Bayesian Analysis of Stochastic and Deterministic Processes in The Error Correction Model.

Rodney W. Strachan[1], University of Liverpool

Brett Inder, Monash University

ABSTRACT

In this article a method for joint estimation of the number of stochastic trends and the deterministic processes in a multivariate error correction model is presented. This approach takes advantage of the Laplace method of approximating integrals and, the second important contribution of the paper, careful elicitation of the prior for the cointegrating vectors from a prior on the cointegrating space. The approach follows the classical approaches of James (1969), Anderson (1951) and Johansen (1988 and 1991) and performs well when used to estimate the number of stochastic trends compared with information criteria in finite samples in Monte Carlo experiments.

JEL classification: C11; C32.

Keywords: Stochastic trend; Deterministic trend; Posterior probability; Grassman manifold; Stiefel manifold.

[1]Correspondoing author: Rodney W. Strachan, Department of Economics and Accounting, University of Liverpool, Liverpool, L69 7ZH, Great Britain. Ph: +44 (0)151 794 3043, Fx: +44 (0)151 794 3028, email: rstrach@liverpool.ac.uk

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 and has found applications in many theories such as for real business cycle models, term structure of interest rates, purchasing power parity, and money demand to name just a few. 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 (forthcoming) also used an MCMC approach but produced a posterior distribution for the rank. While there have been a few other approaches (see for example Strachan, forthcoming), none of these has produced a simple, efficient test which performs consistently well.

The first aim of this paper is to present a simple Bayesian test for the rank which 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 and 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.

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 as evidenced by the range of distributions necessary to conduct classical inference in cointegrated models, the presence (or absence) of various deterministic processes affects inference about the number of stochastic trends. This influence is reflected in the changing location of the posterior distribution of r for models with different deterministic processes. Therefore, we extend the analysis to consideration of various models of deterministic terms within the ECM and present a method for estimation of the joint posterior distribution of (r, i), where i is an indicator of the deterministic process present in the model.

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 Two the models are described. The likelihood, the priors and a general form for the posterior are given in Section

Three. 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 Four introduces the inferential tool, the Bayes factor and the application of Laplace approximation is presented in Section Five. Monte Carlo experiments are reported in Section Six and the technique is applied to an actual data set in Section Seven. Section Eight concludes.

Throughout this paper the notation '$a \equiv b$' implies that models a and b are equivalent. The space spanned by a matrix A is denoted $sp(A)$ and its j^{th} largest eigenvalue is $\lambda_j(A)$, $V_{r,n} = \{V(n \times r) : V'V = I_r\}$ is the Stiefel manifold, $O(r) = \{C(r \times r) : C'C = I_r\}$ denotes the orthogonal group of $r \times r$ orthogonal matrices. $\mathfrak{p} \in G_{r,n-r}$ denotes that $\mathfrak{p}$ is an $r-$dimensional plane in $n-$space, passing through the origin of that space, and hence $\mathfrak{p}$ is an element of a Grassman manifold, $G_{r,n-r}$. Importantly, 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 reader is referred to Muirehead (1982) and in particular James (1954) for more details on these matrix spaces.

2 The model.

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

$$\begin{aligned}
\Delta y_t &= y_{t-1}\beta^+\alpha + d_t\mu + \Delta y_{t-1}\Gamma_1 + \ldots + \Delta y_{t-l}\Gamma_l + \varepsilon_t & (1) \\
&= y_{t-1}\beta^+\alpha + d_t\mu_1\alpha + d_t\mu_2\alpha_\perp + \Delta y_{t-1}\Gamma_1 + \ldots + \Delta y_{t-l}\Gamma_l + \varepsilon_t \\
&= z_{1,t}\beta\alpha + z_{2,t}\Phi + \varepsilon_t & (2)
\end{aligned}$$

where $\Delta y_t = y_t - y_{t-1}$, $z_{1,t} = (d_t, y_{t-1})$, $z_{2,t} = (d_t, \Delta y_{t-1}, \ldots, \Delta y_{t-l})$, $\Phi = (\alpha'_\perp \mu'_2, \Gamma'_1, \ldots, \Gamma'_l)'$ and $\beta = \left(\mu', \beta^{+\prime}\right)'$. 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 $n_i \times r$, where n_i depends upon the deterministic processes present and is defined in the next section, and $rank(\beta\alpha) = r \leq n$. When $r < n$ this implies y_t is cointegrated. Expressing $z_{1,t}\beta$ as $z_{1,t}\beta = d_t\mu_1 + y_{t-1}\beta^+$, then β^+ is the matrix of cointegration coefficients and α is the matrix of factor loading coefficients or adjustment coefficients.

The $(j+1)^{th}$ element of the vector d_t, is t^j such that d_t contains the deterministic terms such as constants and trends. We will restrict ourselves to considering $j \in (0, 1)$ such that $d_t = (1, t)$ and $\mu = (a \quad \delta)' = \mu_1\alpha + \mu_2\alpha_\perp$. Restricitions on these trends and constants entering either the levels, y_t, or the cointegrating relations, $y_t\beta^+$, are discussed further in the next subsection.

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

$$\Delta y_t = \widetilde{z}_t B + \varepsilon. \tag{3}$$

2.1 Deterministic terms

It is well known that simplistic treatment of the deterministic terms by testing whether μ or some elements of μ are zero leads to the strange and unsatisfactory situation that very different trending behaviour is implied in the levels of the process for differing values of r. For example, a nonzero intercept, a, in (1) simply produces a nonzero mean when $r = n$, but it could induce a linear drift in y_t and a nonzero mean for the error correction term, $y_t\beta^+$, when $r < n$. It is for the purpose of incorporating a range of deterministic behaviours, such as drifts and trends in the cointegrating relations and in the levels, that μ is decomposed into $\mu = \mu_1\alpha + \mu_2\alpha_\perp$ where $\mu_1 = \mu\alpha' (\alpha\alpha')^{-1}$ and $\mu_2 = \mu\alpha'_\perp (\alpha_\perp\alpha'_\perp)^{-1}$ (see Johansen, 1995 Section 5.7 for further discussion).

Economists are commonly interested in the presence or absence of deterministic processes in y_t or $y_t\beta^+$. Important are questions such as whether linear or quadratic drifts are present in y_t and whether nonzero constant terms and deterministic trends are present in $y_t\beta^+$. Assuming $d_t = (1, t)$, then for each $j = 1, 2$, $d_t\mu_j = \mu_{j,\iota} + t\mu_{j,\delta}$. Although a wider range of models are clearly available, the five most commonly considered may be stated as follows, where $M_{r,i}$ is the i^{th} model of deterministic terms at given rank r :

$$
\begin{aligned}
M_{r,1} &: d_t\mu = \mu_{1,\iota}\alpha + \mu_{2,\iota}\alpha_\perp + \left(\mu_{1,\delta}\alpha + \mu_{2,\delta}\alpha_\perp\right) t \\
M_{r,2} &: d_t\mu = \mu_{1,\iota}\alpha + \mu_{2,\iota}\alpha_\perp + \mu_{1,\delta}\alpha t \\
M_{r,3} &: d_t\mu = \mu_{1,\iota}\alpha + \mu_{2,\iota}\alpha_\perp \\
M_{r,4} &: d_t\mu = \mu_{1,\iota}\alpha \\
M_{r,5} &: d_t\mu = 0
\end{aligned}
$$

A total of $5(n+1)$ models of deterministic terms and numbers of stochastic terms are considered in this article. Notice that at $r = n$, $M_{n,1} \equiv M_{n,2}$ and $M_{n,3} \equiv M_{n,4}$ since $\alpha_\perp = 0$. Similarly, at $r = 0$, $M_{0,2} \equiv M_{0,3}$ and $M_{0,4} \equiv M_{0,5}$ since $\alpha = 0$. Finally, $n_i = n + 2$ for $i = 1, 2$, $n_i = n + 1$ for $i = 3, 4$ and $n_i = n$ for $i = 5$.

3 Priors and posteriors.

In this section the forms of the priors and resultant posterior are presented. We restrict ourselves to flat priors where possible, although consideration is given to informative priors when discussing the parameters of interest. For the model in (3), assume the rows of $\varepsilon = (\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_T')'$ are $\varepsilon_t \sim iidN(0, \Sigma)$. The likelihood can then be written as $L\left(y|\Sigma, B, \beta, r, \widetilde{Z}\right) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}\varepsilon'\varepsilon\right)\right\}$.

3.1 The prior for (Σ, B, r, i).

Throughout this paper, the prior for the rank r is $p(r) = (n+1)^{-1}$and for the deterministic models $p(i) = 1/5$. The standard diffuse prior for Σ, $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$, is used.

The matrix B changes dimensions across the different models $M_{r,i}$. Thus if the prior on B is $p(B|\beta, r, i) = \mathfrak{c}_{r,i} h(B)$ where $\mathfrak{c}_{r,i}^{-1} = \int h(B)(dB)$, then clearly $\mathfrak{c}_{r,i}$ depends upon (r, i). As discussed in O'Hagan (1995), Bayes factors for $M_{r,i}$ to M_{r^*,i^*}, from which the posterior probabilities are derived, are proportional to the ratio $\iota = \mathfrak{c}_{r,i}/\mathfrak{c}_{r^*,i^*}$ and therefore knowledge of ι is required. If an improper prior on B such as $h(B) = 1$, were used, then $\mathfrak{c}_{r,i}$ does not exist but can be treated as an unspecified constant such that $\mathfrak{c}_{r,i}/\mathfrak{c}_{r,i} = 1$ and as a result the posterior will be well defined. However, as ι will be unspecified, the resulting Bayes factors can not be obtained. For this reason a (weakly) informative proper prior for B must be used. In this article, the prior for B conditional upon (Σ, β, r, i) is normal with zero mean and covariance $\Sigma \quad \left(\widetilde{\beta}' \underline{H} \widetilde{\beta}\right)^{-1}$ where$\underline{H} = 0.01 I_{(n+k_i)}$ and

$$\widetilde{\beta} = \begin{bmatrix} \beta & 0 \\ 0 & I_{k_i} \end{bmatrix}.$$

3.2 Eliciting a prior on β.

Linear restrictions and the cointegrating space: It is well known that as β and α appear as a product in (2), r^2 restrictions need to be imposed on the elements of β and α to just identify these elements. These restrictions are commonly imposed upon β by assuming $c\beta$ is invertible for known matrix c and the restricted β is $\overline{\beta} = \beta\,(c\beta)^{-1}$. Thus the free elements are collected in $\beta_2 = c_{\perp}\overline{\beta}$ where $c_{\perp}c' = 0$. A common choice in theoretical work is $c = [I_r\ 0]$ such that $\overline{\beta} = [I_r\ \beta_2']'$. A prior is then specified for β_2.

The practical problems in classical analysis of incorrectly selecting c were discussed in Boswijk (1996) and Luukkonen, Ripatti and Saikkonen (1999) and in Bayesian analysis by Strachan (forthcoming). In each of these papers compelling examples were provided of the importance of correctly determining c. Assumng 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) impropriety of the posterior have been detailed by Kleibergen and van Dijk (1994) and Bauwens and Lubrano (1996). In addition, unpublished notes by Bauwens and Lubrano showed nonexistence of the posterior when another important and commonly employed restriction on (2), exogeneity, is imposed. Further, from the discussion on the prior for B it is clear a flat prior on β_2 cannot be employed to obtain posterior probabilities for (r,i), since the dimensions of β_2 depend upon (r,i).

As argued in the introduction, an advantage of the Bayesian approach is the ability to explicitly incorporate prior beliefs into the analysis. As a flat prior is generally intended to reflect ignorance about the parameter of interest, the above issues with the posterior at least, may be resolved by use of an informative prior on β_2. However, to preserve the options of both informative and uninformative priors, as the issue of selecting c is not resolved by a proper prior, and as we do not see β_2 as the parameter of interest, we therefore diverge at this point from much of the earlier literature (except Villani 2000) in both specifying our parameter of interest and eliciting uninformative and informative priors on that parameter.

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 (ignoring for now the dependence on n_i and assume $n_i = n$) 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 (2000) as $|I_r + \beta_2'\beta_2|^{-n/2}$. Although Villani (2000) uses $c = [I_r \ 0]$, this form holds for general c and is the kernal of a Cauchy density. 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 when an element of the Steifel manifold - which defines an element of a Grassman manifold - is renormalised by imposing linear restrictions. That is Phillips (1994) shows 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 nonzero - more generally linearly independent from eachother - and then normalise on these coefficients. This is a necessary assumption to ensure $(c\beta)^{-1}$ exists. By using these linear restrictions, however, the Jacobian for $\beta_2 \rightarrow \mathfrak{p}$ places 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. To use linear restrictions a normalisation must be chosen by choice of c. 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 making it safe to normalise on β_1, 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

$$J\left(\beta_2 : \beta\right) = \left|I_r + \left(c\beta\right)'^{-1}\beta' c_\perp' c_\perp \beta \left(c\beta\right)^{-1}\right|^{n/2}.$$

As $\mathfrak{p} = sp\left(\beta\right) \to sp\left(c\right)$, $c_\perp \beta \to 0_{(n-r)\times r}$ and $c\beta \to O\left(r\right)$ and $J\left(\beta_2 : \beta\right) \to 1$. However, as vectors in β approach the null space of c, that is $\det\left(c\beta\right) \to 0$, then $\left(c\beta\right)^{-1} \to \infty$, and thus $J\left(\beta_2 : \beta\right) \to \infty$. As a result the prior will more heavily weight regions where $\det\left(c\beta\right) = \det\left(\beta_1\right) \approx 0$, contrary to the intention of the economist. As a trivial example, if $r = 1$, we would choose $c = \left(1, 0, \ldots, 0\right)$ as we believe $\beta_1 \neq 0$. Yet the Jacobian places infinite weight in the region of $\beta_1 = 0$.

A uniform prior on the cointegrating space: Clearly then there is reason to consider another approach to eliciting priors for β. Our recommendation is, if the economist wishes to incorporate prior beliefs about the cointegrating relations, these should be expressed in the prior distribution for the cointegrating space.

As we have claimed the cointegrating space to be the parameter of interest, we propose working directly with $\mathfrak{p} = sp\left(\beta\right)$ and avoiding the linear restrictions. Initially a distribution and identifying restrictions for β from the uniform distribution for $\mathfrak{p}$ over $G_{r,n-r}$ is derived using the results of James (1954). This prior has the form

$$p\left(\beta\right)\left(\beta' d\beta\right) = \frac{1}{\int \left(\beta' d\beta\right)}\left(\beta' d\beta\right)$$

where $\left(\beta' d\beta\right)$ is the exterior product differential form for the free elements of β and defines the invariant (to left and right orthogonal translations) measure on $G_{r,n-r}$ and is equivalent to the product of the Jacobian for the transformation from $\mathfrak{p}$ to β and the differential for the elements of β (see James 1954, Muirhead 1982, Ch. 2). This expression for the prior defines a probability measure on the space of β. Throughout the paper, to save on notation we omit the differential term for all parameters except β as this is the focus of the analytical results.

As the support of β is a function of (r, i), so will be $\int (\beta' d\beta)^2$. Therefore to employ the above distribution it is necessary to find the explicit form for $\int (\beta' d\beta)$. This is obtained by using the relationship between $G_{r,n-r}$ and the Steifel manifold and orthogonal group. We reproduce this result from James (1954) as we will rely on some of its implications later. If $A \in V_{r,n_i}$ and $\mathfrak{p} = sp(A)$, then $\mathfrak{p} \in G_{r,n_i-r}$ and A is determined uniquely given $\mathfrak{p}$ and orientation of A in $\mathfrak{p}$ by $C \in O(r)$, such that $A = \beta C$ where $\beta \in V_{r,n_i}$, $sp(\beta) = \mathfrak{p}$. James (1954)[3] shows

$$\begin{aligned} \int_{G_{r,n_i-r}} (\beta' d\beta) &= \frac{\int_{V_{r,n_i}} (A' dA)}{\int_{O(r)} (C' dC)} \\ &= \pi^{-(n_i-r)r} \Pi_{j=1}^{r} \frac{\Gamma\left[(n_i+1-j)/2\right]}{\Gamma\left[(r+1-j)/2\right]} \end{aligned} \tag{4}$$

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

In early work in this area, Villani (2000) also began with a uniform prior upon the cointegrating space from which he derived the prior distribution for β_2. However

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

[3] There is an error in (5.23) of James (1954). The sums, Σ, should be products, Π.

his interest was in estimation and tests related to $sp(\beta)$ and as such, the analysis was conditional upon (r, i). For reasons discussed earlier, we wish to avoid using linear restrictions to identify β and thus must find an alternative set of restrictions that do not require knowledge of c and which avoid the issues associated with the posterior for β_2. Fortunately the definition of A provides a natural solution to this question. That is use $\beta \in V_{r,n_i}$ which implies $r(r+1)/2$ restrictions, with the usual assumption about fixing the sign of the element in the first row. This latter assumption simply restricts the vectors of β to one half hemisphere but in no way restricts the estimable space. Although the dimension of the Grassman manifold is only $(n_i - r)r$, the remaining $r(r-1)/2$ restrictions come from the orientation by C. The prior, the posterior (as is made clear later) and the differential form for β are all invariant to translations of the form $\beta \to \beta H$, $H \in O(r)$. Therefore it is possible to work directly with β as an element of the Steifel manifold and adjust the integrals with respect to β by $\left(\int_{O(r)} (C'dC)\right)^{-1}$ as shown in (4). Note that these identifying restrictions do not distort the weight on the space of the parameter of interest, $\mathfrak{p}$.

An informative prior on the cointegrating space: Although the uniform prior (4) is used in this paper, it is common to employ informative priors for parameters and so one is specified here for $\mathfrak{p}$.

If an economist 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 $\mathfrak{p}$, we will denote the likely value as $\mathfrak{p}^H = sp(H\kappa)$ where $H \in V_{s,n}$ (again we ignore the dependence on n_i and assume $n_i = n$) is a known $n \times s$ $(s \geq r)$ matrix, $H_\perp \in V_{n-s,n}$ its orthogonal compliment 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 $H = H^g \left(H^{g\prime} H^g\right)^{-1/2}$.

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 prior in (4) for V_κ. This prior assigns probability one to the point $\mathfrak{p} = \mathfrak{p}^H$.

Often, however, the economist will want to employ a less dogmatic prior such that there is some weight away from the likely value. A possible 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

$$\begin{aligned} P_\tau &= HH' + H_\perp H_\perp' \tau \\ &= [H \ H_\perp] \begin{bmatrix} I_r & 0 \\ 0 & I_{n-r}\tau \end{bmatrix} \begin{bmatrix} H' \\ H_\perp' \end{bmatrix} \end{aligned}$$

and let the elements of the $n \times r$ matrix Z be independently distributed as standard normal, $N(0,1)$. 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 $N(0,1)$. Integrating with respect to $K = \kappa\kappa'$ which is distributed as Wishart, the form of the resultant prior for β and the hyperparameter τ is

$$p(\tau,\beta)(\beta' d\beta)(d\tau) = \tau^{-(n_i - r)r} \exp\left\{-\frac{\tau^2}{2}\right\} \left|\beta' P_{\tau^2}^{-1} \beta\right|^{-n/2} \mathfrak{c}_{(r,i)} (\beta' d\beta)(d\tau)$$

where $\mathfrak{c}_{(r,i)} = 2^{-r-1/2} \pi^{r(r-1)/4-(n+1)r/2} \Pi_{j=1}^{r} \Gamma\left[(n_i + 1 - j)/2\right]$. This prior treats the area around $\mathfrak{p}^{H_\perp}$, which occurs at $\tau = \infty$, as an extreme (practically impossible) event regardless of the choice of σ. This is desirable since at $\tau = \infty$ the dimension of the cointegrating space, $\dim(\mathfrak{p})$, would become $\dim\left(\mathfrak{p}^{H_\perp}\right) = \min(p - r, r)$ rather than r.

As an alternative, if the researcher would prefer to assign more weight in the direction of $\mathfrak{p}^{H_\perp}$ but preserve $\dim(\mathfrak{p}) = r$ with probability one, she may choose $P_\tau = HH'\left(1 - \tau^2\right)^{1/2} + H_\perp H_\perp' \tau$ with $\tau \in [-1, 1]$. Again the choice of $E(\tau) = 0$ would make sense and $E(\tau^2) = \sigma^2$ controls the tightness of the prior around $\mathfrak{p}^H$. A possible choice of a distribution for $\eta = \tau + 1$ may be Beta over $\eta \in [0, 2]$ which allows some mass to be distributed around $\mathfrak{p}^{H_\perp}$ by appropriate choice of parameter values.

An MC(MC) scheme for obtaining draws from the posterior with either the uniform or informative prior could be developed using the form of the above informative prior as a candidate density in which H is set near to the mode of the posterior (see Appendix). Draws of β could then be drawn using the above outline by drawing τ

and Z, then constructing X then β, giving a draw of (τ, β). Finally, the value of σ could be calibrated to a preferred level of dispersion using the span variation measure (sv) of Larsson and Villani (2001), developed further in Villani (2000), and MC(MC) draws. Villani (2000) shows how sv can be used to express the degree of variation in a distribution for as a proportion of the variation under the uniform distribution - the uniform providing equal variation in every direction.

3.3 The posteriors.

Using the priors specified above, the general form of the posterior is then

$$\begin{aligned} p(B, \Sigma, \beta, r, i|y) \quad \propto \quad & |\Sigma|^{-(T+n+r+k_i+1)/2} \\ & \times \exp\left\{-\frac{1}{2}tr\Sigma^{-1}\left[TS + \left(B - \widetilde{B}\right)' V \left(B - \widetilde{B}\right)\right]\right\} \\ & \times (2\pi)^{-n(k_i+r)/2}\, 100^{-n(k_i+r)/2} p(\beta)(\beta' d\beta) \end{aligned} \tag{5}$$

where $S = S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10}$, $\widetilde{B} = \begin{bmatrix} \widetilde{\alpha}' & \widetilde{\Phi}' \end{bmatrix}'$, $\widetilde{\alpha} = \left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10}$, $\widetilde{\Phi} = S_{22}^{-1} S_{20}$, and $V = \widetilde{\beta}'\left(\Sigma_{t=1}^{T} z_t' z_t + \underline{H}\right)\widetilde{\beta}$ where $z_t = (z_{1,t}\ z_{2,t})$. The values for the S_{ij} are defined as

$$\begin{aligned} \nu M_{ij} &= h_{ij} + \Sigma_{t=1}^{T} z_{i,t}' z_{j,t} \qquad \text{for } i \text{ and } j = 1, 2, \\ h_{ij} &= 0 \text{ if } i \neq j \text{ and } h_{ii} = 0.01I, \end{aligned}$$

$$\begin{aligned} \nu M_{20} &= \Sigma_{t=1}^{T} z_{2,t}' \Delta y_t, \qquad \nu M_{10} = \Sigma_{t=1}^{T} z_{1,t}' \Delta y_t, \\ \nu M_{00} &= \Sigma_{t=1}^{T} \Delta y_t' \Delta y_t \quad \text{and so } S_{ij} = M_{ij} - M_{i2} M_{22}^{-1} M_{2j} \text{ for } ij = 0, 1, 2, \end{aligned}$$

except $i = j = 2$ where

$$S_{22} = M_{22} - M_{21}M_{11}^{-1}M_{12} \text{ and } S_{20} = M_{20} - M_{21}M_{11}^{-1}M_{10}.$$

For later use we also define $D_0 = D_1 - D_2$, $D_1 = S_{11}$ and $D_2 = S_{01}S_{11}^{-1}S_{10}$.

4 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 model $M_{i,r}$, $p\left(M_{i,r}|y\right) = p\left(i, r|y\right)$. Let B_{kl} be the Bayes factor for the model M_k, $k = (r, i)$ to the model M_l, $l = (r^*, i^*)$. The posterior probabilities and the Bayes factors are linked through the expression for the posterior odds ratio. First consider the posterior odds ratio for model M_k to model M_l with parameters θ_k and θ_l respectively,

$$\begin{aligned} \frac{p\left(k|y\right)}{p\left(l|y\right)} &= \frac{p\left(k\right)\int L\left(\theta_k\right)p\left(\theta_k|k\right)d\left(\theta_k\right)}{p\left(l\right)\int L\left(\theta_l\right)p\left(\theta_l|l\right)d\left(\theta_l\right)} \\ &= \frac{p\left(k\right)}{p\left(l\right)} \times \frac{m\left(y|k\right)}{m\left(y|l\right)} = \frac{p\left(k\right)}{p\left(l\right)} \times B_{k,l}. \end{aligned}$$

where $p\left(k\right)$ is the prior probability of the model k and

$$m\left(y|k\right) = \int L\left(\theta_k\right)p\left(\theta_k|k\right)d\left(\theta_k\right) \tag{6}$$

is the marginal likelihood for the model k. As the prior odds are known, we need only estimate $B_{k,l}$.

To estimate the relevant Bayes factors for the models of interest, estimates of the marginal likelihoods in (6) are required. To perform the integration in (6) of $\theta_k = (\Sigma, B, \beta)$, first analytically integrate (5) with respect to (Σ, B). From the expression in (5), it is straight forward to show that the posterior for (Σ, B) conditional on (β, r, i) has a standard form which may be integrated analytically (see for example Zellner, 1971). The resulting posterior for the remaining parameters is

$$p(\beta, r, i|y) \propto \mathfrak{g}_{(r,i)} |S_{00}|^{-T/2} |M_{22}|^{-n/2} |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2} (\beta' d\beta) \qquad (7)$$

where in this case $\mathfrak{g}_{(r,i)} = T^{-nr/2} \pi^{-(n_i - r)r/2} 100^{-n(k_i+r)/2}$. The conditional density for β given (r, i) has the form

$$\begin{aligned} p(\beta|r, i, y)(\beta' d\beta) &\propto |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2} (\beta' d\beta) \\ &= k(\beta)(\beta' d\beta) \end{aligned}$$

where $k(\beta) = |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2}$. $(\beta' d\beta)$ is invariant to $\beta \to \beta C$ for $C \in O(r)$ and the above form makes it clear that so is $k(\beta)$. The eigenvalues $\lambda_j(D_l)$ for $l = 0, 1$, will be positive and finite with probability one. By the Poincaré separation theorem, since $\beta \in V_{r,n}$, then $\Pi_{j=1}^{r} \lambda_{n-r+j}(D_l) \leq |\beta' D_l \beta| \leq \Pi_{j=1}^{r} \lambda_j(D_l)$ and so $k(\beta)$ is bounded above (and below) by some positive finite constant. Also, $(\beta' d\beta)$ is integrable and therefore finite almost everywhere. Thus $k(\beta)(\beta' d\beta)$ has a finite upper bound, M. As the elements of β, b_{ij}, have compact support, the integral $\int_{V_{r,n}} b_{ij}^m k(\beta)(\beta' d\beta)$ for $m = 0, 1, ...$ will be bounded above almost everywhere by the integral $M \int_{-1}^{1} b_{ij}^m db_{ij}$.

These conditions are sufficient to ensure the posterior for β will be proper and all finite moments exist (see Billingsley 1979, pp. 174 and 180).

To obtain the posterior distribution of (r,i), $p(r,i|y)$, it is necessary to integrate (7) with respect to β and so obtain an expression for

$$\begin{aligned} p(r,i|y) &= \int p(\beta,r,i|y)\,d\beta \\ &= \int p(\beta|r,i,y)\,p(r,i|y)\,d\beta \\ &= \int p(\beta|r,i,y)\,d\beta p(r,i|y)\,. \end{aligned} \tag{8}$$

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

$$p(\beta|r,i,y) = \mathfrak{c}_{(r,i)}^{-1} k(\beta)(\beta' d\beta) \tag{9}$$

which is not of standard form. Although one may exist, we do not currently know of a simple, general analytical solution for $\mathfrak{c}_{(r,i)} = \int_{V_{r,n}} k(\beta)(\beta' d\beta)$ and so we estimate $\mathfrak{c}_{(r,i)}$.

Two possible approaches to estimating $\mathfrak{c}_{(r,i)}$ are either to use Markov Chain Monte Carlo (MCMC) methods or numerical integration. Kleibergen and Paap (forthcoming) and Bauwens and Lubrano (1996) demonstrate how to evaluate similar integrals using MCMC when β has been identified using linear restrictions rather than those used in this paper. Strachan (forthcoming) demonstrates the MCMC approach when β has been identified using related restrictions, how ever the posterior has a very different form as an embedding approach similar to Kleibergen and Paap 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 a general integral is briefly outlined and applied to obtain an estimate of $\mathfrak{c}_{(r,i)}$.

5 Laplace approximation.

Let θ be a m-dimensional vector of parameters. If $\ln f = \ln f(\theta)$ is a smooth, positive function with a maximum at $\overline{\theta}$, then by the Laplace method the integral $\int g f^{\kappa} d\theta$ can be approximated by $g\left(\overline{\theta}\right) f^{\kappa}\left(\overline{\theta}\right)\left(\frac{2\pi}{\kappa}\right)^{\frac{m}{2}} |\Psi|^{-\frac{1}{2}}$ where Ψ is the Hessian of $-\ln f$, evaluated at $\theta = \overline{\theta}$ ($g = g(\theta)$ is a continuous nonzero function around $\overline{\theta}$). There are a number of papers on applications of the Laplace approximation in econometrics (see for example Lindley 1980, Tierney & Kadane 1986, Tierney, Kass & Kadane 1989, Kass & Raftery 1995). However for more relevant references for our application to an integral over the Stiefel manifold the reader is directed to Muirehead (1982, Ch. 9), G.A. Anderson (1965) and James (1969). In these applications the aim was to derive distributions of latent roots of covariance matricies 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 will work well if the mode and Hessian, Ψ, are easy to obtain and if the posterior is reasonably peaked around the mode (Tierney and Kadane 1986, Tierney, Kass & Kadane 1989, Kass and Raftery 1995). 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, $\mathfrak{c}_{(r,i)}$, by the Laplace approximation, there is considerable precedent in the literature for using this method for this purpose (Lindley 1980, Kass & Vaidyanathan 1992, Raftery 1994, Kass & Raftery 1995, Lewis & Raftery 1997). Further, the results presented in this paper support the application of Laplace approximation at least for estimation of r.

To apply the Laplace approximation, let $k = f^T g$, $\theta = \beta$, $m = \frac{r}{2}(2n_i - r - 1)$, $\kappa = T$, $f = |\beta' D_0 \beta|^{-1/2} |\beta' D_1 \beta|^{1/2}$ and $g = |\beta' D_1 \beta|^{-n/2}$. The value of β at the mode of f will be denoted as $\overline{\beta}$ and the Hessian matrix for rank r evaluated at $\overline{\beta}$ will be $\Psi = \Psi_r$. Johansen (1991) presents a modal estimator for $f(\beta)$, $\widehat{\beta}$, from which we could obtain $\overline{\beta} = \widehat{\beta}\left(\widehat{\beta}'\widehat{\beta}\right)^{-1/2} \in V_{r,n}$. However, a slightly different derivation is presented in the Appendix such that $\overline{\beta} = D_1^{1/2}\widehat{\beta}$ so $\overline{\beta}$ is the r eigenvectors associated with the eigenvalues $\lambda_i\left(D_1^{-1/2} D_2 D_1^{-1/2}\right)$ for $i = 1, \ldots, r$. This approach simplifies derivation of the Hessian which is also presented in the Appendix.

Using (7), (8), and (9), approximate $p(r, i|y)$ by

$$\begin{aligned} \widehat{p}(r, i|y) &= \left(\frac{2\pi}{T}\right)^{r(2n_i - r - 1)/2} g\left(\overline{\beta}\right) f^{\kappa}\left(\overline{\beta}\right) |\Psi_r|^{-\frac{1}{2}} \\ &\quad \times \mathfrak{g}_{(r,i)} |S_{00}|^{-T/2} |M_{22}|^{-n/2} \text{ (see Appendix).} \end{aligned}$$

The classical maximum eigenvalue test statistic, which is a likelihood ratio test statistic for the hypothesis $H_0 : rank\,(\Pi) = r$ versus $H_1 : rank\,(\Pi) = r + 1$, has the form $m_{r,r+1} = -T \ln \left(1 - \widehat{\lambda}_{r+1}\right)$. From this expression it is possible to show the link between the classical test statistic, $m_{r,r+1}$, and the Bayes factor, $B_{(r,i)(r+1,i)}$, for the diffuse prior as $\widehat{B}_{(r,i)(r+1,i)} = c_r \exp\left(-0.5 m_{r,r+1}\right)$ where c_r depends on the data, r, and n. Denote the classical trace test statistic, which is a likelihood ratio test statistic for the hypothesis $H_0 : rank\,(\Pi) = r$ versus $H_1 : rank\,(\Pi) = n$, as $m_{r,n}$. Similarly, it is possible to present the link between $m_{r,n}$ and the Bayes factor, $B_{(r,i)(n,i)}$, for the diffuse prior as $\widehat{B}_{(r,i)(n,i)} = \Pi_{j=r}^{n} c_j \exp\left(-0.5 m_{r,n}\right)$.

6 Monte Carlo experiment.

To investigate the small sample performance of our estimator for $p\,(r, i|y)$, we conduct Monte Carlo experiments and compare these results to those for the a range of information criteria and the classical trace test. In the next section, the test is applied to a set of real data.

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 $\varepsilon_{j,t} \backsim iidN\left(0, \sigma^2\right)$ 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 $\delta = \begin{bmatrix} \delta_1 + \delta_2\beta_2, & \delta_2 \end{bmatrix}$, $\mu = \begin{bmatrix} \mu_1 + \mu_2\beta_2 - \delta_0, & \mu_2 \end{bmatrix}$, $\alpha = \left[(\rho - 1)\, I_r \quad 0\right]$, $\beta = \begin{bmatrix} -\mu_0' & -\delta_0' & I_r & -\beta_2' \end{bmatrix}'$, $\Gamma_1 = \left[\Gamma_{11}',\ \Gamma_{21}'\right]$, $\Gamma_{11} = 0$, $\Gamma_{21} = \begin{bmatrix} \rho\beta_2 & I_{n-r}\rho \end{bmatrix}$, and $\varepsilon_t = (\varepsilon_{1,t} + \varepsilon_{2,t}\beta_2, \quad \varepsilon_{2,t})$.

The μ_j and δ_j are set equal to 0.35ι in which ι is a vector of ones, $\rho = 0.35$, $\sigma = 1.5$, $T = 100$, 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 will have Monte Carlo standard errors of at most 0.005.

The range of models simulated is for each $i = 1, \ldots, 5$, $n = 2, 3, 4$ and $r = 1, 2$ for a total of 30 experiments. For each model the combination of (r, i) 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 $p\,(r, i|y)$ is also compared to that of $m_{r,n}$ at the 5% ($m_{r,n}^{5\%}$) and the 1% ($m_{r,n}^{1\%}$) level of significance. However, when using $m_{r,n}$ it is assumed i is known. While this assumption is expected to advantage the classical test, the results indicate that the Laplace estimator still performs, generally, as well and often better than $m_{r,n}$.

The marginal relative selection frequencies of the correct r are compared for the five techniques and marginal relative selection frequencies for i are compared for the information criteria and LP. 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 a combination (i, r) on which subsequent inference can condition. This does not indicate performance of the method in model averaging.

In selecting r LP was first or equal first in 20 of the 30 experiments. For the other techniques the same result was: AIC 5; BIC 6; HQ 6; $m_{r,n}^{5\%}$ 9; and $m_{r,n}^{1\%}$ 7. Ranking the techniques on frequencies of correct selection of r from 1 (best) to 6 (worst), the average ranks were LP 1.97, AIC 4.9, BIC 2.77, HQ 2.8, $m_{r,n}^{5\%}$ 2.93, and $m_{r,n}^{1\%}$ 3.33. The results were fairly consistent across the range of models although there were some patterns evident. Figure 1 shows a sample result for correct selection frequencies of r in this case for $n = 3$, $r = 2$ and over the five models of i. LP tended to perform better for the models with fewer deterministic processes ($i = 3, 4, 5$), the AIC was most frequently the worst, while $m_{r,n}^{5\%}$ performed markedly better and $m_{r,n}^{1\%}$ performed slightly better when $i = 1, 3$ or 5.

In selecting i LP was first or equal first in 13 of the 30 experiments. For the other techniques the same result was: AIC 4; BIC 13; and HQ 6. The average ranks were

LP 2.67, AIC 3.03, BIC 1.9, and HQ 2.17. The performance of LP was not consistent over i, as LP came equal last 16 times (AIC 11 times, BIC 1 and HQ 0). Figure 2 shows the sample results for selection of i. In this case the information criteria tended to perform better when $i = 1$ or 5, but poorly otherwise (particularly at $i = 4$). BIC and HQ tend to over-select $i = 5$ when in fact $i = 4$. LP performs very well when $i = 1$ or 3 (with relative frequencies near 1), but only occassionally selects correct i when $i = 2, 4$, or 5. The better performance of the information criteria is due to the treatment in the penalty function of the change in the number of parameters. The information criteria treat changes in the dimensions of β and B symmetrically, however the Laplace method distributes the changes among gamma functions and exponents. An alternative specification of the prior for i could remove this problem, however the effect diminishes with increased sample size.

The Monte Carlo results suggest that LP is useful for selecting r from the joint distribution of (r, i), however when selecting i it performs well in only some cases.

7 An illustrative example: Interest rates.

In this section we demonstrate testing for the (r, i) for four U.S. treasury bill rates. The four interest rates are the 5 year (i_5) and 1 year (i_1) Treasury Bond rates (Capital Market) and the 1 and 3 month and 1 and 5 year Treasury constant maturity rates (i_{30}, i_{180}, i_{1YR} and i_{5YR} respectively). The data are annualised monthly rates for the period January 1982 to January 1999 ($T = 214$).

These variables are useful for the study of the various theories for the term structure of interest rates. Common implications of many of these theories is that, while the rates themselves may be integrated of order one, we would expect to find in this case, three cointegrating relations. It is unlikely that interest rates would contain linear drifts suggesting $i = 4$ or 5, however over the period in this sample rates showed a clear downwards movement suggesting we may find $i = 3$. It is commonly assumed, and there is strong empirical evidence in support of this assumption, that the rates enter the cointegrating relations through the spreads. With this assumption, choosing between $i = 4$ and 5 depends upon our beliefs about the long run or equilibrium term structure of the interest rates. If we believe the term structure to be flat, this would support $i = 5$, if we believe it is sloping (up or down) this would suggest $i = 4$.

Classical pretesting suggests each series is integrated of order one and we find an ECM with two lags of differences is sufficient to model the process. The residuals in the ECM, particularly for the short rates, do not appear normal and this is largely due to excess kurtosis, however, following earlier studies using interest rates to demonstrate an application, such as Luukkonen, Ripatti and Saikkonen (1999), we ignore this feature as modelling this behaviour is outside the scope of this paper.

The information criteria select combinations of (r, i) of AIC $(4, 5)$, BIC $(1, 5)$ and HQ $(2, 5)$. Likelihood ratio tests at $r = 2$ suggest $i = 5$, and assuming $i =$

5 and using $m_{r,n}^{5\%}$ and $m_{r,n}^{1\%}$, $r = 2$ is accepted. Posterior probabilities using LP suggest $(r = 2, i = 3)$ to be the most likely combination with posterior probability of $\Pr(r = 2, i = 3|y) = 0.63$. The conditional probabilities for r also support $r = 2$ with $\Pr(r = 2|i = 3, y) = 0.68$ and $\Pr(r = 2|i = 5, y) = 0.50$. Conditioning on $i = 3$, $m_{r,n}^{1\%}$ supports $r = 2$ whereas $m_{r,n}^{5\%}$ supports $r = 4$. The acceptance by the classical test of $r = 4$ when $i = 3$ at the 5% level of significance is reflected in a Bayesian conditional posterior probability of $\Pr(r = 4|i = 3, y) = 0.25$.

8 Conclusion.

In this paper a method of finding approximations to Bayes factors has been demonstrated for models of stochastic and deterministic 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. However, the approach does not perform consistently well when used to determine the deterministic processes in the data.

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.

9 Appendix.

The Laplace approximation.

Before applying the Laplace approximation to the integral in (8), 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 $l = 0, 1, 2,$ $H_1 = U'\beta$ and since $U \in O(n)$ then $(\beta' d\beta) = (H_1' dH_1)$ by invariance of $(\beta' d\beta)$. Therefore

$$\int_{V_{r,n}} f(\beta)\, g(\beta)\, (\beta' d\beta) = \int_{V_{r,n}} f(UH_1)\, g(UH_1)\, (H_1' dH_1)$$

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

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

Since $D_1^{-1/2} D_0 D_1^{-1/2} = I_n - A$, and $\lambda_{n-r+j}\left(D_0 D_1^{-1}\right) = \lambda_{n-r+j}\left(D_1^{-1/2} D_0 D_1^{-1/2}\right)$, then this equals $1 - \lambda_j(A) = 1 - \lambda_j(\Lambda)$. Therefore, $\min f^{-2} = \Pi_{j=1}^{r}(1 - \lambda_j(\Lambda)) = \min |I_r - H_1 \Lambda H_1|$ 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,

$$\mathfrak{c} \approx 2^r \int_{\mathcal{N}\left(\overline{H}_1\right)} k(UH_1)(H_1' dH_1)$$

where $\mathcal{N}\left(\overline{H}_1\right)$ denotes a neighbourhood of the matrix $\overline{H}_1$ (see Muirehead 1982, Ch. 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

$$\begin{aligned} \Psi_H &= -\frac{\partial^2 \ln f}{(\partial vec H_1)' (\partial vec H_1)} \\ &= \Psi_0 - \Psi_1. \end{aligned}$$

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

$$\begin{aligned} \Psi_l &= \left[(\beta' D_l \beta)^{-1} \quad \left(D_l - D_l \beta (\beta' D_l \beta)^{-1} \beta' D_l \right) \right] \\ &\quad - \left[(\beta' D_l \beta)^{-1} \beta' D_l \quad D_l \beta (\beta' D_l \beta)^{-1} \right] K_{n,r} \end{aligned}$$

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

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

$$\begin{aligned}
[H_1 : -] &= \exp(X) \\
&= I_n + X + \frac{1}{2}X^2 + \frac{1}{3!}X^3 + \ldots, \qquad (10) \\
X &= \begin{bmatrix} X_{11} & X_{12} \\ -X_{12}' & 0 \end{bmatrix}
\end{aligned}$$

where X and X_{11} are skew symmetric. If H has ij^{th} element h_{ij} and X has x_{ij}, then

$$\begin{aligned}
h_{ii} &= 1 - \frac{1}{2}\Sigma_{j=1}^{n} x_{ij}^2 + \text{higher order terms, } i \leq r \text{ and} \\
h_{ij} &= x_{ij} + \text{higher order terms } (i \neq j),\ x_{ij} = -x_{ji}
\end{aligned}$$

(see James 1969 for details). In the neighbourhood $\mathcal{N}\left(\overline{H}_1\right)$, $X_{11} = 0$ and $X_{12} = 0$. Differentiate (10) once and set all remaining $x_{ij} = 0$ to obtain $\frac{dvec(H)}{dvec(h_{ij})}$ and thus the Jacobian from H_1 to h_{ij} at $H_1 = \overline{H}_1$.

10 Acknowledgements.

The authors would like to thank Herman van Dijk, Charles Bos, Frank Kleibergen, Brendan McCabe and participants of the workshop 'Recent Advances in Bayesian Econometrics' in Marseilles, June 2001 and two anonymous referees for useful discussion that has improved this paper.

11 References.

Akaike, H., 1974, A new look at the statistical model identification, I.E.E.E. 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. and M. Lubrano, 1996, Identification restrictions and posterior densities in cointegrated Gaussian VAR systems, in Advances in Econometrics, Vol. 11B, Bayesian Methods Applied to Time Series Data, T.B. Fomby, ed., (JAI Press) 3-28.

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

Billingsley, P. 1979, Probability and Measure (John Wiley and Sons, Inc., New York).

Engle, R.F. and C.W.J. Granger, 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, Discussion paper, University of California, San Diego.

Hannan, EJ and B.G. Quinn, 1979, The determination of the order of an autoregression, Journal of the Royal Statistical Society:B, 41:2, 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: P.R. Krishnaiah, ed., Multivariate Analysis, Vol. II (Academic Press, New York) 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.

Johansen, S., 1995, Likelihood-based Inference in Cointegrated Vector Autoregressive Models (Oxford University Press, New York).

Kass, R.E. and A.E. Raftery, 1995, Bayes Factors, Journal of the American Statistical Association, 90, 773-795.

Kass, R.E. and S.K. Vaidyanathan, 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. and R. Paap, forthcoming, Priors, posteriors and Bayes factors for a

Bayesian analysis of cointegration, Journal of Econometrics.

Kleibergen, F. and H.K. van Dijk, 1994, On the shape of the likelihood/posterior in cointegration models, Econometric Theory, 10, 514-551.

Larsson, R. and Villani, M. 2001, A distance measure between cointeration spaces, Economics Letters, 70, 21-27.

Lewis, S.M. and A.E. Raftery, 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: J.M. Bernardo, M.H. Degroot, D.V. Lindley, and A.M.F. Smith, eds., Bayesian Statistics, (Valencia, Spain: University Press).

Luukkonen, R, A. Ripatti and P. Saikkonen, 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. and H. Neudecker, 1988, Matrix Differential Calculus with Applications in Statistics and Econometrics (John Wiley and Sons, New York).

Muirhead, R. J., 1982, Aspects of Multivariate Statistical Theory (John Wiley and Sons, New York).

O'Hagan, A., 1995, Fractional Bayes factors for model comparison, Journal of the

Royal Statistical Society, Series B, 57, 1, 99-138.

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.

Schwarz, G., 1978, Estimating the dimension of a model, Annals of Statistics, 6:2, 461-464.

Strachan, R.W., 2003, Valid Bayesian estimation of the cointegrating error correction model, Journal of Business and Economic Statistics, 21:1, 185-195.

Strachan, R.W., and van Dijk, H. K. 2002, The value of structural information in the VAR, Discussion paper, University of Liverpool.

Tierney, L., R.E. Kass, and J.B. Kadane, 1989, Approximate marginal densities of nonlinear functions, Biometrika, 76, 425-433.

Tierney, L. and J.B. Kadane, 1986, Accurate approximations for posterior moments and marginal densities, Journal of the American Statistical Association, 81, 82-86.

Villani, M., 2000, Aspects of Bayesian Cointegration. PhD. Thesis, University of Stockholm, Sweden.

Zellner, A., 1971, An Introduction to Bayesian Inference in Econometrics (John Wiley and Sons Inc., New York).

12 Figures.

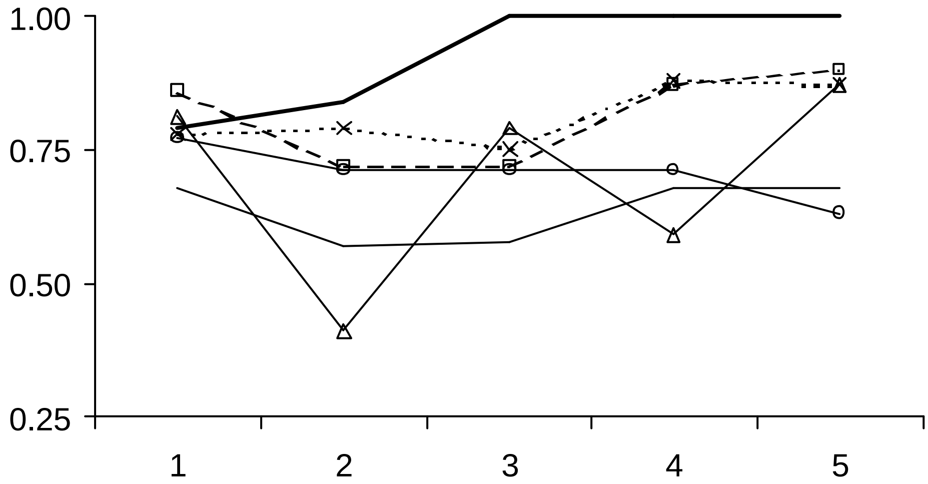

Figure 1: The above figures show the relative selection frequencies for r when $n = 3$, $r = 2$. The labels are given in the figure and the *x-axis* shows the value of i.

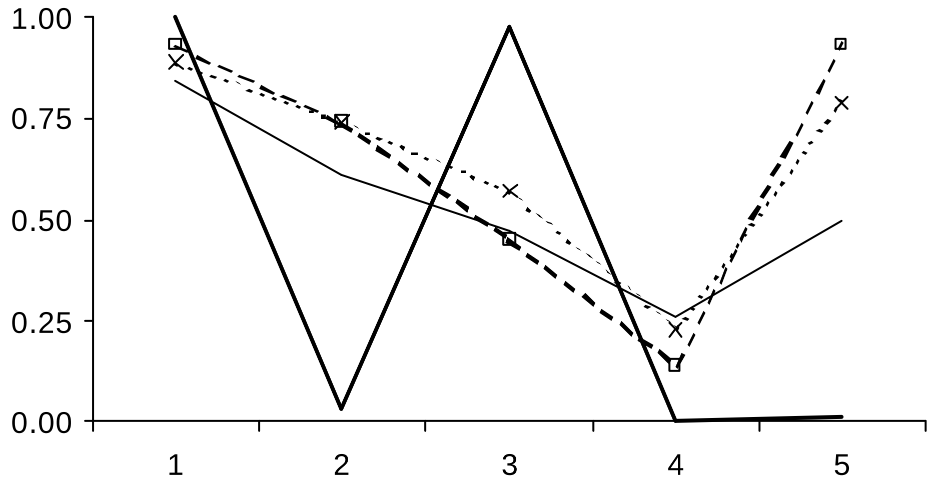

Figure 2: The above figures show the relative selection frequencies for i when $n = 3$, $r = 2$. The legend is given in Figure 1 and the *x-axis* shows the value of i.

13 Figure Legends.

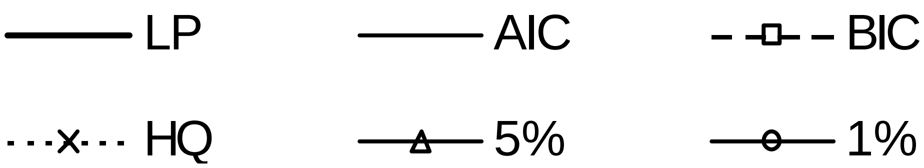

Figure 3: Legend for Figures 1 and 2. LP - Laplace approximation method; AIC, BIC and HQ are the information criteria and 5% and 1% are the classical trace tests.