BAYESIAN ANALYSIS OF SYSTEMS OF SEEMINGLY UNRELATED REGRESSION EQUATIONS UNDER A RECURSIVE EXTENDED NATURAL CONJUGATE PRIOR DENSITY

by

J.F. Richard and M.F.J. Steel

ISDS, Duke University
CORE, Catholic University of Louvain

DP#: 87-06

Abstract

A "Recursive Extended Natural Conjugate" prior density is proposed for systems of seemingly unrelated regression equations. Compared to other classes of Extended Natural Conjugate prior densities it offers the advantage that analytical expressions are available for the prior covariance matrix of the regression coefficients. The posterior analyses combines together analytical and numerical techniques. Numerical integration is based on Monte-Carlo importance sampling. An application to a system of four wage equations for EEC countries serves to evaluate different techniques discussed in the paper.

Acknowledgements

We are grateful to John Geweke and to an anonymous referee for many useful comments on an earlier draft of the paper. The usual disclaimer applies.
1. Introduction

The Bayesian analysis of systems of linear equations has proved challenging to statisticians and econometricians over the last twenty-five years if only for two main reasons:

(i) The only case for which a straightforward analytical treatment of the exact posterior and predictive densities is available is that of an unrestricted System of Seemingly Unrelated Regression Equations (SURE) under a so-called Natural Conjugate (NC) prior density.\(^{(1)}\) Unfortunately, as noted initially by Rothenberg (1963), and as briefly discussed in the course of the present paper, the latter greatly lacks flexibility and is, therefore, of limited interest for applications.

(ii) Attempts to define more flexible prior densities which we shall generically refer to as "Extended Natural Conjugate" (ENC) prior densities lead to posterior densities whose analysis requires partial numerical treatment in parameter spaces that can easily be of (relatively) "large" dimensionality.

Though in the present paper for reasons that are developed below, we restrict our attention to SURE models, we can obviously rely upon techniques that have been developed for Simultaneous Equation Models (SEM) since, from an analytical point of view, the former can be viewed as a special, and simpler case of the latter.

The Bayesian analysis of SEM's dates back to unpublished papers by Drèze (1962), Rothenberg (1963), and Zellner (1965) and has been surveyed by Drèze

\(^{(1)}\) Analytical asymptotic approximations are available under a broader class of (non-dogmatic) prior densities. See e.g. Zellner (1971, p. 240) for an example. Other approximations, in the form of so-called importance functions, are discussed in the literature on Monte Carlo numerical integration to which we refer below.
and Richard (1983) where additional references can be found. Substantial progress has been achieved in a number of complementary directions.

(i) In the early seventies, when the scope for numerical integration techniques was still limited, emphasis was put on generalizations of the NC approach leading to the introduction of a variety of ENC prior densities. References of interest in this context are Harkema (1971), Morales (1971), Richard (1973) or Drèze and Morales (1976). These attempts led to a number of pioneering applications but have received little attention from practitioners, if only because, if we except the contribution in Richard, no software packages were made available (or, in the case of Richard, it lacks the user friendliness that has become a must nowadays).

(ii) Starting with the pioneering work of Klock and van Dijk (1978) numerical integration based on Monte Carlo importance sampling procedures have been applied to SEM’s with increasing success, which takes full advantage of the last ten years’ explosion in computing power. Recent contributions are van Dijk (1984), Bauwens (1984) and, in the present volume, Geweke (1987) and Zellner et al. (1987). Nevertheless, though substantial progress has been achieved, in our opinion much remains to be done before importance sampling procedures can be routinely applied to systems of equations.

(iii) The class of distributions which can be efficiently handled by means of appropriate combinations of numerical and analytical techniques has been broadened over the years. In particular, poly-t densities, as defined in Drèze (1976), are now fully operational and there exist efficient software for evaluating some of their characteristics and performing random drawings. See Richard and Tompa (1980), Bauwens (1984) or Bauwens and Richard (1985). Poly-t densities can be extremely skewed or even bimodal so that they can usefully be considered as importance functions for ill-behaved problems.
Procedures for selecting poly-t importance functions are on our research agenda.

(iv) Throughout the years, applications of varying complexity have been analysed and have largely contributed to improving our understanding of the difficulties practitioners are likely to encounter when designing their own applications. See the above mentioned references.

The object of our paper which positions itself in the continuation of these lines of research is to propose a "Recursive Extended Natural Conjugate" (RENC) prior density for SURE systems whose coefficients may be subject to linear restrictions. Several reasons motivate our interest in this problem:

(i) SURE systems can be of interest in their own right for applications such as the one we discuss in section 5 below. Multivariate time series techniques and causality analysis, as discussed e.g. in Sims (1977) are also domains of application. Furthermore the analysis of the "reduced" form associated with a "structural" SEM is as important as ever. Samples are naturally informative on the reduced form coefficients and, in a sense, the structural equations serve to define functions of those coefficients that are of interest to the model builder (e.g. because, on the basis of economic theory, they are thought to be invariant in worlds subject to policy interventions). The contribution by Zellner et al. (1987) in the present volume takes full advantage of this basic feature of SEM's.

(ii) Reduced form analysis plays a central role in the "Limited Information" (LI) or "Instrumental Variables" (IV) analysis of a subsystem of structural equations, the fully operational case being that of a single structural equation. Except for the equation(s) of interest, the structural equations are (implicitly) replaced by their reduced form counterparts, whose coefficients are treated as nuisance parameters. Efficient marginalisation
with respect to these nuisance parameters is the cornerstone of Bayesian LI or IV techniques and is achieved under an unrestricted NC prior density or under a limiting "non-informative" prior density. Technical details are found e.g. in Zellner (1971), Drèze and Richard (1983), Richard (1984), Lubrano et al. (1986) and Steel (1987b). As discussed in Richard, there are good reasons for trying to extend these techniques to cases where the reduced form equations are subject to restrictions (degrees of freedom, noncausality assumptions,...)

(iii) The range of applicability of RENC prior densities is limited by the fact that no analytical expressions exist for the prior covariance matrix of the regression coefficients (except for the case where the coefficients associated with different equations are a priori independent of each others). The class of RENC prior densities we discuss below does not suffer from that limitation.

(iv) The search for families of importance functions that are easy to handle and yet flexible enough is, more than ever, of relevance. We intend to argue that RENC prior densities could be useful candidates since, in particular, we shall demonstrate that their first and second order moments are fully flexible and can accommodate exact linear prior restrictions (or, in the context of local approximations, more general non linear restrictions).

The class of RENC prior densities we discuss in our paper is closely related to that proposed by Steel (1987a) though it is modified in a way which facilitates its elicitation and yet requires little or no additional computation (depending on which method of analysis is adopted).

The paper is organized as follows:

In Section 2 we describe the SURE model and establish notation, special attention being paid to the treatment of exact prior restrictions. RENC prior densities are introduced and discussed in Section 3. The posterior analysis
is conducted in Section 4, where two approaches are considered, depending on whether one first integrates the regression coefficients or the covariance matrix. These techniques are applied in Section 5 to a system consisting of four wage equations. Conclusions are drawn in Section 6.

2. The SURE MODEL

2.1 Likelihood Function

The domain of application of SURE systems covers both cross-sectional and time series problems. Which of the two is emphasized does not affect the algebra of our analysis. We shall, however, adopt here a time series framework which is relevant to a number of applications we are interested in, such as exogeneity or non-causality analysis.

Consider a sequential conditional model whereby a vector of endogenous variables \( y(t) \in \mathbb{R}^m \) is generated at time \( t \) conditionally on an information set \( I_t \) which consists of a vector of currently dated (weakly) exogenous variables \( z(t) \in \mathbb{R}^k \) and of the lagged values of the \( y \)'s and the \( z \)'s including initial conditions that are assumed to be known. For practical purposes it is assumed that \( I_t \) can be represented by a fixed dimensional vector \( x(t) \in \mathbb{R}^m \) in the sense that, conditionally on \( x(t) \), \( y(t) \) is independent of the other variables in \( I_t \).

Normality is assumed with linear conditional expectations and a covariance matrix which is constant across observations

\[
y(t) \mid I_t \sim N(\Pi'x(t), V) \quad t: 1 \to T
\]  

(2.1)

where \( \Pi \) is an mxn matrix function of a vector of unknown parameters \( \alpha \in \mathbb{R}^k \) and \( V > 0 \) is an unrestricted symmetric positive definite covariance matrix. Hence, the disturbances
\[ \epsilon(t) = y(t) - \Pi^* x(t) \quad t: 1 \rightarrow T \] (2.2)

are "innovations" relative to the information set \( I_L \) and, in particular, are linearly independent of \( \{ x(s); s: 1 \rightarrow t-1 \} \).

We shall restrict our attention to the case where \( \Pi \) is subject only to linear restrictions and, in order to avoid conditionalization paradoxes [see, for example, Kolmogorov (1950)] we shall use an explicit representation of these restrictions. Hence, it is assumed that \( \pi \), the column vector expansion of \( \Pi \), can be written as

\[ \pi = \text{vec} \Pi - S\alpha \] (2.3)

where \( S \) is an \( m \times l \) matrix of known constants of rank \( l \) and \( \alpha \) consists of \( l \) "free" parameters.

We assume that a sample of size \( T \) is available. Let \( Y \) and \( X \) denote the \( T \times n \) and \( T \times m \) matrices of observations on \( y(t) \) and \( x(t) \) respectively.\(^{(3)}\) In order to be able to single out individual observations or equations we adopt the following partitionings:

\[ Y = (y_1 \ldots y_1 \ldots y_n) = \begin{pmatrix} y(1) \\ \vdots \\ y(t) \\ \vdots \\ y(T) \end{pmatrix} \quad X = (x_1 \ldots x_j \ldots x_m) = \begin{pmatrix} x(1) \\ \vdots \\ x(t) \\ \vdots \\ x(T) \end{pmatrix} \] (2.4)

\(^{(2)}\) Non homogeneous restrictions can be transformed into homogeneous restrictions as in (2.3) through a transformation of the \( y \)'s and, hence, need not to be considered explicitly.

\(^{(3)}\) No explicit assumption is made regarding the rank of \( X \). It is, however, implicitly assumed that prior and sample information complement each other in such a way that the first and second order moments of whatever functions of \( \alpha \) are of interest to us exist.
\[ \Pi = (\pi_1 \ldots \pi_i \ldots \pi_n) \] so that \[ \pi' = (\pi_1' \ldots \pi_i' \ldots \pi_n') \] (2.5)

Two equivalent representations are available for the likelihood function depending on whether we use the multivariate normal density associated with the matrix \( Y \) or the normal density associated with the vector \( y \). The corresponding expressions are

\[ L(\alpha, V; Y, X) \propto f_M^{TXn}(y| X\Pi, V \otimes I_n) \] (2.6)

with \( \pi = \text{vec}\Pi = \sigma \), or

\[ L(\alpha, V; Y, X) \propto f_N^{Tn}(y| (I_n \otimes X)\sigma, V \otimes I_n) \] (2.7)

Expressions for the density functions are presented in Appendix A.

2.2 Prior Exact Restrictions: Special Cases

The formulation (2.3) covers restrictions within equations, such as long term unit elasticity restrictions, or across equations, such as adding up constraints, or any combination of the two types. Two cases deserve special attention.

(i) The NC approach we discuss in section 3.1 below applies only to cases where equation (2.3) can be rewritten in matrix form which requires that only those restrictions which apply either to every row or to every column of \( \Pi \) can be taken into consideration. In such a case equation (2.3) can be rewritten as

\[ \Pi = BAG \] (2.8)

where \( A \) is a pxq matrix of "free" coefficients while \( B \) and \( C \) are respectively
and \( \mathbf{Q} \) matrices of known constants with rank \( B-p \) and rank \( C-q \). The correspondence between formulae (2.3) and (2.8) is given by

\[
\alpha - \text{vec} \mathbf{A} \quad \mathbf{S} = (\mathbf{C} \otimes \mathbf{B})
\]  

(2.9)

As discussed e.g. in Richard (1984), the matrixvariate normal distribution is stable under linear (matrix) transformations of the form (2.8).

(ii) A case of special interest to us is that which corresponds to the I or IV analysis of a single structural equation, whereby it is assumed that the conditional expectation of \( y(t) \) belongs to a linear manifold of dimension \( n-1 \) whose equation is given by

\[
\beta' \mathbb{E}(y(t) | I_T) = \gamma' x(t) \quad t: 1 \to T
\]  

(2.10)

where \( \beta \) and \( \gamma \) are vectors of unknown coefficients. For notational convenience the equation is normalized by setting the first element in \( \beta \) equal to one so that \( \beta' = (1:b') \). The vectors \( b \) and \( \gamma \) are then subject to linear indentifying restrictions (at least \( n-1 \) of them when \( \Pi \) is left otherwise unrestricted). Let \( B \) denote the set of a prior admissible values of \( (\beta, \gamma) \). Compatibility between formulation (2.1) and assumption (2.10) requires that \( \Pi \) is restricted to the set

\[
\mathbb{P}_B = \{ \Pi \mid \exists (\beta, \gamma) \in B \; ; \; \Pi \beta = \gamma \}
\]  

(2.11)

Conditionally on \( b \), which has to be dealt with numerically anyway, we can reformulate condition (2.11) in a way which is compatible with our approach. We partition \( \Pi \) into \( (\pi_1: \pi_2) \) and substitute \( \gamma \) for \( \pi_1 \), obtaining thereby
\[ \Pi = \Pi Q_b \] (2.12)

with

\[ \Pi = (\gamma: \Pi_2) \quad Q_b = \begin{pmatrix} 1 & 0 \\ -b & I_{n-1} \end{pmatrix} \] (2.13)

or, equivalently

\[ \pi = (Q_b' \otimes I_m) \tilde{\pi} \] (2.14)

The conventional LI framework sets restrictions on \( \gamma \) but leaves \( \Pi_2 \) unrestricted while our approach allows for the possibilities that, conditionally on \( b \), \( \Pi \) is subject to arbitrary linear restrictions. In particular identification could be achieved by means of restrictions linking together \( b, \gamma \) and \( \Pi_2 \). Hence, in all generality, we may now assume that

\[ \tilde{\pi} = \tilde{S}_b \alpha \] (2.15)

It follows from formulae (2.14) and (2.15) that the "restricted" LI or IV framework is covered by formula (2.3), where, conditionally on \( b \), \( S \) is set equal to \( (Q_b' \otimes I_m) \tilde{S}_b \). See Steel (1987a) for more details.

2.3 A Recursive Form

As is well known (see, for example, Strotz and Wold (1960)), the system (2.1) can be rewritten in a recursive form. Let \( Y_1, \Pi_1 \) and \( V_1 \) denote the (endogenous) data matrix and the matrices of parameters associated with the
subsystem consisting of the first i equations in the SURE model (2.6).

Following, in particular, equations (2.4) and (2.5), the recursion from i-1 to i is given by

\[ Y_i = (Y_{i-1}; y_i) \quad \Pi_i = (\Pi_{i-1}; \pi_i) \quad (2.16) \]

and say

\[ V_i = \begin{pmatrix} v_{i-1} & \delta_i \\ \delta_i & v_i \end{pmatrix} \quad (2.17) \]

A recursive application of the standard factorization of the multivariate normal density function yields the following factorization of the likelihood function (2.6)

\[ L(\alpha, (\lambda_1, \omega_1^2); Y, X) \propto \prod_{i=1}^{n} f_N^{T}(y_i \mid Xp_i + Y_{i-1} \lambda_i, \omega_i^2, T) \quad (2.18) \]

where, for i = 2 → n,

\[ \lambda_i = v_{i-1}^{-1} \delta_i, \quad p_i = \pi_i - \Pi_{i-1} \lambda_i, \quad \omega_i^2 = v_i^2 - \lambda_i^2 v_{i-1} \lambda_i \quad (2.19) \]

and where it proves notationally convenient to allow for \( \lambda_1 \) being a zero dimensional vector so that \( p_1 = \pi_1 \) and \( \omega_1^2 = v_1^2 \). The p's are functions of \((\alpha, (\lambda_1))\) and their expressions are given in formula (2.27) below. The matricvariate representation of the likelihood function (2.18) is

\[ L(\alpha, (\lambda_1, \omega_1^2); Y, X) \propto f_{MN}^{T}(Y \Lambda^\top X, D, D \odot I_T) \quad (2.20) \]

where \( \Lambda \) is the lower triangular matrix whose diagonal elements are set equal to one and whose i-th row, for i:2 → n, is given by \((-\lambda_i : 1 : 0^\top)\). P and D are given by
\[ P = (p_1 \ldots p_n) = \Pi \Lambda^*, \quad D_\omega = \begin{bmatrix} \omega_1^2 & 0 \\ \vdots & \ddots \\ 0 & \omega_n^2 \end{bmatrix} = \Lambda V \Lambda^* \quad (2.21) \]

2.4 A Recursive Representation of the a priori Restrictions

Since the prior density we introduce in section 3.2 below is defined in terms of the recursive form parameters we have to reformulate accordingly the exact prior restrictions (2.3). Note first that, since \( S \) is an \( m \times l \) matrix of rank \( l \), we can postmultiply it by an \( l \times l \) non singular upper triangular matrix \( U \), representing an appropriate sequence of elementary transformations of the columns of \( S \), and, thereby, reduce \( S \) to a lower block triangular \( m \times l \) matrix of rank \( l \). Hence, without loss of generality, we shall restrict our attention to cases where \( S \) is such a lower block triangular matrix, possibly at the cost of transforming \( \alpha \) into \( U^{-1} \alpha \) (though no additional notation is introduced to cover such a transformation). We next partition \( S \) and \( \alpha \) conformably with the columns of \( \Pi \) and, also, with each other in such a way that

\[
\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_n \end{pmatrix} \quad S = \begin{pmatrix} S_{11} & 0 & \ldots & 0 \\ S_{21} & S_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \ldots & S_{nn} \end{pmatrix} \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \quad (2.22)
\]

where \( \pi_i \in \mathbb{R}^m \) is the \( i \)-th column of \( \Pi \), \( S_{ij} \) is an \( m \times l_j \) submatrix, \( S_{i:} \) is (block) lower triangular of rank \( l_i \geq 0 \) and \( \alpha_i \in \mathbb{R}^{l_j} \) with \( \sum_{j=1}^{n} l_j = l \). It follows that \( l_i \) measures the number of additional "free" elements in \( \pi_i \) given those already associated with \( (\pi_1 \ldots \pi_{i-1}) \). \( S \) being of full column rank, its Moore-Penrose inverse \( S^+ \) is given by
\[ S^+ = (S'S)^{-1}S' \tag{2.23} \]

It follows that for any a priori admissible \( \pi \), i.e. for any \( \pi \) such that

\[(I_{m} - SS^+)\pi = 0, \text{ the system (2.3) has a unique solution in } \alpha \text{ which is given by } \alpha = S^+\pi \text{ and whose recursive representation is} \]

\[ \alpha_1 = S_{11}^+(\pi_1 \cdot \ldots \cdot \sum_{j=1}^{i-1} S_{1j} \alpha_j) \quad \text{with } S_{11}^+ = (S_{11}^+ S_{1i})^{-1} S_{11}^' \tag{2.24} \]

In section 3.2 below we shall introduce the prior distribution of \( \alpha_1 \) conditionally on \((\alpha_1, \ldots, \alpha_{i-1})\). Hence let \( \pi_{(i)} \) and \( \alpha_{(i)} \) denote the subvectors regrouping the coefficients associated with the first \( i \) equations and let \( S_{(i)} \) be the corresponding leading submatrix of \( S \) in (2.3) so that \( \pi_{(i)} = S_{(i)}\alpha_{(i)} \).

Let \( \pi_{(i)} \), \( \alpha_{(i)} \) and \( S_{(i)} \) be partitioned conformably with each other as follows

\[ \pi_{(i)} = \begin{pmatrix} \pi_{(i-1)} \\ \pi_1 \end{pmatrix}, \quad \alpha_{(i)} = \begin{pmatrix} \alpha_{(i-1)} \\ \alpha_1 \end{pmatrix}, \quad S_{(i)} = \begin{pmatrix} S_{(i-1)} & 0 \\ S_{1(i-1)} & S_{11} \end{pmatrix} \tag{2.25} \]

so that

\[ \pi_1 = S_{11} \alpha_1 + S_{1(i-1)} \alpha_{(i-1)}, \quad \text{and } \alpha_1 = S_{11}^+ (\pi_1 - S_{1(i-1)} \alpha_{(i-1)}) \tag{2.26} \]

Note finally that, conditionally on \( \lambda_1 \), the linear restrictions on \( \Pi \) are transformed into linear restrictions on \( P \) and we have

\[ p = \text{vec}P = (\Lambda \otimes I_m) S\alpha = \text{say } S_\lambda \cdot \alpha \tag{2.27} \]

so that \( S_\lambda \) has the same block lower diagonal structure as \( S \) in equation (2.22). The blocks of \( S_\lambda \) are
\[ S_{i1}(\lambda) = S_{i1} \quad \text{for } i: 1 \rightarrow n, \quad \text{and} \]
\[ S_{ij}(\lambda) = S_{ij} - \sum_{k=j}^{i-1} \lambda_{ik} S_{kj} \quad \text{for } 1 \leq j < i \leq n \quad (2.29) \]

The recursive inverse relationship between \( \alpha \) and \( p \) is given by (2.24) where \((S_{ij}, \pi_i)\) are replaced by \((S_{ij}(\lambda), p_i)\).

3. A Class of RENC Prior Densities

3.1 The NC Prior Density

For later ease of reference, we briefly recall here the definition of the NC prior density for the unrestricted version of the "reduced form" model (2.1). It is defined as the product of the following two densities

\[ D(\Pi \mid V) = f_{MN}^{\Pi \times n}(\Pi \mid \Pi_o, V \otimes M_o^{-1}) \quad (3.1) \]
\[ D(V) = f_{IV}^{n}(V \mid V_o, \nu_o) \quad (3.2) \]

Its limitations are well documented—see e.g. Rothenberg (1962) or Zellner (1971) and also Richard (1984) for technical details that are relevant to the object of the present discussion. Since it requires that, conditionally on \( V \), the prior covariance matrix of \( \tau = \text{vec}\Pi \) should be of the form \( V \otimes M_o^{-1} \) it can only accommodate block linear restrictions of the form (2.8). As pointed out by Zellner (1971), the large samples posterior density for \( \Pi \) and \( V \) has the same structure as the NC prior density whatever non-dogmatic prior is employed.

3.2 The RENC Prior Density

Let \( \theta(1) = ((\alpha_j, \lambda_j, \omega_j^2); j: 1 \rightarrow i) \) denote the parameters in the first \( i \)
recursive equations associated with the factorization (2.18) of the likelihood function. The complete set of parameters is \( \theta = \theta^{(n)} \). As in (2.20), we do not treat the first equation differently from the others at the cost of introducing the dummy notation \( \theta^{(o)} \) which represents an empty set of parameters.

The RENC prior density we consider here is given by

\[
D(\theta) = \prod_{i=1}^{n} D(\theta_i^{(i-1)})
\]

where \( D(\theta_i^{(i-1)}) \) is essentially Natural-Conjugate for the \( i \)-th recursive equation in (2.18). More specifically it is given by the product of the following three densities

\[
D(\omega_i^2 | \theta^{(i-1)}) = D(\omega_i^2) - f_i \gamma(\omega_i^2 | \omega_{i0}^2, v_i^2) 
\]

\[
D(\lambda_i | \omega_i^2, \theta^{(i-1)}) = D(\lambda_i | \omega_i^2) = f_i^{-1}(\lambda_i | \lambda_i^0, \omega_i^2, \lambda_i^0 - 1) 
\]

\[
D(\alpha_i | \lambda_i, \omega_i^2, \theta^{(i-1)}) = D(\alpha_i | \omega_i^2, \alpha_i^{(i-1)}) 
= f_i^{-1}(\alpha_i | \alpha_i^0 + F_i^0(\alpha_i^{(i-1)}), \omega_i^2, \alpha_i^0 - 1) 
\]

The prior hyperparameters consist of the scalars \( \omega_{i0}^2 > 0 \) and \( v_i^2 > 0 \), the vectors \( \lambda_i^0 \in \mathbb{R}^{i-1} \) and \( \alpha_i^0 \in \mathbb{R}^i \) together with the matrix \( F_i^0(\alpha_i^{(i-1)} = (F_{i1}^0, \ldots, F_{i,i-1}^0) \), where \( F_{i,j} \in \mathbb{R}^{i \times j} \), and the relative precision matrices \( \lambda_i^2 > 0 \) and \( \alpha_i^2 > 0 \) which are respectively \( (i-1) \)-and \( l_i \)-dimensional.

Note that the specification (3.4)-(3.6) incorporates two sets of conditional independence assumptions whose object is essentially to facilitate the derivation of the first and second order unconditional prior moments of \( \theta \):

1. Conditionally on \( \omega_i^2 \) and \( \theta^{(i-1)} \), \( \lambda_i \) and \( \alpha_i \) are independent of each other;

14
(ii) \( \lambda_1 \) and \( \omega_1^2 \) are independent of \( \theta_{(i-1)} \) and, conditionally on \( \lambda_1 \) and \( \omega_1^2 \), \( \alpha_1 \) depends solely (and linearly) on \( \alpha_{(i-1)} \).

These assumptions could conceivably be removed and, in particular, the prior hyperparameters could be functions of \( \theta_{(i-1)} \). The recursive techniques which are described below for the analysis of the posterior densities would be generalized accordingly. However, the derivation of the prior moments of \( \theta \) and, hence, the elicitation of the prior density would become more complicated. Furthermore, there is no need to discuss in an expository paper all the refinements a general computer software might incorporate.

In the course of the discussion below, we shall briefly evoke the case where \( P_{1(i-1)}^0 \) is a linear function of \( \lambda_1 \) covering thereby as special cases the NC prior and the version of the RENC prior used by Steel (1987a).

The density functions (3.4) and (3.5) could originate from a joint Inverted-Wishart prior density on \( V \), say

\[
D(V) = f_{1W}^n(V|V_o, \nu_o) \tag{3.7}
\]

in which case the hyperparameters \( \lambda_1^0, \omega_1^2 \) and \( \nu_1^0 (= \nu_{i-1}^0) \) are derived from \( V_o \) through the prior equivalent of the identities (2.17) and (2.19). Also \( \nu_1^0 = \nu_o - n + 1 \) under the condition that \( \nu_o > n - 1 \). Compared to the more general specification in formulae (3.4) and (3.5), the choice of an Inverted-Wishart density for \( V \), as in (3.7), reduces the number of prior hyperparameters from ambiguous; change into \( n(n^2+3n+8)/6 \) and \( (n^2+n+2)/2 \). In fact, except for \( \nu_o \), the density (3.7) is uniquely determined by the prior expectation of \( V \) (for \( \nu_o > n + 1 \)). In practice, however, \( V \) is often a nuisance parameter and achieving maximum flexibility in the choice of its prior density does not seem to be critical, if only for two related reasons:

(i) It is rarely the case that the prior information on \( V \) is so sharp
that it dominates the sample information;

(ii) Our practical experience, based on applications such as that
discussed in section 5 below, seems to suggest that the posterior
density of α is robust against the specification of the prior on V.
Note also, that going for the more general specification in (3.4) and (3.5)
does not affect our posterior analysis since, in particular, random drawings
from an Inverted-Wishart density are based anyway on a recursive decomposition
of V.

It is well-known that the use of a conditional prior density such as
(3.6) in combination with a "diffuse" prior density on ω_1^2 may result in too
much weight being given to the prior relative to the sample information since,
for any preassigned value of the unconditional (relative to ω_1^2) covariance
matrix of α_1, the relative precision matrix G_i^0 is a monotonically increasing
function of ω_{10}^2. (See formulae (4.6) and (4.7) below or Richard (1975, p. 81)
for additional details). Practical experience suggests, however, that the
problem largely vanishes with the use of moderately informative prior
densities for ω_1^2 and, in particular, the application which is discussed in
section 5 below raises no difficulties in this respect. Alternatively, we may
replace the normal density (3.8), which is conditional on ω_1^2, by the following
unconditional t density

\[ D(\alpha_1 | \alpha_{(1-1)}) = \int_t (\alpha_1 | \alpha_1^0 + F_{1(1-1)}(\alpha_{(1-1)} - \alpha_{(1-1)}), \omega_{10}^2, \nu_1) \]  

Though partially redundant, the choice of hyperparameters in (3.8) offers the
advantage that, for ν_1 > 2, the first and second order prior moments of α are
unaffected by the choice between (3.6) and (3.8). However, as discussed in
section 4.2 below, the use of the prior density (3.8) instead of (3.6)
typically results in an increase in the cost of computation for the analysis
of the posterior densities. On the other hand, posterior distributions based on (3.8) may be quite different from those based on (3.6), especially in situations where the prior information conflicts with the sample information.\(^{(4)}\)

Under (3.6) the joint distribution of \(\alpha\), conditionally on \((\omega_1^2 \ldots \omega_n^2)\) is given by

\[
D(\alpha \mid \omega_1^2 \ldots \omega_n^2) = f_N(\alpha, (I-F_o)^{-1}D_\omega (I-F_o)^{-1})
\]

(3.9)

where

\[
D_\omega = \begin{bmatrix}
\omega_1^2 & 0 & \cdots & 0 \\
0 & \omega_2^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \omega_n^2
\end{bmatrix}, \quad F_o = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
F_{21} & 0 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
F_{n1} & F_{n2} & \cdots & F_{nn}
\end{bmatrix}
\]

(3.10)

The unconditional covariance matrix of \(\alpha\) is then obtained by replacing the \(\omega_i^2\)'s in \(D_\omega\) by their expectations under (3.4), i.e. by \((\omega_i^2-1)\omega_i^2\). The factorization of the covariance matrix in (3.9) is standard and indicates that we can select the matrices \(G_i\) and \(F_{ij}\) (for \(j < i\)) so as to obtain any arbitrary covariance matrix for \(\alpha\), achieving thereby full flexibility in the choice of the RENC prior density on \(\alpha\).

We conclude our discussion of the RENC prior specification with four remarks.

(i) The RENC prior is formulated directly in terms of the \(\alpha\)'s in order to avoid the usual conditionalization paradoxes. This being said we can easily drive from the density (3.8), the distribution of \(\pi_i\), conditionally

\(^{(4)}\) In a single equation framework, for example, replacing (3.6) by (3.8) implies replacing a unimodal posterior \(t\) distribution by a potentially bimodal posterior 2-0 poly-\(t\) distribution. Clearly bimodality may provide a more sensible treatment of a conflict of information than unimodality does.
on $\pi_{(i-1)}$ and $\omega_1^2$, through the integrand transformation (2.26). We find that

$$
D(\pi_1 \mid \pi_{(i-1)}, \omega_1^2) = f_m(\pi_1 \mid \pi_{(i-1)}^0 + R_{(i-1)}^0 (\pi_{(i-1)} - \pi_{(i-1)}^0), \omega_1^2 2^0 \nu_1) \tag{3.11}
$$

with

$$
R_{(i-1)}^0 = \left( S_{i(1-1)} + S_{ii} E_{ii(1-1)} \right) S_{ii(1-1)}^+ \text{ and } \Sigma_1^0 = S_{ii} G_{ii}^{-1} S_{ii} \tag{3.12}
$$

Note that, under (3.11), the joint prior density of $\pi = \text{vec} \Pi$, conditionally on $\omega_1^2$, is a multivariate Normal density with a fully flexible covariance matrix which, as intended, can accommodate arbitrary linear restrictions. That it is conditional on the $\omega_1^2$'s represent the price we have to pay for partially preserving the operational characteristics of NC prior densities.

In some circumstances we might wish to relax the exact prior restrictions on $\pi$ by substituting for the singular distribution (3.11) a non-singular one that would be "informative" in the directions which correspond to the prior restrictions and relatively "non-informative" in the other directions. Comparisons between the corresponding constrained and unconstrained posterior densities could serve as the basis of a Bayesian test of the exact (overidentifying) restrictions on $\pi$ in the spirit of Florens and Mouchart (1987).

(ii) The NC prior density (3.1) belongs to our class of RENC prior densities provided we let $R_{(i-1)}^0$ depend linearly on $\lambda_1$. Applying recursively to the joint density (3.1) the usual decomposition formulae for the matrixmultivariate normal distribution, as described e.g. in the appendix of Drèze and Richard (1983), we find that the NC prior corresponds to the case where
\[ F_{i(1-1)}^O = \lambda_1' \otimes I_m \quad \text{for } i: 2 \to n \] (3.13)

(iii) Steel (1987a) studies the case where the conditional expectation of 
\[ \pi_1 | \pi_{(i-1)}, \lambda_1 \] is of the form

\[ E(\pi_1 | \pi_{(i-1)}, \lambda_1) = \pi_{i(1)}^O + (\lambda_1' \otimes I_m + R_i^O) (\pi_{(i-1)} - \pi_{(i-1)}^O) \] (3.14)

in which case the NC prior is obtained by setting \( R_i^O = 0 \). On the basis of
formula (3.12) and taking advantage of the full column rank conditions for \( S_{i1} \)
and \( S_{(i-1)} \) we find that (3.14) is compatible with our RENC formulation in
(3.6) under the following choice for \( F_{i(1-1)}^O \)

\[ F_{i(1-1)}^O = S_{1i}^+ \left( (\lambda_1' \otimes I_m + R_i^O) S_{(i-1)} - S_{i(1-1)} \right) \] (3.15)

We note that in the constrained case where \( S_{(i-1)} \) and \( S_{1i} \) are not square, the
inverse mapping from \( F_{i(1-1)}^O \) to \( R_i^O \) is not one-to-one. Hence the results in
Steel (1987a, section 3.3) formally apply to the unconstrained case and their
explicit extension to the constrained case requires reasoning in terms of \( \alpha \),
as in (3.6), instead of \( \pi \).

(iv) The ENC prior density, as surveyed in Drèze and Richard (1983) can
be reformulated as follows using our notation:

\[ D(\alpha | V) = f_N^L(\alpha | \alpha_o, (V^{-1} \Box \mathbb{M}_o)^{-1}) \] (3.16)

where the \( l \times l \) relative precision matrix \( \mathbb{M}_o \) is partitioned conformably with \( \alpha \)
into \( m^2 \) blocks \( M_{ij}^O \) and where the corresponding blocks in the precision matrix
\( V^{-1} \Box \mathbb{M}_o \) are defined to be
\[(V^{-1} \otimes M_0)_{ij} = v^{ij} M^0_{ij}\]  \hspace{1cm} (3.17)

with \(v^{ij}\) being the \((i,j)\)-th element of \(V^{-1}\). We could conceivably reformulate (3.16) as a special case of (3.6) at the cost of letting \(C^0_i\) and \(P^0_i(i-1)\) be functions of \(V^{-1}\). Doing so goes beyond the objectives of our present paper.

4. Posterior Analysis

4.1 Introduction

Having abandoned the NC formulation we already know that we are bound to face numerical integration on at least a subvector of the parameters. We shall restrict our attention to Monte Carlo integration based on importance sampling as discussed by Hammersley and Handscomb (1964, section 5.4) or, within an econometric framework by Kloek and van Dijk (1978). These techniques have proved increasingly successful for handling the class of problems with which we are dealing. Hence, our main concern in this section will be that of constructing importance functions, i.e. approximations to the densities to be analyzed from which coefficients can be drawn at random, preferably at a "reasonable" cost of computation.

We shall discuss in turn two different approaches for conducting the analysis of the posterior density on \(\alpha\) and \(V\).

The first approach starts by integrating out analytically \(V\), taking advantage from the fact that conditionally on \(\alpha\), the posterior density of \(V\) is either an Inverted-Wishart density or an obvious generalization thereof. This approach has occasionally been advocated on the grounds that \(\alpha\) is typically a parameter "of interest" while \(V\) is often treated as a "nuisance" parameters though there are cases where the latter is not true as, for example, in the context of an exogeneity analysis. The resulting posterior density on \(\alpha\) is
generally analytically untractable and we then face the task of constructing operational importance functions. Multivariate t densities or extensions thereof—such as split t’s in Geweke (1986) or poly-t’s in Bauwens (1984)—have often been used in related contexts. Experience suggests, however, that the posterior densities for the coefficients of systems of equations can be ill-behaved in several important aspects (fat tails, skewness, non unimodality,...) so that much remains to be done before Monte Carlo techniques can be routinely applied in this context.

The second approach proceeds in reverse order and initially takes advantage of the fact that, conditionally on V, \( \alpha \) is normally distributed. Hence marginal posterior densities for components of \( \alpha \), moments and/or fractiles can be analytically evaluated, conditionally on V. Numerical integration is now required to evaluate the posterior density of V and to marginalize results that were obtained conditionally on V. This approach has been used e.g. by Richard (1973) in the context of a Bayesian full-information analysis of two equation systems where the relative dimensionalities of \( \alpha \) and V make it attractive to restrict numerical integration to the posterior density of V. (In fact, the numerical analysis of the posterior density of V was then based on infinite expansions in terms of hypergeometric series but could nowadays be based instead on Monte Carlo techniques.) Also much of the later work on poly-t densities, as described e.g. in Richard and Tompa (1980) or Bauwens and Richard (1985) follows a similar route whereby poly-t densities are construed as marginal densities which are derived from conditional normal densities or mixtures thereof. Our practical experience in these situations seem to suggest that the posterior densities of V are typically better behaved than those of \( \alpha \) so that the task of designing importance functions is substantially less demanding under the second approach. This will be
illustrated in section 5 in the context of an application.

We now discuss the two approaches in turn. We restrict our attention to their main features, leaving out technical details that would only matter at the programming stage.

4.2 Recursive Monte Carlo Analysis of the Posterior Density of \( \alpha \)

The starting point of our analysis consists in introducing the coefficients \( \alpha \) in the expression of the recursive equations in (2.18).

Following definitions (2.19) and (2.26) we have

\[
y_1^* - X_{i-1}^\alpha - Y_{i-1}^\lambda  = y_1^* - X_{i-1}^\alpha - Y_{i-1}^\lambda
\]

with

\[
y_1^* = y_1^* - Xs_{i-1(i-1)}^\alpha, \quad X_{i-1}^\alpha = X_{i-1}^\alpha, \quad Y_{i-1}^\lambda = Y_{i-1}^\lambda - Xs_{i-1(i-1)}^\lambda
\]

Obviously \( \Pi_{i-1} \) itself is a function of \( \alpha_{i-1} \) since \( \pi_{i-1} = \text{vec} \Pi_{i-1} = S_{i-1}^{\alpha_{i-1}} \). The posterior density of \( \theta = (\alpha_1, \lambda_1, \omega_1^2); i: 1 \rightarrow n \) is then given by the product of the likelihood function (2.18) with the prior density (3.3).

We can essentially apply to each recursive equations in turn the standard formulae for the (single equation) regression model, as given e.g. in Zellner (1971, section 3.2), obtaining thereby a product of Normal-gamma densities.

However, in adopting the Normal-gamma formulation we implicitly add into the expression of the posterior density integrating constants that, in our problem, depend on the parameters \( \alpha_{i-1} \) for \( i > 1 \). These additional integrating constants have to cancel out in the expression of the joint posterior density of \( \theta \), which is, therefore, written as
$$D(\theta \mid Y) \propto K(\alpha_{(n-1)}; Y) \cdot J_1(\theta \mid Y)$$  \hspace{1cm} (4.3)$$

where $J_1(\theta \mid y)$ regroups the standard normal-gamma expressions, adapted to the specificities of formulae (3.6) and (4.1), while $K(\alpha_{(n-1)}; Y)$ regroups the reciprocal of the parts of the integrating constants that depend on $\{\alpha_{(i-1)}\}$. $K(\alpha_{(n-1)}; Y)$ and $J_1(\theta \mid Y)$ are specifically given by

$$K(\alpha_{(n-1)}; Y) = \prod_{i=2}^{n} \left| H_i^* \right|^{-1/2} \left( \frac{\omega_2^2}{\omega_{1,i}^*} \right)^{-(1/2)v_i^*}$$  \hspace{1cm} (4.4)$$

$$J_1(\theta \mid Y) = \prod_{i=1}^{n} \int_{\mathbb{R}} \left( \frac{\alpha_i^{o} m_i^*}{\lambda_i^{o} H_i^*} \right) f_{i,Y}(\omega_i^2, \omega_i^2, v_i^*)$$  \hspace{1cm} (4.5)$$

where

$$H_i^* = \begin{pmatrix} G_i^{o} + X_{i,1}^{*} & X_{i,1-i}^{*} \\ Y_{i-1,1}^{*} & L_i^{o} + Y_{i-1,1}^{*} \end{pmatrix}$$  \hspace{1cm} (4.6)$$

$$H_i^{o,m*} = \begin{pmatrix} G_i^{o} \left[ \alpha_i^{o} + P_i^{o} (\alpha_{(i-1)} - \alpha_{(i-1)}^{o}) \right] X_{i,1}^{*} \\ L_i^{o} + Y_{i-1,1}^{*} \end{pmatrix}$$  \hspace{1cm} (4.7)$$

$$\omega_1^2 - \omega_2^2 + \left[ \frac{\alpha_i^{o} + P_i^{o} (\alpha_{(i-1)} - \alpha_{(i-1)}^{o})}{\lambda_{1,1}^{o} H_i^{o} m_i^{*}} \right] G_i^{o} \left[ \alpha_i^{o} + P_i^{o} (\alpha_{(i-1)} - \alpha_{(i-1)}^{o}) \right]$$

$$+ \lambda_{1,i}^{o} L_i^{o} + Y_{i-1,1}^{*} - m_i^{*} H_i^{o} m_i^{*}$$

$$v_i^* - v_i^{o} + T$$  \hspace{1cm} (4.8)$$

It is precisely the dependence of $K(\alpha_{(n-1)}; Y)$ on $\alpha_{(n-1)}$ that prevents analytical integration of the posterior density (4.3) with respect to $\alpha_{(n-1)}$.\[23\]
Marginalization of the posterior density with respect to \((\lambda_i, \omega_i^2); i: 1 \rightarrow n\) is straightforward and we obtain

\[
D(\alpha \mid Y) \propto K(\alpha^{(n-1)}; Y). I_1(\alpha^{(n)} \mid Y) \tag{4.10}
\]

with

\[
I_1(\alpha \mid Y) = \prod_{i-1}^n \int_1^{\mathfrak{l}_1} \left( \alpha_i, \omega_i^2, \mu_i^*, \sigma_i^2, \nu_i^* \right)
\]

(4.11)

where \(m_{11}^*\) regroups the first \(\ell_1\) elements of \(m_1^*\), and

\[
G_1^* = \sigma_1^2 \cdot X_1^*X_1^* - X_1^*Y_1^* (I_1^0 + Y_1^*Y_1^*)^{-1} Y_1^*X_1
\]

(4.12)

Note that we can still analytically integrate the posterior density with respect to \(\alpha_n\) since only \(I_1(\alpha \mid Y)\) depends on \(\alpha_n\). It is, however, unclear that doing so reduces the cost of computation in any substantial way since whatever moments are obtained at this stage for \(\alpha_n\) are conditional on \(\alpha^{(n-1)}\) and, hence, have yet to be marginalized with respect to \(\alpha^{(n-1)}\). Therefore, we only discuss here Monte Carlo integration with respect to the entire vector \(\alpha\).

Steel (1987a) discusses a number of ways in which importance functions can be constructed for the posterior density (4.3) in the case where, as the result of assumption (3.18), our model can be reparameterized in such a way that the counterpart of \(H_1^*\) in (4.6) does not depend anymore on \(\alpha^{(n-1)}\).

We can extend his analysis to the case under consideration here.

\((^5)\) Specifically, under assumption (3.4) we find that the parameters \(\lambda_i\) and, say, \(\xi_i = \pi_i(\Pi_i - \Pi_i^0)\lambda_i\) are a priori independent. Equation (4.1) is then rewritten as \(y_i - X_i\lambda_i = y_i - X_i\lambda_i - X_i\lambda_i\), where \(\xi_i\) has the same form as \(H_i^*\) in (4.1), except that \(Y^*_i\) is replaced by \(Y^0_i\), which does not depend on \(\Pi_i\). This suggests drawing sequentially from the posterior distribution of \((\xi_i, \lambda_i \mid \Pi_i)\) rather than from that of \((\pi_i, \lambda_i \mid \Pi_i)\)
Firstly, we can choose \( I_1(\alpha | y) \) itself as the importance function. It offers the advantage that random drawings of \( \alpha \) are easy to generate from the factorization in (4.11). A drawing of the entire vector \( \alpha \) requires \( n \) successive rounds, of which the \( i \)-th consists of drawing \( \alpha_i \) from the corresponding \( t \) density in (4.11), conditionally on \( \alpha_{(i-1)} \) being set at the value drawn in the \( i-1 \) first rounds (the drawing of \( \alpha_1 \) in the first round is unconditional). Two remarks ought to be made at this stage of the discussion:

(i) The ratio between the posterior density (4.10) and the importance function \( I_1(\alpha | y) \) is given by \( K(\alpha_{(n-1)}; y) \), as defined in (4.4). It includes factors whose (sampling) variance might be large in which case the convergence of the Monte Carlo procedure would be slow. The applications in Bauwens (1984) where similar techniques were experimented with in the context of SEM's and the application we discuss in section 5 below confirm that impression.

(ii) The hyperparameters \( H_i^* \) depend on \( \alpha_{(i-1)} \) for \( i > 1 \) and, hence, have to be recomputed for every new drawing. Efficient ways of organizing the computations are discussed in Appendix B.

The importance function \( I_1(\alpha | y) \) has to be modified if we replace the conditional (relative to \( \omega_i^2 \)) normal density (3.6) by the unconditional \( t \) density in (3.8). It essentially amounts to replacing the individual \( t \) densities in (4.11) by 2-0 poly-\( t \) densities, one of the two \( t \) kernels being that of the prior density (3.8) and the other being the one associated with the hyperparameters in formulae (4.6) through (4.9), where \( G_i^0 \) is set equal to zero. Let \( I_1^*(\alpha | y) \) denote the correspondingly modified importance function.

Note that the hyperparameters of the prior \( t \) kernel do not depend on \( \alpha_{(i-1)} \) and, hence, do not have to be reevaluated for each new drawing. Also setting \( G_i^0 \) equal to zero in the other kernel results in substantial simplifications. It follows that using \( I_1^*(\alpha | y) \) as importance function instead of \( I_1(\alpha | y) \) should
not increase substantially the cost of computation. Furthermore, even if we used the prior (3.8), we could still consider using $I_1(\alpha \mid Y)$ as the importance function, since the difference between (3.6) and (3.8) might not matter much as long as the prior and the sample information on $(\omega_1^2 \ldots \omega_n^2)$ are not conflicting.

Finally there are ways of trying to improve upon the design of the importance function itself, essentially by trying to incorporate within its expression additional factors taken from the function $K(\alpha_{(n-1)}; Y)$. For example, in the case considered by Steel (1987a) where $H_i^*$ does not depend on $\alpha_{(i-1)}$, $\omega_i^{2*}$ is quadratic in $\alpha_{(i-1)}$ and may then be reformulated as a quadratic form in $\alpha_{i-1}$ conditionally on $\alpha_{(i-2)}$, say

$$\omega_i^{2*} = q_i^*(\alpha_{i-1} \mid \alpha_{(i-2)})$$  \hspace{1cm} (4.13)

Let then

$$I_2(\alpha \mid Y) = I_1(\alpha \mid Y), \left\{ \prod_{i=2}^n \left[ q_i^*(\alpha_{i-1} \mid \alpha_{(i-2)}) \right]^{-(1/2)} v_i^* \right\}$$  \hspace{1cm} (4.14)

Using $I_2(\alpha \mid Y)$ as the importance function requires drawing $\alpha_i$ conditionally on $\alpha_{(i-1)}$ from 2-0 poly-t densities, for $i$: 2 $\rightarrow$ n, whereas under $I_1(\alpha \mid Y)$ they were drawn from t densities. Algorithms for generating random drawings from 2-0 poly-t densities are byproducts of the analysis in Dickey (1968) or in Richard and Tompa (1980, formulas [4.3] to [4.7]).

The situation where $H_i^*$ depends on $\alpha_{(i-1)}$ is more complicated to handle and is briefly discussed in Appendix B.

4.3 Monte Carlo Analysis of the Posterior Density of $V$

The approach we now discuss consists in deriving analytical results for $\alpha$
conditionally on \( V \) and then using Monte Carlo integration on \( V \). The conditional posterior density of \( \alpha \) given \( V \) is proportional to the product of the likelihood function (2.7) with the prior density (3.9). For the ease of notation let \( M_o \) denote the (conditional) precision matrix in (3.9)

\[
M_o = (I - F_o)D_0^{-1}D_\omega (I - F_o)
\]  

(4.15)

Standard algebraic manipulations similar to those described in Zellner (1971) for the analysis of the single equation regression model lead to the following expression for the conditional posterior density of \( \alpha \)

\[
D(\alpha \mid V, Y) \propto f_N^L(\alpha \mid \alpha^*_x, M^{-1}_x) 
\]  

(4.16)

where

\[
M_x = M_o + S'(V^{-1} \otimes X'X)S 
\]  

(4.17)

\[
\alpha_x = M_x^{-1} \left[ M_o \alpha_o + S'(V^{-1} \otimes X'y) \right] 
\]  

(4.18)

We note that both \( M_x \) and \( \alpha_x \) depend on \( V \).

The posterior density of \( V \) regroups all the factors which depend on \( V \) in the product of the prior density with the likelihood function, excluding those already in the expression of the conditional density (4.16). For notational convenience we restrict our attention to the case where the prior density functions in (3.4) and (3.5) all originate from the joint Inverted-Wishart density in (3.7) but the extension to the more general case is conceptually straightforward. The posterior density of \( V \) is then given by
\[ D(V \mid Y) \propto |M_o|^{-1/2} |M_\kappa|^{-1/2} \exp -1/2(\alpha_o M_o \alpha_o - \alpha_\kappa M_\kappa \alpha_\kappa) f_{\text{IW}}^n(V \mid V_*, v_*) \] (4.19)

where \( v_* = v_o + T \) and

\[ V_* = V_o + Y'Y \quad \text{and} \quad |M_o|^{-1} = |B_o|^{-1} \alpha \prod_{i=1}^n (\omega_i^2)^{\lambda_i} \] (4.20)

Of course, under the NC prior densities (3.1) and (3.2) we find that \( M_o = V^{-1} \otimes N_o \) and \( M_\kappa = V^{-1} \otimes N_\kappa \) with \( N_\kappa = N_o + X'X \). The posterior density (4.19) then simplifies into the well-known expression

\[ \tilde{D}(V \mid Y) = f_{\text{IW}}^n(V \mid \tilde{V}_*, \tilde{v}_*) \] (4.21)

where

\[ \tilde{V}_* = V_* + \prod_{i=0}^{\kappa} N_o \Pi_\kappa - \Pi_\kappa' N_\kappa' \Pi_\kappa \] (4.22)

\[ \Pi_\kappa = N_\kappa^{-1}(N_o \Pi_0 + X'Y) \] (4.23)

One is, therefore, naturally led to consider using an Inverted-Wishart (IW) density of the form (4.21) as an importance function for the more general posterior density in (4.19). An additional (heuristic) reason for doing so lies in the fact that the posterior density of \( V \) need not be extremely sensitive to the choice of a prior on \( \Pi \), at least as long as the latter does not conflict in major ways with the sample information relative to \( \Pi \).

Furthermore, we can easily iterate on the choice of an IW importance function. Specifically let us assume that at the \( j \)-th step of such an iterative procedure we are using an IW importance function with hyperparameter \( \tilde{V}_k^j \).

Based on a limited number of drawings we then compute a rough estimate of the posterior mean of \( V \) under the (exact) RENC prior under consideration. Let
\( E_k^1(V) \) denote that estimate. We may then choose the next step hyperparameter as

\[
\tilde{V}_{k+1}^* = (\nu_k - n - 1) E_k^1(V) \tag{4.24}
\]

The use of this iterative procedure is illustrated in section 5 below in the context of an application. Steel (1987a) offers two suggestions for the choice of an initial value \( \tilde{V}_k^O \). A first initial value, say \( \tilde{V}_k^{FI} \), is obtained by application of formula (4.22) where \( \Pi_o \) and \( V_o \) are set at their RENC values while \( N_o \) is given by the arithmetic mean of the individual relative precision matrices \( G_{1i}^O \) in (3.16), which are enlarged into \( mm \times mm \) conformable matrices, say

\[
N_o^{FI} = \frac{1}{n} \sum_{i=1}^{n} S_i^{-1} G_{1i}^O S_i^{-1} \tag{4.25}
\]

A justification for this enlargement procedure is provided in Appendix C within the context of a single equation regression model. Steel next suggests factorizing \( \tilde{V}_k^{FI} \) as in (2.21) into, say

\[
\tilde{V}_k^{FI} = (\lambda_k^{FI})^{-1} D_{\omega_k}^{FI} (\lambda_k^{FI})^{-1} \tag{4.26}
\]

and replacing the \( \omega_k^2 \)'s in \( D_{\omega_k}^{FI} \) by their equivalents as derived from a single equation analysis of each constrained reduced form equation. Let \( D_{\omega_k}^{LI} \) denote the corresponding transformation of the matrix \( D_{\omega_k}^{FI} \). The corresponding initial value is then defined as

\[
\tilde{V}_k^{FI/LI} = (\lambda_k^{FI})^{-1} D_{\omega_k}^{LI} (\lambda_k^{FI})^{-1} \tag{4.27}
\]

These two choices for the initial value of \( \tilde{V}_k \) have been used with success in the context of the application we discuss in section 5 below. It seems,
however, that on the basis of the discussion in Appendix C we might prefer averaging the individual covariance matrices in (3.12) instead of the precision matrices. The corresponding precision matrix, say \( \tilde{N}_o^{FI} \), is then defined as

\[
\tilde{N}_o^{FI} = \left( \frac{1}{n} \sum_{i=1}^{n} S_{i1} G_{i}^{-1} S_{i1}' \right)^{-1}
\]  

(4.28)

Though it now seems to us that \( \tilde{N}_o^{FI} \) might provide us with a better initial value for \( \tilde{V}_x \) than \( N_o^{FI} \) in (4.25) does, we have not yet tested this conjecture in the context of an application such as the one we now discuss.

5. An Application to a System of Wage Equations

5.1 The Programs

Before engaging into a major programming effort aiming at providing practitioners with a user friendly and numerically reliable computer program, we wish to investigate first which of the options we have discussed seems to perform best on a range of pilot applications. The motivation for doing so is obvious: the Bayesian analysis of systems of equations has proved quite challenging and, over the years, many heuristically attractive procedures have often failed to perform as hoped for.

At this exploratory stage programming has been kept extremely simple and highly modular so that no attempts have been made to organize computations efficiently. The test program is written in APL but has to access the Fortran subroutines that are available for drawing random numbers from poly-t densities. Each new drawing is treated completely independently from the others. These simplifying options have resulted in a test program that is extremely inefficient and, hence, computing time is largely irrelevant. Based
on a rough evaluation of the impact of a fully efficient program and on previous experience in related contexts—see e.g. Bauwens (1984)—we expect an efficient program to require of the order of 1/10 of a second or less per drawing in the application we describe below (with 12 coefficients in $\alpha$ and 10 in $V$). The computations are run on a Data General MV8000 minicomputer equipped with a floating point accelerator.

5.2 The Application

The object of this section is to briefly report on the results of a pilot application. The interested reader is referred to Steel (1987a, chapters 5 and 6) for additional details and results as well as for other applications. Since several of the options we have developed in the present paper have not yet been incorporated in our pilot program, the analysis is conducted under the RENC prior specification (3.14). As discussed in sections 3.2 and 4.2 this choice complicates the selection of the prior hyperparameters but simplifies the recursive Monte Carlo analysis of the posterior density of $\alpha$. Since, however, the information content of the prior densities we use here is essentially that which is associated with exact prior restrictions (in the sense that we are only vaguely informative on $\alpha$) we do not expect the posterior results to be sensitive to the choice of a specific RENC prior specification. Hence, the results we report here should be fairly representative of what we expect to find under alternative formulations of the RENC prior.

The application is taken from Bauwens (1984) and consists of four seemingly unrelated equations describing real wage formation in four different EEC countries (Germany: DB; France: FR; the Netherlands: NL and Belgium: BE). In each country the growth rate of real wage is explained by the growth rates
of domestic prices and productivity and by the country's own rate of unemployment. We use yearly data for the period 1955 to 1976 and, unlike Bauwens (1984), constant terms are deleted so that autonomous real wage growth is excluded. The regressors being specific to each country, the prior exact restrictions are within equation and consists of 9 distinct exclusion restrictions per equation. Hence, it is obvious that a NC prior density would be totally inappropriate in this context.

The prior first and second order moments of $\alpha$ are chosen as follows. The prior means are set equal to 0.5 for the productivity coefficient, -0.5 for the unemployment variable and zero for the price coefficients. The prior variances are set to 1 and the covariances to 0.5. Other choices could equally be considered at this exploratory stage but we do not intend here to conduct a sensitivity analysis with respect to the choice of a non-dogmatic prior on $\alpha$. We wish instead to emphasize the fact that RENC priors can incorporate prior information that is specific to individual equations (in the limit form of our 36 exclusion restrictions).

We shall use an Inverted-Wishart prior density on $\Sigma$ and apply an elicitation procedure that has already proved successful in related contexts—see e.g. the application in Lubrano et al. (1986). Specifically we equate the prior mean of $\Sigma$ to a certain proportion of the unconditional sample covariance matrix of the dependent variables, the proportion being meant to represent our prior belief regarding the overall expected fit of the system. As discussed in Appendix D, this procedure does not aim at selecting a "data-based" prior. It is meant instead, to approximate a piece of genuine prior information whose exact treatment would require numerical integration. Also, our main objective here is to protect ourselves against the risk of selecting relatively diffuse priors on $\Sigma$ thereby, as discussed in section 3.2, implicitly giving too much
weight to the (RE)NC prior on $\alpha$. Three different values of the prior expectation of $V$ are considered here. $E_0^{40}$ and $E_0^{60}$ represent respectively 40% and 60% of the sample covariance matrix of the $y$'s while $D_0^{40}$ is a diagonal matrix whose diagonal elements are roughly taken from $E_0^{40}$. The "degrees of freedom" hyperparameter $v_0$ is chosen on the basis of the size of the "hypothetical" sample which is implicit in (RE)NC prior densities. We have mainly used such values at 10, 20 or 30 (the effective sample size is 22).

The results which are reported in Table 1 below consist of the posterior means and standard deviations for the 12 coefficients in $\alpha$ (in the order price, productivity and unemployment for each country) together with estimated relative error bounds, expressed in percentage points and calculated at a 95% confidence level, for the posterior means ($\epsilon_\alpha$) and for the integrating constant ($\epsilon_{\text{inc}}$). See e.g. Kloek and van Dijk (1978) for a discussion of these error bounds.

The recursive approach on $\alpha$, as described in section 4.2, has generally proved very unsatisfactory. The error bounds are large and often increase with the number of drawings. Column 1 in Table 1 describe a typical outcome corresponding to a joint t prior density\(^{(6)}\) for $\alpha$ with 20 degrees of freedom and evaluated under the importance function $I_1^T(\alpha \mid Y)$ as described in section 4.2. Trial runs including up to 20,000 drawings have resulted in similar outcomes, even when $I_1^T(\alpha \mid Y)$ is replaced by the heuristically more attractive $I_2(\alpha \mid Y)$. The problem seems to be that these importance functions fail to

---

\(^{(6)}\) Since the product of the conditional RENC densities in (3.8) does not define a joint t density, the RENC prior serves here only as an approximation to the joint t prior and is instrumental in the construction of importance functions that are then applied to the genuine posterior density.
### Table 1: Posterior Results on Four Wages Equations

<table>
<thead>
<tr>
<th>Country</th>
<th>( E(V) )</th>
<th>( E_0 )</th>
<th>Mean s.d.</th>
<th>( E_\alpha )</th>
<th>Mean s.d.</th>
<th>( E_\delta )</th>
<th>Mean s.d.</th>
<th>( E_\delta )</th>
<th>Mean s.d.</th>
<th>( E_\delta )</th>
<th>Mean s.d.</th>
<th>( E_\delta )</th>
<th>Mean s.d.</th>
<th>( E_\delta )</th>
<th>Mean s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
</tr>
<tr>
<td></td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td>not applicable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2500</td>
<td>278.1</td>
<td>1500</td>
<td>14.4</td>
<td>2000</td>
<td>24.0</td>
<td>2000</td>
<td>20.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB</td>
<td>0.52 0.17 55.7</td>
<td>0.61 0.23 4.9</td>
<td>0.58 0.21 2.6</td>
<td>0.63 0.25 10.4</td>
<td>0.62 0.25 8.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.73 0.08 17.4</td>
<td>0.66 0.13 1.2</td>
<td>0.68 0.12 0.6</td>
<td>0.66 0.13 1.3</td>
<td>0.68 0.13 1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.31 0.13 65.5</td>
<td>-0.46 0.41 9.4</td>
<td>-0.40 0.38 6.1</td>
<td>-0.49 0.44 20.9</td>
<td>-0.48 0.45 19.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR</td>
<td>0.32 0.08 43.9</td>
<td>0.29 0.19 1.1</td>
<td>0.28 0.18 0.7</td>
<td>0.29 0.18 1.3</td>
<td>0.31 0.19 2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.57 0.12 17.6</td>
<td>0.50 0.28 2.8</td>
<td>0.59 0.26 1.7</td>
<td>0.61 0.27 4.1</td>
<td>0.61 0.26 2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.21 0.20 82.4</td>
<td>-0.16 0.72 19.5</td>
<td>-0.13 0.68 14.4</td>
<td>-0.18 0.77 28.7</td>
<td>-0.18 0.78 25.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>0.38 0.09 43.9</td>
<td>0.43 0.16 1.1</td>
<td>0.43 0.15 0.8</td>
<td>0.43 0.16 2.1</td>
<td>0.44 0.16 1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.71 0.05 8.2</td>
<td>0.82 0.16 0.5</td>
<td>0.83 0.14 0.5</td>
<td>0.82 0.16 0.7</td>
<td>0.84 0.15 0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.84 0.28 48.0</td>
<td>-0.55 0.47 2.5</td>
<td>-0.55 0.45 2.1</td>
<td>-0.56 0.50 4.9</td>
<td>-0.59 0.47 4.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>0.36 0.07 32.8</td>
<td>0.29 0.17 1.7</td>
<td>0.28 0.16 1.3</td>
<td>0.30 0.18 3.2</td>
<td>0.32 0.18 2.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.30 0.06 19.0</td>
<td>0.62 0.27 2.5</td>
<td>0.59 0.24 1.8</td>
<td>0.63 0.28 3.1</td>
<td>0.64 0.28 3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.24 0.15 62.9</td>
<td>0.08 0.46 32.9</td>
<td>0.16 0.40 12.4</td>
<td>0.07 0.49 52.6</td>
<td>0.04 0.49 119.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
1. For each country, the coefficients are in the order: price, productivity and unemployment.
2. The error bounds are calculated at a 95% confidence level and expressed in percentage
   \( \varepsilon_\alpha \) for the posterior means and \( \varepsilon_\text{int} \) for the integrating constants
capture the tail behavior of the posterior density—whether the RENC prior serves as the genuine prior or only as an approximation of the prior.

In contrast the approach based on a Monte Carlo analysis of the posterior density of $V$, as described in section 4.3, has proved much more satisfactory under a range of prior assumptions on $V$. Some results are reported in columns 2 to 5 of table 1 under a genuine RENC prior for $\alpha$. The iterations on the choice of $\tilde{V}_\alpha$ were based on 200 drawings each. The number of such iterations, based here on the initial condition (4.27) ran from 3 to 5. Hence the total number of drawings lies between 2000 and 3000, compared to the 10,000 drawings used in Bauwens (1984). (The results in Bauwens are based on data expressed in deviations from their sample means and, hence, are not directly comparable to the results in table 1). The results are largely satisfactory at the light of the following comments:

(i) The error bounds $\epsilon_{\text{int}}$ are largely irrelevant if we are only interested in the posterior moments of $\alpha$ (and $V$);

(ii) As discussed e.g. by Kloek and van Dijk (1978), the error bounds $\epsilon_\alpha$ are "conservative" in the sense that they do not take into account the typically positive correlation between the estimators of the numerator and the denominator in the expression of the posterior means. This is confirmed by the fact that we have found substantially less variation in the posterior means across replications of similar runs of computations than the $\epsilon_\alpha$'s seems to indicate;

(iii) The $\epsilon_\alpha$'s are relative error bounds and some of their largest values are naturally associated with small posterior means. The error of approximation is always much smaller than the corresponding posterior standard deviation and is generally negligible when the
posterior density under consideration is "sharp".

Additional results as well as other pilot applications are found in Steel (1987a). They all confirm the marked superiority of the "V approach" on the "α approach."

6. Conclusions

We have proposed a class of RENC prior densities for SURE models. These priors are fully flexible in the sense that they can accommodate arbitrary joint t densities on the regression coefficients as well as exact linear restrictions on these coefficients. The prior density on the covariance matrix V is typically an Inverted-Wishart but recursive generalizations of it can be handled. Compared to other classes of RENC prior densities that are available in the literature, ours offers the advantage that simple analytical expressions exist for the first and second order prior moments.

We have experimented with two basic procedures for analyzing the posterior densities of α and V. Our first approach combines analytical integration on the elements of V with numerical recursive Monte Carlo integration on α. In the second approach we proceed in inverse order and use Monte Carlo integration on V based on Inverted-Wishart importance functions (or on obvious generalizations thereof).

Applications have demonstrated the superiority of the second approach over the first one. Furthermore the second approach appears to produce stable results with relatively small numbers of drawings compared to other procedures. Additional work is needed regarding, in particular, the choice of initial values for the hyperparameter $\tilde{V}_\alpha$ of the importance function, for selecting the number of drawings of the auxiliary iterations and for choosing a convergence criterion. Altogether the preliminary results we report here
appear promising. We might well be on our way towards designing a flexible, easily applicable and fully automatic Monte Carlo procedure for the Bayesian analysis of restricted SURE models, a long awaited for development.
Appendix A: Notation for Density Functions

We have regrouped here the expressions for the density functions that appear in the course of our paper. For more details, properties and proofs see, for example, the appendices in Zellner (1971) or Drèze and Richard (1983). $\mathcal{G}^p$ denotes the set of pxp symmetric positive definite matrices

Inverted-gamma-2

$$f_{1\gamma}(\sigma | s^2, v) = \left[ \Gamma\left( \frac{v}{2} \right) \right]^{-1} \left[ \frac{s^2}{2} \right]^{1/2v} (\sigma^2)^{-1/2(v+2)} \exp \left[ -\frac{1}{2} \frac{s^2}{\sigma^2} \right]$$

$s^2 > 0, \sigma^2 > 0$ and $v > 0$

Multivariate Normal

$$f_{\mathcal{N}}(x | \mu, \Sigma) = (2\pi)^p |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2} (x-\mu)' \Sigma^{-1} (x-\mu) \right]$$

$x \in \mathbb{R}^p$, $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathcal{G}^p$

Multivariate t

$$f_{t}(x | \mu, \Pi, v) = \left[ \pi^{-1/2p} \Gamma\left( \frac{v+p}{2} \right) / \Gamma\left( \frac{v}{2} \right) \right]^{1/2} |\Pi|^{1/2} \left[ 1 + (x-\mu)' \Pi^{-1} (x-\mu) \right]^{-1/2(v+p)}$$

$x \in \mathbb{R}^p$, $\mu \in \mathbb{R}^p$, $\Pi \in \mathcal{G}^p$ and $v > 0$

Inverted-Wishart

$$f_{\text{IW}}(\Sigma | s, v) = \left[ 2^{1/2vq} \pi^{1/4q(q-1)} \prod_{i=1}^{q} \Gamma\left( \frac{v+1-q}{2} \right) \right]^{-1} \left[ s \right]^{1/2v} |\Sigma|^{-1/2(v+q+1)} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} s \right]$$

$\Sigma \in \mathcal{G}^q, s \in \mathcal{G}^q$ and $v > q - 1$

Matrix-variate Normal

$$f_{\text{MN}}(x | \Sigma, P) = (2\pi)^{pq} |\Sigma|^\frac{1}{2} |P|^\frac{1}{2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} (x-\bar{x})' P^{-1} (x-\bar{x}) \right]$$

$x \in \mathbb{R}^{pq}$, $\bar{x} \in \mathbb{R}^{pq}$, $\Sigma \in \mathcal{G}^q$ and $P \in \mathcal{G}^p$
Appendix B: Importance sampling when \( \beta_{i}^{*} \) depends on \( \alpha_{i-1} \)

Note first that \( \beta_{i}^{*} \) depends on \( \alpha_{i-1} \), only in its i - 1 last rows and columns. We can take advantage of this property and substantially reduce the cost of computation. Specifically, let \( \Phi_{i}^{*} \) be the following square matrix of order \( \ell_{1} + 1 \).

\[
\Phi_{i}^{*} = \begin{pmatrix}
\beta_{i}^{*} & \beta_{i}^{*} \\
\beta_{i}^{*} & \beta_{i}^{*} + \omega_{i}^{*} \beta_{i}^{*}
\end{pmatrix}
\]  \hspace{1cm} (B.1)

Consider next its Cholesky decomposition into \( \Phi_{i}^{*} = BB' \) where \( B \) is a lower triangular matrix which can be partitioned conformably with \( \Phi_{i}^{*} \) into

\[
B = \begin{pmatrix}
b_{11} & 0 \\
b_{12} & b_{22}
\end{pmatrix}
\]  \hspace{1cm} (B.2)

where \( b_{11} \) is of order \( \ell_{1} + 1 \), \( b_{12} \) and \( b_{22} \) is a non-negative scalar. The hyperparameters of the posterior distribution (4.10) are then given by

\[
\beta_{i}^{*} = b_{11} \beta_{11}, \quad \beta_{i}^{*} = b_{11} \beta_{12} \quad \text{and} \quad \omega_{i}^{2} = b_{22}^{2}
\]  \hspace{1cm} (B.3)

The first \( \ell_{1} \) rows and columns of \( B_{11} \) do not depend on \( \alpha_{i-1} \) so that each new drawing only requires adjusting the last \( i \) rows of \( B \), independently of the number of components in \( \alpha_{i} \). We note that, at this stage, we might probably as well consider drawing jointly the \( \alpha \)'s and the \( \lambda \)'s instead of just the \( \alpha \)'s especially if the \( \lambda \)'s are themselves parameters of interest, e.g. in the context of analyzing whether or not a SURE model is "block-recursive." We note also that \( X_{1}^{*}Y_{1-1}^{*} \) and \( Y_{1-1}^{*} \) may be rewritten as

\[
X_{1}^{*}Y_{1-1}^{*} = S_{1}^{*}X_{1}^{*} (\hat{\Pi}_{1-1} - \Pi_{1-1})
\]  \hspace{1cm} (B.4)

\[
Y_{1-1}^{*}Y_{1-1}^{*} = Y_{1-1}^{*} M_{1} Y_{1-1}^{*} + (\hat{\Pi}_{1-1} - \Pi_{1-1}) X_{1}^{*} (\hat{\Pi}_{1-1} - \Pi_{1-1})
\]  \hspace{1cm} (B.5)

where \( \hat{\Pi}_{1-1} = (X'X)^{-1}X'Y_{1-1} \) and \( M_{1} = I_{T} - X(X'X)^{-1}X' \). This suggests that we ought to be able to reduce further the cost of computation by introducing additional simplifications in the design of the importance function. We might, for example, abandon revising \( \beta_{i}^{*} \) for each new drawing and instead fix \( \Pi_{i-1} \) at some initial estimate of its posterior mean. Or we could use a first order Taylor approximation of the matrix \( B \) with respect to the differences \( \hat{\Pi}_{i-1} - \Pi_{i-1} \). We do not intend to elaborate further here on such potential simplifications because the approach we have just described appears, presently at least, to be dominated by the alternative we discuss in section 4.3.

Regarding the design of \( I_{2}(\alpha | Y) \) in (4.14) we note that, if \( \beta_{i}^{*} \) depends on \( \alpha_{i-1} \), then \( \omega_{i}^{*} \) is no longer quadratic in \( \alpha_{i-1} \). In fact, following formula (4.13), \( \omega_{i}^{*} \) can be rewritten as
\[
\omega_1^2 \left| \Phi_1^* \right. \left| H_1^{* \perp} \right.
\]

Conditionally on \( \alpha_{(i-2)} \), \( |H_1^{*}| \) is quadratic in \( \alpha_{i-1} \) but \( |\Phi_1^*| \) is of degree four in \( \alpha_{i-1} \). Hence, if we use a (local) quadratic approximation for \( |\Phi_1^*| \), we may then construct an importance function \( I_3(\alpha \mid Y) \) that now requires drawing \( \alpha_i \), conditionally on \( \alpha_{(i-1)} \), from 2-1 poly-t densities. Algorithms for drawing from 2-1 poly-t densities or, at a lower cost, from 1-1 poly-t approximations thereof, are proposed in Bauwens and Richard (1985).

Appendix C: Exact Restrictions and the Use of the Moore-Penrose Inverse

Consider the standard regression model

\[
y = X \beta + u \quad u \sim N(0, \sigma^2 I_n)
\]  

(\text{G.1})

We assume further that \( \beta \) is subject to exact linear restrictions. We can transform the \( X \)'s and partition \( \beta \) into \( \beta_1 \) and \( \beta_2 \) in such a way that the restrictions read as \( \beta_2 = \beta_{20} \). Let then the (conditional) density on the unrestricted coefficients \( \beta_1 \) and \( \sigma^2 \) be

\[
D(\beta_1 \mid \sigma^2) = f_N^k(\beta_1 \mid \beta_{10}, \sigma^2 I_n^{-1})
\]  

(\text{G.2})

\[
D(\sigma^2) = f_{17}(\sigma^2 \mid \sigma_{0}, v_0)
\]  

(\text{G.3})

It is well-known that the posterior densities belong to the same class of densities, their hyperparameters being updated as follows

\[
N_x = N_o + \sum_{i=1}^N y_i X_i \quad N_x \beta_\ast = N_o \beta_\ast + \sum_{i=1}^N (y_i - X_i \beta_{20})
\]

(\text{G.4})

\[
\sigma^2 + \beta_\ast N_x \beta_\ast = \sigma^2 + \beta_\ast N_o \beta_\ast + \sum_{i=1}^N (y_i - X_i \beta_{20})' (y_i - X_i \beta_{20})
\]

(\text{G.5})

The prior density (G.2) can be represented in the \( \beta \)-space by a singular normal distribution with zero variances for \( \beta_2 \) and, hence, arbitrarily large values for the corresponding elements of the joint precision matrix. This would, however, clearly be inappropriate for such purposes as that of averaging precision matrices across equations as in formula (4.25). Let us, therefore, consider what happens if we use instead the Moore-Penrose inverse of the joint singular covariance matrix so that the densities in (G.2) and (G.3) are reinterpreted as

\[
D(\beta \mid \sigma^2) = f_N^k(\beta \mid \beta_o, \sigma^2 I_n^o)
\]

(\text{G.6})

\[
D(\sigma^2) = f_{17}(\sigma^2 \mid \sigma_{o}, v_o - k_2)
\]

(\text{G.7})

with
\[
M_o = \begin{bmatrix} N_0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad M_o^+ = \begin{bmatrix} N_o^{-1} & 0 \\ 0 & 0 \end{bmatrix}
\] 
(C.8)

Note the correction of degrees of freedom in (C.7) reflecting the fact that \(D(\sigma^2)\) in (C.3) is implicitly conditional on \(\beta_2 = \beta_{20}\) while it is marginal in (C.7). The hyperparameters in (C.6) and (C.7) are updated as follows:

\[
M_x = M_o + X'X \quad \quad M_x b_x = M_o \beta_o + X'y
\] 
(C.9)

\[
s_x^2 + b_x M_x b_x = \sigma_o^2 + \beta_o M_o \beta_o + y'y
\] 
(C.10)

We can now condition this joint posterior density on \(\beta_2 = \beta_{20}\). It is easily shown that the hyperparameters of that conditional distribution coincide with those in (C.4) and (C.5), i.e. that

\[
M_{11} = N_x, \quad b_1 = M_{11}^{-1} N_x (\beta_{20} - \beta_{2*}) - \beta_1
\] 
(C.11)

\[
(u_x - k_2) + k_2 = u_x \quad \text{and} \quad s_x^2 + (\beta_{20} - \beta_{2*}) M_{22.1} (\beta_{20} - \beta_{2*}) = \sigma_{x*}^2
\] 
(C.12)

where notations are standard and self-explanatory. Hence, reasoning in the joint parameter space using the Moore-Penrose inverse of the (partially non-informative) covariance matrix in (C.6) leaves unaffected the conditional posterior distribution of the parameters of interest (it obviously affects the distribution of the parameters subject to the restrictions).

Appendix D: Prior information on residual variances

Consider first the case of a single equation

\[
y = X\beta + u, \quad u \sim N(0, \sigma^2 I_T)
\] 
(D.1)

If we approximate the covariance matrix of the x's by its sample estimate, say \(Q\), the ratio between the conditional and the unconditional variances of \(y\) is given by

\[
\frac{\sigma^2}{\sigma^2 + \beta Q \beta} = \text{say } 1 - R^2
\] 
(D.2)

Let us assume we have already chosen the prior \(D(\beta \mid \sigma^2)\). We now wish to select for \(\sigma^2\) an Inverted-gamma-2 with hyperparameters \(\sigma_0^2\) and \(v_0\), where \(v_0\) is chosen, for example, on the basis of the prior squared variation coefficient for \(\sigma^2\). Our proposed procedure aims at selecting \(\sigma_0^2\) in such a way that the corresponding prior expectation of \(R^2\) is equal to a preassigned value \(R_0^2\), i.e. such that

41
\[ 1 - E_{\beta, \sigma^2} \left( \frac{\sigma^2}{\sigma^2 + \beta'Q\beta} \right) = R^2_o \]  

(D.3)

An exact procedure requires evaluating the expectation in (D.3) for different values of \( \sigma^2_o \) until one is found for which (D.3) holds at the desired level of approximation. Though Monte Carlo sampling from the tentative prior is likely to prove efficient, the exact procedure remains tedious and may not be needed when the posterior density of the coefficients of interest \( \beta \) is found to be relatively insensitive to the choice of \( \sigma^2_o \). Our procedure amounts to approximating \( \sigma^2 + \beta'Q\beta \) by the sampling variance of \( y \), say \( s^2_y \). The prior expectation of \( \sigma^2 \) is then approximated by

\[ E(\sigma^2) \approx (1 - R^2_o) \cdot s^2_y \]  

(D.4)

We intend to investigate further the properties of such approximations, considering such other candidates as

\[ E(\sigma^2) \approx \frac{1 - R^2_o}{R^2_o} \cdot \beta'_o Q \beta_o \]  

(D.5)

The case of a system of equations is naturally more complicated. We can first assign prior means to the individual error variances, following the procedure we have just described. We would then have to assign correlations on whatever grounds would be thought to be relevant. At a higher level of approximation we can proceed as described in section 5.1, at least as long as the posterior density of the parameters of interest is found to be relatively unaffected by the precise choice of \( V_o \).
References


