AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

MONASH

WP 9/98

ISSN 1440-771X
ISBN 0 7326 1046 X

MONASH UNIVERSITY

ANCORA · IMPARO

AUSTRALIA

LIBRARY
WITHDRAWN
NOV 2 4 1998
GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS

Bayesian Estimation of the Reduced Rank Regression Model without Ordering Restrictions

Rodney W. Strachan

Working Paper 9/98
August 1998

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

Bayesian Estimation of the Reduced Rank Regression Model without Ordering Restrictions.*

Rodney W. Strachan†
Department of Econometrics and Business Statistics,
Monash University,
Clayton, Vic., 3168,
Australia
email: Rodney.Strachan@BusEco.monash.edu.au

The latest version of this paper may be found at
http://www.buseco.monash.edu.au/Depts/EBS/ESAM98/

August 12, 1998

Key Words: error correction model; identification; cointegration; Bayesian analysis.

ABSTRACT

Estimation of the parameters of the reduced rank regression model in a Bayesian method requires the solution of two identification problems: global or strong identification and local identification. Traditionally Bayesians, and

*This work is based on work from the author's Ph.D. thesis and relates to a paper presented at the Econometric Society Australasian Meeting 1998, A.N.U., Canberra, Australia.

†The author would like to thank Brett Inder, David Harris, and Gael Martin for valuable discussions that have improved this paper.

to a large extent frequentists, have relied on zero-one identifying restrictions which require the researcher to impose an order on the variables to achieve global identification. Examples of this approach include Geweke (1996), Bauwens and Lubrano (1993), Kleibergen (1997), Kleibergen and Paap (1997), and Kleibergen and van Dijk (1994). This ordering relies on *a priori* knowledge of which variables enter the reduced rank relations. For example, the cointegrating error correction model requires knowledge of which variables are I(0) or cointegrate. Incorrect ordering may result in an estimated space for the cointegrating vectors that does not have the true cointegrating space as a subset, effectively misspecifying the model. In this paper, we present an estimation method which does not require *a priori* ordering by using restrictions similar to those used in maximum likelihood estimation by Anderson (1951) of the reduced rank regression model generally, and by Johansen (1988) in an error correction model specifically. As with much of the recent work, we focus on the cointegrating error correction model to show our approach.

Local identification is achieved by nesting the reduced rank model within a full rank model with a well behaved posterior distribution. This approach is due to Kleibergen (1997) and is consistent with the principle of a "data-translated likelihood" suggested by Box and Tiao (1973). In nesting the reduced rank model in a full rank model, we use a transformation from the potentially reduced rank matrix Π to the matrices α, β and λ where $\lambda = 0$ restricts Π to a lower rank. Results from Roy (1952) enable us to derive the Jacobian for this transformation.

1 Introduction.

In this paper, a method for Bayesian estimation without *a priori* ordering of the variables in the reduced rank multivariate regression model is presented. The reduced rank regression model has received significant attention in econometrics with both frequentist and, more recently, Bayesian treatment. Important examples of applications include the cointegrating error correction model (ECM) (see for example Johansen 1988, Kleibergen and van Dijk 1994, Kleibergen 1997, Kleibergen and Paap 1997 - hereafter referred to as K&P, and Geweke 1996) and the incomplete simultaneous equations model (SEM) (Dreze 1976, Dreze and Richard 1983, and Zellner, Min and Dallaire 1993, Geweke 1996). Classical likelihood based analysis of the reduced rank multivariate regression model was first presented by Anderson (1951).

The reduced rank multivariate regression model is,

$$Y = X\Pi + ZA + E = \underline{X}B + E. \tag{1}$$

where $\underline{X} = [X \quad Z]$ and $B = [\Pi' \quad A']'$. Further, Y is a $T \times L$ matrix of dependent variables, X and Z are, respectively, $T \times p$ and $T \times k$ matrices of explanatory variables, and E is a $T \times L$ matrix of errors with covariance matrix $\Sigma \otimes I_T$. The coefficient matrix Π is of rank $r \leq min(L,p)$, while A is full rank. When Π has reduced rank it can be expressed as $\Pi = \beta\alpha$ where β and α are $p \times r$ and $r \times L$ and it is assumed $rank\,(\alpha) = rank\,(\beta) = r$. An alternative representation is to transform the potentially reduced rank matrix Π to the matrices α, β and λ where $\lambda = 0$ restricts Π to a lower rank.

Identifying restrictions are necessary for global identification of the elements of β and α. Regardless of whether these identifying restrictions have been imposed, there may exist areas of local nonidentification in the space of $(\beta\ \alpha)$. As outlined below, the problem of local nonidentification has been extensively investigated, whereas global identification has been solved in the Bayesian approaches by restrictions which impose an order on the variables in X.

1.1 Why consider (non)ordering restrictions?

A common method of achieving strong identification, and in fact the only method we are aware that has been employed in Bayesian studies to date, is to apply r^2 zero-one restrictions on either $\beta = [\,\beta_1' \mid \beta_2']'$ or α, such as setting $\beta^* = \beta\beta_1^{-1} = [\,I_r \mid \beta_2^{*\prime}]'$ (Geweke 1996, Bauwens and Lubrano 1993, Kleibergen and van Dijk 1994, and K&P 1997). This method of achieving identification 'forces' a set of r variables to appear in the reduced rank relationships. For example, in the cointegrating ECM the researcher is required to make assumptions about which variables cointegrate (or are *I(0)*) (Johansen 1995a p.94). The researcher must then appropriately 'order' the variables. However, we may wish to investigate precisely this question of which variables enter the reduced rank relations, or at least we may wish to avoid making such strong assumptions.

It is important to consider ordering, as the normalisation $\beta^* = \beta\beta_1^{-1}$ is only valid if β_1^{-1} exists. Consider a cointegrating bivariate system

$$\Delta w_t = w_{t-1}\beta\alpha + e_t = z_{t-1}\alpha + e_t, \qquad (2)$$

where the vectors $\alpha = (\alpha_1\ \alpha_2)$ and $\beta = \left[\begin{array}{cc} \beta_1 & \beta_2 \end{array}\right]'$. Let $\Delta w_t = (\Delta x_t\ \Delta y_t)'$,and $w_t = (x_t\ y_t)'$. Next assume $x_t \sim I(1)$, $y_t \sim I(0)$ and that $w_t\beta = y_t\beta_2 = z_t \sim I(0)$ is the only cointegrating relationship. By incorrectly normalizing and estimating β by $\tilde{\beta}\left(=\beta\beta_1^{-1}\right) = \left[\begin{array}{cc} 1 & \tilde{\beta}_2 \end{array}\right]'$, the variable $\tilde{z}_t = w_t\tilde{\beta} = x_t + y_t\tilde{\beta}_2$ is included in the estimated equation, which is not an estimate of z_t. Unless $\alpha = 0$ to exclude $\tilde{z}_t$ from the equation, (2) will be an unbalanced equation as $\tilde{z}_t \sim I(1)$ but $\Delta w_t \sim I(0)$. This result would indicate there is no cointegrating relation. Regardless of the value of α, the variable z_t has been excluded, effectively misspecifying the model. Forecasting is an application which clearly shows the implications of incorrect normalisation. Without a correct estimate for the disequilibrium errors, z_t, it is unlikely the forecast at $T+h$, $h > 0$, will be forced back to the space spanned by $\beta_\perp$, the attractor set (Johansen, 1995, p41). If, however, β is correctly normalized and estimated by $\hat{\beta}\left(=\beta\beta_2^{-1}\right) = \left[\begin{array}{cc} \hat{\beta}_1 & 1 \end{array}\right]'$, it is now possible to estimate $\hat{\beta}_1 = 0$ and correctly choose and estimate one cointegrating relation. The forecasts are then more likely to follow a path that is correct in its response to disequilibrium errors, that is, the forecasts can move *toward* the attractor set. Figures 1 (a) to 1 (f) show the consequences for one step ahead forecasts of incorrectly normalizing β for the model in (2). Figures 1 (a) and 1 (d) show the true error correction term and forecast error, z_t and e_t. Figures 1 (b) and 1 (e) show the estimates $\hat{z}_t$ and $\hat{e}_t$ where β has been correctly normalised and estimated as $\hat{\beta} = \left[\hat{\beta}_1, 1\right]'$, and finally, Figures 1 (c) and 1 (f) show the estimates $\tilde{z}_t$ and $\tilde{e}_t$ where β has been incorrectly normalised and estimated as $\tilde{\beta} = \left[1, \tilde{\beta}_2\right]'$. Both z_t and $\hat{z}_t$ display $I(0)$ behavior, as do the errors, e_t and $\hat{e}_t$. The estimates $\tilde{z}_t$ and $\tilde{e}_t$ inherit the stochastic trend as $sp(\beta) \subsetneq sp\left(\tilde{\beta}\right)$, and so $w_t\tilde{\beta} \sim I(1)$. These results clearly show how incorrectly normalizing to achieve global identification can have serious implications for forecasting performance.

Insert Figures 1 (a) to 1 (f) here.

This is a simple example of an exclusion restriction (that is a variable is excluded from the relation) which requires the variables in w_t to be correctly ordered such that β_1^{-1} exists. Another way β_1 may be singular is if a subset of the variables enter all reduced rank relations in the same way. Juselius (1995) considers two important economic questions: long run evidence of purchasing power parity (PPP) and uncovered interest rate parity (UIP). She uses Danish and German data on the variables log consumer prices in Denmark and Germany, p_d and p_g, log exchange rate e_{dg}, the Danish bond rate i_d, and the German bond rate i_g. If the variables are ordered as $w_t = (p_d, p_g, e_g, i_d, i_g)_t$, and we assume that the cointegrating relations contain (p_d, p_g, e_g) only through the PPP relation, $p_d - p_g - e_{dg}$, this implies $\beta = H\varphi$ where φ is a general $3 \times r$ matrix, with $r = 1, 2, 3$,

$$H = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ and } \quad \beta = \begin{bmatrix} \varphi_1 \\ -\varphi_1 \\ -\varphi_1 \\ \varphi_2 \\ \varphi_3 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

such that $sp(\beta) \subset sp(H)$. This implies if $r > 1$, then β_1 will be singular. While we may spuriously estimate $\tilde{\beta} = \left[I_r \mid \tilde{\beta}_2' \right]'$, $sp(\tilde{\beta})$ is not restricted to be in $sp(H)$ (Johansen 1995a, p.73) and PPP cannot be imposed on all cointegrating vectors. We are then unable to estimate $sp(\beta)$, and so there will be no mechanism to allow forecasts to respond to disequilibriums errors $w_t\beta - E(w_t\beta)$. Further general examples in which β_1 is singular can be found in Johansen (1995a).

As we stated earlier, identifying restrictions are necessary for global identification of the elements of β and α. Clearly, however, how these restrictions are imposed on α and β matters. In Section 2 we discuss the concept of identification and its treatment in the frequentist and Bayesian approaches to date. We use this discussion to motivate our approach.

In Section 3 we present the nonordinal transformation of the potentially reduced rank square matrix Π in (1) and the Jacobian for this transformation. The details of the derivation of the Jacobian and the transformation in the case where Π is not square, are left to the appendices I and II respectively. A virtue of the method presented in this paper is that we may test hypotheses about the space spanned by β using Bayes factors. To do this, we need to be able to estimate the model with β subject to these restrictions. The necessary transformations to estimate these models are presented in Section 4. It turns out that these transformations are very similar to those presented in Appendix II for the nonsquare Π. The diffuse and natural conjugate priors, the likelihood, the resulting posteriors and the sampling scheme are presented in Section 5. An illustrative example is presented in Section 6 and Section 7 concludes.

2 Identification.

In the classical approach, identification is largely treated as a property of the likelihood function, $L = L(Y|\theta)$. Let θ be a k-dimensional vector in R^k. Two points θ^1 and θ^2 are said to be *observationally equivalent* if $L(Y|\theta^1) = L(Y|\theta^2)$ for *all* values of Y. The point θ^1 in R^k is *(globally) identified* if there exists no other $\theta \neq \theta^1$ in R^k which is observationally equivalent to θ^1 (Hsiao, 1983). The point θ^1 is *locally identified* if there exists an open neighbourhood, $\Im$ in R^k, containing θ^1 such that no other θ in $\Im$ is observationally equivalent to θ^1 (Hsiao, 1983). It is easy to show the unrestricted matrices α and β are unidentified by setting $\beta^1 = \beta^2\kappa^{-1}$ and $\alpha^1 = \kappa\alpha^2$ such that $\beta^1\alpha^1 = \beta^2\kappa^{-1}\kappa\alpha^2$ for nonzero scaler or nonsingular $r \times r$ matrix κ. Therefore, $L(Y|\beta^1, \alpha^1) = L(Y|\beta^2, \alpha^2)$ for all Y.

Whether another definition of identification is necessary for the Bayesian is not clear. Several authors have addressed this question and a recent useful overview of this discussion can be found in Aldrich (1994a). While the implications of nonidentification for the estimator can vary significantly between the two approaches, for the purposes of this paper, we accept Kadane's (1974, p. 175) frequently quoted view that "identification is a property of the likelihood function and is the same whether considered classically or from the Bayesian approach." Identification can then be determined for the parameter prior to choosing an estimation method. Local identification is discussed first as this is a necessary condition for global identification and because local identification is relatively easy to determine.

2.1 Local identification.

Research relating to identification appears to fall into three broad, overlapping areas. First, there has been considerable Bayesian work investigating the pathologies associated with local nonidentification (see for example Kleibergen and van Dijk 1994, Kleibergen 1997, Bauwens and Lubrano 1993, and Dreze 1974).

Both classical and Bayesian discussions of local nonidentification relate to the flatness in the likelihood in the direction of the unidentified parameters, θ. This flatness or observational equivalence results in a singular information matrix, $\mathcal{I}_\theta$, which in turn results from linear dependence of the elements of the score vector (Rothenberg 1971, p.580). Haavelmo presents a related result that if the first derivatives of L (rather than $\ln L$) are linearly independent "for every point in the interior of the parameter space, then every parameter point is *locally nonarbitrary*" (Aldrich 1994b, p. 210). This singularity of the information matrix is used in the classical approach to determine local identification. Bayesian approaches use this singularity to counter the effect of the flatness in the likelihood.

Much of the work on Bayesian analysis of cointegration has necessarily focused on the issue of local nonidentification of the elements in $\beta\alpha$ in (1) with respect to the likelihood at the point where $rank\,(\beta\alpha) < r$ (however clearly this is not the only cause of local nonidentification), and the near nonidentification in this region. Although this point has measure zero, using a flat prior for β and α in the cointegrating ECM will result in possibly infinite posterior moments and improper posteriors (see Kleibergen and van Dijk, 1994) and therefore make it difficult to obtain meaningful inferences about these models. These pathologies are due to the flatness in the likelihood being transmitted to the posterior.

The second area of research has covered possible solutions to these pathologies (Kleibergen and van Dijk 1994, Kleibergen 1997, Geweke 1996, and K&P 1997). Methods used to offset the flatness and so enable inference, have included using an informative prior (Geweke, 1996) or a Jeffrey's prior (Kleibergen and van Dijk, 1994) for the nonidentified parameters θ, in the restricted model. The former method produces proper posteriors. However, as Kadane (1974) shows in theorem 2 for a two point prior, the posterior probability for an hypothesis on the locally nonidentified parameters at a point of local nonidentification, will simply be the prior probability, that is, the experiment that generated the data is said to be uninformative for θ. The use of the Jeffreys' prior is motivated by the fact that the local nonidentification manifests itself as a singular information matrix (Rothenberg 1971). The Jeffreys prior, $p_J(\theta)$, uses this singularity (or more accurately the approach towards singularity) to offset the problematic effect of nonidentification - flatness in the likelihood - since $p_J(\theta) = |\mathcal{I}_\theta|^{\frac{1}{2}}$.

The method in this paper uses Kleibergen's (1997 and K&P 1997) 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 well a behaved posterior distribution with a uniform prior, and then the model is transformed such as to parameterize the restriction to a lower rank, effectively conditioning on the given rank r. This is the transformation from the potentially reduced rank matrix Π to the matrices α, β and λ where $\lambda = 0$ restricts Π to a lower rank. Specifying a uniform prior on Π will not translate into a uniform prior for (β, α, λ). However, this approach is consistent with Box and Tiao's (p.26 1973) suggestion that a uniform prior should be used for the parameterization for which the likelihood is 'data translated' - that is, the likelihood "is in location form in terms of sufficient statistics" (Chao and Phillips 1996, p. 17). In this case, the variable Y enters the likelihood through the location parameter for Π, which is the OLS estimate, eg. $\hat{\Pi} = (X'X)^{-1} X'Y$, and the location parameter for Σ, which is the OLS estimate $\hat{S} = Y'Y - Y'X\,(X'X)^{-1} X'Y$. The implied prior for (β, α, λ) given Σ, is then the Jacobian for the transformation from the parameters of the general model to the parameters of the nested model.

If the parameters of interest are population moments, Rothenberg's Theorem 4 (1971) shows they will be (globally) identified. As the conditional mean of $(\underline{X}'\underline{X})^{-1} \underline{X}'Y$ is B, then $\pi = vec\,(B)$ is identified. As global identification implies local identification, this result implies $\mathcal{I}_\pi$ is nonsingular. For the matrices E_i, $i = 1, \ldots, j$ of varying dimensions, let the notation $\delta = vec\,(E_1, E_2, \ldots, E_j)$ imply $\delta = \left(vec\,(E_1)', vec\,(E_2)', \ldots, vec\,(E_j)'\right)'$. Using Haavelmo's definition of reducibility (Aldrich, 1994b) and the transformation of Π to (β, α, λ), $\pi = vec\,(\Pi, A, \Sigma)$ is regarded as the reduced form of the reduced rank regression model, and $\theta = vec\,(\beta, \alpha, \lambda, A, \Sigma)$ as the structural parameters. Let the Jacobian matrix for the transformation between π and θ be J. If π is an $m \times 1$ vector and θ is $n \times 1$, then if J is full rank, θ is locally identified since $\mathcal{I}_\theta = J'\mathcal{I}_\pi J$ will be full rank n. Thus θ is 'statistically' locally identified according to Rothenberg (1971) since $\mathcal{I}_\theta^{-1}$ exists and, further, θ is locally identified according to Koopmans since J is full rank (Aldrich 1994a).

As discussed in Poirier (1995), Kolmogorov (1950) and at length in this context in Kleibergen (1997), there is no unique way of conditioning on an hypothesis with zero measure (such as $\lambda = 0$). This implies that it is possible to get different results for Bayes factors for different specifications of the restricted model, such as reordering the variables in X. Kleibergen (1997) presents results that allow us to determine specifications of the transformation for which the posterior, and therefore the Bayes factors, are invariant. The invariant specification with nonordinal identifying restrictions is presented in Section 2.

The third area of Bayesian research has outlined similarities between Bayesian posterior distributions and finite sample distributions of classical estimators in certain cases. Chao and Phillips (1996) show the similarity of the form of the posterior pdf of the coefficients in an SEM with a Jeffreys prior with the classical finite sample distribution of the LIML estimator for the same coefficients - both distributions have Cauchy-like tails. Phillips (1994) shows that the exact finite sample distribution of the classical reduced rank regression estimator $\hat{B} = \hat{\beta}_2\hat{\beta}_1^{-1}$(ie., Johansen's) has Cauchy-like tails and no finite moments of integer order. Whether the Bayesian estimator of $B = \beta_2\beta_1^{-1}$, and the classical reduced rank regression estimator $\hat{B}$ show such similarities as in the SEM is an interesting area for possible future research.

2.2 Global or Strong Identification.

The elements of β and α are not identified in the strong sense and so a number of restrictions are required to achieve global identification. We generalise the results of Bauwens and Lubrano (1993) who formally show the following result for the case where $L = p$. If $rank(\Pi) = r < \min(L, p)$ such that we can write $\Pi = \beta\alpha$ where β and α are $p \times r$ and $r \times L$ matrices respectively, $Lr+pr-r^2$ elements of β and α are identified. As there are Lp elements in Π from which we wish to obtain the $Lr+pr$ elements in β and α, r^2 restrictions need to be applied. Another way of looking at the number of necessary restrictions consistent with the method used in this paper, is as follows. Generally $Lp \neq Lr + pr$. Taking Π to be a general matrix, decompose this into three matrices β, α and λ of dimensions $p \times r$, $r \times L$ and $(L-r) \times (p-r)$ respectively such that if Π is of rank r, then $\lambda = 0$ and so $\Pi = \beta\alpha$. This decomposition requires a number of appropriate identifying restrictions, $\Re$, such that $Lp = Lr + pr + (L-r) \times (p-r) - \Re$. Therefore apply $\Re = r^2$ prior restrictions to β and α to achieve strong identification. Determining the number of restrictions required says nothing of the form those restrictions must take. Therefore we need to discuss when the parameters α and β are identified and then ask the question: do our restrictions identify the elements of α and β?

The restrictions used in this paper resemble those used by Johansen in his maximum likelihood estimator of the space spanned by cointegrating vectors. This method of estimation "implies an identification of the individual coefficients, as the coefficients of the eigenvectors of a suitable eigenvalue problem" (Johansen 1995b, p.123). This method of estimation is equivalent to performing a singular value decomposition (SVD) of the unrestricted OLS estimator $\hat{\Pi}$, which has been normalised with respect to its estimate of the covariance matrix, $S_{00.1} \otimes S_{11}^{-1}$ (Johansen 1995a, p. 94). That is, an SVD is taken of the matrix $\hat{\Pi}^* = S_{00.1}^{-\frac{1}{2}} \hat{\Pi} S_{11}^{\frac{1}{2}}$. The Bayesian equivalent, presented in this paper, is to perform an SVD of Π with respect to its covariance matrix, $\Sigma \otimes S_{11}^{-1}$ (this term is further defined when necessary in Appendix III).

Geweke (1996) addressed this ordering problem by estimating all $\sum_{r=0}^{L} \begin{pmatrix} L \\ r \end{pmatrix}$ possible orders and estimating predictive probabilities for each order. This approach provides a valid means of obtaining the expected values for β, however it could become cumbersome if L becomes large. In this paper a method for Bayesian estimation without ordering restrictions is presented in the reduced rank multivariate regression model. This alternative method of achieving strong identification is to define r^2 elements of β (or α) as deterministic functions of the remaining $(p-r)\,r$ (or $(L-r)\,r$) elements in that matrix. That is, $\beta = [\,\beta_1' \mid \beta_2']'$, where β_1 is an $r \times r$ matrix and β_2 is a $(p-r) \times r$ matrix, is defined such that $\beta_1 = f(\beta_2)$. This nomination of β_1 as a function β_2 of is somewhat artificial in that if *any* $(p-r)\,r$ elements of β are known, then the values of the remaining r^2 elements are known through the identifying equations. However, this nomination does help us to derive the Jacobian. This approach to implementing the restrictions, which is well established in the frequentist literature, was used by Anderson (1951) and Johansen (1988) and has the advantage that it removes the necessity for *a priori* consideration of the ordering of the variables in X. As cointegration has become an important concept in econometrics over the past decade and it implies a reduced rank structure on a matrix of parameters in the ECM, the ECM will be used throughout the paper as an example to demonstrate the application of the above method of identification.

3 The Nonordinal Transformation.

In this section we present the transformation which avoids consideration of the order of the variables. This is achieved by nesting the restricted model within a general model which has a well behaved posterior, such as the full rank model. To nest the reduced rank regression model in a full rank model, reparameterize the model by the following transformation

$$\Pi = S_{11}^{-\frac{1}{2}} \Pi^* \Sigma^{\frac{1}{2}} \tag{3}$$

where Σ is the variance-covariance matrix of the errors in the model and the covariance of Π is $\Sigma \otimes T^{-1} S_{11}^{-1}$. The decomposition in (3) is equivalent to taking an SVD of Π with respect to its variance-covariance matrix, and is therefore equivalent to the approach Johansen (1988) takes in his maximum likelihood estimator. The SVD depends on the dimensions of Π and decompositions for nonsquare matrices are presented in Appendix II. Here it is assumed Π is square ($p = L$) as in the ECM and 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:

$$\begin{aligned} \Pi^* &= U\underline{S}V' = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \underline{S}_1 & 0 \\ 0 & \underline{S}_2 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \\ &= \begin{bmatrix} U_1\overline{T} & U_2 \end{bmatrix} \begin{bmatrix} \overline{T}'\underline{S}_1 & 0 \\ 0 & \underline{S}_2 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \\ &= U_1\overline{T}\overline{T}'\underline{S}_1 V_1' + U_2 \underline{S}_2 V_2' \end{aligned} \tag{4}$$

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

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

in which D is a $p \times p$ positive definite, symmetric matrix. We choose D to conform with Johansen's matrix at the mode of the likelihood. For the diffuse prior specification, this matrix will be exactly the same as Johansen's matrix. That is, $D = S_{10} S_{00}^{-1} S_{01}$ where S_{10}, S_{01}' and S_{00} are defined such that with a diffuse prior, at the mode of the likelihood β will be equal to the value given by the Johansen maximum likelihood estimator. Denote the eigenvalues of the $p \times p$ matrix $S_{11}^{-\frac{1}{2}} D S_{11}^{-\frac{1}{2}}$, by $(\lambda_1, ..., \lambda_p)$ where $\lambda_1 \geq \ldots \geq$

$\lambda_p \geq 0$, although in practice $\lambda_1 > \ldots > \lambda_p > 0$. These eigenvalues are also the squared canonical correlation coefficients. The diagonal elements of Γ, $(\gamma_1, ..., \gamma_r)$, are bounded by $\lambda_i \geq \gamma_i \geq \lambda_{p-r+i}$ (Schott 1997, p.111), and at the mode of the likelihood, $\lambda_i = \gamma_i$.

This gives the following expressions for β and α,

$$\beta = S_{11}^{-\frac{1}{2}} U_1 \overline{T} \qquad \alpha = \overline{T}' \underline{S}_1 V_1' \Sigma^{\frac{1}{2}}. \tag{5}$$

Next, define the following matrices

$$\begin{aligned} T_1' U_2' S_{11}^{-1} U_2 T_1 &= \Lambda_{p-r} = diag(\eta_1, ..., \eta_{p-r}) \\ T_2' V_2' \Sigma V_2 T_2 &= \Lambda_{p-r} = diag(\varsigma_1, ..., \varsigma_{p-r}) \end{aligned}$$

so that T_1 and T_2 are $(p-r) \times (p-r)$.

The last term in (4) becomes,

$$U_2 S_2 V_2' = U_2 T_1 T_1' \underline{S}_2 T_2 T_2' V_2'$$

so that from (3),

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

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

and finally

$$\lambda = T_1' \underline{S}_2 T_2. \tag{6}$$

The resultant transformation is

$$\Pi = \beta\alpha + S_{11}^{-1} \beta_\perp \lambda \alpha_\perp \Sigma = S_{11}^{-\frac{1}{2}} \Pi^* \Sigma^{\frac{1}{2}} \tag{7}$$

where β is $p \times r$, α is $r \times p$, λ is $(p-r) \times (p-r)$, $\beta_\perp$ is $p \times (p-r)$ and $\alpha_\perp$ is $(p-r) \times p$ such that $\beta' \beta_\perp = 0$ and $\alpha_\perp \alpha' = 0$. Further, we have $\alpha_\perp \Sigma \alpha_\perp' = I_{p-r}$ and $\beta_\perp' S_{11}^{-1} \beta_\perp = I_{p-r}$. In this transformation, Π is restricted to lower rank where $\lambda = 0$. The Borel-Kolmogorov paradox states that there is no unique form for the distribution of any generally specified β and α conditional on this restriction. This implies that Bayes factors will not be unique, making inference dependent upon the specification of β, α and λ. However, by defining λ such that it is locally uncorrelated at $\lambda = 0$ with β

and α, as in (7), the resulting posterior will be invariant as will the Bayes factors.

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

$$\beta' S_{11} \beta = I \tag{8}$$

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

$$\beta' D \beta = \Gamma = diag(\gamma_1, ..., \gamma_r) \tag{9}$$

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. In the expression (9), Γ is random with an *implied* posterior distribution in the Bayesian method and $\gamma_1 > ... > \gamma_r > 0$. The resultant model is now

$$Y = X\beta\alpha + XS_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma + ZA + E \tag{10}$$

where the reduced rank model occurs at $\lambda = 0$,

$$Y = X\beta\alpha + ZA + E. \tag{11}$$

The values of the matrices S_{11}, S_{10}, and S_{00} are chosen such that they conform in the posterior with a diffuse prior to the values used for the maximum likelihood estimator. These matrices are defined in Appendix III.

3.1 The Jacobian for the transformation.

The above transformation gives the following equations.

$$\begin{aligned} \Pi &= \beta\alpha + S_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma, \\ \Sigma &= \Sigma. \end{aligned} \tag{12}$$

The restrictions are expressed in the r^2 equations in

$$Lvec(\beta' S_{11} \beta - I) = 0 \tag{13}$$

and

$$\bar{L}vec(\beta' D \beta - \Gamma) = 0 \tag{14}$$

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).

As Σ is symmetric, the following results use only its lower triangular elements. For the purpose of finding an expression for the Jacobian, let β_2 contain the free parameters such that given (β_2, X, Y, D), then β_1 is known. Using expression (3.1) of Roy (1952), the Jacobian matrix can be expressed as

$$\begin{aligned} J((\Pi,\Sigma):(\beta_2,\alpha,\lambda,\Sigma)) &= \frac{\partial vec(\Pi,\Sigma)}{\partial vec(\beta_2,\alpha,\lambda,\Sigma)'} \\ &= \begin{bmatrix} \frac{\partial vec(\Pi)}{\partial vec(\beta_2)'} & \frac{\partial vec(\Pi)}{\partial vec(\alpha)'} & \frac{\partial vec(\Pi)}{\partial vec(\lambda)'} & \frac{\partial vec(\Pi)}{\partial vec(\Sigma)'} \\ \frac{\partial vec(\Sigma)}{\partial vec(\beta_2)'} & \frac{\partial vec(\Sigma)}{\partial vec(\alpha)'} & \frac{\partial vec(\Sigma)}{\partial vec(\lambda)'} & \frac{\partial vec(\Sigma)}{\partial vec(\Sigma)'} \end{bmatrix} \end{aligned}$$

where the forms of these expressions are presented in Appendix I. From Appendix I,

$$\begin{aligned} \frac{\partial vec(\Sigma)}{\partial vec(\beta_2)} &= 0_{\frac{p(p+1)}{2}\times r(p-r)} & \frac{\partial vec(\Sigma)}{\partial vec(\alpha)} &= 0_{\frac{p(p+1)}{2}\times pr} \\ \frac{\partial vec(\Sigma)}{\partial vec(\lambda)} &= 0_{\frac{p(p+1)}{2}\times (p-r)^2} & \frac{\partial vec(\Sigma)}{\partial vec(\Sigma)} &= I_{\frac{p(p+1)}{2}\times \frac{p(p+1)}{2}} \end{aligned}$$

and so the Jacobian has the form:

$$\begin{aligned} |J((\Pi,\Sigma):(\beta_2,\alpha,\lambda,\Sigma))| &= |J| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \\ &= |A_{22}|\left|A_{11} - A_{12}A_{22}^{-1}A_{21}\right| = |A_{11}| \end{aligned} \tag{15}$$

where

$$A_{11} = \frac{\partial vec(\Pi)}{\partial vec(\beta_2)'} \quad \frac{\partial vec(\Pi)}{\partial vec(\alpha)'} \quad \frac{\partial vec(\Pi)}{\partial vec(\lambda)'} \qquad A_{21} = \frac{\partial vec(\Pi)}{\partial vec(\Sigma)'}$$

$$A_{12} = \frac{\partial vec(\Sigma)}{\partial vec(\beta_2)'} \quad \frac{\partial vec(\Sigma)}{\partial vec(\alpha)'} \quad \frac{\partial vec(\Sigma)}{\partial vec(\lambda)'} \qquad A_{22} = \frac{\partial vec(\Sigma)}{\partial vec(\Sigma)'}.$$

3.2 Linear Restrictions on β.

In the Introduction it was argued that normalizing β by $\beta\beta_1^{-1}$ will not always be appropriate as β_1 may be singular, depending on the order of the variables in X. Further, this normalization forces the researcher to declare which r variables *must* enter the relationship $X\beta$, again ordering so the first r variables are not excluded from $X\beta$. These issues raise the question of what is the appropriate order for the columns of X and, in particular, what is the form of β? That is, 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 (1995b) 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) $\quad H_0 : \beta = H\psi$

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

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

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

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

$$Y = X\Pi + ZA + \varepsilon = XH\varphi + ZA + \varepsilon$$

where $\varphi = \psi\alpha + (H'S_{11}H)^{-1}\psi_\perp \lambda \alpha_\perp \Sigma$. The Jacobian for this transformation is of the same form as the Jacobian in Appendix I with S_{11}, β, and Π replaced by $H'S_{11}H$, ψ, and φ respectively. The definition of S_{11} remains unchanged in this case.

For R_2, the model becomes

$$\begin{aligned} Y &= X\Pi + ZA + \varepsilon \\ &= Xb\alpha_1 + Xb_\perp\varphi + ZA + \varepsilon \\ &= Xb_\perp\varphi + \underset{\sim}{Z}\underset{\sim}{A} + \varepsilon \end{aligned}$$

where $\underset{\sim}{Z} = [Xb\ Z]$, $\underset{\sim}{A} = [\alpha_1'\ A']'$ and $\varphi = \psi\alpha_2 + (b_\perp' S_{11} b_\perp)^{-1} \psi_\perp \lambda \alpha_{2\perp}\Sigma$. Again the Jacobian can be found by using the form of the Jacobian in Appendix I with S_{11}, β, Π and α replaced by $b_\perp' S_{11} b_\perp$, ψ, φ, and α_2 respectively. The definition of S_{11} in the simple case of a diffuse prior, becomes $S_{11} = T^{-1}(X'X - X'\ \underset{\sim}{Z}\ (\underset{\sim}{Z}'\underset{\sim}{Z})^{-1}\ \underset{\sim}{Z}'\ X)$.

Finally, for R_3, the model becomes

$$\begin{aligned} Y &= X\Pi + ZA + \varepsilon \\ &= XH_1\varphi_1 + XH_2\varphi_2 + \cdots + XH_l\varphi_l + ZA + \varepsilon \end{aligned}$$

where $\varphi_i = \psi_i\alpha_i + (H_i' S_{11i} H_i)^{-1} \psi_{i\perp}\lambda_i\alpha_{i\perp}\Sigma$, $i = 1, \ldots, l$, $l \leq r$. Each of the l Jacobians, J_i, can be found by using the form of the Jacobian in Appendix I with S_{11}, β, Π, α and λ replaced by $H_i' S_{11i} H_i$, ψ_i, φ_i, α_i, and λ_i respectively. The definition of S_{11i} again when we have a diffuse prior, becomes $S_{11i} = T^{-1}(X'X - X'\ \underset{\sim}{Z_i}\ (\underset{\sim}{Z_i}'\underset{\sim}{Z_i})^{-1}\ \underset{\sim}{Z_i}'\ X)$ where the matrix $\underset{\sim}{Z_i} = [XH_1 \cdots XH_l\ Z]$ does not include XH_i. Estimation of these models requires little extra computer coding beyond that required for the general model using the transformation (7).

4 Priors and Posteriors.

In this section we present the forms of the posterior for the general reduced rank regression model and the ECM representation of the vector autoregressive model with one lag (VAR(1)) with no deterministic terms for a simple application.

4.1 The Model: ECM for a VAR(1).

The relevant functions for the general model in (1) are presented in the following discussion. However, in some places it is necessary to present relevant features of these functions in a simple case only, therefore the ECM for a VAR(1) with one lag and no deterministic terms is used. This model has the form:

$$\begin{aligned} Y_t &= X_{t-1}\Pi + e_t \qquad (16) \\ &= X_{t-1}\beta\alpha + X_{t-1}\left(X'X\right)^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma + e_t, \quad t = 1, ..., T \end{aligned}$$

where α' and $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ are $p \times r$ matrices, β_1 is $r \times r$, β_2 is $(p-r) \times r$, λ is $(p-r) \times (p-r)$, and the $1 \times p$ vector e_t is distributed as independent $N_p(0, \Sigma)$. Let $Y_t = \Delta X_t$, $Y = (\Delta X_1', \Delta X_2', ..., \Delta X_T')'$, $E = (e_1', e_2', ..., e_T')'$, and $\underline{X} = X = \left(X_0', X_1', ..., X_{T-1}'\right)'$ so $Y = X\Pi + E$. The restrictions in (8) and (9) are imposed on the system such that the parameters α and β will be identified.

4.2 Priors

The information matrix, $\mathcal{I}_\theta$, for the structural parameters $\theta = vec\,(\alpha, \beta, \lambda, \Sigma)$ in (16) is presented in Appendix I. The Jeffreys prior is proportional to $|\mathcal{I}|^{1/2}$. Therefore, the form of the diffuse (d) or Jeffreys prior for the parameters in the reduced form models $\pi = vec\,(\Pi, A, \Sigma)$ (1) and (16) is:

$$p_\pi\,(\Sigma, B)_d \propto |\mathcal{I}_\pi|^{1/2} \propto |\Sigma|^{-(L+p+k+1)/2}.$$

The natural conjugate (n) prior for the unrestricted linear model in (1) and (16) is:

$$\begin{aligned} p_\pi\,(\Sigma, B)_n \;\propto\; & |\Sigma|^{-(\underline{\nu}+L+1)/2} \exp\left\{-\frac{1}{2} tr \Sigma^{-1}\underline{S}\right\} \\ & |\Sigma|^{-(p+k)/2}\,|\underline{H}|^{L/2} \exp\left\{-\frac{1}{2} tr\Sigma^{-1}\left(B - \underline{B}\right)'\,\underline{H}\left(B - \underline{B}\right)\right\} \end{aligned}$$

where $B = [\Pi' \quad A']'$ and $\underline{B} = [\underline{\Pi}' \quad \underline{A}']'$ for (1) and $B = \Pi$ and $\underline{B} = \underline{\Pi}$ for (16). Using the expression for the Jacobian in Appendix I, J, the resultant priors for the structural parameters θ in the transformed nonlinear regression model (10) are

$$p_\theta(\Sigma,\alpha,\beta,\lambda,A)_d = p_\pi(\Sigma,\Pi(\alpha,\beta,\lambda),A)_d\,|J| \propto |J\mathcal{I}_\pi J'|^{1/2} \propto |\mathcal{I}_\theta|^{1/2} \quad (17)$$

and

$$p_\theta(\Sigma,\alpha,\beta,\lambda,A)_n = p_\pi(\Sigma,\Pi(\alpha,\beta,\lambda),A)_n\,|J|\,. \quad (18)$$

The priors for the reduced rank models are found by evaluating the above expressions at $\lambda=0$ and we use the expression $|J|_{\lambda=0}|$ to represent the value of the Jacobian at $(\alpha,\beta,\lambda=0,\Sigma)$. As in was found in K&P, the above priors cannot be decomposed to provide useful marginal and conditional densities. So the same simulation scheme as used in K&P is used in this paper. This is a Metropolis-Hastings algorithm with the posterior for the unrestricted model as the candidate density for the reduced rank models.

4.3 The likelihood.

For the full rank models in (1) and (16), assume the rows of $E=(e_1',e_2',\ldots,e_T')'$ are independently and normally distributed as $N(0,\Sigma)$. Under these assumptions, the likelihood can be written as

$$L(Y|\Sigma,\beta,\alpha,\lambda,A,\underline{X}) \propto |\Sigma|^{-\frac{T}{2}}\exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}E'E\right)\right\} \quad (19)$$

where

$$E=\left(Y-X\beta\alpha-XS_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma-ZA\right)$$

for (1), and

$$E=\left(Y-X\beta\alpha-X\left(X'X\right)^{-1}\beta_\perp\lambda\alpha_\perp\Sigma\right)$$

for the VAR(1) ECM in (16).

4.4 The posteriors.

The posterior for model (1) when we use the diffuse prior is:

$$\begin{aligned} p_\pi(\Sigma,B|Y,\underline{X})_d &\propto p_\pi(\Sigma)_d\,p_\pi(B|\Sigma)_d\,L(Y|\Sigma,B,\underline{X}) \\ &\propto |\Sigma|^{-(T+L+p+k+1)/2}\exp\left\{-\frac{1}{2}tr\Sigma^{-1}(Y-\underline{X}B)'(Y-\underline{X}B)\right\} \end{aligned}$$

$$\propto \quad |\Sigma|^{-(p+k)/2} \exp\left\{-\frac{1}{2} tr\Sigma^{-1}\left(B-\hat{B}\right)' \underline{X}'\underline{X}\left(B-\hat{B}\right)\right\}$$

$$\times |\Sigma|^{-(T+L+1)/2} \exp\left\{-\frac{1}{2} tr\Sigma^{-1} Y'\underline{X}\left(\underline{X}'\underline{X}\right)^{-1} \underline{X}'Y\right\}.$$

The posterior for the transformed model as parameterised in (10) with the diffuse prior may be decomposed as follows (see Appendix III).

$$\begin{aligned}
p_\pi\left(\Sigma, B|Y, \underline{X}\right)_d &= p_\pi\left(A, \lambda, \alpha, \beta, \Sigma|Y, \underline{X}\right)_d |J| \\
&= p\left(A|\lambda, \alpha, \beta, \Sigma, Y, \underline{X}\right)_d p\left(\lambda|\alpha, \beta, \Sigma, Y, \underline{X}\right)_d p\left(\alpha, \beta|\Sigma, Y, \underline{X}\right)_d \\
&\quad \times p\left(\Sigma|Y, \underline{X}\right)_d |J| \\
&\propto \exp\left(-\frac{1}{2} tr\Sigma^{-1}\left(A-\Delta\right)' Z'Z\left(A-\Delta\right)\right) |\Sigma|^{-k/2} \\
&\quad \times \exp\left(-\frac{T}{2} tr\left(\lambda-\widehat{\lambda}\right)'\left(\lambda-\widehat{\lambda}\right)\right) |\Sigma|^{-(p-r)/2} \\
&\quad \times \exp\left(-\frac{T}{2} tr\Sigma^{-1}\left(\alpha-\widehat{\alpha}\right)'\left(\alpha-\widehat{\alpha}\right)\right) |\Sigma|^{-r/2} \\
&\quad \times \exp\left(-\frac{1}{2} tr\Sigma^{-1}\left[\tilde{S}+T\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}\right]\right) |\Sigma|^{-(T+L+1)/2} |J|.
\end{aligned}$$

The terms used in this section are defined in Appendix III. This provides an expression for the posterior for the reduced rank model (10) at $\lambda = 0$ as:

$$\begin{aligned}
& p_\pi\left(A, \alpha, \beta, \Sigma|\lambda=0, Y, \underline{X}\right)_d |J|_{\lambda=0}| \\
=\; & p\left(A|\lambda=0, \alpha, \beta, \Sigma, Y, \underline{X}\right)_d p\left(\alpha, \beta|\Sigma, Y, \underline{X}\right)_d p\left(\Sigma|Y, \underline{X}\right)_d |J|_{\lambda=0}| \\
\propto\; & \exp\left(-\frac{1}{2} tr\Sigma^{-1}\left(A-\Delta|_{\lambda=0}\right)' Z'Z\left(A-\Delta|_{\lambda=0}\right)\right) |\Sigma|^{-k/2} \\
& \times \exp\left(-\frac{T}{2} tr\widehat{\lambda}'\widehat{\lambda}\right) |\Sigma|^{-(p-r)/2} \\
& \times \exp\left(-\frac{T}{2} tr\Sigma^{-1}\left(\alpha-\widehat{\alpha}\right)'\left(\alpha-\widehat{\alpha}\right)\right) |\Sigma|^{-r/2} \\
& \times \exp\left(-\frac{1}{2} tr\Sigma^{-1}\left[\tilde{S}+T\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}\right]\right) |\Sigma|^{-(T+L+1)/2} |J|_{\lambda=0}|.
\end{aligned}$$

The posterior for model (1) when we use the natural conjugate prior is:

$$p_\pi\left(\Sigma, B|Y, X\right)_n \propto |\Sigma|^{-(p+k)/2} \exp\left\{-\frac{1}{2} tr\Sigma^{-1}\left(B-\widetilde{B}\right)'\left(\underline{H}+\underline{X}'\underline{X}\right)\left(B-\widetilde{B}\right)\right\}$$

$$\times |\Sigma|^{-(T+L+\underline{\nu}+1)/2} \exp\left\{-\frac{1}{2} tr \Sigma^{-1} \tilde{S}\right\}.$$

The posterior for (10) with the n prior may be decomposed as follows (see Appendix III).

$$\begin{aligned}
p_\pi(\Sigma, B|Y, \underline{X})_n &= p(A|\lambda, \alpha, \beta, \Sigma, Y, \underline{X})_n \, p(\lambda|\alpha, \beta, \Sigma, Y, \underline{X})_n \\
&\quad \times p(\alpha, \beta|\Sigma, Y, \underline{X})_n \, p(\Sigma|Y, \underline{X})_n |J| \\
&\propto \exp\left(-\frac{1}{2} tr \Sigma^{-1} (A - \Delta)' (\underline{H}_{22} + Z'Z)(A - \Delta)\right) |\Sigma|^{-k/2} \\
&\quad \times \exp\left(-\frac{T}{2} tr \left(\lambda - \widehat{\lambda}\right)' \left(\lambda - \widehat{\lambda}\right)\right) |\Sigma|^{-(p-r)/2} \\
&\quad \times \exp\left(-\frac{T}{2} tr \Sigma^{-1} (\alpha - \widehat{\alpha})' (\alpha - \widehat{\alpha})\right) |\Sigma|^{-r/2} \\
&\quad \times \exp\left(-\frac{1}{2} tr \Sigma^{-1} \left[\tilde{S} + T \widetilde{\Pi}' \widetilde{S}_{11} \widetilde{\Pi}\right]\right) |\Sigma|^{-(T+L+\underline{\nu}+1)/2} |J|.
\end{aligned}$$

This provides an expression for the posterior for model (10) at $\lambda = 0$ as:

$$\begin{aligned}
& p_\pi(A, \alpha, \beta, \Sigma|\lambda = 0, Y, \underline{X})_n \, |J|_{\lambda=0}| \\
= \; & p(A|\lambda = 0, \alpha, \beta, \Sigma, Y, \underline{X})_n \, p(\alpha, \beta|\Sigma, Y, \underline{X})_n \, p(\Sigma|Y, \underline{X})_n \, |J|_{\lambda=0}| \\
\propto \; & \exp\left(-\frac{1}{2} tr \Sigma^{-1} (A - \Delta|_{\lambda=0})' (\underline{H}_{22} + Z'Z)(A - \Delta|_{\lambda=0})\right) |\Sigma|^{-k/2} \\
& \times \exp\left(-\frac{T}{2} tr \widehat{\lambda}' \widehat{\lambda}\right) |\Sigma|^{-(p-r)/2} \\
& \times \exp\left(-\frac{T}{2} tr \Sigma^{-1} (\alpha - \widehat{\alpha})' (\alpha - \widehat{\alpha})\right) |\Sigma|^{-r/2} \\
& \times \exp\left(-\frac{1}{2} tr \Sigma^{-1} \left[\tilde{S} + T \widetilde{\Pi}' \widetilde{S}_{11} \widetilde{\Pi}\right]\right) |\Sigma|^{-(T+L+\underline{\nu}+1)/2} |J|_{\lambda=0}|.
\end{aligned}$$

These decompositions do not provide conditional distributions useful for sampling, except for A which has a conditional normal posterior distribution in all cases. However, they do provide us with amenable forms for finding expressions of the weights in the Metropolis-Hastings algorithm outlined in the next section.

4.5 Sampling scheme.

The aim of this section is to outline a procedure for obtaining a draw from the posterior $p_\theta(A, \alpha, \beta, \Sigma | \lambda = 0, Y, \underline{X})$. The forms of the joint and conditional posteriors of α, λ, β and Σ are not from known classes of probability density functions. To sample from these posteriors we use the same approach as detailed in K&P. That is, a Metropolis-Hasting algorithm is used with the candidate density being that for the full rank model, $p_\pi(A, \lambda, \alpha, \beta, \Sigma | Y, \underline{X}) |J|$. This sampling scheme is outlined here. Let $k = n$ or d depending on which prior we wish to use, $\theta = vec(\alpha, \lambda, \beta, \Sigma)$ and $\theta_{-\lambda} = vec(\alpha, \beta, \Sigma)$. First we use the following steps to draw from the candidate density. At iteration i, for $i = 1, \ldots, m$,

Step 1: Draw $\Sigma^{(i)}$ from $p_\pi(\Sigma | Y, \underline{X})_k$.
Draw $\Pi^{(i)}$ from $p_\pi\left(\Pi | \Sigma^{(i)}, Y, \underline{X}\right)_k$.

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 (5) and (6).

As we are unable to find known conditional pdfs for the parameters in (10), we sample θ at once, then sample A conditional on α, $\lambda = 0$, β and Σ. The matrix A has a conditional normal posterior pdf in every model considered in the previous section. As we have sampled λ which only appears in (10), to get draws from (11), we augment the posterior for the reduced rank model, $p_\pi(\theta_{-\lambda} | \lambda = 0, Y, \underline{X})_k \, |J|_{\lambda=0}|$, with a proper distribution for λ, $g(\lambda | \theta_{-\lambda}, Y, \underline{X})$. Again, the reader is directed to K&P for details of this method. The weights used to accept or reject draws are

$$\begin{aligned} w_i &= w(\theta | Y, \underline{X}) \\ &= \frac{g\left(\lambda^{(i)} | \theta_{-\lambda}, Y, \underline{X}\right) p_\pi\left(\theta_{-\lambda}^{(i)} | \lambda = 0, Y, \underline{X}\right)_k |J|_{\lambda=0}|}{p_\pi\left(\theta^{(i)} | Y, \underline{X}\right)_k |J|}. \end{aligned} \tag{20}$$

As discussed in K&P, Geweke (1989) shows $m^{-1} \sum_{i=1}^{m} w_i$ converges to the ratio of the integrals

$$\frac{\int g\left(\lambda^{(i)} | \theta_{-\lambda}^{(i)}, Y, \underline{X}\right) p_\pi\left(\theta_{-\lambda}^{(i)} | \lambda = 0, Y, \underline{X}\right)_k |J|_{\lambda=0}| \, d\theta}{\int p_\pi\left(\theta^{(i)} | Y, \underline{X}\right)_k |J| \, d\theta}. \tag{21}$$

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

Step 1: Draw $\theta^{*(i+1)}$ from $p_\pi\left(\theta|Y,\underline{X}\right)_k|J|$.

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

Step 3: Draw $A^{(i+1)}$ from $p\left(A|\theta^{(i+1)},Y,\underline{X}\right)$.

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

$$p\left(A,\lambda,\theta^{(i+1)}|Y,\underline{X}\right)=p\left(A|\theta^{(i+1)},Y,\underline{X}\right)g\left(\lambda|\theta_{-\lambda A},Y,\underline{X}\right)p_\pi\left(\theta_{-\lambda}|\lambda=0,Y,\underline{X}\right)_k|J|_{\lambda=0}|,$$

and the set $vec\left(\theta_{-\lambda}^{(i+1)},A^{(i+1)}\right)$ will be a draw from

$$p_\theta\left(A,\theta_{-\lambda}|Y,\underline{X}\right)_k=p_\pi\left(A,\theta_{-\lambda}|\lambda=0,Y,\underline{X}\right)_k|J|_{\lambda=0}|.$$

5 Application.

To demonstrate the applicability of the method presented in this paper we investigate the real-business-cycle model with permanent productivity shocks proposed in King, Plosser, Stock, and Watson (1991), and present results for this model using the consumption, investment and output data considered in Harris (1997). These results are compared with the results from the classical approach. To investigate the support for various hypotheses, we calculate the Bayes factor which is the ratio of the marginal likelihoods for the model under one hypothesis and the model under an alternative hypothesis, $m\left(y|H_0\right)$ and $m\left(y|H_A\right)$ respectively. That is,

$$BF(0|A)=BF(H_0|H_A)=m\left(y|H_0\right)/m\left(y|H_A\right).$$

The marginal likelihood for a particular model i, with parameters θ, is defined by the expression $p(\theta|y,H_i)=p(\theta|H_i)L(y|\theta,H_i)/m\left(y|H_i\right)$, where, for this model, $p(\theta|y,H_i)$ is the posterior density for θ, $p(\theta|H_i)$ is the prior density for θ, and $L(y|\theta,H_i)$ is the likelihood function for the model. Therefore, for all θ,

$$m\left(y|H_i\right)=\int p(\theta|H_i)L(y|\theta,H_i)d\theta=p(\theta|H_i)L(y|\theta,H_i)/p(\theta|y,H_i). \quad (22)$$

The ratio of the posterior probability for the hypothesis H_0, $P\left(H_0|y\right)$, and an alternative H_A, $P\left(H_A|y\right)$, is a function of the Bayes factor and the prior probabilities for these models, $P(H_0)$ and $P(H_A)$ respectively. That is, $P(H_0|y)/P(H_A|y) = P(H_0)/P(H_A) \times BF(0|A)$. Therefore, to estimate posterior probabilities for the models of interest, estimates of their marginal likelihoods or the relevant Bayes factors are required. A sampling based estimator of $BF(r|p)$ where H_r : $rank = r$, was suggested in K&P and uses the draws w_i of in (20). That is, we estimate $BF(r|p)$ by $B\hat{F}_w(r|p) = (c_r m)^{-1}\sum_{i=1}^{m} w_i$, where $c_r = (2\pi)^{-\frac{(p-r)^2}{2}}$ (K&P).

5.1 Consumption, Investment and Output for Australia.

In this section we provide an illustrative application of the methods presented in this paper. Harris (1997) investigates the evidence in Australian data for the real-business-cycle model with permanent productivity shocks proposed in King, Plosser, Stock, and Watson (1991). This model implies two cointegrating relationships. The differences between the log of consumption (c_t) and the log of output (y_t), $(c_t - y_t)$, and the log of investment (i_t) and the log of output, $(i_t - y_t)$, will be $I(0)$. So the vector $x_t = (c_t, i_t, y_t)'$ will be cointegrated with rank of $r = 2$ and cointegrating vectors $\beta = H$ where

$$H = (h_1\ h_2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}. \tag{23}$$

To demonstrate the applicability of the method, the analysis will investigate support for the following hypotheses.

$H_r : rank = r$ for $r = 0, 1, 2, 3$, $H_4 : \beta = H\psi_4$,
$H_5 : \beta = (h_1\ h_{1\perp}\psi_5)$, $H_6 : \beta = (h_2\ h_{2\perp}\psi_6)$, and
$H_7 : \beta = H_3\psi_7$,

where $H_3 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$, ψ_4 and ψ_7 are 2×2 matrices, and ψ_5 and ψ_6 are 2×1 vectors. The hypotheses H_0 to H_3 state the number of stochastic trends in x_t is $3 - r$. H_4 states that c_t, i_t and y_t enter the cointegrating relations only through $(c_t - y_t)$ and $(i_t - y_t)$ and some combination of these two terms is $I(0)$. H_5 and H_6 state that at least one of $(c_t - y_t)$ and $(i_t - y_t)$ respectively, are cointegrating relations and so are $I(0)$, and H_8 is a test of whether i_t can be excluded from the cointegrating relations.

Insert Figure 2 here.

We use the same data as Harris but the sample is extended to cover June 1971 to March 1997. The data, shown in Figure 2, are per capita, quarterly, seasonally adjusted observations and measured in constant 1989/1990 dollars. The details on construction of the series can be found in Harris (1997). King *et al.* estimate using a VAR(6) with a constant term for the U.S. data. We find a restricted VAR with 3 lags for i_t and 2 lags for c_t and y_t, with a constant is appropriate for the Australian data.

The Bayes and maximum likelihood estimates of β, assuming a rank of 2, are respectively:

$$\beta_{BAYES} = \begin{pmatrix} 14.88 & 2.55 \\ 2.23 & -2.19 \\ -17.54 & -0.42 \end{pmatrix} \text{ and } \beta_{ML} = \begin{pmatrix} -0.34 & 6.76 \\ -1.50 & 0.56 \\ 1.86 & -7.56 \end{pmatrix}.$$

The normalised estimates,

$$\beta_{BAYES} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1.028 & -1.008 \end{pmatrix} \text{ and } \beta_{ML} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1.035 & -1.011 \end{pmatrix},$$

are comparable with (23) and suggest support for the real-business-cycle model proposed in King *et al.* (1991). However, this model also implies there is one stochastic trend, and therefore the hypotheses about r need to be investigated. The calculated likelihood ratio test statistics and Bayesian posterior probabilities of the ranks are presented in Table 1. The classical results suggest acceptance of $r = 1$, however the test statistic is close enough to the critical value to suggest some support for $r = 2$. The Bayesian evidence in these results strongly suggests there is one stochastic trend, and so $r = 2$. These results, particularly the Bayesian results, do support the hypothesis of a real-business-cycle model.

In Tables 2, 3, 4, and 5, the results of the tests of hypotheses H_4, H_5, H_6, and H_7 conditional on the ranks, are presented. If the true rank were 2, then we would conclude from the results for H_4 that there is strong evidence in support of the real-business-cycle model. The classical results for the hypotheses H_5 and H_6 indicate that $(c_t - y_t)$ and $(i_t - y_t)$ enter the model as error correction terms only if the true rank were two, however, the Bayesian results suggest support for H_5 and H_6 only if we condition on a rank of one. The ambiguity in these conditional results for different ranks lead us to consider unconditional results. The Bayesian unconditional posterior probabilities of each H_i, $P(H_i|y) = \sum_{r=0}^{3} P(H_i|H_r, y)P(H_r|y)$, are included at the bottom of Tables 2 to 5. These results, particularly $P(H_4|y)$, suggest the real-business-cycle model with permanent productivity shocks is valid as there appears to be one stochastic trend and, although $(c_t - y_t)$ and $(i_t - y_t)$ are not valid error correction terms, the variables do enter the long run relations through some combination of these terms regardless of the rank. While we can compare frequentist and Bayesian conclusions from the conditional tests in Tables 2 to 5, no classical equivalent to this unconditional inference exists. Finally, the Bayesian conditional and unconditional results indicate i_t cannot be excluded from the cointegrating relations. However, using a classical level of significance less than 2% we would accept this restriction at the accepted rank of one.

Insert Tables 1 to 5 here.

6 Conclusion.

In this paper we have demonstrated the implications of incorrectly normalising the parameters of a reduced rank regression model to achieve global identification, and presented a method for estimating this model without using the ordering restrictions imposed in previous Bayesian and frequentist approaches. 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). 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 $\lambda = 0$. Further, this method is consistent with the principle suggested by Box and Tiao (1973) that, if a diffuse prior is preferred, this prior should be on the parameterisation for which the likelihood is data translated, that is Π. By this approach, the issue of local nonidentification is avoided as the Jacobian for the transformation offsets the flatness in the likelihood that occurs near problem areas, much like the Jeffreys' prior does for the nonnested reduced rank model. Further, since Π is a population moment and the Jacobian of the transformation is full rank, the elements of (α, β, λ) are locally identified. The form of the restrictions on β imply the elements of (α, β, λ) are globally identified.

The form of the restrictions allow sampling based estimates of Bayes factors for various hypotheses. Estimates of coefficients and posterior probabilities are presented for a simple model and compared with classical methods of inference. In this empirical example we find strong evidence, both conditional on the rank and unconditional in the Bayesian approach, in support of the real-business-cycle model of King *et al.* (1991). An advantage of the Bayesian approach is that it allows access to unconditional evidence on economic relations. There exists no classical equivalent to this unconditional inference. The approach presented in this paper can be extended to incorporate a range of models and allows inference in these models in a unified approach.

7 References.

Aldrich, J. (1994a) "Identification and Bayesian theory: an old debate revisited." Mimeograph, University of Southampton.

Aldrich, J. (1994b) "Haavelmo's identification theory." *Econometric Theory*, **10**, 198-219.

Anderson, T.W. (1951) "Estimating linear restrictions on regression coefficients for multivariate normal distributions." *Annals of Mathematical Statistics*, **22**, 327-351.

Box, G.E.P and G.C. Tiao (1973) *Bayesian Inference in Statistical Analysis,* John Wiley and Sons, Inc., New York.

Bauwens, L. and M. Lubrano (1993) "Identification restrictions and posterior densities in cointegrated Gaussian VAR systems." Mimeograph, Core, University Catholique de Louvain.

Chao, J.C. and P.C.B. Phillips (1996) "Bayesian posterior distributions in limited information analysis of the simultaneous equations model using the Jeffreys' prior." Cowles Foundation discussion paper number 1137, November 1996.

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

Dreze, J.H. (1974) "Bayesian theory of identification in simultaneous equations models." In S.E. Fienberg and A. Zellner (eds.) *Studies in Bayesian Econometrics and Statistics*, Chapter 5.1. Amsterdam: North-Holland, 159-174.

Dreze, J.H. (1976) "Bayesian limited information analysis of the simultaneous equations model." *Econometrica*, **44**, 1045-1075.

Dreze, J.H. and J.H. Richard (1983) "Bayesian analysis of simultaneous equations systems" in: Z. Griliches and M.D. Intriligater, eds., *Handbook of Econometrics*, Vol. 1 (North-Holland, Amsterdam).

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.

Harris, D. (1997) "Principal components analysis of cointegrated time series." *Econometric Theory*, **13**, 529-557.

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.

Hsiao, C (1983) "Identification." in: Z. Griliches and M.D. Intriligater, eds., *Handbook of Econometrics*, Vol. 1 (North-Holland, Amsterdam).

Johansen, S. (1988) "Statistical analysis of cointegration vectors." *Journal of Economic Dynamics and Control*, **12**, 231-254.

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

Johansen, S. (1995b) "Identifying restrictions of linear equations with applications to simultaneous equations and cointegration." *Journal of Econometrics*, **69**, 111-132.

Juselius, K. (1995) "Do purchasing power parity and uncovered interest rate parity hold in the long run? An example of likelihood inference in a multivariate time-series model." *Journal of Econometrics*, **69**, 211-240.

Kadane, J.B. (1974). "The role of identification in Bayesian theory." In S.E. Fienberg and A. Zellner (eds.) *Studies in Bayesian Econometrics and Statistics*, Chapter 5.2. Amsterdam: North-Holland, 175-191.

King, R.G., C.I. Plosser, J.H. Stock, and M.W. Watson (1991). "Stochastic trends and economic fluctuations." *The American Economic Review*, 81, 819-840.

Kleibergen, F. (1997). "Equality restricted random variables: Densities and sampling algorithms." *Econometric Institute Reports*, no. **9662**, Econometric Institute, Erasmus University Rotterdam.

Kleibergen, F. (1998). "Conditional Densities in Econometrics." Working paper, Econometric Institute, Erasmus University Rotterdam.

Kleibergen, F. and R. Paap (1997). "Priors, posterior odds and Lagrange multiplier statistics in Bayesian analysis of cointegration." *Econometric Institute Reports*, no. **9668**, 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.

Kolmogorov, A.N. (1950) *Foundations of the Theory of Probability*, Chelsea, New York.

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.

Phillips, P.C.B. (1994) "Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models." *Econometrica*, **62**, 73-93.

Poirier, D.J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach,* MIT Press, Cambridge.

Rothenberg, T.J. (1971) "Identification in parametric models." *Econometrica*, **39**, 577-591.

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.

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.

Zellner, A., C. Min, and D. Dallaire (1993) "Bayesian analysis of simultaneous equation and related models using the Gibbs sampler and convergence checks" H.G.B. Alexander Research Foundation working paper (University of Chicago, Chicago, IL).

8 Appendix I: The Jacobian for the Transformation.

From the transformation (12) in Section 2 we have the following equations

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

and the r^2 restrictions equations in (13) and (14)

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

and

$$\bar{L}vec(\beta' D\beta - \Gamma) = 0.$$

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

"**Theorem 1.** If $y_i = f_i(x_1, \ldots, x_m; x_{m+1}, \ldots, x_{m+n}) \quad (i = 1, \ldots, m)$ when $x_j's \quad (j = 1, \ldots, m+n)$ 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

$$\begin{aligned} |J(y_1, \ldots, y_m : x_1, \ldots, x_m)| &= \left| \frac{\delta y_i}{\delta x_j} \right| i, j = 1, \ldots, m \qquad (24) \\ &= \left| \frac{\delta f_i}{\delta x_j} + \sum_{k=m+1}^{m+n} \frac{\delta f_i}{\delta x_k} \frac{\delta x_k}{\delta x_j} \right| i, j = 1, \ldots, m \end{aligned}$$

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, λ and Σ as the $x_j's$ in (24). Therefore the form of the Jacobian matrix for (12) is

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

where

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

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

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

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

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

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

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

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

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

To find the partial derivative $\frac{\partial(vec(\beta_1))}{\partial(vec(\beta_2))'}$, express β_1 as a function of β_2. The r^2 restrictions in (8) and (9) provide this expression. These r^2 restrictions, therefore, enter the Jacobian from the following expressions:

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

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

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

which is an $\frac{r(r+1)}{2} \times 1$ vector.

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

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

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

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

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

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,

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

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,

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

Next,

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

From the transformation in (12), 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

$$\frac{\partial\left(vec\left(\beta_{\perp}\right)\right)}{\partial\left(vec\left(\beta\right)\right)'} = \frac{\partial(vec(\beta_{\perp}))}{\partial\left(vec\left(\Gamma_2\right)\right)'}\frac{\partial\left(vec\left(\Gamma_2\right)\right)}{\partial(vec(\tilde{X}))'}\frac{\partial(vec(\tilde{X}))}{\partial\left(vec\left(\Gamma_1\right)\right)'}\frac{\partial\left(vec\left(\Gamma_1\right)\right)}{\partial(vec(\beta))'}$$

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

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

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

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

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

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

Finally,

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

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

$$\frac{\partial vech\left(\Sigma\right)}{\left(\partial vec\left(\alpha\right)\right)'} = 0_{\frac{L(L+1)}{2}\times Lr} \qquad \frac{\partial vech\left(\Sigma\right)}{\left(\partial vec\left(\beta_2\right)\right)'} = 0_{\frac{L(L+1)}{2}\times r(p-r)}$$

$$\frac{\partial vech\left(\Sigma\right)}{\left(\partial vec\left(\lambda\right)\right)'} = 0_{\frac{L(L+1)}{2}\times(p-r)(L-r)} \qquad \frac{\partial vech\left(\Sigma\right)}{\left(\partial vech\left(\Sigma\right)\right)'} = I_{\frac{L(L+1)}{2}\times\frac{L(L+1)}{2}}$$

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

$$\begin{aligned} |J((\Pi, \Sigma) : (\beta_2, \alpha, \lambda, \Sigma))| &= |J| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \\ &= |A_{22}| \left| A_{11} - A_{12} A_{22}^{-1} A_{21} \right| = |A_{11}| \end{aligned}$$

where

$$A_{11} = \frac{\partial vec(\Pi)}{(\partial vec(\alpha))'} \quad \frac{\partial vec(\Pi)}{(\partial vec(\beta_2))'} \quad \frac{\partial vec(\Pi)}{(\partial vec(\lambda))'} \qquad A_{12} = \frac{\partial vec(\Pi)}{(\partial vech(\Sigma))'}$$

$$A_{21} = \frac{\partial vech(\Sigma)}{(\partial vec(\alpha))'} \quad \frac{\partial vech(\Sigma)}{(\partial vec(\beta_2))'} \quad \frac{\partial vech(\Sigma)}{(\partial vec(\lambda))'} \qquad A_{22} = \frac{\partial vech(\Sigma)}{(\partial vech(\Sigma))'}.$$

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

$$A_{11} = \frac{\partial vec(\Pi)}{(\partial vec(\alpha))'} \quad \frac{\partial vec(\Pi)}{(\partial vec(\beta_2))'} \quad \frac{\partial vec(\Pi)}{(\partial vec(\lambda))'}$$

where

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

and

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

from which it should be apparent $|J|_{\lambda=0}| = |A_{11}|_{\lambda=0}| \neq 0$.

8.1 The form of the information matrix for the structural parameters.

The information matrix in K&P for their invariant specification is block diagonal at the point $\lambda = 0$. This implies the posterior of α and β will be invariant. This relationship between the form of the information matrix ($\mathcal{I}_\theta$)

for the structural parameters of interest, $\theta = (\alpha, \beta, \lambda)$, and invariance provides a method to show that the specification for θ in this paper is invariant. Further, local identification for the parameters of the structural model is established via a link with the information matrix for the (globally identified) parameters of the reduced form model, π. We use the ECM model in (16) to establish invariance in a simple case.

Let J_{FR} and J_R be the Jacobians in K&P and this paper respectively. Next, let $\mathcal{I}_{\theta FR}$ and $\mathcal{I}_{\theta R}$ be the information matrices for θ in K&P and this paper respectively, and $\mathcal{I}_\pi = \Sigma^{-1} \otimes X'X$, the information matrix for the reduced form model in (16), which is equation (6) in K&P. For the ECM in (16) with specification (12) at $\lambda = 0$, the information matrix becomes

$$\begin{aligned} \mathcal{I}_{\theta R} &= J_R' \mathcal{I}_\pi J_R \\ &= \begin{bmatrix} \mathcal{I}_{\beta\beta R} & \mathcal{I}_{\beta\alpha R} & 0 & 0 \\ * & \mathcal{I}_{\alpha\alpha R} & 0 & 0 \\ * & * & \mathcal{I}_{\lambda\lambda R} & 0 \\ * & * & * & \mathcal{I}_{\Sigma\Sigma R} \end{bmatrix}, \end{aligned} \tag{25}$$

where

$$\begin{aligned} \mathcal{I}_{\beta\beta R} &= \Sigma^{-1} \otimes I_r, \\ \mathcal{I}_{\beta\alpha R} &= \left(\Sigma^{-1}\alpha' \otimes \beta' \begin{pmatrix} X_1'X_2 \\ X_2'X_2 \end{pmatrix}\right) + \left(\Sigma^{-1}\alpha' \otimes \beta' \begin{pmatrix} X_1'X_1 \\ X_1'X_2 \end{pmatrix}\right) \ddot{E}, \\ \mathcal{I}_{\alpha\alpha R} &= \left\{(I_r \otimes X_1)\ddot{E} + (I_r \otimes X_2)\right\}' \left(\alpha\Sigma^{-1}\alpha' \otimes I_T\right) \left\{(I_r \otimes X_1)\ddot{E} + (I_r \otimes X_2)\right\}, \\ \mathcal{I}_{\lambda\lambda R} &= \alpha_\perp \Sigma \alpha_\perp' \otimes \beta_\perp' (X'X)^{-1} \beta_\perp = I_{p-r} \otimes I_{p-r} \\ \mathcal{I}_{\Sigma\Sigma R} &= \frac{1}{2} D_L' \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) D_L. \end{aligned}$$

Thus $\mathcal{I}_{\theta R}$ is block diagonal at $\lambda = 0$ which establishes that (12) is an invariant specification. As discussed in the introduction, as $\mathcal{I}_{\theta R}$ is nonsingular, and π are globally identified as population moments, then θ is locally identified. Replacing J_R with J_{FR} and $\mathcal{I}_{\theta R}$ with $\mathcal{I}_{\theta FR}$, shows the specification in K&P is invariant, as shown in K&P, and θ in K&P is identified. Kleibergen and van Dijk (1994) present the Jeffreys' prior for a cointegrating ECM at rank r using the ordering restrictions discussed in the Introduction. Let I_θ be the information matrix this model but using the nonordering restrictions presented in this paper. If we develop the Jeffreys' prior for this model and denote this prior by $p_J(\theta) = |I_\theta|^{\frac{1}{2}}$, we find $p_J(\theta) = |I_\theta|^{\frac{1}{2}} = |I_{\theta R}|^{\frac{1}{2}}$.

9 Appendix II: Singular Value Decompositions of Nonsquare φ.

Here the SVD of the $(p \times L)$ matrix $\varphi^* = S_{11}^{-\frac{1}{2}} \varphi \Sigma^{\frac{1}{2}} = \psi\alpha + S_{11}^{-1}\psi_\perp \lambda \alpha_\perp \Sigma$ is presented. First consider the case where $L > p$. Let $\underset{p\times p}{U} = \left[\underset{p\times r}{U_1} \quad \underset{p\times(p-r)}{U_2} \right]$, $\underset{L\times L}{V} = \left[\underset{L\times r}{V_1} \quad \underset{L\times(p-r)}{V_2} \quad \underset{L\times(L-p)}{V_3} \right]$, 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,

$$\psi = S_{11}^{-\frac{1}{2}} U_1 T \qquad \alpha = T_1' \underline{S}_1 V_1' \Sigma^{\frac{1}{2}}.$$

Next, define the following matrices

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

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

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 (3)

$$S_{11}^{-1}\psi_\perp = S_{11}^{-\frac{1}{2}}U_2T_1' \qquad \Rightarrow \qquad \psi_\perp = S_{11}^{\frac{1}{2}}U_2T_1',$$

$$\lambda = T_1\underline{S}_2T_{21}',$$

and

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

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:

$$\begin{aligned} \varphi^* &= U\underline{S}V' = \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix}\begin{bmatrix} \underline{S}_1 & 0 \\ 0 & \underline{S}_2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \\ &= \begin{bmatrix} U_1T & U_2 & U_3 \end{bmatrix}\begin{bmatrix} T'\underline{S}_1 & 0 \\ 0 & \underline{S}_2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \\ &= U_1TT'\underline{S}_1V_1' + U_2\underline{S}_2V_2' \end{aligned} \tag{27}$$

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

$$T'U_1'S_{11}^{-\frac{1}{2}}DS_{11}^{-\frac{1}{2}}U_1T = \Gamma = diag(\gamma_1, ..., \gamma_r)$$

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

$$\psi = S_{11}^{-\frac{1}{2}}U_1T \qquad \alpha = T_1'\underline{S}_1V_1'\Sigma^{\frac{1}{2}}.$$

Next, define the following matrices

$$\begin{aligned} T_1'\begin{bmatrix} U_2' \\ U_3' \end{bmatrix}S_{11}^{-1}\begin{bmatrix} U_2 & U_3 \end{bmatrix}T_1 &= \Lambda_{p-r} = diag(\eta_1, ..., \eta_{p-r}) \\ T_2'V_2'S_{11}^{-\frac{1}{2}}S_{10}S_{00}^{-1}S_{01}S_{11}^{-\frac{1}{2}}V_2T_2 &= \Lambda_{L-r} = diag(\varsigma_1, ..., \varsigma_{L-r}) \end{aligned}$$

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 $(L-r)\times(L-r)$.

The last term in (27) becomes,

$$U_2\underline{S}_2V_2' = \begin{bmatrix} U_2 & U_3 \end{bmatrix} T_1'T_{11}\underline{S}_2T_2T_2'V_2'$$

so that from (3)

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

$$\lambda = T_{11}\underline{S}_2T_2,$$

and

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

10 Appendix III: Decomposition of the Exponent in the Posterior.

Recall the model in (10):

$$\begin{aligned} Y &= X\Pi + ZA + E = \underline{X}B + E \\ &= X\beta\alpha + XS_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma + ZA + E \end{aligned}$$

where $\underline{X} = [X \quad Z]$ and $B = [\Pi' \quad A']' = \left[(\beta\alpha + S_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma)' \quad A'\right]'$. Using the transformation (3), for the natural conjugate prior the form of the trace in the exponent in the posterior is

$$\begin{aligned} & tr\Sigma^{-1}\left[(Y - \underline{X}B)'(Y - \underline{X}B) + \underline{S} + (B - \underline{B})'\underline{H}(B - \underline{B})\right] \qquad (28) \\ = \; & tr\Sigma^{-1}\left[\underline{S} - \widetilde{B}'(\underline{H} + \underline{X}'\underline{X})\widetilde{B} + \underline{B}'\underline{H}\underline{B} + Y'Y + T\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}\right] \\ & -Ttr\alpha_\perp\widetilde{\Pi}'\beta_\perp\beta_\perp'\widetilde{\Pi}\alpha_\perp - Ttr\Sigma^{-1}\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10} \\ & +Ttr\Sigma^{-1}(\alpha - \widehat{\alpha})'(\alpha - \widehat{\alpha}) \\ & +Ttr\left(\lambda - \widetilde{\lambda}\right)'\left(\lambda - \widetilde{\lambda}\right) \\ & +tr\Sigma^{-1}(A - \Delta)'(\underline{H}_{22} + Z'Z)(A - \Delta) \end{aligned}$$

where

$$\widetilde{B} = (\underline{H} + \underline{X}'\underline{X})^{-1}(\underline{H}\underline{B} + \underline{X}'Y) = \left[\widetilde{\Pi}' \quad \widetilde{A}'\right]'$$

$$
\begin{aligned}
\widetilde{\Pi} &= \widetilde{S}_{11}^{-1}\widetilde{S}_{10} \\
\widetilde{\alpha} &= \beta'\widetilde{S}_{10} \\
\widetilde{\lambda} &= \left(\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\right)^{-1}\beta_\perp'\widetilde{\Pi}\alpha_\perp'\left(\alpha_\perp\Sigma\alpha_\perp'\right)^{-1} = \beta_\perp'\widetilde{\Pi}\alpha_\perp' \\
\Delta &= \widetilde{A} - \left(\underline{H}_{22} + Z'Z\right)^{-1}\left(\underline{H}_{21} + Z'X\right)\left(\Pi - \widetilde{\Pi}\right) \\
\tilde{S} &= \underline{S} - \widetilde{B}'\left(\underline{H} + \underline{X}'\underline{X}\right)\widetilde{B} + \underline{B}'\underline{H}\underline{B} + Y'Y \\
M_{22} &= \underline{H}_{22} + Z'Z \qquad M_{20} = \underline{H}_2\underline{B} + Z'Y \qquad M_{21} = \underline{H}_{21} + Z'X \\
R_0 &= Y - ZM_{22}^{-1}M_{20} \qquad R_1 = X - ZM_{22}^{-1}M_{21} \\
T\widetilde{S}_{ij} &= T\widetilde{S}_{ji}' = R_i'R_j \\
\underline{H} &= \begin{bmatrix} \underline{H}_1 \\ \underline{H}_2 \end{bmatrix} = \begin{bmatrix} \underline{H}_{11} & \underline{H}_{12} \\ \underline{H}_{21} & \underline{H}_{22} \end{bmatrix} \\
\underline{B} &= [\underline{\Pi}' \quad \underline{A}']'.
\end{aligned}
$$

The form of the trace in the exponent in the posterior for the diffuse prior can be found by setting $\underline{H} = 0$ and $\underline{S} = 0$. Finally, the model with $\underline{X} = X$ and $B = \Pi$ is retrieved by ignoring the last line in (28) and making the following changes.

$$
\begin{aligned}
\widetilde{B} &= \widetilde{\Pi} \qquad\qquad \widetilde{\alpha} = \beta'\widetilde{S}_{10} \\
T\widetilde{S}_{10} &= T\widetilde{S}_{01}' = \underline{H}\underline{B} + X'Y \\
T\widetilde{S}_{11} &= \underline{H} + X'X \qquad \underline{B} = \underline{\Pi}.
\end{aligned}
$$

Proof:

$$
\begin{aligned}
& (Y - \underline{X}B)'(Y - \underline{X}B) + \underline{S} + (B - \underline{B})'\underline{H}(B - \underline{B}) \\
= \ & \underline{S} - \widetilde{B}'\left(\underline{H} + \underline{X}'\underline{X}\right)\widetilde{B} + \underline{B}'\underline{H}\underline{B} + Y'Y + \left(B - \widetilde{B}\right)'\left(\underline{H} + \underline{X}'\underline{X}\right)\left(B - \widetilde{B}\right) \\
= \ & \tilde{S} + T\left(\Pi - \widetilde{\Pi}\right)'\widetilde{S}_{11}\left(\Pi - \widetilde{\Pi}\right) + (A - \Delta)'\left(\underline{H}_{22} + Z'Z\right)(A - \Delta)
\end{aligned}
$$

As A does not appear in the Jacobian, we can see from the above expression that the distribution of $vec(A)$ conditional on $(\alpha, \beta, \lambda, \Sigma)$ will be $N\left(vec(\Delta), \Sigma \otimes \left(\underline{H}_{22} + Z'Z\right)^{-1}\right)$. As setting $\lambda = 0$ changes only the location of its distribution, A can be ignored in the sampling scheme and integrated out of the posterior so we can focus on inference directly from the marginal for $(\alpha, \beta, \lambda, \Sigma)$, which is the parameter set usually of interest, using the following results.

$$
\begin{aligned}
& \left(\Pi-\widetilde{\Pi}\right)' \widetilde{S}_{11}\left(\Pi-\widetilde{\Pi}\right) \\
= & \left(\beta\alpha+\widetilde{S}_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma-\widetilde{\Pi}\right)' \widetilde{S}_{11}\left(\beta\alpha+\widetilde{S}_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma-\widetilde{\Pi}\right) \\
= & \left(\beta\alpha-\widetilde{\Pi}\right)' \widetilde{S}_{11}\left(\beta\alpha-\widetilde{\Pi}\right) \\
& +\left(\widetilde{S}_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma-\widetilde{\Pi}\right)' \widetilde{S}_{11}\left(\widetilde{S}_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma-\widetilde{\Pi}\right)-\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}
\end{aligned}
$$

$$
\begin{aligned}
& \left(\beta\alpha-\widetilde{\Pi}\right)' \widetilde{S}_{11}\left(\beta\alpha-\widetilde{\Pi}\right) \\
= & \left((\beta\alpha-\beta\widetilde{\alpha})+\left(\beta\widetilde{\alpha}-\widetilde{\Pi}\right)\right)' \widetilde{S}_{11}\left((\beta\alpha-\beta\widetilde{\alpha})+\left(\beta\widetilde{\alpha}-\widetilde{\Pi}\right)\right) \\
= & (\alpha-\widetilde{\alpha})'\beta'\widetilde{S}_{11}\beta(\alpha-\widetilde{\alpha})+\left(\beta\widetilde{\alpha}-\widetilde{\Pi}\right)' \widetilde{S}_{11}(\beta\alpha-\beta\widetilde{\alpha}) \\
& \quad +(\beta\alpha-\beta\widetilde{\alpha})' \widetilde{S}_{11}\left(\beta\widetilde{\alpha}-\widetilde{\Pi}\right)+\left(\beta\widetilde{\alpha}-\widetilde{\Pi}\right)' \widetilde{S}_{11}\left(\beta\widetilde{\alpha}-\widetilde{\Pi}\right) \\
= & (\alpha-\widetilde{\alpha})'(\alpha-\widetilde{\alpha}) \\
& +\widetilde{\alpha}'\beta'\widetilde{S}_{11}\beta\alpha-\widetilde{\alpha}'\beta'\widetilde{S}_{11}\beta\widetilde{\alpha}-\widetilde{\Pi}'\widetilde{S}_{11}\beta\alpha+\widetilde{\Pi}'\widetilde{S}_{11}\beta\widetilde{\alpha} \\
& +\alpha'\beta'\widetilde{S}_{11}\beta\widetilde{\alpha}-\alpha'\beta'\widetilde{S}_{11}\widetilde{\Pi}-\widetilde{\alpha}'\beta'\widetilde{S}_{11}\beta\widetilde{\alpha}+\widetilde{\alpha}'\beta'\widetilde{S}_{11}\widetilde{\Pi} \\
& +\widetilde{\alpha}'\beta'\widetilde{S}_{11}\beta\widetilde{\alpha}-\widetilde{\alpha}'\beta'\widetilde{S}_{11}\widetilde{\Pi}-\widetilde{\Pi}'\widetilde{S}_{11}\beta\widetilde{\alpha}+\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi} \\
= & (\alpha-\widetilde{\alpha})'(\alpha-\widetilde{\alpha}) \\
& +\widetilde{S}_{01}\beta\alpha-\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10}-\widetilde{S}_{01}\beta\alpha+\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10} \\
& +\alpha'\beta'\widetilde{S}_{10}-\alpha'\beta'\widetilde{S}_{10}-\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10}+\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10} \\
& +\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10}-\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10}-\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10}+\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi} \\
= & (\alpha-\widetilde{\alpha})'(\alpha-\widetilde{\alpha})-\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10}+\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}.
\end{aligned}
$$

Next, we isolate the terms involving λ.

$$
\begin{aligned}
& \left(\widetilde{S}_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma-\widetilde{\Pi}\right)' \widetilde{S}_{11}\left(\widetilde{S}_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma-\widetilde{\Pi}\right) \\
= & \left(\widetilde{S}_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma-\widetilde{S}_{11}^{-1}\beta_{\perp}\widetilde{\lambda}\alpha_{\perp}\Sigma+\widetilde{S}_{11}^{-1}\beta_{\perp}\widetilde{\lambda}\alpha_{\perp}\Sigma-\widetilde{\Pi}\right)' \\
& \widetilde{S}_{11}\left(\widetilde{S}_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma-\widetilde{S}_{11}^{-1}\beta_{\perp}\widetilde{\lambda}\alpha_{\perp}\Sigma+\widetilde{S}_{11}^{-1}\beta_{\perp}\widetilde{\lambda}\alpha_{\perp}\Sigma-\widetilde{\Pi}\right) \\
= & \Sigma\alpha_{\perp}'\left(\lambda-\widetilde{\lambda}\right)'\beta_{\perp}'\widetilde{S}_{11}^{-1}\beta_{\perp}\left(\lambda-\widetilde{\lambda}\right)\alpha_{\perp}\Sigma
\end{aligned}
$$

$$
\begin{aligned}
& +\left(\widetilde{S}_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma-\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma\right)'\widetilde{S}_{11}\left(\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma-\widetilde{\Pi}\right)\\
& +\left(\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma-\widetilde{\Pi}\right)'\widetilde{S}_{11}\left(\widetilde{S}_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma-\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma\right)\\
& +\left(\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma-\widetilde{\Pi}\right)'\widetilde{S}_{11}\left(\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma-\widetilde{\Pi}\right)\\
= & \ \Sigma\alpha_\perp'\left(\lambda-\widetilde{\lambda}\right)'\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\left(\lambda-\widetilde{\lambda}\right)\alpha_\perp\Sigma\\
& +\Sigma\alpha_\perp'\lambda'\beta_\perp'\widetilde{S}_{11}^{-1}\widetilde{S}_{11}\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma-\Sigma\alpha_\perp'\lambda'\beta_\perp'\widetilde{S}_{11}^{-1}\widetilde{S}_{11}\widetilde{\Pi}\\
& -\Sigma\alpha_\perp'\widetilde{\lambda}'\beta_\perp'\widetilde{S}_{11}^{-1}\widetilde{S}_{11}\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma+\Sigma\alpha_\perp'\widetilde{\lambda}'\beta_\perp'\widetilde{S}_{11}^{-1}\widetilde{S}_{11}\widetilde{\Pi}\\
& +\Sigma\alpha_\perp'\widetilde{\lambda}'\beta_\perp'\widetilde{S}_{11}^{-1}\widetilde{S}_{11}\widetilde{S}_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma\\
& \qquad -\Sigma\alpha_\perp'\widetilde{\lambda}'\beta_\perp'\widetilde{S}_{11}^{-1}\widetilde{S}_{11}\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma\\
& -\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{S}_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma+\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma\\
& +\Sigma\alpha_\perp'\widetilde{\lambda}'\beta_\perp'\widetilde{S}_{11}^{-1}\widetilde{S}_{11}\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma-\Sigma\alpha_\perp'\widetilde{\lambda}'\beta_\perp'\widetilde{S}_{11}^{-1}\widetilde{S}_{11}\widetilde{\Pi}\\
& -\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma+\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}\\
= & \ \Sigma\alpha_\perp'\left(\lambda-\widetilde{\lambda}\right)'\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\left(\lambda-\widetilde{\lambda}\right)\alpha_\perp\Sigma\\
& +\Sigma\alpha_\perp'\lambda'\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma-\Sigma\alpha_\perp'\lambda'\beta_\perp'\widetilde{\Pi}\\
& -\Sigma\alpha_\perp'\widetilde{\lambda}'\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\widetilde{\lambda}\alpha_\perp\Sigma+\Sigma\alpha_\perp'\widetilde{\lambda}'\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma\\
& -\widetilde{\Pi}'\beta_\perp\lambda\alpha_\perp\Sigma+\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}\\
= & \ \Sigma\alpha_\perp'\left(\lambda-\widetilde{\lambda}\right)'\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\left(\lambda-\widetilde{\lambda}\right)\alpha_\perp\Sigma\\
& +\Sigma\alpha_\perp'\lambda'\beta_\perp'\widetilde{\Pi}\alpha_\perp'\left(\alpha_\perp\Sigma\alpha_\perp'\right)^{-1}\alpha_\perp\Sigma-\Sigma\alpha_\perp'\lambda'\beta_\perp'\widetilde{\Pi}\\
& -\Sigma\alpha_\perp'\left(\alpha_\perp\Sigma\alpha_\perp'\right)^{-1}\alpha_\perp\widetilde{\Pi}'\beta_\perp\left(\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\right)^{-1}\beta_\perp'\widetilde{\Pi}\alpha_\perp'\left(\alpha_\perp\Sigma\alpha_\perp'\right)^{-1}\alpha_\perp\Sigma\\
& +\Sigma\alpha_\perp'\left(\alpha_\perp\Sigma\alpha_\perp'\right)^{-1}\alpha_\perp\widetilde{\Pi}'\beta_\perp\lambda\alpha_\perp\Sigma-\widetilde{\Pi}'\beta_\perp\lambda\alpha_\perp\Sigma+\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}.
\end{aligned}
$$

Using the fact that the above expressions appear in the exponent premultiplied by Σ^{-1} and as a trace, we can simplify further as follows.

$$
\begin{aligned}
& tr\Sigma^{-1}\left(\widetilde{S}_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma-\widetilde{\Pi}\right)'\widetilde{S}_{11}\left(\widetilde{S}_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma-\widetilde{\Pi}\right)\\
= & \ tr\alpha_\perp\Sigma\alpha_\perp'\left(\lambda-\widetilde{\lambda}\right)'\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\left(\lambda-\widetilde{\lambda}\right)\\
& +tr\lambda'\beta_\perp'\widetilde{\Pi}\alpha_\perp'-tr\lambda'\beta_\perp'\widetilde{\Pi}\alpha_\perp'\\
& -tr\left(\alpha_\perp\Sigma\alpha_\perp'\right)^{-1}\alpha_\perp\widetilde{\Pi}'\beta_\perp\left(\beta_\perp'\widetilde{S}_{11}^{-1}\beta_\perp\right)^{-1}\beta_\perp'\widetilde{\Pi}\alpha_\perp'
\end{aligned}
$$

$$
\begin{aligned}
& +tr\alpha_\perp\widetilde{\Pi}'\beta_\perp\lambda - tr\alpha_\perp\widetilde{\Pi}'\beta_\perp\lambda + tr\Sigma^{-1}\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi} \\
= \; & tr\left(\lambda - \widetilde{\lambda}\right)'\left(\lambda - \widetilde{\lambda}\right) \\
& -tr\alpha_\perp\widetilde{\Pi}'\beta_\perp\beta'_\perp\widetilde{\Pi}\alpha_\perp + tr\Sigma^{-1}\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}
\end{aligned}
$$

which shows that the form of the distribution of $vec(\lambda)$ conditional on (α, β, Σ) will be $N(0, I_{(p-r)(L-r)})$ multiplied by the Jacobian.

Therefore

$$
\begin{aligned}
& tr\Sigma^{-1}\left(\Pi - \widetilde{\Pi}\right)'\widetilde{S}_{11}\left(\Pi - \widetilde{\Pi}\right) \\
= \; & tr\left(\lambda - \widetilde{\lambda}\right)'\left(\lambda - \widetilde{\lambda}\right) + tr\Sigma^{-1}\left(\alpha - \widetilde{\alpha}\right)'\left(\alpha - \widetilde{\alpha}\right) \\
& -tr\alpha_\perp\widetilde{\Pi}'\beta_\perp\beta'_\perp\widetilde{\Pi}\alpha'_\perp - tr\Sigma^{-1}\widetilde{S}_{01}\beta\beta'\widetilde{S}_{10} \\
& +tr\Sigma^{-1}\widetilde{\Pi}'\widetilde{S}_{11}\widetilde{\Pi}.
\end{aligned}
$$

Table 1: *LR statistics and posterior probabilities for ranks.*

Rank (r)	$LR(H_r \mid H_3)$	5% Critical value	$P(H_r \mid y)$
0	33.04	29.68	0.00
1	14.29	15.41	0.03
2	1.972	3.76	0.96
3	-	-	0.01

Table 2: *LR statistics and posterior probabilities of H_4 given H_r.*

Rank (r)	$LR(H_4 \mid H_r)$	p-value	$P(H_4 \mid H_r, y)$
0	-		0.50
1	0.02	0.90	0.00
2	0.85	0.66	0.98
3	-	-	0.00
			$P(H_4 \mid y) = 0.93$

Table 3: *LR statistics and posterior probabilities of H_5 given H_r.*

Rank (r)	$LR(H_5 \mid H_r)$	p-value	$P(H_5 \mid H_r, y)$
0	-	-	0.00
1	20.51	0.00	1.00
2	2.66	0.10	0.00
3	-	-	0.00
			$P(H_5 \mid y) = 0.03$

Table 4: *LR statistics and posterior probabilities of H_6 given H_r.*

Rank (r)	$LR(H_6 \mid H_r)$	p-value	$P(H_6 \mid H_r, y)$
0	-	-	0.00
1	14.31	0.00	1.00
2	1.98	0.16	0.00
3	-	-	0.00
			$P(H_6 \mid y) = 0.03$

Table 5: *LR statistics and posterior probabilities of H_7 given H_r.*

Rank (r)	$LR(H_7 \mid H_r)$	p-value	$P(H_7 \mid H_r, y)$
0	-	-	0.00
1	5.66	0.02	0.00
2	15.73	0.00	0.00
3	-	-	0.00
			$P(H_7 \mid y) = 0.00$

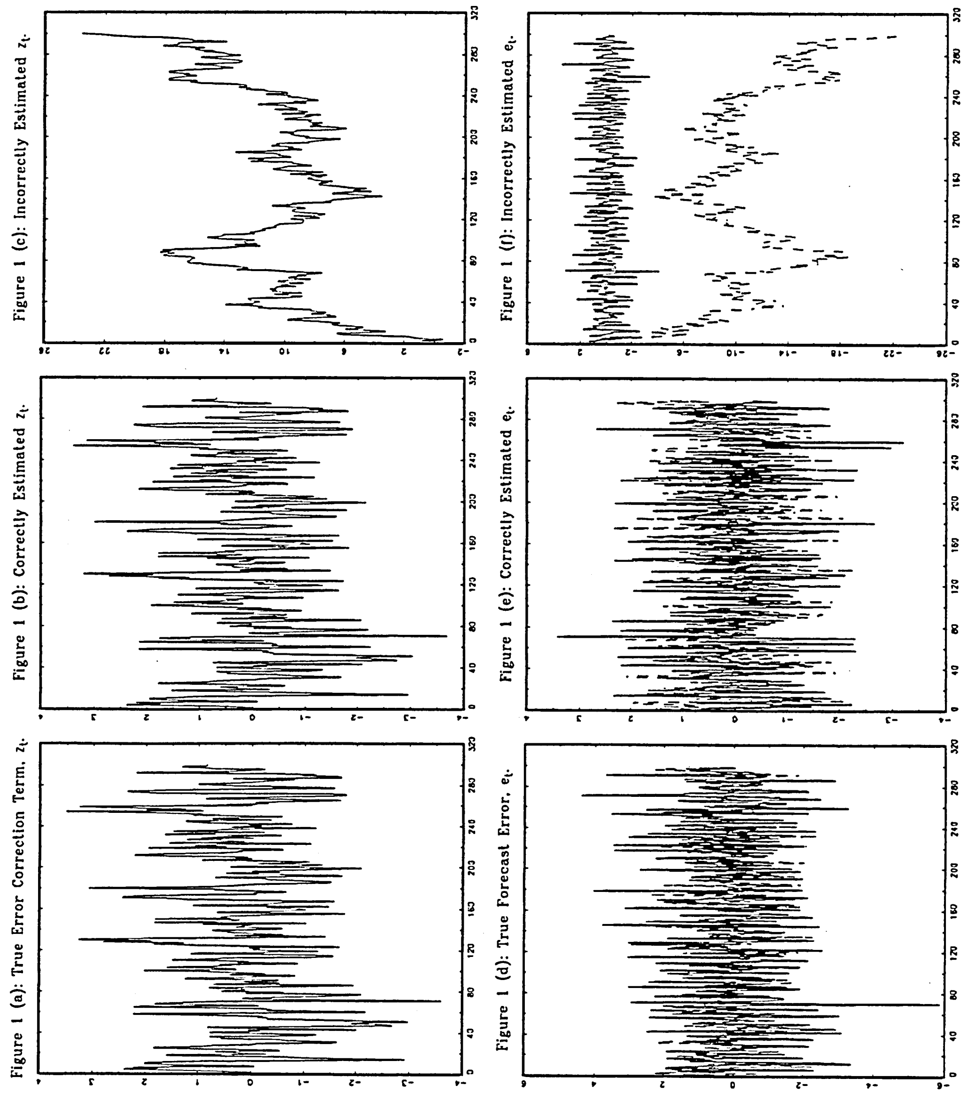

Figure 1: (a) to Figure 1: (f) - Actual and estimated z_t and e_t for the model (2) with DGP $\alpha = (-0.5\ 0)\,, \beta = (1\ 0)'$ and $e_t \sim N\,(0, I_2)$. The first 150 observations are 'in sample' in the sense that they were used to estimate α, β and e_t. The remaining 150 observations are 'out of sample'.

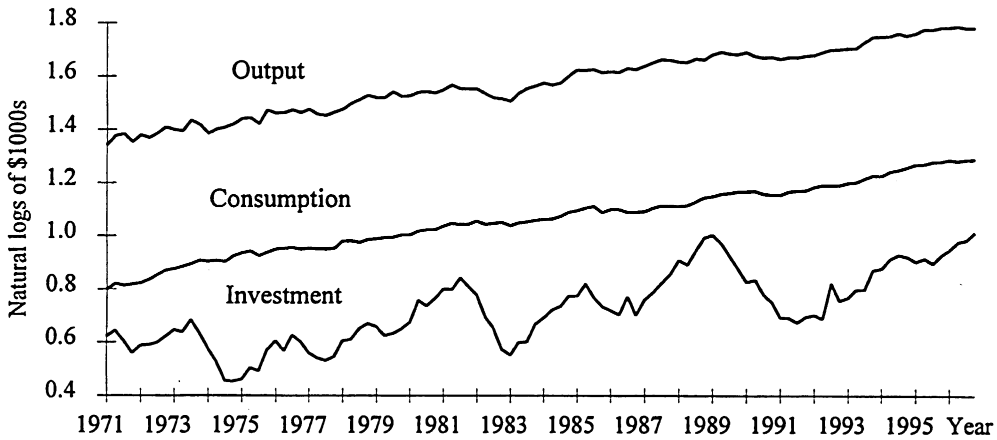

Figure 2: Logs of seasonally adjusted, real private per capita consumption (c_t), investment (i_t) and output (y_t). The data are obtained from the Australian dX database and the series identifiers and construction are detailed in Harris (1997, Figure 1). In this figure, 0.9 has been added to i_t for presentation purposes.