta}}_1 = 0$. The Bayesian decision is to reject the null if a $(1-\alpha)\%$ posterior probability region centered at $\hat{\boldsymbol{\eta}}_1$ does not contain the point $\boldsymbol{\eta}_1 = 0$. An exact Bayesian decision procedure relies on the marginal posterior density $p_1(\boldsymbol{\eta}_1|D)$ rather than on the conditional density $p_1(\boldsymbol{\eta}_1|\hat{\Pi}_1, D)$. Some illustrative results on exogeneity testing are presented in subsection 4.1.

Next, we discuss how we can check whether the overidentifying restrictions in (2.12a) and (2.12b) seem acceptable. It follows from the discussion, given in subsection 2.2 [between eqs. (2.20) and (2.21)], that the degree of overidentification is equal to the number k_0 of omitted predetermined variables in eq.

(2.9b) minus the number m_1 of included endogenous variables on the right-hand side of (2.9b). Thus, we may include some predetermined variables in (2.9b) that were, at first, excluded from this equation. If we add $k_0 - m_1$ predetermined variables to the right-hand side of (2.9b), then we have an exactly identified equation instead of an overidentified equation. As a consequence, one can make use of the URF approach and compute highest posterior density (HPD) regions around zero for the parameters of the $k_0 - m_1$ included variables. This yields a check on the value of the overidentifying restrictions. If we add fewer than $k_0 - m_1$ predetermined variables to (2.9b), then this equation is still overidentified and the RRF approach can be used to analyze the HPD regions around zero of the parameters of the included variables.

Several other diagnostic checks may be constructed, i.e., restricted reduced form moments may be compared with unrestricted reduced form moments. Diagnostic checks on autocorrelation and outliers may be constructed from posterior distributions of realized error terms [see van Dijk (1987)]. Further, one may compute posterior odds relating to exogeneity hypotheses. There are thus ample opportunities for much applied work using the methods discussed above.

4. Applications of methods

In this section we illustrate the methods of sections 2 and 3 for the case of an exactly identified simultaneous equation model and for the case of an overidentified model. As an example of an exactly identified model we consider the Belgian beef market model [see Drèze and Richard (1983, sect. 2.4)] which is given as

$$Q_t = \alpha_1 + \beta_1 P_t + \gamma_1 Y_t + u_{1t}, \tag{4.1}$$

$$Q_t = \alpha_2 + \beta_2 P_t + \gamma_2 S_t + u_{2t}, \tag{4.2}$$

where Q_t is the quantity of beef consumed per capita in period t; P_t is the price index; Y_t is real national income per capita; and S_t is the cattle stock per capita (measured as the number of heads at the beginning of each period). The variables Q_t and P_t are endogenous, and the variables Y_t, S_t and the constant term are assumed exogenous. The data are annual observations for the period 1950–1965. Given our uniform prior with $\nu_0 = 0$ and given that the model is exactly identified, posterior first- and higher-order moments do not exist. In fig. 1 we present the marginal posterior density of β_1 and give the computed quartiles of the posterior distribution. The density is concentrated around the mode but has a long tail to the left. We note that the mode and the median are almost equal; however, the first and fourth quartiles indicate that the density is skewed to the left. Further, we find evidence that the exogeneity of the price

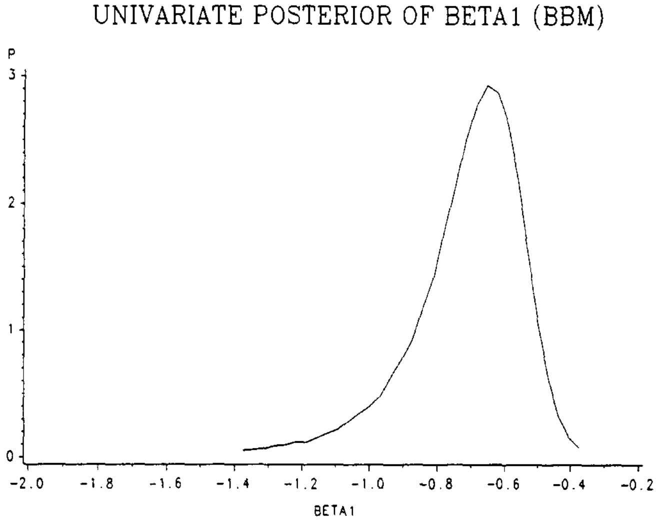


Fig. 1. Marginal posterior density of β_1 in the Belgian beef market model [eq. (4.1)].

variable is rejected. The results reported are based on $N = 100{,}000$ drawings in order to obtain an accurate figure. We emphasize, however, that the figure is already rather accurate with $N = 10{,}000$ or $N = 20{,}000$.

As an example of an overidentified simultaneous equation model we take Klein's Model I [see Klein (1950)], which is given as

$$C = \alpha_1 P + \alpha_2 P_{-1} + \alpha_3 W + \alpha_0 + u_1, \tag{4.3}$$

$$I = \beta_1 P + \beta_2 P_{-1} + \beta_3 K_{-1} + \beta_0 + u_2, \tag{4.4}$$

$$W_1 = \gamma_1 X + \gamma_1 X_{-1} + \gamma_3 t + \gamma_0 + u_3, \tag{4.5}$$

$$X = C + I + G, \tag{4.6}$$

$$P = X - W_1 - T, \tag{4.7}$$

$$K = K_{-1} + I, \tag{4.8}$$

$$W = W_1 + W_2. \tag{4.9}$$

Consumption expenditure (C) is structurally dependent on profits (P), on profits lagged one year (P_{-1}) and on total wages (W). Net investment expenditure (I) depends on profits, lagged profits and on the capital stock at the beginning of the year (K_{-1}). Finally, private wage income (W_1) depends on net private product at market prices (X), the same variable lagged (X_{-1}) and a trend term (t). The model is closed by four identities, which provide links with three exogenous variables: the government wage bill (W_2), government non-wage expenditure, including the net foreign balance, (G) and

Table 1

Investment equation of Klein's Model I: Posterior expectations and standard deviations.[a]

	β_1	β_2	β_3	β_0	η_{1P}
Prior with $\nu_0 = 0$					
URFA with prior (2.17)					
GILS	0.59 (0.45)	0.48 (0.45)	−0.05 (0.11)	0.75 (22.76)	
TSLS	0.35 (0.19)	0.45 (0.18)	−0.13 (0.04)	14.20 (8.40)	0.31 (0.15)
RRFA with prior (2.32) or (3.6)					
conditional on $\hat{\Pi}_1^b$	0.15 (0.11)	0.62 (0.10)	−0.16 (0.02)	20.28 (4.86)	0.57 (0.15)
marginal	0.20 (0.17)	0.57 (0.17)	−0.15 (0.04)	19.02 (8.56)	0.50 (0.19)
Prior with $\nu_0 = k$					
URFA with prior (2.17)					
GILS	0.54 (0.42)	0.52 (0.39)	−0.06 (0.10)	3.81 (20.84)	
TSLS	0.27 (0.16)	0.51 (0.15)	−0.14 (0.03)	16.54 (6.83)	0.39 (0.13)
RRFA with prior (2.32) or (3.6)					
conditional on $\hat{\Pi}_1^b$	0.15 (0.09)	0.62 (0.09)	−0.16 (0.02)	20.28 (3.92)	0.57 (0.12)
marginal	0.17 (0.15)	0.60 (0.15)	−0.16 (0.03)	19.83 (6.95)	0.55 (0.16)

[a] Quantities in parentheses are posterior standard deviations.
[b] The conditional posterior means are equal to the non-Bayesian two-stage least squares estimates as explained in section 2.3.

Table 2

Wage equation of Klein's Model I: Posterior expectations and standard deviations,[a] prior with $\nu_0 = 0$.

	γ_1	γ_2	γ_3	γ_0	η_{1X}
URFA with prior (2.17)					
GILS	0.50 (0.15)	0.03 (0.13)	0.17 (0.19)	−4.37 (15.18)	
TSLS	0.44 (0.04)	0.15 (0.05)	0.13 (0.03)	1.52 (1.16)	0.00 (0.05)
RRFA with prior (2.32) or (3.6)					
conditional on $\hat{\Pi}_1$	0.44 (0.04)	0.15 (0.05)	0.13 (0.03)	1.50 (1.36)	0.00 (0.07)
marginal	0.44 (0.07)	0.15 (0.07)	0.13 (0.04)	1.47 (1.52)	0.00 (0.11)

[a]Quantities in parentheses are posterior standard deviations.

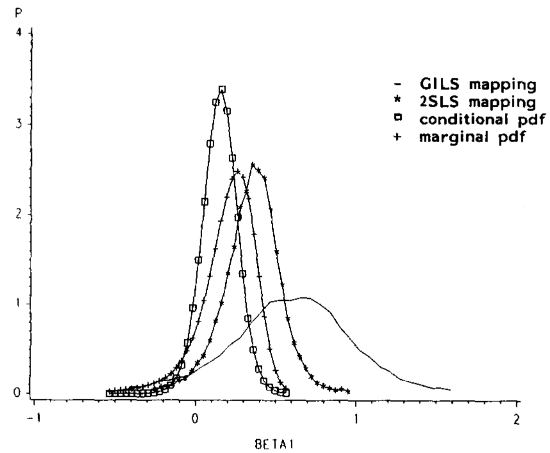
Table 3

Consumption equation of Klein's Model I: Posterior expectations and standard deviations.[a]

	α_1	α_2	α_3	α	η_{1P}	η_{1W}
Prior with $\nu_0 = 0$						
URFA with prior (2.17)						
GILS	0.47 (0.34)	0.04 (0.24)	0.41 (0.25)	15.74 (10.73)		
TSLS	0.10 (0.14)	0.15 (0.12)	0.81 (0.04)	16.33 (1.36)	0.47 (0.22)	−0.33 (0.21)
RRFA with prior (2.32) or (3.6)						
conditional on $\hat{\Pi}_1$	0.02 (0.10)	0.22 (0.09)	0.81 (0.03)	16.55 (1.07)	0.69 (0.21)	−0.45 (0.24)
marginal	−0.08 (0.16)	0.29 (0.17)	0.83 (0.05)	16.15 (2.09)	0.79 (0.17)	−0.47 (0.11)
Prior with $\nu_0 = k$						
URFA with prior (2.17)						
GILS	0.42 (0.32)	0.07 (0.21)	0.42 (0.23)	18.11 (9.77)		
TSLS	0.07 (0.11)	0.18 (0.10)	0.81 (0.03)	16.43 (1.07)	0.54 (0.19)	−0.37 (0.18)
RRFA with prior (2.32) or (3.6)						
conditional on $\hat{\Pi}_1$	0.02 (0.08)	0.22 (0.07)	0.81 (0.03)	16.55 (0.86)	0.69 (0.17)	−0.45 (0.19)
marginal	−0.08 (0.13)	0.33 (0.13)	0.80 (0.03)	16.62 (1.05)	0.79 (0.15)	−0.44 (0.07)

[a]Quantities in parentheses are posterior standard deviations.

UNIVARIATE POSTERIOR OF BETA1 (KLEIN 1)



UNIVARIATE POSTERIOR OF ETA1 (KLEIN 1)

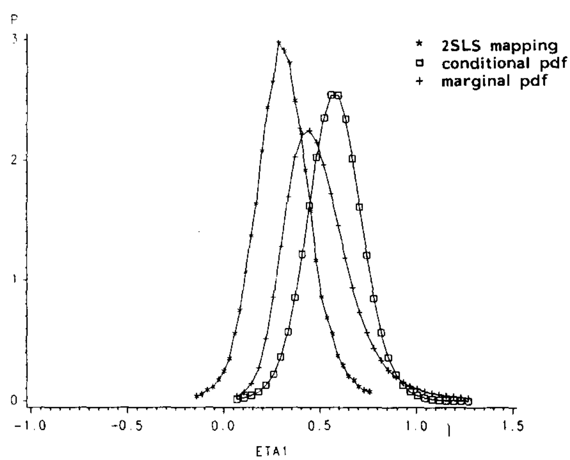


Fig. 2. Univariate marginal posterior densities of β_1 and η_{1P} in the investment equation of Klein's Model I.

business taxes (T). The model has seven jointly dependent variables (C, I, W_1, X, P, W) and eight predetermined variables ($1, P_{-1}, X_{-1}, K_{-1}, G, T, W_2, t$). All variables (except 1 and t) are measured in constant dollars. Posterior moments for Klein's Model I are reported in tables 1–3 and univariate and bivariate marginal posterior densities of a structural parameter and an exogeneity parameter in the investment equation are given in figs. 2 and 3. It is seen from the results on the investment equation in table 1 that the

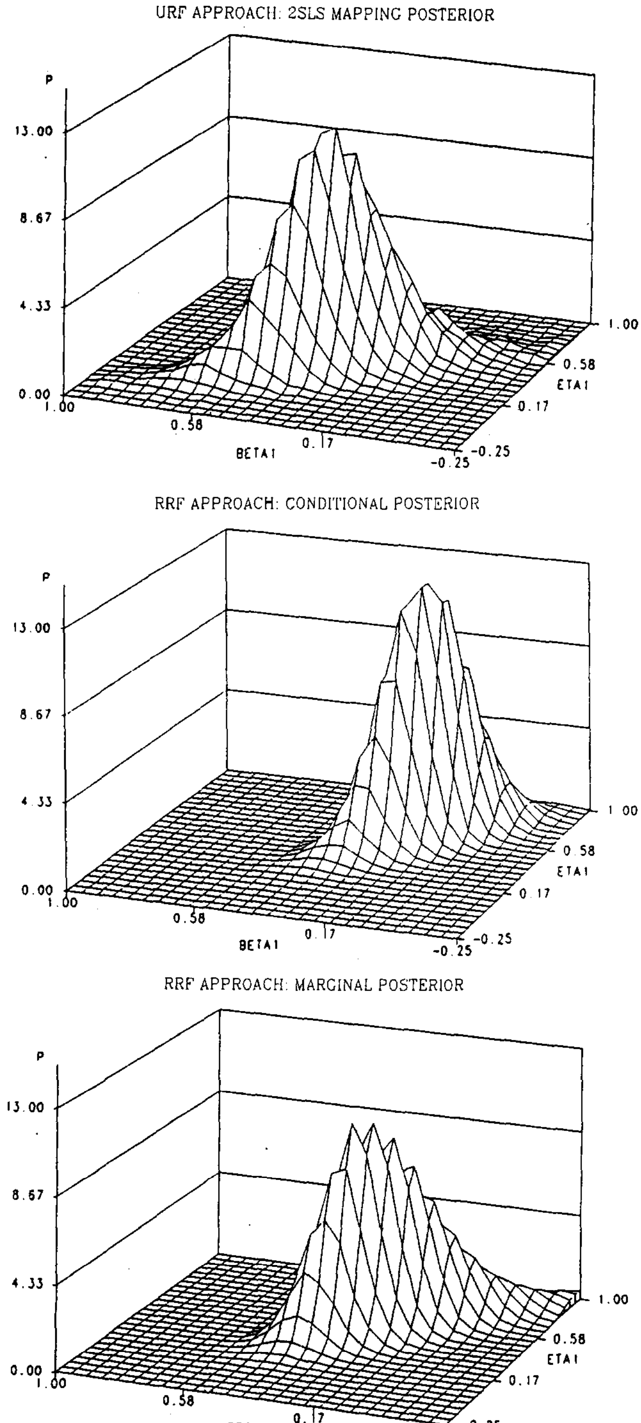


Fig. 3. Bivariate marginal posterior density functions for (β_1, η_{1P}) in the investment equation of Klein's Model I.

URF approach, in particular the GILS mapping, yields gross approximation errors for several parameters. The posterior means and standard deviations of the parameter of the included endogenous variable, of the constant term, and of the exogeneity parameter differ substantially from the results of the restricted reduced form approach. The results of the latter approach are based on $N = 20{,}000$ drawings. We note that the marginal results differ also from the conditional results in the RRF approach but less than from the results given by the URF approach. The sensitivity with respect to the particular choices of $\nu_0 = 0$ and $\nu_0 = k$ is as expected. A larger value of ν_0 implies smaller variances due to lighter tails. It is of interest that the exogeneity of profits appears to be rejected while some preliminary results on overidentifying restrictions (not reported) suggest that these restrictions are not to be rejected. More details will be reported in future work. It is also of interest that conditional standard deviations are always smaller than the asymptotic TSLS standard deviations. The reason is that in the conditional approach the values of s_1^2 is smaller than in the non-Bayesian approach. The results for the wage income equation given in table 2 produced by different methods are similar. The hypothesis that net private product is exogenous is not rejected while, for preliminary results, it appears that the overidentifying restrictions are rejected. The consumption function was the most complex case to analyze. The posterior means differ substantially for the different approaches. The posterior standard deviations for the exogeneity parameters for profits and wage income show a surprising result. The marginal standard deviations are smaller than the conditional ones. It appears that the effect of the weight function $f(H_1)$ (see subsection 2.3) is very non-linear. This is a topic of current research. Exogeneity and preliminary results on over identification, not reported here, suggest that both hypotheses are rejected. Figs. 2 and 3 show the skewness of the marginal pdf's and differences between the results of the URF, the conditional RRF and the marginal RRF approaches.

5. Concluding remarks

In this paper, we have shown how Monte Carlo numerical methods can be employed to compute exact posterior densities of the parameters of a structural equation using diffuse or informative prior distributions. In addition, operational procedures for Bayesian diagnostic checking or specification analysis were described. For example, discrepancy parameter vectors were introduced to represent departures from exact identifying restrictions and it was shown how to compute posterior densities for them and interesting functions of their elements which we refer to as discrepancy functions. In addition, a Bayesian procedure for evaluating exogeneity hypotheses was described. That diagnostic checking or specification analysis be performed is quite important and the fact that operational Bayesian procedures for diagnostic checking or specification analysis can be carried through without much difficulty is fortunate.

Applications of our methods were presented and yielded useful results. In particular, it was found in several instances that certain specifying assumptions, exogeneity hypotheses and identifying restrictions, were of doubtful validity. Also, it was found that exact marginal posterior densities differed considerably from conditional posterior densities based on conditioning assumptions which are often employed in non-Bayesian procedures, for example in the 2SLS approach or other K-class estimation approaches. Thus we consider it very important to use appropriate marginal posterior densities for structural parameters rather than approximate conditional posterior densities. That the former can be computed using Monte Carlo techniques without much difficulty is indeed fortunate.

In future research, we plan to extend our consideration of diagnostic checking procedures to consider checks for autocorrelation of error terms, outliers and possible left out variables. Also, the single-equation analysis in this paper will be extended to provide results for sets of structural equations and complete structural equation systems.

Appendix: The generation of pseudo-random drawings from a matrix Student distribution

Because the matrix Student (Mt) distribution is related to the matrix Normal (MN) and to the inverted Wishart (iW) distributions, we define these three families of distributions through their density functions and state a few properties that are useful to build an algorithm for generating a pseudo-random drawing from an Mt distribution.

A.1. Definitions

Let $\Pi \in R^{km}$ be a $k \times m$ random matrix.

(i) Π has an MN distribution if its density function is

$$\begin{aligned} p(\Pi) &= f_{\mathrm{MN}}^{k\times m}(\Pi|\bar{\Pi}, \Omega \otimes M^{-1}) \\ &:= \left[(2\pi)^{km}|\Omega|^{k}|M|^{m}\right]^{-1/2} \\ &\quad \times \exp - \tfrac{1}{2}\operatorname{tr}\left[\Omega^{-1}(\Pi - \bar{\Pi})' M(\Pi - \bar{\Pi})\right], \end{aligned} \tag{A.1}$$

where $\bar{\Pi} \in R^{km}$ is a $k \times m$ constant matrix, Ω is an $m \times m$ PDS constant matrix and M is a $k \times k$ PDS constant matrix.

From here on, let Ω be a random PDS matrix.

(ii) Ω has an iW distribution if its density function is

$$p(\Omega) = f_{\mathrm{iW}}^{m}(\Omega | W, \nu)$$

$$:= \left[2^{\nu m/2} \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma\left(\frac{\nu + 1 - i}{2} \right) \right]^{-1}$$

$$\times |W|^{\nu/2} |\Omega|^{-(\nu+m+1)/2} \exp - \tfrac{1}{2} \operatorname{tr} \Omega^{-1} W, \tag{A.2}$$

where W is an $m \times m$ constant matrix and the constant $\nu > m - 1$.

(iii) Π has an Mt distribution if its density function is

$$p(\Pi) = f_{\mathrm{Mt}}^{k \times m}(\Pi | \overline{\Pi}, W, M, \nu)$$

$$= \left[\pi^{km/2} \prod_{i=1}^{m} \Gamma\left(\frac{\nu + 1 - i}{2} \right) \Big/ \Gamma\left(\frac{\nu + k + 1 - i}{2} \right) \right]^{-1}$$

$$\times |W|^{\nu/2} |M|^{m/2} \left| W + (\Pi - \overline{\Pi})' M (\Pi - \overline{\Pi}) \right|^{-(\nu+k)/2}, \tag{A.3}$$

where $\overline{\Pi}$, W, M and ν are defined as in (i) and (ii).

A.2. Some properties of these distributions

(1) If $p(\Pi|\Omega) = f_{\mathrm{MN}}^{k \times m}(\Pi | \overline{\Pi}, \Omega \otimes M^{-1})$ and $p(\Omega) = f_{\mathrm{iW}}^{m}(\Omega | W, \nu)$, then $p(\Pi)$ is given by formula (A.3). This property states that an Mt distribution is a marginal distribution from an MN–iW one.

(2) Let Π have the density (A.1).

(i) If A is an $r \times k$ matrix of rank $r \leq k$, and B is an $m \times s$ matrix of rank $s \leq m$, then

$$p(A\Pi B) = f_{\mathrm{MN}}^{r \times s}(A\Pi B | A\overline{\Pi} B, B'\Omega B \otimes A M^{-1} A'). \tag{A.4}$$

(ii) In particular, if $B'\Omega B = I_m$ and $AM^{-1}A' = I_k$, $Z := A(\Pi - \overline{\Pi})B$ is a matrix of independent standard normal variables.

(3) Let Ω have the density (A.2).

(i) If C is an $m \times s$ matrix of rank $s \leq m$, then

$$p(C'\Omega C) = f_{\mathrm{iW}}^{s}(C'\Omega C | C'WC, \nu - m + s). \tag{A.5}$$

(ii) Partition Ω into $\Omega_{11}(m_1 \times m_1, \mathrm{PDS})$, $\Omega_{22}(m_2 \times m_2, \mathrm{PDS})$, $\Omega_{21}(m_2 \times m_1)$, $\Omega_{12} = \Omega_{21}'$ and let $\Omega_{22 \times 1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$. Then Ω and $(\Omega_{11}, \Omega_{11}^{-1}\Omega_{12}, \Omega_{22 \times 1})$

are in one-to-one correspondence and

$$p\left(\Omega_{11}, \Omega_{11}^{-1}\Omega_{12}, \Omega_{22\times1}\right) = p(\Omega_{11})\, p\left(\Omega_{11}^{-1}\Omega_{12}|\Omega_{22\times1}\right) p(\Omega_{22\times1}), \quad \text{(A.6)}$$

with

$$p(\Omega_{11}) = f_{\mathrm{iW}}^{m_1}(\Omega_{11}|W_{11}, \nu - m_2), \quad \text{(A.7)}$$

$$p\left(\Omega_{11}^{-1}\Omega_{12}\right) = f_{\mathrm{MN}}^{m_1\times m_2}\left(\Omega_{11}^{-1}\Omega_{12}|W_{11}^{-1}W_{12}, \Omega_{22\times1} \otimes W_{11}^{-1}\right), \quad \text{(A.8)}$$

$$p(\Omega_{22\times1}) = f_{\mathrm{iW}}^{m_2}(\Omega_{22\times1}|W_{22\times1}, \nu), \quad \text{(A.9)}$$

where W_{11}, W_{22} and $W_{22\times1}$ are defined from W as Ω_{11}, Ω_{22} and $\Omega_{22\times1}$ are defined from Ω.

(iii) In particular, if $C'WC = I_m$ in (A.5), $\Psi := C'\Omega C$ is in one-to-one correspondence with $\frac{1}{2}m(m+1)$ independent random variables: $\frac{1}{2}m(m-1)$ standard normal variables, plus m variables λ_i, each of them having an inverted-gamma density defined as

$$f_{\mathrm{iW}}^{1}(\lambda_i|1, \nu - i + 1) \quad \text{for} \quad i = 1, 2, \ldots, m.$$

This follows from the property 3(ii) applied to Ψ m times: one starts, e.g., with $m_2 = 1$ and $m_1 = m - 1$ and notices that 3(ii) can be applied again to $p(\Psi_{11})$ which is an iW density with parameters I_{m_1} and $\nu - 1$.

Other properties from these distributions can be found in Zellner (1971, app. B.4, B.5), Drèze and Richard (1983, app. A) and Bauwens (1984, app. A.1, A.II, who gives separate algorithms for the generation of random numbers from MN and iW distribution). These algorithms can be combined to draw from an Mt distribution, with density given by (A.3), by drawing firstly an iW Ω matrix with density (A.2), and by drawing subsequently an MN matrix with density (A.1) where Ω is the iW matrix obtained at the iW step.

A.3. Mt algorithm

To obtain a drawing Π from the Mt distribution defined by (A.3):

(1) Compute the lower triangular (LT) matrices Q' and P such that $W = Q'Q$ and $M^{-1} = PP'$.

(2) iW step:

(i) Generate $\frac{1}{2}m(m-1)$ standard normal drawings and m inverted gamma drawings λ_i, with $p(\lambda_i) = f_{\mathrm{iW}}^{1}(\lambda_i|1, \nu - i + 1)$.

(ii) Compute the $m \times m$ LT matrix Φ such that $\Phi\Phi' =: \Omega$ is a drawing from the iW distribution of Ω defined by (A.2) (*but one does not need to compute*

$\Phi\Phi'$). Let $\Phi = (\phi_{ij})$: then $\phi_{ij} = 0$ for $i < j$. The lower triangle of Φ can be filled by the following steps:

1) $i \leftarrow 0$; $1 \leftarrow \frac{1}{2}m(m+1)+1$; let ϕ be a vector of $l-1$ elements that will finally contain the column expansion of the LT of Φ, i.e., $\phi = (\phi_{11}\phi_{21}, \dots, \phi_{m1}\phi_{22}\phi_{32}, \dots, \phi_{m-1m-1}\phi_{mm-1}\phi_{mm})$.
2) $i \leftarrow i+1$; if $i > m$, stop.
3) $l \leftarrow l - i\phi(l) = \sqrt{\lambda_i}$ [λ_i obtained at step 2(i)].
4) If $i = 1$, go to 2); or else go to 5).
5) Pick $i-1$ standard normal drawings obtained at step 2(i) and assign them in a vector u. Compute $y = \sqrt{\lambda_i}.\Phi_{i-1}.u$ where y is a vector of $i-1$ elements and Φ_{i-1} denotes the LT matrix whose column expansion of the lower triangle is stored in the last $i(i-1)/2$ elements of the vector ϕ (but Φ_1 is the scalar $\phi_{mm} = \sqrt{\lambda_1}$). Finally, $\phi(1+k) \leftarrow y(k)$, $k = 1,2,\dots,i-1$, and go to 2).

(3) MN step:

(i) Generate km standard normal drawings z_{ij} ($i = 1,2,\dots,k$ and $j = 1,2,\dots,m$). Let $Z = (z_{ij})$.

(ii) Compute $\Pi = \bar{\Pi} + PZ\Phi'Q'$ where Φ is the LT matrix obtained at step 2(ii).

To draw standard normal variables, one can use the polar algorithm – see, e.g., Knuth (1971). To draw inverted gamma variables, one can use the GRUB algorithm of Kinderman and Monahan (1980) that is efficient since the computer time required to obtain one inverted gamma drawing is almost perfectly independent of the value of ν (as is *not* the case if one generates gamma drawings as sums of ν independent squared normal drawings). To get one drawing Π, one needs $\frac{1}{2}m(m-1)+mk$ univariate standard normal drawings, plus the m inverted gamma drawings; all these drawings must be independent.

The proposed Mt algorithm has the advantage that the marginal cost of a drawing (steps 2 and 3) is not affected by the value of the degrees of freedom parameter ν. For a similar type of algorithm, where use is made of the Wishart instead of the inverted Wishart distribution, we refer to Geweke (1988).

Provided ν is an integer, one could replace the implementation of the iW step by (i) drawing a Wishart matrix Ω^{-1} as $\Sigma_{j=1}^{\nu} Z_j Z_j'$ where the $m \times 1$ independent vectors Z_j have a multivariate normal density with zero expected value and covariance matrix given by W, (ii) inverting Ω^{-1} and (iii) computing the LT matrix Φ such that $\Omega = \Phi\Phi'$. This implementation requires νm standard normal drawings at the iW step, instead of $\frac{1}{2}m(m-1)$ of these plus the m inverted gamma drawings. So for very small values of ν and m, this implementation may be more efficient. Notice however that a Cholesky decomposition of Ω, giving Φ, has to be performed, whereas Φ is obtained directly in the implementation we use.

Another method to generate from the Mt distribution (A.3) that is expected to be less efficient, is to use the property that

$$p(\Pi)=p(\Pi_1|\Pi_2\Pi_3\cdots\Pi_m)\,p(\Pi_2|\Pi_3\cdots\Pi_m)\cdots p(\Pi_m), \qquad \text{(A.10)}$$

where Π_i $(i=1,2,\dots,m)$ is the ith column of Π, and each of the densities on the right of (A.10) is a multivariate Student density [see Zellner (1971, p. 397) or Drèze and Richard (1983, p. 589)]. Formula (A.10) suggests a sequential drawing procedure.

References

Anderson, T.W., 1984, An introduction to multivariate statistical analysis, 2nd ed. (Wiley, New York).

Bauwens, L., 1984, Bayesian full information analysis of simultaneous equation models using integration by Monte Carlo (Springer-Verlag, Berlin).

Bauwens, L. and J.F. Richard, 1985, A 1-1 poly-t random variable generator with applications to Monte Carlo integrations, Journal of Econometrics 29, 19–46.

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

Dickey, J.M., 1967, Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution, Annals of Mathematical Statistics 38, 511–518.

Drèze, J.H., 1976, Bayesian limited information analysis of the simultaneous equation model, Econometrica 44, 1045–1075.

Drèze, J.H. and J.F. Richard, 1983, Bayesian analysis of simultaneous equation systems, Ch. 9 in: Z. Griliches and M. Intriligator, eds., Handbook of econometrics, Vol. 1 (North-Holland, Amsterdam).

Engle, R.F., 1984, Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, Ch. 13 in: Z. Griliches and M. Intriligator, eds., Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam).

Engle, R.F., D.F. Hendry and J.F. Richard, 1983, Exogeneity, Econometrica 51, 277–304.

Geisser, S., 1965, Bayesian estimation in multivariate analysis, Annals of Mathematical Statistics 36, 150–159.

Geweke, J., 1988, Antithetic acceleration of Monte Carlo integration in Bayesian inference, Journal of Econometrics, this issue.

Graybill, F.A., 1969, Introduction to matrices with applications in statistics (Wadsworth, Belmont, CA).

Hausman, J.A., 1983, Specification and estimation of simultaneous equation models, Ch. 7 in: Z. Griliches and M. Intriligator, eds., Handbook of econometrics, Vol. 1 (North-Holland, Amsterdam).

Holly, A., 1982, A remark on Hausman's specification test, Econometrica 50, 749–759.

Khazzoom, J.D., 1976, An indirect least squares estimator for overidentified equations, Econometrica 44, 741–750.

Kinderman, A.J. and J.F. Monahan, 1980, New methods for generating Student's and gamma variables, Computing 25, 369–377.

Klein, L.R., 1950, Economic fluctuations in the United States, 1921–1941 (Wiley, New York).

Knuth, D.E., 1971, The art of computer programming, Vol. 2 (Addison-Wesley, Reading, MA).

Reynolds, R.A., 1980, A study of the relationship between health and income, Unpublished doctoral dissertation (Department of Economics, University of Chicago, Chicago, IL).

Reynolds, R.A., 1982, Posterior odds for the hypothesis of independence between stochastic regressors and disturbances, International Economic Review 23, 479–490.

Richard, J.F., 1984, Classical and Bayesian inference in incomplete simultaneous equation models, Ch. 4 in: D.F. Hendry and K.F. Wallis, eds., Econometrics and quantitative economics (Blackwell, London).

Tsurumi, H., 1985, Limited information Bayesian analysis of a structural coefficient in a simultaneous equation system, Communications in Statistics 14, 1103–1120.

Tsurumi, H., 1987, Comparing Bayesian and non-Bayesian limited information estimators, Manuscript (Department of Economics, Rutgers University, New Brunswick, NJ).

van Dijk, H.K., 1985, A note on simple algebraic relations between structural and reduced form parameters, Report no. 8549 (Econometric Institute, Erasmus University, Rotterdam).

van Dijk, H.K., 1987, Some advances in Bayesian estimation methods using Monte Carlo integration, in: T.B. Fomby and G.F. Rhodes, eds., Advances in econometrics, Vol. 6 (JAI Press, Greenwich, CT).

Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York); reprinted in 1987 (Krieger, Malabar, FL).

Zellner, A., 1975, Bayesian analysis of regression error terms, Journal of the American Statistical Association 70, 138–144.

Zellner, A., 1979, Statistical analysis of econometric models, Journal of the American Statistical Association 74, 628–651; reprinted in: A. Zellner, 1984, Basic issues in econometrics (University of Chicago Press, Chicago, IL) 83–119.

Zellner, A., 1986, Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients, Advances in Econometrics 5, 171–182.