Econometric Theory, **21**, 2005, 326–357. Printed in the United States of America.
DOI: 10.1017/S026646660505019X

BAYESIAN REFERENCE ANALYSIS OF COINTEGRATION

MATTIAS VILLANI
Sveriges Riksbank
and
Stockholm University

A Bayesian reference analysis of the cointegrated vector autoregression is presented based on a new prior distribution. Among other properties, it is shown that this prior distribution distributes its probability mass uniformly over all cointegration spaces for a given cointegration rank and is invariant to the choice of normalizing variables for the cointegration vectors. Several methods for computing the posterior distribution of the number of cointegrating relations and distribution of the model parameters for a given number of relations are proposed, including an efficient Gibbs sampling approach where all inferences are determined from the same posterior sample. Simulated data are used to illustrate the procedures and for discussing the well-known issue of local nonidentification.

1. INTRODUCTION

Many macroeconomic time series behave in a random walk–like fashion and tend to move around wildly. Typically, such variables move around together, striving to fulfill one or several economic laws, or long-run equilibria, which tie them together. A random walk is often referred to as an *integrated* process, and integrated processes that move around together have therefore been termed *cointegrated* (Engle and Granger, 1987).

The present work is concerned with estimation of both the *number* of equilibria, the so-called cointegration rank, and the *form* of the equilibria conditional on the rank. Inferences regarding the error correcting coefficients and other short-run dynamics are also treated.

Several non-Bayesian statistical treatments of cointegration have been presented during the last two decades, most notably Ahn and Reinsel (1990), Johansen (1991), Phillips (1991), and Stock and Watson (1988).

The author thanks Luc Bauwens, Anant Kshirsagar, Peter Phillips, Herman van Dijk, four anonymous referees, and especially Daniel Thorburn for helpful comments. Financial support from the Swedish Council of Research in Humanities and Social Sciences (HSFR) grant F0582/1999 and Swedish Research Council (Vetenskapsrådet) grant 412-2002-1007 is gratefully acknowledged. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank. Address correspondence to Mattias Villani, Sveriges Riksbank, SE-103 37, Stockholm, Sweden; e-mail: mattias.villani@riksbank.se.

  0266-4666/05 $12.00

More recently, a handful of Bayesian analyses of cointegration have been developed; see Bauwens and Giot (1998), Bauwens and Lubrano (1996), Geweke (1996), Kleibergen and Paap (2002), Kleibergen and van Dijk (1994), Strachan (2003), and Villani (2000); see also Corander and Villani (2004) for a fractional Bayes approach and Chao and Phillips (1999) for an information criterion with a Bayesian flavor. Philosophical issues aside, a Bayesian approach is advantageous for many reasons: it produces whole probability distributions for each unknown parameter that are valid for any sample size, it affords straightforward handling of the inferences on the cointegration rank and tests of restrictions on the model parameters (Geweke, 1996; Kleibergen and Paap, 2002; Strachan, 2003; Villani, 2000), and it makes a satisfactory treatment of the prediction problem possible (Villani, 2001b).

The crucial step in a Bayesian analysis is the choice of prior distribution, and in each of the previously mentioned papers a new prior distribution has been introduced. The degree of motivation of the priors has varied, but the authors seem to have been more or less focused on vague priors that add only a small amount of information to the analysis, i.e., priors largely dominated by data.

This paper will be less concerned with whether or not a prior is "noninformative." The aim here is to propose a Bayesian analysis based on a sound prior that appeals to practitioners. Such a prior must consider several partially conflicting aspects of actual econometric practice. First, the number of parameters in cointegration models is usually very large, and it is not realistic to demand a detailed subjective specification of priors on such high-dimensional spaces, at least not at the current state of elicitation techniques for multivariate distributions. A prior with relatively few hyperparameters, each with a clear interpretation, is thus mandatory. Second, priors will not, or at least should not, be used by practitioners unless they are transparent in the sense that one can easily understand the kind of information they convey. Third, the prior must lead to straightforward posterior calculations that can be performed on a routine basis without the need for fine tuning in each new application. Finally, the posterior distribution of the cointegration rank can only be obtained if some parameter matrices are given proper integrable priors. A prior that fulfills these objectives will probably not coincide with the investigator's actual prior beliefs but should nevertheless be useful as point of reference, or an agreed standard, and is called a *reference* prior accordingly.

The organization of the paper is as follows. The cointegrated vector autoregressive (VAR) model is presented in Section 2. A reference prior is proposed in Section 3, and its properties are discussed in detail. Sections 4 and 5 treat the posterior distribution conditional on the cointegration rank and the posterior distribution of the rank itself, respectively. The methods are illustrated in Section 6, and the final section gives some concluding remarks. The proofs have been collected in an Appendix. Some of the more straightforward, but tedious, proofs have been omitted and may be found in Villani (2001c).

2. THE MODEL

Let $\{x_t\}_{t=1}^T$ be a p-dimensional process modeled by a cointegrated error correction (EC) model with r stationary long-run relations

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Psi_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t, \tag{2.1}$$

where $\Pi = \alpha\beta'$, β is the $p \times r$ matrix with the cointegration vectors as columns, and α is the $p \times r$ matrix of adjustment coefficients. The number of long-run relations is equal to the rank of Π, which has therefore been termed the *cointegration rank*. Both α and β are assumed to be of full rank. Here Ψ_i $(p \times p)$ $(i = 1, \ldots, k-1)$ govern the short-run dynamics of the process, d_t $(w \times 1)$ is a vector of trend, seasonal dummies, or other exogenous variables with coefficient matrix Φ $(p \times w)$, and ε_t $(p \times 1)$ contains the disturbances at time t that are assumed to follow the $N_p(0, \Sigma)$ distribution with independence across time periods.

The lag length, k, will be assumed known or determined before the analysis; see Villani (2001a) for a Bayesian approach. Alternatively, the lag length can be estimated jointly with the cointegration rank (Phillips, 1996; Chao and Phillips, 1999; Corander and Villani, 2004) or even analyzed via its posterior distribution given that all model parameters have been assigned proper prior distributions.

It is well known that only the space spanned by the cointegration vectors (sp β), the *cointegration space*, is identified, i.e., β is only determined up to arbitrary linear combinations of its columns. We will follow the traditional route in Bayesian analyses of cointegration by using a linear normalization

$$\beta = \begin{pmatrix} I_r \\ B \end{pmatrix} \tag{2.2}$$

to settle this indeterminacy, where B is a $(p - r) \times r$ matrix of fully identified parameters. When β is used as an argument in density functions it must be remembered that some of its elements are known with probability one as a result of the normalization.

The linear normalization is very convenient for computational reasons (see Sections 4 and 5), and the Bayesian analysis in this paper is shown to be invariant to the choice of normalizing variables. It should be noted, however, that the linear normalization implicitly assumes that the last $p - r$ components of x_t are not cointegrated among themselves; see Luukkonen, Ripatti, and Saikkonen (1999) for a test if this is indeed the case. Although this event is of measure zero it may have some effect on the numerical evaluation of the posterior distribution in situations where the data are located near this region.

The following compact form of the cointegrated EC model in (2.1) is useful:

$$Y = X\beta\alpha' + Z\Psi + E, \tag{2.3}$$

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

where the tth row of Y, X, Z, and E is given by $\Delta x_t'$, x_{t-1}', $(\Delta x_{t-1}', \ldots, \Delta x_{t-k+1}', d_t')$ and ε_t', respectively, and $\Psi = (\Psi_1, \ldots, \Psi_{k-1}, \Phi)'$. The expression $\mathcal{D} = \{Y, X, Z\}$ will be used as shorthand for the available data, and $d = (k-1)p + w$ denotes the number of columns in Z. We shall also use the notation

$$M_H = I_m - H(H'H)^{-1}H' \tag{2.4}$$

for any $m \times s$ matrix H of full column rank.

3. THE PRIOR DISTRIBUTION

The prior distribution is conveniently decomposed as

$$p(\alpha, \beta, \Psi, \Sigma, r) = p(\alpha, \beta, \Psi, \Sigma | r) p(r),$$

where $\Psi = (\Psi_1, \ldots, \Psi_{k-1}, \Phi)'$ and $p(r)$ is any probability distribution over the possible cointegration ranks, $r = 0, 1, \ldots, p$.

The essential conceptual difficulty in a Bayesian approach to cointegration is the prior distribution of α and β. Kleibergen and van Dijk (1994) criticized the uniform prior on α and β (see Section 6) and suggested the Jeffreys (1961) prior as a plausible alternative. The Jeffreys prior turns out to be dependent on the expected value of a data matrix, and none of the four ways of computing this expectation discussed by Kleibergen and van Dijk led to a convenient form of the posterior distribution. Bauwens and Lubrano (1996) worked with a more general class of identifying restrictions coupled with a uniform prior on α and student t priors on the free elements of the cointegration vectors. The prior was chosen out of convenience and does not consider the fact that the space of the cointegration vectors is nonstandard as a result of the identification problem discussed in Section 2. Geweke (1996) used normal shrinkage priors and obtained the posterior distribution numerically with the Gibbs sampler. The choice of prior is not motivated but seems to have been mainly chosen to assure the convergence of Gibbs sampling algorithm. Recently, Kleibergen and Paap (2002) proposed a reference prior on α and β that is essentially a prior on Π in the full rank EC model projected down to the subspace where $\text{rank}(\Pi) = r$; Strachan (2003) extended this idea to more general identifying restrictions. It is an approach that is rather common and well understood in linear models, but its implications in nonlinear models, such as the EC model with reduced rank in (2.1), are not as transparent; see also Section 6.

The approach taken here differs from the previously mentioned works by focusing directly on the actual structure of the parameter space of β. We introduce the proposed reference prior now and spend the rest of this section motivating its particular form. Let $\text{etr}(H) = \exp(-\frac{1}{2}\,\text{tr}\,H)$ for any square matrix H. The prior can then be written

$$p(\alpha, \beta, \Psi, \Sigma | r) = c_r |\Sigma|^{-(p+r+q+1)/2} \,\text{etr}[\Sigma^{-1}(A + v\alpha\beta'\beta\alpha')], \tag{3.1}$$

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

where $v > 0$, $q \geq p$, and A, a $p \times p$ positive definite matrix, are the three hyperparameters to be specified by the investigator. The normalizing constant is

$$c_r = |A|^{q/2} \frac{\Gamma_r(p)}{\Gamma_p(q)\Gamma_r(r)} \frac{2^{-qp/2}\pi^{-p(p-1)/4}}{(2\pi/v)^{pr/2}\pi^{(p-r)r/2}},$$

where $\Gamma_b(a) = \prod_{i=0}^{b-1} \Gamma[(a-i)/2]$, for positive integers a and b satisfying $a \geq b-1$.

Note that Ψ is uniformly distributed over $\mathbb{R}^{(p-1)k+w}$, which makes the overall prior $p(\alpha, \beta, \Psi, \Sigma | r)$ improper. The prior on α, β, and Σ conditional on Ψ is proper, however. The uniform prior for Ψ is used here for simplicity, but a general multivariate normal prior on vec Ψ (e.g., a structured shrinkage prior as in Litterman, 1986) leads to essentially the same posterior computations.

Implicit in (3.1) is the assumption of common A, q, and v for all r; the ensuing analysis proceeds in the same manner in the general case with varying A, q, and v.

3.1. Marginal and Conditional Prior Distributions

Throughout this section, we will assume that $k = 1$ and $w = 0$ for notational convenience. The results will still be valid for $k > 1$ and $w > 0$ as long as prior independence between Ψ and the other parameter matrices is assumed. All probability distributions in this section will be conditional on a given cointegration rank, though this will not be written out explicitly.

The space of β is not euclidean because of the nonidentification of the cointegration vectors. It is deceptive to think in terms of the free parameter space of β under some arbitrarily chosen normalization, e.g., the linear normalization in Section 2, without regard to the fact that actual parameter space is non-euclidean. In the following paragraphs we shall describe the true parameter space of β and show that the prior in (3.1) implies a uniform distribution over this abstract space.

Let $\mathcal{X}$ denote the set of $p \times r$ real matrices of rank $r (\leq p)$ and define the group of transformations $X \to XL$, where $X \in \mathcal{X}$ and L is any nonsingular $r \times r$ matrix. This group defines an equivalence relation $\overset{\text{sp}}{=}$ in $\mathcal{X}$ such that for any $X, Y \in \mathcal{X}$, $X \overset{\text{sp}}{=} Y$ if and only if $\text{sp}(X) = \text{sp}(Y)$. Thus, the points of the resulting coset space of equivalence classes, usually denoted by $\mathbb{R}^{p\times r}/\overset{\text{sp}}{=}$, stand in a 1-1 correspondence with the r-dimensional subspaces of $\mathbb{R}^p$. The set of r-dimensional subspaces of $\mathbb{R}^p$ is an analytic manifold of dimension $(p-r)r$ (James, 1954), which has been termed the *Grassman manifold* and is denoted by $\mathcal{G}_{r,p-r}$.

The uniform distribution on $\mathcal{G}_{r,p-r}$ is naturally defined as the (unique) invariant distribution under the group of transformations of $\mathcal{G}_{r,p-r}$ induced by the group of orthonormal transformations of $\mathbb{R}^p$ (James, 1954).

As a result of the nonidentification of the cointegration vectors explained in Section 2, the actual parameter space of β is the Grassman manifold. We shall

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

now prove that the distribution in (3.1) implies that β is marginally uniformly distributed over $\mathcal{G}_{r,p-r}$. First we need a definition and a few lemmas.

DEFINITION 3.1. *An $m \times s$ matrix D follows the matrix t distribution, $D \sim t_{m\times s}(\mu, \Upsilon, \Theta, g)$, if its density is given by*

$$\frac{\Gamma_s(g+m+s-1)}{\Gamma_s(g+s-1)\pi^{ms/2}|\Upsilon|^{s/2}|\Theta|^{m/2}}\,|I_s+\Theta^{-1}(D-\mu)'\Upsilon^{-1}(D-\mu)|^{-(g+m+s-1)/2}.$$

See Box and Tiao (1973) and Bauwens, Lubrano, and Richard (1999) for properties of the matrix t distribution.

LEMMA 3.2. *Let R be a $p \times r$ matrix of independent N(0,1) variables. Then* $\mathrm{sp}(R)$ *is uniformly distributed over $\mathcal{G}_{r,p-r}$.*

Proof. See James (1954).

LEMMA 3.3. *If N_1 and N_2 are independent $s \times s$ and $m \times s$ matrices of independent $N(0,1)$ variables, then*

$N_2 N_1^{-1} \sim t_{m\times s}(0, I_m, I_s, 1)$.

Proof. See Phillips (1989) and Dickey (1967).

LEMMA 3.4. *If $\beta = (I_r\ B')'$ and $B \sim t_{(p-r)\times r}(0, I_{p-r}, I_r, 1)$, then* $\mathrm{sp}(\beta)$ *is uniformly distributed over $\mathcal{G}_{r,p-r}$.*

With the preceding definitions and lemmas out of the way, we are now prepared to state an important property of the distribution in (3.1).

THEOREM 3.5. *β is marginally uniformly distributed over $\mathcal{G}_{r,p-r}$.*

To illustrate this rather abstract uniform distribution, let us consider the bivariate case with a single cointegration vector $\beta = (1, B)'$. According to the proof of Theorem 3.5 in the Appendix, the distribution in (3.1) implies a Cauchy(0,1) distribution on B. This is not surprising given that B is a ratio of two independent $N(0,1)$ variates under the uniform distribution over $\mathcal{G}_{r,p-r}$ (see Lemmas 3.2–3.4). A more natural, but computationally inconvenient, parametrization of β is the polar parametrization

$$\tilde{\beta} = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}, \qquad -\frac{\pi}{2} \le \theta < \frac{\pi}{2}, \tag{3.2}$$

where θ is the angle of the cointegration vector. In this parametrization the distribution in Theorem 3.5 reduces to a constant density for θ (James, 1954). Slightly more generally, in the p-dimensional case with a single cointegration

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

vector, the distribution in Theorem 3.5 reduces to the conventional uniform distribution over the p-dimensional hemisphere with unit radius (Mardia and Jupp, 2000). In the general case, we may say that the prior in (3.1) assigns equal probability to every possible cointegration space of dimension r. Although more informative prior information on the cointegration vectors may be available, the marginal prior on β implied by the prior in (3.1) satisfies all four of the desiderata stated in the Introduction and should therefore be a suitable reference prior.

It should be noted that the prior in (3.1) is by no means the only distribution on α and β that implies a uniform distribution on the Grassman manifold. The prior in (3.1) is especially interesting, however, in that it is both conceptually relevant and, as will be shown later, very convenient from a computational viewpoint.

THEOREM 3.6. *The marginal prior of* Σ *is*

$$\Sigma \sim IW(A, q),$$

where IW denotes the inverted Wishart distribution (Zellner, 1971).

Proof. This follows directly from the proof of Theorem 3.5. ■

From (3.1) we immediately obtain

$$\alpha|\beta,\Sigma \sim N_{p\times r}[0,(\beta'\beta)^{-1},v^{-1}\Sigma], \tag{3.3}$$

where $A \sim N_{m\times s}(\mu,\Omega_1,\Omega_2)$ means that $\text{vec}\, A \sim N_{ms}(\text{vec}\,\mu,\Omega_1 \otimes \Omega_2)$. The linear normalization of β makes α difficult to interpret, however, and the conditional prior in (3.3) may not shed much light on the prior in (3.1).

Consider instead the prior of α conditional on β and Σ when β is orthonormal. Restricting β to be orthonormal is not sufficient to identify the model, however, as any orthonormal version of β can be rotated to a new one by postmultiplying it with an $r \times r$ orthonormal matrix. This need not concern us here as β only enters $p(\alpha|\beta,\Sigma)$ in the form $\beta'\beta$ and $p(\alpha|\beta,\Sigma)$ is therefore invariant under these rotations. Define $\tilde{\beta} = \beta(\beta'\beta)^{-1/2}$ and note that $\tilde{\beta}$ is orthonormal. For $\Pi = \alpha\beta'$ to remain unchanged by the transformation $\beta \to \tilde{\beta}$, we must make the corresponding transformation of the adjustment matrix from α to $\tilde{\alpha} = \alpha(\beta'\beta)^{1/2}$. In the following theorem, let $\tilde{\alpha}_i$ denote the ith column of $\tilde{\alpha}$ and note that $\tilde{\alpha}_i$ describes how the p response variables are affected by the ith cointegrating relation under the orthonormal normalization.

THEOREM 3.7. $\tilde{\alpha}_i|\Sigma \overset{iid}{\sim} N_p(0,v^{-1}\Sigma), \qquad i = 1,2,\ldots,r.$

The rather restrictive form of the prior in Theorem 3.7 must be motivated. First, the restriction to conditional normal priors on α (and thereby also on $\tilde{\alpha}$) is necessary for an efficient numerical evaluation of the posterior; see

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

Sections 4 and 5. Second, nonidentical priors on the columns of $\tilde{\alpha}$ do not make sense unless overidentifying restrictions on the columns of β are used to give a unique meaning to each cointegration vector. Another way to see this is that within the class of matrix normal priors $\tilde{\alpha}|\tilde{\beta},\Sigma \sim N_{p\times r}(\mu,\Omega_1,\Omega_2)$, only the priors with $\mu = 0$, $\Omega_1 = I_r$ are invariant to rotations of $\tilde{\beta}$. Third, the scale matrix in the conditional prior may be any positive definite matrix; the posterior computations remain nearly the same. By making the conditional covariance matrix proportional to Σ we are taking the possibly differing scales of the time series into account. Finally, the reason for centering the conditional prior over zero is motivated by the invariance requirement just stated. It has the effect of centering the prior over $\Pi = 0$, which is often a good starting point in an analysis; see the discussion of the "sum of coefficients" prior in Doan, Litterman, and Sims (1984) and Section 3.2.

In his influential development of Bayesian reference tests of sharp null hypotheses Jeffreys (1961, Ch. 5) argued that the prior on the parameters under the alternative hypothesis should be centered over the point in the null and that the prior spread around this point should be an increasing function of the model's scale parameter; see also Berger (1985, Sect. 4.3.3). Although the situation is quite a bit more complex here, the prior in Theorem 3.7, which is centered over the hypothesis $\Pi = 0$, or $r = 0$, with a prior scale depending on Σ, has the same flavor and should therefore be appropriate for inference on the cointegration rank; see Section 5.

Further clarification of the hyperparameters A, q, and v is obtained from the marginal prior of $\tilde{\alpha}$. By multiplying $p(\tilde{\alpha}|\Sigma)$ with the marginal inverted Wishart prior of Σ and integrating with respect to Σ, we obtain

$$\tilde{\alpha} \sim t_{p\times r}(0, v^{-1}A, I_r, q-p+1). \tag{3.4}$$

Results in Box and Tiao (1973, pp. 446–447) then give

$$E(\tilde{\alpha}) = 0 \quad \text{and} \quad \operatorname{Cov}(\operatorname{vec}\tilde{\alpha}) = I_r \otimes v^{-1}E(\Sigma),$$

where $E(\Sigma) = A/(q-p-1)$ is the expected value of Σ a priori; see, e.g., Bauwens et al. (1999, p. 306).

The hyperparameter A is determined from $E(\Sigma)$ and q, and the investigator thus faces subjective specification of (i) the expected value of Σ, (ii) the degree of certainty regarding Σ (large values of q imply large certainty), and (iii) the tightness around the point zero for $\tilde{\alpha}$ (large values of v give high concentration of probability mass around zero). Note that whether a value for v is large or not depends on $E(\Sigma)$, which should therefore be specified before v.

The main difficulty for the investigator is likely to be the specification of $E(\Sigma)$. If interest only centers on the posterior of α,β,Ψ,Σ conditional on a given cointegration rank, then A may be set equal to the zero matrix and $q = 0$. This corresponds to using the usual improper prior $p(\Sigma) \propto |\Sigma|^{-(p+1)/2}$. If we also aim at analyzing the cointegration rank, but are either unable or unwilling

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

to state our beliefs about Σ, then $A = \hat{\Sigma}$ and $q = p + 2$ may be used, where $\hat{\Sigma}$ is the maximum likelihood estimate of Σ in the full rank model; note that this implies that $E(\Sigma) = \hat{\Sigma}$. This suggestion is of course not a proper Bayesian solution as the prior then becomes dependent on the observed data. The consequences of this side step are minimized by the choice of the smallest possible q (maximum uncertainty) subject to a finite expected value of Σ.

3.2. Prior Stability

Define

$$\Pi_C = \begin{pmatrix} I_p + \alpha\beta' + \Psi_1 & \Psi_2 - \Psi_1 & \dots & \Psi_{k-1} - \Psi_{k-2} & -\Psi_{k-1} \\ I_p & 0 & \dots & 0 & 0 \\ 0 & I_p & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_p & 0 \end{pmatrix}.$$

The assumption of $\text{rank}(\Pi) = r$ implies that r of the eigenvalues of Π_C are equal to one. A cointegrated process is stable if all the remaining eigenvalues of Π_C are smaller than one in modulus. It is clearly of interest to know what prior probability is implicitly being placed on the set of stable processes if the prior in (3.1) is used. This could be investigated either by analytical approximation or by simulation methods for different models, i.e., by varying p, r, and k. We shall here be content with simulating the special case $p = 2$ and $r = k = 1$. Table 1 displays the prior probability that the process is stable for $A = I_2$ as a function of q and $\sigma = v^{-1/2}$ (note that σ is on a standard deviation scale). Experiments with other choices of A with strong positive and negative correlation structure did not have a large impact on the probability. Note also that it is unnecessary to increase the magnitude of the diagonal elements in A as this has the same effect as increasing σ.

TABLE 1. Implied prior probability that the process is stable

	σ									
	0.01	0.1	0.25	0.5	0.75	1	5	10	50	100
$q = 2$	0.48	0.46	0.40	0.35	0.30	0.26	0.08	0.04	0.01	0.00
$q = 4$	0.49	0.49	0.47	0.46	0.42	0.40	0.15	0.08	0.02	0.01
$q = 10$	0.50	0.50	0.49	0.48	0.47	0.45	0.26	0.14	0.03	0.02
$q = 20$	0.50	0.49	0.49	0.48	0.47	0.47	0.33	0.19	0.05	0.02

Note: $r = k = 1$ and $A = I_2$.

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

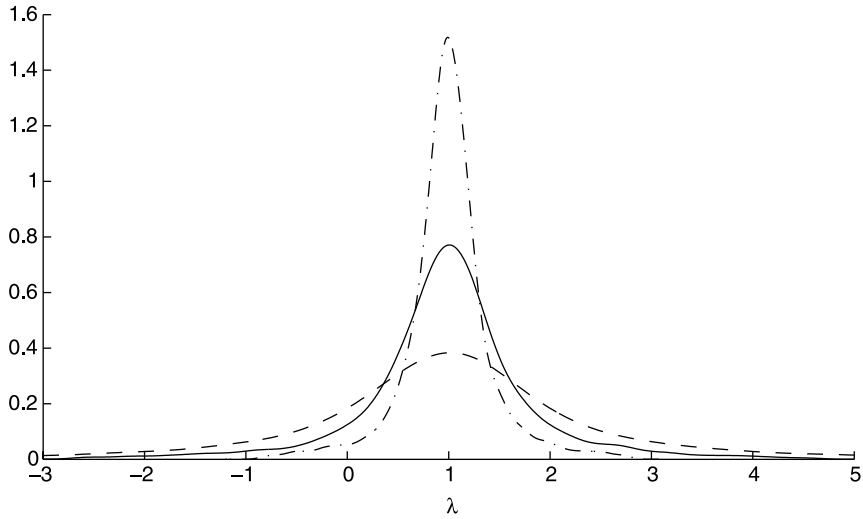

FIGURE 1. Implied prior distribution on the unrestricted eigenvalue for $A = I_2$ and $q = 4$. Here $\sigma = 0.25$ (– · –), $\sigma = 0.5$ (——), and $\sigma = 1$ (– – –).

Densities of the unrestricted eigenvalue (λ) are displayed in Figure 1 for different values of σ. The densities are symmetric around the modal value $\lambda = 1$. A nonsymmetric density for λ that places more mass to the left of $\lambda = 1$ than to the right of this point would perhaps better represent actual beliefs. The gain from a nonsymmetric prior is probably less than the loss in computational efficiency in the posterior calculations, however.

A crude way to obtain a nonsymmetric prior is to simply exclude explosive processes a priori (or "too explosive" processes, e.g., with eigenvalues larger than 1.1 in modulus) by restricting the domain of the prior in (3.1) to the space of α, β, and Ψ where the process is stable. This is neatly handled in the posterior calculations for a given cointegration rank by simply rejecting the draws from the posterior corresponding to nonstable processes; see Section 4. Note that the latter region will be small if the process actually is stable and data informative and most draws will then be accepted. The posterior distribution of the rank will require heavier numerical computations, however.

4. THE POSTERIOR DISTRIBUTION CONDITIONAL ON THE RANK

4.1. Normalization Issues

The choice of variables used for normalizing β may be somewhat arbitrary, and it is important to show that the posterior distribution corresponding to the prior in (3.1) is invariant to this choice. Let $\mathcal{N}_1 = \{i_1, \ldots, i_r\}$ denote the set of

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

indices for the r variables used to normalize β. Consider the change in normalization $\mathcal{N}_1 \to \mathcal{N}_2$, where $\mathcal{N}_2$ equals $\mathcal{N}_1$ with jth variable in the normalized set replaced by the kth variable in the nonnormalized set. This change in normalizing variables is accomplished by the transformation $T_U:(\alpha,\beta,\Sigma) \to (\bar{\alpha},\bar{\beta},\Sigma)$, where $\bar{\alpha} = \alpha U'$, $\bar{\beta} = \beta U^{-1}$, and U is an $r \times r$ invertible transformation matrix whose elements are functions of the kth row of B. The exact form of U need not concern us for the moment; it is sufficient to note that such a matrix always exists, and is unique, if the kth variable in the nonnormalized set has a nonzero coefficient in the jth cointegrating vector; see the proof of Theorem 4.2 in the Appendix. Such a change of normalizing variables will be termed *valid*. It is important to note that $\Pi = \alpha\beta'$ is unchanged by the transformation.

The next definition, adapted from Drèze and Richard (1983), formalizes the idea that the inference should not depend on whether we (i) work directly with $\mathcal{N}_1$ or (ii) start with $\mathcal{N}_2$ and then transform to $\mathcal{N}_1$.

DEFINITION 4.1. *A density $p(\alpha,\beta,\Sigma)$ is said to be* invariant with respect to normalization *if and only if its functional form is invariant with respect to the valid parameter transformation $T_U:(\alpha,\beta,\Sigma) \to (\bar{\alpha},\bar{\beta},\Sigma)$.*

THEOREM 4.2. *The posterior distribution corresponding to the prior (3.1) is invariant with respect to normalization.*

The main advantages of the linear normalization are that the prior that assigns the same probability to every cointegration space is of rather simple form and that easily implemented numerical methods (see Sections 4.3 and 5) can be used to compute the posterior results. Note also that we are free to transform the posterior distribution of α and β as long as the space spanned by the columns of β and the matrix of long-run multipliers $\Pi = \alpha\beta'$ remain unchanged, i.e., the class of allowable transformations is $(\alpha,\beta) \to (\alpha V', \beta V^{-1})$, for any invertible $r \times r$ matrix V. For example, an orthonormal β is obtained with $V = (\beta'\beta)^{1/2}$. The transformation is conveniently performed directly on the posterior draws of α and β. Thus, as long as the initial linear normalization is valid (dubious normalizations may be excluded with the test of Luukkonen et al., 1999), the restriction to the linear normalization is no restriction at all as the final results may be transformed to any desired normalization.

4.2. Marginal Posterior Distribution of β

The next result gives the marginal posterior of the cointegration vectors.

THEOREM 4.3. *The marginal posterior distribution of β is*

$$p(\beta|\mathcal{D},r) \propto \frac{|\beta' C_1 \beta|^{(T+q-d-p)/2}}{|\beta' C_2 \beta|^{(T+q-d)/2}}, \tag{4.1}$$

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

where $C_1 = X'M_Z X + vI_p$, $C_2 = vI_p + X'Q[I_T - Z(Z'QZ)^{-1}Z'Q]X$, *and* $Q = I_T - Y(A + Y'Y)^{-1}Y'$.

The expression $p(\beta|\mathcal{D}, r)$ in Theorem 4.3 is a 1-1 poly-matrix-*t* density (Bauwens and van Dijk, 1990). Theorem 3.1 in Bauwens and Lubrano (1996) is the limiting special case of Theorem 4.3 with $A = 0$ and $q = v = 0$ (which corresponds to a constant prior on α and β). Contrary to the family of multivariate poly-*t* densities (see, e.g., Dickey, 1968; Drèze, 1977; Bauwens et al., 1999), poly-matrix-*t* densities have remained largely unexplored. The following result can be shown, however.

THEOREM 4.4. *The marginal posterior of B is integrable but possesses no finite integer moments.*

Proof. The result follows from a trivial modification of the proof of Corollary 3.2 in Bauwens and Lubrano (1996).

The nonexistence of integer moments is not a consequence of the prior distribution in (3.1) but rather of the linear normalization of β, where each element of B is a matrix quotient with the upper $r \times r$ submatrix of β in the denominator. Phillips (1994) makes the same point about the distribution of the maximum likelihood estimator in the linear normalization, which he shows has Cauchy-like tails.

It is also possible to derive the marginal posterior distribution of α as in Kleibergen and van Dijk (1994, eq. (29)) in closed form. It is a complicated nonstandard distribution (see Section 6 for further discussion) and is not conveniently used in the numerical posterior evaluations discussed in the next section.

4.3. Numerical Posterior Evaluation

The marginal posterior distribution of the cointegration vectors in Theorem 4.3 is of the same 1-1 poly-matrix-*t* form as the distribution in Theorem 3.1 in Bauwens and Lubrano (1996). Bauwens and Lubrano discuss both importance sampling (Kloek and van Dijk, 1978) and Gibbs sampling (Smith and Roberts, 1993) approaches to evaluating such a density; Bauwens and Giot (1998) implement the Gibbs sampling approach and give details on convergence issues. The key properties used in those exercises are (i) the conditional distribution of one of the cointegration vectors conditional on all other cointegration vectors is a vector 1-1 poly-*t*, (ii) the 1-1 poly-*t* is amenable to direct simulation using the algorithm of Bauwens and Richard (1985), and (iii) the posteriors of α, Ψ, and Σ conditional on β are all standard. Once the marginal posterior of β has been evaluated by sampling methods the marginal posteriors of α, Ψ, and Σ may therefore be computed by averaging their posteriors conditional on β over the

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

posterior sample of β. We refer the reader to Bauwens and Lubrano (1996) and Bauwens and Giot (1998) for details.

A major disadvantage of building the numerical posterior evaluations on the analytical form of $p(\beta|\mathcal{D}, r)$ is the inability to handle posterior distributions of quantities with intractable posterior distribution conditional on β, such as impulse response functions or forecasts. The Gibbs sampler is a convenient algorithm for sampling from the joint posterior distribution of α, β, Ψ, and Σ and may thus be used in such situations; Geweke (1996) seems to have been the first to use Gibbs sampling in cointegration models. It turns out that the posterior distribution for the prior in (3.1) is amenable to an algorithm similar to the one in Geweke (1996). The Gibbs sample may also be used to efficiently compute the posterior distribution of the cointegration rank (Section 5 and Theorem 4.6, which follows).

The Gibbs sampler is an easily implemented method for generating observations from complex multidimensional distributions by sampling iteratively from the so-called full conditional posterior distributions. The full conditional posterior distribution of a subset of parameters in a model is the posterior distribution of the subset conditional on all other parameters. Initial values for all parameters are needed to start up the Gibbs sampler. The maximum likelihood estimates in Johansen (1995) are natural candidates. The sampled parameter values are not independent but can be shown to converge in distribution to the target posterior distribution independently of the choice of initial values (Tierney, 1994). Furthermore, the expected value of any well-behaved transformation of the parameters may be consistently estimated by sampling averages.

The full conditional posteriors of α, β, Ψ, and Σ are given in the next theorem.

THEOREM 4.5.

- *The full conditional posterior of* Σ

  $$\Sigma|\alpha, \beta, \Psi, \mathcal{D}, r \sim IW_p(E'E + A + v\alpha\beta'\beta\alpha', T + q + r),$$

  where $E = Y - X\beta\alpha' - Z\Psi$.

- *The full conditional posterior of* Ψ

  $$\Psi|\alpha, \beta, \Sigma, \mathcal{D}, r \sim N_{d\times p}[\mu_\Psi, \Sigma, (Z'Z)^{-1}],$$

  where $\mu_\Psi = (Z'Z)^{-1}Z'(Y - X\beta\alpha')$.

- *The full conditional posterior of* α

  $$\alpha|\beta, \Psi, \Sigma, \mathcal{D}, r \sim N_{p\times r}\{\mu_\alpha, [\beta'(X'X + vI_p)\beta]^{-1}, \Sigma\},$$

  where $\mu_\alpha = (Y - Z\Psi)'X\beta[\beta'(vI_p + X'X)\beta]^{-1}$.

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

- *The full conditional posterior of B*

$$B|\alpha,\Psi,\Sigma,\mathcal{D},r \sim N_{(p-r)\times r}[\mu_B,(\alpha'\Sigma^{-1}\alpha)^{-1},(X_2'X_2+vI_{p-r})^{-1}],$$

where $\mu_B = (X_2'X_2 + vI_{p-r})^{-1}X_2'(Y - X_1\alpha' - Z\Psi)\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}$ *and* X_1 *denotes the r first columns of X and* X_2 *the* $p - r$ *last ones.*

Most of the model parameters are located in Ψ and Σ, and the Gibbs updating steps for these two matrices usually dominate the total computing time. The time to convergence of the Gibbs sampler also increases as the dimensions of Ψ and Σ grow. The next theorem gives the conditional posteriors necessary to perform a (marginal) Gibbs sampler to generate samples directly from $p(\alpha, B|\mathcal{D}, r)$. This Gibbs sampler is also used in Section 5 to calculate the posterior distribution of the rank.

THEOREM 4.6.

- *The posterior of* α *conditional on* β *and r*

$$\alpha|\beta,\mathcal{D},r \sim t_{p\times r}[\hat{\alpha},A + Y'M_Z(Y - X\beta\hat{\alpha}'),(\beta'C_1\beta)^{-1},$$
$$T + q - (d + p) + 1],$$

where $\hat{\alpha} = Y'M_ZX\beta(\beta'C_1\beta)^{-1}$.

- *The posterior of B conditional on* α *and r*

$$B|\alpha,\mathcal{D},r \sim t_{(p-r)\times r}[\hat{B},G_3 - G_2'G_1^{-1}G_2,C_3,T + q + r - (d + p) + 1],$$

where $\hat{\beta} = \hat{\Pi}'S^{-1}\alpha(\alpha'S^{-1}\alpha)^{-1}$, $\hat{\Pi} = Y'M_ZXC_1^{-1}$, $S = A + Y'M_ZY - \hat{\Pi}C_1\hat{\Pi}'$, $\hat{\beta}_1$ *contains the r first rows of* $\hat{\beta}$ *and* $\hat{\beta}_2$ *the* $p - r$ *remaining ones, and*

$$C_1^{-1} + \hat{\Pi}'S^{-1}\hat{\Pi} - \hat{\beta}\alpha'S^{-1}\alpha\hat{\beta}' = \begin{pmatrix} \underset{r\times r}{G_1} & \underset{r\times(p-r)}{G_2} \\ \underset{(p-r)\times r}{G_2'} & \underset{(p-r)\times(p-r)}{G_3} \end{pmatrix}$$

is decomposed conformably, $C_3 = (I_r - \hat{\beta}_1)'G_1^{-1}(I_r - \hat{\beta}_1) + (\alpha'S^{-1}\alpha)^{-1}$ *and* $\hat{B} = \hat{\beta}_2 + G_2'G_1^{-1}(I_r - \hat{\beta}_1)$.

The posterior densities of Ψ and Σ are obtainable by marginalizing their densities conditional on α and β, which belong to the matrix t and inverted Wishart family, respectively, using draws from the marginal Gibbs sampler in Theorem 4.6; Bauwens and Lubrano (1996) and Bauwens and Giot (1998) provide the details.

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

5. THE POSTERIOR DISTRIBUTION OF THE COINTEGRATION RANK

The posterior distribution of the cointegration rank is

$$p(r|\mathcal{D}) = \frac{p(\mathcal{D}|r)p(r)}{\sum_{r=0}^{p} p(\mathcal{D}|r)p(r)}, \tag{5.1}$$

where $p(r)$ is the prior probability of r cointegrating relations and

$$p(\mathcal{D}|r) = \iiiint p(\mathcal{D}|\alpha,\beta,\Psi,\Sigma,r)p(\alpha,\beta,\Psi,\Sigma|r)\, d\Sigma d\Psi d\alpha d\beta \tag{5.2}$$

is the marginal likelihood of the data given $\text{rank}(\Pi) = r$.

The marginal likelihoods for $r = 0$ and $r = p$ are analytically tractable if the prior in (3.1) is used also for the zero and full rank models. These priors agree with our earlier prior in the reduced rank case and do not introduce any new prior hyperparameters. If $r = 0$, then $\alpha = \beta = 0$ and the prior in (3.1) becomes

$$p(\Psi,\Sigma|r) = c_0|\Sigma|^{-(p+q+1)/2}\,\text{etr}(\Sigma^{-1}A), \tag{5.3}$$

which is an $IW(A,q)$ prior on Σ, and $p(\Psi)$ is a constant density. For $r = p$, $\Pi = \alpha\beta'$ is of full rank and

$$p(\Pi,\Psi,\Sigma|r) = c_p|\Sigma|^{-(2p+q+1)/2}\,\text{etr}[\Sigma^{-1}(A + v\Pi\Pi')], \tag{5.4}$$

which implies $\Sigma \sim IW(A,q)$, $\text{vec}\,\Pi|\Sigma \sim \text{N}_{p^2}(0, I_p \otimes v^{-1}\Sigma)$ and a constant prior on Ψ. If the Kronecker structure on the prior covariance matrix of Π is too restrictive, a general normal-Wishart distribution may be used as a prior for Π and Σ.

The marginal likelihoods for $r = 0$ and $r = p$ are given in the next theorem.

THEOREM 5.1. *For the priors in (5.3) and (5.4)*

$$p(\mathcal{D}|r = 0) \propto \Gamma_p(T + q - d)|A + Y'M_ZY|^{-(T+q-d)/2},$$

$$p(\mathcal{D}|r = p) \propto \Gamma_p(T + q - d)v^{p^2/2}|S|^{-(T+q-d)/2}|C_1|^{-p/2},$$

where S is defined in Theorem 4.6 and C_1 is given in Theorem 4.3.

The proportionality signs in Theorem 5.1 are used to denote that the multiplicative constant $|A|^{q/2}|Z'Z|^{-p/2}\pi^{-(T-d)p/2}\Gamma_p^{-1}(q)$ has been discarded as it enters all marginal likelihoods of r. This practice is followed throughout this section.

For $1 \le r \le p - 1$ at least one of the integrals in (5.2) must be handled by numerical methods. We shall here discuss three possible simulation-based approaches: Monte Carlo integration, importance sampling, and the marginal likelihood identity approach of Chib (1995).

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

5.1. Monte Carlo Integration

The integrals in (5.2) with respect to α, Ψ, and Σ may be computed analytically, leading to the 1-1 poly-matrix-t density in Theorem 4.3, which is repeated here (along with its proportionality constant)

$$p(\mathcal{D}|r) = \frac{\Gamma_r(p)\Gamma_p(T+q-d)v^{pr/2}\pi^{-(p-r)r/2}}{\Gamma_r(r)|A+Y'M_ZY|^{(T+q-d)/2}} \times \int |\beta'C_1\beta|^{(T+q-d-p)/2}|\beta'C_2\beta|^{-(T+q-d)/2}\,dB.$$

The final integral with respect to B must be computed numerically. A Monte Carlo integration approach is suggested by the following lemma, which is proved by expanding $\beta'C_2\beta$ in B and completing the square (see the proof of Corollary 3.2 in Bauwens and Lubrano, 1996).

LEMMA 5.2. *For* $1 \le r \le p-1$

$$p(\mathcal{D}|r) \propto \frac{\Gamma_r(p)\Gamma_p(T+q-d)\Gamma_r(T+q+r-p-d)v^{pr/2}E(|\beta'C_1\beta|^{(T+q-d-p)/2})}{\Gamma_r(r)\Gamma_r(T+q-d)|A+Y'M_ZY|^{(T+q-d)/2}|K_3|^{r/2}|U|^{(T+q+r-d-p)/2}},$$

where the expectation is taken with respect to the $t_{(p-r)\times r}(\tilde{B},K_3^{-1},U,T+q-p-d+1)$ *distribution,* C_2 *(see Theorem 4.3) is partitioned as*

$$C_2 = \begin{pmatrix} \underset{r\times r}{K_1} & \underset{r\times(p-r)}{K_2} \\ \underset{(p-r)\times r}{K_2'} & \underset{(p-r)\times(p-r)}{K_3} \end{pmatrix},$$

$\tilde{B} = -K_3^{-1}K_2'$, *and* $U = K_1 - K_2K_3^{-1}K_2'$.

The expected value in Lemma 5.2 may be computed by generating variates from the $t_{(p-r)\times r}(\tilde{B},K_3^{-1},U,T+q-p-d+1)$ distribution, computing $|\beta'C_1\beta|^{(T+q-d-p)/2}$ for each draw, and averaging over all draws.

5.2. Importance Sampling

Another method that may be used to approximate the integral with respect to β in (4.1) is importance sampling (Kloek and van Dijk, 1978; Geweke, 1989). In cases where the importance function well approximates the target integrand, importance sampling can be quite efficient as it produces independent draws without wasting an initial burn-in sample. The fact that the draws are independent makes a central limit theorem directly applicable, and the precision of the estimates is easily assessed (Geweke, 1989).

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

Given the heavy tails of the marginal posterior of β (Theorem 4.4), a natural suggestion for an importance function is the matrix Cauchy density. The maximum likelihood estimate of B and an estimate of its asymptotic covariance matrix (Johansen, 1995, Theorem 13.4) may be used as location and scale matrix, respectively. That is, we suggest the density $t_{(p-r)\times r}[\hat{B}, (X_2'X_2)^{-1}, (T\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha})^{-1}, 1]$ as an importance function. Further fine tuning may be introduced by multiplying $(X_2'X_2)^{-1}$ by a scale factor.

Poly-t densities may be substantially skew and even bimodal. In such cases the matrix Cauchy may not perform well as an importance function. An alternative may be to generate each of the r cointegration vectors conditional on the maximum likelihood estimates of the remaining $r-1$ vectors. These conditional posteriors are 1-1 poly-t (Bauwens and Lubrano, 1996) and may be generated by one of the algorithms in Bauwens and Richard (1985).

5.3. Marginal Likelihood Identity Approach

By a slight rearrangement of Bayes' theorem we obtain what Chib (1995) has termed the *basic marginal likelihood identity*:

$$p(\mathcal{D}|r) = \frac{p(\mathcal{D}|\alpha, B, r)p(\alpha, B|r)}{p(\alpha, B|\mathcal{D}, r)} = \frac{p(\mathcal{D}|\alpha, B, r)p(\alpha, B|r)}{p(B|\alpha, \mathcal{D}, r)p(\alpha|\mathcal{D}, r)}. \tag{5.5}$$

Chib (1995) suggested using this identity in combination with a Gibbs sampler to estimate the marginal likelihood. The expression for $p(\mathcal{D}|r)$ in (5.5) clearly holds for any α and B. Let $(\tilde{\alpha}, \tilde{B})$ be the point where $p(\mathcal{D}|r)$ is evaluated. As explained in Chib (1995), this point should preferably be of high posterior density; the posterior mode and median are good candidates (the posterior mean does not exist; see Theorem 4.4). The term $p(B|\alpha, \mathcal{D}, r)$ in (5.5) is given in the second part of Theorem 4.6, and the next result gives the expression for the numerator of (5.5).

LEMMA 5.3.

$$p(\mathcal{D}|\alpha, B, r)p(\alpha, B|r) \propto \frac{\Gamma_p(T+q+r-d)\Gamma_r(p)}{\Gamma_r(r)\pi^{(2pr-r^2)/2}v^{-pr/2}} \times |A + v\alpha\beta'\beta\alpha' + W'M_Z W|^{-(T+q+r-d)/2},$$

where $W = Y - X\beta\alpha'$.

The final term of the marginal likelihood identity $p(\alpha|\mathcal{D}, r)$ is not available in closed form, but its value in a point $\alpha = \tilde{\alpha}$, which is all we need, can be computed from a posterior sample $B^{(1)}, \ldots, B^{(n)}$ of B by

$$\hat{p}(\tilde{\alpha}|\mathcal{D}, r) = \frac{1}{n}\sum_{i=1}^{n} p(\tilde{\alpha}|B^{(i)}, \mathcal{D}, r),$$

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

where $p(\alpha|B,\mathcal{D},r)$ is given in the first part of Theorem 4.6. From the ergodic theorem (Tierney, 1994), $\hat{p}(\tilde{\alpha}|\mathcal{D},r) \to p(\tilde{\alpha}|\mathcal{D},r)$ almost surely. This procedure for computing $p(\mathcal{D}|r)$ will be named the marginal likelihood identity (MLI) algorithm.

The posterior sample from $p(B|\mathcal{D},r)$ needed in the MLI approach can be obtained from (i) a Gibbs sampler for the 1-1 poly-matrix-t density in (4.1) as described in Bauwens and Lubrano (1996) and Bauwens and Giot (1998), (ii) the marginal Gibbs sampler in Theorem 4.6, which samples from $p(\alpha,B|\mathcal{D},r)$, and (iii) the full Gibbs sampler in Theorem 4.5, which samples from $p(\alpha,B,\Psi,\Sigma|\mathcal{D},r)$.

The matrix t conditional posteriors $p(B|\alpha,\mathcal{D},r)$ and $p(\alpha|B,\mathcal{D},r)$ in Theorem 4.6 are easily sampled using, e.g., the algorithm in Bauwens et al. (1999). Even though the second approach samples α in addition to β it is likely to be faster than the first approach, which requires draws from a 1-1 poly-t distribution for each of the cointegration vectors (for an algorithm, see Bauwens and Richard, 1985). The third approach is clearly not as fast as the second but has the advantage of yielding both the posterior distribution of the cointegration rank and the joint posterior $p(\alpha,\beta,\Psi,\Sigma|\mathcal{D},r)$ at the same time.

6. AN ILLUSTRATION

A single data set of length $T = 100$ was simulated from a bivariate model, without short-run dynamics and constant term, with parameters $\alpha = (0,0.1)$, $\beta = (1,-1)$, and $\Sigma = I_2$. Note that α is close to the zero vector and the model is thus close to the zero rank model. This difficult setup has been chosen to accentuate some features of the posterior distribution in cointegration models that were initially raised by Kleibergen and van Dijk (1994). The simulated time series are displayed in Figure 2.

The sequential testing procedure based on the so-called trace test (Johansen, 1995) estimates the cointegration rank to $r = 0$ and $r = 2$ on the 1% and 5% significance levels, respectively. The maximum eigenvalue test (Johansen, 1995) fails to reject the zero rank hypothesis at the 5% level but rejects $r = 1$ when tested against $r = 2$. The Bayesian information criterion (BIC) derived by Schwarz (1978) favors $r = 0$. The zero rank model is also favored by the posterior information criterion (PIC) (Chao and Phillips, 1999), whereas two other well-known information criteria, the Akaike information criterion (AIC) (Akaike, 1974) and the Hannan and Quinn information criterion (HQ) (Hannan and Quinn, 1979), are both in favor of the full rank model. The inconclusive evidence regarding the cointegration rank is of course expected as we purposely simulated data from a very difficult parametric setup.

To compute the posterior distribution of the cointegration rank, a uniform distribution on the ranks was used a priori, q was set to 4, and the maximum likelihood estimate

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

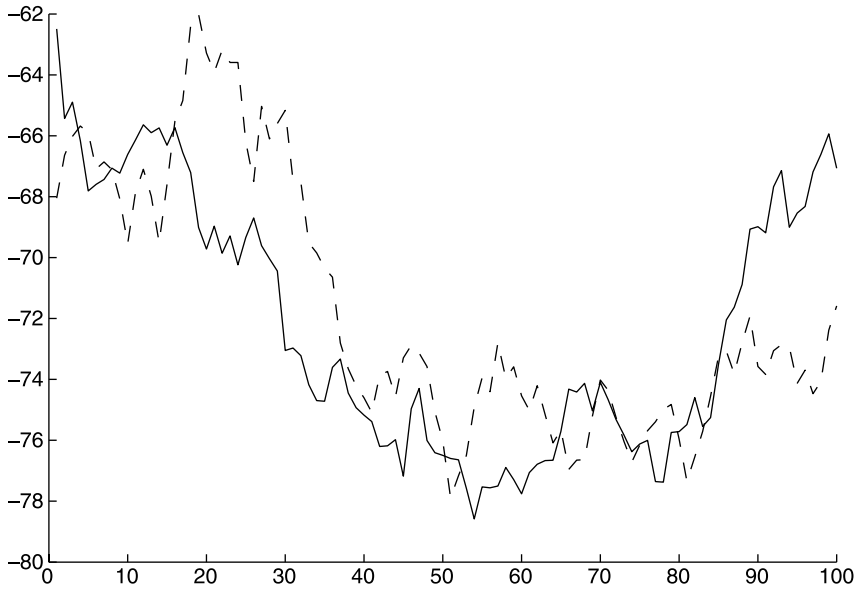

FIGURE 2. The simulated bivariate process.

$$\hat{\tilde{\Sigma}} = \begin{pmatrix} 0.83 & -0.10 \\ -0.10 & 1.02 \end{pmatrix}$$

was used for A as discussed in Section 3.1; other choices of A with larger positive and negative off-diagonal elements had only minor effects on the results. Note that as $\hat{\tilde{\Sigma}} \approx I_2$, $\sigma = v^{-1/2}$ corresponds roughly to the prior standard deviation of $\tilde{\alpha}$ as can be seen from (3.4).

Figure 3 displays the posterior probabilities of the three possible cointegration ranks as a function of σ. The MLI algorithm based on 25,000 draws from the marginal Gibbs sampler in Theorem 4.6 (see Section 5.3) was used for the computations. For small values of σ, the full rank model is most probable a posteriori, and as σ grows the posterior mass shifts rather quickly first in favor of $r = 1$ and subsequently to the zero rank model. The behavior of $p(r|\mathcal{D})$ as a function of σ follows the usual pattern in Bayesian analysis where the prior distributions of the model parameters in the larger models (higher rank) are centered over the smallest model ($r = 0$); see the discussion following Theorem 3.7. For such priors, the logic of Bayesian inference dictates the following intuitively reasonable behavior at the extremes of σ: $p(r|\mathcal{D}) \to p(r)$ for all r as $\sigma \to 0$ (all models/hypotheses approach the zero rank model) and

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

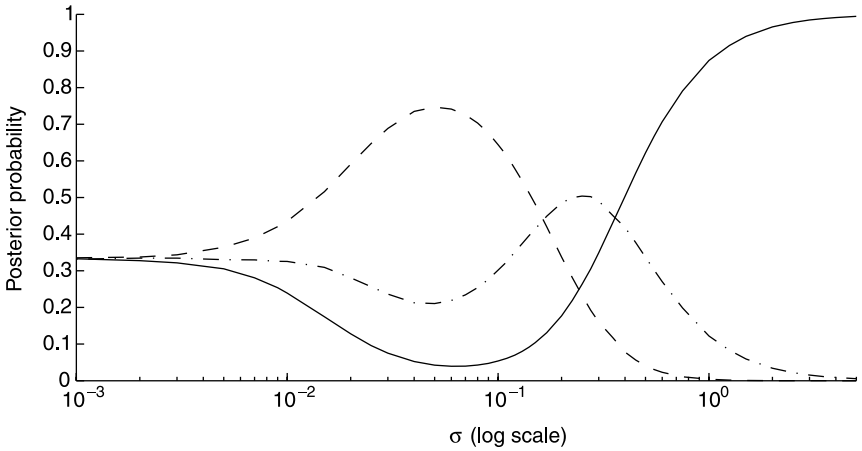

FIGURE 3. Posterior probabilities of the three possible cointegration ranks: $r = 0$ (——), $r = 1$ (–·–), and $r = 2$ (– – –) as a function of $\sigma = v^{-1/2}$.

$$p(r|\mathcal{D}) \to \begin{cases} 1 & \text{for } r = 0 \\ 0 & \text{for } r > 0 \end{cases} \quad \text{as } \sigma \to \infty$$

(all models with $r > 0$ give too much weight to regions in parameter space that are grossly at odds with the data), both of which are clearly borne out in Figure 3.

Note also from Figure 3 that the unit rank model is the most probable model only in the rather narrow interval $\sigma \in (0.16, 0.37)$. This fits well with the behavior of the traditional methods discussed earlier, which all favored either $r = 0$ or $r = 2$.

To investigate the efficiency of the three methods for computing the posterior distribution of the cointegration rank proposed in Section 5 we compute the marginal likelihood of $r = 1$ for different number of iterations of the respective algorithm. The matrix Cauchy density is used as importance function, and the marginal Gibbs sampler is used in the MLI algorithm. For each pair of methods and number of iterations we repeated the estimation 10,000 times. The upper graph in Figure 4 displays the evolution of the mean of the estimates $\hat{p}(\mathcal{D}|r = 1)$ over the 10,000 replications. The lower graph gives the numerical standard error of the estimators. Two main observations from Figure 4 are (i) the Monte Carlo integration approach converges extremely slowly toward the true value and (ii) the MLI algorithm outperforms the importance sampling method, despite the fact that the marginal posterior of β is symmetric and unimodal (see Figure 5) and is therefore favorable for the importance sampling algorithm. Even if we adjust for the faster execution time of the importance sampling approach (roughly three times faster than the MLI algorithm when the number of iterations exceeds 1,000), the MLI algorithm is still the preferred method.

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

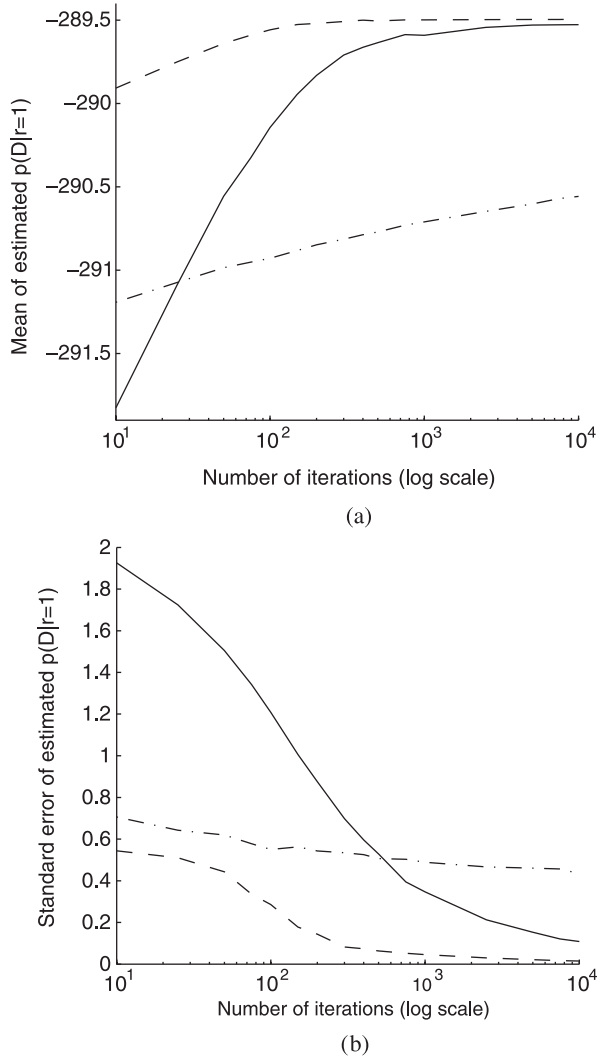

FIGURE 4. (a) Mean and (b) standard error of the estimated $p(\mathcal{D}|r)$ as a function of the number of iterations used in the three numerical algorithms: Monte Carlo integration (– · –), importance sampling (——), and MLI approach (– – –)

To discuss the issue of local nonidentification, the simulated data set is analyzed conditional on $r = 1$. The solid curves in Figure 5 display the inferences for α_1, α_2, and B. Figure 6 gives the prior and posterior distribution of the unrestricted eigenvalue of the companion matrix; see Section 3.2.

The local mode at point zero in the marginal posterior of α_2 in Figure 5 (which is actually an asymptote and thereby a global mode, a fact not visible in the

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

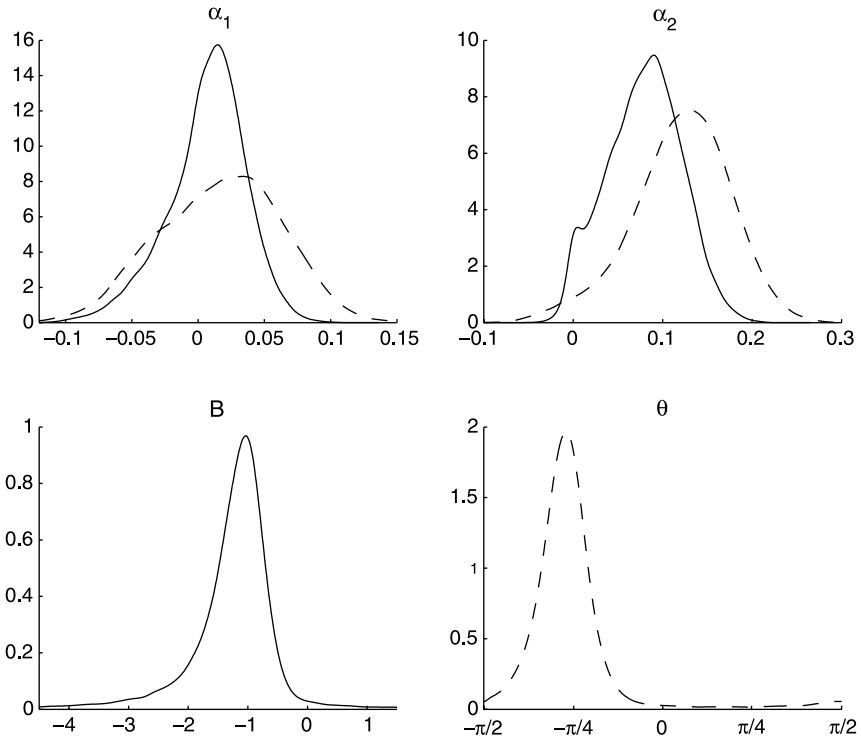

FIGURE 5. The posterior distribution of α and β for $\sigma = 0.5$ conditional on $r = 1$ in both the linear (——) and the orthonormal (– – –) normalizations. Here $\theta = \arctan(B)$ is the angle of the cointegration vector in the orthonormal normalization. In the density estimation, 2% of the draws from each tail of the posterior distribution of B were excluded.

figure because of the numerical approximation of the posterior; see the discussion that follows) is an effect of the local nonidentification discussed in Kleibergen and van Dijk (1994). They pointed out that when $\alpha = (0,0)'$, β drops out of the likelihood function and the likelihood is then constant along the B-axis (which has infinite length) and all values for B are observationally equivalent; B is said to be *locally nonidentified* when $\alpha = (0,0)'$. The posterior distribution based on the prior in (3.1) has the same property as it is flat in the direction of B when α is the zero vector. This is illustrated in Figures 7 and 8, which show the joint posterior density of α_2 and B for the simulated data set. Note how the conditional variance of B grows as $\alpha_2 \to 0$. The posterior variance of B given $\alpha = 0$ is actually infinite, as can be seen from the second part of Theorem 4.6. This of course is as it should be: if the processes do not react at all to past deviations from the equilibrium, then the data are necessarily uninformative regarding the cointegration vector.

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

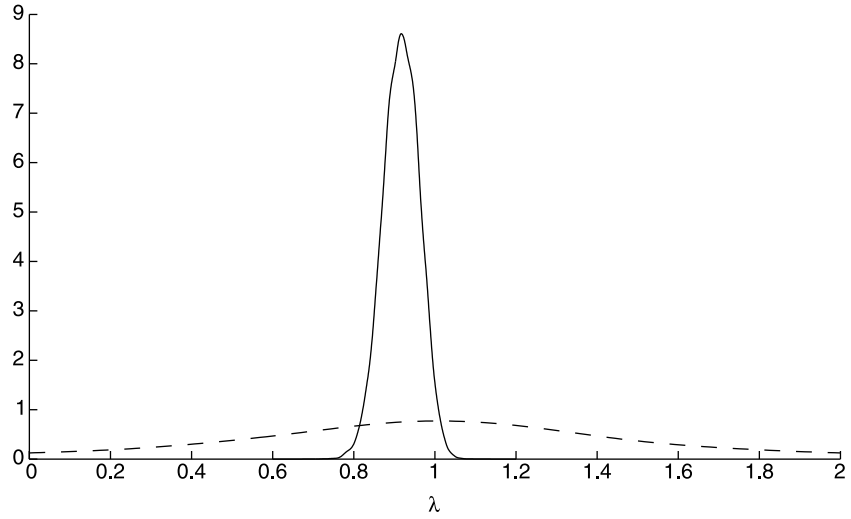

FIGURE 6. Prior ($\sigma = 0.5$, – – –) and posterior (——) distribution of the unrestricted eigenvalue of the companion matrix.

Kleibergen and van Dijk (1994) argue that this local nonidentification causes problems for a Bayesian analysis with uniform improper priors on α and B. Their argument is as follows: the marginal posterior of α is obtained by integrating the posterior $p(\alpha, B|\mathcal{D})$ with respect to B. As the posterior under a uni-

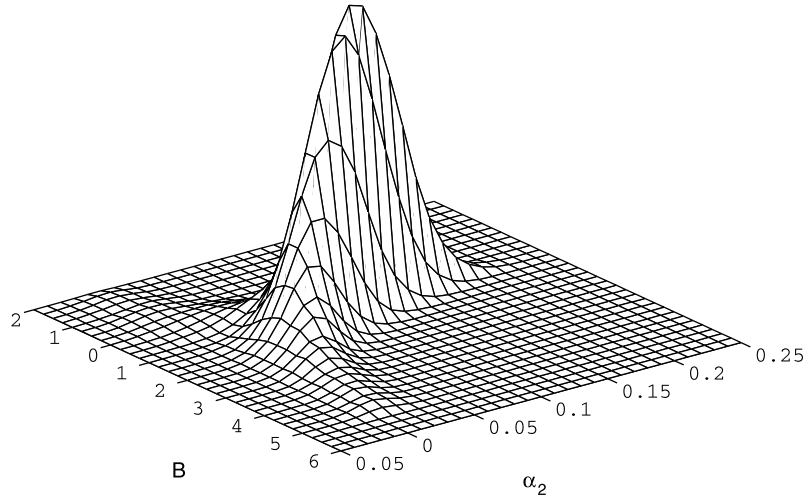

FIGURE 7. Joint posterior density of α_2 and B for $\sigma = 0.5$ conditional on $r = 1$.

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

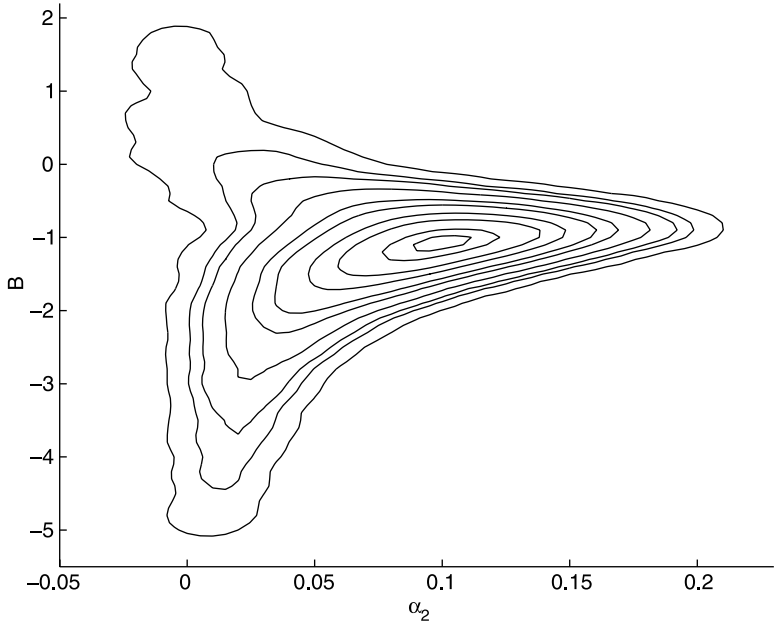

FIGURE 8. Contours of equal density height in the joint posterior distribution of α_2 and B for $\sigma = 0.5$ conditional on $r = 1$.

form prior is flat along the B-axis when $\alpha = (0,0)'$, the marginal posterior density of α in the point $\alpha = (0,0)'$ is proportional to the integral of a constant over an unbounded region $(-\infty < B < \infty)$, i.e., infinity. The marginal posterior of α is thus expected to have an asymptote in the point $(0,0)'$ that is entirely created by the local nonidentification.

Kleibergen and van Dijk suggest the Jeffreys prior to counterattack the unwanted asymptote as this prior is zero in the locally nonidentified points. The prior in Kleibergen and Paap (2002) has the same property.

Our view on the local nonidentification problem is best illustrated by transforming the posterior results so that β is restricted to a half-circle with unit radius, i.e., parameterizing β as in (3.2). This change in normalization is accomplished by the transformation $\theta = \arctan B$ and $\tilde{\alpha} = \alpha(1 + B^2)^{1/2}$; note that the product $\alpha\beta'$ is unchanged. The dashed curves in Figure 5 display the marginal posteriors in the new normalization. Note that there is no longer a mode at $\tilde{\alpha}_2 = 0$ after the transformation.

To explain this effect, note that B is a ratio of the two elements of β and that the tails in the marginal posterior of B are therefore heavy. Heavy tails in $p(B|\mathcal{D}, r)$ correspond to very small values for α, in the sense that a large β must be matched by a small α to keep the product $\Pi = \alpha\beta'$ at a reasonable magnitude. When we transform to the more natural orthonormal normalization

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

we are multiplying α with $(1 + B^2)^{1/2}$, which is large if B is drawn far out in the tails of $p(B|\mathcal{D}, r)$ and has the effect of spreading out the extra mode at $\alpha = (0,0)'$ and thereby producing a more well-behaved surface.

Alternatively, because the value of the marginal posterior of α in the point zero is proportional to the volume of the parameter region of β, this is a finite number if the normalization of β in (3.2) is used as θ is bounded. More generally, the volume of the Grassman manifold is finite (James, 1954) and there will be no asymptotes in the marginal posterior of $\tilde{\alpha}$.

Theorem 3.5 and the proof of Theorem 3.7 together show that the prior on $\tilde{\beta}$, the orthonormal matrix of cointegration vectors, is uniformly distributed over the Grassman manifold independently of $\tilde{\alpha}$. This means that the prior on $\tilde{\beta}$ conditional on $\tilde{\alpha} = 0$ is still uniform over the Grassman manifold. Thus, given the information that $\tilde{\alpha} = 0$, the prior in (3.1) represents the belief that every possible cointegration space of dimension r has the same probability a priori. This seems sensible.

One of the referees correctly pointed out that although the marginal prior on α is integrable, it has an asymptote in the point $\alpha = 0$. This is entirely natural, using the same argument as before for the posterior, as the heavy tails in the implied matrix Cauchy prior on B (a consequence of the uniformity of $\text{sp}(\beta)$ over the Grassman manifold) must again be matched by very small values on α to keep $\Pi = \alpha\beta'$ (whose interpretation does not, in contrast to α and β, depend on the chosen normalization) at a reasonable magnitude. As mentioned earlier, the linear normalization is a computationally convenient, but rather unnatural, way to solve the identification problem, and we have argued that the properties of the prior distribution are more clearly understood in the orthonormal normalization. With this in mind, note that the marginal prior on $\tilde{\alpha}$ follows a well-behaved matrix t distribution; see Section 3.1.

7. CONCLUDING REMARKS

This paper has introduced a practicable Bayesian analysis of cointegration based on a prior that is convenient both in elicitation and computation and could serve as a standard for inference reporting. The posterior distributions of both the cointegration rank and the model parameters conditional on the rank are obtained from the same Gibbs sampler.

Although a reference prior provides a good starting point in an analysis, and usually ends up in the final communication of results as a benchmark, it is clearly important to move beyond the reference case and consider more informative priors. Several informative distributions on the Grassman manifold are available for this purpose (see, e.g., Mardia and Jupp, 2000), and the major challenge is the construction of numerical algorithms for evaluating the posterior distribution.

The focus here has been on the case of just-identifying restrictions on β. The special case where the same overidentifying restrictions are imposed on each

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

of the cointegration vectors has the same geometry of the parameter space as the just-identified case, and all the results in this paper thus apply. We are currently working on the extension to general overidentifying restrictions on β and a Bayesian analysis of the validity of such restrictions within the framework proposed here.

REFERENCES

Ahn, S.K. & G.C. Reinsel (1990) Estimation for partially non-stationary multivariate autoregressive processes. *Journal of the American Statistical Association* 85, 813–823.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19, 716–723.

Bauwens, L. & P. Giot (1998) A Gibbs sampler approach to cointegration. *Computational Statistics* 13, 339–368.

Bauwens, L. & M. Lubrano (1996) Identification restrictions and posterior densities in cointegrated Gaussian VAR systems. In T.B. Fomby & R.C. Hill (eds.), *Advances in Econometrics*, vol. 11, part B, pp. 3–28. JAI Press.

Bauwens, L., M. Lubrano, & J.-F. Richard (1999) *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.

Bauwens, L. & J.-F. Richard (1985) A 1-1 poly-*t* random variable generator with application to Monte Carlo integration. *Journal of Econometrics* 29, 19–46.

Bauwens, L. & H.K. van Dijk (1990) Bayesian limited information analysis revisited. In Gabszewicz, J.J., Richard, J.-F., & Wolsey, L. (eds.), *Economic Decision-Making: Games, Econometrics and Optimisation*, pp. 385–424. North-Holland.

Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag.

Box, G.E.P. & G.C. Tiao (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley.

Chao, J.C. & P.C.B. Phillips (1999) Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* 91, 227–271.

Chib, S. (1995) Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90, 1313–1321.

Corander, J. & M. Villani (2004) Bayesian assessment of dimensionality in reduced rank regression. *Statistica Neerlandica* 58, 255–270.

Dickey, J.M. (1967) Matric-variate generalizations of the multivariate *t* distribution and the inverted multivariate *t* distribution. *Annals of Mathematical Statistics* 38, 511–518.

Dickey, J.M. (1968) Three multidimensional integral identities with Bayesian applications. *Annals of Mathematical Statistics* 39, 1615–1627.

Drèze, J.H. (1977) Bayesian regression analysis using poly-*t* densities. *Journal of Econometrics* 6, 329–354.

Drèze, J.H. & J.-F. Richard (1983) Bayesian analysis of simultaneous equation systems. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 1.

Doan, T., R.B. Litterman, & C.A. Sims (1984) Forecasting and conditional projection using realistic prior distributions. *Econometrics Reviews* 3, 1–100.

Engle, R.F. & C.W.J. Granger (1987) Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.

Geweke, J. (1989) Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57, 1317–1340.

Geweke, J. (1996) Bayesian reduced rank regression in econometrics. *Journal of Econometrics* 75, 121–146.

Hannan, E.J. & B.J. Quinn (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B* 41, 190–195.

Harville, D.A. (1997) *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag.

James, A.T. (1954) Normal multivariate analysis and the orthogonal group. *Annals of Mathematical Statistics* 25, 40–74.

Jeffreys, H. (1961) *Theory of Probability*, 3rd ed. Oxford University Press.
Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
Johansen, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
Kleibergen, F. & R. Paap (2002) Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics* 111, 223–249.
Kleibergen, F. & H.K. van Dijk (1994) On the shape of the likelihood/posterior in cointegration models. *Econometric Theory* 10, 514–551.
Kloek, T. & H.K. van Dijk (1978) Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica* 46, 1–19.
Litterman, R.B. (1986) Forecasting with Bayesian vector autoregressions—Five years of experience. *Journal of Business & Economic Statistics* 4, 25–38.
Luukkonen, R., A. Ripatti, & P. Saikkonen (1999) Testing for a valid normalization of cointegration vectors in vector autoregressive processes. *Journal of Business & Economic Statistics* 17, 195–204.
Mardia, K.V. & P.E. Jupp (2000) *Directional Statistics*. Wiley.
Phillips, P.C.B. (1989) Spherical matrix distributions and Cauchy quotients. *Statistics and Probability Letters* 8, 51–53.
Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
Phillips, P.C.B. (1994) Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models. *Econometrica* 62, 73–93.
Phillips, P.C.B. (1996) Econometric model determination. *Econometrica* 64, 763–812.
Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.
Smith, A.F.M. & G.O. Roberts (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B* 55, 3–24.
Stock, J.H. & M.W. Watson (1988) Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.
Strachan, R.W. (2003) Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business & Economic Statistics* 21, 185–195.
Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* 22, 1701–1762.
Villani, M. (2000) Aspects of Bayesian Cointegration. Ph.D. thesis, Stockholm University, Sweden.
Villani, M. (2001a) Fractional Bayesian lag length inference in multivariate autoregressive processes. *Journal of Time Series Analysis* 22, 67–86.
Villani, M. (2001b) Bayesian prediction with cointegrated vector autoregressions. *International Journal of Forecasting* 17, 585–605.
Villani, M. (2001c) Bayesian Reference Analysis of Cointegration. Research report 2001:1, Department of Statistics, Stockholm University, Sweden. Available at www.statistics.su.se.
Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics*. Wiley.

APPENDIX: PROOFS

Proof of Lemma 3.4. From Lemma 3.3,

$$\mathrm{sp}\begin{pmatrix} I_r \\ B \end{pmatrix} \stackrel{d}{=} \mathrm{sp}\begin{pmatrix} I_r \\ N_2 N_1^{-1} \end{pmatrix},$$

where $\stackrel{d}{=}$ denotes equality in distribution and N_1 and N_2 are independent $r \times r$ and $(p-r) \times r$ matrices of independent $N(0,1)$ variables. Postmultiplication of an arbitrary matrix A by a nonsingular matrix does not affect $\mathrm{sp}(A)$. Thus, postmultiplying

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

$$\begin{pmatrix} I_r \\ N_2 N_1^{-1} \end{pmatrix}$$

by N_1, which is nonsingular with probability one, yields

$$\mathrm{sp}\begin{pmatrix} I_r \\ N_2 N_1^{-1} \end{pmatrix} = \mathrm{sp}\begin{pmatrix} N_1 \\ N_2 \end{pmatrix},$$

almost surely. The result now follows from Lemma 3.2. ■

Proof of Theorem 3.5. To obtain the marginal distribution of β, we first derive the marginal distribution of B. The joint prior of B and Σ is

$$p(B,\Sigma) = \int p(\alpha, B, \Sigma)\, d\alpha = c_r |\Sigma|^{-(p+r+q+1)/2} \operatorname{etr}(\Sigma^{-1}A) \int \operatorname{etr}(\Sigma^{-1} v\alpha\beta'\beta\alpha')\, d\alpha.$$

Substituting the relation (Harville, 1997, Theorem 16.2.2)

$$\operatorname{tr}(\Sigma^{-1} v\alpha\beta'\beta\alpha') = \operatorname{vec}(\alpha)'(\beta'\beta \otimes v\Sigma^{-1})\operatorname{vec}(\alpha)$$

and integrating with respect to α using properties of the normal distribution we obtain

$$\begin{aligned} p(B,\Sigma) &= c_r |\Sigma|^{-(p+r+q+1)/2} \operatorname{etr}(\Sigma^{-1}A)(2\pi)^{pr/2} |\beta'\beta \otimes v\Sigma^{-1}|^{-1/2} \\ &= c_r (2\pi/v)^{pr/2} |\Sigma|^{-(p+q+1)/2} \operatorname{etr}(\Sigma^{-1}A) |I_r + B'B|^{-p/2}. \end{aligned}$$

This shows that B and Σ are independent and marginally $B \sim t_{(p-r)\times r}(0, I_{p-r}, I_r, 1)$. Thus, using Lemma 3.4, β is uniformly distributed over $\mathcal{G}_{r,p-r}$. ■

Proof of Theorem 3.7. From (3.3)

$$\alpha|\beta,\Sigma \sim N_{p\times r}[0, (\beta'\beta)^{-1}, v^{-1}\Sigma].$$

As $\tilde{\alpha} = \alpha(\beta'\beta)^{1/2}$ we have (see, e.g., Bauwens et al., 1999, p. 302)

$$\tilde{\alpha}|\tilde{\beta},\Sigma \sim N_{p\times r}(0, I_r, v^{-1}\Sigma).$$

The density $p(\tilde{\alpha}|\tilde{\beta},\Sigma)$ is not a function of $\tilde{\beta}$, and we may write $\tilde{\alpha}|\Sigma \sim N_{p\times r}(0, I_r, v^{-1}\Sigma)$. The statement of the theorem now follows from the usual independence property of the multivariate normal distribution. ■

Proof of Theorem 4.2. It is well known that the likelihood function is invariant with respect to normalization (Johansen, 1995). It is therefore sufficient to prove that the prior is invariant. Let $\mathcal{N}_1$ denote that β is normalized on the r first variables and $\mathcal{N}_2$ that β is normalized on variables $1,2,\ldots,r-1$ and $r+1$, i.e., the change in normalizing variables from $\mathcal{N}_1$ to $\mathcal{N}_2$ is accomplished by replacing the last variable of the normalizing set with the first variable in the nonnormalizing set. It will be evident that the lemma holds generally under any valid change of normalizing variables. We shall first prove that $J(\bar{\alpha},\bar{\beta},\Sigma \to \alpha,\beta,\Sigma) = 1$. Let

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

$$B = \begin{pmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,r} \\ b_{2,1} & b_{2,2} & \dots & b_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p-r,1} & b_{p-r,2} & \dots & b_{p-r,r} \end{pmatrix}$$

denote the matrix of free coefficients in β under $\mathcal{N}_1$. The transformation matrix in this case is

$$U = \begin{pmatrix} J \\ b_{1,1}, b_{1,2}, \dots, b_{1,r} \end{pmatrix}, \quad \text{if } r > 1 \quad \text{and} \quad U = b_{1,1} \quad \text{if } r = 1,$$

where J denotes the $r-1$ first rows of I_r. It is easy to see that $|U| = b_{1,r}$ and

$$U^{-1} = \begin{pmatrix} J \\ -b_{1,1}b_{1,r}^{-1}, -b_{1,2}b_{1,r}^{-1}, \dots, b_{1,r}^{-1} \end{pmatrix}, \quad \text{if } r > 1 \quad \text{and} \quad U^{-1} = b_{1,1}^{-1} \quad \text{if } r = 1.$$

Note that the restriction to *valid* changes in normalizing variables is equivalent to the condition $b_{1,r} \neq 0$, which ensures the existence of U^{-1}. It is straightforward to check that U actually produces the intended change in normalization and that the matrix of free coefficients under $\mathcal{N}_2$ is

$$\bar{B} = \begin{pmatrix} -b_{1,1}b_{1,r}^{-1} & -b_{1,2}b_{1,r}^{-1} & \dots & b_{1,r}^{-1} \\ b_{2,1} - b_{2,r}b_{1,1}b_{1,r}^{-1} & b_{2,2} - b_{2,r}b_{1,2}b_{1,r}^{-1} & \dots & b_{2,r}b_{1,r}^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p-r,1} - b_{p-r,r}b_{1,1}b_{1,r}^{-1} & b_{p-r,2} - b_{p-r,r}b_{1,2}b_{1,r}^{-1} & \dots & b_{p-r,r}b_{1,r}^{-1} \end{pmatrix}. \tag{A.1}$$

The change in normalization from $\mathcal{N}_2$ to $\mathcal{N}_1$ is thus given by the transformation $\bar{\alpha}, \bar{B}, \Sigma \to \alpha, B, \Sigma$, where $\bar{\alpha} = \alpha U'$. The Jacobian of this transformation is

$$J(\bar{\alpha}, \bar{\beta}, \Sigma \to \alpha, \beta, \Sigma) = \begin{vmatrix} \dfrac{d\,\mathrm{vec}(\bar{\alpha})}{d\,\mathrm{vec}(\alpha)'} & \dfrac{d\,\mathrm{vec}(\bar{\alpha})}{d\,\mathrm{vec}(B)'} \\ \dfrac{d\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(\alpha)'} & \dfrac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(B)'} \end{vmatrix} = |U|^p \left| \frac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(B)'} \right|, \tag{A.2}$$

as Σ is unaffected by the transformation, $d\,\mathrm{vec}(\bar{B})/d\,\mathrm{vec}(\alpha)' = 0$ and $d\,\mathrm{vec}(\bar{\alpha})/d\,\mathrm{vec}(\alpha)' = U \otimes I_p$. Let b_i and $\bar{b}_i$ denote the ith columns of B and $\bar{B}$, respectively. It is easily seen from (A.1) that $d\bar{b}_i/db_j = 0$ for $i > j$, and thus

$$\left| \frac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(B)'} \right| = \left| \frac{d\bar{b}_1}{db_1} \right| \left| \frac{d\bar{b}_2}{db_2} \right| \cdots \left| \frac{d\bar{b}_r}{db_r} \right|, \tag{A.3}$$

where

$$\frac{d\bar{b}_i}{db_i} = \begin{pmatrix} -b_{1,r}^{-1} & 0 \\ \cdot & I_{p-r-1} \end{pmatrix}, \quad \text{for } i = 1, \dots, r-1, \quad \text{and} \quad \frac{d\bar{b}_r}{db_r} = \frac{d\bar{b}_1}{db_1} b_{1,r}^{-1} \tag{A.4}$$

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

and the dot replaces an expression that is unnecessary to calculate. Thus, from (A.2)–(A.4)

$$J(\bar{\alpha},\bar{\beta},\Sigma\to\alpha,\beta,\Sigma)=|U|^{p}\left|\frac{d\bar{b}_1}{db_1}\right|\left|\frac{d\bar{b}_2}{db_2}\right|\cdots\left|\frac{d\bar{b}_r}{db_r}\right|=b_{1,r}^{p}(b_{1,r}^{-1})^{r}(b_{1,r}^{-1})^{p-r}=1.$$

Now, because the transformation T_U is one-to-one and differentiable, the implied prior obtained from the transformation from $\mathcal{N}_2$ to $\mathcal{N}_1$ is

$$\begin{aligned}p_{\mathcal{N}_2\to\mathcal{N}_1}(\alpha,\beta,\Sigma)&=p(\bar{\alpha},\bar{\beta},\Sigma)J(\bar{\alpha},\bar{\beta},\Sigma\to\alpha,\beta,\Sigma)\\&=c_r|\Sigma|^{-(p+r+q+1)/2}\operatorname{etr}[\Sigma^{-1}(A+v\alpha U'U^{-1\prime}\beta'\beta U^{-1}U\alpha')]\\&=c_r|\Sigma|^{-(p+r+q+1)/2}\operatorname{etr}[\Sigma^{-1}(A+v\alpha\beta'\beta\alpha')],\end{aligned}$$

which is exactly the same density as would have been obtained by specifying the prior directly in the $\mathcal{N}_1$ normalization. ■

Proof of Theorem 4.5. All full conditional posteriors are proportional to the likelihood function multiplied with the prior in (3.1), i.e., proportional to

$$|\Sigma|^{-(T+p+r+q+1)/2}\operatorname{etr}[\Sigma^{-1}(E'E+A+v\alpha\beta'\beta\alpha')], \tag{A.5}$$

where $E=Y-X\beta\alpha'-Z\Psi$.

It follows directly from (A.5) that the full conditional posterior of Σ is the $IW_p(E'E+A+v\alpha\beta'\beta\alpha',T+q+r)$ density.

The full conditional posterior of Ψ follows from the treatment of the multivariate regression in Zellner (1971); see also Geweke (1996).

To obtain the full conditional posterior of B, let $X=(X_1,X_2)$, where X_1 contains the r first columns of X and X_2 contains the $p-r$ remaining ones, and $W=Y-X_1\alpha'-Z\Psi$. The full conditional likelihood of B is then

$$\begin{aligned}p(\mathcal{D}|\alpha,\beta,\Psi,\Sigma,r)&\propto\operatorname{etr}[\Sigma^{-1}(W-X_2B\alpha')'(W-X_2B\alpha')]\\&=\exp\left\{-\frac{1}{2}[\operatorname{vec}(W\Sigma^{-1/2})-H\operatorname{vec}B]'[\operatorname{vec}(W\Sigma^{-1/2})-H\operatorname{vec}B]\right\},\end{aligned}$$

where $H=(\Sigma^{-1/2}\alpha\otimes X_2)$. Thus,

$$p(\mathcal{D}|\alpha,\beta,\Psi,\Sigma,r)\propto\exp\left\{-\frac{1}{2}(\operatorname{vec}B-\operatorname{vec}\hat{B})'(\alpha'\Sigma^{-1}\alpha\otimes X_2'X_2)(\operatorname{vec}B-\operatorname{vec}\hat{B})\right\}, \tag{A.6}$$

where, after some simplifications,

$$\operatorname{vec}\hat{B}=\operatorname{vec}[(X_2'X_2)^{-1}X_2'W\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}].$$

The prior in (3.1) can be rewritten as

$$p(\alpha,B,\Sigma,\Psi|r)\propto\exp\left\{-\frac{1}{2}(\operatorname{vec}B)'(\alpha'\Sigma^{-1}\alpha\otimes vI_{p-r})(\operatorname{vec}B)\right\}. \tag{A.7}$$

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

By multiplying (A.6) by $p(\alpha, B, \Sigma, \Psi | r)$ in (A.7) and completing the square in the exponential (see Box and Tiao, 1973, Lemma 1, p. 418), it is seen that

$$p(B|\alpha, \Psi, \Sigma, \mathcal{D}, r) \propto \exp\left\{-\frac{1}{2}(\text{vec}\, B - \text{vec}\, \mu_B)'\Omega_B^{-1}(\text{vec}\, B - \text{vec}\, \mu_B)\right\},$$

where $\Omega_B^{-1} = \alpha'\Sigma^{-1}\alpha \otimes (X_2'X_2 + vI_{p-r})$ and

$$\text{vec}\, \mu_B = \text{vec}[(X_2'X_2 + vI_{p-r})^{-1}X_2'W\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}].$$

Thus, $B|\alpha, \Psi, \Sigma, D \sim N_{(p-r)\times r}[\mu_B, (\alpha'\Sigma^{-1}\alpha)^{-1}, (X_2'X_2 + vI_{p-r})^{-1}]$.

The full conditional posterior of α is derived in essentially the same way as the full conditional posterior of B. ■

Proof of Theorem 4.6. Integrating (A.5) with respect to Ψ and Σ yields

$$\begin{aligned} p(\alpha|\beta, \mathcal{D}, r) &\propto |(Y - X\beta\alpha')'M_Z(Y - X\beta\alpha') + A + v\alpha\beta'\beta\alpha'|^{-(T+q+r-d)/2} \\ &= |A + Y'M_Z(Y - X\beta\hat{\alpha}') + (\alpha' - \hat{\alpha}')'(\beta'C_1\beta)(\alpha' - \hat{\alpha}')|^{-(T+q+r-d)/2}, \end{aligned}$$

where $\hat{\alpha}' = (\beta'C_1\beta)^{-1}\beta'X'M_ZY$. Thus, $\alpha' \sim t_{r\times p}[\hat{\alpha}', (\beta'C_1\beta)^{-1}, A + Y'M_Z(Y - X\beta\hat{\alpha}'), T + q - (d + p) + 1]$. From Box and Tiao (1973, p. 442), $\alpha \sim t_{p\times r}[\hat{\alpha}, A + Y'M_Z(Y - X\beta\hat{\alpha}'), (\beta'C_1\beta)^{-1}, T + q - (d + p) + 1]$.

Because $\Pi = \alpha\beta'$, the posterior of β conditional on α can be written

$$\begin{aligned} p(\beta|\alpha, \mathcal{D}, r) &\propto |(Y - X\Pi')'M_Z(Y - X\Pi') + A + v\Pi\Pi'|^{-(T+q+r-d)/2} \\ &= |S + (\Pi - \hat{\Pi})C_1(\Pi - \hat{\Pi})'|^{-(T+q+r-d)/2}, \end{aligned}$$

where $S = A + Y'M_ZY - \hat{\Pi}C_1\hat{\Pi}'$ and $\hat{\Pi} = Y'M_ZXC_1^{-1}$. Thus,

$$\begin{aligned} p(\beta|\alpha, \mathcal{D}, r) &\propto |C_1^{-1} + (\alpha\beta' - \hat{\Pi})'S^{-1}(\alpha\beta' - \hat{\Pi})|^{-(T+q+r-d)/2} \\ &= |R + (\beta - \hat{\beta})(\alpha'S^{-1}\alpha)(\beta - \hat{\beta})'|^{-(T+q+r-d)/2} \\ &\propto |(\alpha'S^{-1}\alpha)^{-1} + (\beta - \hat{\beta})'R^{-1}(\beta - \hat{\beta})|^{-(T+q+r-d)/2}, \end{aligned} \tag{A.8}$$

where $\hat{\beta} = \hat{\Pi}'S^{-1}\alpha(\alpha'S^{-1}\alpha)^{-1}$ and $R = C_1^{-1} + \hat{\Pi}'S^{-1}\hat{\Pi} - \hat{\beta}(\alpha'S^{-1}\alpha)\hat{\beta}'$. Let $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')'$, where $\hat{\beta}_1$ contains the r first rows of $\hat{\beta}$ and $\hat{\beta}_2$ the $p - r$ remaining ones and R is conformably decomposed as

$$R = \begin{pmatrix} \underset{r\times r}{G_1} & \underset{r\times(p-r)}{G_2} \\ \underset{(p-r)\times r}{G_2'} & \underset{(p-r)\times(p-r)}{G_3} \end{pmatrix}.$$

By using the result (see, e.g., Harville, 1997)

$$R^{-1} = \begin{pmatrix} (G_1 - G_2G_3^{-1}G_2')^{-1} & -(G_1 - G_2G_3^{-1}G_2')^{-1}G_2G_3^{-1} \\ -G_3^{-1}G_2'(G_1 - G_2G_3^{-1}G_2')^{-1} & (G_3 - G_2'G_1^{-1}G_2)^{-1} \end{pmatrix}$$

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press

it is straightforward to show that

$$(\beta - \hat{\beta})'R^{-1}(\beta - \hat{\beta}) = (I_r - \hat{\beta}_1)'G_1^{-1}(I_r - \hat{\beta}_1) + (B - \hat{B})'(G_3 - G_2'G_1^{-1}G_2)^{-1}(B - \hat{B}),$$

where $\hat{B} = \hat{\beta}_2 + G_2'G_1(I_r - \hat{\beta}_1)$. From (A.8)

$$p(B|\alpha, \mathcal{D}, r) \propto |C_3 + (B - \hat{B})'(G_3 - G_2'G_1^{-1}G_2)^{-1}(B - \hat{B})|^{-(T+q+r-d)/2},$$

where $C_3 = (I_r - \hat{\beta}_1)'G_1^{-1}(I_r - \hat{\beta}_1) + (\alpha'S^{-1}\alpha)^{-1}$. This is proportional to the matrix t density in Theorem 4.6. ■

https://doi.org/10.1017/S026646660505019X Published online by Cambridge University Press