Chapter 18

BAYESIAN LIMITED INFORMATION ANALYSIS REVISITED

Luc BAUWENS

GREQE, Ecole des Hautes Etudes en Sciences Sociales, Marseille, France

Herman K. van DIJK

Erasmus University Rotterdam, Rotterdam, The Netherlands

1. Introduction

The econometric analysis of the laws of demand and supply makes, in many cases, use of the assumption that prices and quantities traded of economic commodities are *jointly* determined. Well known examples are the demand and supply for agricultural and financial commoditie. Econometric research is also directed towards the analysis of the *joint* dynamic behaviour of such variables as gross national product, investment, consumption, money supply, inflation, and unemployment. In particular, the secular and cyclical properties of these variables are of interest. The econometric study of market processes and of business cycle phenomena is, in this century, greatly advanced by the formulation of the *Simultaneous Equations Model* (*SEM*) (see Haavelmo, 1943).[1]

L. Bauwens acknowledges financial support from the Netherlands organisation for scientific research (NWO) for the preparation of this paper, which was started during a visit at Erasmus University in 1988. H.K. van Dijk acknowledges financial support from NWO and from GREQE/EHESS in 1989. The authors are indebted to Frank Kleibergen for research assistance and to Gerrit Draisma for assistance in preparing the figures.

[1] For a historical review and a critical evaluation of the role of the SEM we refer to Epstein (1987).

Economic Decision-Making: Games, Econometrics and Optimisation
Edited by J.J. Gabszewicz, J.-F. Richard and L.A. Wolsey

Given the specification of the SEM, estimation methods for its parameters were developed using the maximum likelihood and least squares principles. However, in an unpublished paper in 1962, Drèze argued that these classical estimation methods were inadequate in two respects. First, available information on parameters of interest is ignored. A classic example is the marginal propensity to consume, which is an unrestricted parameter in the simple Keynesian consumption function, while past experience and economic knowledge restricts it, for most countries, to a subinterval of the unit interval. Second, too much prior information is used in the sense that some variables are omitted from an equation without proper justification. For instance, the interest rate is deleted from the consumption function mentioned above. Thus, the interest elasticity of consumption is assumed to be zero in the long run. Starting from these limitations of classical estimation methods, Drèze made several contributions to the econometric analysis of the SEM from a Bayesian point of view.[2] A major result, contained in Drèze's 1976 *Econometrica* paper is the derivation of the functional form of the posterior density of the parameters of a single structural equation, which is analyzed from a limited information point of view (i.e., ignoring information on the parameters of the other structural equations).[3] This posterior density is proportional to a ratio of multivariate-*t* densities and is defined as the class of (1-1) *poly-t* densities (Dickey, 1968; Drèze, 1977; Zellner, 1971, p. 269).

In the present paper we reanalyze and extend some of the results obtained by Drèze. The organization of this paper is as follows. We start in Section 2 with an analysis of the exact form of the likelihood function of an underidentified SEM, in the structural parameter space. Using a two-step integration procedure we show that a uniform prior density on the structural parameters gives explosive behaviour of the marginal posterior densities of several parameters of an underidentified model. So, *noninformative priors may give sharp, albeit, pathological behaviour of posteriors.* For expository purposes, the integration steps are spelled out in detail since they are repeatedly used in the sequel of the paper.

In Section 3, we give a proof that, for a standard class of noninformative prior densities, the posterior density of the parameters of a single structural equation derived in a limited information framework, is a ratio-form poly-*t* density *if, and only if,* the prior degrees of freedom parameter has the value suggested by Drèze (1976). The noninformative priors that are discussed by Zellner (1971, p. 225) and Malinvaud (1978, p. 122) give different classes of posterior densities.

In Section 4 we discuss the approach where a single structural equation is completed with the unrestricted reduced form equations of the endogenous variables. This so-called *incomplete simultaneous equations model* has been used by, e.g., Hendry and Richard (1983), Richard (1984), Zellner (1971, Section 9.4), Drèze and Richard (1983, Sections 2 and 5) and Zellner, Bauwens and Van Dijk (1988). The advantage of this alternative approach is that it fits naturally to the field of modelling a single equation when one intends to investigate whether some explanatory variables are exogenous. We discuss several representations of the incomplete simultaneous equations model and show that the representation of Drèze and Richard and the representation of Zellner, Bauwens and Van Dijk are in a certain sense *dual* to each other. The duality follows from two different decompositions of the likelihood function. One may argue that the Drèze and Richard representation yields the Bayesian counterpart of the limited information maximum likelihood estimator or least variance ratio estimator and that the Zellner, Bauwens and Van Dijk representation yields the Bayesian counterpart of the instrumental variable estimator (in particular two-stage least squares and *k*-class).

In Section 5 we state the conditions under which the prior specification is invariant under the different representations of the model that are introduced in Section 4. Next, we derive the posterior densities of the equation system parameters for the different model representations. It is shown that Drèze and Richard (1983) use a conditional *matrix-t* density for the reduced form coefficients given a value of the structural coefficients and a marginal (1-1) poly-*t* density for the structural coefficients while Zellner, Bauwens and Van Dijk (1988) make use of a conditional *multivariate-t* density in the structural coefficients given a value of the reduced form coefficients and a so-called marginal (2-1) *poly-matrix-t* density in the reduced form coefficients. As a next step we discuss, briefly, in Section 6 how the different model representations of Section 4 and the distribution theoretic results of

[2] Drèze (1975, 1976), Drèze and Morales (1976) describe results on identification, limited information and full information estimation. An extensive survey is given by Drèze and Richard (1983).

[3] The reason for the focus on Bayesian limited information estimation was the difficulty of deriving computationally tractable and flexible results in the full information case. It is to some extent the same reason why, earlier, at the Cowles Commission, several researchers (e.g., Anderson, 1949) developed the limited information maximum likelihood estimator (LIML) as an alternative to the full information maximum likelihood estimator (FIML). We note that several computational procedures that are useful for Bayesian limited information analysis are given in Richard and Tompa (1980), Bauwens and Richard (1985) and Bauwens et al. (1981).

Section 5 can be used for Bayesian inference on the validity of overidentifying restrictions and exogeneity assumptions. Some conclusions and suggestions for further work are given in Section 7. The appendices contain technical details and proofs.

2. The likelihood function of an underidentified simultaneous equations model

The complete SEM can be written in the structural form as

$$YB + Z\Gamma = U, \tag{2.1}$$

where Y is a $T \times m$ matrix of observations on m endogenous variables, Z is a $T \times k$ matrix of observations on k predetermined variables, and the data matrix $(Y\ Z)$ has full column rank. The matrix B is an $m \times m$ non-singular matrix of unknown coefficients and Γ is a $k \times m$ matrix of unknown coefficients. U is a $T \times m$ matrix of unobserved disturbances. The T rows of U are assumed to be independent, each of them being normal with expectation zero and identical positive definite symmetric (PDS) covariance matrix Σ. The predetermined variables in Z that are not lagged values of variables in Y are assumed to be weakly exogenous (see Engle et al., 1983). At this stage, no restrictions are imposed on the matrices of parameters B, Γ and Σ (except for the PDS restriction). So, the model is underidentified.

The reduced form of (2.1) is

$$Y = Z\Pi + V, \tag{2.2}$$

with Π and V given as

$$\Pi = -\Gamma B^{-1}, \qquad V = UB^{-1}. \tag{2.3}$$

It follows immediately from (2.1)–(2.3) that the SEM is an example of a nonlinear regression model in the sense that $E(Y|Z)$ is nonlinear in the parameters B and Γ, and also that given a particular value of B, the expected value of Y is linear in the parameter matrix Γ. So, given a value of B, one can make use of results from the statistical analysis of the multivariate linear regression model (see, e.g., Anderson, 1984, Chapter 8; Zellner, 1971, Chapter 8).

We study the functional form of the likelihood function of an underidentified SEM in the structural parameter space. Consider the likelihood function of (B, Γ, Σ) given the data matrix $D = (Y\ Z)$:

$$L(B, \Gamma, \Sigma \mid D) = (2\pi)^{mT/2} \|B\|^{T} |\Sigma|^{-T/2} \exp\{-\tfrac{1}{2}\operatorname{tr}[Q(B, \Gamma)\Sigma^{-1}]\}, \tag{2.4}$$

$L(B, \Gamma, \Sigma \mid D) = L_c(\Sigma \mid B, \Gamma, D) L_m(B, \Gamma \mid D)$

↓

inverted Wishart step on Σ: equation (2.6)

↓

$L_m(B, \Gamma \mid D) = L_c(\Gamma \mid B, D) L_m(B \mid D)$

↓

complete squares on Γ: equations (2.7)–(2.9)

↓

matrix-t step on Γ: equations (2.10)–(2.12)

↓

$L_m(B \mid D)$: equation (2.13)

Figure 1. Marginalization of the likelihood function of (B, Γ, Σ). (One may interchange the order of integration and first integrate with respect to Γ by making use of the matrix normal distribution. As a next step one integrates with respect to Σ.)

where

$$Q(B, \Gamma) = (YB + Z\Gamma)'(YB + Z\Gamma). \tag{2.5}$$

For a derivation of (2.4) we refer to standard textbooks in econometrics, e.g., Zellner (1971, Chapter 9). One can marginalize (2.4) with respect to Σ and Γ as indicated in Figure 1.

The inverted Wishart step on Σ (loss of $m+1$ degrees of freedom)

By making use of the definition of the inverted Wishart density function (Anderson, 1984, Chapter 7; Zellner, 1971, Appendix B.4) one can integrate (2.4) with respect to Σ as follows:

$$\begin{aligned} L_m(B, \Gamma \mid D) &\propto \|B\|^{T} \int |\Sigma|^{-T/2} \exp\{-\tfrac{1}{2}\operatorname{tr}[Q(B, \Gamma)\Sigma^{-1}]\}\, d\Sigma \\ &\propto \|B\|^{T} |Q(B, \Gamma)|^{-(T-m-1)/2} \end{aligned} \tag{2.6}$$

under the conditions $|Q(B, \Gamma)| > 0$ and $T > 2m$. The proportionality signs indicate that the normalization factors of the likelihood and the inverted Wishart density, that do not depend on the parameters (B, Γ), have been deleted for notational convenience. The exponent $-(T-m-1)/2$ indicates the loss of $m+1$ degrees of freedom due to the marginalization with respect to Σ. If one conditions (2.4) on the maximum likelihood estimator $(YB + Z\Gamma)'(YB + Z\Gamma)/T$ for Σ, one obtains the same functional form as (2.5), except that the exponent is $-T/2$. This is the well-known concentrated likelihood function. We emphasize this difference between conditionalization and marginalization since it plays a major role in the sequel.

Complete squares on Γ

As a next step, we analyze the functional form of (2.6) by factorizing it as the product of a conditional function of Γ, given a value of B, and a function of B. Consider the matrix of sums of squares and cross-products $(YB+Z\Gamma)'(YB+Z\Gamma)$ and complete the squares in Γ, given a value of B. Then one obtains

$$(YB+Z\Gamma)'(YB+Z\Gamma)=B'\hat{\Omega}B+(\Gamma+\hat{\Pi}B)'Z'Z(\Gamma+\hat{\Pi}B), \tag{2.7}$$

where $\hat{\Omega}$ is a function of the data only, given as

$$\hat{\Omega}=Y'M_zY, \qquad M_z=I_T-Z(Z'Z)^{-1}Z', \tag{2.8}$$

and $\hat{\Pi}$ is a function of the data, given as

$$\hat{\Pi}=(Z'Z)^{-1}Z'Y. \tag{2.9}$$

The matrix-t step on Γ (loss of k degrees of freedom)

We make use of the definition of the matrix-t density. (See, e.g., Drèze and Richard, 1983, p. 589; Zellner, 1971, Appendix B5; Dickey, 1967.) Given this definition and using equations (2.7)-(2.9), one can write (2.6) as

$$L_m(B,\Gamma\,|\,D)\propto|B'\hat{\Omega}B+(\Gamma+\hat{\Pi}B)'Z'Z(\Gamma+\hat{\Pi}B)|^{-(T-m-1)/2}$$
$$\times|B'\hat{\Omega}B|^{(T-m-k-1)/2}|B'B|^{(m+k+1)/2}. \tag{2.10}$$

Note that we have used the equality $\|B\|^2=|B'\hat{\Omega}B|\,|\hat{\Omega}|^{-1}$ and dropped the factor in $|\hat{\Omega}|$ in the second line. The first two factors on the right-hand side of (2.10) form a kernel of a matrix-t density in Γ. So, one can write

$$L_m(B,\Gamma\,|\,D)=L_c(\Gamma\,|\,B,D)\,L_m(B\,|\,D), \tag{2.11}$$

where

$$L_c(\Gamma\,|\,B,D)\propto f_{\mathrm{MT}}^{k\times m}(\Gamma\,|\,-\hat{\Pi}B,\,B'\hat{\Omega}B,\,Z'Z,\,T-m-k-1), \tag{2.12}$$

which is a conditional matrix-t density[4] under the usual conditions *and the condition* $0<\|B\|<\infty$. So, one can write

$$L_m(B\,|\,D)=\int L_m(B,\Gamma\,|\,D)\,\mathrm{d}\Gamma\propto|B'B|^{(m+k+1)/2}. \tag{2.13}$$

Clearly, if the elements of B are not restricted in an adequate way, the function (2.13), which can be interpreted as the marginal posterior density of B based on a uniform prior on all the parameters, is not integrable in $\mathbb{R}^{m\times m}$. This reflects simply the lack of identification of the structural parameters assumed at the beginning of this section. As a consequence, the marginal posterior density of Γ is in principle also not integrable. In practice, one could define (2.13) on a region of integration that is finite, e.g., the set $\{B\,|\,\varepsilon\leq|B'B|\leq M\}$, where ε is a small positive constant and M is a large finite constant. The results of the integrations of the functions (2.13) (or (2.12) times (2.13)), with respect to B may of course be sensitive with respect to the choice of ε and M, depending on the data.

In the maximum likelihood framework one substitutes $-\hat{\Pi}B$ for Γ in the likelihood function which is already concentrated with respect to Σ. Then one obtains

$$L_c(B\,|\,D)\propto\text{constant} \tag{2.13$'$}$$

as is well known since the structural form is not identified. The result that the likelihood function concentrated with respect to both Σ and Γ, i.e., (2.13)$'$, is flat, but that the marginal likelihood of B, given as (2.13), is not flat was implicit in Drèze (1976); see also Maddala (1976). A summary of the results on marginalization and conditionalization of the likelihood of an underidentified SEM is given in Table 1.

Table 1
Likelihood functions of equation system parameters B and Γ

	Marginal likelihood L_m	Conditional likelihood L_c
$L(B,\Gamma\,|\,D)\propto$	$\|B\|^T|Q(B,\Gamma)|^{-(T-m-1)/2}$	$\|B\|^T|Q(B,\Gamma)|^{-T/2}$
$L(B\,|\,D)\propto$	$\|B\|^{(m+k+1)/2}$	constant

[4] In this paper, we use the notation of Drèze and Richard (1983, Appendix A) for density functions: the subscripts of f are a mnemonic for the name of the density (e.g., MT for matrix-t), the superscripts indicate the dimension of the random variable the name of which is given as argument; then follow the parameters.

To get (2.10)-(2.13), we have implicitly used a truncated uniform prior density on the parameters (B, Γ, Σ). In order to show the effect of marginalization of Γ and Σ, one can introduce an extra positive parameter h through the following noninformative prior

$$p(B, \Gamma, \Sigma) \propto |\Sigma|^{-h/2}. \tag{2.14}$$

Then, after multiplication of (2.4) by (2.14), and after repetition of the steps from (2.4) to (2.12), one obtains the result that the posterior density of B, Γ and Σ can be decomposed as follows:

$$p(B, \Gamma, \Sigma \mid D) = p_c(\Sigma \mid B, \Gamma, D) p_c(\Gamma \mid B, D) p_m(B \mid D), \tag{2.15}$$

where

$$p_c(\Sigma \mid B, \Gamma, D) \propto f^m_{\mathrm{IW}}(\Sigma \mid Q(B, \Gamma), T+h-m-1), \tag{2.15a}$$

$$p_c(\Gamma \mid B, D) \propto f^{k \times m}_{\mathrm{MT}}(\Gamma \mid -\hat{\Pi}B, B'\hat{\Omega}B, Z'Z, T+h-m-k-1), \tag{2.15b}$$

$$p_m(B \mid D) \propto |B'B|^{-(h-m-k-1)/2}. \tag{2.15c}$$

A kernel of the conditional posterior density $p_c(\Gamma \mid B, D)$, equation (2.15b), is shown in Figure 2 for the case where $T = 15$, $h = m+k+1$ with $m = 1$ and $k = 1$. Note that for this value of h the conditional density $p_c(\Gamma \mid B, D)$

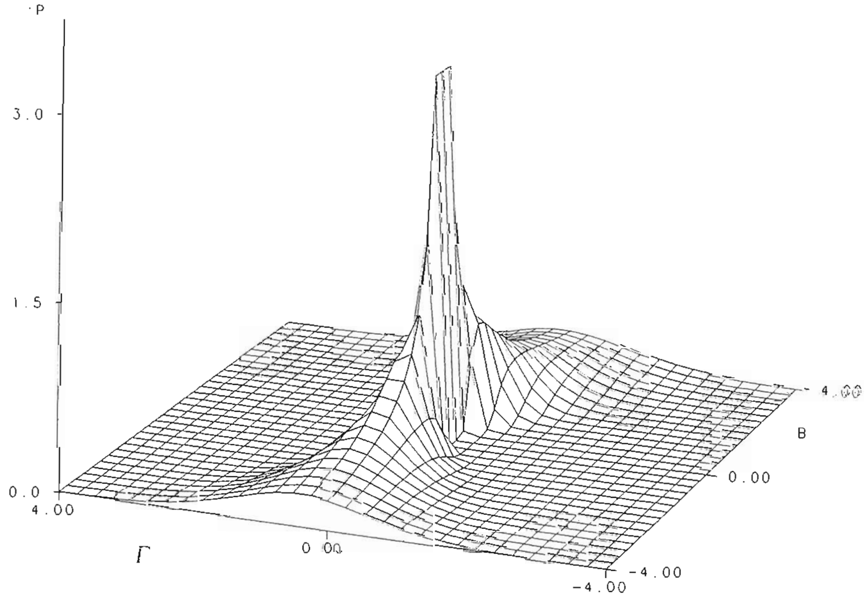

Figure 2. Kernel of $p(B, \Gamma \mid D)$ or $p_c(\Gamma \mid B, D)$ for $h = m+k+1$ and $T = 15$.

and the joint density $p(B, \Gamma \mid D)$ are the same, since the marginal density $p_m(B \mid D)$ is uniform. The choice of the parameter values and the data is discussed in Appendix A. Clearly, the conditional densities become more concentrated when $B \to 0$ and have thick tails when B becomes large in absolute value. The conditional density is not defined for $B = 0$. Kernels of the marginal density $p_m(B \mid D)$ and of the joint density $p(B, \Gamma \mid D)$ are shown in Figure 3 and Figure 4, respectively, for different values of the degrees of freedom parameter h. One may distinguish three cases. If h is less than $m+k+1$, one has a "polynomial" weighting function. See Figure 3, at the point $B = 0$, the function p_m is equal to 0, while it tends to ∞ in the areas where B tends to $\pm\infty$. If h is greater than $m+k+1$, one has an "exponential" weighting function. See Figure 3, when B tends to 0, the function p_m tends to ∞, while it tends to 0 when B becomes large in absolute value. Finally, if $h = m+k+1$, one has a flat weighting function. This is the value used by Drèze (1976). In practice, it appears a sensible strategy to experiment with a sequence of prior densities.

In order to illustrate further the difference between marginalization and conditionalization, one can work with the implied reduced form parameters

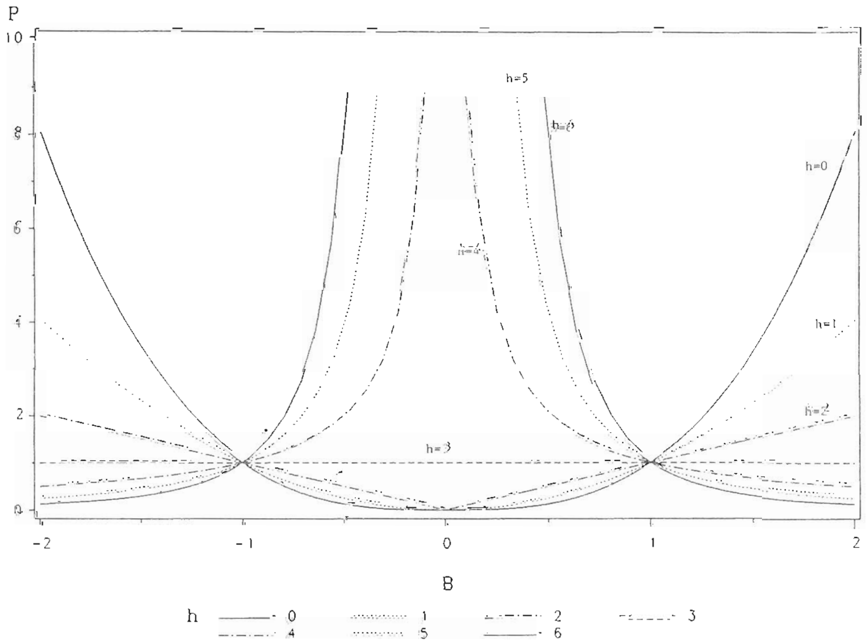

Figure 3. Kernels of $p_m(B \mid D)$ for different values of h.

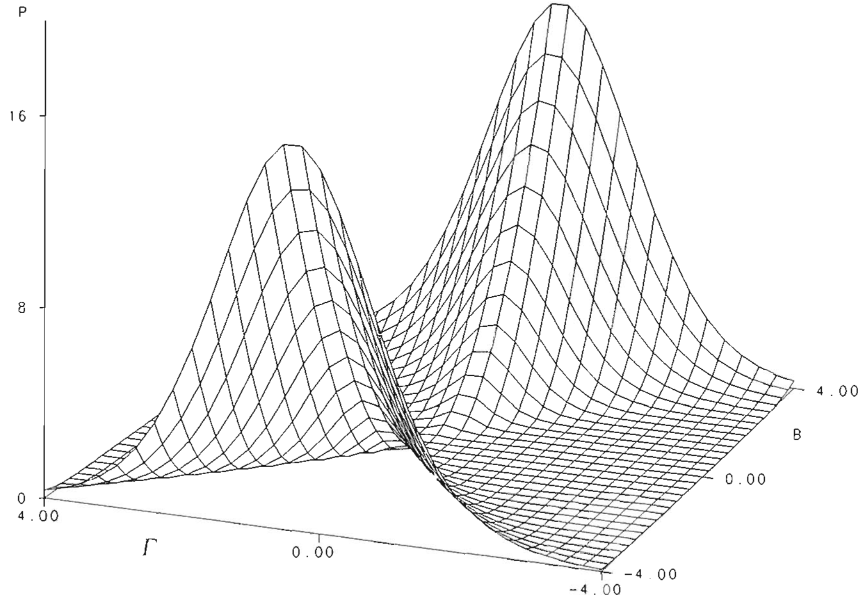

Figure 4a. Kernels of $p(B, \Gamma | D)$ for $h = 0$ and $T = 15$.

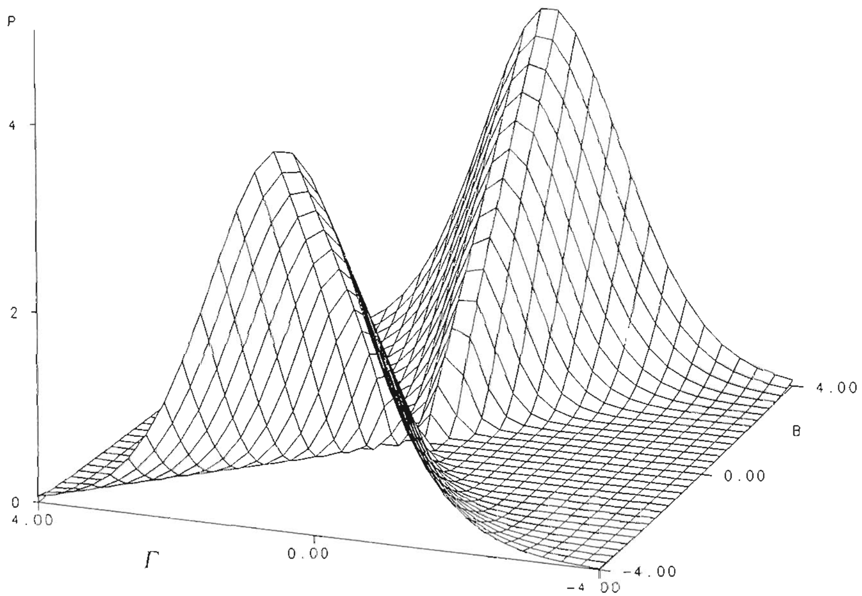

Figure 4b. Kernels of $p(B, \Gamma | D)$ for $h = 1$ and $T = 15$.

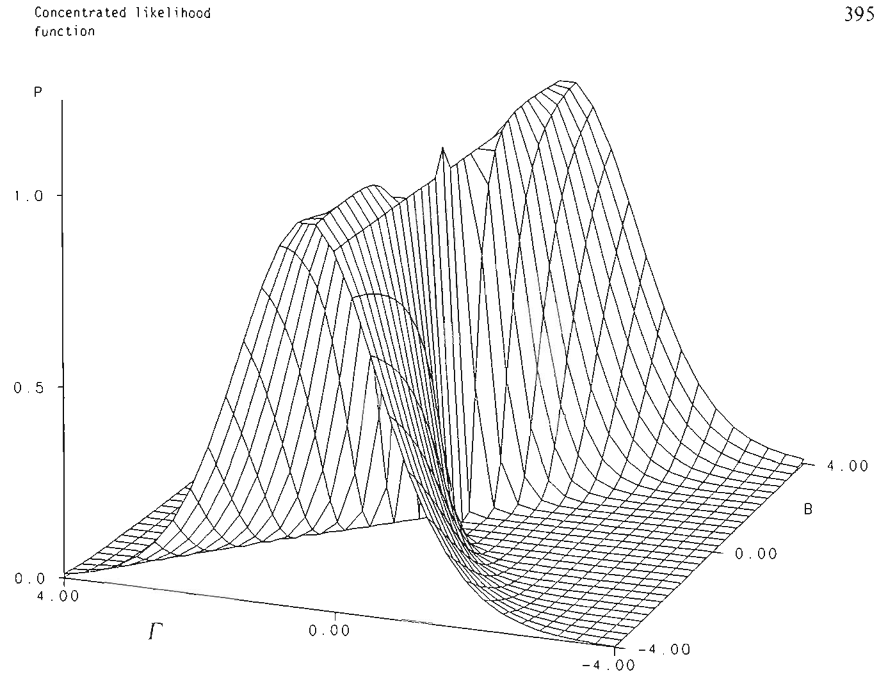

Figure 4c. Kernels of $p(B, \Gamma | D)$ for $h = 2$ and $T = 15$.

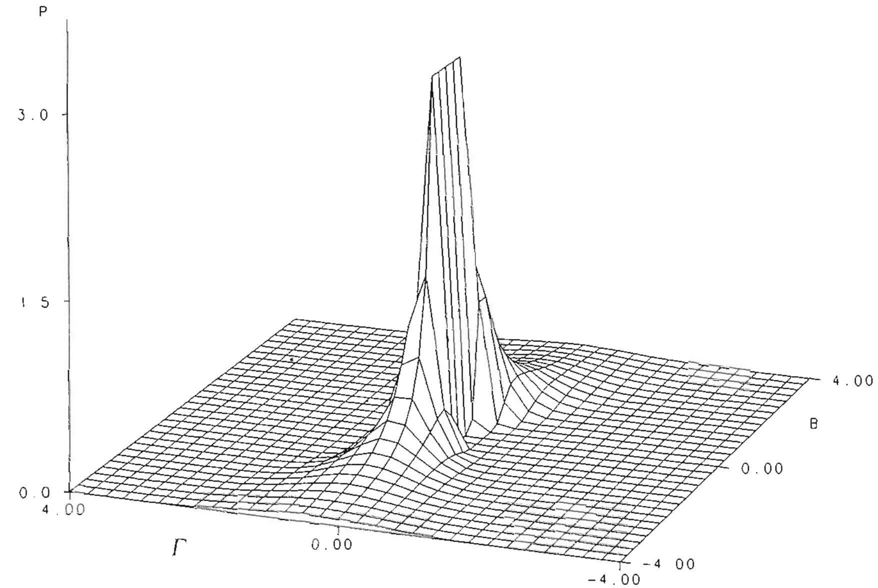

Figure 4d. Kernels of $p(B, \Gamma | D)$ for $h = 4$ and $T = 15$.

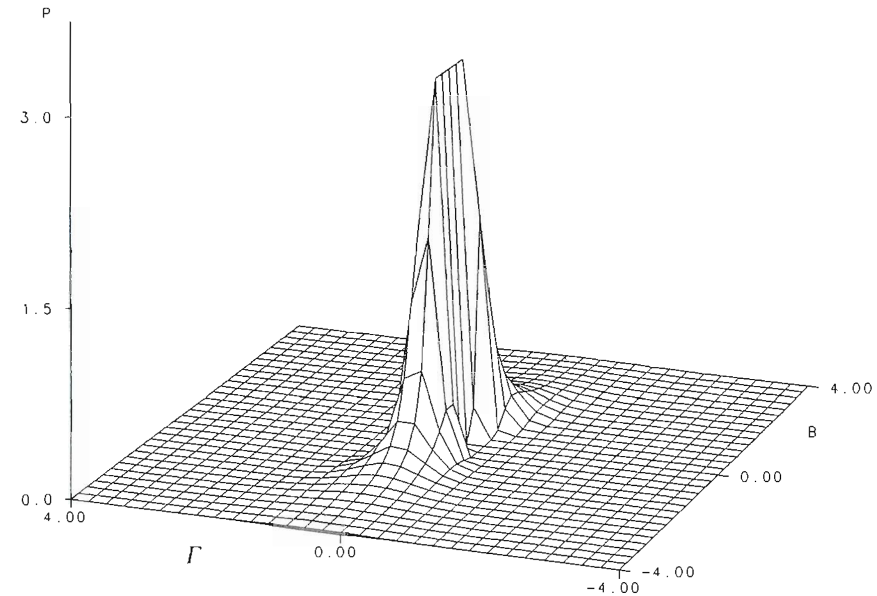

Figure 4e. Kernels of $p(B, \Gamma \mid D)$ for $h=5$ and $T=15$.

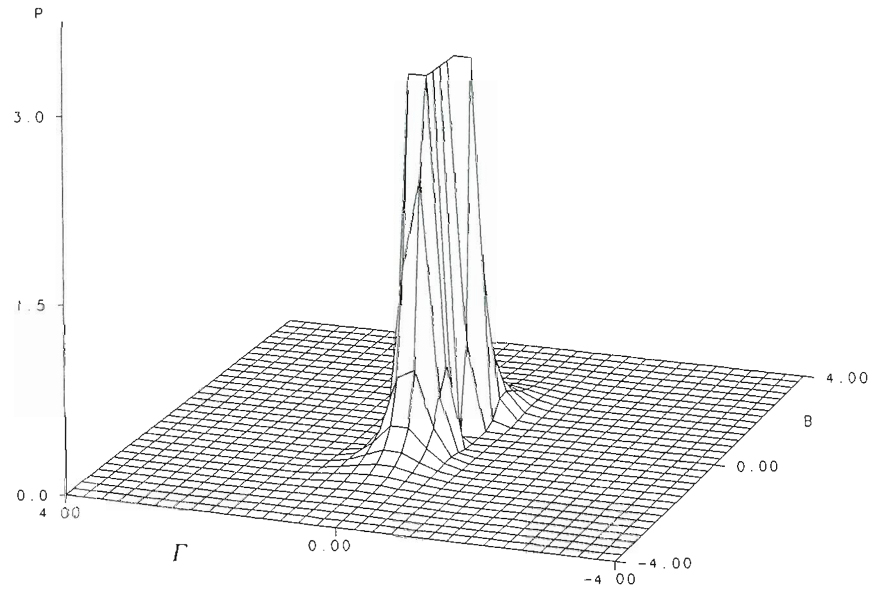

Figure 4f. Kernels of $p(B, \Gamma \mid D)$ for $h=6$ and $T=15$.

Π and Ω, where Ω is the PDS covariance matrix of each row of V, and is therefore related to the structural parameters by the relation

$$\Omega = B'^{-1}\Sigma B^{-1}. \tag{2.16}$$

As the transformation from (B, Γ, Σ) to (Π, Ω) is not one-to-one, we proceed as follows. First, perform the transformation of random variables that are elements of (B, Γ, Σ) to variables that are elements of (B, Π, Ω). The Jacobian determinant of this transformation is $\|B\|^{m+k+1}$ (see, e.g., Magnus and Neudecker, 1988, pp. 30–31 and Chapter 9). Second, marginalize the posterior density of (B, Π, Ω) with respect to B. In order to study the effect of this marginalization we transform the prior (2.14) to

$$p(B, \Pi, \Omega) \propto \|B\|^{m+k+1-h} \cdot |\Omega|^{-h/2}. \tag{2.17}$$

The likelihood function (2.4) can be rewritten as

$$L(B, \Pi, \Omega \mid D) \propto |\Omega|^{-T/2} \cdot \exp\{-\tfrac{1}{2}\mathrm{tr}[(Y - Z\Pi)'(Y - Z\Pi)\Omega^{-1}]\}. \tag{2.18}$$

The posterior density of (B, Π, Ω), which is proportional to the product of (2.17) and (2.18), can be marginalized with respect to Ω by using again the definition of the inverted Wishart density. Then one obtains

$$p(B, \Pi \mid D) \propto p_{\mathrm{u}}(\Pi \mid D) p_{\mathrm{m}}(B \mid D), \tag{2.19}$$

where

$$p_{\mathrm{u}}(\Pi \mid D) = f_{\mathrm{MT}}^{k\times m}(\Pi \mid \hat{\Pi}, \hat{\Omega}, Z'Z, T+h-m-k-1) \tag{2.20}$$

is a matrix-t density with parameters that depend on the data and h but *that do not depend on* B, and

$$p_{\mathrm{m}}(B \mid D) \propto |B'B|^{(m+k+1-h)/2}, \tag{2.21}$$

which is equivalent to (2.13) if $h=0$. Thus, we conclude that inference on Π is not sensitive to the truncation of $\|B\|$ to a finite range. However, as noted before, structural inference may be sensitive to the truncation of the range of $\|B\|$. In other words, the sample is informative on Π but not on B.

In the maximum likelihood framework, there is no Jacobian involved and there is no loss of degrees of freedom. That is, after concentrating (2.18) with respect to Ω, one obtains the concentrated likelihood function of Π, which is proportional to the same function as (2.20), except that $T+h-m-k-1$ is replaced by T, and one obtains the "concentrated" likelihood of B, which is proportional to a constant; compare (2.13′).

Summarizing, we have reviewed the implications of the lack of identification of the structural parameters of a SEM through marginalization and conditionalization of its likelihood function. Given the results (2.12), (2.13) and (2.15), there is a need for prior information, since a noninformative prior on the parameters of a nonidentified SEM may typically give an explosive behaviour of the posterior density in some directions of the parameter space while it is more regular in other directions. For work on informative prior densities in a full information framework we refer to, e.g., Zellner (1971), Drèze and Morales (1976), Bauwens (1984), Van Dijk (1984), Van Dijk (1987, and the references cited there, in particular, Kloek and Van Dijk, 1978; Van Dijk and Kloek, 1980). A survey is given by Drèze and Richard (1983).

3. Structural single equation analysis

Suppose one considers only prior information on the parameters of the first equation of (2.1), chosen without loss of generality, and no prior information on equations 2 to m. So, we have restricted matrices B, Γ, U, Σ, Π, V and Ω. In particular, we consider the case where (2.1) to (2.3) hold with B and Γ replaced by

$$B_r = \begin{pmatrix} 1 & \\ -\beta_1 & B_2 \\ 0 & \end{pmatrix}, \qquad \Gamma_r = \begin{pmatrix} -\gamma_1 & \\ & \Gamma_2 \\ 0 & \end{pmatrix}. \tag{3.1}$$

The submatrices B_2 and Γ_2 correspond to equations 2 to m and are unrestricted matrices of dimension $m \times (m-1)$ and $k \times (m-1)$, respectively. The first column of B has 1 as a first element by a normalization constraint, and the remaining $m-1$ elements have been partitioned into an $m_1 \times 1$ vector $-\beta_1$ of unrestricted parameters and an $m_0 \times 1$ vector of parameters, say β_0, restricted to 0. The first column of Γ_r has been partitioned into a $k_1 \times 1$ vector of unrestricted parameters $-\gamma_1$ and a $k_0 \times 1$ vector of parameters, say γ_0, restricted to 0. The zero restrictions in the first equation correspond to the exclusion of variables from it, usually for the purpose of identification. Let Y and Z be partitioned as

$$Y = (y_1 \;\; Y_1 \;\; Y_0), \qquad Z = (Z_1 \;\; Z_0), \tag{3.2}$$

where y_1 is $T \times 1$, Y_1 is $T \times m_1$, Y_0 is $T \times m_0$, Z_1 is $T \times k_1$ and Z_0 is $T \times k_0$, while $1 + m_1 + m_0 = m$ and $k_1 + k_0 = k$. In summary, the system in structural form can be written as

$$y_1 = Y_1\beta_1 + Z_1\gamma_1 + u_1, \tag{3.3a}$$

$$YB_2 + Z\Gamma_2 = U_2. \tag{3.3b}$$

Note that u_1 is different from the first column of U in (2.1). The covariance matrix of the rows of the matrix $(u_1 \;\; U_2)$ is denoted by Σ_r.

The likelihood function of the model has the same form as (2.4), but one makes use of the definitions of B_r and Γ_r given by (3.1). Suppose we keep the prior on Σ_r as given in (2.14), then one can write the posterior density $p(\beta_1, \gamma_1, B_2, \Gamma_2, \Sigma_r | D)$ in a straightforward way as

$$p(\beta_1, \gamma_1, B_2, \Gamma_2, \Sigma_r | D) \propto \| B_r \|^T |\Sigma_r|^{-(T+h)/2} \times \exp\{-\tfrac{1}{2}\operatorname{tr}[(u_1 \;\; U_2)'(u_1 \;\; U_2)\Sigma_r^{-1}]\}, \tag{3.4}$$

where $(u_1 \;\; U_2)$ is restricted by (3.3a)–(3.3b).

In order to derive the marginal posterior density of the parameters of interest (β_1, γ_1), one has to integrate the join posterior with respect to Σ_r, Γ_2, and B_2. The integration with respect to Σ_r is done in the same way as in Section 2, equation (2.6), with the extra parameter h. The next step is to complete the squares on Γ_r as was done in Section 2, equations (2.7)–(2.9) for the case of Γ. Then one follows the steps given in (2.10)–(2.12), or in (2.15)–(2.15c). Finally, one obtains

$$p(\beta_1, \gamma_1, B_2, \Gamma_2 | D) \propto f_{\text{RMT}}^{k \times m}(\Gamma_r | -\hat{\Pi}B_r, B_r'\hat{\Omega}B_r, Z'Z, T+h-m-k-1) \times |B_r'B_r|^{(m+k+1-h)/2}. \tag{3.5}$$

The mnemonic[5] RMT stands for a restricted (nonnormalized) matrix-t density in the sense that Γ_r is restricted as given in (3.1). We impose the condition $M \geq |B_r'\hat{\Omega}B_r| \geq \varepsilon$, with $\varepsilon > 0$ and $M < \infty$.

A next step is to integrate (3.5) with respect to Γ_2 by making use of properties of the matrix-t density. This involves some tedious derivations. A relatively easy procedure is the following. Assume that the restrictions on B_r and Γ_r are not yet imposed. We denote the first column of unrestricted elements of B by $\tilde{\beta}_1$ and the first column of unrestricted elements of Γ by $\tilde{\gamma}_1$. So, one can replace (3.1) by

$$B = (\tilde{\beta}_1, B_2), \qquad \Gamma = (\tilde{\gamma}_1, \Gamma_2). \tag{3.1$'$}$$

If the restrictions of (3.1) are not imposed on (3.5), this density is a conditional matrix-t in Γ given a value of B (see (2.15b)). By making use of properties of the matrix-t density (see Zellner, 1971, Appendix B5; Drèze and Richard, 1983, p. 589), one can decompose the right-hand side of (3.5) in two factors. The first factor is a conditional matrix-t density of Γ_2 given a value of $\tilde{\gamma}_1$ (the unrestricted first column of Γ) and a value of B. The second factor is a marginal $k \times 1$ matrix-t density of $\tilde{\gamma}_1$, given a value of

[5] See footnote 4.

the first column of B, denoted by $\tilde{\beta}_1$. This latter density is defined as

$$p(\tilde{\gamma}_1|\tilde{\beta}_1, D) = f_{\mathrm{MT}}^{k\times 1}(\tilde{\gamma}_1|-\hat{\Pi}\tilde{\beta}_1, \tilde{\beta}_1'\hat{\Omega}\tilde{\beta}_1, Z'Z, T+h-2m-k) \tag{3.6}$$

under the condition $M \geq \tilde{\beta}_1'\hat{\Omega}\tilde{\beta}_1 \geq \varepsilon$, with $\varepsilon > 0$ and $M < \infty$. We emphasize that the density (3.6) is *conditional on* $\tilde{\beta}_1$ *only, and not on* B_2, *i.e. it is conditionally independent* of B_2. Note that (3.6) can be interpreted as a multivariate-t density.

Integration of (3.5) with respect to Γ_2 gives a posterior density of $(\tilde{\beta}_1, \tilde{\gamma}_1, B_2)$ that is proportional to the right-hand side of (3.6) multiplied by the factor $|B'B|^{(m+k+1-h)/2}$ (compare (2.15b)–(2.15c) for the full information case). The factor $|B'B|^{(m+k+1-h)/2}$ can be deleted from (3.5) *if and only if* $h = m+k+1$. Suppose that the prior on B_2 is uniform on a region of integration where $|B'B|$ is restricted to be finite. Using the definition of a matrix-t density, one can derive the marginal posterior density of $(\tilde{\beta}_1, \tilde{\gamma}_1)$ as

$$p(\tilde{\beta}_1, \tilde{\gamma}_1|D) \propto (\tilde{\beta}_1'\hat{\Omega}\tilde{\beta}_1)^{(T-m+1)/2}$$
$$\times [\tilde{\beta}_1'\hat{\Omega}\tilde{\beta}_1 + (\tilde{\gamma}_1 + \hat{\Pi}\tilde{\beta}_1)'Z'Z(\tilde{\gamma}_1 + \hat{\Pi}\tilde{\beta}_1)]^{-(T-m+k+1)/2}. \tag{3.7}$$

We emphasize that the derivation of (3.7) does not depend on whether the restrictions on the first columns of B and Γ are imposed. That is, one may start with (3.4), with the restrictions imposed, and one can obtain the same functional form as given in (3.7) with $(\tilde{\beta}_1, \tilde{\gamma}_1)$ replaced by (β_1, γ_1) and with a proper adjustment of the data dependent matrices $\hat{\Omega}$, $\hat{\Pi}$, $Z'Z$, see below. However, the properties of the functional form of (3.7), in particular whether the function is integrable on a large region, depend on the restrictions. We distinguish two major cases.

First, suppose no restrictions are imposed. That is, equation (3.7) is maintained as it stands. The functional form of (3.7) has in this case the same properties as discussed in Section 2 for the full information case; compare equations (2.15)–(2.15c) with $h = m+k+1$. The conditional multivariate-t density of $\tilde{\gamma}_1$ given a value of $\tilde{\beta}_1$ degenerates at $\tilde{\beta}_1 = 0$. Kernels of the joint posterior of $(\tilde{\beta}_1, \tilde{\gamma}_1)$ are given in Figure 5a for $T=2$ and $T=15$. Note that Figure 5a(ii) is the same as Figure 2. Further details on the choice of the parameter values and data are given in Appendix A.

As an intermediate case, consider the case where $\tilde{\beta}_1$ is replaced by the first column of B_r; see (3.1). Then one makes use of (2.8) and $\tilde{\beta}_1'\hat{\Omega}\tilde{\beta}_1$ in (3.7) is replaced by

$$(1 \;\; -\beta_1' \;\; 0')\hat{\Omega}\begin{pmatrix}1\\ -\beta_1\\ 0\end{pmatrix} = (y_1 - Y_1\beta_1)'M_z(y_1 - Y_1\beta_1), \tag{3.8}$$

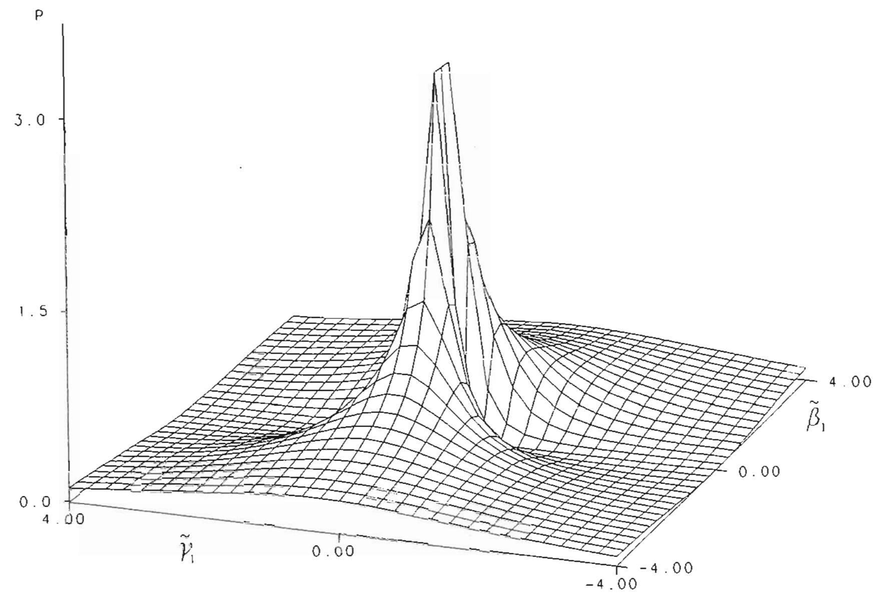

Figure 5a(i). Kernels of $p(\tilde{\beta}_1, \tilde{\gamma}_1|D)$; no restrictions on parameters, $T=2$.

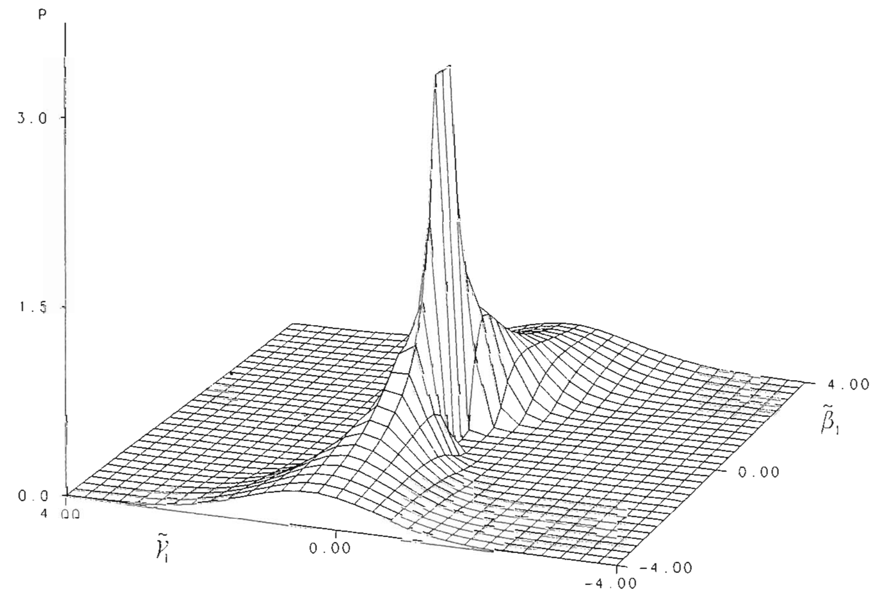

Figure 5a(ii). Kernels of $p(\tilde{\beta}_1, \tilde{\gamma}_1|D)$; no restrictions on parameters, $T=15$.

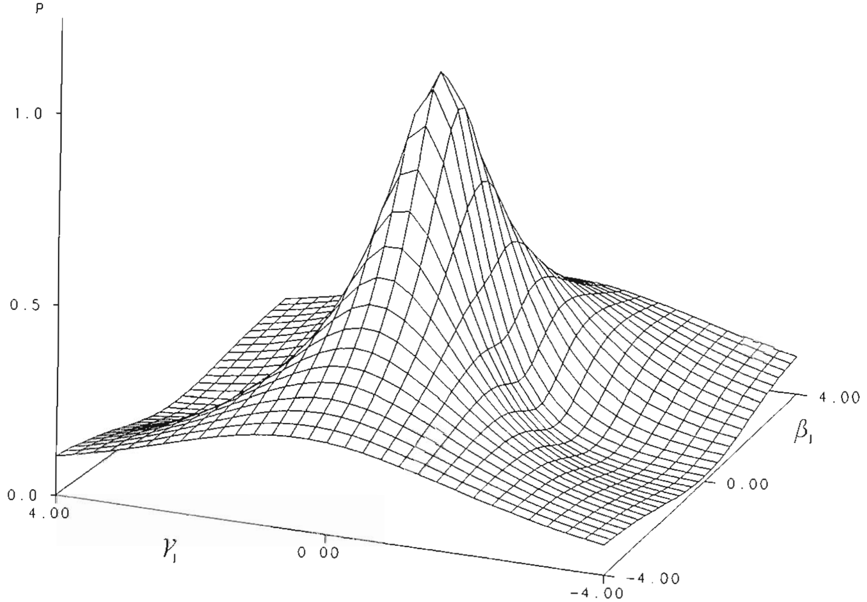

Figure 5b(i). Kernels of the (1-1) poly-t density $p(\beta_1, \gamma_1 | D)$, $T=2$.

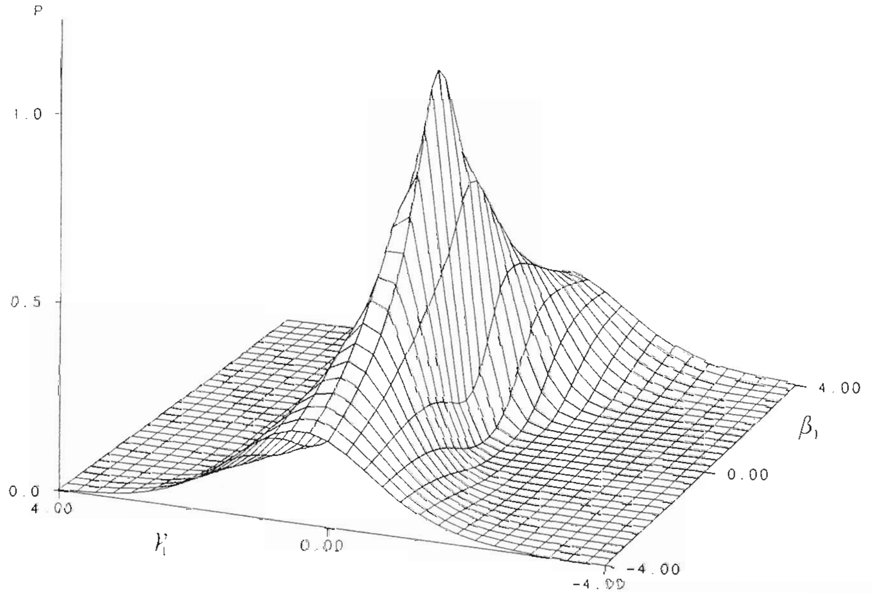

Figure 5b(ii). Kernels of the (1-1) poly-t density $p(\beta_1, \gamma_1 | D)$, $T=15$.

where

$$M_Z = I - Z(Z'Z)^{-1}Z'. \tag{3.9}$$

Given the condition of full column rank of $(Y_1 \;\; Z)$ and given the normalization restriction that the first element of the vector $(1 \;\; -\beta_1' \;\; 0')$ is equal to one, the right hand side of (3.8) is positive everywhere in the region of integration. The functional form of $p(\beta_1, \tilde\gamma_1 | D)$ is such that no degeneracy occurs as β_1 tends to zero.

Second, consider the case where the first columns of both B_r and Γ_r, given in (3.1), are substituted in (3.7) for the unrestricted vectors $\tilde\beta_1$ and $\tilde\gamma_1$. We make use of

$$\tilde\beta_1'\hat\Omega\tilde\beta_1 + (\tilde\gamma_1 + \hat\Pi\tilde\beta_1)'Z'Z(\tilde\gamma_1 + \hat\Pi\tilde\beta_1) = (Y\tilde\beta_1 + Z\tilde\gamma_1)'(Y\tilde\beta_1 + Z\tilde\gamma_1). \tag{3.10}$$

Imposing the restrictions implies that (3.10) is replaced by $(y_1 - Y_1\beta_1 - Z_1\gamma_1)'(y_1 - Y_1\beta_1 - Z_1\gamma_1)$. If one completes the squares on β_1 and γ_1 one obtains

$$(y_1 - Y_1\beta_1 - Z_1\gamma_1)'(y_1 - Y_1\beta_1 - Z_1\gamma_1)$$
$$= s_1^2 + \begin{pmatrix}\beta_1 - \hat\beta_1\\ \gamma_1 - \hat\gamma_1\end{pmatrix}' \begin{pmatrix} Y_1'Y_1 & Y_1'Z_1 \\ Z_1'Y_1 & Z_1'Z_1\end{pmatrix} \begin{pmatrix}\beta_1 - \hat\beta_1\\ \gamma_1 - \hat\gamma_1\end{pmatrix}, \tag{3.11}$$

where

$$\begin{pmatrix}\hat\beta_1\\ \hat\gamma_1\end{pmatrix} = \begin{pmatrix} Y_1'Y_1 & Y_1'Z_1 \\ Z_1'Y_1 & Z_1'Z_1\end{pmatrix}^{-1} \begin{pmatrix} Y_1'y_1 \\ Z_1'y_1\end{pmatrix},$$
$$s_1^2 = (y_1 - Y_1\hat\beta_1 - Z_1\hat\gamma_1)'(y_1 - Y_1\hat\beta_1 - Z_1\hat\gamma_1). \tag{3.12}$$

In a similar way one may complete the squares on β_1 in (3.8). Then one obtains

$$(y_1 - Y_1\beta_1)'M_Z(y_1 - Y_1\beta_1) = s_2^2 + (\beta_1 - \beta_1^*)'Y_1'M_ZY_1(\beta_1 - \beta_1^*), \tag{3.13}$$

where

$$\beta_1^* = (Y_1'M_ZY_1)^{-1}Y_1'M_Zy_1, \quad s_2^2 = (y_1 - Y_1\beta_1^*)'M_Z(y_1 - Y_1\beta_1^*). \tag{3.14}$$

Using the definition of a multivariate-t density one can write (3.7) as

$$p(\beta_1, \gamma_1 | D)$$
$$\propto \frac{\left[s_1^2 + \begin{pmatrix}\beta_1 - \hat\beta_1\\ \gamma_1 - \hat\gamma_1\end{pmatrix}' \begin{pmatrix} Y_1'Y_1 & Y_1'Z_1 \\ Z_1'Y_1 & Z_1'Z_1\end{pmatrix} \begin{pmatrix}\beta_1 - \hat\beta_1\\ \gamma_1 - \hat\gamma_1\end{pmatrix}\right]^{-(\nu_1+m_1+k_1)/2}}{\left[s_2^2 + (\beta_1 - \beta_1^*)'Y_1'M_ZY_1(\beta_1 - \beta_1^*)\right]^{-(\nu_2+m_1)/2}}, \tag{3.15}$$

where

$$\nu_1 = T - m - m_1 + k - k_1 + 1, \quad \nu_2 = T - m - m_1 + 1.$$

The right hand side of (3.15) is a ratio of multivariate-t kernels and is defined as a (1-1) poly-t density function.

We emphasize that this density is integrable if $k - k_1 > m_1$, which is the order condition for identification. This can be seen as follows. The tail behaviour of the marginal multivariate-t density of β_1 in the numerator depends on the degrees of freedom parameter which is equal to $T - m - m_1 + k - k_1 + 1$. In the denominator one has a quadratic form in β_1 raised to the power $-\frac{1}{2}(T - m + 1)$. The tail behaviour of the ratio of the two quadratic forms in β_1 depends on whether the difference in the exponents $T - m - m_1 + k - k_1 + 1$ and $T - m + 1$ is positive. This condition is equal to the order condition for identification $k - k_1 > m_1$. More details are given in Drèze (1976) or Dickey (1968). In Figure 5b we show a simple case of a (1-1) poly-t density with parameter values that are specified in Appendix A. We emphasize that the (1-1) poly-t density can have many different shapes than the one shown in Figure 5b, in particular, bimodality may occur.

We summarize this section in the following theorem.

Theorem 1. *Given the model* (3.1) *and the standard assumptions of a SEM, and given the class of noninformative priors defined in* (2.14), *the posterior density of the parameters of interest* (β_1, γ_1) *of a single structural equation is a* (1-1) *poly-t density, defined in* (3.15) *under the following conditions*:

(i) *The order condition for identification is satisfied*: $k - k_1 > m_1$;

(ii) *The prior degrees of freedom parameter h is given as* $h = m + k + 1$;

(iii) $|B_r' B_r|$ *is finite.* □

We end this section with some remarks.

First, we have not used the rank condition for identification but only the order condition. So uniform priors on a model that is underidentified may give sharp posteriors, see Maddala (1976). Note that we impose the condition that $|B_r' B_r|$ is finite. The sensitivity of the posterior results for this condition has to be investigated for each case. This remains an area for further research.

Second, in this section we have analyzed a single restricted structural equation while the remaining structural equations are unrestricted. In the next section we study the case where a single structural equation is completed with a set of unrestricted reduced form equations of the endogenous variables that appear in the structural equations. This is the so-called *incomplete simultaneous equations model* (see, e.g., Richard, 1984; Hendry and Richard, 1983). The connection between these two models can be studied as follows. Consider the posterior $p(\beta_1, \gamma_1, B_2, \Gamma_2 | D)$, equation (3.5). First, one applies the transformation of variables from $(\beta_1, \gamma_1, B_2, \Gamma_2)$ to $(\beta_1, \gamma_1, B_2, \Pi_2)$, where Π_2 is the matrix of reduced form parameters of the equations 2 to m. Next, one integrates the posterior of $(\beta_1, \gamma_1, B_2, \Pi_2)$ with respect to B_2 and with respect to those elements of Π_2 that do correspond to endogenous variables, which are deleted from the first structural equation. Then one has a posterior density $p(\beta_1, \gamma_1, \Pi_1 | D)$. The parameters Π_1 are in this case the implied reduced form parameters that correspond to the endogenous variables that are included in the structural equation. We note that one needs a special value of the prior parameter h in order to obtain a posterior of $p(\beta_1, \gamma_1, \Pi_1 | D)$ that is equal to the posterior studied in the next section. Details of such a rather tedious exercise are left to the interested reader and are omitted here, partly for space consideration.

Third, one can derive the result of Theorem 1 also if the prior (2.14) is replaced by

$$p(\beta_1, \gamma_1, \beta_2, \Gamma_2, \Sigma_r) \propto \|B_r\|^{\tau} |\Sigma_r|^{-h/2} \tag{3.16}$$

and one imposes $\tau - h = m + k + 1$. The justification for this prior is not trivial, since the posterior tends to infinity if B_r tends to infinity. In the case of no restrictions on the structural form one may start with a noninformative prior on the reduced form parameters (Π, Ω) and obtain (3.16) through an enlargement of the parameter space to (Π, B, Ω) and a transformation of variables from (B, Π, Ω) to (B, Γ, Σ).

Fourth, our derivation of Theorem 1 is based on a sequence of integration steps spelled out above. For a different sequence of integration steps based on, a.o., a decomposition of the Wishard density we refer to Drèze (1976).

4. Three representations of the incomplete simultaneous equations model

Consider again the first structural equation of the model given in Section 3 (see (3.3a)),

$$y_1 = Y_1 \beta_1 + Z_1 \gamma_1 + u_1. \tag{4.1}$$

The unrestricted reduced form equations corresponding to the right-hand side variables of Y_1 can be written as

$$Y_1 = Z\Pi_1 + V_1 = Z_1 \Pi_{11} + Z_0 \Pi_{10} + V_1, \tag{4.2}$$

where $\Pi_1 = (\Pi_{11}' \;\; \Pi_{10}')'$ is a $k \times m_1$ matrix of parameters, Π_{11} and Π_{10} being $k_1 \times m_1$ and $k_0 \times m_1$, respectively, and V_1 is a $T \times m_1$ matrix of unobserved disturbances.

The model defined by equations (4.1) and (4.2) incorporates no reduced form equations for other variables than the endogenous variables Y_1 of the right-hand side of (4.1). This will be typically the case when one starts an investigation with (4.1) considered as a regression model (i.e., model y_1 conditionally on Y_1 and Z_1), and later questions the exogeneity status of a subset of variables (Y_1) and adds equations (4.2). This is referred to as *the incomplete simultaneous equations model* (see Richard, 1984; Hendry and Richard, 1983; Drèze and Richard, 1983; Zellner, Bauwens and Van Dijk, 1988).

We note that, if one starts with a structural single equation analysis, some variables of Y, say Y_0, do not appear in the structural equation of y_1, but the reduced form (4.2) does contain equations for the endogenous variables Y_0 that do not appear in (4.1). Further, the incomplete simultaneous equations model contains, usually, less predetermined variables than the structural single equation analysis. A model with extra reduced form equations and extra predetermined variables is relevant when limited information inference is considered as a way to simplify the computational problem of estimation of a structural form.

By substituting $Z\Pi_1 + V_1$ for Y_1 in (4.1), one obtains the restricted reduced form equation of y_1, which can be written as

$$y_1 = Z\pi_1 + v_1 = Z_1\pi_{11} + Z_0\pi_{10} + v_1, \tag{4.3a}$$

where $\pi_1 = (\pi_{11}', \pi_{10}')'$ is a $k \times 1$ vector restricted by

$$\pi_{11} = \gamma_1 + \Pi_{11}\beta_1, \tag{4.3b}$$

$$\pi_{10} = \Pi_{10}\beta_1, \tag{4.3c}$$

and $v_1 = u + V_1\beta$. Thus, one can interpret (4.3a)-(4.3c) as a *nonlinear regression* model (see Zellner, 1971, Chapter 9). There are several other interpretations for the model (4.1)-(4.2), for instance, the *unobserved independent variables model* (see, e.g., Zellner, 1970; Zellner, Bauwens and Van Dijk, 1988).

We make use of the following notation. We write (4.1)-(4.2) in matrix notation as

$$Y^*L + Z(\gamma \;\; \Pi_1) = (u_1 \;\; V_1), \tag{4.1$'$}$$

where

$$Y^* = (y_1 \;\; Y_1), \qquad L = \begin{pmatrix} 1 & 0' \\ -\beta_1 & I_{m_1} \end{pmatrix}.$$

We assume that the rows of the $T \times (m_1 + 1)$ matrix $(u_1 \;\; V_1)$ are independently, identically and normally distributed, with expectation equal to zero and PDS covariance matrix Ω^*, given as

$$\Omega^* = \begin{pmatrix} \sigma^2 & \omega_1' \\ \omega_1 & \Omega_1 \end{pmatrix}.$$

The likelihood function can be written as

$$L(\beta_1, \gamma_1, \Pi_1, \Omega^* \mid D)$$

$$\propto |\Omega^*|^{-T/2} \exp\{-\tfrac{1}{2}\operatorname{tr}[(u_1 \;\; V_1)'(u_1 \;\; V_1)\Omega^{*-1}]\} \tag{4.4a}$$

$$\propto p(Y^*L) = f_{\mathrm{MN}}^{T\times m}(Y^*L \mid Z(\gamma \;\; \Pi_1), \Omega^* \otimes I_T), \tag{4.4b}$$

where $(u_1 \;\; V_1)$ is restricted as in (4.1)$'$, and $f_{\mathrm{MN}}(\;)$ denotes a matrix-normal density function as defined in Drèze and Richard (1983, Appendix A). In the sequel, (4.1) and (4.2) will be called *Representation I* of the incomplete simultaneous equations model. We shall now define two other representations from alternative parametrizations.

Representation II is defined by decomposing $p(Y^*L)$, the matrix-normal density of $Y^*L = (y_1 - Y_1\beta_1 \;\; Y_1)$, as

$$p(Y^*L) = p(Y_1 \mid y_1 - Y_1\beta_1)\,p(y_1 - Y_1\beta_1), \tag{4.5a}$$

where

$$p(y_1 - Y_1\beta_1) = f_{\mathrm{N}}^{T}(y_1 - Y_1\beta_1 \mid Z_1\gamma_1, \sigma^2 I_T) \tag{4.5b}$$

and

$$p(Y_1 \mid y_1 - Y_1\beta_1) = f_{\mathrm{MN}}^{T\times m_1}(Y_1 \mid Z\Pi_1 + (y_1 - Y_1\beta - Z_1\gamma_1)\delta', \Phi \otimes I_T), \tag{4.5c}$$

with

$$\delta = \sigma^{-2}\omega_1, \tag{4.5d}$$

$$\Phi = \Omega_1 - \omega_1\sigma^{-2}\omega_1'. \tag{4.5e}$$

In (4.5b), $f_{\mathrm{N}}(\;)$ denotes a multivariate normal density function. The parameters β_1, γ_1, Π_1, δ, σ^2 amd Φ are in one-to-one correspondence with those of Representation I (see 4.4a) and the likelihood function $L(\beta_1, \gamma_1, \sigma^2, \Pi_1, \delta, \Phi \mid D)$ is the product of (4.5b) and (4.5c). This way of parametrizing the incomplete simultaneous equations model has been used by Drèze and Richard (1983, Sections 2 and 5).

Representation III is defined by the dual decomposition

$$p(Y^*L) = p(y_1 - Y_1\beta_1 \mid Y_1)\,p(Y_1), \tag{4.6a}$$

where

$$p(y_1 - Y_1\beta_1 | Y_1) = f_N^T(y_1 - Y_1\beta_1 | Z_1\gamma_1 + (Y_1 - Z\Pi_1)\eta, \tau^2 I_T) \tag{4.6b}$$

and

$$p(Y_1) = f_{MN}^{T\times m_1}(Y_1 | Z\Pi_1, \Omega_1 \otimes I_T), \tag{4.6c}$$

with

$$\eta = \Omega_1^{-1}\omega_1, \tag{4.6d}$$

$$\tau^2 = \sigma^2 - \omega_1'\Omega_1^{-1}\omega_1. \tag{4.6e}$$

The likelihood function $L(\beta_1, \gamma_1, \eta, \tau^2, \Pi_1, \Omega_1 | D)$ is proportional to the product of (4.6b) and (4.6c). This way of parametrizing the model has been used by Zellner, Bauwens and Van Dijk (1988).

Both reparametrizations have the advantage of cutting explicitly the likelihood function (4.4a) of the original parametrization into the product of two factors. This proves to be useful for defining some classes of prior densities and for integrating analytically some (possibly) nuisance parameters in order to derive posterior densities of parameters of interest, as we shall see in the next section.

In regression form the three models are summarized in Table 2. In Representation II the structural disturbances are added as artificial regressors in the reduced form equation. In Representation III the reduced form disturbances are added as artificial regressors to the structural equation. This turns out to be useful if one analyzes exogeneity; see Engle (1984) and Section 6. Note that v_{1t}' denotes the tth row of V_1; e_{1t}' is the tth row of E_1; u_{1t} is the tth element of u_1; and ε_{1t} is the tth element of ε_1.

Table 2
Three representations of an incomplete simultaneous equations model

Representation	I	II	III
Structural form	$y_1 - Y_1\beta_1 - Z_1\gamma_1 = u_1$	u_1 remains	u_1 replaced by $(Y_1 - Z\Pi_1)\eta + \varepsilon_1$
Reduced form	$Y_1 - Z\Pi_1 = V_1$	V_1 replaced by $(y_1 - Y_1\gamma_1 - X_1\beta_1)\delta' + E_1$	V_1 remains
	$\mathrm{cov}(u_{1t}, v_{1t}) = \omega_1$	$\mathrm{cov}(u_{1t}, e_{1t}) = 0$	$\mathrm{cov}(\varepsilon_{1t}, v_{1t}) = 0$

5. Prior and posterior densities for the incomplete simultaneous equations model

When one uses the incomplete simultaneous equations model, one is interested mainly in doing inference on the coefficients β_1 and γ_1 of the structural equation (4.1) and in predicting future values of y_1; for the latter purpose, it is necessary to do inference on π_1 and Π_1; see (4.1)-(4.3). In this section, we discuss several classes of prior densities that may prove to be useful in practice and derive the corresponding posterior densities of the parameters $\beta_1, \gamma_1, \Pi_1, \sigma^2$ and Ω_1. Inference on τ^2 and Φ is not discussed in this paper. Inference on ω_1, δ and η is briefly discussed in Section 6.

5.1. Prior densities

To remain in the spirit of limited information, we assume that prior information may be available on the structural parameters β_1 and γ_1 and assume prior densities that are noninformative on the other parameters. We review three cases:

(i) noninformative prior densities on all structural parameters;

(ii) natural-conjugate prior densities on either β_1, γ_1 and σ^2 or β_1, γ_1, η and τ^2 — natural-conjugate with respect to (4.5b) or (4.6b);

(iii) independent Student prior densities on β_1, γ_1 and β_1, γ_1 and η.

Case (i). Non-informative prior densities

For Representation I, it is given by

$$p(\beta_1, \gamma_1, \Pi_1, \Omega^*) \propto |\Omega^*|^{-(\alpha_0+m_1+2)/2}. \tag{5.1}$$

That is, the elements of β_1, γ_1 and Π_1 are distributed uniformly and independently of Ω^* which has a degenerate inverted Wishart distribution with α_0 degrees of freedom; Drèze (1976) proposes the value $\alpha_0 = k$ (the number of predetermined variables) by reference to an invariance argument that is specific to the underidentified simultaneous equations model and is reformulated by Drèze and Richard (1983, Section 5). Zellner (1971) proposes $\alpha_0 = 0$ by reference to Jeffreys' invariance principle applied to the reduced form.

For Representation II, a noninformative prior can be defined as the product of a noninformative prior $p(\beta_1, \gamma_1, \sigma^2)$, defined with respect to (4.5b), and a noninformative prior $p(\Pi_1, \delta, \Phi | \beta_1, \gamma_1, \sigma^2)$, defined independently with respect to (4.5c). The two noninformative priors are defined as if the regression equations of Representation II in Table 2 are unrelated and one assumes that β_1 and γ_1 are known in the enlarged reduced form equation of Representation II. We write this prior as

$$p(\beta_1, \gamma_1, \sigma^2, \Pi_1, \delta, \Phi) \propto (\sigma^2)^{-(\mu_0+2)/2}|\Phi|^{-(\nu_0+m_1+1)/2}. \tag{5.2}$$

By setting $\mu_0=0$ and $\nu_0=0$, one applies Jeffreys' invariance principle to (4.1) and the enlarged reduced form equation of Representation II. By choosing $\mu_0 = m_1 + k_1$, the number of regressors of (4.1), and $\nu_0 = k+1$, the number of regressors in the enlarged reduced form equation, one obtains degenerate limits of natural-conjugate densities. In (5.2), the elements of β_1, γ_1, Π_1 and δ are distributed uniformly and independently of σ^2 and Φ; σ^2 has a degenerate inverted-gamma density with μ_0 degrees of freedom and is independent of Φ, and the latter has a degenerate inverted-Wishart density with ν_0 degrees of freedom.

For Representation III a noninformative prior can be defined in the same was as for Representation II. This yields

$$p(\beta_1, \gamma_1, \eta, \tau^2, \Pi_1, \Omega_1) \propto (\tau^2)^{-(\kappa_0+2)/2}|\Omega_1|^{-(\lambda_0+m_1+1)/2}. \tag{5.3}$$

For Jeffreys' invariance principle, applied to the enlarged structural equation of Representation III where Π_1 is assumed to be known, one sets $\kappa_0=0$, and for (4.2), one sets $\lambda_0=0$. For degenerate limits of natural-conjugate densities, one sets $\kappa_0 = 2m_1 + k_1$ and $\lambda_0 = k$.

The noninformative priors (5.1) to (5.3) can differ since they are defined for different parametrizations; (5.2) and (5.3) have two parameters but (5.1) has only one. These priors can be made compatible under certain conditions.

Theorem 2. *The noninformative prior measures* (5.1), (5.2) *and* (5.3) *are identical if, and only if, their parameters* α_0, κ_0, λ_0, μ_0 *and* ν_0 *satisfy the following relations*:

$$\alpha_0 = \kappa_0 - m_1 = \lambda_0 + 1 = \mu_0 + m_1 = \nu_0 - 1. \tag{5.4}$$

Proof. The Jacobian of the transformation from the parameters of Representation I to those of Representation II is $(\sigma^2)^{m_1}$, and the Jacobian of the transformation of the parameters of Representation I to those of Representation III is $|\Omega_1|$. Since $|\Omega^*| = \sigma^2|\Phi| = \tau^2|\Omega_1|$, one can easily make transformations of random variables from (5.1) to the implied prior measures on the parameters of Representation II and Representation III, and then compare the exponents with those of (5.2) and (5.3) to get (5.4). □

Table 3
Prior values of degrees of freedom parameters

Parameters	Drèze's invariance ($\alpha_0 = k$)	Jeffreys' invariance ($\alpha_0 = 0$)
μ_0	$k - m_1$	$-m_1$
ν_0	$k + 1$	1
κ_0	$k + m_1$	m_1
λ_0	$k - 1$	-1

Notice that it is not possible to satisfy (5.4) and at the same time to choose both μ_0 and ν_0 (or κ_0 and λ_0) according to Jeffreys' invariance principle or so as to obtain degenerate limits of natural-conjugate prior densities. Table 3 gives the values of the prior parameters κ_0, λ_0, μ_0 and ν_0 that satisfy (5.4) for the values of α_0 that have been proposed by Drèze and Zellner.

We give two remarks on noninformative prior densities.

First, in the incomplete simultaneous equations model one can choose different values of the degrees of freedom parameter and obtain the (1-1) poly-t class of posterior densities; see Subsection 5.2, while in the structural single equation analysis of Section 3 one may take only one value of the degrees of freedom parameter for the derivation of the (1-1) poly-t density.

Another class of noninformative prior densities is one where the prior is taken as proportional to the information matrix. We leave it to the interested reader to work this out for the three model representations; see also Zellner, Bauwens and Van Dijk (1988).

Note: the reader not interested in informative prior densities can go directly to Subsection 5.2.

Case (ii). Partially natural-conjugate prior densities

They can be defined only in the parametrizations of Representations II and III; this is one advantage of these reparametrizations. In fact, (5.2) and (5.3) are limiting câses of these partially natural-conjugate prior densities for Representations II and III. The natural-conjugate prior densities that we shall define are "partial", in the sense that they are defined with respect to a part of the likelihood function, e.g., (4.5b) for β_1, γ_1 and σ^2 — not taking into account the occurrence of β_1 and γ_1 in the other part (4.5c) — and (4.5c) for Π_1, δ and Φ, assuming that β_1 and γ_1 are known; for the complete set of parameters, the product of the two partial natural-conjugate prior densities is therefore not natural-conjugate. However, because we assume that there is usually no prior information except on the parameters of the structural equation, we shall use the limiting degenerate form of the partial natural-conjugate prior densities that could be defined on the parameters Π_1, δ and Φ (Representation II) or Π_1 and Ω_1 (Representation III), i.e., the second factor that appear in the right-hand side of (5.2) or (5.3).

For Representation II, the prior is given by

$$p(\beta_1, \gamma_1, \sigma^2, \Pi_1, \delta, \Phi) \propto f_N^l\left(\begin{bmatrix}\beta_1\\ \gamma_1\end{bmatrix} \middle| \theta_0, \sigma^2 M_0^{-1}\right) \times f_{ig}(\sigma^2 | s_0^2, \mu_0) |\Phi|^{-(\nu_0+m_1+1)/2}, \tag{5.5}$$

where $f_{ig}(\ )$ denotes an inverted gamma density. From (5.5), one sees that conditionally on σ^2, β_1 and γ_1 are jointly normally distributed, with an expectation equal to θ_0 and a covariance matrix proportional to σ^2, and σ^2 is marginally distributed as an inverted gamma variable with μ_0 degrees of freedom and scale parameter s_0^2.

For Representation III, the prior is given by

$$p(\beta_1, \gamma_1, \eta, \tau^2, \Pi_1, \Omega_1) \propto f_N^n\left(\begin{bmatrix}\beta_1\\ \gamma_1\\ \eta\end{bmatrix} \middle| \begin{bmatrix}\theta_0\\ \eta_0\end{bmatrix}, \tau^2 G_0^{-1}\right) \times f_{ig}(\tau^2 | \tau_0^2, \kappa_0) |\Omega_1|^{-(\lambda_0+m_1+1)/2}, \tag{5.6}$$

where $n = l + m_1 = 2m_1 + k_1$. Conditionally on τ^2, β_1, γ_1 and η are jointly normally distributed, with an expectation independent of τ^2 and a covariance matrix proportional to τ^2, and τ^2 is marginally distributed as an inverted-gamma variable with κ_0 degrees of freedom and scale parameter τ_0^2. To be noninformative on η, one has to fix the appropriate elements of G_0 to 0.

Notice that (5.5) and (5.6) can easily be defined so that the marginal prior distribution of θ (i.e., β_1 and γ_1) is identical since it is a multivariate-t distribution.

Case (iii). Independent multivariate-t densities

It is defined by taking the marginal prior density of β_1, γ_1 (and possibly η for Representation II) obtained from (5.5) or (5.6).

For representation I, we write it as

$$p(\beta_1, \gamma_1, \Pi_1, \Omega^*) \propto f_t^l\left(\begin{bmatrix}\beta_1\\ \gamma_1\end{bmatrix} \middle| \theta_0, M_0, \zeta_0\right) |\Omega^*|^{-(\alpha_0+m+2)/2}. \tag{5.7}$$

That is, β_1 and γ_1 are distributed independently of the other parameters, according to a multivariate-t distribution with expectation equal to θ_0, covariance matrix equal to $\zeta_0 M_0^{-1}/(\zeta_0 - 2)$ (if $\zeta_0 > 2$), and ζ_0 degrees of freedom, and the other parameters have the same noninformative prior as in (5.1).

For Representation II, we define the prior as

$$p(\beta_1, \gamma_1, \sigma^2, \Pi_1, \delta, \Phi) \propto p(\beta_1, \gamma_1)(\sigma^2)^{-(\mu_0+2)/2} |\Phi|^{-(\nu_0+m_1+1)/2}, \tag{5.8}$$

where $p(\beta_1, \gamma_1)$ is the multivariate-t density of (5.7); again, the prior on the other parameters is the same noninformative one as in (5.2).

For Representation III, the prior is

$$p(\beta_1, \gamma_1, \eta, \tau^2, \Pi_1, \Omega_1) \propto f_t^n\left(\begin{bmatrix}\beta_1\\ \gamma_1\\ \eta\end{bmatrix} \middle| \begin{bmatrix}\theta_0\\ \eta_0\end{bmatrix}, G_0, \zeta_0\right)(\tau^2)^{-(\kappa_0+2)/2} \times |\Omega_1|^{-(\lambda_0+m_1+1)/2}. \tag{5.9}$$

That is, β_1, γ_1 and η have a joint multivariate-t distribution, but one can easily remain noninformative on η; the prior on the other parameters is the same as in (5.3). We assume that the marginal density of β_1, γ_1 from (5.9) is exactly the multivariate-t density that appears in (5.7), denoted $p(\beta_1, \gamma_1)$ in the sequel.

5.2. *Posterior densities*

In this subsection we present two theorems that define different classes of posterior densities of the parameters β_1, γ_1 and Π_1. We concentrate on posterior densities for the case of the noninformative prior densities of Subsection 5.1 and for the models of Section 4. The posterior results for the informative prior densities of Subsection 5.1 and the models of Section 4 are similar to the results for the noninformative prior densities. That is, the same classes of posterior densities can be derived, only the parameter values of these posteriors differ from the noninformative case. Due to space considerations we have omitted the posterior results for the informative prior densities.

In order to make the sequence of integration steps in the derivation of the posterior densities of β_1, γ_1 and Π_1 more transparent, two summaries of these steps are shown in Figures 6 and 7. Figure 6 gives the steps according to the specification of Representation I. Figure 7 gives the steps after the reparametrizations that define Representations II and III.

$p(\beta_1, \gamma_1, \Pi_1, \Omega^* | D)$ (equation (5.1) × equation (4.4a))

↓

inverted Wishart step on Ω^*

↓

$p(\beta_1, \gamma_1, \Pi_1 | D)$

↓

expand determinant $|(u_1 \;\; V_1)'(u_1 \;\; V_1)|$ as:

↓	↓
$(u_1'u_1)\|V_1'M_{u_1}V_1\|$	$\|V_1'V_1\|(u_1'M_{V_1}u_1)$
↓	↓
complete squares on Π_1	complete squares on β_1, γ_1
↓	↓
$p(\Pi_1 \| \beta_1, \gamma_1, D) \propto$ matrix-t	$p(\beta_1, \gamma_1 \| \Pi_1, D) \propto$ multivariate-t
↓	↓
matrix-t step on Π_1	multivariate-t step on β_1, γ_1
↓	↓
$p(\beta_1, \gamma_1 \| D) \propto$ (1-1) poly-t	$p(\Pi_1 \| D) \propto$ (2-1) poly-matrix-t (Zellner, Bauwens and Van Dijk, 1988)

Figure 6. Scheme of integration steps for the posterior densities of β_1, γ_1, Π_1.

We start in Figure 6 with the joint posterior of $(\beta_1, \gamma_1, \Pi_1, \Omega^*)$ for Model Representation I, i.e., the product of equations (5.1) and (4.4a). In a similar way as done in Sections 2 and 3, one can integrate this density with respect to Ω^* using the inverted Wishart density function. This yields

$$p(\beta_1, \gamma_1, \Pi_1 | D) \propto |(u_1 \;\; V_1)'(u_1 \;\; V_1)|^{-(T+\alpha_0)/2}. \tag{5.10}$$

Note that the determinant in (5.10) is given as

$$|(u_1 \;\; V_1)'(u_1 \;\; V_1)| = \begin{vmatrix} (y_1 - Y_1\beta_1 - Z_1\gamma_1)'(y_1 - Y_1\beta_1 - Z_1\gamma_1) & (y_1 - Y_1\beta_1 - Z_1\gamma_1)'(Y_1 - Z\Pi_1) \\ (Y_1 - Z\Pi_1)'(y_1 - Y_1\beta_1 - Z_1\gamma_1) & (Y_1 - Z\Pi_1)'(Y_1 - Z\Pi_1) \end{vmatrix}$$

and the projection matrices are defined as

$$M_{u_1} = I - u_1(u_1'u_1)^{-1}u_1', \qquad M_{V_1} = I - V_1(V_1'V_1)^{-1}V_1'.$$

The two equivalent ways to expand this determinant correspond to two ways to decompose the joint density $p(\beta_1, \gamma_1, \Pi_1 | D)$ into a conditional density and a marginal density. That is, the decomposition $p(\beta_1, \gamma_1, \Pi_1 | D) = p_c(\beta_1, \gamma_1 | \Pi_1, D) p_m(\Pi_1 | D)$ is used by Zellner, Bauwens and Van Dijk (1988) and the decomposition $p(\beta_1, \gamma_1, \Pi_1 | D) = p_c(\Pi_1 | \beta_1, \gamma_1, D) p_m(\beta_1, \gamma_1 | D)$ is similar to the derivation of Drèze and Richard (1983). We note that the poly-matrix-t class of densities that we derive in this paper was already mentioned as a possible class of densities by Dickey (1968). Due to space considerations we omit the detailed steps of the derivations.

A second approach to derive the posterior densities of β_1, γ_1 and Π_1 is to reparametrize the basic Representation I as done in Subsection 5.2 into the Representations II and III and to use the noninformative prior densities corresponding to these model representations. We summarize the different steps in Figure 7. An explicit proof is given in Appendix B and Appendix C.

The results of Figure 7 are summarized in the following two theorems.

parameter set $(\beta_1, \gamma_1, \Pi_1, \Omega^*)$

↓

REPARAMETRIZATION

↓

↓		↓	
Representation II		Representation III	
↓		↓	
$(\beta_1, \gamma_1, \sigma^2)$ and (Π_1, δ, Φ)		$(\beta_1, \gamma_1, \eta, \tau^2)$ and (Π_1, Ω_1)	
↓		↓	
$p(\beta_1, \gamma_1, \sigma^2 \| D) p(\Pi_1, \delta, \Phi \| \beta_1, \gamma_1, D)$		$p(\beta_1, \gamma_1, \eta, \tau^2 \| \Pi_1, D) p(\Pi_1, \Omega_1 \| D)$	
↓	↓	↓	↓
inverted gamma-2 step on σ^2	inverted Wishart step on Φ	inverted gamma-2 step on τ^2	inverted Wishart step on Ω_1
↓	↓	↓	↓
$p(\beta_1, \gamma_1 \| D) p(\Pi_1, \delta \| \beta_1, \gamma_1, D)$		$p(\beta_1, \gamma_1, \eta \| \Pi_1, D) p(\Pi_1 \| D)$	
↓		↓	
conditional matrix-t step on (Π_1, δ), given (β_1, γ_1)		conditional multivariate-t step on $(\beta_1, \gamma_1, \eta)$, given Π_1	
↓		↓	
$p(\beta_1, \gamma_1 \| D) \propto$ (1-1) poly-t Drèze and Richard (1983)		$p(\Pi_1 \| D) \propto$ (2-1) poly-matrix-t	

Figure 7. Reparametrizations and scheme of integration steps for posterior densities of β_1, γ_1, Π_1.

Theorem 3. *Given the class of noninformative prior densities of Theorem 2 and given Model Representation II, the marginal posterior density of the structural parameters* (β_1, γ_1) *is a* (1–1) *poly-t density, and the conditional posterior density of* Π_1 *given a value of* (β_1, γ_1), *is a matrix-t density.*

The posterior of (β_1, γ_1) *is defined, through its kernel,*

$$p(\beta_1, \gamma_1 | D) \propto [(y_1 - Y_1\beta_1)' M_Z (y_1 - Y_1\beta_1)]^{(T+\alpha_0-m_1-k)/2} \times [(y_1 - Y_1\beta_1 - Z_1\gamma_1)'(y_1 - Y_1\beta_1 - Z_1\gamma_1)]^{-(T+\alpha_0-m_1)/2}. \tag{5.11}$$

The density of Π_1 *is defined as*

$$p(\Pi_1 | \beta_1, \gamma_1, D) = f_{\mathrm{MT}}^{k \times m_1}(\Pi_1 | \tilde{\Pi}_1, Y_1' \tilde{M}_{u_1} Y_1, Z' M_{u_1} Z, T + \alpha_0 - k), \tag{5.12}$$

where $\tilde{\Pi}_1 = (Z' M_{u_1} Z)^{-1} Z' M_{u_1} Y_1$ *and*

$$\tilde{M}_{u_1} = M_{u_1} - M_{u_1} Z (Z' M_{u_1} Z)^{-1} Z' M_{u_1}, \qquad M_{u_1} = I_T - u_1 (u_1' u_1)^{-1} u_1'.$$

A proof of Theorem 3 is given in Appendices B and C. We note that $p(\beta_1, \gamma_1 | D)$ has the same functional form as given in Section 3, equation (3.15); compare equations (3.11)–(3.14). We remark that this method of derivation is, essentially, one followed by Drèze and Richard (1983).

Note that if one conditions $p(\Pi_1 | \beta_1, \gamma_1, D)$ on the limited information maximum likelihood estimator (LIML) of β_1, γ_1, then it can be shown that $\tilde{\Pi}_1$, the conditional posterior expectation of Π_1 is equal to the LIML estimator of Π_1.

The second theorem on posterior densities for β_1, γ_1 and Π_1 is based on Model Representation III. It is stated as follows.

Theorem 4. *Given the class of noninformative prior densities of Theorem 2 and given Model Representation III, the conditional posterior density of the structural parameters* (β_1, γ_1), *given a value of* Π_1, *is a multivariate-t density, and the marginal posterior density of* Π_1 *is a* (2–1) *poly-matrix-t density.*

The density of (β_1, γ_1) *is defined as*

$$p(\beta_1, \gamma_1 | \Pi_1, D) = f_t^{m_1+k_1}((\beta_1', \gamma_1')' | \hat{\theta}_1, (y_1' \tilde{M}_{V_1} y_1)^{-1} X' M_{V_1} X, T + \alpha_0 - m_1 - k_1), \tag{5.13}$$

where $X = (Y_1 \;\; Z_1)$ *and*

$$\hat{\theta}_1 = (X' M_{V_1} X)^{-1} X' M_{V_1} y_1, \qquad M_{V_1} = I_T - V_1 (V_1' V_1)^{-1} V_1',$$
$$V_1 = Y_1 - Z\Pi_1, \qquad \tilde{M}_{V_1} = M_{V_1} - M_{V_1} X (X' M_{V_1} X)^{-1} X' M_{V_1}.$$

The density of Π_1 *is defined through its kernel as*

$$p(\Pi_1 | D) \propto | Y_1' \bar{M}_X Y_1 + (\Pi_1 - \bar{\Pi}_1)' Z' M_X Z (\Pi_1 - \bar{\Pi}_1)|^{(T+\alpha_0-1-l)/2} \times | Y_1' \tilde{M}_X^* Y_1 + (\Pi_1 - \Pi_1^*)' Z' M_X^* Z (\Pi_1 - \Pi_1^*)|^{-(T+\alpha_0-l)/2} \times | Y_1' M_Z Y_1 + (\Pi_1 - \hat{\Pi}_1)' Z' Z (\Pi_1 - \hat{\Pi}_1)|^{-(T+\alpha_0-1)/2}, \tag{5.14}$$

where $l = m_1 + k_1$ *and*

$$\bar{M}_X = M_X - M_X Z (Z' M_X Z)^{-1} Z' M_X, \quad \bar{\Pi}_1 = (Z' M_X Z)^{-1} Z' M_X Y_1,$$
$$\tilde{M}_X^* = M_X^* - M_X^* Z (Z' M_X^* Z)^{-1} Z' M_X^*, \quad \Pi_1^* = (Z' M_X^* Z)^{-1} Z' M_X^* Y_1,$$
$$M_X^* = M_X - M_X y_1 (y_1' M_X y_1)^{-1} y_1' M_X, \quad \hat{\Pi}_1 = (Z'Z)^{-1} Z' Y_1. \qquad \square$$

A proof of Theorem 4 is omitted and left to the interested reader. It follows the similar steps as taken in Appendix B and C.

We end this section with two remarks.

First, one may extend Theorems 3 and 4 to the case where one has a conditional normal-inverted gamma-2 prior on the structural parameters and a locally uniform prior on the reduced form parameters. Similarly, one may use a multivariate-*t* prior on the structural parameters and be locally uniform on reduced form parameters. Details are omitted for space considerations.

Second, we have not stated explicitly the conditions that are sufficient for the existence of the integrals defined in the posterior distributions. These conditions are similar to the ones discussed in Section 3 and will be investigated explicitly in future work.

6. Remarks on Bayesian inference on exogeneity and overidentification

The posterior densities of β_1, γ_1 and Π_1 of Section 5 can be used for a diagnostic analysis of overidentifying restrictions and exogeneity restrictions. Of course, another approach is to specify prior odds and derive posterior odds for the restrictions mentioned above. However, some simple checks are the following.

The (1–1) poly-*t* density of (β_1, γ_1) (see (3.15) and Theorems 3 and 4) has as numerator the posterior density of the equation parameters of the standard linear model. Therefore, the variation in the denominator is an indication of the plausibility of exogeneity of the variables Y_1. If this function is constant, one may conclude that the variables Y_1 are exogenous.

A parametric analysis of exogeneity can be conducted in each of the three model representations. In Representation III one may investigate the posterior density of the parameter η. Algorithmic procedures for this purpose are given in Zellner, Bauwens and Van Dijk (1988). In Model Representation II one may analyze the posterior density of the parameter δ (see Lubrano, Pierce and Richard, 1986). In Model Representation I, one may investigate the posterior density of ω_1. In small samples, different results may emerge for the different representations. This is a topic for further research.

The validity of the overidentifying restrictions may be verified by comparing the (2-1) poly-matrix-t of Π_1 and the unrestricted matrix-t of Π_1.

One may also add variables to the structural equation and check how "close" the posterior density of its parameter is concentrated around zero.

Clearly, this is only a tentative list. More research in this area needs to be done. The use of a predictive approach is a worthwhile topic of research in this context.

7. Conclusions

We have reviewed Bayesian limited information analysis of the simultaneous equations model. Our results can be summarized as follows.

The marginal likelihood function of an underidentified SEM is not constant, but is explosive in some directions of the parameter space while it is more regular in other directions. Locally uniform priors are, therefore, not a suitable class of prior densities and there is a need for informative prior information.

Drèze's result that the parameters of a single structural equation have a (1-1) poly-t density holds under the condition that the prior degrees of freedom parameter has a particular value. It is of interest that in the incomplete simultaneous equations model one can derive the class of (1-1) poly-t densities for several different values of the degrees of freedom parameter.

One can show that the marginal posterior density of the implied reduced form parameters in the incomplete SEM is a so-called (2-1) poly-matrix-t density.

It appears that the different representations of the incomplete simultaneous equations model have each particular properties that are useful for the analysis of exogeneity and overidentification.

Clearly, the results are to be used in empirical studies. In applied work there is a need for diagnostic analysis on the plausibility of the prior assumptions. If one is interested in posterior results on the nuisance parameters such as the covariance matrix of the disturbances, one needs a decomposition of the inverted Wishart density along the lines discussed by Drèze (1976).

Appendix A

In this appendix we give some details on the Figures 2 to 5. First, we discuss the full information case; Figures 2 to 4. From (2.15b) and (2.15c) one can write the joint posterior density of (B, Γ) as

$$p(B, \Gamma \mid D) \propto |B'\hat{\Omega}B + (\Gamma + \hat{\Pi}B)'Z'Z(\Gamma + \hat{\Pi}B)|^{-(T+h-m-1)/2}|B'B|^{T/2}. \tag{A.1}$$

For Figures 2-4 we interpret B and Γ as scalars and take the data matrices $\hat{\Omega}$, $Z'Z$ and $\hat{\Pi}$ as $\hat{\Omega}=1$, $Z'Z=1$, $\hat{\Pi}=0.10$. Further, $m=1$ and $k=1$. Then (A.1) becomes

$$p(B, \Gamma \mid D) \propto [B^2 + (\Gamma + 0.1B)^2]^{-(T+h-2)/2}(B^2)^{T/2}. \tag{A.2}$$

Clearly, if $h=0$ or $h=1$, the right-hand side of (A.2) tends to infinity when B tends to infinity. If $h>2$ then (A.2) tends to infinity when B tends to zero. If $h=2$, then one has the concentrated likelihood function that is only undefined when $B=0$ and $\Gamma=0$. Note that in the case of $h=2$, the least upperbound for (A.2) in the point (0, 0) is equal to one. One can rewrite (A.2) as

$$p(B, \Gamma \mid D) \propto \left[1 + \left(\frac{\Gamma + 0.1B}{B}\right)^2\right]^{-(T+h-2)/2} |B|^{-1}(B^2)^{-(h-3)/2}. \tag{A.3}$$

Apart from the last factor, (A.3) is proportional to a conditional *univariate* t density of Γ given a value of B. A kernel of this conditional density is shown in Figure 2 for values of B in the interval $[-4, 4]$.

In the limited information case, Figure 5, we start with equation (3.7) and interpret $\tilde{\beta}_1$ and $\tilde{\gamma}_1$ as scalars. Further $m=1$, $k=1$, $\hat{\Omega}=1$, $Z'Z=1$, and $\hat{\Pi}=0.1$. Then one has

$$p(\tilde{\beta}_1, \tilde{\gamma}_1 \mid D) \propto (\tilde{\beta}_1^2)^{T/2}[\tilde{\beta}_1^2 + (\tilde{\gamma}_1 + 0.1\tilde{\beta}_1)^2]^{-(T+1)/2}. \tag{A.4}$$

Clearly, this is the same function as (A.2) with $h=3$. If one replaces $\tilde{\beta}_1$ by β_1 and $\tilde{\gamma}_1$ by γ_1, and if we impose on the right-hand side of (A.4) that $\tilde{\beta}_1^2 = (1+\beta_1^2)$, then one has a kernel of a (1-1) poly-t density as shown in Figure 5.

Appendix B

Proof of Theorem 3.

– Model Representation II is given in Table 2 and repeated here for convenience with some new notation,

$$y_1 = (Y_1 \quad Z_1)\begin{pmatrix}\beta_1\\ \gamma_1\end{pmatrix} + u_1 := X\theta + u_1, \tag{B.1}$$

$$Y_1 = (Z \quad u_1)\begin{pmatrix}\Pi_1\\ \delta'\end{pmatrix} + E := W\Delta + E. \tag{B.2}$$

Given the assumptions stated in Section 4, the likelihood function of this representation can be written as the product of (4.5b) and (4.5c),

$$L(\beta_1, \gamma_1, \sigma^2, \Pi_1, \delta, \Phi \mid D) = f_{\rm N}^T(y_1 \mid Y_1\beta_1 + Z_1\gamma_1, \sigma^2 I_T) \times f_{\rm MN}^{T\times m}(Y_1 \mid Z\Pi_1 + u\delta', \Phi \otimes I_T). \tag{B.3}$$

– The prior density is given by (5.2) together with (5.4), i.e.,

$$p(\beta_1, \gamma_1, \sigma^2, \Pi_1, \delta_1, \Phi) \propto (\sigma^2)^{-(\alpha_0+2-m_1)/2} |\Phi|^{-(\alpha_0+2+m_1)/2}. \tag{B.4}$$

– Multiply $L(\cdot)$ with $p(\cdot)$; use the definition of the MN density function; and complete the squares on Δ. Then one obtains

$$p(\beta_1, \gamma_1, \sigma^2, \Pi_1, \delta, \Phi \mid D) \propto (\sigma^2)^{-(\alpha_0+2-m_1)/2} f_{\rm N}^T(y_1 \mid Y_1\beta_1 + Z_1\gamma_1, \sigma^2 I_T)|\Phi|^{-(T+\alpha_0+m_1+2)/2} \times \exp\{-\tfrac{1}{2}\,\mathrm{tr}\,\Phi^{-1}[Y_1'M_W Y_1 + (\Delta - \hat{\Delta})'W'W(\Delta - \hat{\Delta})]\}, \tag{B.5}$$

where

$$M_W = I_T - W(W'W)^{-1}W' \text{ and } \hat{\Delta} = (W'W)^{-1}W'Y_1. \tag{B.6}$$

– From the definition of the MN and IW density functions, one recognizes in the formula above that

$$p(\Delta, \Phi \mid \beta_1, \gamma_1, \sigma^2, D) = p(\Delta \mid \Phi, \beta_1, \gamma_1, \sigma^2, D)\,p(\Phi \mid \beta_1, \gamma_1, \sigma^2, D), \tag{B.7}$$

where

$$p(\Delta \mid \Phi, \beta_1, \gamma_1, \sigma^2, D) = f_{\rm MN}^{(k+1)\times m_1}(\Delta \mid \hat{\Delta}, \Phi \otimes (W'W)^{-1}) \tag{B.8}$$

and

$$p(\Phi \mid \beta_1, \gamma_1, \sigma^2, D) = f_{\rm IW}^{m_1}(\Phi \mid Y_1'M_W Y_1, T + \alpha_0 - k). \tag{B.9}$$

– Integrating the joint density (B.5) with respect to Δ and Φ gives

$$p(\beta_1, \gamma_1, \sigma^2 \mid D) \propto (\sigma^2)^{-(\alpha_0+2-m_1)/2} f_{\rm N}^T(y_1 \mid Y_1\beta_1 + Z_1\gamma_1, \sigma^2 I_T) \times |W'W|^{-m_1/2} |Y_1'M_W Y_1|^{-(T+\alpha_0-k)/2}. \tag{B.10}$$

The last two factors are the parts of the normalizing constants of the MN and IW conditional densities above, that depend on β_1 and γ_1. Then, we make use of

$$|W'W| = \begin{vmatrix} Z'Z & Z'u_1\\ u_1'Z & u_1'u_1 \end{vmatrix} = |Z'Z|(u_1'M_Z u_1). \tag{B.12}$$

This follows from the usual expansion of determinants. Only $u_1'M_Z u_1$ depends on the parameters; furthermore

$$u_1'M_Z u_1 = (y_1 - Y_1\beta_1 - Z_1\gamma_1)'M_Z(y_1 - Y_1\beta - Z_1\gamma) = (1 \quad -\beta_1')'Y'M_Z Y(1 \quad -\beta_1')', \tag{B.13}$$

where $Y = (y_1 \quad Y_1)$ and $M_Z Z_1 = 0$. In addition

$$|Y_1'M_W Y_1| = |Y_1'Y_1|\,|W'M_{Y_1}W|\,|W'W|^{-1},$$

where $M_{Y_1} = I_T - Y_1(Y_1'Y_1)^{-1}Y_1'$, and

$$\begin{vmatrix} Y_1'Y_1 & Y_1'W\\ W'Y & W'W \end{vmatrix} = |Y_1'Y_1|\,|W'M_{Y_1}W| = |W'W|\,|Y_1'M_W Y_1|.$$

The determinant $|Y_1'Y_1|$ does not depend on the parameters, $|W'W|$ has already been analyzed, and $|W'M_{Y_1}W|$ does not depend on the parameters (see Appendix C).

– Collecting the factors depending on the parameters, using the definition of the normal density $f_{\rm N}(\ )$, and completing squares in $\theta = (\beta_1' \quad \gamma_1')'$

$$p(\beta_1, \gamma_1, \sigma^2 \mid D) \propto [(1 - \beta_1')Y'M_Z Y(1 - \beta_1')']^{(T+\alpha_0-m_1-k)/2} \times (\sigma^2)^{-(T+\alpha_0-m_1+2)/2} \times \exp\left\{-\frac{1}{2\sigma^2}[y_1'M_X y_1 + (\theta - \hat{\theta})'X'X(\theta - \hat{\theta})]\right\} \tag{B.14}$$

($\hat{\theta} = (X'X)^{-1}X'y_1$ and M_X is the projection matrix of X). From the definition of the inverted gamma density (i.e., the one-dimensional IW density), one can then easily integrate out σ^2; the result is (5.11).

– To derive (5.12), we make use of the MN-IW conditional density of Δ, Φ. Notice that this density is in fact conditional on θ and not on σ^2; so we can drop σ^2 from the conditions.

The first step is to integrate out δ from $p(\Delta \mid \Phi, \theta, D)$ by standard properties of the normal density. This yields:

$$p(\Pi_1 \mid \Phi, \theta, D) = f_{\mathrm{MN}}^{k \times m_1}(\Pi_1 \mid \tilde{\Pi}_1, \Phi \otimes (Z' M_{u_1} Z)^{-1}) \tag{B.15}$$

($\tilde{\Pi}_1$ is the submatrix of $\hat{\Delta}$ corresponding to Π_1 and is defined explicitly in Theorem 3, together with M_{u_1}).

The second and last step is to marginalize the latter density with respect to Φ. A well known property of the MN–IW density yields the result stated in (5.12) (see, e.g., Drèze and Richard, 1983, p. 589). □

Appendix C

We show that $|W' M_{Y_1} W|$ does not depend on the parameters β_1 and γ_1. Consider

$$\begin{aligned} M_{Y_1} W &= (M_{Y_1} Z \mid M_{Y_1} y_1 - M_{Y_1} Y_1 \beta_1 - M_{Y_1} Z_1 \gamma_1) \\ &= (M_{Y_1} Z \mid M_{Y_1}(y_1 - Z_1 \gamma_1)), \end{aligned}$$

since $M_{Y_1} Y_1 = 0$. Hence

$$\begin{aligned} |W' M_{Y_1} W| &= \begin{vmatrix} (y_1 - Z_1\gamma_1)' M_{Y_1} (y_1 - Z_1 \gamma_1) & (y_1 - Z_1 \gamma_1)' M_{Y_1} Z \\ Z' M_{Y_1} (y_1 - Z_1 \gamma_1) & Z' M_{Y_1} Z \end{vmatrix} \\ &= |Z' M_{Y_1} Z| \\ &\quad \times [(y_1 - Z_1 \gamma_1)' M_{Y_1} (y_1 - Z_1 \gamma_1) \\ &\qquad - (y_1 - Z_1 \gamma_1)' M_{Y_1} Z (Z' M_{Y_1} Z)^{-1} Z' M_{Y_1} (y_1 - Z_1 \gamma_1)]. \end{aligned}$$

The first factor on the right-hand side depends only on the data. The second is equal to

$$\begin{aligned} & y_1' M_{Y_1} y_1 - y_1' M_{Y_1} Z (Z' M_{Y_1} Z)^{-1} Z' M_{Y_1} y_1 \\ & \quad + \gamma_1' Z_1' M_{Y_1} Z_1 \gamma_1 - \gamma_1' Z_1' M_{Y_1} Z (Z' M_{Y_1} Z)^{-1} Z' M_{Y_1} Z_1 \gamma_1 \\ & \quad - 2\gamma_1' Z_1' M_{Y_1} y_1 + 2 \gamma_1' Z_1' M_{Y_1} Z (Z' M_{Y_1} Z)^{-1} Z' M_{Y_1} y_1 \\ & = y_1' M_{Y_1} y_1 - y_1' M_{Y_1} Z (Z' M_{Y_1} Z)^{-1} Z' M_{Y_1} y_1 \\ & \quad + \gamma' Z' M_{Y_1} Z \gamma - \gamma' Z' M_{Y_1} Z (Z' M_{Y_1} Z)^{-1} Z' M_{Y_1} Z \gamma \\ & \quad - 2 \gamma' Z' M_{Y_1} y_1 + 2 \gamma' Z' M_{Y_1} Z (Z' M_{Y_1} Z)^{-1} Z' M_{Y_1} y_1, \end{aligned}$$

because $Z_1 \gamma_1 = Z\gamma$ upon defining $\gamma = \binom{\gamma_1}{0}$. So, the two terms in the second line cancel, and also the two terms on the last line, while the two terms on the first line depend only on data, and not on parameters.

References

Anderson, T.W. (1949) 'Estimation of the parameters of a single equation in a complete system of stochastic equations', *Annals of Mathematical Statistics*, 20:46–68.

Anderson, T.W. (1984) *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Bauwens, L. (1984) *Bayesian Full Information Analysis of Simultaneous Equation Models using Integration by Monte Carlo*, Berlin: Springer.

Bauwens, L., J.-P. Bulteau, P. Gille, L. Longrée, M. Lubrano and H. Tompa (1981) 'Bayesian regression program (BRP) user's manual', CORE Computing Report 81-A-01.

Bauwens, L. and J.-F. Richard (1985) 'A 1-1 poly-*t* random variable generator with applications to Monte Carlo integration', *Journal of Econometrics*, 29:19–46.

Dickey, J.M. (1967) "Matricvariate generalizations of the multivariate *t* distribution and the inverted multivariate *t* distribution', *Annals of Mathematical Statistics*, 38:511–518.

Dickey, J.M. (1968) 'Three multidimensional-integral identities with Bayesian applications', *Annals of Mathematical Statistics*, 39:1615–1628.

Dijk, H.K. van (1984) 'Posterior analysis of econometric models using Monte Carlo integration', Thesis, Erasmus University Rotterdam.

Dijk, H.K. van (1987) 'Some advances in Bayesian estimation methods using Monte Carlo integration', in: T.B. Fomby and G.F. Rhodes, eds., *Advances in Econometrics 6*. Connecticut: JAI Press, pp. 215–261.

Dijk, H.K. van and T. Kloek (1980) 'Further experience in Bayesian analysis using Monte Carlo integration', *Journal of Econometrics*, 14:307–328.

Drèze, J.H. (1962) 'The Bayesian approach to simultaneous equations estimation', ONR Research Memorandum 67, The Technological Institute, Northwestern University.

Drèze, J.H. (1975) 'Bayesian theory of identification in simultaneous equations models', in: S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*. Amsterdam: North-Holland.

Drèze, J.H. (1976) 'Bayesian limited information analysis of the simultaneous equations model', *Econometrica*, 44:1045–1075.

Drèze, J.H. (1977) 'Bayesian regression analysis using poly-*t* densities', *Journal of Econometrics*, 6:329–354.

Drèze, J.H. and J.-A. Morales (1976) 'Bayesian full information analysis of simultaneous equations', *Journal of the American Statistical Association*, 71:919–923.

Drèze, J.H. and J.-F. Richard (1983) 'Bayesian analysis of simultaneous equation systems', in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, Vol. I*. Amsterdam: North-Holland.

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

Engle, R.F., D.F. Hendry and J.-F. Richard (1983) 'Exogeneity', *Econometrica*, 51:277–304.

Epstein, R.J. (1987) *A History of Econometrics*. Amsterdam: North-Holland.

Haavelmo, T. (1943) 'The statistical implications of a system of simultaneous equations', *Econometrica*, 11:1–12.

Hendry, D.F. and J.-F. Richard (1982) 'The econometric analysis of economic time series', *International Statistical Review*, 51:111–163.

Kloek, T. and H.K. van Dijk (1978) 'Bayesian estimates of equation systems parameters: an application of integration by Monte Carlo', *Econometrica*, 46:1–19.
Lubrano, M., R.G. Pierse and J.-F. Richard (1986) 'Stability of a UK money demand equation: A Bayesian approach to testing exogeneity', *Review of Economic Studies*, 53:603–634.
Maddala, G.S. (1986) 'Weak priors and sharp posteriors in simultaneous equation models', *Econometrica*, 44:345–351.
Magnus, J.R. and H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.
Malinvaud, E. (1978) *Méthodes Statistiques de l'Économétrie*. Paris: Dunod.
Richard, J.-F. (1984) 'Classical and Bayesian inference in incomplete simultaneous equation models', in: D.F. Hendry and K.F. Wallis, eds., *Econometrics and Quantitative Economics*. Oxford: Basic Blackwell.
Richard, J.-F. and H. Tompa (1980) 'On the evaluation of poly-*t* density functions', *Journal of Econometrics*, 12:335–351.
Zellner, A. (1970) 'Estimation of regression relationships containing unobservable independent variables', *International Economic Review*, 11:441–454.
Zellner, A. (1971) *An Introduction to Bayesian inference in Econometrics*. New York: Wiley.
Zellner, A., L. Bauwens and H.K. van Dijk (1988) 'Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo integration', *Journal of Econometrics*, 38:39–72.

PART III

OPTIMISATION