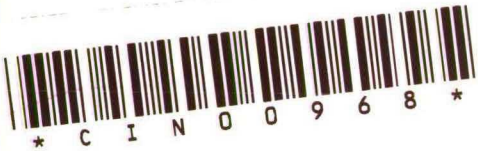
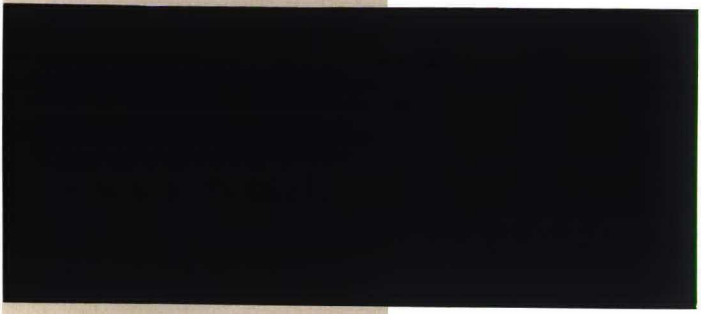
entER

for

nomic Research

Discussion paper





No. 8805

SEEMINGLY UNRELATED REGRESSION EQUATION SYSTEMS UNDER DIFFUSE STOCHASTIC PRIOR INFORMATION: A RECURSIVE ANALYTICAL APPROACH

by

Mark F.J. Steel

October 1988

Seemingly Unrelated Regression Equation Systems under Diffuse Stochastic Prior Information: A Recursive Analytical Approach

Mark F.J. Steel
Department of Econometrics
Tilburg University
P.O.Box 90153
5000 LE Tilburg
The Netherlands

JEL classification: 211

Keywords: Bayesian analysis; SURE models; analytical inference; general linear restrictions

Abstract: (=headnote)
The scope of exact analytical results in Bayesian econometrics is known to be quite limited. It is, however, shown here to be broader than the simple natural-conjugate framework. Restricting the coefficients of a SURE model in a recursive linear way can not be accommodated in a natural-conjugate analysis, but still allows for analytical inference, exploiting the recursive characteristics. These findings help us to clearly define the boundaries of the analytical domain in Bayesian analysis, and subsequent research suggests this may reduce the numerical burden in the analysis of more complicated models than the one under scrutiny here.

Acknowledgements:
The research presented here was conducted under a fellowship of the Royal Netherlands Academy of Arts and Sciences. Useful comments and suggestions by Luc Bauwens, Teun Kloek, Jacek Osiewalski, and Jean-François Richard are gratefully acknowledged. The usual disclaimer applies.

1. Introduction

A major obstacle for the implementation of Bayesian methods in econometrics has undoubtedly been the very limited scope of available straightforward analytical solutions, as was recently stressed in Richard and Steel (1988). For this reason, we have seen quite an upsurge of numerical methods for conducting inference in a Bayesian framework.

On the one hand, Monte Carlo integration techniques were adopted following Kloek and van Dijk (1978) and these methods proved quite successful in numerous empirical applications [see, inter alia, van Dijk (1984), Bauwens (1984), and Steel (1987)], although naturally prone to problems of specificity and imprecision as discussed in Hendry (1984). In particular, the search for a useful importance function in the so-called importance sampling technique [see Hammersley and Handscomb (1979)] can not yet be performed in a fully automatic way and remains an important issue to be resolved. Geweke (1988), Richard and Steel (1988), and Zellner et al. (1988) all address this problem in more detail.

On the other hand, many Bayesian analyses naturally give rise to the occurrence of poly-t densities as defined in Drèze (1977). Efficient methods for their analysis were developed, e.g. in Richard and Tompa (1980), and are now incorporated in a standard software package, called the Bayesian regression program (BRP), as described in Bauwens et al. (1981). Unfortunately, these procedures, that employ both analytical and numerical techniques, can only cover a rather limited class of models if we are looking for full information marginal results, as explained in Drèze and Richard (1983).

In spite of the considerable progress achieved in numerical methods and the explosion in computational facilities, which has recently made these methods more attractive for empirical work, exact analytical results are, of course, more satisfactory from a theoretical point of view, and typically much cheaper in computing time.
Therefore, the present paper is an attempt to clearly define the boundaries of the analytical domain in Bayesian analysis. As we already suspect

this to be quite limited, we focus upon a very simple type of model, namely systems of seemingly unrelated regression equations under noninformative[1] stochastic prior notions.

We do, however, introduce a rather general type of restrictions on the coefficients and try to ascertain which class of restrictions can be accommodated by fully analytical methods.

As we feel it clarifies the exposition and reduces the complexity of the formulas, we use a recursive transformation of the model, which is introduced in the next section. Section 3 briefly discusses the reason why analytical methods do not work in general cases and describes some of the standard ways of dealing with the problem numerically, as described e.g. in Drèze and Richard (1983). Some analytical results and their limitations are presented in Subsection 3.3, together with a few suggestions for the analysis of those cases that are not covered by our analytical formulas. The last section groups some conclusions and topics for further research, whereas Appendices A, B, and C contain some technical details, and Appendix D tabulates the various assumptions concerning rank properties that are made in the course of this paper.

2. The SURE Model

2.1. Description

The main results of this paper will pertain to the standard system of seemingly unrelated regression equations (SURE) as introduced in Zellner (1962).

We consider a sequential conditional model describing a vector of n endogenous variables $y_{(t)}$ at time t, given an available information set I_t which contains weakly exogenous variables and lagged values of both endogenous and exogenous variables. Let us assume we can group all variables in I_t that are relevant for the description of $y_{(t)}$ into a vector of finite dimension $x_{(t)} \in \mathbb{R}^m$. We further assume normality with linear conditional expectations and a constant covariance matrix over time, leading to the data density

$$D(y_{(t)}|I_t) = f_N^n(y_{(t)}|\Pi' x_{(t)}, V), \quad t: 1 \to T, \tag{2.1}$$

where we have adopted the notation for density functions found in Appendix A of Drèze and Richard (1983). In line with most applications using SURE systems, we let V be any unrestricted symmetric positive definite (PDS) matrix, but put restrictions on the $m \times n$ matrix of coefficients Π.
In particular, we consider Π as a matrix function of a vector of unrestricted parameters $\alpha \in \mathbb{R}^{\ell}$, and will focus upon linear restrictions of the type

$$\pi = \text{vec}\ \Pi = S\alpha, \tag{2.2}$$

where S is a $mn \times \ell$ matrix of known constants and of full column rank. Extensive discussion of the general nature of (2.2), as well as its form in several special cases can be found in Richard and Steel (1988). Suffice it to say here that it allows for restrictions across equations as well as within equations and it is certainly more general than the very restrictive matrix form

$$\Pi = BAC, \tag{2.3}$$

where A is a $p \times q$ matrix of unrestricted coefficients and B and C are known matrices of respectively full column and row rank. The matrix form (2.3) is the only one which is compatible with a natural-conjugate prior structure, which is therefore of very limited use in SURE models. In fact, in a natural-conjugate (NC) framework, the prior covariance matrix of Π given V is restricted to be of the Kronecker form $V \otimes M_0^+$, where the superscript $^+$ denotes the Moore-Penrose inverse and the relative precision matrix M_0 can be singular. For compatibility with (2.3) we require that the restrictions expressed by C reflect a redundancy (e.g. through identities) in $y_{(t)}$, so that instead $C^{+\prime} y_{(t)}$ has a nondegenerate distribution with PDS covariance matrix Σ. We then have a proper NC prior structure on A, the unrestricted coefficients of Π, and Σ, given by

$$D(A|\Sigma) = f_{MN}^{p \times q}(A|A_0, \Sigma \otimes N_0^{-1}) \tag{2.4}$$

$$D(\Sigma) = f_{IW}^{q}(\Sigma|\Sigma_0, \nu_0), \tag{2.5}$$

from which we can deduce

$$D(\Pi|V) = f_{MN}^{m\times n}(\Pi|BA_0C, V \otimes BN_0^{-1}B') \tag{2.6}$$

$$D(V) = f_{IW}^{n}(V|C'\Sigma_0C, \nu_0 - q + n), \tag{2.7}$$

both singular distributions, where $V = C'\Sigma C$ is no longer PDS, but positive semidefinite symmetric (PSDS).

Usually, we shall assume that the model is defined in terms of a nondegenerate distribution on $y_{(t)}$ as in (2.1) with V a PDS matrix, possibly after redefining $y_{(t)}$ so as to take redundancies into account from the start. This would imply $C = I_n$ in (2.3) and would clearly put the main problem of the NC approach into focus, namely the fact that the coefficients of each equation are restricted in exactly the same way by $BN_0^{-1}B'$ in (2.6). This implies that e.g. we can't exclude a variable in one equation without doing the same for every other equation, clearly a very undesirable feature in practice, except for reduced form time series modelling as e.g. in the unrestricted vector autoregressions approach of Sims (1980). Therefore, we shall focus upon the representation of the restrictions in (2.2), using a <u>vector</u> of free parameters, instead of a <u>matrix</u>.

If we wish to impose the more general linear restrictions in (2.2), we need to use e.g. extended natural-conjugate [see Drèze and Richard (1983)] or recursive extended natural-conjugate prior structures [see Richard and Steel (1988)] if we possess additional stochastic prior information. If the latter is not the case, we can naturally use the simple prior structure

$$D(\alpha,V) \propto |V|^{-\frac{1}{2}\nu_0}, \tag{2.8}$$

in combination with (2.2). In this paper we shall analyse the posterior structure of α and V under a noninformative prior of the type in (2.8).

In order to write the likelihood function of the model as it was defined here, let us introduce the obvious matrix notation for all T observations:

$$Y = (y_1 \ \dots \ y_i \ \dots \ y_n) = \begin{bmatrix} y'_{(1)} \\ \vdots \\ y'_{(t)} \\ \vdots \\ y'_{(T)} \end{bmatrix} \tag{2.9}$$

$$X = (x_1 \ \dots \ x_j \ \dots \ x_m) = \begin{bmatrix} x'_{(1)} \\ \vdots \\ x'_{(t)} \\ \vdots \\ x'_{(T)} \end{bmatrix}, \tag{2.10}$$

assumed to be of full column rank, and let us partition Π as

$$\Pi = (\pi_1 \ \dots \ \pi_i \ \dots \ \pi_n), \tag{2.11}$$

and, finally, define $y = \text{vec } Y$. Then we can write the relevant information contained in the data as

$$\mathcal{L}(\alpha, V; Y, X) \propto f_{MN}^{T \times n}(Y | X\Pi, V \otimes I_T) \tag{2.12}$$

with $\pi = S\alpha$, or, alternatively, as

$$\mathcal{L}(\alpha, V; Y, X) \propto f_{N}^{Tn}(y | (I_n \otimes X)S\alpha, V \otimes I_T), \tag{2.13}$$

where we remark that (2.12) and (2.13) are only partial likelihoods under the hypothesis of weak exogeneity of $x_{(t)}$ [for definitions of the exogeneity concepts used here we refer to Engle et al. (1983)]. This means they contain all the information relevant to the parameters of our model in (2.1) and (2.2), and are thus sufficient for our analysis. The entire joint data density can in fact only be written as the expressions in (2.12) and (2.13) under the more demanding hypothesis of strong exogeneity, where feedback effects are excluded.

2.2. A Recursive Transformation

From classical econometrics we know that a fully recursive system simplifies estimation considerably. More in particular, in such a triangular

structure with independent error terms full maximum likelihood procedures boil down to applying OLS equation by equation.

In order to investigate whether such simplifications can be used to our advantage in a Bayesian framework, let us define the following transformation of our SURE model in (2.1) and (2.2) into a recursive system.

We introduce the n × n lower triangular matrix Λ defined as the matrix that diagonalizes V, a PDS matrix, and partitioned as

$$\Lambda = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\lambda_{21} & 1 & & \vdots \\ -\lambda_{31} & -\lambda_{32} & 1 & \vdots \\ \vdots & & & 0 \\ -\lambda_{n1} & -\lambda_{n2} & \cdots & 1 \end{bmatrix}, \tag{2.14}$$

leading to

$$\Lambda V \Lambda' = D_{\omega}, \tag{2.15}$$

a diagonal matrix with $\omega_i^2 > 0$ (i: 1 → n) on the main diagonal, reflecting the independent disturbances of the recursive form equations. As $|\Lambda| = 1$, we can rewrite the partial likelihood function in (2.12) in terms of the recursive form as

$$\mathcal{L}(\alpha, \{\lambda_i, \omega_i^2\}; Y, X) \propto f_{MN}^{T\times n}(Y\Lambda' | XP, D_{\omega} \otimes I_T), \tag{2.16}$$

where $\lambda_i = (\lambda_{i1} \cdots \lambda_{i,i-1})'$, and $P = \Pi\Lambda'$ is now a matrix function of α and $\{\lambda_i\}$, given through

$$\text{vec } P = p = (\Lambda \otimes I_m)S\alpha. \tag{2.17}$$

Let us now denote by an index i in parentheses those quantities that refer to the first i equations, and introduce a recursion[2] as follows

$$\pi_{(i)} = \begin{bmatrix} \pi_{(i-1)} \\ \pi_i \end{bmatrix} \qquad \alpha_{(i)} = \begin{bmatrix} \alpha_{(i-1)} \\ \alpha_i \end{bmatrix}, \tag{2.18}$$

and $$S_{(i)} = \begin{bmatrix} S_{(i-1)} & 0 \\ S_{i(i-1)} & S_{ii} \end{bmatrix}, \tag{2.19}$$

as S can always be reduced to a lower block-triangular matrix of full column rank [see Richard and Steel (1988)], where each block S_{ii} is of full column rank ℓ_i. This ℓ_i gives the number of additional elements in α due to equation i, here defined as the dimension of α_i. Finally, introduce the notation

$$Y_i = (Y_{i-1} \; y_i) \tag{2.20}$$

to recursively split up the data matrix Y into a part referring to the (i-1) previous equations (Y_{i-1}) and a part referring to the present equation i [as in (2.9)].

Rewriting (2.16) equation by equation, it is then easily seen to factorize into

$$\mathcal{L}(\alpha, \{\lambda_i, \omega_i^2\}; Y, X) \propto \prod_{i=1}^{n} f_N^T(y_i | Y_{i-1}\lambda_i + X p_i, \omega_i^2 I_T) \tag{2.21}$$

where $P = (p_1 \ldots p_n)$, and where imposing the restrictions in (2.17) implies that

$$\alpha = S^+(\Lambda^{-1} \otimes I_m)p \tag{2.22}$$

under the condition that $SS^+\pi = \pi$, i.e. that a solution for α exists from (2.2). Using our recursion formulas, we can deduce from (2.22) a solution for α_i, the unrestricted coefficients of equation i, as follows:

$$\alpha_i = S_{ii}^+(p_i + [(\lambda_i' \otimes I_m)S_{(i-1)} - S_{i(i-1)}]\alpha_{(i-1)}), \tag{2.23}$$

which implies that, given λ_i and p_i at each i, we can recursively solve for the vector α. Of course, p_i is still implicitly restricted, which has to be incorporated in the analysis. One way of expliciting these restrictions would be to focus upon an unrestricted ℓ_i dimensional vector, say

p_i^α, which would, given the unrestricted λ_i and $\alpha_{(i-1)}$, exactly determine α_i, or to find a p_i^α such that

$$\alpha_i = p_i^\alpha + S_{ii}^+[(\lambda_i' \otimes I_m)S_{(i-1)} - S_{i(i-1)}]\alpha_{(i-1)}, \tag{2.24}$$

and, at the same time, assuring that this p_i^α enters our analysis in a very natural way, e.g. by equating Xp_i in our likelihood (2.21) to $X_i p_i^\alpha$, where X_i is defined as XS_{ii}, which we assume to be of full column rank ℓ_i. We then need that both

$$S_{ii}^+ p_i = p_i^\alpha \tag{2.25}$$

and

$$Xp_i = XS_{ii}p_i^\alpha \tag{2.26}$$

hold simultaneously. This obviously holds if S_{ii} is square and non-singular, which corresponds to the unrestricted SURE case, and also if $S_{ii} = c_{ii}B$ with B of full column rank by rewriting XP as $(XB)(AC\Lambda')$ corresponding to the matrix form of restrictions in (2.3).

A more interesting case where (2.25) and (2.26) hold simultaneously makes full use of the recursive features introduced here. If we denote all X variables used in the first i equations by $X_{(i)}$, defined in such a way as to have full column rank, then a sufficient condition for (2.25) and (2.26) to hold is

$$P_i^X \; X_{(i)} = X_{(i)}, \tag{2.27}$$

where $P_i^X = X_i X_i^+ = X_i(X_i'X_i)^{-1}X_i'$.

A proof is given in Appendix A.

This is satisfied if all variables in the previous (i-1) equations also appear (with unrestricted coefficients) in the i-th equation, as in the limited information (LI) case of Lubrano and Richard (1981), where it leads to substantial simplifications in a two-equation instrumental varia-

bles analysis. Let us, by analogy, call this the "recursive limited information" (RLI) case, which certainly seems more important, at this stage, for expository and reference purposes than for serious applications.

In the three cases described above we have from (2.24) a unique solution for α_i, given $\alpha_{(i-1)}$, once we know the (ℓ_i+i-1) dimensional vector $(p_i^{\alpha'} \lambda_i')'$, which we redefine as

$$\beta_i = \begin{bmatrix} p_i^{\alpha} \\ \lambda_i \end{bmatrix}, \tag{2.28}$$

and reasoning in terms of β_i automatically imposes the restrictions in their general vector form (2.2).
In addition, from (2.26) we notice that we can easily rewrite the partial likelihood function (2.21) as

$$\mathcal{L}(\alpha, \{\lambda_i, \omega_i^2\}; Y, X) \propto \prod_{i=1}^{n} f_N^T(y_i|W_i\beta_i, \omega_i^2 I_T), \tag{2.29}$$

where we have implicitly defined

$$W_i = (X_i \; Y_{i-1}). \tag{2.30}$$

3. Posterior Analysis under a Noninformative Prior Structure

3.1. The Problem Stated

The mere fact that the restrictions as expressed in (2.2) are generally not compatible with a NC analysis already indicates that the posterior densities will not belong to those families that we can easily examine analytically. No matter how simple the structure of the stochastic prior information is, the exact prior restrictions will complicate matters considerably.

Let us explain this more in detail by referring to the family of diffuse prior densities described in (2.8)

$$D(\alpha,V) \propto |V|^{-\frac{1}{2}\nu_0},$$

where the choice of ν_0 is open to some difference of opinion. Drèze (1976) suggests that if our prior information is noninformative, it seems reasonable to impose invariance with respect to transformations between the reduced and structural forms of the model, and therefore suggests $\nu_0 = m+n+1$. Although we only treat the SURE model here, we feel that this argument might be of general validity. In contrast, Zellner (1971) uses Jeffreys' invariance principles to motivate choosing $\nu_0 = n+1$, Stewart (1987) derives $\nu_0 = 2$ from a data-relocated likelihood function, conform to Jeffreys' rule under a certain prior independence assumption, and Malinvaud (1981) advocates the use of $\nu_0 = 2n$ in order to retain the same marginal results per equation as in the single equation case. Bauwens (1984) remarks that the choice of ν_0 will affect the existence of posterior moments for α. A formal theory for obtaining a diffuse "reference prior" based on expected information is provided in Bernardo (1979). The resulting reference distributions are invariant both under one-to-one transformations of the parameter space and under reductions to sufficient statistics, but, unfortunately, involve integration over the sample space and thus violate the likelihood principle, as explained in the discussion following Bernardo (1979). Here we shall limit ourselves to the simple prior structure in (2.8), which seems somewhat less controversial and certainly easier to use in practice, although it has less convincing invariance properties.

Only under asymptotic normality of the posterior distribution, which will not generally hold in view of (2.2), and provided there are no nuisance parameters, does Bernando's approach lead to the type of improper prior densities used here.

Drèze and Richard (1983) and Van Dijk (1985) address the latter issue in more detail. A sufficient condition for the existence of posterior moments of order r is now given as

$$\nu_0 > 2n + \mu + r - T \tag{3.1}$$

with $\mu = \sup\{\ell_i;\ i: 1 \rightarrow n\}$, based on Lemma 6.6 of Drèze and Richard (1983) applied to diffuse priors as in (2.8) on SURE models. Provided $\mu = m$, i.e.

at least one equation is exactly identified, we get the same degrees of freedom condition as given in Van Dijk (1985) except for the last term T, which represents the advantage of SURE over simultaneous equation models (SEM) in terms of existence of moments.

The main problem we face in the analysis of the posterior structure, however, can be seen to depend mainly on the presence of general linear restrictions in vec form as in (2.2). Combining the (partial) likelihood function in (2.12) with the simple diffuse prior (2.8), we obtain for the posterior density on V given α a well-known form

$$D(V|\alpha, Y, X) = f_{IW}^{n}(V|V_*, \nu_*-n-1) \tag{3.2}$$

with

$$V_* = (Y - X\Pi)'(Y-X\Pi)$$
$$\nu_* = \nu_0 + T,$$

where we have implicitly assumed that V_* is PDS and $\nu_0 > 2n - T$, and we get a posterior structure for α

$$D(\alpha|Y, X) \propto |(Y-X\Pi)'(Y-X\Pi)|^{-\frac{1}{2}(\nu_*-n-1)}, \tag{3.3}$$

which looks like a matrix-t kernel, but has to be treated conditionally on the restrictions on Π. Now, again, it becomes obvious that under the matrix form (2.3) the properties of α are known analytically and we are still in a NC framework. Then, indeed, we have a matrix-t kernel in A, expressing the restrictions on Y by $Y = \tilde{Y}C$, and reasoning in terms of $\tilde{Y}$ and XB.

If, however, we wish to restrict Π, in a more useful way by imposing (2.2), we lose the nice analytical properties of a matrix-t kernel and the problem then seems to require numerical procedures. In view of the potentially large dimension of α, one could suggest the use of Monte Carlo procedures, as described in Kloek and Van Dijk (1978) and Hendry (1984). More in particular, the method of Importance Sampling [see e.g. Hammersley and Handscomb (1979)] seems appropriate as it has previously led to satisfactory results in similar models. In this context, we can refer to Bauwens (1984) and Richard and Steel (1988).

3.2. Analysis in Terms of the Original Parameterisation

If we wish to remain in the parameterisation (α,Σ) of the untransformed SURE model, we can attempt to isolate the posterior density of the coefficients of one equation α_i, given all other $\ell-\ell_i$ elements of α, along the lines of Drèze and Richard (1983) and Bauwens (1984).

We start out from the restricted matrix-t form for the posterior density of α in (3.3), and explicit the restrictions by using (2.18) and (2.19) to replace π_i by an expression in α_i, given $\alpha_{(i-1)}$ of the previous equations:

$$\pi_i = S_{ii}\alpha_i + S_{i(i-1)}\alpha_{(i-1)}. \tag{3.4}$$

As V_* is a PDS matrix composed of the elements $(y_i-X\pi_i)'(y_j-X\pi_j)$, we can easily split up the determinant into a part that depends on α_i given $\alpha_{(i-1)}$ and a marginal part that depends only on $\alpha_{(i-1)}$, letting i run from n to 2 in a recursive way. We now partition $\Pi_i = (\Pi_{i-1}\ \pi_i)$ as in (2.20), and start at i = n, while we first rewrite V_* as a quadratic from in Π and base our factorisation upon the matrix

$$\begin{bmatrix} I_{n-1} & 0 \\ -(y_n'M_xY_{n-1}) & (Y_{n-1}'M_xY_{n-1})^+ \quad 1 \end{bmatrix},$$

which implies the reverse factorisation of the one used in Lemma B.1 of Drèze and Richard (1983).
Using the PSDS nature of $Y'M_xY$, where we have defined

$$M_x = I_T - X(X'X)^{-1}X',$$

and having assumed X'X to be nonsingular in (2.10), we can write

$$|(Y-X\Pi)'(Y-X\Pi)|$$

$$= |Y'M_xY + (\Pi-\hat{\Pi})'X'X\,(\Pi-\hat{\Pi})|$$

$$= |Y'_{n-1}M_xY_{n-1} + (\Pi_{n-1}-\hat{\Pi}_{n-1})'X'X\,(\Pi_{n-1}-\hat{\Pi}_{n-1})|$$

$$\left[q_n^* + (\pi_n-\pi_n^*)'\,H_n^*\,(\pi_n-\pi_n^*)\right], \tag{3.5}$$

with

$$\hat{\Pi} = (X'X)^{-1}X'Y = (\hat{\Pi}_{n-1}\hat{\pi}_n),$$

partitioned conformably with Π, and

$$H_n^* = X'X - X'X\,(\Pi_{n-1}-\hat{\Pi}_{n-1})\,[Y'_{n-1}M_xY_{n-1} + (\Pi_{n-1}-\hat{\Pi}_{n-1})'X'X$$

$$(\Pi_{n-1}-\hat{\Pi}_{n-1})]^{-1}\,(\Pi_{n-1}-\hat{\Pi}_{n-1})'X'X,$$

$$q_n^* = y_n'M_xy_n - y_n'M_xY_{n-1}\,(Y'_{n-1}M_xY_{n-1})^{+}\,Y'_{n-1}M_xy_n,$$

$$\pi_n^* = \hat{\pi}_n + (\Pi_{n-1}-\hat{\Pi}_{n-1})\,(Y'_{n-1}M_xY_{n-1})^{+}\,Y'_{n-1}M_xy_n.$$

Next, we substitute (3.4) to examine the posterior density for α_n given $\alpha_{(n-1)}$, corresponding to the last factor in (3.5), and we obtain the kernel of a Student t density

$$D\,(\alpha_n|\alpha_{(n-1)},Y,X)$$

$$= f_t^{\ell_n}\,(\alpha_n|\alpha_n^*, s_n^{*-1}\,(S'_{nn}H_n^*S_{nn}),\ \nu_*-n-\ell_n-1), \tag{3.6}$$

with

$$\alpha_n^* = (S'_{nn} H_n^* S_{nn})^{-1} S'_{nn} H_n^* (\pi_n^* - S_{n(n-1)} \alpha_{(n-1)}),$$

$$s_n^* = q_n^* + (\pi_n^* - S_{n(n-1)} \alpha_{(n-1)})' [H_n^* - H_n^* S_{nn} (S'_{nn} H_n^* S_{nn})^{-1} S'_{nn} H_n^*]$$

$$(\pi_n^* - S_{n(n-1)} \alpha_{(n-1)}),$$

provided $|S'_{nn} H_n^* S_{nn}|$ and s_n^* are both nonzero and $\nu_0 > n + \ell_n - T + 1$. This gives us a very simple form for the conditional posterior density of α_n, but the crucial point is that the reciprocal of the integrating constant of (3.6), which is given by $(s_n^*)^{-\frac{1}{2}(\nu_* - n - \ell_n - 1)} |S'_{nn} H_n^* S_{nn}|^{-\frac{1}{2}}$, depends on $\alpha_{(n-1)}$ as is obvious from the expressions above. However, in the RLI case the last equation incorporates all X variables, and S_{nn} is therefore square and, as it is, by assumption, of full column rank, it is also invertible.

This implies $s_n^* = q_n^*$, which does not depend on $\alpha_{(n-1)}$, and $|S'_{nn} H_n^* S_{nn}| \propto |H_n^*|$, which is shown in Appendix B to be proportional to the inverse of the first factor in (3.5), a quantity that has to be nonzero since V_* was assumed a PDS matrix. This means that we are left with a nondegenerate posterior density

$$D(\alpha_{(n-1)} | Y, X)$$

$$\propto |Y'_{n-1} M_x Y_{n-1} + (\Pi_{n-1} - \hat{\Pi}_{n-1})' X' X (\Pi_{n-1} - \hat{\Pi}_{n-1})|^{-\frac{1}{2}(\nu_* - n - 2)}.$$

At this stage, we can reduce the X space to only those variables used in the first (n-1) equations, denoted by $X_{(n-1)}$, by first partitioning

$$X = (X_{(n-1)} \; \tilde{X}_n), \text{ and correspondingly}$$

$$\Pi_{n-1} = \begin{bmatrix} \Pi_{(n-1)} \\ 0 \end{bmatrix},$$

and

$$\hat{\Pi}_{n-1} = \begin{bmatrix} \hat{\Pi}_{(n-1)} \\ \hat{\pi}_{n-1} \end{bmatrix},$$

possibly after rearranging the remaining equations. We then use the matrix

$$\begin{bmatrix} I_{\ell_{n-1}} & 0 \\ -\tilde{X}_n' X_{(n-1)} [X_{(n-1)}' X_{(n-1)}]^{-1} & I_{m-\ell_{n-1}} \end{bmatrix}$$

as in Drèze and Richard (1983, p. 593) in order to rewrite the posterior in terms of $\Pi_{(n-1)}$:

$$D(\alpha_{(n-1)} | Y, X)$$

$$\propto |Q^*_{(n-1)} + (\Pi_{(n-1)} - \Pi^*_{(n-1)})' X'_{(n-1)} X_{(n-1)} (\Pi_{(n-1)} - \Pi^*_{(n-1)})|^{-\frac{1}{2}(\nu_* - n - 2)}, \tag{3.7}$$

with

$$Q^*_{(n-1)} = Y'_{n-1} M^X_{(n-1)} Y_{n-1},$$

and

$$\Pi^*_{(n-1)} = (X'_{(n-1)} X_{(n-1)})^{-1} X'_{(n-1)} Y_{n-1},$$

where we have defined

$$M^X_{(n-1)} = I_T - X_{(n-1)} (X'_{(n-1)} X_{(n-1)})^{-1} X'_{(n-1)}.$$

In the RLI framework, equation (n-1) will, in its turn, contain all variables in $X_{(n-1)}$, which means we can deduce a Student kernel from (3.7) for α_{n-1} without contaminating the functional form of the posterior for the remaining (n-2) equations by the integrating constant. Moving through the system in this recursive fashion we can, thus, factorize the conditional matrix-t kernel in (3.3) into a product of conditional t densities for α_i

given $\alpha_{(i-1)}$, i=1,...n, provided the model has an RLI structure. Unfortunately, in the general case the presence of the integrating constants prevents such a convenient analytical solution. Only in the special case of a two-equation SURE system, we retain an analytical expression for the marginal posterior density of e.g. α_1, which can under certain conditions be written as a 2-1 poly-t density [see Drèze and Morales (1976)].

Drèze and Richard (1983) propose to use a Normal approximation to the posterior density for α in (3.3) along the lines of Morales (1971) and then draw values for α_i conditional on the values obtained from the Normal approximation for the rest of α, using Student t densities as in (3.6). This way of conducting the Monte Carlo drawings is labeled the "poly-t drawn conditions" (PTDC) approach in Bauwens (1984), who proposes several alternative importance functions as well. It remains, however, unclear which importance function should be chosen in any particular application and how closely the one opted for will approximate the actual posterior structure, whereas we know that if the far tails of the posterior density are not dominated by the importance function our results can be very deceptive. Including stochastic prior information as well, Richard and Steel (1988) find that the posterior density is hard to capture by such conditional Student t or poly-t importance functions if we reason in terms of α. Therefore, the next subsection will focus on the possibility of obtaining analytical results using the recursive reparameterisation introduced in subsection 2.2, as this seems to allow for a slightly less cumbersome presentation than the one given in the present subsection and is more in line with the recursive treatment of the model.

3.3. Analytical Results Using the Recursive Model

In this subsection we shall limit ourselves to the three particular cases (unrestricted, matrix restricted, and RLI models) discussed in subsection 2.2 under which the partial likelihood function takes the very simple form (2.29), in an effort to explore the boundaries of the analytical domain when treating SURE models under noninformative stochastic prior information. This very simple case will then hopefully provide us with some suggestions for the analysis of more complicated models, which will be the

object of future research. The use of the recursive parameterisation is not entirely crucial, but we feel it enhances clarity of the exposition and it somewhat facilitates the analysis of exogeneity along the lines of Engle et al. (1983).

In all three particular cases examined here, the condition $P_i^x X_{(i)} = X_{(i)}$ holds and the partial likelihood function takes the simple form (2.29) in terms of unrestricted parameter vectors β_i. Therefore, we shall perform the posterior analysis in terms of this recursive parameterisation, and later use (2.24) in order to examine the properties of α. As we do not wish to express informative prior notions about α in the context of the present paper, we are not required to transform informative prior densities on α into their counterparts for $\{\beta_i\}$, which greatly simplifies the procedure.

Let us first rewrite the diffuse prior (2.8) in terms of the parameters of the recursive model. From (2.14) and (2.15) we easily obtain

$$|V| = |\Lambda^{-1} D_\omega \Lambda'^{-1}| = |D_\omega| = \prod_{i=1}^{n} (\omega_i^2), \tag{3.8}$$

so that the relevant prior density becomes

$$D(\{\beta_i, \omega_i^2\}) \propto \prod_{i=1}^{n} (\omega_i^2)^{-\frac{1}{2}\nu_0} . J, \tag{3.9}$$

where J is the Jacobian of the parameter transformation from (α, V) to $(\{\beta_i, \omega_i^2\})$ or $(\{p_i^\alpha, \lambda_i, \omega_i^2\})$. As a result of the recursive structure described in (2.24), this Jacobian will amount to

$$J = \prod_{i=1}^{n-1} (\omega_i^2)^{n-i} \tag{3.10}$$

(for a proof, see Appendix C), so that (3.9) can be written as

$$D(\{\beta_i, \omega_i^2\}) \propto \prod_{i=1}^{n} (\omega_i^2)^{-\frac{1}{2}(\nu_0 - 2(n-i))} \tag{3.11}$$

Combining this with the likelihood function (2.29) gives for each equation i the following simple posterior structure

$$D(\omega_i^2|\beta_i, Y, X) = f_{i\chi}(\omega_i^2|\omega_{i*}^2, \nu_* - 2(n-i+1)) \tag{3.12}$$

with

$$\omega_{i*}^2 = (y_i - W_i\beta_i)'(y_i - W_i\beta_i), \tag{3.13}$$

and

$$D(\beta_i|Y, X) = f_t^{\ell_i+i-1}(\beta_i|\hat{\beta}_i, (y_i'M_i^W y_i)^{-1}W_i'W_i, \nu_*-\ell_i-2n+i-1) \tag{3.14}$$

with

$$\hat{\beta}_i = (W_i'W_i)^{-1}W_i'y_i \tag{3.15}$$

$$M_i^W = I - W_i(W_i'W_i)^{-1}W_i', \tag{3.16}$$

where W_i is assumed to have full column rank (i.e. we add the assumption that Y_{i-1} is of full column rank and has no columns that are linear combinations of those of X_i). Further, we assume that there exists no linear dependence between y_i and the columns in W_i, or that $(X_i\ Y_i)$ has full column rank, which implies $y_i'M_i^W y_i > 0$, and we require $\nu_0 > 2n + \mu - T$. Remark that now, due to the fact that we can reason in terms of β_i, given our specific condition (2.27), the integrating constants from (3.14) do not involve any parameters. The joint posterior density for all vectors β_i (i: 1 → n) will thus just be given by the product of Student t densities as in (3.14). Furthermore, the Jacobian of the transformation from $\beta_i = \begin{bmatrix} p_i^\alpha \\ \lambda_i \end{bmatrix}$ to $\begin{bmatrix} \alpha_i \\ \lambda_i \end{bmatrix}$ is unity in view of (2.24). The density (3.14) can therefore simply be interpreted as the posterior density of (α_i, λ_i) given $\alpha_{(i-1)}$, provided the necessary transformations in (2.24) are carried out. Remark that we can always sequentially partition the entire posterior density for α and Λ into

$$D(\alpha, \Lambda|Y, X) = \prod_{i=1}^{n} D(\alpha_i, \lambda_i|\alpha_{(i-1)}, Y, X),$$

as can be seen directly by considering the posterior structure in (3.2) and (3.3) and by virtue of the recursive decomposition properties of Inverted-Wishart densities, as documented e.g. in Drèze and Richard (1983).

What is particular about the special cases treated here, however, is that the actual form of $D(\alpha_i, \lambda_i|\alpha_{(i-1)}, Y, X)$ is known analytically to be Student t for each equation i. This is certainly not the case if we reason in terms of (p_i, λ_i) since p_i is implicitly restricted, leading to a conditional Student t which has no known analytical properties. This is exactly the same problem as described in the previous subsection, only now expressed in terms of the recursive parameterisation.

If we are prepared to conduct a Monte Carlo analysis, we could ignore the problems that occur in the general case in order to construct an importance function as was examined in Richard and Steel (1988) in the context of informative stochastic prior notions. One possibility would be to draw values for $(S_{ii}^{+}p_i, \lambda_i)$ from simple Student t densities like (3.14) and use (2.23) to construct drawings for (α_i, λ_i) keeping $\alpha_{(i-1)}$ fixed at the value obtained by drawing from the coefficients for previous equations. This entails that we require n drawings from Student t densities in a recursive ordering so as to obtain one full drawing for α and Λ. Using this recursive Monte Carlo procedure we can evaluate any function of (α, Λ) that is of interest to us, e.g. moments of α or Λ, but also more complicated functions like moments of certain elements in Λ^{-1} for examining weak exogeneity in conditional SURE models as in Steel (1987), or moments of other nonlinear transformations, fractiles etc.
Of course, the outcome of such a Monte Carlo procedure would generally, at least to some extent, depend upon the ordering of the equations that we happen to choose, and its merits in various applications can only be ascertained in an empirical way.

Let us, therefore, concentrate here on the possibility of obtaining analytical results as these are, by their very nature, much more interesting theoretically and do not suffer from problems of specificity and lack of numerical accuracy [see Hendry (1984)] that can, in some cases, produce very misleading inferences. We thus limit ourselves to our three special

cases and remark that from (3.14) the posterior moments of β_i exist up to order r for all equations if

$$\nu_0 > 2n + \mu + r - T,$$

which is exactly the same as the sufficient condition (3.1) for the existence of posterior moments of α. As the dependence on previous equations comes in solely through p_i^{α}, the posterior density for λ_i is independent of $\alpha_{(i-1)}$, since

$$D(\beta_i|Y,X) = D(\lambda_i|Y,X).D(p_i^{\alpha}|\lambda_i,Y,X) = D(\lambda_i|Y,X).D(\alpha_i|\lambda_i,\alpha_{(i-1)},Y,X),$$

where the marginal posterior density of λ_i is simply of the Student form:

$$D(\lambda_i|Y,X) = f_t^{i-1}(\lambda_i|\hat{\lambda}_i,(y_i'M_i^W y_i)^{-1}Y_{i-1}'M_i^X Y_{i-1},\nu_*-\ell_i-2n+i-1), \quad (3.17)$$

and we define $\hat{\beta}_i = \begin{bmatrix} \hat{p}_i^{\alpha} \\ \hat{\lambda}_i \end{bmatrix}$, while M_i^X is defined analogously to M_i^W in (3.16) and is thus given by $M_i^X = I - P_i^X$.

Using this independence property, we can evaluate the posterior mean of α in a recursive way, following (2.24),

$$E(\alpha_i|Y,X) = \hat{p}_i^{\alpha} + S_{ii}^{+}[(\hat{\lambda}_i' \otimes I_m)S_{(i-1)}$$

$$- S_{i(i-1)}]E(\alpha_{(i-1)}|Y,X), \quad \forall i: 1 \to n. \quad (3.18)$$

Remark that it is easily verified from (3.6) that the analysis in terms of the original parameterisation produces exactly the same results, provided, of course, the RLI structure is imposed there as well. Appendix B provides some further details. At the cost of somewhat more cumbersome calculations, we can deduce the following recursive analytical expression for the posterior variance of α_i (i: 1 → n)

$$\mathrm{Var}(\alpha_i|Y,X) = \frac{1}{\nu_* - \ell_i - 2n + i - 3}\, y_i' M_i^W y_i \{(X_i'X_i)^{-1} +$$

$$X_i^+ [\sum_{k,\ell=1}^{i-1} (Y_{i-1}' M_i^X Y_{i-1})^{k\ell} (y_k y_\ell' - XE(\pi_k|Y,X)y_\ell' - y_k E(\pi_\ell'|Y,X)X'$$

$$+ XE(\pi_k\pi_\ell'|Y,X)X')]X_i^{+\prime}\} + S_{ii}^+ (S_{i(i-1)} - (\hat{\lambda}_i' \otimes I_m)S_{(i-1)})$$

$$\mathrm{Var}(\alpha_{(i-1)}|Y,X)(S_{i(i-1)} - (\hat{\lambda}_i' \otimes I_m)S_{(i-1)})' S_{ii}^{+\prime}, \qquad (3.19)$$

where, with $S_k = (S_{k(k-1)} : S_{kk})$, we can use

$$E(\pi_k|Y,X) = S_k E(\alpha_{(k)}|Y,X)$$

and

$$E(\pi_k\pi_\ell'|Y,X) = S_k E(\alpha_{(k)}\alpha_{(\ell)}'|Y,X)S_\ell',$$

and where superscript $k\ell$ refers to the corresponding element of the inverse. A considerably simpler formula can be found for the covariance between α_i and α_j, $i > j$, namely

$$\mathrm{Cov}(\alpha_i,\alpha_j'|Y,X) = S_{ii}^+ \left[(\hat{\lambda}_i' \otimes I_m)S_{(i-1)} - S_{i(i-1)}\right] \mathrm{Cov}(\alpha_{(i-1)}, \alpha_j'|Y,X). \qquad (3.20)$$

With the help of (3.19) and (3.20) we can construct the posterior covariance structure of α in a fully analytical way.

If our interest centres upon λ_i or certain transformations of Λ, we can use the posterior in (3.17). A particularly important quantity for evaluating weak exogeneity, for example, is the matrix Λ^{-1}. Denoting its elements by λ^{ij} and realizing that Λ^{-1} will also be of lower triangular form with unitary elements on the diagonal we can easily calculate the (i,j)th element of $\Lambda\Lambda^{-1}$ for $i > j$, which, of course, has to be equal to zero. This leads to the following recursive expression for λ^{ij} $(i > j)$:

$$\lambda^{ij} = \sum_{m=1}^{i-1} \lambda_{im}\lambda^{mj}, \qquad (3.21)$$

which only involves products of elements belonging to different equations, and those are independent from (3.14). In particular, we obtain

$$E(\lambda^{ij}|Y,X) = \sum_{m=1}^{i-1} \hat{\lambda}_{im} E(\lambda^{mj}|Y,X), \tag{3.22}$$

which can be evaluated recursively, bearing in mind that $\lambda^{jj} = 1$ and $\lambda^{mj} = 0 \ \forall m < j$, and where $\hat{\lambda}_{im}$ is the m-th element of $\hat{\lambda}_i$. The same goes for the variance of λ^{ij}; in particular, if we denote by ${}_{(j-1)}\lambda_i$ the vector λ_i without its first (j-1) elements (i > j), and by $\lambda^{\cdot j}_{(n-i+1)}$ the free elements on the j-th column of Λ^{-1} without the last (n-i+1) elements, or, equivalently, obtained by just taking the (i-j-1) first elements of

$$\lambda^{\cdot j} = \begin{bmatrix} \lambda^{j+1,j} \\ \vdots \\ \lambda^{nj} \end{bmatrix}, \tag{3.23}$$

then we can rewrite (3.21) as

$$\lambda^{ij} = {}_{(j-1)}\lambda_i' \begin{bmatrix} 1 \\ \lambda^{\cdot j}_{(n-i+1)} \end{bmatrix} \tag{3.24}$$

for i > j. If we now realize that $\lambda^{\cdot j}_{(n-i+1)}$ only refers to the equations previous to equation i, we can use the independence of λ_i with respect to these previous equations in deriving the following recursive relationship for the covariance between λ^{ij} and $\lambda^{i\ell}$ (i > j, i > ℓ)

$$\begin{aligned} &\mathrm{Cov}(\lambda^{ij}, \lambda^{i\ell}|Y,X) = \\ &\mathrm{tr}\ \mathrm{Cov}({}_{(j)}\lambda_i, {}_{(\ell)}\lambda_i'|Y,X)\mathrm{Cov}(\lambda^{\cdot \ell}_{(n-i+1)}, \lambda^{\cdot j'}_{(n-i+1)}|Y,X) \\ &+ [1: E(\lambda^{\cdot j'}_{(n-i+1)}|Y,X)]\mathrm{Cov}({}_{(j-1)}\lambda_i, {}_{(\ell-1)}\lambda_i'|Y,X)\begin{bmatrix} 1 \\ E(\lambda^{\cdot \ell}_{(n-i+1)}|Y,X) \end{bmatrix} \\ &+ {}_{(j)}\hat{\lambda}_i' \mathrm{Cov}(\lambda^{\cdot j}_{(n-i+1)}, \lambda^{\cdot \ell'}_{(n-i+1)}|Y,X)\,{}_{(\ell)}\hat{\lambda}_i, \end{aligned} \tag{3.25}$$

which can be evaluated by use of (3.22) and (3.23), and from which we find the variance of λ^{ij} by taking $j = \ell$. The covariance of two elements from different rows of Λ^{-1} simplifies somewhat to

$$\mathrm{Cov}(\lambda^{ij}, \lambda^{k\ell}|Y,X) = {}_{(j)}\hat{\lambda}_i \mathrm{Cov}(\lambda^{\cdot j}_{(n-i+1)}, \lambda^{k\ell}|Y,X) \qquad (3.26)$$

for $i > k$, which allows a recursive solution knowing (3.25).

In addition, we should be able to derive analytical expressions for higher order moments of α or Λ, although we have not yet done so at the present stage. From (3.14) we know that such moments exist at least up to the order $\nu_* - 2n - \mu$.

It should again be stressed that all analytical results here formally only apply to the very restrictive cases treated here, i.e. unrestricted, matrix restricted or RLI models, within the domain of SURE models under diffuse stochastic prior information. They could very well provide useful first approximations in more general SURE models, that can be obtained at a very low computational cost[3], but such matters remain to be investigated in an empirical fashion. Of course, the analysis also suggests a rather natural way of performing the conditional Monte Carlo drawings, using the recursive transformation of the model.

Finally, we could remark that a possible case of some practical meaning covered by our analytical analysis here is the two-equation LI model found in Lubrano and Richard (1981), which can be of some use in an instrumental variables analysis of weak exogeneity, conditionally upon the coefficient of the variable tested for exogeneity in the structural equation [see also Steel (1987)]. In addition, this case can not be treated by the software available in the Bayesian regression program (BRP) as described in Bauwens et al. (1981), since this package relies on the 2-1 poly-t form mentioned in the previous subsection. Whenever both equations have regressors in common and we have diffuse prior stochastic information, one of the conditions given in Drèze and Morales (1976) for obtaining a 2-1 poly-t density is violated, and BRP cannot give us marginal posterior results. Using the methods described here, however, these results are easily obtained.

4. Concluding Remarks

We feel the present analysis may contribute to a somewhat better understanding of the scope of analytical Bayesian results. To the author, at least, the fact that in SURE models of the RLI form analytical inference is possible was not at all obvious from the start, the other two (NC) special cases being, of course, well known. Although the impact of the present results may be more at a theoretical level than at an empirical one, we feel that it might increase our grasp on more complicated models, where our simple analytical analysis formally does not apply.

In particular, we think of help in the choice of importance functions for Monte Carlo integration, but also of simple first approximations, possibly made less crude by iterating on the order of the equations, as we know that the actual results should be invariant to changing the order of the equations. Such methods could be considered for the analysis of general SURE models under diffuse prior densities as examined here, but we could also think of extensions to SURE models under stochastic prior information on α or V, or to simultaneous equation models (SEM), as used in Drèze and Richard (1983). As the cases that allow a formal analytical treatment do not seem to cover many models used in empirical work, such extensions seem an important topic for future research.

Appendix A: Proof of sufficiency of (2.27) for (2.25) and (2.26).

If we want a solution of (2.26) in terms of p_i^{α} to exist, a necessary and sufficient condition is that

$$P_i^X \, X \, p_i = X \, p_i, \tag{A.1}$$

whereas we know from the recursive structure in (2.17)-(2.19) that p_i will only depend on $\alpha_{(i)}$, from which we can rewrite

$$X \, p_i = X_{(i)} \, \tilde{p}_i, \tag{A.2}$$

implicitly defining $\tilde{p}_i$ as the vector of nonzero elements in p_i, corresponding to the explanatory variables used in the first i equations. Obviously, a sufficient[4] condition for (A.1) to hold is then

$$P_i^X \, X_{(i)} = X_{(i)}. \tag{2.27}$$

Under (2.27) the existence of a solution to (2.26) is thus established, while we know it will be unique due to the full column rank of X_i and given by

$$p_i^{\alpha} = X_i^{+} \, X_{(i)} \, \tilde{p}_i. \tag{A.3}$$

From the definitions of X_i and $X_{(i)}$ we obtain

$$X_i = X \, S_{ii} = X_{(i)} \, \tilde{S}_{ii}, \tag{A.4}$$

where $\tilde{S}_{ii}$ groups the nonzero rows of S_{ii}. It is well known [see e.g. Magnus and Neudecker (1988) p. 33] that X_i and P_i^X have the same rank ℓ_i, and, under (2.27), we obtain

$$r(X_{(i)}) = r(P_i^X \, X_{(i)}) \leq r(P_i^X) = \ell_i,$$

whereas, from (A.4),

$$r(X_i) = \ell_i \leq r(X_{(i)}),$$

so that

$$r(X_{(i)}) = r(X_i) = \ell_i. \tag{A.5}$$

As both X_i and $X_{(i)}$ were assumed to be of full column rank, the latter implies that $\tilde{S}_{ii}$ is square and nonsingular and we get from (A.3) and (A.4)

$$p_i^{\alpha} = \tilde{S}_{ii}^{-1} \tilde{p}_i, \tag{A.6}$$

leading directly to the expression in (2.25), which completes the proof.

Appendix B: Some technicalities of the original parameterisation.

A first result that will be shown here is the proportionality of $|H_n^*|$, defined in the main text following (3.5), to the inverse of the first factor in (3.5), corresponding to the posterior density for $\alpha_{(n-1)}$, the unrestricted coefficients in the first (n-1) equations.

Using the definition of H_n^*, we see clearly that

$$|H_n^*| = |X'X||I_m - (\Pi_{n-1}-\hat{\Pi}_{n-1}) [Y'_{n-1}M_x Y_{n-1} +$$

$$(\Pi_{n-1}-\hat{\Pi}_{n-1})'X'X (\Pi_{n-1}-\hat{\Pi}_{n-1})]^{-1} (\Pi_{n-1}-\hat{\Pi}_{n-1})'X'X|, \tag{B.1}$$

which implies that, given the data, $|H_n^*|$ is proportional to the second determinant in (B.1), which can be rewritten via a theorem in Zellner (1971, p.231) as

$$|H_n^*| \propto |I_{n-1} - (\Pi_{n-1}-\hat{\Pi}_{n-1})'X'X (\Pi_{n-1}-\hat{\Pi}_{n-1})$$

$$[Y'_{n-1}M_x Y_{n-1} + (\Pi_{n-1}-\hat{\Pi}_{n-1})'X'X (\Pi_{n-1}-\hat{\Pi}_{n-1})]^{-1}|, \tag{B.2}$$

where the expression in square brackets, say E, is, of course, nonsingular as V_* in (3.2) is PDS, which is required for a proper posterior density on α.

Replacing I_{n-1} by EE^{-1} in (B.2), it is then easily seen that

$$|H_n^*| \propto |Y'_{n-1}M_xY_{n-1}||E^{-1}| \propto |E|^{-1}, \tag{B.3}$$

which is the result we set out to prove.
Remark that (B.3) only holds if $Y'_{n-1}M_xY_{n-1}$ is taken to be nonsingular as well, a restriction that is implicitly imposed in view of the full column rank of W_n [see the discussion following (3.16)] and the RLI assumption, the latter of which implies that S_{nn} is invertible and thus that

$$M_n^x = I - XS_{nn}(S'_{nn}X'XS_{nn})^{-1}S'_{nn}X' = M_x. \tag{B.4}$$

A second result is aimed at a quick verification of the equivalence of the analysis in terms of the original and the recursive parameterisations. We shall limit ourselves to comparing the posterior mean for α_n as implied by (3.6) adding an RLI assumption, and as stated in (3.18), where the same assumption was already made.

From (3.6) we obtain under an RLI structure:

$$\begin{aligned} &E(\alpha_n|\alpha_{(n-1)},Y,X) \\ &= S_{nn}^{-1}\,(\pi_n^* - S_{n(n-1)}\alpha_{(n-1)}) \end{aligned} \tag{B.5}$$

with

$$\pi_n^* = \hat{\pi}_n + (\Pi_{n-1}-\hat{\Pi}_{n-1})\,(Y'_{n-1}M_xY_{n-1})^{-1}Y'_{n-1}M_xy_n, \tag{B.6}$$

which can be rewritten as

$$\pi_n^* = \hat{\pi}_n - \hat{\Pi}_{n-1}\hat{\lambda}_n + \Pi_{n-1}\hat{\lambda}_n$$

in view of (3.15) and (B.4). Using the fact that, from (3.15),

$$\hat{p}_n^{\alpha} = (X_n'X_n)^{-1} X_n'Y \begin{bmatrix} -\hat{\lambda}_n \\ 1 \end{bmatrix}$$

$$= S_{nn}^{-1} \hat{\Pi} \begin{bmatrix} -\hat{\lambda}_n \\ 1 \end{bmatrix},$$

we see clearly that (B.5) becomes

$$E(\alpha_n | \alpha_{(n-1)}, Y, X)$$

$$= \hat{p}_n^{\alpha} + S_{nn}^{-1} [(\hat{\lambda}_n' \otimes I) S_{(n-1)} - S_{n(n-1)}] \alpha_{(n-1)}, \qquad (B.7)$$

from which (3.18) follows by taking expectations with respect to $\alpha_{(n-1)}$. Similarly, more complicated expressions, like (3.19), can be checked, using the theory of partitioned matrices and the binomial inverse theorem of Woodbury as documented e.g. in Press (1972, p. 23).

Appendix C: Derivation of the Jacobian in (3.10).

Taking into account the symmetry of V, we wish to consider the parameter transformation of $(\{\alpha_i, v_{ij}\})$ to $(\{p_i^{\alpha}, \lambda_i, \omega_i^2\})$ for $i > j$, say. If we split this up into two steps, namely

$$(\{\alpha_i, v_{ij}\}) \rightarrow (\{\alpha_i, \lambda_i, \omega_i^2\})$$

with Jacobian J_1, and

$$(\{\alpha_i, \lambda_i, \omega_i^2\}) \rightarrow (\{p_i^{\alpha}, \lambda_i, \omega_i^2\})$$

with Jacobian J_2, it is clear that we wish to find $J = J_1 . J_2$.

The matrix of partial derivations corresponding to the second part of the transformation will take the simple form:

$$
\begin{array}{c|cccc|cccccc}
 & p_1^{\alpha} & \cdots & p_n^{\alpha} & \omega_1^2 & \lambda_{21} & \omega_2^2 & \lambda_{31} & \lambda_{32} & \cdots & \omega_n^2 \\
\alpha_1 & 1 & & 0 & 0 & & 0 & & & & 0 \\
\vdots & & \ddots & & \vdots & & \vdots & & & & \vdots \\
\alpha_n & & & 1 & 0 & & 0 & & & & 0 \\
\omega_1^2 & & & & & & & & & & \\
\lambda_{21} & & & & & & & & & & \\
\omega_2^2 & & & & & & & & & & \\
\lambda_{31} & & 0 & & & & & I_{\frac{n(n+1)}{2}} & & & \\
\lambda_{32} & & & & & & & & & & \\
\vdots & & & & & & & & & & \\
\omega_n^2 & & & & & & & & & &
\end{array}
\qquad (C.1)
$$

from (2.24), where the shaded areas contain nonzero elements that do not influence J_2, which is seen to be equal to one:

$$J_2 = |(C.1)| = 1. \qquad (C.2)$$

In order to obtain J_1, we consider the corresponding matrix of partial derivatives

$$
\begin{array}{c|ccc|cccccc}
 & \alpha_1 & \cdots & \alpha_n & \omega_1^2 & \lambda_{21} & \omega_2^2 & \lambda_{31} & \lambda_{32} & \cdots & \omega_n^2 \\
\alpha_1 & & & & & & & & & & \\
\vdots & & I_{\ell} & & & & & 0 & & & \\
\alpha_n & & & & & & & & & & \\
\hline
v_{11} & & & & & & & & & & \\
v_{21} & & & & & & & & & & \\
v_{22} & & 0 & & & & & D & & & \\
v_{31} & & & & & & & & & & \\
v_{32} & & & & & & & & & & \\
\vdots & & & & & & & & & & \\
v_{nn} & & & & & & & & & &
\end{array}
\qquad (C.3)
$$

where D can be obtained by exploring the relationship given in (2.15), from which we derive

$$v_{ii} = \omega_i^2 + 2\,\lambda_i' v_{(i-1),i} - \lambda_i' V_{(i-1),(i-1)}\lambda_i \tag{C.4}$$

and

$$v_{ij} = \lambda_i'(v_{(i-1),j} - V_{(i-1),(i-1)}\lambda_j) + v_{i,(j-1)}\lambda_j \tag{C.5}$$

if we partition the upper left i × j block of V as

$$\begin{bmatrix} V_{(i-1),(j-1)} & v_{(i-1),j} \\ v_{i,(j-1)} & v_{ij} \end{bmatrix}. \tag{C.6}$$

From (C.4) and (C.5) the partial derivatives are found to be

$$\frac{\partial v_{ii}}{\partial \omega_i^2} = 1; \quad \frac{\partial v_{ii}}{\partial \omega_k^2} = 0 \quad \forall\, k > i$$

$$\frac{\partial v_{ij}}{\partial \omega_k^2} = 0 \quad \forall\, k \geq i > j$$

$$\frac{\partial v_{i,(i-1)}'}{\partial \lambda_i'} = V_{(i-1),(i-1)}$$

$$\frac{\partial v_{i,(i-1)}'}{\partial \lambda_k'} = 0 \quad \forall\, k > i$$

and, finally,

$$\frac{\partial v_{ii}}{\partial \lambda_k'} = 0 \quad \forall\, k > i.$$

This implies that D becomes

$$
\begin{array}{c}
 \\
v_{11} \\
v_{21} \\
v_{22} \\
v_{3(2)} \\
v_{33} \\
\vdots \\
v_{nn}
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} \omega_1^2 & \lambda_{21} & \omega_2^2 & \lambda_3' & \omega_3^2 & \cdots & \omega_n^2 \end{array} \\
\left[\begin{array}{ccccccc}
1 & & & & & & \\
 & v_{11} & & & & 0 & \\
 & & 1 & & & & \\
 & & & V_{(2),(2)} & & & \\
 & & & & 1 & & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{array}\right]
\end{array}
, \qquad \text{(C.7)}
$$

a block-lower triangular matrix with nonzero elements in the shaded area, the determinant of which leads to

$$
J_1 = \left| \prod_{i=1}^{n-1} |V_{(i),(i)}| \right| = \prod_{i=1}^{n-1} (\omega_i^2)^{n-i}, \qquad \text{(C.8)}
$$

using (3.8), and $J = J_1$ thus becomes equal to the expression in (3.10).

Appendix D: Rank assumptions made throughout the paper.

object (i:1→ n)	assumption (implication)	discussed in
S	full column rank	2.1
⇒ S_{ii}	(full column rank)	2.1
X	full column rank	2.1; 3.2
⇒ $X_{(i)}$	(full column rank)	2.2; App. A
V	PDS	2.1
V_*	PDS	3.1; 3.2; App. B
$(X_i \; Y_i)$	full column rank	3.3
⇒ W_i	(full column rank)	3.3; App. B
⇒ X_i	(full column rank)	2.2; App. A
⇒ Y_{i-1}	(full column rank)	3.3
⇒ $y_i' M_i^W y_i$	(> 0)	3.3

Footnotes

1) The term "noninformative" might be considered abuse of language in view of the numerous problems associated with expressing prior ignorance [see e.g. Zellner (1971) and Bernardo (1979)]. We could, of course, replace it by "diffuse" or "vague", or we might choose to use the denomination "reference prior". Subsection 3.1 briefly comments on these issues.

2) It proves notationally convenient to use these recursions for i: 1 → n and define $\pi_{(0)}$, $\alpha_{(0)}$, $S_{(0)}$, $S_{1(0)}$ and Y_0 to be of zero dimension.

3) It is certainly much easier to compute than the FIML solution, used as a starting point for a numerical analysis in e.g. Morales (1971) and Bauwens (1984).

4) This condition is not necessary unless $\tilde{p}_i$ is a scalar different from zero.

References

Bernardo, J.M., 1979, Reference posterior distributions for Bayesian inference, Journal of the Royal Statistical Society B 41, 113-147.

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

Bauwens, L., J-P. Bulteau, P. Gille, L. Longree, M. Lubrano and H. Tompa, 1981, Bayesian regression program (BRP) user's manual, CORE Computing Report 81-A-01 (Catholic University of Louvain, Louvain-la-Neuve).

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, Ch. 9 in: Z. Griliches and M.D. Intriligator, eds., Handbook of Econometrics (North-Holland, Amsterdam).

Engle, R.F., D.F. Hendry and J.F. Richard, 1983, Exogeneity, Econometrica 51, 227-304.

Geweke, J., 1988, Antithetic acceleration of Monte Carlo integration in Bayesian inference, Journal of Econometrics 38, 73-89.

Hammersley, J.M. and D.C. Handscomb, 1979, Monte Carlo methods (Chapman and Hall, London).

Hendry, D.F., 1984, Monte Carlo experimentation in econometrics, Ch. 16 in: Z. Griliches and M.D. Intriligator, eds., Handbook of Econometrics (North-Holland, Amsterdam).

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

Lubrano, M. and J.F. Richard, 1981, Specification of the prior density in single equation errors-in-variables models: an application to a U.K. money demand equation, CORE Discussion Paper 8101 (Catholic University of Louvain, Louvain-la-Neuve).

Magnus, J.R. and H. Neudecker, 1988, Matrix differential calculus with applications in statistics and econometrics (Wiley, Chichester).

Malinvaud, E., 1981, Méthodes statistiques de l'économétrie (Dunod, Paris).

Morales, J-A., 1971, Bayesian full information structural analysis (Springer Verlag, Berlin).

Press, S.J., 1972, Applied multivariate analysis (Holt, Rinehart and Winston, New York).

Richard, J.F. and M.F.J. Steel, 1988, Bayesian analysis of systems of seemingly unrelated regression equations under a recursive extended natural conjugate prior density, Journal of Econometrics 38, 7-37.

Richard, J.F. and H. Tompa, 1980, On the evaluation of poly-t density functions, Journal of Econometrics 12, 335-351.

Sims, C.A., 1980, Macroeconomics and reality, Econometrica 48, 1-48.

Steel, M.F.J., 1987, A Bayesian analysis of multivariate exogeneity: a Monte Carlo approach, Nouvelle série no. 168 (Catholic University of Louvain, Louvain-la-Neuve).

Stewart, W.E., 1987, Multiresponse parameter estimation with a new and noninformative prior, Biometrika 74, 557-562.

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

van Dijk, H.K., 1985, Existence conditions for posterior moments of simultaneous equation model parameters, Econometric Institute Report 8551/A (Erasmus University, Rotterdam).

Zellner, A., 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association 57, 348-368.

Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York).

Zellner, A., L. Bauwens and H.K. van Dijk, 1988, Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods, Journal of Econometrics 38, 39-72.

Already published in this series:

No.	Author(s)	Title
8801	Th. van de Klundert and F. van der Ploeg	Fiscal Policy and Finite Lives in Interdependent Economies with Real and Nominal Wage Rigidity.
8802	J.R. Magnus and B. Pesaran	The Bias of Forecasts from a First-order Autoregression.
8803	A.A. Weber	The Credibility of Monetary Policies, Policymakers' Reputation and the EMS-Hypothesis: Empirical Evidence from 13 Countries.
8804	F. van der Ploeg and A.J. de Zeeuw	Perfect Equilibrium in a Model of Competitive Arms Accumulation.

In publication:

8806	Th. Ten Raa and E.N. Wolff	Secondary Products and the Measurement of Productivity Growth.