Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Econometrics 123 (2004) 307–325
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Bayesian analysis of the error correction model

Rodney W. Strachan[a,*], Brett Inder[b]

[a] *Department of Economics and Accounting, University of Liverpool, Liverpool L69 7ZH, UK*
[b] *Department of Econometrics and Business Statistics, Monash University, Victoria 3800, Australia*

Abstract

This paper presents a method for estimating the posterior probability density of the cointegrating rank of a multivariate error correction model. A second contribution is the careful elicitation of the prior for the cointegrating vectors derived from a prior on the cointegrating space. This prior obtains naturally from treating the cointegrating space as the parameter of interest in inference and overcomes problems previously encountered in Bayesian cointegration analysis. Using this new prior and Laplace approximation, an estimator for the posterior probability of the rank is given. The approach performs well compared with information criteria in Monte Carlo experiments.

JEL classification: C11; C32

Keywords: Cointegration; Posterior probability; Grassman manifold; Stiefel manifold; Error correction model

1. Introduction

Since its development by Granger (1983) and Engle and Granger (1987), the concept of cointegration has proven a valuable tool in economic analysis. An important consideration in cointegration is the accurate determination, or estimation, of the number of stationary combinations, r. Many classical tests have been developed for this purpose, although relatively few Bayesian tests exist. In an early study, Geweke (1996) proposed the use of predictive probabilities estimated using a Markov chain Monte Carlo (MCMC) method to determine the rank. Other work by Kleibergen and Paap (2002) (hereafter referred to as KP) also used an MCMC approach but produced a posterior distribution for the rank. While there have been a few other approaches (see

* Corresponding author. Department of Economics, Keele University, Keele STS 5BG, UK. Tel.: +44-151-795-3714; fax: +44-151-794-3028.

E-mail address: rstrach@liverpool.ac.uk (R.W. Strachan).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2003.12.004

for example Strachan, 2003), none of these has produced a simple, efficient test which performs consistently well.

The first aim of this paper is to present a Bayesian test for the rank which is simple to apply. The test is related to the classical trace test developed by Anderson (1951) for the general reduced rank regression model and applied to the cointegrating ECM by Johansen (1988, 1991). This test has similar performance to the trace test when used to estimate the number of stochastic trends, which is not surprising as it can be shown to be a simple function of the trace statistic for particular common priors.

The second aim of the paper is to develop both uninformative and informative priors for the cointegrating space. These priors are simple to develop, allow MCMC or Laplace approximations to integration, and avoid many of the issues encountered in the literature (and some not in the literature) with previous work on Bayesian cointegration analysis, particularly when uninformative priors are employed.

Important advantages of the Bayesian approach over the classical approach are the treatment of model uncertainty and greater flexibility to explicitly incorporate prior beliefs. Classical methods such as hypothesis tests on parameter values or the use of information criteria for model selection, result in subsequent inference being conditional upon the chosen model, regardless of the information content of 'nearby' models. Bayesian posterior probabilities, however, allow inference to be averaged over a range of models if the econometrician so desires.

The outline of the article is as follows. In Section 2 the model is described, and the likelihood, the priors and a general form for the posterior are given. An important contribution of this section, and indeed this paper, is the careful elicitation of the prior distribution on the cointegrating coefficients from a prior on the cointegrating space. An outline of the identifying restrictions which arise naturally from the prior elicitation is also provided. Section 3 introduces the inferential tool and the Bayes factor. The application of Laplace approximation is presented in Section 4. Monte Carlo experiments are reported in Section 5. Section 6 concludes.

As our Bayesian cointegration analysis involves integration over cointegrating spaces, we must introduce some relevant concepts related to matrix spaces and their measures. The following definitions are useful in the development of the prior and subsequently the posterior on the cointegrating space. The reader is referred to Muirhead (1982), Phillips (1994) and in particular James (1954) for more details on these matrix spaces.

We denote the space spanned by a matrix A by $sp(A)$. If an $r \times r$ matrix C is orthonormal, such that $C'C = I_r$, then this matrix is an element of the orthogonal group of $r \times r$ orthogonal matrices denoted by $O(r) = \{C(r \times r) : C'C = I_r\}$, that is $C \in O(r)$. A feature of these matrices is that each column vector is of unit length and as C varies over $O(r)$, each vector describes an r-dimensional sphere centred at the origin of the coordinate system. If an $n \times r$ matrix V is semi-orthonormal such that $V'V = I_r$, then this matrix is an element of the Stiefel manifold denoted by $V_{r,n} = \{V(n \times r) : V'V = I_r\}$, that is $V \in V_{r,n}$. As with C, each column vector of V is of unit length and as V varies over $V_{r,n}$, each vector describes an n-dimensional sphere centred at the origin

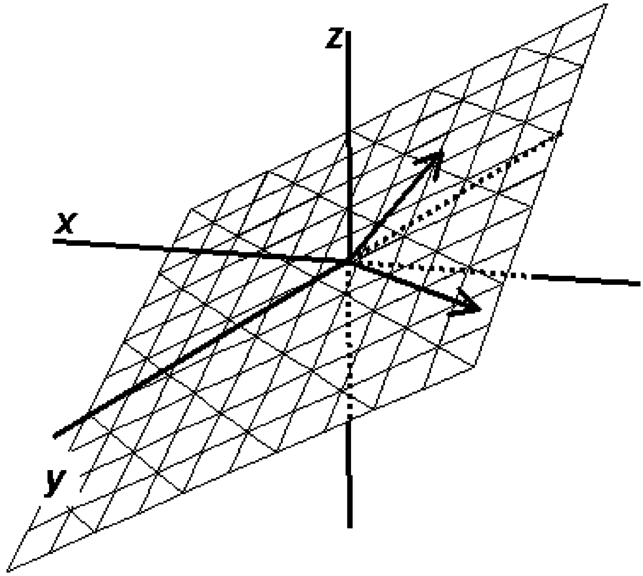

Fig. 1. Two dimensional plane in three-dimensional space representing $G_{2,1}$.

of the coordinate system. As the vectors of V are linearly independent (since they are orthogonal) the columns of V lie in—and in fact define—an r-dimensional plane in the n-dimensional coordinate system, and this plane passes through the origin of this coordinate system.

To demonstrate these matrices and their spaces, and to introduce the Grassman manifold, we use the following example. Consider the case of $r = 2$, $n = 3$, and the 3×2 matrix $V \in V_{2,3}$. Fig. 1 shows a three-dimensional coordinate system in which are drawn the two unit-length vectors in the matrix V. These two vectors define the two-dimensional plane which is shown as a cross-hatched plane in Fig. 1. This 2D plane, which we will denote by $\mathfrak{p}$, is one of the set of possible 2D planes that may exist in this coordinate system and this set is the $(n - r)r = 2$ dimensional Grassman manifold, denoted $G_{2,1}$ such that $\mathfrak{p} \in G_{2,1}$. For general r and n, we denote the Grassman manifold as $G_{r,n-r}$ and an element of this manifold is denoted as $\mathfrak{p} \in G_{r,n-r}$. As the vectors in V in Fig. 1 are allowed to move over the whole of the Stiefel manifold $V_{2,3}$ (that is, they are allowed to point in any of the directions from the origin) the plane they define ($\mathfrak{p}$) moves over the whole of the Grassman manifold $G_{2,1}$. Note that for any $V \in V_{r,n}$, there exists a plane $\mathfrak{p} = sp(V)$ such that $\mathfrak{p} \in G_{r,n-r}$. The cointegrating space for an n-dimensional system with cointegrating rank r is an example of an element of the $(n - r)r$-dimensional Grassman manifold.

A final piece of notation: throughout the paper let the jth largest eigenvalue of the matrix A be denoted $\lambda_j(A)$.

2. Model, priors and posteriors

The error correction model (ECM) of the $1 \times n$ vector time-series process $y_t = (y_{1t}, \dots y_{nt})$, $t = 1, \dots, T$, conditioning on the l observations $t = -l+1, \dots, 0$, is

$$\Delta y_t = y_{t-1}\beta\alpha + d_t\mu + \Delta y_{t-1}\Gamma_1 + \cdots + \Delta y_{t-l}\Gamma_l + \varepsilon_t \quad (1)$$

$$= z_{0,t} = z_{1,t}\beta\alpha + z_{2,t}\Phi + \varepsilon_t, \quad (2)$$

where $z_{0,t} = \Delta y_t = y_t - y_{t-1}$, $z_{1,t} = y_{t-1}$, $z_{2,t} = (d_t, \Delta y_{t-1}, \dots, \Delta y_{t-l})$, and $\Phi = (\mu', \Gamma_1', \dots, \Gamma_l')'$. The matrices β and α' are $n \times r$ and assumed to have rank r.

Of interest when considering the number of stochastic trends is the coefficient matrix β which is of dimension $n \times r$ and we have $rank(\beta\alpha) = r \leqslant n$. When $0 < r < n$, y_t is cointegrated, β is the matrix of cointegration coefficients and α is the matrix of factor loading coefficients or adjustment coefficients.

Finally, we introduce the following terms to simplify the expressions in the posteriors. Let $\tilde{z}_t = (z_{1,t}\beta \; z_{2,t})$, and the $(r+k) \times n$ matrix $B = [\alpha' \;\; \Phi']'$. The model may now be written as $z_{0,t} = \tilde{z}_t B + \varepsilon_t$.

In considering the priors in this section, we restrict ourselves to flat priors where possible, although consideration is given to informative priors when discussing the parameters of interest.

2.1. The prior for (Σ, B, r)

Throughout this paper, the prior for the rank r is $p(r) = (n+1)^{-1}$ and the standard diffuse prior for (Σ, Φ), $p(\Sigma, \Phi) \propto |\Sigma|^{-(n+1)/2}$, is used.

Each element of the matrix α has the real line as its support and α changes dimensions across the different models. Thus if the prior on α is $p(\alpha|\beta, r) = h(\alpha)\mathfrak{c}_r^{-1}$ where $\mathfrak{c}_r = \int h(\alpha)(\mathrm{d}\alpha)$, then clearly $\mathfrak{c}_r$ depends upon r. As discussed in O'Hagan (1995), the Bayes factor for a model with rank r to a model with rank r^*, $r \neq r^*$, from which the posterior probabilities are derived, is proportional to the ratio $\varsigma = \mathfrak{c}_{r^*}/\mathfrak{c}_r$ and therefore knowledge of ς is required. If an improper prior on α such as $h(\alpha) = 1$, were used, then $\mathfrak{c}_r = \infty$ but can be treated as an unspecified constant. Thus when we combine the prior with the likelihood, $L(\alpha)$, to form the posterior as $L(\alpha)h(\alpha)\mathfrak{c}_r^{-1} / \int L(\alpha)h(\alpha)\mathfrak{c}_r^{-1}(\mathrm{d}\alpha)$, then the constants $\mathfrak{c}_r$ cancel and as a result the posterior will be well defined. However, as ς will be unspecified, the resulting Bayes factor for the model rank cannot be obtained. For this reason a (weakly) informative proper prior for α must be used. In this article, the prior for α conditional upon (Σ, β, r) is normal with zero mean and covariance $\Sigma \otimes (\beta' \underline{H} \beta)^{-1}$ where in our case we choose $\underline{H} = \upsilon I_n$ and so $\beta' \underline{H} \beta = \upsilon I_r$. As υ approaches zero, the prior becomes less informative about the value of α. As is well known, the choice of the prior variance, represented by the value of υ, can influence inference on a sharp null such as the hypothesis on r in which we are interested. This effect in our model is made explicit in Section 3 in which the marginal posterior distributions of (β, r) in Eq. (8), and therefore the Bayes factor for r, can be seen to be a function of υ. For this reason in Section 5 we specify a range of values of υ for the Monte Carlo experiments.

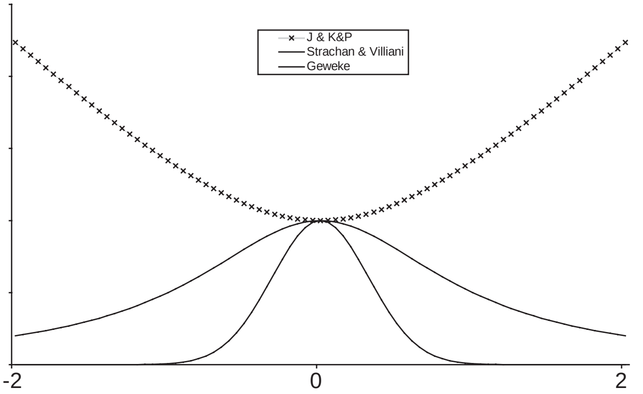

Fig. 2. Plot of priors for β_2 for the case $n=2$, $r=1$.

2.2. Eliciting a prior on β

2.2.1. Literature on priors for β

A range of priors for β have been proposed and their implications investigated in the literature. As β and α appear as a product in (2), r^2 restrictions need to be imposed on their elements to just identify these elements. These restrictions are commonly imposed upon β by assuming $c\beta$ is invertible for known matrix c (such as $c=[I_r\ 0]$) and the restricted β is $\bar{\beta}=\beta(c\beta)^{-1}$. Thus the free elements are collected in $\beta_2=c_\perp\bar{\beta}$ where $c_\perp c'=0$ and a prior is then specified for β_2.

Using the results of Drèze (1978), Bauwens and Lubrano (1996) show that with a flat prior the posterior for β_2 is a 1–1 poly-t density which is potentially bimodal and proper but, for $r=1$, has no moments. To demonstrate the alternative priors to which we refer, Fig. 2 plots some of the priors we will discuss for β_2 with $n=2$, $r=1$. In response to the issue of local nonidentification, Kleibergen and van Dijk (1994) (hereafter referred to as KVD) propose the approximate Jeffreys prior which downweights regions of local non-identification and leads to what they call balanced posteriors. An example of this prior assuming full rank α can be seen in Fig. 2 labelled J&KP. Although this is generally a rather complex prior, the important features of its general form are that it is convex in β_2 and at points of local non-identification ($\alpha=0$) it is zero.

To avoid the issues of nonuniqueness of the posterior and the issue of local non-identification KP, following the approach of Kleibergen and van Dijk (1998) in a related model, propose a prior derived from the Jacobian for the transformation from the parameters of the full rank model to α, β_2 and a parameter incorporating the rank reduction. This prior for β_2 (J and KP in Fig. 2 given a full rank value of α) has the same features as the approximate Jeffreys prior in that it is convex in β_2 and zero at points of local non-identification (e.g., $\alpha=0$).

Other priors have been proposed in the literature, such as Geweke (1996) who proposed the multivariate normal prior for β_2 and α, however a common feature of all these priors is that the prior has been placed upon β_2. A recent exception is Strachan (2003) where the uniform prior is specified for β using identifying restrictions similar to those of the Johansen maximum likelihood estimator. This approach has the advantage that posterior moments of the cointegrating vector always exist, this posterior is always proper with a unique mode thereby simplifying inference on the cointegrating space and the identifying restrictions will always be valid. However, as Strachan (2003) uses a data based prior, this is not a strictly Bayesian approach. To plot the density in the same space as the other priors, we require simply the Jacobian for the transformation from this specification to β_2 which has the form of a Cauchy density, $|I_r + \beta_2'\beta_2|^{-n/2}$ (Phillips, 1994). This Jacobian is plotted in Fig. 2 as Strachan and Villani.

Recent work by Villani (2001) has shifted the focus of prior attention from the cointegrating vectors to the actual parameter of interest, the cointegrating space. Villani (2001) places a uniform prior on the cointegrating space rather than the cointegrating vectors. He then goes on to prove that the implied prior for β_2 is Cauchy and this prior is plotted in Fig. 2 as Strachan and Villani.

We extend the work of Villani (2001) and thus begin with the prior on the cointegrating space, however, following Strachan (2003) we diverge from much of the literature in that we do not use linear identifying restrictions. In the following subsections we discuss some implications of using the linear identifying restrictions and then develop our priors for the cointegrating space.

2.2.2. Linear restrictions and the cointegrating space

A first issue with linear identifying restrictions is the specification of c. The practical problems in classical analysis of incorrectly selecting c are discussed in Boswijk (1996) and Luukkonen et al. (1999) and in Bayesian analysis by Strachan (2003). Assuming known c, the pathologies and complicating features (for analysis) of the posterior for β_2 with a flat prior, such as multimodality, nonexistence of moments and (under some specifications) non-integrability of the posterior have been detailed by KVD and Bauwens and Lubrano (1996). In addition, unpublished notes by Bauwens and Lubrano show the posterior is improper when another important and commonly employed restriction on (2), exogeneity, is imposed. Further, from the discussion on the prior for α it is clear a flat prior on β_2 cannot be employed to obtain posterior probabilities for r, since the dimensions of β_2 depend upon r. We would argue that working with β_2 is unnecessary and poses unnecessary complications for obtaining inference on the cointegrating space.

In cointegration analysis it is not the values of the elements of β that are the object of interest, rather the space spanned by β, $\mathfrak{p} = sp(\beta)$, and this space is in fact all we are able to uniquely estimate. The parameter $\mathfrak{p}$ is an r-plane in n-space and as such an element of the Grassman manifold $G_{r,n-r}$. Before we derive the priors for $\mathfrak{p}$ we briefly comment on the relationship between priors on β_2 and on $\mathfrak{p}$.

The Jacobian for the transformation from $\mathfrak{p} \in G_{r,n-r}$ to $\beta_2 \in R^{(n-r)r}$ is presented in Villani (2001) as $|I_r + \beta_2'\beta_2|^{-n/2}$. From this Jacobian we can clearly see that a flat prior on $\mathfrak{p}$ is informative with respect to β_2 and vice versa. This result reflects that found by Phillips (1994) in classical analysis, that the finite sample distribution of the maximum likelihood estimator with linear restrictions imposed has

Cauchy tails and that this Cauchy behaviour is a direct result of imposing the linear restrictions.

Next, consider the implications of a flat prior on β_2 for the prior on $\mathfrak{p}$. A common justification for the linear restrictions is that an economist will usually have some idea about which variables will enter the cointegrating relations and so she chooses c to select the rows of coefficients most likely to be non-zero—more generally linearly independent from each other—and then normalises on these coefficients. By using these linear restrictions, however, the Jacobian for $\beta_2 \to \mathfrak{p}$ places infinitely more weight in the direction where the coefficients thought most likely to be different from zero are, in fact, zero (or linearly dependent).

To demonstrate this claim, consider a n-dimensional system for $y=(x',z')'$ where x is a r vector. It is believed that if a cointegrating relationship exists then it will most likely involve the elements of x in linearly independent relations. That is, in $y\beta = x\beta_1 + z\beta_2 \backsim I(0)$, $\det(\beta_1)$ is believed far from zero and so choose $c=[I_r\ 0]$ and estimate $\beta_2 = c_\perp \beta(c\beta)^{-1}$. If $\mathfrak{p} = sp(\beta)$, $\beta \in V_{r,n}$, the Jacobian for the transformation $\beta_2 \to \mathfrak{p}$ is proportional to $J(\beta_2 : \beta) = |I_r + (c\beta)^{'-1}\beta' c_\perp' c_\perp \beta (c\beta)^{-1}|^{n/2}$. As vectors in β approach the null space of c, then $(c\beta)^{-1} \to \infty$, and thus $J(\beta_2 : \beta) \to \infty$. As a result the prior will more heavily weight regions where $\det(c\beta) = \det(\beta_1) \approx 0$, contrary to the intention of the researcher.

In summary, inference using β_2 relies on the investigator having some prior knowledge of the cointegrating space (c) and implementing this information has strange, even contrary, implications for the expression of prior beliefs about the cointegrating space. We know in many cases the posterior for β_2 is potentially bimodal, has no moments (except with highly informative priors such as in, possibly, Geweke, 1996) and is proper, but it is not known if these properties are general and we know of cases where the posterior is improper (see KVD). Further, combining cointegration with other model restrictions can produce an improper posterior for β_2.

Clearly then, there is reason to consider another approach to eliciting priors for β. We present one approach in the following subsections.

2.2.3. A uniform prior on the cointegrating space

The previous subsection presents arguments for why prior beliefs about the cointegrating relations should be expressed in the prior distribution for the cointegrating space. Villani (2001) begins with a uniform prior upon the cointegrating space from which he derives the prior density for β_2. However his interest was in estimation and tests related to $sp(\beta)$ and as such, the analysis was conditional upon r. Although we begin with the same uniform prior for the cointegrating space as Villani, as indicated in the discussion above we use a different parameterisation of the cointegrating vectors than those implied by linear identifying restrictions.

As we have claimed the cointegrating space to be the parameter of interest, we propose working directly with $\mathfrak{p} = sp(\beta)$. To derive a prior density for $\mathfrak{p}$, we need a measure on its support, $G_{r,n-r}$. Using the results of James (1954), a distribution and identifying restrictions for β are derived from the uniform distribution for $\mathfrak{p}$ over $G_{r,n-r}$. First we must introduce some notation on measures.

As discussed in James (1954) (see also Muirhead 1982, Chapter 2), the invariant measures on the orthogonal group, the Stiefel manifold and the Grassman manifold are defined in exterior product differential forms. For brevity we denote these measures as follows. For the $(n \times n)$ orthogonal matrix $[b_1, b_2, \ldots, b_n] \in O(n)$ such that $\beta = [b_1, b_2, \ldots, b_r] \in V_{r,n}$, $r < n$, the measure on the orthogonal group $O(n)$ is denoted $\mathrm{d}v_n^n \equiv \Lambda_{i=1}^n \Lambda_{j=i+1}^n b_j' \,\mathrm{d}b_i$, the measure on the Stiefel manifold $V_{r,n}$ is denoted $\mathrm{d}v_r^n \equiv \Lambda_{i=1}^r \Lambda_{j=i+1}^n b_j' \,\mathrm{d}b_i$, and the measure on the Grassman manifold $G_{r,n-r}$ is denoted $\mathrm{d}g_r^n \equiv \Lambda_{i=1}^r \Lambda_{j=n-r+1}^n b_j' \,\mathrm{d}b_i$. These measures are invariant (to left and right orthogonal translations).

The space $G_{r,n-r}$ is compact. The flat, proper prior density on $G_{r,n-r}$ has the form $p(\beta) = 1/c_r^n$ where $c_r^n = \int_{G_{r,n-r}} \mathrm{d}g_r^n$ is the volume $G_{r,n-r}$ and $\beta \in V_{r,n}$. This expression for the prior defines a probability measure on the space of β. As the posterior distribution and therefore the posterior probabilities for r will depend upon c_r^n, which itself is a function of r, we need an expression.[1] The following discussion is based upon James (1954).

Any matrix element of $V_{r,n}$ spans one element of $G_{r,n-r}$, however each element of $G_{r,n-r}$ is spanned by many elements of $V_{r,n}$. That is, the matrix β spans the space $\mathfrak{p} = sp(\beta)$. If we post-multiply β by $C \in O(r)$ to obtain the resulting matrix $A = \beta C$, where $A \in V_{r,n}$, since the operation βC simply rotates the vectors of β within the plane $\mathfrak{p}$, the matrix A will span the same space as β, that is $\mathfrak{p} = sp(\beta) = sp(A)$. Thus there is a many to one relationship between $V_{r,n}$ and $G_{r,n-r}$. The dimension of $V_{r,n}$ is $nr - r(r+1)/2$ which exceeds the dimension of $G_{r,n-r}$, $nr - r^2$, by $r(r-1)/2$. To derive c_r^n, fix the orientation of β in the plane $\mathfrak{p}$ such that although $\beta \in V_{r,n}$, as $\mathfrak{p}$ varies over all of $G_{r,n-r}$, β is not free to vary over all of $V_{r,n}$. This defines β to have only $nr - r^2$ free elements and there is a one to one relationship between β and $\mathfrak{p}$. Postmultiplying β by $C \in O(r)$ produces $A = \beta C$ where $A \in V_{r,n}$ and as $\mathfrak{p}$ varies over all of $G_{r,n-r}$, C is free to vary over all of $O(r)$ and A over all of $V_{r,n}$. Taking differentials of A, β and C gives the measures $\mathrm{d}v_r^n$, $\mathrm{d}g_r^n$ and $\mathrm{d}v_r^r$, respectively. These measures are related by $\mathrm{d}v_r^n = \mathrm{d}g_r^n \times \mathrm{d}v_r^r$. As $c_r^n = \int_{G_{r,n-r}} \mathrm{d}g_r^n$, we integrate to obtain $\int_{V_{r,n}} \mathrm{d}v_r^n = \int_{G_{r,n-r}} \mathrm{d}g_r^n \times \int_{O(r)} \mathrm{d}v_r^r$ and so $c_r^n = \int_{G_{r,n-r}} \mathrm{d}g_r^n = \int_{V_{r,n}} \mathrm{d}v_r^n / \int_{O(r)} \mathrm{d}v_r^r$. Thus c_r^n is the ratio of the volume of a Stiefel manifold, $V_{r,n}$, to the volume of an orthogonal group, $O(r)$, (James, 1954)[2] giving

$$c_r^n = \pi^{-(n-r)r} \prod_{j=1}^{r} \frac{\Gamma[(n+1-j)/2]}{\Gamma[(r+1-j)/2]}, \tag{3}$$

where $\Gamma[q] = \int_0^\infty u^{q-1} \mathrm{e}^{-u} \,\mathrm{d}u$ $q > 0$.

Although the derivation of c_r^n requires β to have a fixed orientation in $\mathfrak{p}$, we can ignore this restriction for obtaining integrals with respect to β (and therefore $\mathfrak{p}$) as the prior, the posterior and the measure $\mathrm{d}g_r^n$ are all invariant to translations of the form $\beta \to \beta C$, $C \in O(r)$. Therefore it is possible to work directly with β as an element of

[1] The authors are grateful to an anonymous referee for pointing out the importance of this dependence in development of the posterior.

[2] We note that the sums, $\sum$, in (5.23) of James (1954) should be products, $\prod$.

the Stiefel manifold and adjust the integrals with respect to β by $Vol(O(r))=\int_{O(r)} \mathrm{d}v_r^r$. To make this process explicit, if we have a function $f(\beta)=f(\beta C)$ where $C \in O(r)$, then the integral of $f(\beta)$ with respect to $G_{r,n-r}$ is obtained by integrating over $V_{r,n}$ and adjusting this integral by $Vol(O(r))$. Thus we have

$$\int_{G_{r,n-r}} f(\beta)\,\mathrm{d}g_r^n = \frac{\int_{V_{r,n}} f(A)\,\mathrm{d}v_r^n}{\int_{O(r)} \mathrm{d}v_r^r}. \tag{4}$$

Note that the identifying restrictions $\beta'\beta=I_r$ do not distort the weight on the space of the parameter of interest, $\mathfrak{p}$.

To demonstrate the simplicity of the implementation of this approach we explain how to obtain a draw from the prior. For general $r<n$, a draw from the uniform prior is obtained by drawing an $(n\times r)$ matrix Z from $vec(Z)\sim \mathrm{N}(0,I_{nr})$. Next we decompose Z into $\beta\in V_{r,n}$ and a square positive definite matrix T. This may be achieved in a number of ways one of which is to take $\beta=Z(Z'Z)^{-1/2}$. We now have in β a draw from the uniform prior on $V_{r,n}$ and the space $\mathfrak{p}=sp(\beta)$ is uniformly distributed on $G_{r,n-r}$.

2.2.4. *An informative prior on the cointegrating space*

Although the uniform prior (3) is used in this paper, it is common to employ informative priors for parameters and so one is specified here for $\mathfrak{p}$.

If a researcher believes a parameter is likely to have a particular value, to incorporate this prior belief she places more prior mass around this likely point. For the parameter $\mathfrak{p}$, denote the likely value as $\mathfrak{p}^H=sp(H\kappa)$ where $H\in V_{s,n}$ is a known $n\times s$ $(s\geqslant r)$ matrix, $H_\perp\in V_{n-s,n}$ its orthogonal complement and κ is an $s\times r$ full rank r matrix. To obtain H, specify the general matrix H^g with the desired coefficient values, then map this to $V_{r,n}$ by the transformation $H=H^g(H^{g'}H^g)^{-1/2}$.

As an example, consider the system with $n=2$, $r=1$ and the preferred space spanned by the vector $h^g=(1,1)$. This space is also spanned by the vector $h=h^g(h^{g'}h^g)^{-1/2}=(2^{-1/2},2^{-1/2})'$ which has the orthogonal complement $h_o=(-2^{-1/2},2^{-1/2})'$. Both the space of h, which we will denote by $\mathfrak{p}^h=sp(h)$, and the vector h are presented in Fig. 3. The space orthogonal to $\mathfrak{p}^h$, $\mathfrak{p}^{h_\perp}=sp(h_o)$, and the vector h_o, are also presented in Fig. 3. Fig. 3 gives a clear indication of what is meant by the vector h lying in, or spanning, the space $\mathfrak{p}^h=sp(h)$. The vectors h and h_o have unit length and as such $h,h_o\in V_{1,2}$. Our interest in this example is to place a prior on $\mathfrak{p}^h$.

A dogmatic prior for $\mathfrak{p}$ could be obtained by letting $\beta=H\kappa V$, $V\in O(r)$. Define $\kappa V=V_\kappa\in V_{r,s}$ and specify the flat prior density in (3) for V_κ. This prior assigns probability one to the point $\mathfrak{p}=\mathfrak{p}^H$. For the example in Fig. 3, we would choose $\beta=h$ such that the vector could not span any other space than $\mathfrak{p}^h$.

Often, however, the researcher will want to employ a less dogmatic prior. In our example, we may wish the space to have a mean under the prior of $\mathfrak{p}^h$, but be allowed to vary over the entire one-dimensional Grassman manifold. Consider a vector $Z=(z_1,z_2)$ distributed as $Z\sim \mathrm{N}(0,I_2)$. The space of Z is uniformly distributed over $G_{1,1}$. An example of such a vector drawn from the bivariate standard normal is $z=(0.5,1.5)$

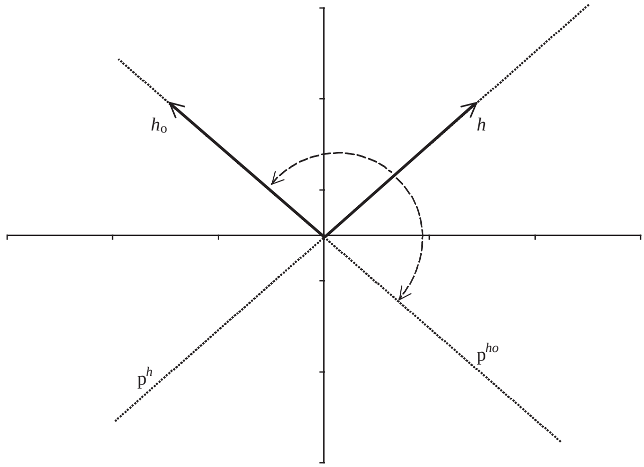

Fig. 3. The spaces $\mathfrak{p}^h$ and $\mathfrak{p}^{ho}$.

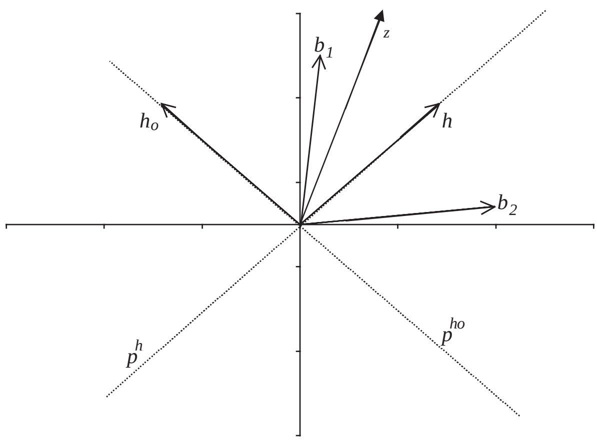

Fig. 4. Antithetic draws of b_1 and b_2 from the informative prior.

which is presented in Fig. 4.[3] We can project z into $\mathfrak{p}^h$ by the operation $h_z = hh'z$ and into $\mathfrak{p}^{ho}$ by the operation $h_{oz} = h_o h_o' z$. If we take a weighted sum of these two vectors $x = h_z w_1 + h_{zo} w_2$ where the weights are random with supports $w_1 \in [0, 1]$ and $w_2 \in [-1, 1]$ and are *not* constrained to sum to one, then x will vary over the half-hemisphere described by the arc in Fig. 3. Next, take $b = x(x'x)^{-1/2}$ as a draw from our prior and the random space $\mathfrak{p} = sp(b)$ can take any direction, but, if $E(w_2) = 0$, will have location $\mathfrak{p}^h$ and dispersion determined by the distribution on w_1 and w_2.

[3] Unit vectors in Fig. 4 have open arrow heads $>$, while general length vectors have closed heads, ▶.

A possible choice of a distribution for $\eta = w_2 + 1$ with $w_1 = (1 - w_2^2)^{1/2}$, is Beta over $\eta \in [0, 2]$. What is important is that the relative weights on the spaces $\tau = w_2/w_1$ vary over $(-\infty, \infty)$ such that all of the Grassman manifold is accessible. Thus we may alternatively take $x = h_z + h_{zo}\tau$ and specify the distribution directly on τ, so long as $E(\tau) = 0$ to centre the distribution of $\mathfrak{p}$ on $\mathfrak{p}^h$. In Fig. 4 we present examples of b that result from antithetic draws of w_2. For a draw of $w_2 = 0.85$, such that $w_1 = (1 - w_2)^{1/2} = 0.527$ $(\tau = 1.61)$ we obtain the vector marked as b_1. The vector b_2 results from $w_2 = -0.85$ $(\tau = -1.61)$. A possible general specification for this prior follows.

Let the random scalar τ have $E(\tau)=0$ and $E(\tau^2)=\sigma^2$. The value of σ will control the tightness of the prior around $\mathfrak{p}^H$. Next construct $P_\tau = HH' + H_\perp H_\perp' \tau$ and let the $n \times r$ matrix Z be distributed as $vec(Z) \sim \mathrm{N}(0, I_{nr})$. The matrix $X = P_\tau Z$ can be decomposed as $X = \beta\kappa$ where $\beta \in V_{r,n}$ and κ is an $r \times r$ lower triangular matrix. For $\tau \neq 0$ and $|\tau| < \infty$, the space of β, $\mathfrak{p} = sp(\beta)$, is a direct weighted sum of the spaces $\mathfrak{p}^H$ and $\mathfrak{p}^{H_\perp}$ with the weight determined by τ. At $\tau = 0$ and $\tau = \pm\infty$, $\mathfrak{p}$ is respectively $\mathfrak{p}^H$ and $\mathfrak{p}^{H_\perp}$. It is for this reason that we chose $E(\tau) = 0$ such that with respect to τ, the space will on average be $\mathfrak{p}^H$.

One choice for τ is $\mathrm{N}(0, \sigma^2)$. To obtain the joint prior for (β, τ), we form the joint distribution for (Z, τ), transform Z by $Z = P_\tau^{-1} X$ to obtain the joint distribution for (X, τ). Next, we transform X by $X = \beta\kappa$ and integrate with respect to $K = \kappa\kappa'$ which is distributed as Wishart. The form of the resultant prior for β and the hyperparameter τ is then

$$p(\tau, \beta) = \tau^{-(n-r)r} \exp\left\{-\frac{\tau^2}{2\sigma^2}\right\} |\beta' P_{\tau^2}^{-1} \beta|^{-n/2} \mathfrak{m}_r, \tag{5}$$

where $\mathfrak{m}_r = 2^{-1/2}\pi^{-(n-r)r/2-1/2}\sigma^{-1}\prod_{j=1}^{r} \Gamma[(n+1-j)/2]/\Gamma[(r+1-j)/2]$. Note that the normalising constant $\mathfrak{m}_r$ incorporates the adjustment in (4). The conditional density for $\beta|\tau$ is an example of the matrix angular central Gaussian distribution on $G_{r,n-r}$ first introduced by Chikuse (1990).

Prior (5) treats the area around $\mathfrak{p}^{H_\perp}$, which occurs at $\tau = \infty$, as an unlikely (measure zero) event regardless of the choice of σ. This is desirable since at $\mathfrak{p}^{H_\perp}$ the dimension of the cointegrating space, $\dim(\mathfrak{p})$, would become $\min(p - r, r)$ rather than r.

Obtaining a draw from (5) is relatively straightforward. First draw τ from $\mathrm{N}(0, \sigma^2)$ and then draw $Z = \{z_{ij}\}$ with $z_{ij} \sim \mathrm{N}(0, 1)$. Construct $X = P_\tau Z$ and decompose X into β and κ (we then discard κ). The resulting draw of (τ, β) will be a draw from (5).

A final remark on the choice of σ is required. The value of σ could be calibrated to a preferred level of dispersion using the span variation measure (sv) of Villani (2000), and (MC)MC draws. Villani (2000) shows how sv can be used to express the degree of variation in a distribution as a proportion of the variation under the uniform distribution—the uniform providing equal variation in every direction.

2.3. The posterior

We assume the rows of $\varepsilon = (\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_T')'$ are $\varepsilon_t \sim \text{iid}\, \mathrm{N}(0, \Sigma)$. The likelihood can then be written as $L(y|\Sigma, B, \beta, r, \tilde{Z}) \propto |\Sigma|^{-T/2} \exp\{-\frac{1}{2} tr(\Sigma^{-1}\varepsilon'\varepsilon)\}$. Using the priors specified above, the general form of the posterior is then

$$p(B, \Sigma, \beta, r|y) \propto |\Sigma|^{-(T+n+r+k+1)/2}(2\pi)^{-n(k+r)/2}\upsilon^{n(k+r)/2}p(\beta) \times \exp\left\{-\frac{1}{2} tr\Sigma^{-1}[TS + (B-\tilde{B})'V(B-\tilde{B})]\right\}, \tag{6}$$

where $S = S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}$, $\tilde{B} = [\tilde{\alpha}' \ \ \tilde{\Phi}']'$, $\tilde{\alpha} = (\beta' S_{11}\beta)^{-1}\beta' S_{10}$, $\tilde{\Phi} = S_{22}^{-1}S_{20}$, and $V = \beta'(\sum_{t=1}^{T} z_t' z_t + \underline{H})\beta$, where $z_t = (z_{1,t} \ z_{2,t})$. The values for the S_{ij} are defined as $vM_{ij} = h_{ij} + \sum_{t=1}^{T} z_{i,t}' z_{j,t}$ for i and $j = 1, 2$, $h_{ij} = 0$ if $i \neq j$ and $h_{ii} = \upsilon I$, $vM_{20} = \sum_{t=1}^{T} z_{2,t}' z_{0,t}$, $vM_{10} = \sum_{t=1}^{T} z_{1,t}' z_{0,t}$, $vM_{00} = \sum_{t=1}^{T} z_{0,t}' z_{0,t}$ and so $S_{ij} = M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$ for $ij = 0, 1, 2$, except $i = j = 2$ where $S_{22} = M_{22} - M_{21}M_{11}^{-1}M_{12}$ and $S_{20} = M_{20} - M_{21}M_{11}^{-1}M_{10}$. For later use we also define $D_0 = D_1 - D_2$, $D_1 = S_{11}$ and $D_2 = S_{10}S_{00}^{-1}S_{01}$.

3. Bayes factors and posterior probabilities

In this section the tool for Bayesian inference in this paper—the posterior probabilities of the ranks—is introduced. Our objective is to report estimates of the posterior probabilities of the rank r, $p(r|y)$. Let B_{r,r^*} be the Bayes factor for the model with the rank r to the model with rank r^*. These models have parameters θ_r and θ_{r^*}, respectively. The terms $p(r|y)$, $p(r^*|y)$ and B_{r,r^*} are linked through the expression for the posterior odds ratio, $p(r|y)/p(r^*|y) = p(r)/p(r^*) \times B_{r,r^*}$, where $p(r)$ is the prior probability of the model r, $B_{r,r^*} = m(y|r)/m(y|r^*)$ and

$$m(y|r) = \int L(\theta_r)p(\theta_r|r)\,\mathrm{d}\theta_r \tag{7}$$

is the marginal likelihood for the model r. As the prior odds are known, we need only estimate $m(y|r)$.

To perform the integration in (7) of $\theta_r = (\Sigma, B, \beta)$, first analytically integrate (6) with respect to (Σ, B). From the expression in (6), it is straightforward to show that the posterior for (Σ, B) conditional on (β, r) has a standard form which may be integrated analytically (see for example Zellner, 1971). The resulting posterior for the remaining parameters is

$$p(\beta, r|y) \propto \mathfrak{g}_r |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2}, \tag{8}$$

where in this case

$$\mathfrak{g}_r = T^{-nr/2}\pi^{-(n-r)r/2}\upsilon^{nr/2}\prod_{j=1}^{r}\frac{\Gamma[(n+1-j)/2]}{\Gamma[(r+1-j)/2]}. \tag{9}$$

Again the normalising constant $\mathfrak{g}_r$ incorporates the adjustment in (4).

The object of interest is the cointegrating space, however the tool for obtaining inference on this space is the cointegrating vectors. As all inference on the cointegrating space is derived from the cointegrating vectors, it is important that the posterior for these vectors is well behaved and that we know as much as possible about this distribution. Both these properties of the posterior for β and the generality of what can be said of this posterior contrast with what is known about the posterior when linear identifying restrictions are used. We present some of these properties here.

The conditional density for β given r has the form $p(\beta|r,y) \propto k(\beta)$ where $k(\beta) = |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2}$. $k(\beta)$ is invariant to $\beta \to \beta C$ for $C \in O(r)$. The eigenvalues $\lambda_j(D_l)$ for $l = 0, 1$, are positive and finite with probability one. By the Poincaré separation theorem, since $\beta \in V_{r,n}$, then $\prod_{j=1}^{r} \lambda_{n-r+j}(D_l) \leqslant |\beta' D_l \beta| \leqslant \prod_{j=1}^{r} \lambda_j(D_l)$ and so $k(\beta)$ is bounded above (and below) by some positive finite constant. Thus $k(\beta)$ has a finite upper bound, M. As the elements of β, b_{ij}, have compact support, the integral $\int_{V_{r,n}} b_{ij}^m k(\beta)\, \mathrm{d}g_r^n$ for $m = 0, 1, \ldots$ is bounded above almost everywhere by the integral $M \int_{-1}^{1} b_{ij}^m \, \mathrm{d}b_{ij}$. These conditions are sufficient to ensure the posterior for β is proper and all finite moments exist (see Billingsley, 1979, pp. 174, 180).

To obtain the posterior distribution of r, $p(r|y)$, it is necessary to integrate (8) with respect to β and so obtain an expression for

$$p(r|y) = \int p(\beta, r|y)\, \mathrm{d}g. \tag{10}$$

The marginal density of β conditional on r has the same form in all cases as

$$p(\beta|r,y) = \mathfrak{c}_r^{-1} k(\beta) \tag{11}$$

which is not of standard form. Although one may exist, we do not currently know of a simple, general analytical solution for $\mathfrak{c}_r = \int_{G_{r,n-r}} k(\beta)\, \mathrm{d}g_r^n = \int_{V_{r,n-r}} k(A)\, \mathrm{d}v_r^n / \int_{O(r)} \mathrm{d}v_r^r$, and so we estimate $\mathfrak{c}_r$.

Two possible approaches to estimating $\mathfrak{c}_r$ are either to use Markov chain Monte Carlo (MCMC) methods or numerical integration. KP and Bauwens and Giot (1998) demonstrate how to evaluate similar integrals using MCMC when β has been identified using linear restrictions rather than those used in this paper. Strachan (2003) demonstrates the MCMC approach when β has been identified using restrictions related to those in this paper, however the posterior has a very different form as an embedding approach similar to KP is used. An alternative approach commonly used in classical work to approximate integrals over $V_{r,n}$, is to use the Laplace approximation which is computationally much faster. In the following section the Laplace approximation to $\mathfrak{c}_r$ is briefly described.

4. Laplace approximation

Let $m = (r/2)(2n - r - 1)$, $f = |\beta' D_0 \beta|^{-1/2} |\beta' D_1 \beta|^{1/2}$ and $g = |\beta' D_1 \beta|^{-n/2}$. Let $\bar{\beta}$ denote the value of β at the mode of f and the Hessian matrix for $-\ln f$ evaluated at $\bar{\beta}$ be $\Psi = \Psi_r$. By the Laplace method the integral $\int g f^T \, \mathrm{d}\beta$ in (10) can be approximated by $g(\bar{\beta})f^T(\bar{\beta})(2\pi/T)^{m/2}|\Psi|^{-1/2}g_r$ (see the appendix). There are a number of papers on applications of the Laplace approximation in Bayesian econometrics (see for example Lindley, 1980; Tierney and Kadane, 1986; Tierney et al., 1989; Kass and Raftery, 1995). However for more relevant references for our application to an integral over the Stiefel manifold the reader is directed to Muirhead (1982, Chapter 9), Anderson (1965) and James (1969). In these applications the aim was to derive distributions of latent roots of covariance matrices and the Laplace approach was used to provide asymptotic representations of hypergeometric functions of zonal polynomials which can be represented as integrals over the orthogonal group or, in some cases, the Stiefel manifold.

The Laplace method works well if the mode and Hessian, Ψ, are easy to obtain and if the posterior is reasonably peaked around the mode. Expressions for the mode and Hessian are presented in the appendix. The posterior tends to be peaked for reasonable sample sizes and the mode dominates as T increases. On the general question of approximating the marginal likelihood by the Laplace approximation, there is considerable precedent in the literature for using this method for this purpose (see Kass and Vaidyanathan, 1992; Raftery, 1994; Lewis and Raftery, 1997 in addition to previous references). Further, the results presented in this paper support the application of Laplace approximation for estimation of r.

The classical maximum eigenvalue test statistic, which is a likelihood ratio test statistic for the hypothesis $\mathrm{H}_0: rank(\Pi)=r$ versus $\mathrm{H}_1: rank(\Pi)=r+1$, has the form $m_{r,r+1}=-T\ln(1-\hat{\lambda}_{r+1})$. From this expression it is possible to show the link between the classical test statistic, $m_{r,r+1}$, and the log Bayes factor, $\ln B_{r,r+1}$, for the diffuse prior as $\ln\hat{B}_{r,r+1}=k_0+k_1 m_{r,r+1}$ where the k_i depend on the data, r, and n. Denote the classical trace test statistic, which is a likelihood ratio test statistic for the hypothesis $\mathrm{H}_0: rank(\Pi)=r$ versus $\mathrm{H}_1: rank(\Pi)=n$, as $m_{r,n}$. Similarly, it is possible to present the link between $m_{r,n}$ and the log Bayes factor, $\ln B_{r,n}$, for the diffuse prior as $\ln\hat{B}_{r,n}=k_2+k_3 m_{r,n}$.

5. Monte Carlo experiment

To investigate the small sample performance of our estimator for $p(r|y)$, we conduct Monte Carlo experiments and compare these results to those for a range of information criteria and the classical trace test. In applied analysis, economic time series are commonly modelled as combinations of both stochastic and deterministic trends. Deterministic processes have implications for (both theoretical and) applied analysis of time series in that the presence (or absence) of various deterministic processes affects inference about the number of stochastic trends. Therefore we consider estimation of $p(r|y)$ for a range of deterministic processes. The issue of selection of deterministic processes is the subject of ongoing research.

The general DGP for the experiments is a VAR with 2 lags and deterministic processes $\mu_{jt}=\mu_j+\delta_j t$ for $j=0,1,2$. Let β_2 be a $(n-r)\times r$ matrix, $w_{1,t}$ be a $1\times r$ random vector and $w_{2,t}$ be a $1\times(n-r)$ random vector is generated by $w_{j,t}=\mu_{jt}+w_{j,t-1}\rho_j+\varepsilon_{j,t}$, $j=1,2$ where $\varepsilon_{j,t}\backsim$ iid $\mathrm{N}(0,\sigma^2)$ and ρ_j is an identity matrix times ρ. The $1\times n$ vector of variables in the system is $y_t = (y_{1,t}\ \ y_{2,t})$ where $y_{1,t}$ is a $1 \times r$ random vector and $y_{2,t}$ is a $1 \times (n-r)$ random vector jointly generated by $y_{1,t} = \mu_{0t} + y_{2,t}\beta_2 + w_{1,t}$, $y_{2,t} = y_{2,t-1} + w_{2,t}$.

This specification corresponds to the ECM in (1) as $\Delta y_t = \mu + \delta t + y_{t-1}\beta\alpha + \Delta y_{t-1}\Gamma_1 + \varepsilon_t$ with $\delta = [\delta_1 + \delta_2\beta_2 - \delta_0\alpha_1,\ \delta_2]$, $\mu = [\mu_1 + \mu_2\beta_2 - \delta_0 - \mu_0\alpha_1,\ \mu_2]$, $\alpha = [\alpha_1\ 0] = [(\rho-1)I_r\ 0]$, $\beta = [I_r\ -\beta_2']'$, $\Gamma_1 = [\Gamma_{11}',\ \Gamma_{21}']$, $\Gamma_{11} = 0$, $\Gamma_{21} = [\rho\beta_2\ I_{n-r}\rho]$, and $\varepsilon_t = (\varepsilon_{1,t} + \varepsilon_{2,t}\beta_2,\ \varepsilon_{2,t})$.

The range of deterministic processes is indexed by i and can be represented as non-zero means and trends in $\mathrm{E}(y_t\beta)$ or $\mathrm{E}(\Delta y_{j,t}), j=1,2$. These processes are: for $i=1$, $\mathrm{E}(\Delta y_{j,t}) = \mu_j + \delta_j t$ and $\mathrm{E}(y_t\beta) = \mu_0 + \delta_0 t$; for $i=2$, $\mathrm{E}(\Delta y_{j,t}) = \mu_j$ and $\mathrm{E}(y_t\beta) = \mu_0 + \delta_0 t$; for $i=3$, $\mathrm{E}(\Delta y_{j,t}) = \mu_j$ and $\mathrm{E}(y_t\beta) = \mu_0$; for $i=4$, $\mathrm{E}(\Delta y_{j,t}) = 0$ and $\mathrm{E}(y_t\beta) = \mu_0$; and for $i=1$, $\mathrm{E}(\Delta y_{j,t}) = \mathrm{E}(y_t\beta) = 0$.

The values of μ_j and δ_j are set equal to $0.35\imath$ in which $\imath$ is a vector of ones or zeros depending upon whether this term is included, $\rho = 0.35$, $\sigma = 1.5$, $T = 100$, and each element of β_2 is 1. All of the following results come from 10,000 draws of y_t for each model such that the following probabilities and relative frequencies have Monte Carlo standard errors of at most 0.005.

The range of models simulated is for each $i = 1,\dots,5$, $n = 2,3,4$ and $r = 0,1,2$ with prior variance term $\upsilon = 0.01, 0.1, 1$ for a total of 135 experiments. For each model the value of r is selected using the highest estimated posterior probability for the Laplace method (LP) and three commonly employed information criteria: the Akaike (1974) (AIC); Schwarz (1978) (BIC); and the Hannan and Quinn (1979) (HQ). The estimator's performance in selecting r by using the mode of $p(r|y)$ is also compared to that of $m_{r,n}$ at the 5% ($m_{r,n}^{5\%}$) level of significance.

The marginal relative selection frequencies of the correct r are compared for the six techniques. Full results are available from the authors and a selection is reported here. Again it should be noted at this point that the reporting of selection frequencies aids only in comparison with other techniques for the purpose of selecting an r on which subsequent inference can condition.

In these experiments, LP was first or equal first in 63 of the 135 experiments. For the other techniques the same result was: AIC 15; BIC 54; HQ 29; and $m_{r,n}^{5\%}$ 27. Different values of υ affect the relative performance of LP. The effect of this value can been seen as with $\upsilon = 0.01$, where LP was first or equal first in 15 of 45 experiments, with $\upsilon = 0.1$, we obtain a count of 23 of 45 experiments, and with $\upsilon = 1$, we obtain 25 of 45 experiments.

For the second experiment, we vary the largest root over the stationary region. This is intended to give an indication of the ability of the test to select the correct rank, r, when the data displays behaviour suggestive of the rank being $r-1$. This is analogous to a stationary AR(1) process with a root that approaches one in absolute size. As the root approaches one, the data will appear to have been generated by a unit root process.

For the process with $n = 2$, $\upsilon = 1$, $i = 4$ and $r = 1$ we take ρ_1 and vary this over the range $\rho_1 \in (0.15, 0.35, 0.55, 0.75, 0.95, 0.99)$. A value of $\rho_1 = 1$ implies a rank $r = 0$. Figs. 5 and 6 show the relative selection frequencies for each technique for the ranks $r = 1$ and 0, respectively. As expected, LPs performance deteriorates as ρ_1

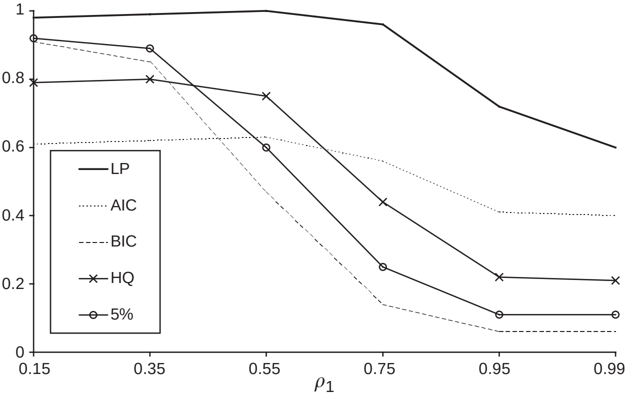

Fig. 5. Plot of relative selection frequencies for $r = 1$.

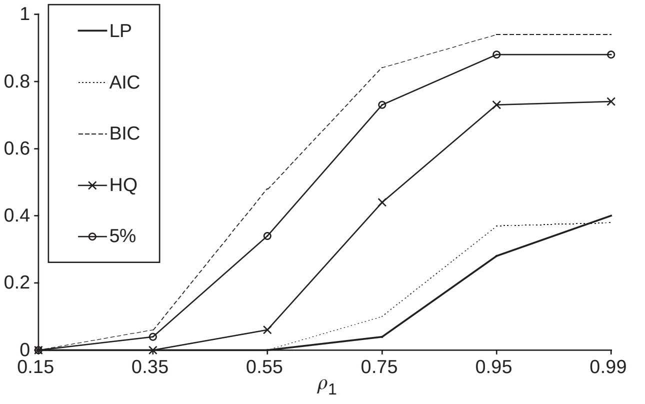

Fig. 6. Plot of relative selection frequencies for $r = 0$.

approaches unity, however, we see LP correctly selects $r = 1$ more often than the other techniques.

6. Conclusion

In this paper a method of finding approximations to Bayes factors has been demonstrated for models of stochastic processes of a cointegrating error correction model. These approximations use both analytical integration and the Laplace method of approximating integrals. Although the Laplace method has been employed in many Bayesian studies, the approach in this article owes more to the classical literature on obtaining distributions of latent roots of covariance matrices. The Monte Carlo results suggest the Laplace approach performs well at selecting the number of stochastic trends when compared with the equivalent classical test statistics and information criteria.

An important contribution of this article is the approach to eliciting priors for cointegrating vectors. As the object of interest is the cointegrating space, a prior is placed upon this parameter and this defines the implied prior for the elements in the cointegrating vectors.

Acknowledgements

The authors would like to thank Charles Bos, Herman van Dijk, Soren Johansen, Frank Kleibergen, Brendan McCabe, Don Poskitt and participants of the workshop 'Recent Advances in Bayesian Econometrics' in Marseilles, June 2001 and two anonymous referees for useful discussions that have improved this paper. We would particularly like to thank Luc Bauwens for many helpful comments.

Appendix. The Laplace approximation

Before applying the Laplace approximation to the integral in (10), define by $U = [U_1 \;\; U_2] \in O(n)$ the eigenvectors of $A = D_1^{-1/2} D_2 D_1^{-1/2}$ such that $A = U \Lambda U'$ and $\Lambda = diag(\lambda_1(A), \ldots, \lambda_p(A))$. Next let $D_l^u = U' D_l U$ for $l = 0, 1, 2$, $\beta = [b_1, b_2, \ldots, b_r]$, $H_1 = [h_1, h_2, \ldots, h_r] = U'\beta$ and since $U \in O(n)$ then $\mathrm{d}v_r^n = \Lambda_{i=1}^r \Lambda_{j=i+1}^n b_j' \,\mathrm{d}b_i = \Lambda_{i=1}^r \Lambda_{j=i+1}^n h_j' \,\mathrm{d}h_i$ by invariance of $\mathrm{d}v_r^n$. Therefore

$$\int_{V_{r,n}} f(\beta) g(\beta) \,\mathrm{d}v_r^n = \int_{V_{r,n}} f(UH_1) g(UH_1) \,\mathrm{d}v_r^n,$$

where $f(UH_1) = |H_1' D_0^u H_1|^{-1/2} |H_1' D_1^u H_1|^{1/2}$. The Laplace approximation is then applied to this integral with respect to H_1. This application requires the mode of f, $\overline{H}_1$, and an expression for the Hessian of $-\ln f$ at $\overline{H}_1$, Ψ_r.

Maximising $f > 0$ is equivalent to minimising $f^{-2} = |H_1' D_0^u H_1 (H_1' D_1^u H_1)^{-1}|$. Note for a $m \times m$ matrix E, $|E| = \prod_{j=1}^m \lambda_i(E) = \prod_{j=1}^m \lambda_i(U'EU)$. From an extension of the Poincaré separation theorem (see Schott, 1997, p. 116)

$$\begin{aligned} \min f^{-2} &= \prod_{j=1}^{r} \lambda_{n-r+j}(D_0^u D_1^{u-1}) \\ &= \prod_{j=1}^{r} \lambda_{n-r+j}(D_0 D_1^{-1}) \quad \text{since } D_0^u D_1^{u-1} = U' D_0 D_1^{-1} U. \end{aligned}$$

Since $D_1^{-1/2} D_0 D_1^{-1/2} = I_n - A$, and $\lambda_{n-r+j}(D_0 D_1^{-1}) = \lambda_{n-r+j}(D_1^{-1/2} D_0 D_1^{-1/2})$, then this equals $1 - \lambda_j(A) = 1 - \lambda_j(\Lambda)$. Therefore, $\min f^{-2} = \prod_{j=1}^r (1 - \lambda_j(\Lambda)) = \min |I_r - H_1 \Lambda H_1|$ which occurs at $H_1 = [\pm I_r \;\; 0]'$ where $\pm I_r$ means one of the 2^r matrices with zero off-diagonal elements and diagonal elements either $+1$ or -1. Therefore if $\overline{H}_1 = [I_r \;\; 0]'$ for large T, $\mathfrak{c} \approx 2^r \int_{\mathcal{N}(\overline{H}_1)} k(UH_1) \,\mathrm{d}v_r^n$ where $\mathcal{N}(\overline{H}_1)$ denotes a neighbourhood of the matrix $\overline{H}_1$ (see Muirhead (1982, Chapter 9, p. 394) for a more detailed explanation of this point). This result will allow a simple form for the Hessian of $-\ln f(UH_1)$ at $\overline{H}_1$.

First note that the Hessian of $-\ln f = \frac{1}{2}\ln|H_1' D_0^u H_1| - \frac{1}{2}\ln|H_1' D_1^u H_1|$ has the form $\Psi_r = J_{H,h}\Psi_H J_{H,h}$ where

$$\Psi_H = -\partial^2 \ln f/(\partial\, vec\, H_1)'(\partial\, vec\, H_1) = \Psi_0 - \Psi_1.$$

Using standard results for obtaining matrix differentials (see Magnus and Neudecker, 1988), for $l = 0, 1$ and using $\beta = UH_1$,

$$\Psi_l = [(\beta' D_l \beta)^{-1} \otimes (D_l - D_l\beta(\beta' D_l \beta)^{-1}\beta' D_l)]$$
$$- [(\beta' D_l \beta)^{-1}\beta' D_l \otimes D_l \beta(\beta' D_l \beta)^{-1}]K_{n,r},$$

where for the $(n \times r)$ matrix E, $K_{n,r} vec(E) = vec(E')$.

The $nr \times (r/2)(2n - r - 1)$ matrix $J_{H,h}$ contains the partial differentials of H_1 with respect to the free elements of H_1 denoted by h_{ij}, $\mathrm{d}\, vec(H)/\mathrm{d}\, vec(h_{ij})$. From Muirhead (1982), since $H_1 \in V_{r,n}$ there exists a $n \times n$ orthogonal matrix $H = [H_1 : -]$ given by

$$[H_1 : -] = \exp(X) = I_n + X + \frac{1}{2}X^2 + \frac{1}{3!}X^3 + \cdots, \tag{A.1}$$

where X and the submatrix formed by the first r rows and columns of X are skew symmetric. If H has ijth element h_{ij} and X has x_{ij}, then $h_{ii} = 1 - \frac{1}{2}\sum_{j=1}^{n} x_{ij}^2 +$ higher order terms, $i \leqslant r$ and $h_{ij} = x_{ij}+$higher order terms $(i \neq j)$, $x_{ij} = -x_{ji}$ (see James, 1969 for details). In the neighbourhood $\mathcal{N}(\overline{H}_1)$, $X_{11} = 0$ and $X_{12} = 0$. Differentiate (A.1) once and set all remaining $x_{ij} = 0$ to obtain $\mathrm{d}\, vec(H)/\mathrm{d}\, vec(h_{ij})$ and thus the matrix $J_{H,h}$ at $H_1 = \overline{H}_1$.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19 (6), 716–723.

Anderson, G.A., 1965. An asymptotic expansion for the distribution of the latent roots of the estimated covariance matrix. Annals of Mathematical Statistics 36, 1153–1173.

Anderson, T.W., 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals of Mathematical Statistics 22, 327–351.

Bauwens, L., Giot, P., 1998. A Gibbs sampling approach to cointegration. Computational Statistics 13, 339–368.

Bauwens, L., Lubrano, M., 1996. Identification restrictions and posterior densities in cointegrated Gaussian VAR systems. In: Fomby, T.B. (Ed.), Advances in Econometrics, Vol. 11B, Bayesian Methods Applied to Time Series Data, JAI Press, Greenwich, pp. 3–28.

Billingsley, P., 1979. Probability and Measure. Wiley, New York.

Boswijk, H.P., 1996. Testing identifiability of cointegrating vectors. Journal of Business and Economic Statistics 14, 153–160.

Chikuse, Y., 1990. The matrix angular central Gaussian distribution. Journal of Multivariate Analysis 33, 265–274.

Drèze, J.H., 1978. Bayesian regression analysis using poly-t densities. Journal of Econometrics 6, 329–354.

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

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

Granger, C.W.J., 1983. Cointegrated variables and error correction models. Unpublished UCSD Discussion paper, 83-13.
Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society B 41, 190–195.
James, A.T., 1954. Normal multivariate analysis and the orthogonal group. Annals of Mathematical Statistics 25, 40–75.
James, A.T., 1969. Test of equality of the latent roots of the covariance matrix. In: Krishnaiah, P.R. (Ed.), Multivariate Analysis, Vol. II, Academic Press, New York, pp. 205–218.
Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.
Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 69, 111–132.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.
Kass, R.E., Vaidyanathan, S.K., 1992. Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. Journal of the Royal Statistical Society B 54, 129–144.
Kleibergen, F., Paap, R., 2002. Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. Journal of Econometrics 111, 223–249.
Kleibergen, F., van Dijk, H.K., 1994. On the shape of the likelihood/posterior in cointegration models. Econometric Theory 10, 514–551.
Kleibergen, F., van Dijk, H.K., 1998. Bayesian simultaneous equations analysis using reduced rank structures. Econometric Theory 14, 701–743.
Lewis, S.M., Raftery, A.E., 1997. Estimating Bayes factors via posterior simulation with the Laplace–Metropolis estimator. Journal of the American Statistical Association 92, 648–655.
Lindley, D.V., 1980. Approximate Bayesian methods. In: Bernardo, J.M., Degroot, M.H., Lindley, D.V., Smith, A.M.F. (Eds.), Bayesian Statistics. University Press, Valencia, Spain.
Luukkonen, R., Ripatti, A., Saikkonen, P., 1999. Testing for a valid normalisation of cointegrating vectors in vector autoregressive processes. Journal of Business and Economic Statistics 17, 195–204.
Magnus, J.R., Neudecker, H., 1988. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York.
Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.
O'Hagan, A., 1995. Fractional Bayes factors for model comparison. Journal of the Royal Statistical Society B 57, 99–138.
Phillips, P.C.B., 1994. Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models. Econometrica 62, 73–93.
Raftery, A.E., 1994. Approximate Bayes factors and accounting for model uncertainty in generalized linear models. Technical Report No. 255, Department of Statistics, University of Washington.
Schott, J.R., 1997. Matrix Analysis for Statistics. Wiley, New York.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Strachan, R.W., 2003. Valid Bayesian estimation of the cointegrating error correction model. Journal of Business and Economic Statistics 21, 185–195.
Tierney, L., Kadane, J.B., 1986. Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association 81, 82–86.
Tierney, L., Kass, R.E., Kadane, J.B., 1989. Approximate marginal densities of nonlinear functions. Biometrika 76, 425–433.
Villani, M., 2000. Aspects of Bayesian cointegration. PhD. Thesis, University of Stockholm, Sweden.
Villani, M., 2001. Bayesian reference analysis of cointegration. Working paper, University of Stockholm, Sweden.
Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.