Valid Bayesian Estimation of the Cointegrating Error Correction Model

Rodney W Strachan

Working Paper 6/2000

July 2000

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

MONASH UNIVERSITY

AUSTRALIA

Valid Bayesian Estimation of the Cointegrating Error Correction Model.*

Rodney W. Strachan[†]

Department of Econometrics and Business Statistics,

Monash University,

Clayton, Vic., 3168,

Australia

email: Rodney.Strachan@BusEco.monash.edu.au

July 25, 2000

ABSTRACT

Two methods of identifying cointegrating vectors are commonly used linear restrictions and the nonlinear method of Johansen's maximum likelihood procedure. That the linear method can produce invalid estimates while the Johansen approach always produces valid estimates has been recognised in several recent articles. As all Bayesian studies to date have used linear restrictions, this article presents a Bayesian method for obtaining estimates of cointegrating vectors that will always be valid.

Key Words: Identification restrictions; singular value decomposition; error correction model; cointegration; Bayesian analysis.

---

1   Introduction.

A significant advance in multivariate time series methods was the development of the concept of cointegration, first proposed by Granger (1983). It is well known that as cointegrating vectors are not unique, identifying restrictions must be imposed to allow their estimation. Two common methods of imposing such restrictions are by imposing linear restrictions with normalisations on particular coefficients, or using the eigenvalue-eigenvector method of identification first developed by Anderson (1951) for the reduced rank regression model, and used by Johansen (1988, 1991) in the cointegrating error correction model.

The relative strengths of these two methods of identification were recently emphasised by Boswijk (1996) and Luukkonen, Ripatti and Saikkonen (1999). They both make the point that the restrictions used by Johansen do not generally produce interpretable estimates, but they will always be valid. In contrast, imposing linear restrictions has the advantage that it can produce interpretable coefficients, however, the chosen restrictions may be invalid and, in this case, any analysis based on them will be meaningless (Luukkonen, *et al.*, 1999). These authors stress the importance of considering the validity of the restrictions used and they develop procedures for testing the validity of the restrictions imposed.

In their applications these authors present convincing empirical examples of plausible normalisations that were in fact found to be invalid. Using data for Finland, Luukkonen *et al.* show the own yield of broad money does not cointegrate with one and three month money market rates and the five year bond rate. Boswijk shows for a money demand relation for the UK, that real money and income may only enter a cointegrating relation together as the velocity of money which implies unit-income elasticity of money demand. The evidence from these studies is that the normalisation matters and an invalid normalisation will suggest the presence of more stochastic trends than actually exist (Boswijk, 1996). This occurs as the space spanned by the cointegrating vectors is excluded from the space spanned by the normalised estimates of the cointegrating vectors. These problems provide strong arguments for using the Johansen approach rather than the linear restrictions when we wish to avoid having to choose a coefficient on which to normalise.

The above work has been developed using the classical approach. Bayesian work lags classical efforts in this area in that it has, to date, relied solely on the linear restrictions to achieve identification and there has been little

consideration in the Bayesian literature of the validity of the identification restrictions and normalisations used. Examples of studies using linear restrictions include Kleibergen and van Dijk (1994), Bauwens and Lubrano (1996), Geweke (1996) and Kleibergen and Paap (1998). These papers focus upon estimation, determining the number of stochastic trends in a system, and the implications of just identifying and over identifying restrictions for the posterior. All of this work assumes that the chosen linear restrictions are valid. With the possible exception of Geweke (1996) who suggests a procedure for determining the most likely set of restrictions based upon posterior probabilities, none of this work carefully considers procedures for testing the validity of such restrictions or methods of estimating cointegrating vectors that will always be valid.

This article partly addresses this imbalance between the classical and Bayesian literature by presenting a procedure for Bayesian estimation of cointegrating vectors that will always produce valid estimates. This procedure uses a method equivalent to the identification method of Johansen. It allows for inference regarding the rank, estimation of both unrestricted and restricted cointegrating vectors, as well as tests of restrictions on the cointegrating vectors. As interesting hypotheses about cointegrating vectors can be specified as linear restrictions, I specify restrictions of the form suggested by Johansen and Juselius (1992) for testing of the space spanned by the cointegrating vectors. Using the Bayesian approach has some advantages over the classical approach in that exact finite sample inference is possible, and this inference is obtained from the posterior which uses all of the information in the likelihood, rather than the conditional likelihood. Using this approach we are able to obtain unconditional probabilities for the rank and need not select the rank based on a sequence of possibly conflicting tests.

The article is organized as follows. Section 2 introduces the model, the alternative forms of identifying restrictions and some explanation of the implications of these alternatives. The aim of this section is to present some implications of using the Johansen restrictions and to contrast these with the linear restrictions. A method for implementing the Johansen restrictions is presented in Section 3 along with the specification of the linear restrictions upon the cointegrating vectors. We use the method of Kleibergen and van Dijk (1998) of embedding the reduced rank cointegrating model within a full rank model. Section 4 contains an empirical application and Section 5 concludes.

2   The Model.

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

$$
\begin{aligned}
\Delta y_t &= y_{t-1}\Pi + \Delta y_{t-1}\Gamma_1 + \ldots + \Delta y_{t-l}\Gamma_l + \mu + t\delta + \varepsilon_t \qquad (1) \\
&= y_{t-1}\Pi + x_t\Phi + \varepsilon_t
\end{aligned}
$$

where $\Delta y_t = y_t - y_{t-1}$, $x_t = (\Delta y_{t-1}, \ldots, \Delta y_{t-l}, 1, t)$ and $\Phi = (\Gamma_1', \ldots, \Gamma_l', \mu', \delta')'$. For later use we define the vector $\pi = vec(\Pi)$.

Of interest in this paper is the coefficient matrix Π which is of $rank(\Pi) = r \leq p$, while we assume Γ_i, $i = 1, \ldots, l$ and Φ are full rank. When Π has reduced rank this implies y_t is cointegrated and Π can be expressed as $\Pi = \beta\alpha$ where β and α' are $p \times r$ and assumed to have rank r. Then β is the matrix of cointegration coefficients and α is the matrix of factor loading coefficients or adjustment coefficients. The $p \times (p - r)$ matrices $\beta_\perp$ and $\alpha'_\perp$ are orthogonal to β and α' respectively such that $\beta'\beta_\perp = 0$ and $\alpha_\perp\alpha' = 0$. We assume that

$$
\left| (1-z)\left(I_p - \Sigma_{i=1}^l z^i\Gamma_i\right) - z\beta\alpha \right| = 0
$$

implies $z = 1$ or $z > 1$, and we further assume

$$
\left| \beta'_\perp \left(I_p - \Sigma_{i=1}^l \Gamma_i\right)\alpha'_\perp \right| \neq 0,
$$

such that each element of y_t is either $I(1)$ or $I(0)$ (Johansen, 1991).

To identify the elements of β and α, we need to apply at least r^2 restrictions. A common method of imposing identifying restrictions is to apply r^2 linear restrictions on β by assuming the matrix $c'\beta$ is nonsingular for some known $p \times r$ full column rank matrix c. The general specification of these restrictions and the resultant unique normalised cointegrating vectors, β^*, is

$$
\begin{aligned}
c'\beta\,(c'\beta)^{-1} &= I_r \quad \text{and} \qquad\qquad (2) \\
\beta^* &= \beta\,(c'\beta)^{-1}.
\end{aligned}
$$

This approach takes advantage of the fact that there will be r linearly independent rows of β, as selected by c, but makes the assumption that we know which rows these will be so that we can specify c. Consider the partition

$$
\beta = \left[\begin{array}{c} \beta_1 \\ \beta_2 \end{array} \right]
$$

where β_1 is the first r rows of β. A common example of the linear method in (2) is the zero-one restrictions on β which result from specifying

$$c' = [I_r \quad 0] \tag{3}$$

such that

$$\beta^* = \beta \beta_1^{-1} = \begin{bmatrix} I_r \\ b^* \end{bmatrix} \tag{4}$$

where b^* is a $(p-r) \times r$ matrix to be estimated. This method assumes the set of the first r variables cointegrate (or are $I(0)$) (Johansen 1995 p.94), or, in other words, that the rows of β_1 are linearly independent, thus imposing an order on the variables. However, we may wish to investigate precisely this question of which variables enter the cointegrating relations and how they enter, or at least we may wish to avoid making such strong assumptions.

Before we specify the restrictions used in this paper, we need to define the matrices S_{ij} for $ij = 0, 1, 2$. These are moment matrices when we use a diffuse prior and are analogous to moment matrices when we use an informative prior. The S_{ij} for the diffuse prior are defined as follows. Let

$$\kappa = T + (l+3)p + 1,$$
$$z_{0,t} = \Delta y_t, \qquad z_{1,t} = y_{t-1}, \quad \text{and} \quad z_{2,t} = x_t.$$

Then

$$\kappa M_{ij} = \Sigma_{t=1}^{T} z'_{i,t} z_{j,t}, \qquad \text{for } (i,j) = (1,1), (2,1), (2,2),$$

$$\kappa M_{20} = \Sigma_{t=1}^{T} z'_{2,t} z_{0,t}, \qquad \kappa M_{10} = \Sigma_{t=1}^{T} z'_{1,t} z_{0,t},$$

$$\kappa M_{00} = \Sigma_{t=1}^{T} z'_{0,t} z_{0,t}, \quad \text{and so}$$

$$S_{ij} = M_{ij} - M_{i2} M_{22}^{-1} M_{2j}, \qquad \text{for } ij = 0, 1, 2,$$

except when $i = j = 2$ where

$$S_{22} = M_{22} - M_{21} M_{11}^{-1} M_{12}.$$

The identifying restrictions used in this paper are all imposed on β in the following normalisations,

$$\beta' S_{11} \beta = I \tag{5}$$

which implies $\frac{r(r+1)}{2}$ restrictions, and

$$\beta' S_{10} S_{00}^{-1} S_{01} \beta = \Lambda = diag(\gamma_1, ..., \gamma_r) \tag{6}$$

which implies a further $\frac{r(r-1)}{2}$ restrictions for a total of r^2 restrictions and $(p-r)r$ free parameters in β. With an assumption about the sign of the first element in each row of β, (5) and (6) define r^2 elements of β as deterministic functions of the remaining $(p-r)r$ elements in that matrix. Thus, we may write

$$\beta_1 = f(\beta_2), \tag{7}$$

where $f(\cdot)$ is a deterministic function.

In the expression (6), Λ is random with an *implied* posterior distribution in the Bayesian method and $\gamma_1 > ... > \gamma_r > 0$. As this set of restrictions does not impose an order on the variables, we will call this the *nonordinal* method.

Validity of the normalisation used is not the only implication of the chosen method of identification. Bauwens and Lubrano (1996) show that the posterior distribution of the cointegrating coefficients subject to just identifying linear restrictions has a Cauchy form and no moments. For a further explanation of this result see Villani (2000). In contrast, using the nonordinal identification method the posterior for these coefficients has bounded support and all moments are defined (Strachan, 2000). This difference in the form of the posteriors and the ability to produce valid cointegrating vectors results from how the two methods restrict the vectors for identification. Essentially, the linear method restricts *the length of one coefficient* by tying it to an axis and so excludes directions orthogonal to the axes to which we have tied our coefficients. However, the nonordinal method achieves identification by restricting *the length of the cointegrating vector,* allowing all possible directions to be obtained.

Figures (1) and (2) demonstrate these restrictions for the cointegrating vector from a bivariate system. The solid lines represent the cointegrating vector in R^2 and the dashed lines represent the identifying restrictions. The figures illustrate how the support for the coefficients is bounded in one case but not for the other. Using the linear restrictions we assume $\beta_1 = 1$ and $\beta_2 \in R^1$. Using the nonordinal restrictions we assume $\beta_1 \in [0,1]$ and $\beta_2 \in [-1,1]$. The bounds for β_2 in the Figure 2 were arbitrarily set at 1. In reality these bounds depend upon the data and the boundary represented by the dashed line will generally form an elipse (Strachan, 2000).
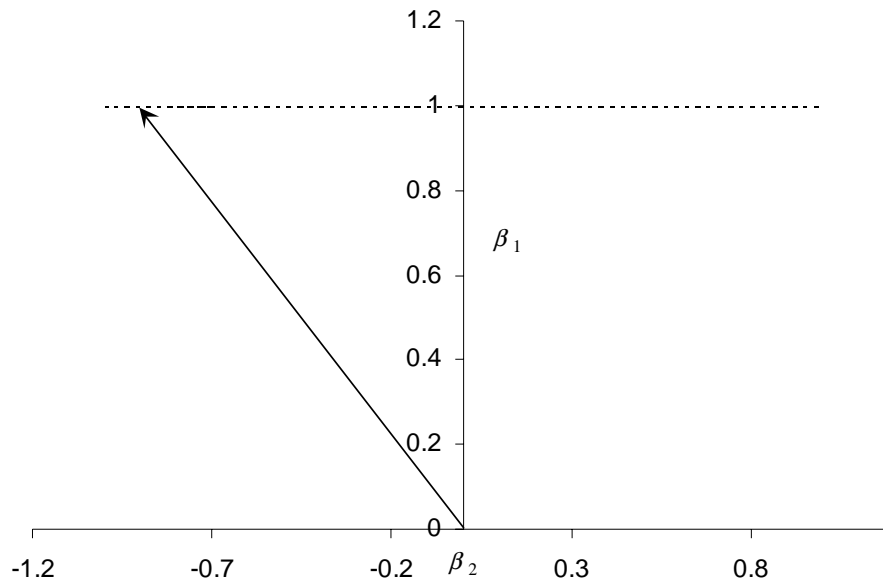
Figure 1: Representation of the linear identifying restrictions in a bivariate cointegrating system. The dashed line represents the restriction and the solid arrow represents a possible vector. The solid arrow must always reach the dashed line.

It is well known that any linear combination of a set of cointegrating vectors constitutes a valid cointegrating vector. This implies that the length of any cointegrating vector is of no real interest, since we can post multiply β by a full rank diagonal matrix to produce whatever length for each vector we desire. What is of interest is the *direction* or the angle of the vector.

Selecting a dependent variable for a cointegrating relation and normalising on the coefficient for this variable is not necessary and may in fact complicate estimation even when such a normalisation is valid. Under this normalisation, the expectation of the quantity of interest may not exist due to the Cauchy form of the posterior. For example, consider a system with the endogenous variables m, the logarithm of the $M3$ measure of money, and p, the logarithm of the level of prices, P and possibly some other variables. Suppose this system contains these variables only in the term
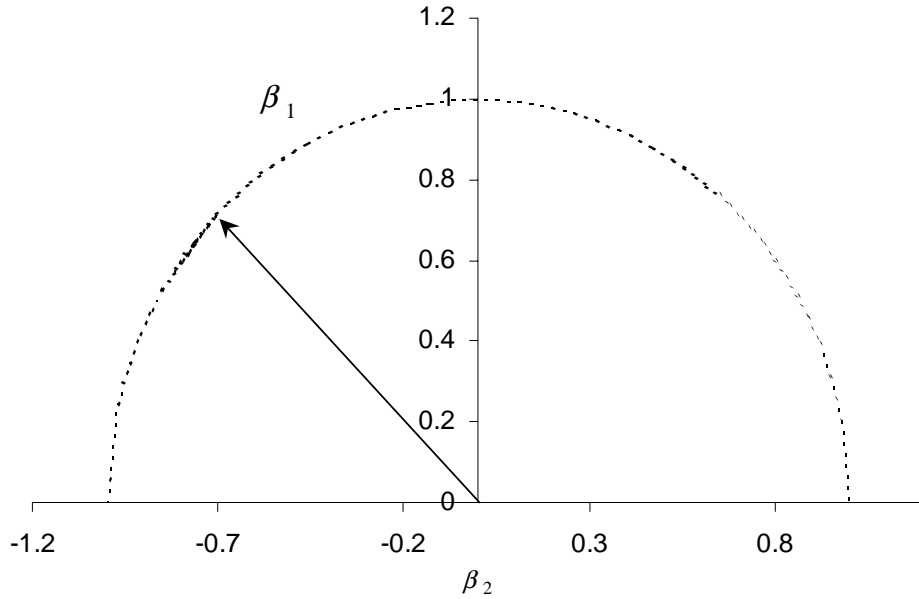
$$\alpha_0 m + \alpha_1 p$$

7

Figure 2: Representation of the nonordinal identifying restrictions in a bivariate cointegrating system. The dashed line represents the restriction and the solid arrow represents a possible vector. The solid arrow must always reach the dashed line.

and we are interested in the prices elasticity of money given as

$$\eta = \frac{dm}{dp} = \frac{dM3}{dP}\frac{P}{M}.$$

If we normalise before estimation by

$$m + \frac{\alpha_1}{\alpha_0}p = m + \eta p,$$

the posterior for η may have no moments and so we need to resort to the posterior mode or median as estimators. For a simple univariate problem these may be relatively straightforward to obtain, however, they tend not to be estimated as well as the mean of a distribution using current methods. Further, if the parameter of interest is a vector, posterior modes and medians become more difficult to estimate accurately relative to the mean.

8

As the quantity of interest is the ratio of two coefficients in a system, then this quantity can alternatively be retrieved from means of the two coefficients. We could use the normalisation given in (5) and (6) so that the posterior means $\widetilde{\alpha}_0 = E\left(\alpha_0|y\right)$ and $\widetilde{\alpha}_1 = E\left(\alpha_1|y\right)$ exist, and report the quantity of interest as

$$\eta = \frac{\widetilde{\alpha}_1}{\widetilde{\alpha}_0}.$$

If we are interested in the support for some restriction such as

$$\alpha_1 = \kappa\alpha_0,$$

with known κ, so that

$$\alpha_0 m + \alpha_1 p = \alpha_0\left(m + \kappa p\right),$$

this constraint can be tested and imposed within (5) and (6), and the posterior for α_0 will have all its moments defined.

3   Identification.

The method in this paper uses Kleibergen and van Dijk's (1998) approach of transforming from a general model to a restricted model to overcome the problem of local nonidentification even when a uniform prior is used. That is, the reduced rank model is nested within a more general (full rank) model which has a well behaved posterior distribution with a reasonable prior, and then the model is transformed such as to parameterize the restriction to a lower rank, effectively conditioning on the given rank r. Beginning with the ECM in (1) with posterior $p\left(\omega|y\right),$ we use the transformation

$$\Pi = \beta\alpha + S_{11}^{-1}\beta_\perp\lambda\alpha_\perp\widetilde{\Sigma} \tag{8}$$

where $\widetilde{\Sigma} = S_{00} - S_{01}S_{11}^{-1}S_{10}$. This gives us the intermediate transformed, but unrestricted model

$$\Delta y_t = y_{t-1}\beta\alpha + y_{t-1}S_{11}^{-1}\beta_\perp\lambda\alpha_\perp\widetilde{\Sigma} + x_t\Phi + \varepsilon_t \tag{9}$$

with posterior for

$$\theta = \left(vec\left(\beta\right)', vec\left(\alpha\right)', vec\left(\lambda\right)'\right)'$$

9

given by
$$p(\theta|y) = p(\pi(\theta)|y)|J| \tag{10}$$
where and $|J|$ is the Jacobian for the transformation from π to θ. From this section on we ignore the parameters Σ and Φ as they have standard, well known conditional and marginal posteriors. The cointegrating ECM occurs at $\lambda = 0$,
$$\Delta y_t = y_{t-1}\beta\alpha + x_t\Phi + \varepsilon_t \tag{11}$$
and we obtain the posterior for
$$\theta_0 = \left(vec(\beta)', vec(\alpha)'\right)'$$
as
$$p(\theta_0|y) = p(\theta|y)|_{\lambda=0}. \tag{12}$$

To present the implementation of the transformation in (8) we need the following definition. Let A be a positive definite matrix with singular value decomposition $A = USV'$ such that $\mathbf{S}$ is a diagonal matrix with positive diagonal terms. Let $\mathbf{S}^{\frac{1}{2}}$ be the matrix $\mathbf{S}$ with each of the diagonal elements replaced its square root. We define the square root of A as $A^{\frac{1}{2}} = U\mathbf{S}^{\frac{1}{2}}V'$ such that $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$ and $A^{-\frac{1}{2}}AA^{-\frac{1}{2}} = I$.

The decomposition in (8) is equivalent to taking an SVD of Π with respect to its estimated variance-covariance matrix. This approach is similar to the approach Johansen (1988, 1991) takes in the maximum likelihood estimator. The posterior estimate of the variance-covariance matrix for Π is $\widetilde{\Sigma} \otimes \kappa^{-1}S_{11}^{-1}$. Alternatively we could use Σ instead of $\widetilde{\Sigma}$ in (8), however this results in a far more complex Jacobian and contributes nothing to the method[1]. Define Π^* by
$$\Pi = S_{11}^{-\frac{1}{2}}\Pi^*\widetilde{\Sigma}^{\frac{1}{2}}. \tag{13}$$
The SVD proceeds as follows. Let $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ and $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ be appropriate orthonormal matrices where U_1, V_1 are $p \times r$, U_2, V_2 are $p \times (p-r)$ and $\underline{S}_1$ and $\underline{S}_2$ are diagonal $r \times r$ and $(p-r) \times (p-r)$ respectively. Make the following transformation:
$$\Pi^* = U\underline{S}V' = \begin{bmatrix} U_1 & U_2 \end{bmatrix}\begin{bmatrix} \underline{S}_1 & 0 \\ 0 & \underline{S}_2 \end{bmatrix}\begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \tag{14}$$

---

[1]The author is grateful to an anonymous referee for comments which lead to this much simpler specification.

$$= \begin{bmatrix} U_1\overline{T} & U_2 \end{bmatrix} \begin{bmatrix} \overline{T}'\underline{S}_1 & 0 \\ 0 & \underline{S}_2 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix}$$

$$= U_1\underline{S}_1V_1' + U_2\underline{S}_2V_2'$$

$$= U_1\overline{T}\,\overline{T}'\underline{S}_1V_1' + U_2T_1T_1'\underline{S}_2T_2T_2'V_2'$$

where we have the $r \times r$ orthogonal matrix $\overline{T}$, and the $(p-r) \times (p-r)$ orthogonal matrices T_1 and T_2. The matrices $\overline{T}$, T_1 and T_2 contain the eigenvectors of the (random) matrices $U_1'S_{11}^{-\frac{1}{2}}S_{10}S_{00}^{-1}S_{01}S_{11}^{-\frac{1}{2}}U_1$, $U_2'S_{11}^{-1}U_2$ and $V_2'\widetilde{\Sigma}V_2$ respectively, such that

$$\overline{T}'U_1'S_{11}^{-\frac{1}{2}}S_{10}S_{00}^{-1}S_{01}S_{11}^{-\frac{1}{2}}U_1\overline{T} = \Lambda = diag\,(\gamma_1, ..., \gamma_r)$$

$$T_1'U_2'S_{11}^{-1}U_2T_1 = \Lambda^1 = diag\left(\eta_1, ..., \eta_{p-r}\right)$$

$$T_2'V_2'\widetilde{\Sigma}V_2T_2 = \Lambda^2 = diag\,(\varsigma_1, ..., \varsigma_{p-r})\,.$$

Denote the eigenvalues of the $p \times p$ matrix $S_{11}^{-\frac{1}{2}}S_{10}S_{00}^{-1}S_{01}S_{11}^{-\frac{1}{2}}$, by $(\nu_1, ..., \nu_p)$ where $\nu_1 \geq ... \geq \nu_p \geq 0$, although in practice $\nu_1 > ... > \nu_p > 0$. These eigenvalues are also the squared canonical correlation coefficients. The diagonal elements of Λ, $(\gamma_1, ..., \gamma_r)$, are bounded by $\nu_i \geq \gamma_i \geq \nu_{p-r+i}$ (Schott 1997, p.111), and at the mode of the posterior, $\nu_i = \gamma_i$.

Using the above transformation we obtain the following expressions for β and α,

$$\beta = S_{11}^{-\frac{1}{2}}U_1\overline{T} = S_{11}^{-\frac{1}{2}}\Gamma_1 \tag{15}$$

$$\alpha = \overline{T}'\underline{S}_1V_1'\widetilde{\Sigma}^{\frac{1}{2}}. \tag{16}$$

From the last term in (14) we define the $p \times (p-r)$ matrices $\beta_\perp$ and $\alpha_\perp'$ by

$$\beta_\perp = S_{11}^{\frac{1}{2}}U_2T_1,$$

$$\alpha_\perp = T_2'V_2'\widetilde{\Sigma}^{-\frac{1}{2}},$$

and finally

$$\lambda_r = T_1'\underline{S}_2T_2. \tag{17}$$

Thus we obtain the transformation in (8).

3.1 The Jacobian for the transformation.

The restrictions for the transformation (8) are expressed in the r^2 equations

$$Lvec(\beta' S_{11}\beta - I) = 0 \tag{18}$$

and

$$\bar{L}vec(\beta' D\beta - \Lambda) = 0 \tag{19}$$

where $Lvec(A)$ is the $\frac{r(r+1)}{2}$ vector of lower triangular elements of the $r \times r$ matrix A, and $\bar{L}vec(A)$ is the $\frac{r(r-1)}{2}$ vector of infradiagonal elements of A (Magnus and Neudecker, 1980, Henderson and Searle, 1979).

For the purpose of finding an expression for the Jacobian, let β_2 contain the free parameters such that given β_2 and the data, β_1 is known through the function (7). The Jacobian matrix can be expressed as

$$
\begin{aligned}
J\left(\Pi : (\beta_2, \alpha, \lambda)\right) &= \frac{\partial vec\left(\Pi\right)}{\partial vec\left(\beta_2, \alpha, \lambda\right)'} \\
&= \left[\frac{\partial vec\left(\Pi\right)}{\partial vec\left(\beta_2\right)'} \quad \frac{\partial vec\left(\Pi\right)}{\partial vec\left(\alpha\right)'} \quad \frac{\partial vec\left(\Pi\right)}{\partial vec\left(\lambda\right)'} \right].
\end{aligned}
$$

The forms of these expressions are presented in Appendix I. Where $\lambda = 0$ the Jacobian has the form:

$$
\begin{aligned}
\frac{\partial vec\left(\Pi\right)}{\left(\partial vec\left(\alpha\right)\right)'} &= \left(I_L \otimes \beta\right) \\
\frac{\partial vec\left(\Pi\right)}{\left(\partial vec\left(\beta_2\right)\right)'} &= \left[\left[\alpha' \otimes \begin{pmatrix} 0_{r\times(p-r)} \\ I_{p-r} \end{pmatrix} \right] + \left[\alpha' \otimes \begin{pmatrix} I_r \\ 0_{(p-r)\times r} \end{pmatrix} \right] \ddot{E} \right] \\
\frac{\partial vec\left(\Pi\right)}{\left(\partial vec\left(\lambda\right)\right)'} &= \left(\tilde{\Sigma} \alpha'_\perp \otimes S_{11}^{-1} \beta_\perp \right)
\end{aligned}
$$

3.2 Sampling scheme.

The aim of this section is to outline a procedure for obtaining a draw from the posterior $p\left(\theta_0|y\right)$. The forms of the joint and conditional posteriors of α, λ, and β are not from known classes of probability density functions. To sample from these posteriors we use the same approach as detailed in Kleibergen and Paap (1998). That is, a Metropolis-Hasting algorithm is used with the candidate density being that for the full rank model, $p\left(\theta|y\right)$.

This sampling scheme is outlined here. First we use the following steps to draw from the candidate density. At iteration i, for $i = 1, \ldots, m$,

>**Step 1:** Draw $\Pi^{(i)}$ from $p(\Pi|y)$.
>
>**Step 2:** Perform SVD of $\Pi^{(i)} = U^{(i)} S^{(i)} V^{(i)\prime}$.
>
>**Step 3:** Compute $\alpha^{(i)}$, $\lambda^{(i)}$ and $\beta^{(i)}$ using (15) and (17).

As we have sampled λ which only appears in (10), to get draws from (12), we augment the posterior for the reduced rank model, $p(\theta_0|y)$, with a proper distribution for λ, $g(\lambda|\theta_0, y)$. Again, the reader is directed to Kleibergen and Paap (1998) for details of this method. The weights used to accept or reject draws are

$$
\begin{aligned}
w_i &= w(\theta|y) && (20) \\
&= \frac{g\left(\lambda^{(i)}|\theta_0, y\right) p\left(\theta_0^{(i)}|y\right)}{p\left(\theta^{(i)}|y\right)}.
\end{aligned}
$$

As discussed in Kleibergen and Paap (1998), Geweke (1989) shows $m^{-1} \sum_{i=1}^{m} w_i$ converges to the ratio of the integrals

$$
\frac{\int g(\lambda|\theta_0, y) p(\theta_0|y) \, d\theta}{\int p(\theta|y) \, d\theta}. \qquad (21)
$$

This result is useful for estimating Bayes factors. The steps in the Metropolis-Hastings sampler are:

>**Step 1:** Draw $\theta^{*(i+1)}$ from $p(\theta|y)$.
>
>**Step 2:** Accept $\theta^{(i+1)} = \theta^{*(i+1)}$ with probability $min\left(\frac{w_{i+1}}{w_i}, 1\right)$, otherwise $\theta^{(i+1)} = \theta^{(i)}$.

The resultant set $\theta^{(i+1)}$ will be a draw from the posterior

$$
p(\lambda, \theta_0|y) = g(\lambda|\theta_0, y) p(\theta_0|y),
$$

and the set $\theta_0^{(i+1)}$ will be a draw from $p(\theta_0|y)$.

3.3   Linear restrictions on β.

In the Introduction it was argued that linear restrictions on β will not always be appropriate as $c'\beta$ may be singular depending on the order of the variables in y_t. Therefore, linear restrictions force the researcher to declare which r rows

of β_1 are linearly independent which may involve deciding which variables *must* enter the relationship $y_t\beta$. This raises the question: What is the form of β? Or, in other words, what valid linear restrictions can be imposed on β?

In both the classical and Bayesian approaches, to test the appropriateness of such restrictions and to estimate the restricted model, requires a specification of the model subject to these restrictions. In the classical maximum likelihood approach, Johansen and Juselius (1992) has provided methods for estimation with, and testing of, these restrictions. In this section we present the Bayesian equivalent with the SVDs for the models with various linear restrictions on β. The three restrictions investigated are presented as the following hypotheses.

(R_1) $H_0 : \beta = H\psi$

where the dimensions of the respective matrices are: H $p \times s$, ψ $s \times r$, $r \leq s$.

(R_2) $H_0 : \beta = (b\ \varphi) = (b\ b_\perp\psi)$

where the dimensions of the respective matrices are: b $p \times s$, $b_\perp$ $p \times (p-s)$, ψ $(p-s) \times (r-s)$, $s \leq r$.

(R_3) $H_0 : \beta = (H_1\psi_1, H_2\psi_2, \ldots, H_l\psi_l)$

where the dimensions of the respective matrices are: H_i $p \times s_i$, ψ_i $s_i \times r_i$, $r_i \leq s_i$, $l \leq r$, $\sum_i r_i = r$.

The restriction in R_1 imposes the same restriction on all of the columns in β. The second restriction, R_2, assumes we know the form of the first s columns of β, b, and the remaining $r - s$ columns, $b_\perp\psi$, are unknown except that they are orthogonal to b. The final hypothesis, R_3, generalizes the first two.

The SVD for these hypotheses can be found by using the results in Appendix II which presents the SVD for a nonsquare matrix, φ, and the following transformations.

For R_1, the model becomes

$$
\begin{aligned}
\Delta y_t &= y_{t-1}\Pi + x_t\Phi + \varepsilon_t \\
&= y_{t-1}H\varphi + x_t\Phi + \varepsilon_t \\
&= y^a{}_{t-1}\varphi + x_t\Phi + \varepsilon_t
\end{aligned}
$$

where

$$
\varphi = \psi\alpha + \left(H'S_{11}H\right)^{-1}\psi_\perp\lambda\alpha_\perp\widetilde{\Sigma}.
$$

The Jacobian for this transformation, J^{R_1}, is of the same form as the Jacobian in Appendix I with S_{11}, p, β, and Π replaced by $H'S_{11}H$, s, ψ, and φ

respectively. The definition of S_{11} remains unchanged in this case.

For R_2, the model becomes

$$
\begin{aligned}
\Delta y_t &= y_{t-1}\Pi + x_t\Phi + \varepsilon_t \\
&= y_{t-1}b\alpha_1 + y_{t-1}b_\perp\varphi + x_t\Phi + \varepsilon_t \\
&= y_{t-1}b_\perp\varphi + x^a{}_t\Phi^a + \varepsilon_t \\
&= y^a{}_{t-1}\varphi + x^a{}_t\Phi^a + \varepsilon_t
\end{aligned}
$$

where $x^a{}_t = [y_{t-1}b \ \ x_t]$, $\Phi^a = [\alpha_1' \ \ \Phi']'$ and

$$
\varphi = \psi\alpha_2 + (b'_\perp S_{11}b_\perp)^{-1}\psi_\perp\lambda\alpha_{2\perp}\widetilde{\Sigma}.
$$

Again the Jacobian, J^{R_2}, can be found by using the form of the Jacobian in Appendix I with S_{11}, β, Π, x_t and α replaced by $b'_\perp S_{11}b_\perp$, ψ, φ, $x^a{}_t$, and α_2 respectively.

Finally, for R_3, the model becomes

$$
\begin{aligned}
\Delta y_t &= y_{t-1}\Pi + x_t\Phi + \varepsilon_t \\
&= y_{t-1}H_1\varphi_1 + y_{t-1}H_2\varphi_2 + \cdots + y_{t-1}H_l\varphi_l \\
&\quad + x_t\Phi + \varepsilon_t \\
&= y^a{}_{t-1,l}\varphi_l + x^a{}_t\Phi^a + \varepsilon_t
\end{aligned}
$$

where

$$
\varphi_i = \psi_i\alpha_i + (H_i'S_{11i}H_i)^{-1}\psi_{i\perp}\lambda_i\alpha_{i\perp}\widetilde{\Sigma},
$$

$i = 1, \ldots, l$, $l \leq r$. Each of the l Jacobians, $J_i^{R_3}$, can be found by using the form of the Jacobian in Appendix I with S_{11}, β, p, Π, α, x_t, and λ replaced by $H_i'S_{11i}H_i$, ψ_i, $\Sigma_{i=1}^l s_i$, φ_i, α_i, $x^a{}_{it}$, and λ_i respectively where $x^a{}_{it} = (y_{t-1}H_1 \cdots y_{t-1}H_l \ x_t)$ does not include $y^a{}_{t-1,i} = y_{t-1}H_i$. In the construction of S_{11i}, we use the same definition as for S_{11}, except we replace X with $X^a{}_i = (x^{a'}_{i1} \ x^{a'}_{i2} \cdots x^{a'}_{iT})'$.

Estimation of these models requires little extra computer coding beyond that required for the general model using the transformation (8). That is, we can represent the process of passing $\left(S_{11}, \Pi, \widetilde{\Sigma}\right)$ to a procedure which returns (α, β, λ) by $(\alpha, \beta, \lambda) = proc1\left(S_{11}, \Pi, \widetilde{\Sigma}\right)$ and we get the Jacobian by $J = proc2\left(S_{11}, x_t, \alpha, \beta, \lambda, \widetilde{\Sigma}\right)$. Then for the above restrictions we would have

$$
(\alpha, \psi, \lambda) = proc1\left(H'S_{11}H, \varphi, \widetilde{\Sigma}\right), \ J^{R_1} = proc2\left(H'S_{11}H, x_t, \alpha, \psi, \lambda, \widetilde{\Sigma}\right) \ \text{for} R_1,
$$

$$(\alpha_2, \psi, \lambda) = proc1\left(b'_\perp S_{11} b_\perp, \varphi, \widetilde{\Sigma}\right), \ J^{R_2} = proc2\left(b'_\perp S_{11} b_\perp, x^a{}_t, \alpha_2, \psi, \lambda, \widetilde{\Sigma}\right) \ \text{for} \ R_2,$$

$$\text{and} \ (\alpha_i, \psi_i, \lambda) = proc1\left(H'_i S_{11i} H_i, \varphi_i, \widetilde{\Sigma}\right),$$
$$J_i^{R_3} = proc2\left(H'_i S_{11i} H_i, x^a{}_{it}, \alpha_i, \psi_i, \lambda_i, \widetilde{\Sigma}\right) \ \text{for each} \ i \ \text{for} \ R_3.$$

4   Application.

In this section I demonstrate estimation of a cointegration vector for four Australian interest rates using the nonordinal estimation method. I test the validity of particular linear restrictions using the form R_1 of the hypothesis specified by Johansen and Juselius (1992). The four interest rates are the 5 year (i_5) and 3 year (i_3) Treasury Bond rates (Capital Market) and the 180 day (i_{180}) and 90 (i_{90}) day Bank Accepted Bill (Money Market) rates and so

$$y = (i_5, i_3, i_{180}, i_{90}).$$

The time subscripts have been dropped for notational convenience. The data are annualised monthly rates for the period July 1992 to April 2000 ($T = 94$). A graph of the data is presented in Figure 3.

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 the spreads of these rates to be stationary, and we would expect to find, in this case, three cointegrating relations. Thus the spreads should either form the cointegrating relations or the rates may enter the cointegrating relations via the spreads.

Classical pretesting suggests each series is integrated of order one and we find an ECM with one lag of differences is sufficient to model the process. All classical testing is at the 1% level of significance. Only the 180 day rates show slight evidence of autocorrelation in the squared residuals (the p-value for the LM test is 0.027), however, following Luukkonen *et al.,* we ignore this feature as modelling this behaviour is outside the scope of this paper. For comparison with the classical results, in the Bayesian estimation we use the diffuse prior for the parameters in (1) used by Kleibergen and van Dijk (1994). We use an informative prior to provide interpretable posterior probabilities.

Bayesian posterior probabilities and classical likelihood ratio tests for the rank are presented in Table 1. Both the Bayesian and classical results suggest a rank of one which implies the system contains three stochastic trends, although at the 5% level the classical tests would accept a rank of zero. The finding of a rank of one conflicts with the idea that the interest rates will share a common stochastic trend, however, further analysis provides an explanation for this result. We accept the result that there is one cointegrating vector and the resultant classical and Bayesian estimates of the error correction term are, respectively,

$$
\begin{aligned}
y\beta &= \varphi_1 i_5 + \varphi_2 i_3 + \varphi_3 i_{180} + \varphi_4 i_{90} \\
&\quad \text{Classical} \\
&= -0.382 i_5 + 0.392 i_3 + 0.773 i_{180} - 0.780 i_{90} \\
&\quad \text{Bayesian (diffuse)} \\
&= 0.460 i_5 - 0.506 i_3 - 5.696 i_{180} + 5.643 i_{90} \\
&\quad \text{Bayesian (informative)} \\
&= 0.064 i_5 + 0.065 i_3 - 0.202 i_{180} + 0.014 i_{90}.
\end{aligned}
$$

Normalising on φ_4 we obtain

$$
\begin{aligned}
y\beta/\varphi_4 &= \varphi_1/\varphi_4 i_5 + \varphi_2/\varphi_4 i_3 + \varphi_3/\varphi_4 i_{180} + i_{90} \\
&\quad \text{Classical} \\
&= i_5 - 1.026 i_3 - 2.022 i_{180} + 2.040 i_{90} \\
&\quad \text{Bayesian (diffuse)} \\
&= i_5 - 1.100 i_3 - 12.376 i_{180} + 12.263 i_{90} \\
&\quad \text{Bayesian (informative)} \\
&= i_5 + 1.023 i_3 - 3.159 i_{180} + 0.214 i_{90}.
\end{aligned}
$$

The coefficients on the short rates for the diffuse Bayesian and the classical estimates are much larger than for the long rates, suggesting changes in the short rates have a greater effect upon the error correction term. Also, the estimates suggest we could use the spreads $(i_5 - i_3)$ and $(i_{180} - i_{90})$ as the only variables in the cointegrating vectors since the rates in these spreads appear to have the same coefficients. This grouping would appear sensible as it groups the interest rates into similar markets.

The estimates from the informative prior are quite different from the other two which suggests that this prior has a strong affect on the estimates and is

shifting the posterior away from the values suggested by the data. Further, changing the 'informativeness' of the prior by changing the prior variances did not improve estimation. Therefore, unless using a large sample, the diffuse prior is recommended if the objective is estimation.

If all four rates do enter via the spreads, β will have a space spanned by H_1 where

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

Thus, with φ a 3×1 matrix, our first hypothesis is

$$\mathcal{H}_1 : \beta = H_1\varphi. \tag{22}$$

This implies we should be able to apply a linear normalisation of the form (2) with $c' = (1, 0, 0, 0)$ to obtain estimates of β.

Given a rank of one and the results in Table 2, we would accept this hypothesis with a posterior probability of $P(\mathcal{H}_1|y) = 0.773$ and the classical LR test statistic for $\mathcal{H}_1$ against

$$\mathcal{H}_0 : \beta \text{ unrestricted}$$

of 0.025. The classical and Bayesian estimates of the error correction term subject to $\mathcal{H}_1$ are, respectively,

$$
\begin{aligned}
yH\varphi &= \varphi_1 (i_5 - i_3) + \varphi_2 (i_3 - i_{180}) + \varphi_3 (i_{180} - i_{90}) \\
&\quad \text{Classical} \\
&= 1.932 (i_5 - i_3) + 0.161 (i_3 - i_{180}) - 6.464 (i_{180} - i_{90}) \\
&\quad \text{Bayesian (diffuse)} \\
&= -1.813 (i_5 - i_3) + 0.217 (i_3 - i_{180}) + 5.143 (i_{180} - i_{90}) \\
&\quad \text{Bayesian (informative)} \\
&= 0.080 (i_5 - i_3) + 0.159 (i_3 - i_{180}) + 0.187 (i_{180} - i_{90}).
\end{aligned}
$$

Normalising on φ_3 we obtain

$$
\begin{aligned}
yH\varphi/\varphi_3 &= \varphi_1/\varphi_3 (i_5 - i_3) + \varphi_2/\varphi_3 (i_3 - i_{180}) + (i_{180} - i_{90}) \\
&\quad \text{Classical} \\
&= -0.299 (i_5 - i_3) - 0.002 (i_3 - i_{180}) + (i_{180} - i_{90})
\end{aligned}
$$

$$\text{Bayesian (diffuse)}$$
$$= -0.353\,(i_5 - i_3) + 0.042\,(i_3 - i_{180}) + (i_{180} - i_{90})$$
$$\text{Bayesian (informative)}$$
$$= 0.428\,(i_5 - i_3) + 0.848\,(i_3 - i_{180}) + (i_{180} - i_{90})\,.$$

The informative prior estimate again shows the effect of the prior upon the distribution of the cointegrating coefficients, and we again see the classical and the diffuse prior estimates closely approximate each other.

The support for $\mathcal{H}_1$, which is suggested by the theories of the term structure of interest rates, provides no explanation for the presence of three stochastic trends in the system. Further, we note that the diffuse prior and classical estimates of φ_2 are relatively small and possibly close to zero. Therefore, as a first step to explaining the number of stochastic trends, we next test the hypothesis that the spread between the long and short rates, the capital and money market rates, does not enter the cointegrating relation. This hypothesis is

$$\mathcal{H}_2 : \beta = H_2\varphi \qquad \equiv \qquad \varphi_2 = 0 \tag{23}$$

where

$$H_2 = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

The posterior probability of $\mathcal{H}_2$ and the likelihood ratio test statistic in Table 2 suggest support for $\mathcal{H}_2$. Re-estimating subject to the restriction in $\mathcal{H}_2$ we obtain the estimates

$$yH\varphi = \varphi_1\,(i_5 - i_3) + \varphi_3\,(i_{180} - i_{90})$$
$$\text{Classical}$$
$$= 1.208\,(i_5 - i_3) + 6.968\,(i_{180} - i_{90})$$
$$\text{Bayesian (diffuse)}$$
$$= -1.039\,(i_5 - i_3) + 6.787\,(i_{180} - i_{90})$$
$$\text{Bayesian (informative)}$$
$$= -0.001\,(i_5 - i_3) - 0.108\,(i_{180} - i_{90})\,.$$

Normalising again on φ_3 we obtain

$$yH\varphi/\varphi_3 = \varphi_1/\varphi_3\,(i_5 - i_3) + (i_{180} - i_{90})$$

$$\text{Classical}$$
$$= \quad 0.173\left(i_5 - i_3\right) + \left(i_{180} - i_{90}\right)$$
$$\text{Bayesian (diffuse)}$$
$$= \quad -0.153\left(i_5 - i_3\right) + \left(i_{180} - i_{90}\right)$$
$$\text{Bayesian (informative)}$$
$$= \quad 0.007\left(i_5 - i_3\right) + \left(i_{180} - i_{90}\right).$$

The above estimates suggest an additional restriction which excludes the long rates from the cointegrating relation, and it is this result which would explain the presence of too many stochastic trends in the system. Therefore we test the hypothesis that the long rates do not enter the cointegrating relationship which is implied by the hypothesis

$$\mathcal{H}_3 : \beta = H_3\varphi \qquad \equiv \qquad \varphi_2 = \varphi_3 = 0 \qquad (24)$$

where

$$H_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

$\mathcal{H}_3$ is accepted indicating that the error correction term is simply the spread between the short term rates. A test of cointegration between the long rates using a bivariate ECM confirmed the finding that these rates do not cointegrate.

The finding that the long rates do not enter the cointegrating relation but the short rates do via the spread, explains the presence of three stochastic trends. Further, this result indicates that while it would seem reasonable to identify β using the linear restrictions in (2) with $c' = (1, 0, 0, 0)$, that is by normalising upon φ_1, this would in fact produce an invalid estimate of β since we found i_5 and i_3 are not present in the cointegrating relation and so $\varphi_1 = \varphi_2 = 0$.

Table 1: *LR statistics and posterior probabilities for rank* (r).

$r = rank\,(\Pi)$	$LR\,(r\vert p)$	1% Critical value	$P\,(r\vert y)$
0	59.23	54.46	0.00
1	31.66	35.65	1.00
2	10.26	20.04	0.00
3	2.33	6.65	0.00
4	-	-	0.00
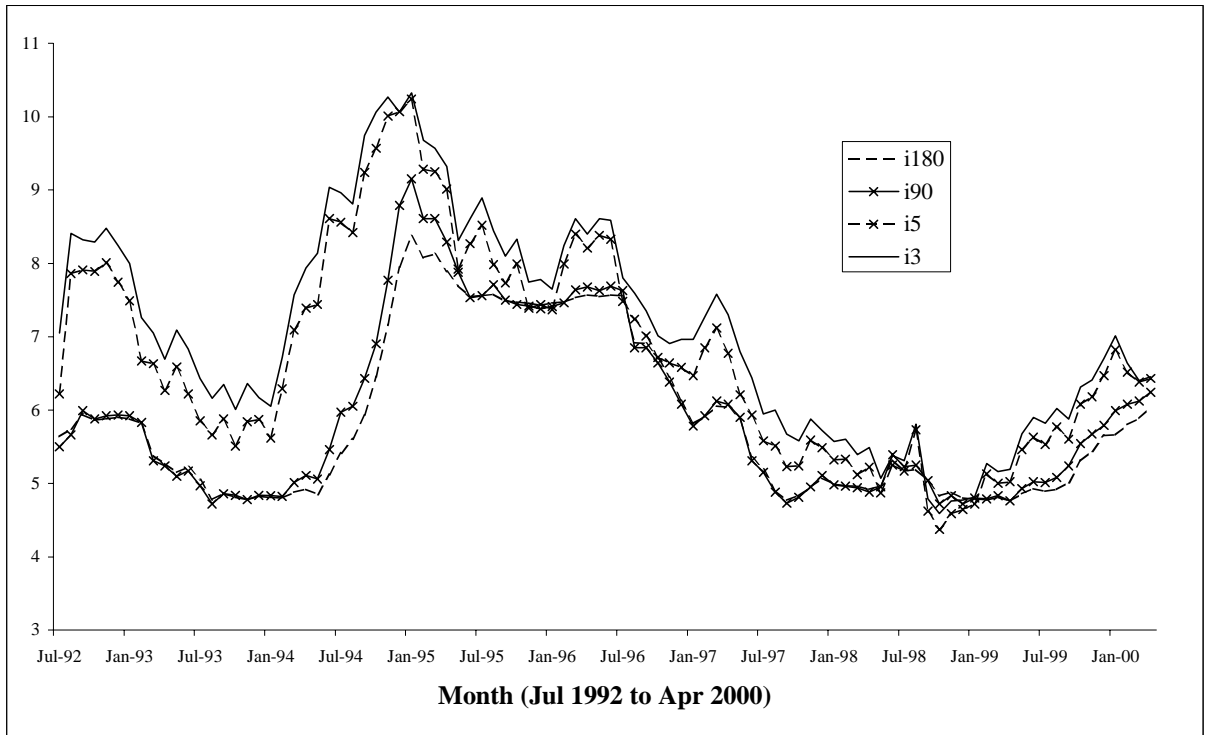
Figure 3: Plot of Australian interest rates. The series are the 5 year (i_5) and 3 year (i_3) Treasury Bond rates (Capital Market), and the 180 day (i_{180}) and 90 (i_{90}) day Bank Accepted Bill (Money Market) rates.

Table 2: *LR statistics and posterior probabilities for hypotheses.*

Hypothesis, $\mathcal{H}_i$	$LR\left(\mathcal{H}_i|\mathcal{H}_0\right)$	1% Critical value	$P\left(H_i|y\right)$
$\mathcal{H}_1$	0.025	6.63	0.773
$\mathcal{H}_2$	4.431	9.21	0.516
$\mathcal{H}_3$	9.328	11.34	0.603

21

5 Conclusion.

In this paper we have demonstrated a method for Bayesian estimation of cointegrating vectors that will always be valid in the sense that it avoids invalid normalisations. This method uses restrictions on the parameters that are equivalent to those used in the classical maximum likelihood estimator for this model developed by Anderson (1951) and applied to cointegration problems by Johansen (1988, 1991). The specification of the model is similar to Kleibergen's approach as it uses a parameterisation of the rank reduction. That is, the potentially reduced rank matrix Π is transformed to (α, β, λ) and the rank of Π is reduced if $\lambda = 0$. We have demonstrated some of the implications of employing linear restrictions for the posterior of the cointegrating coefficients for the estimates and the distribution of the coefficients.

Estimates of coefficients are presented for a simple model and compared with classical methods of inference. In this empirical example we demonstrate a case where a seemingly sensible normalisation would in fact be invalid, justifying the use of the nonordinal method for at least initial estimation.

6   References.

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.

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.

Henderson, H.V. and S.R. Searle (1979) "Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics." *The Canadian Journal of Statistics*, **7**, 65-81.

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.

Johansen, S., and Juselius (1992) "Testing structural hypotheses in a multivariate cointegration analysis of PPP and UIP for UK." *Journal of Econometrics*, **53**, 211-244.

Kleibergen, F. and R. Paap (1998) "Priors, posterior odds and Lagrange multiplier statistics in Bayesian analysis of cointegration." *Econometric Institute Reports*, no. **9821**, Econometric Institute, Erasmus University Rotterdam.

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

Kleibergen, F. and H.K. van Dijk (1998) "Bayesian simultaneous equations analysis using reduced rank structures." *Econometric Theory*, **14**, 701-743.

Luukkonen, R, A. Ripatti and P. Saikkonen (1999) "Testing for a valid normalisation of cointegrating vectors in vector autoregressive processes." *Journal of the American Statistical Association*, **17**, 195-204.

Magnus, J. R. and H. Neudecker (1980) "The elimination matrix: Some lemmas and applications." *Society for Industrial and Applied Mathematics Journal on Discrete and Algebraic Methods*, **4**, 422-449.

Magnus, J. R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons, New York.

Olkin, I. and A.R. Sampson (1972). "Jacobians of matrix transformations and induced functional equations." *Linear Algebra and its Applications* **5**, 257-276.

Roy, S.N. (1952). "Some useful results in Jacobians". *Calcutta Statistical Association Bulletin*, **4**, 117-122.

Schott, J.R. (1997). *Matrix Analysis for Statistics*. John Wiley and Sons, New York.

Strachan, R.W. (2000). *Bayesian Analysis of the Cointegrating Error Correction Model*. PhD Thesis, Monash University.

Villani, M. (2000) *Aspects of Bayesian Cointegration*, PhD thesis, Department of Statistics, Stockholm University.

7   Appendix I: The Jacobian for the Transformation.

From the transformation (8) in Section 2 we have the following equation

$$\Pi = \beta\alpha + S_{11}^{-1}\beta_\perp\lambda\alpha_\perp\widetilde{\Sigma},$$

and the r^2 restrictions equations in (18) and (19)

$$Lvec(\beta'S_{11}\beta - I) = 0$$

and

$$\bar{L}vec(\beta'S_{10}S_{00}^{-1}S_{01}\beta - \Gamma) = 0.$$

24

To find an expression for the Jacobian, use the results of Roy (1952, p. 118) which are reproduced here for convenience.

"**Theorem 1.** If $y_i = f_i(x_1, \ldots, x_m; x_{m+1}, \ldots, x_{m+n})$ $(i = 1, \ldots, m)$ when $x'_j s$ $(j = 1, \ldots, m+n)$ are subject to n constraints

$$f_i(x_1, \ldots, x_m; x_{m+1}, \ldots, x_{m+n}) = 0 \quad (i = m+1, \ldots, m+n),$$

then (under the usual conditions for the existence of the Jacobian, including the non-vanishing of the numerator and the denominator in the following) we have"

$$|J(y_1, \ldots, y_m : x_1, \ldots, x_m)| = \left| \frac{\delta(f_1, \ldots, f_m, f_{m+1}, \ldots, f_{m+n})}{\delta(x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n})} \right| \times \left| \frac{\delta(f_{m+1}, \ldots, f_{m+n})}{\delta(x_{m+1}, \ldots, x_{m+n})} \right|^{-1}.$$

Importantly for our needs, in the proof for the above theorem, Roy presents the Jacobian as

$$|J(y_1, \ldots, y_m : x_1, \ldots, x_m)| = \left| \frac{\delta y_i}{\delta x_j} \right| i, j = 1, \ldots, m \qquad (25)$$

$$= \left| \frac{\delta f_i}{\delta x_j} + \sum_{k=m+1}^{m+n} \frac{\delta f_i}{\delta x_k} \frac{\delta x_k}{\delta x_j} \right| i, j = 1, \ldots, m$$

For the purpose of deriving the Jacobian, we treat $\alpha_\perp$, $\beta_\perp$ and β_1 and as the $x'_k s$, and α, β_2, and λ as the $x'_j s$ in (25). Therefore the form of the Jacobian matrix for (8) is

$$J(\Pi : (\beta_2, \alpha, \lambda)) = \frac{\partial vec(\Pi)}{\partial vec(\beta_2, \alpha, \lambda)'}$$

$$= \left[\frac{\partial vec(\Pi)}{\partial vec(\alpha)'} \quad \frac{\partial vec(\Pi)}{\partial vec(\beta_2)'} \quad \frac{\partial vec(\Pi)}{\partial vec(\lambda)'} \right]$$

where

$$\frac{\partial vec(\Pi)}{(\partial vec(\alpha))'} = (I_L \otimes \beta) + \frac{\partial vec(\Pi)}{(\partial vec(\alpha_\perp))'} \frac{\partial vec(\alpha_\perp)}{(\partial vec(\alpha))'}$$

$$\frac{\partial vec(\Pi)}{(\partial vec(\alpha_\perp))'} = \left(\tilde{\Sigma} \otimes S_{11}^{-1} \beta_\perp \lambda \right).$$

25

Letting $\alpha_\perp = c_\perp' \left[I_L - \alpha' (\alpha c)'^{-1} c' \right]$, where $c = (I_r\ 0)'$ and $c_\perp = (0\ I_{L-r})'$,

$$\frac{\partial \left(vec\left(\alpha_\perp\right)\right)}{\partial \left(vec\left(\alpha\right)\right)'} = \left[\left[c\left(\alpha c\right)^{-1} \otimes \left(c\left(\alpha c\right)^{-1} \alpha c_\perp \right)' \right] - \left[c\left(\alpha c\right)^{-1} \otimes c_\perp' \right] \right] K_{r,L}$$

where for an $m \times n$ matrix A, $K_{mn} vec\left(A\right) = vec\left(A'\right)$ (Magnus and Neudecker 1988, p. 47), we have

$$
\begin{aligned}
\frac{\partial vec\left(\Pi\right)}{\left(\partial vec\left(\alpha\right)\right)'} &= \left(I_L \otimes \beta \right) \\
&\quad + \left(\tilde{\Sigma} \otimes S_{11}^{-1} \beta_\perp \lambda \right) \left[\left[c\left(\alpha c\right)^{-1} \otimes \left(c\left(\alpha c\right)^{-1} \alpha c_\perp \right)' \right] - \left[c\left(\alpha c\right)^{-1} \otimes c_\perp' \right] \right] K_{r,L} \\
&= \left(I_L \otimes \beta \right) + \left(\tilde{\Sigma} c\left(\alpha c\right)^{-1} \otimes S_{11}^{-1} \beta_\perp \lambda c_\perp' \left[\left(c\left(\alpha c\right)^{-1} \alpha \right)' - I_L \right] \right) K_{r,L}.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \left(vec\left(\Pi\right)\right)}{\partial \left(vec\left(\beta_2\right)\right)'} &= \frac{\partial \left(vec\left(\Pi\right)\right)}{\partial \left(vec\left(\beta\right)\right)'} \frac{\partial \left(vec\left(\beta\right)\right)}{\partial \left(vec\left(\beta_2\right)\right)'} \\
&\quad + \frac{\partial \left(vec\left(\Pi\right)\right)}{\partial \left(vec\left(\beta_\perp\right)\right)'} \frac{\partial \left(vec\left(\beta_\perp\right)\right)}{\partial \left(vec\left(\beta\right)\right)'} \frac{\partial \left(vec\left(\beta\right)\right)}{\partial \left(vec\left(\beta_2\right)\right)'}
\end{aligned}
$$

$$\frac{\partial \left(vec\left(\Pi\right)\right)}{\partial \left(vec\left(\beta\right)\right)'} = \left(\alpha' \otimes I_p \right)$$

$$\frac{\partial \left(vec\left(\beta\right)\right)}{\partial \left(vec\left(\beta_2\right)\right)'} = \left[I_r \otimes \left(\begin{array}{c} I_r \\ 0_{(p-r)\times r} \end{array} \right) \right] \frac{\partial \left(vec\left(\beta_1\right)\right)}{\partial \left(vec\left(\beta_2\right)\right)'} + \left[I_r \otimes \left(\begin{array}{c} 0_{r\times(p-r)} \\ I_{p-r} \end{array} \right) \right].$$

To find the partial derivative $\frac{\partial\left(vec\left(\beta_1\right)\right)}{\partial\left(vec\left(\beta_2\right)\right)'}$, express β_1 as a function of β_2. The r^2 restrictions in (5) and (6) provide this expression. These r^2 restrictions, therefore, enter the Jacobian from the following expressions:

$$
\begin{aligned}
\beta' S_{11} \beta &= I_r \Longrightarrow \beta_1' \left(S_{11}\right)_{11} \beta_1 = I_r - \beta_1' \left(S_{11}\right)_{12} \beta_2 - \beta_2' \left(S_{11}\right)_{21} \beta_1 - \beta_2' \left(S_{11}\right)_{22} \beta_2 \\
&\quad\ 2N_r \left(I_r \otimes \beta_1' \left(S_{11}\right)_{11} \right) dvec\left(\beta_1\right) + 2N_r \left(I_r \otimes \beta_2' \left(S_{11}\right)_{21} \right) dvec\left(\beta_1\right) \\
&= -2N_r \left(I_r \otimes \beta_1' \left(S_{11}\right)_{12} \right) dvec\left(\beta_2\right) - 2N_r \left(I_r \otimes \beta_2' \left(S_{11}\right)_{22} \right) dvec\left(\beta_2\right)
\end{aligned}
$$

$$
\begin{aligned}
&\quad\ 2N_r \left[\left(I_r \otimes \beta_1' \left(S_{11}\right)_{11} \right) + \left(I_r \otimes \beta_2' \left(S_{11}\right)_{21} \right) \right] dvec\left(\beta_1\right) \\
&= -2N_r \left[\left(I_r \otimes \beta_1' \left(S_{11}\right)_{12} \right) + \left(I_r \otimes \beta_2' \left(S_{11}\right)_{22} \right) \right] dvec\left(\beta_2\right)
\end{aligned}
$$

$$LN_r\left[(I_r\otimes\beta_1'\,(S_{11})_{11})+(I_r\otimes\beta_2'\,(S_{11})_{21})\right]dvec\,(\beta_1)$$
$$=\ -LN_r\left[(I_r\otimes\beta_1'\,(S_{11})_{12})+(I_r\otimes\beta_2'\,(S_{11})_{22})\right]dvec\,(\beta_2)$$

which is an $\frac{r(r+1)}{2}\times 1$ vector. Next let $D=S_{10}S_{00}^{-1}S_{01}$, then

$$\beta'D\beta\ =\ \Gamma_r\Longrightarrow\beta_1'D_{11}\beta_1=\Gamma_r-\beta_1'D_{12}\beta_2-\beta_2'D_{21}\beta_1-\beta_2'D_{22}\beta_2$$
$$2N_r\,(I_r\otimes\beta_1'D_{11})\,dvec\,(\beta_1)+2N_r\,(I_r\otimes\beta_2'D_{21})\,dvec\,(\beta_1)$$
$$=\ dvec\,(\Gamma_r)-2N_r\,(I_r\otimes\beta_1'D_{12})\,dvec\,(\beta_2)-2N_r\,(I_r\otimes\beta_2'D_{22})\,dvec\,(\beta_2)$$

$$2N_r\left[(I_r\otimes\beta_1'D_{11})+(I_r\otimes\beta_2'D_{21})\right]dvec\,(\beta_1)$$
$$=\ dvec\,(\Gamma_r)-2N_r\left[(I_r\otimes\beta_1'D_{12})+(I_r\otimes\beta_2'D_{22})\right]dvec\,(\beta_2)$$

$$\bar{L}N_r\left[(I_r\otimes\beta_1'D_{11})+(I_r\otimes\beta_2'D_{21})\right]dvec\,(\beta_1)$$
$$=\ -\bar{L}N_r\left[(I_r\otimes\beta_1'D_{12})+(I_r\otimes\beta_2'D_{22})\right]dvec\,(\beta_2)$$

which is an $\frac{r(r-1)}{2}\times 1$ vector. Combine these expressions as

$$\begin{bmatrix}LN_r\,[1]\\\bar{L}N_r\,[2]\end{bmatrix}dvec\,(\beta_1)=-\begin{bmatrix}LN_r\,[3]\\\bar{L}N_r\,[4]\end{bmatrix}dvec\,(\beta_2)$$

where [1] and [2] are full rank r^2, N_r is $r^2\times r^2$ rank $\frac{r(r+1)}{2}$, L is $\frac{r(r+1)}{2}\times r^2$ of rank $\frac{r(r+1)}{2}$, $\bar{L}$ is $\frac{r(r-1)}{2}\times r^2$ of rank $\frac{r(r-1)}{2}$, so $LN_r\,[1]$ is $\frac{r(r+1)}{2}\times r^2$ of rank $\frac{r(r+1)}{2}$, $\bar{L}N_r\,[1]$ is $\frac{r(r-1)}{2}\times r^2$ of rank $\frac{r(r-1)}{2}$ and

$$\begin{aligned}
[1]&=\ (I_r\otimes\beta_1'\,(S_{11})_{11})+(I_r\otimes\beta_2'\,(S_{11})_{21})=(I_r\otimes\beta'\,(S_{11})_1)\\
[2]&=\ (I_r\otimes\beta_1'D_{11})+(I_r\otimes\beta_2'D_{21})=(I_r\otimes\beta'D_1)\\
[3]&=\ (I_r\otimes\beta_1'\,(S_{11})_{12})+(I_r\otimes\beta_2'\,(S_{11})_{22})=(I_r\otimes\beta'\,(S_{11})_2)\\
[4]&=\ (I_r\otimes\beta_1'D_{12})+(I_r\otimes\beta_2'D_{22})=(I_r\otimes\beta'D_2)
\end{aligned}$$

$$D=\begin{bmatrix}D_{11}&D_{12}\\D_{21}&D_{22}\end{bmatrix}=\begin{bmatrix}D_1&D_2\end{bmatrix}\qquad S_{11}=\begin{bmatrix}(S_{11})_{11}&(S_{11})_{12}\\(S_{11})_{21}&(S_{11})_{22}\end{bmatrix}=\begin{bmatrix}(S_{11})_1&(S_{11})_2\end{bmatrix}.$$

Therefore $\begin{bmatrix}LN_r\,[1]\\\bar{L}N_r\,[2]\end{bmatrix}$ is $r^2\times r^2$ and has rank less than r^2 with probability zero. Thus,

$$dvec\,(\beta_1)=-\begin{bmatrix}LN_r\,[1]\\\bar{L}N_r\,[2]\end{bmatrix}^{-1}\begin{bmatrix}LN_r\,[3]\\\bar{L}N_r\,[4]\end{bmatrix}dvec\,(\beta_2)$$

and therefore,

$$\frac{\partial\left(vec\left(\beta_1\right)\right)}{\partial\left(vec\left(\beta_2\right)\right)'} = -\left[\begin{array}{c} LN_r\,[1] \\ \bar{L}N_r\,[2] \end{array}\right]^{-1}\left[\begin{array}{c} LN_r\,[3] \\ \bar{L}N_r\,[4] \end{array}\right],$$

which is $r^2 \times (p-r)\,r$ of less than full rank with probability zero. That this matrix is of full rank is important to ensure the Jacobian matrix has full rank. The other matrices in the Jacobian matrix have relatively simple structures and so it is not difficult to determine they have full rank. Using the above results, we have,

$$\frac{\partial\left(vec\left(\beta\right)\right)}{\partial\left(vec\left(\beta_2\right)\right)'} = \left[I_r \otimes \left(\begin{array}{c} 0_{r\times(p-r)} \\ I_{p-r} \end{array}\right)\right] - \left[I_r \otimes \left(\begin{array}{c} I_r \\ 0_{(p-r)\times r} \end{array}\right)\right]\left[\begin{array}{c} LN_r\,[1] \\ \bar{L}N_r\,[2] \end{array}\right]^{-1}\left[\begin{array}{c} LN_r\,[3] \\ \bar{L}N_r\,[4] \end{array}\right].$$

Next,

$$\frac{\partial\left(vec\left(\Pi\right)\right)}{\partial\left(vec\left(\beta_\perp\right)\right)'} = \left[\tilde{\Sigma}\alpha_\perp'\lambda' \otimes S_{11}^{-1}\right].$$

From the transformation in (8), we have $\beta = S_{11}^{-\frac{1}{2}}\Gamma_1$ and $\beta_\perp = S_{11}^{\frac{1}{2}}\Gamma_2$ where, $\Gamma_1 = U_1 T$, $\Gamma_2 = U_2 T_1$, $\Gamma_1'\Gamma_1 = I_r$, $\Gamma_2'\Gamma_2 = I_{p-r}$ and $\Gamma_1'\Gamma_2 = 0_{r\times(p-r)}$. Therefore we can form the orthonormal matrix $\Gamma = [\Gamma_1\ \Gamma_2]$, such that $\Gamma'\Gamma = I_p$ and $\tilde{X} = (I+\Gamma)^{-1}(I-\Gamma)$ is skew-symmetric, that is, $\tilde{X} = -\tilde{X}'$. This orthomorphic transformation between Γ and $\tilde{X}$ implies $\Gamma = \left(I+\tilde{X}\right)^{-1}\left(I-\tilde{X}\right)$ (Olkin and Sampson, 1972). We can now find an expression for the matrix of partial derivatives

$$\begin{aligned}\frac{\partial\left(vec\left(\beta_\perp\right)\right)}{\partial\left(vec\left(\beta\right)\right)'} &= \frac{\partial(vec(\beta_\perp))}{\partial\left(vec\left(\Gamma_2\right)\right)'}\frac{\partial\left(vec\left(\Gamma_2\right)\right)}{\partial(vec(\tilde{X}))'}\frac{\partial(vec(\tilde{X}))}{\partial\left(vec\left(\Gamma_1\right)\right)'}\frac{\partial\left(vec\left(\Gamma_1\right)\right)}{\partial(vec(\beta))'}\\ &= \left(I_{p-r}\otimes S_{11}^{\frac{1}{2}}\right)A_2 B_1\left(I_r\otimes S_{11}^{\frac{1}{2}}\right)\end{aligned}$$

where

$$\frac{\partial\left(vec\left(\beta_\perp\right)\right)}{\partial\left(vec\left(\Gamma_2\right)\right)'} = \left(I_{p-r}\otimes S_{11}^{\frac{1}{2}}\right) \qquad \frac{\partial\left(vec\left(\Gamma_1\right)\right)}{\partial\left(vec\left(\beta\right)\right)'} = \left(I_r\otimes S_{11}^{\frac{1}{2}}\right)$$

$$\frac{\partial\left(vec\left(\Gamma_2\right)\right)}{\partial\left(vec\left(\tilde{X}\right)\right)'} = A_2 \qquad \frac{\partial\left(vec\left(\tilde{X}\right)\right)}{\partial\left(vec\left(\Gamma_1\right)\right)'} = B_1,$$

28

A_2 is the matrix comprised of the last $p(p-r)$ rows of

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = - \left[(I_p + \Gamma)' \otimes \left(I_p + \tilde{X} \right)^{-1} \right]$$

and B_1 is the matrix comprised of the first pr columns of

$$[B_1 \ B_2] = - \left[\left(I_p + \tilde{X} \right)' \otimes (I_p + \Gamma)^{-1} \right].$$

Finally, let $m = k + \frac{p(p+1)}{2}$

$$\frac{\partial vec\,(\Pi)}{(\partial vec\,(\lambda))'} = \left(\tilde{\Sigma} \alpha'_\perp \otimes S_{11}^{-1} \beta_\perp \right)$$

where for an $n \times n$ matrix A, $D_n vech\,(A) = vec\,(A)$. From the above expressions:

$$|J\,(\Pi : (\beta_2, \alpha, \lambda))| = \left| \frac{\partial vec\,(\Pi)}{(\partial vec\,(\alpha))'} \quad \frac{\partial vec\,(\Pi)}{(\partial vec\,(\beta_2))'} \quad \frac{\partial vec\,(\Pi)}{(\partial vec\,(\lambda))'} \right|.$$

When $\lambda = 0$, then the $Lp \times Lp$ matrix

$$\frac{\partial vec\,(\Pi)}{(\partial vec\,(\alpha))'} \quad \frac{\partial vec\,(\Pi)}{(\partial vec\,(\beta_2))'} \quad \frac{\partial vec\,(\Pi)}{(\partial vec\,(\lambda))'}$$

is

$$\frac{\partial vec\,(\Pi)}{(\partial vec\,(\alpha))'} = (I_L \otimes \beta)$$

$$\frac{\partial vec\,(\Pi)}{(\partial vec\,(\beta_2))'} = (\alpha' \otimes I_p) \left[\left[I_r \otimes \begin{pmatrix} 0_{r \times (p-r)} \\ I_{p-r} \end{pmatrix} \right] + \left[I_r \otimes \begin{pmatrix} I_r \\ 0_{(p-r) \times r} \end{pmatrix} \right] \ddot{E} \right]$$

$$\frac{\partial vec\,(\Pi)}{(\partial vec\,(\lambda))'} = \left(\tilde{\Sigma} \alpha'_\perp \otimes S_{11}^{-1} \beta_\perp \right)$$

and

$$\ddot{E} = - \begin{bmatrix} L N_r\,[1] \\ \bar{L} N_r\,[2] \end{bmatrix}^{-1} \begin{bmatrix} L N_r\,[3] \\ \bar{L} N_r\,[4] \end{bmatrix},$$

from which it should be apparent $|J|_{\lambda=0}| \neq 0$.

8 Appendix II: Singular Value Decompositions of Nonsquare φ.

Here the SVD of the $(p \times L)$ matrix

$$\varphi = S_{11}^{-\frac{1}{2}} \varphi^* \widetilde{\Sigma}^{\frac{1}{2}} = \psi\alpha + S_{11}^{-1}\psi_\perp \lambda\alpha_\perp \widetilde{\Sigma}$$

is presented. First consider the case where $L > p$. Let $\underset{p \times p}{U} = \left[\begin{array}{cc} U_1 & U_2 \\ p \times r & p \times (p-r) \end{array} \right]$, $\underset{L \times L}{V} = \left[\begin{array}{ccc} V_1 & V_2 & V_3 \\ L \times r & L \times (p-r) & L \times (L-p) \end{array} \right]$, and $\underline{S}_1$ and $\underline{S}_2$ be diagonal $r \times r$ and $(p - r) \times (p - r)$ respectively. Make the following transformation:

$$
\begin{aligned}
\varphi^* &= U\underline{S}V' = \left[\begin{array}{cc} U_1 & U_2 \end{array} \right] \left[\begin{array}{ccc} \underline{S}_1 & 0 & 0 \\ 0 & \underline{S}_2 & 0 \end{array} \right] \left[\begin{array}{c} V_1' \\ V_2' \\ V_3' \end{array} \right] \qquad (26)\\
&= \left[\begin{array}{cc} U_1 T & U_2 \end{array} \right] \left[\begin{array}{ccc} T'\underline{S}_1 & 0 & 0 \\ 0 & \underline{S}_2 & 0 \end{array} \right] \left[\begin{array}{c} V_1' \\ V_2' \\ V_3' \end{array} \right] \\
&= U_1 T T' \underline{S}_1 V_1' + U_2 \underline{S}_2 V_2'
\end{aligned}
$$

where the $r \times r$ orthogonal matrix T is chosen such that $T'T = TT' = I$ and

$$T'U_1'S_{11}^{-\frac{1}{2}} D S_{11}^{-\frac{1}{2}} U_1 T = \Gamma = diag(\gamma_1, ..., \gamma_r)$$

where D is a $p \times p$ positive definite, symmetric matrix. Therefore,

$$\psi = S_{11}^{-\frac{1}{2}} U_1 T \qquad \alpha = T'\underline{S}_1 V_1' \widetilde{\Sigma}^{\frac{1}{2}}.$$

Next, define the following matrices

$$
\begin{aligned}
T_1' U_2' S_{11}^{-1} U_2 T_1 &= \Lambda_{p-r} = diag(\eta_1, ..., \eta_{p-r}) \\
T_2' \left[\begin{array}{c} V_2' \\ V_3' \end{array} \right] S_{11}^{-\frac{1}{2}} S_{10} S_{00}^{-1} S_{01} S_{11}^{-\frac{1}{2}} [V_2 \ V_3] T_2 &= \Lambda_{L-r} = diag(\varsigma_1, ..., \varsigma_{L-r})
\end{aligned}
$$

so that T_1 is $(p - r) \times (p - r)$ and $\underset{(L-r) \times (L-r)}{T_2} = \left[\begin{array}{cc} T_{21} & T_{22} \\ (L-r) \times (p-r) & (L-r) \times (L-p) \end{array} \right]$.

The last term in (26) becomes,

$$U_2 \underline{S}_2 V_2' = U_2 T_1' T_1 \underline{S}_2 T_{21}' T_2 \left[\begin{array}{c} V_2' \\ V_3' \end{array} \right]$$

so that from (13)

$$S_{11}^{-1}\psi_\perp = S_{11}^{-\frac{1}{2}}U_2 T_1' \qquad \Rightarrow \qquad \psi_\perp = S_{11}^{\frac{1}{2}}U_2 T_1',$$

$$\lambda = T_1 \underline{S}_2 T_{21}',$$

and

$$\alpha_\perp \widetilde{\Sigma} = T_2 \begin{bmatrix} V_2' \\ V_3' \end{bmatrix} \widetilde{\Sigma}^{\frac{1}{2}} \qquad \Rightarrow \qquad \alpha_\perp = T_2 \begin{bmatrix} V_2' \\ V_3' \end{bmatrix} \widetilde{\Sigma}^{-\frac{1}{2}}.$$

Next, consider the case where $p > L$. Let $\underset{p \times p}{U} = \begin{bmatrix} \underset{p \times r}{U_1} & \underset{p \times (L-r)}{U_2} & \underset{p \times (p-L)}{U_3} \end{bmatrix}$, $\underset{L \times L}{V} = \begin{bmatrix} \underset{L \times r}{V_1} & \underset{L \times (L-r)}{V_2} \end{bmatrix}$ and $\underline{S}_1$ and $\underline{S}_2$ be diagonal $r \times r$ and $(L-r) \times (L-r)$ respectively containing the singular values of φ^*. Make the following transformation:

$$\begin{aligned}
\varphi^* &= U\underline{S}V' = \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix} \begin{bmatrix} \underline{S}_1 & 0 \\ 0 & \underline{S}_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \qquad (27) \\
&= \begin{bmatrix} U_1 T & U_2 & U_3 \end{bmatrix} \begin{bmatrix} T'\underline{S}_1 & 0 \\ 0 & \underline{S}_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \\
&= U_1 T T' \underline{S}_1 V_1' + U_2 \underline{S}_2 V_2'
\end{aligned}$$

where the $r \times r$ orthogonal matrix T is chosen such that $T'T = TT' = I$ and

$$T'U_1'S_{11}^{-\frac{1}{2}}DS_{11}^{-\frac{1}{2}}U_1 T = \Gamma = diag(\gamma_1, ..., \gamma_r)$$

where D is again a $p \times p$ positive definite, symmetric matrix. Therefore,

$$\psi = S_{11}^{-\frac{1}{2}}U_1 T \qquad \alpha = T'\underline{S}_1 V_1' \widetilde{\Sigma}^{\frac{1}{2}}.$$

Next, define the following matrices

$$T_1' \begin{bmatrix} U_2' \\ U_3' \end{bmatrix} S_{11}^{-1} \begin{bmatrix} U_2 & U_3 \end{bmatrix} T_1 = \Lambda_{p-r} = diag(\eta_1, ..., \eta_{p-r})$$

$$T_2'V_2'S_{11}^{-\frac{1}{2}}S_{10}S_{00}^{-1}S_{01}S_{11}^{-\frac{1}{2}}V_2 T_2 = \Lambda_{L-r} = diag(\varsigma_1, ..., \varsigma_{L-r})$$

so that $\underset{(p-r)\times(p-r)}{T_1} = \begin{bmatrix} \underset{(p-r)\times(L-r)}{T_{11}} & \underset{(p-r)\times(p-L)}{T_{12}} \end{bmatrix}$ and T_2 is $(L-r) \times (L-r)$.

31

The last term in (27) becomes,

$$U_2 \underline{S}_2 V_2' = \begin{bmatrix} U_2 & U_3 \end{bmatrix} T_1' T_{11} \underline{S}_2 T_2 T_2' V_2'$$

so that from (13)

$$S_{11}^{-1} \psi_\perp = S_{11}^{-\frac{1}{2}} \begin{bmatrix} U_2 & U_3 \end{bmatrix} T_1' \qquad \Rightarrow \qquad \psi_\perp = S_{11}^{\frac{1}{2}} \begin{bmatrix} U_2 & U_3 \end{bmatrix} T_1',$$

$$\lambda = T_{11} \underline{S}_2 T_2,$$

and

$$\alpha_\perp \widetilde{\Sigma} = T_2' V_2' \widetilde{\Sigma}^{\frac{1}{2}} \qquad \Rightarrow \qquad \alpha_\perp = T_2' V_2' \widetilde{\Sigma}^{-\frac{1}{2}}.$$