ISSN 1440-771X
ISBN 0 7326 1068 0

M O N A S H U N I V E R S I T Y

AUSTRALIA

Bayesian Trace Statistics for the Reduced Rank Regression Model

Rodney W. Strachan and Brett Inder

Working Paper 13/99
October 1999

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Bayesian Trace Statistics for the Reduced Rank Regression Model.*

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

Key Words: reduced rank regression model; error correction model; marginal likelihoods; Bayes factors; Bayesian analysis.

ABSTRACT

Estimation of the reduced rank regression model requires restrictions be imposed upon the model. Two forms of restrictions are commonly used. Earlier Bayesian work relied on the triangular method of identification which imposes an *a priori* ordering on the variables in the system, however, incorrect ordering of the variables can result in model misspecification. Bayesian estimation of the reduced rank regression model without ordering restrictions was presented in Strachan (1998) and follows the classical approach of Anderson (1951) and Johansen (1988). This method of estimation avoids placing restrictions on the space spanned by the reduced rank relations and simplifies testing of restrictions on that space. In this paper, a method for estimating approximate marginal likelihoods and Bayes factors is presented for

*This work is based on work from the first author's Ph.D. thesis.

this model, using Laplace approximations for integrals. These Bayes factors algebraically resemble the Johansen trace statistic (1995), hence the title. We consider the model with rank r and no restrictions on the reduced rank relations.

1 Introduction.

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). Estimation of this model requires restrictions be imposed upon the model. Two forms of restrictions are commonly used. The first form of restrictions is the triangular system (see for example Kleibergen and van Dijk 1994 and Geweke 1996), which imposes an *a priori* ordering on the variables in the system. Incorrect ordering of the variables can result in model misspecification as discussed in Strachan (1988). The second form of restrictions avoid this potential misspecification and was presented by Anderson (1951) and applied to the cointegrating ECM by Johansen (1988).

Earlier Bayesian work in this model relied on the triangular method of identification while Bayesian estimation of the reduced rank regression model without ordering restrictions was presented in Strachan (1998). This second method of estimation avoids placing restrictions on the space spanned by the reduced rank relations, or, in other words, without restrictions that impose an *a priori* ordering on the variables. They have the additional advantage that they simplify testing of restrictions on this space In this paper, a method for estimating approximate marginal likelihoods and Bayes factors is presented for this model, using Laplace approximations for integrals. These Bayes factors algebraically resemble the Johansen trace statistic (1995), hence the title. We consider the model with rank r and no restrictions on the reduced rank relations. Later work will consider the model with rank r and with various restrictions on the reduced rank relations.

2 The model.

The reduced rank multivariate regression model is,

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

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 ε 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 $rank\,(\alpha) = rank\,(\beta) = r$. A familiar example of the reduced rank regression model is the cointegrating error correction model in which β is the matrix of r cointegrating vectors and α is the matrix of adjustment coefficients. 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. We define the matrices S_{ij} for $i, j = 0, 1, 2$ used below according to the chosen prior in Section 3. For our purposes here, these matrices are composed of data and prior values. The identifying restrictions are all imposed on β in the following normalisations,

$$\beta' S_{11} \beta = I \tag{2}$$

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

$$\beta' S_{10} S_{00}^{-1} S_{01} \beta = \Gamma = diag(\gamma_1, ..., \gamma_r) \tag{3}$$

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 method of estimation used in this paper, the parameter β is not directly estimated. Rather we identify and estimate functions of the vectors[(1)] in β, and, specifically, we estimate the basis for the space spanned by β, $sp(\beta)$. As we estimate the basis for this space, none of the space is excluded by the form of the restrictions used. This approach contrasts with that used in the triangular system in which $\beta = [\beta_1' \; \beta_2']'$ is identified prior to estimation by imposing the restriction

[1] The author is grateful to Trevor Breusch for discussion on this point.

$\beta\beta_1^{-1} = \beta^* = [I_r\ \beta_2']'$, which assumes β_1^{-1} exists and excludes the space in which β_1 is singular.

In the expression (3), Γ is random with an *implied* posterior distribution in the Bayesian method and $\gamma_1 > ... > \gamma_r > 0$. Using the transformation presented in Strachan (1998), which is based on the approach taken in K&P, we let

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

the resultant model is now

$$Y = X\beta\alpha + XS_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma + ZA + \varepsilon \tag{4}$$

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

$$Y = X\beta\alpha + ZA + \varepsilon. \tag{5}$$

3 Priors and Posteriors.

In this section we present the forms of the posterior for the general reduced rank regression model with a diffuse prior and then with an informative prior. As was found in K&P, the following priors and posteriors cannot be decomposed to provide marginal and conditional densities useful for sampling. For the rest of the paper we will identify scalars, vectors and matrices associated with the diffuse prior with a caret (for example $\widehat{\beta}$), those associated with the informative prior with a tilde ($\widetilde{\beta}$), and when we wish to use general results that could come from either prior we use a bar ($\overline{\beta}$).

3.1 The likelihood.

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

$$L\left(Y|\Sigma, \beta, \alpha, \lambda, A, \underline{X}\right) = L\left(Y|\theta, \underline{X}\right) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}\varepsilon'\varepsilon\right)\right\} \tag{6}$$

where

$$\varepsilon = Y - X\beta\alpha - XS_{11}^{-1}\beta_\perp\lambda\alpha_\perp\Sigma - ZA.$$

3.2 Priors: A diffuse prior.

Let $\mathcal{I}_\pi$ be the information matrix for the parameters in the reduced form model (1), $\pi = vec(\Pi, A, \Sigma)$. The Jeffreys prior is proportional to $|\mathcal{I}_\pi|^{1/2}$ and therefore the form of the diffuse (d) or Jeffreys prior for the parameters in (1) is:

$$p_\pi(\pi)_d \propto |\mathcal{I}_\pi|^{1/2} \propto 2^{-\frac{L}{2}} |\underline{X}'\underline{X}|^{-\frac{L}{2}} |\Sigma|^{-(L+p+k+1)/2}. \qquad (7)$$

Since $\mathcal{I}_\pi$ is the negative of the expectation of the inverse of the Hessian for (6), the expression in (7) will be valid only for nonrandom $\underline{X}$ such that $E[\underline{X}'\underline{X}] = \underline{X}'\underline{X}$. An important example of the reduced rank regression model is the cointegrating error correction model for which X is random. Kleibergen and van Dijk (1994) present several approaches to approximating $\mathcal{I}_\pi$ in this model. We take the first and most extensively investigated approach suggested by Kleibergen and van Dijk (1994), that is we assume $\underline{X}'\underline{X}$ is a reasonable approximation to the expectation $E[\underline{X}'\underline{X}]$. Next we define the S_{ij} matrices for the diffuse prior, $\widehat{S}_{ij}$. For the following expression for the posterior for the diffuse prior, let $\widehat{\kappa} = T + L + p + k + 1$, $Z_0 = Y$, $Z_1 = X$ and $Z_2 = Z$.

$\widehat{\kappa}\widehat{M}_{ij} = Z_i'Z_j$ for $(i,j) = (1,1)(1,2)(2,2)$,

$\widehat{\kappa}\widehat{M}_{10} = Z_1'Z_0$, $\widehat{\kappa}\widehat{M}_{20} = Z_2'Z_0$, and

$\widehat{\kappa}\widehat{M}_{00} = Z_0'Z_0$.

Finally let $\widehat{S}_{ij} = \widehat{M}_{ij} - \widehat{M}_{i2}\widehat{M}_{22}^{-1}\widehat{M}_{2j}$, $\widehat{S}_{ij} = \widehat{S}_{ji}'$ and $\widehat{M}_{ij} = \widehat{M}_{ji}'$.

3.3 Priors: An informative prior.

The natural conjugate (i) prior for the unrestricted linear model in (1) is:

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

where $\underline{B} = [\underline{\Pi}' \quad \underline{A}']'$ and $\underline{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$. Next we define the S_{ij} matrices for the informative prior, $\widetilde{S}_{ij}$. Let $\widetilde{\kappa} = T + L + p + k + \underline{\nu} + 1$. As for the diffuse prior, let $Z_0 = Y$, $Z_1 = X$ and $Z_2 = Z$,

$\widetilde{\kappa}\widetilde{M}_{ij} = Z_i'Z_j + h_{ij}$ for $(i,j) = (1,1)(1,2)(2,2)$,

$\widetilde{\kappa}\widetilde{M}_{10} = Z_1'Z_0 + h_{11}\underline{\Pi} + h_{12}\underline{A}$,
$\widetilde{\kappa}\widetilde{M}_{20} = Z_2'Z_0 + h_{21}\underline{\Pi} + h_{22}\underline{A}$, and
$\widetilde{\kappa}\widetilde{M}_{00} = Z_0'Z_0 + \underline{B}'\underline{HB} + \underline{S}$.

Finally let $\widetilde{S}_{ij} = \widetilde{M}_{ij} - \widetilde{M}_{i2}\widetilde{M}_{22}^{-1}\widetilde{M}_{2j}$, $\widetilde{S}_{ij} = \widetilde{S}_{ji}'$ and $\widetilde{M}_{ij} = \widetilde{M}_{ji}'$. Using the expression for the Jacobian in Appendix I, J, the resultant priors for the structural parameters $(\theta, \lambda) = (\alpha, \beta, A, \Sigma, \lambda)$ in the transformed regression model (4) are

$$p_\theta(\theta, \lambda)_d = p_\pi(\pi(\theta, \lambda))_d |J| \propto |J\mathcal{I}_\pi J'|^{1/2} \propto |\mathcal{I}_{\theta,\lambda}|^{1/2} \tag{8}$$

where $\mathcal{I}_\theta$ is the information matrix for the parameters in (4), and

$$p_\theta(\theta, \lambda)_i = p_\pi(\pi(\theta, \lambda))_n |J|. \tag{9}$$

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 $(\alpha, \beta, A, \Sigma, \lambda = 0)$.

3.4 The posterior for the diffuse prior.

The posterior for model (1) is:

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

and where $\widehat{B} = \begin{bmatrix} \widehat{\Pi} \\ \widehat{A} \end{bmatrix} = \left(\underline{X}'\underline{X}\right)^{-1}\underline{X}'Y$ and $S = Y'Y - \widehat{B}'\left(\underline{X}'\underline{X}\right)^{-1}\widehat{B}$. The posterior for the transformed model as parameterised in (4) with the diffuse prior may be decomposed as follows,

$$\begin{aligned} p_\theta(\theta, \lambda|Y, \underline{X})_d &= p_\pi(\pi(\theta, \lambda)|Y, \underline{X})_d |J| \\ &= p(\pi(\theta, \lambda))_d L(Y|\theta, \lambda, \underline{X}) |J| \end{aligned}$$

Therefore, posterior for the reduced rank model where $\lambda = 0$ and we use the diffuse prior may be expressed as

$$\begin{aligned} p_\theta(\theta|Y, \lambda = 0, \underline{X})_d &= p_\pi(\pi(\theta, \lambda)|Y, \lambda = 0, \underline{X})_d |J|_{\lambda=0}| \\ &= p_\pi(\pi(\theta, \lambda)|\lambda = 0)_d L(Y|\theta, \lambda = 0, \underline{X}) |J|_{\lambda=0}| \\ &\propto \exp(-\widehat{\kappa}h_d(\theta)) |J|_{\lambda=0}| \end{aligned}$$

where $h_d(\theta)$ is defined in Appendix I. At the mode of the diffuse prior posterior conditional on $\lambda = 0$ (ie., at rank r), $\alpha = \widehat{\alpha} = \widehat{\beta} S_{10}$, and $\beta = \widehat{\beta} = \widehat{S}_{11}^{-\frac{1}{2}} \widehat{U}_r$ where $\widehat{U} = \begin{bmatrix} \widehat{U}_r & \widehat{U}_{p-r} \end{bmatrix}$ are the p eigenvectors of $\widehat{S}_{11}^{-\frac{1}{2}} \widehat{S}_{10} \widehat{S}_{00}^{-1} \widehat{S}_{10} \widehat{S}_{11}^{-\frac{1}{2}}$ with p eigenvalues $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \ldots > \widehat{\lambda}_p > 0$. The columns of $\widehat{U}$ are ordered according to the size of their associated eigenvalue. Further, $\Sigma = \widehat{\Sigma} = \widehat{S}_{00} - \widehat{S}_{01} \widehat{\beta} \left(\widehat{\beta} \widehat{S}_{11} \widehat{\beta} \right)^{-1} \widehat{\beta} \widehat{S}_{10}$. Therefore we may write

$$\begin{aligned} \exp\left(-\widehat{\kappa} h\left(\widehat{\theta}\right)\right) &= \left|\widehat{\Sigma}\right|^{-\frac{\widehat{\kappa}}{2}} \exp\left\{-\frac{L\widehat{\kappa}}{2}\right\} \\ &= \left[\left|\widehat{S}_{00}\right| \prod_{i-1}^{r} (1 - \widehat{\lambda}_i)\right]^{-\frac{\widehat{\kappa}}{2}} \exp\left\{-\frac{L\widehat{\kappa}}{2}\right\}. \end{aligned} \tag{10}$$

3.5 The posterior for the informative prior:

The posterior for model (1) with an informative prior is:

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

and where $\widetilde{B} = \begin{bmatrix} \widetilde{\Pi} \\ \widetilde{A} \end{bmatrix} = (\underline{H} + \underline{X}'\underline{X})^{-1} (\underline{HB} + \underline{X}'Y)$ and $S = Y'Y + \underline{S} + \underline{B}'\underline{HB} - \widetilde{B}' (\underline{H} + \underline{X}'\underline{X})^{-1} \widetilde{B}$. Therefore, posterior for the reduced rank model where $\lambda = 0$ and we use the informative prior may be expressed as

$$\begin{aligned} p_\theta (\theta | Y, \lambda = 0, \underline{X})_i &= p_\pi (\pi(\theta, \lambda) | Y, \lambda = 0, \underline{X})_i \, |J|_{\lambda=0}| \\ &= p_\pi (\pi(\theta, \lambda) | \lambda = 0)_i \, L(Y | \theta, \lambda = 0, \underline{X}) \, |J|_{\lambda=0}| \\ &\propto \exp(-\widetilde{\kappa} h_i(\theta)) \, |J|_{\lambda=0}| \end{aligned}$$

where $h_i(\theta)$ is again defined in Appendix I. At the mode of the diffuse prior posterior conditional on $\lambda = 0$ (ie., at rank r), $\alpha = \widetilde{\alpha} = \widetilde{\beta} \widetilde{S}_{10}$, and $\widetilde{\beta} = \widetilde{S}_{11}^{-\frac{1}{2}} \widetilde{U}_r$ where $\widetilde{U} = \begin{bmatrix} \widetilde{U}_r & \widetilde{U}_{p-r} \end{bmatrix}$ are the p eigenvectors of $\widetilde{S}_{11}^{-\frac{1}{2}} \widetilde{S}_{10} \widetilde{S}_{00}^{-1} \widetilde{S}_{10} \widetilde{S}_{11}^{-\frac{1}{2}}$ with p eigenvalues $\widetilde{\lambda}_1 > \widetilde{\lambda}_2 > \ldots > \widetilde{\lambda}_p > 0$. The columns of $\widetilde{U}$ are ordered according to the size of their associated eigenvalue. Further, $\Sigma = \widetilde{\Sigma} = \widetilde{S}_{00} - \widetilde{S}_{01}\widetilde{\beta}\left(\widetilde{\beta}\widetilde{S}_{11}\widetilde{\beta}\right)^{-1}\widetilde{\beta}\widetilde{S}_{10}$. Therefore we may write

$$\begin{aligned} \exp\left\{-\widetilde{\kappa}h_i\left(\bar{\theta}\right)\right\} &= |\underline{H}|^{\frac{L}{2}}\left|\widetilde{\Sigma}\right|^{-\frac{\kappa}{2}}\exp\left\{-\frac{L\widetilde{\kappa}}{2}\right\} \\ &= |\underline{H}|^{\frac{L}{2}}\left[\left|\widetilde{S}_{00}\right|\prod_{i=1}^{r}\left(1-\widetilde{\lambda}_i\right)\right]^{-\frac{\kappa}{2}}\exp\left\{-\frac{L\widetilde{\kappa}}{2}\right\}. \end{aligned} \tag{11}$$

The expressions (10) for the diffuse prior, and (11) for the informative prior are derived in Appendix I. These expressions are useful for deriving simple expressions for the approximations to the Bayes factors as presented in Section 5. The similarity in the notation and the forms of the expressions used in this paper and that of Johansen (1995a) is deliberate. This similarity implies that although some additional code is required, existing computer code can be easily adapted for this method for much of the necessary computation..

4 Bayes Factors.

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.

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,

$$m\left(y|H_i\right) = \int p(\theta|H_i)L(y|\theta,H_i)d\theta. \tag{12}$$

The Bayes factors for these hypotheses are therefore,

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

Guidelines for the interpretation of twice the log Bayes factors are given in Kass & Raftery (1995) and reproduced here in Table 1.

Insert Table 1 here.

The ratio of the posterior probability for the hypothesis H_0, $P(H_0|y)$, and an alternative H_A, $P(H_A|y)$, 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 used by Strachan (1998). In this paper we use the Laplace approximation to the Bayes factors, or Bayesian trace statistics, and compare these results to the classical trace statistics.

5 Laplace approximation.

Let θ be a p-dimensional vector of parameters. If $-h(\theta)$ is a smooth, positive function with a maximum at $\overline{\theta}$, then by Laplace approximation, the integral

$$\int f(\theta) \exp\left[-\kappa h(\theta)\right] d\theta$$

can be approximated by

$$f\left(\overline{\theta}\right)\left(\frac{2\pi}{n}\right)^{\frac{p}{2}} |\Psi|^{-\frac{1}{2}} \exp\left[-\kappa h\left(\overline{\theta}\right)\right]$$

where Ψ is the Hessian of $h(\theta)$, evaluated at $\theta = \overline{\theta}$ (Lindley 1980, Tierney & Kadane 1986, Tierney, Kass & Kadane 1989, Kass & Raftery 1995).

A weakness of the Laplace method, particularly when estimating marginal likelihoods, is the constant of integration is not estimated as well as a marginal density (Tierney, Kass & Kadane 1989). As the marginal likelihood is the normalising constant for the posterior density, this suggests that $m(y|H_A)$ may not be well approximated, however there is considerable precedent in the literature for using this method (Lindley 1980, Kass & Vaidyanathan 1992, Raftery & Richardson 1993, Raftery 1994, Kass & Raftery 1995, Lewis & Raftery 1997). In the approach presented in this paper $f\left(\overline{\theta}\right) = \left|J|_{\lambda=0}\right|$, so the integral can be viewed as a weighted average of the Jacobian at $\lambda = 0$, with respect to the posterior modulo the Jacobian. To implement the Laplace approximation for the Bayes Factors, we need expressions for $\overline{\theta}$ and Ψ, these are presented in Appendix I and II. Using these results, the Laplace approximation for the marginal likelihood for the reduced rank model with rank r when we use an informative prior can be written as

$$\begin{aligned}\hat{m}\left(y|H_r\right) &= |J|_{\lambda=0}|\left(\frac{2\pi}{\widetilde{\kappa}}\right)^{\frac{L((L+1)/2+k+r)+(p-r)r}{2}} |\underline{H}|^{\frac{L}{2}} \exp\left\{-\frac{L\widetilde{\kappa}}{2}\right\} \\ & 2^{\frac{L}{2}}\left[\left|\widetilde{S}_{00}\right|\prod_{i=1}^{r}\left(1-\widetilde{\lambda}_i\right)\right]^{\frac{(L+k+r-\widetilde{\kappa}+1)}{2}} \left|\widetilde{M}_{22}\right|^{-\frac{L}{2}} \\ & \left|\widetilde{F}'\left(\widetilde{\alpha}\widetilde{\Sigma}^{-1}\widetilde{\alpha}'\otimes\left(\widetilde{S}_{11}-\widetilde{S}_{11}\widetilde{\beta}\widetilde{\beta}'\widetilde{S}_{11}\right)\right)\widetilde{F}\right|^{-\frac{1}{2}}.\end{aligned}$$

Similarly, when we use the diffuse prior we have

$$\begin{aligned}\hat{m}\left(y|H_r\right) &= |J|_{\lambda=0}|\left(\frac{2\pi}{\widehat{\kappa}}\right)^{\frac{L((L+1)/2+k+r)+(p-r)r}{2}} \exp\left\{-\frac{L\widehat{\kappa}}{2}\right\} \\ & 2^{\frac{L}{2}}\left[\left|\widehat{S}_{00}\right|\prod_{i=1}^{r}\left(1-\widehat{\lambda}_i\right)\right]^{\frac{(L+k+r-\widehat{\kappa}+1)}{2}} \left|\widehat{M}_{22}\right|^{-\frac{L}{2}} \\ & \left|\widehat{F}'\left(\widehat{\alpha}\widehat{\Sigma}^{-1}\widehat{\alpha}'\otimes\left(\widehat{S}_{11}-\widehat{S}_{11}\widehat{\beta}\widehat{\beta}'\widehat{S}_{11}\right)\right)\widehat{F}\right|^{-\frac{1}{2}}.\end{aligned}$$

Therefore the general form of the Bayes factor for the hypotheses of rank r to rank p, that is, ratio of the marginal likelihood for the reduced rank model with rank r to the marginal likelihood for the full rank model p, is

$$\begin{aligned}\hat{m}\left(y|H_r\right)/\hat{m}\left(y|H_p\right) &= |J|_{\lambda=0}|\,|J|^{-1}\left(\frac{2\pi}{\overline{\kappa}}\right)^{-\frac{(L-r)(p-r)}{2}} \left|\overline{S}_{11}\right|^{\frac{L}{2}}\left|\overline{S}_{00}\right|^{\frac{(L+k+r+1)}{2}} \\ & \left[\prod_{i=r+1}^{p}\left(1-\bar{\lambda}_i\right)\right]^{\frac{\kappa}{2}}\left[\prod_{i=1}^{r}\left(1-\bar{\lambda}_i\right)\right]^{\frac{(L+k+r+1)}{2}} \\ & \left|\overline{F}'\left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}'\otimes\left(\overline{S}_{11}-\overline{S}_{11}\bar{\beta}\bar{\beta}'\overline{S}_{11}\right)\right)\overline{F}\right|^{-\frac{1}{2}}.\end{aligned}$$

6 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 and those obtained in Strachan (1998) using a Bayesian approach with Monte Carlo integration.

6.1 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 a single common stochastic trend and, therefore, two cointegrating relationships. The cointegrating relationships are such that 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 $r = 2$, where c_t is the log of consumption, y_t is the log of output and i_t is the log of investment. A further implication formally investigated in this paper is that the cointegrating vectors $\beta = H$ where

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}. \tag{13}$$

To demonstrate the applicability of the method, the analysis will investigate support for the following hypotheses, $H_r : rank = r$ for $r = 0, 1, 2, 3$. The hypotheses H_0 to H_3 state the number of stochastic trends in x_t is $3 - r$.

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 with a constant is appropriate for the Australian data. We use the prior values $\underline{S} = I$, $\underline{H} = I$, $\underline{B} = 0$, and $\underline{\nu} = L + p + k + 5$.

We present the two Bayesian estimates of $\beta : \beta_B$ as the average from the posterior distribution with a diffuse prior; and β_{BM} as the mode of the posterior modulo the Jacobian. To investigate $H_4 : \beta = H$ informally, we normalise these estimates by $\beta = \beta^* \beta_1^{*-1}$, and we find the resultant Bayes and maximum likelihood estimates of β, assuming a rank of 2, are:

$$\beta_{BM} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -23.7 & -10.7 \end{pmatrix}, \; \beta_B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1.03 & -1.01 \end{pmatrix} \text{ and } \beta_{ML} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1.04 & -1.01 \end{pmatrix}.$$

The Bayesian mean and maximum likelihood estimates are comparable with (13) and suggest support for the real-business-cycle model proposed in King *et al.* (1991) while the Bayesian modal estimate shows no resembalnce to H. Some investigation shows that the Bayesian estimates of β are very sensitive the prior values and it is this feature and not the different methods of estimation that causes the difference in the estimates.

This real-business-cycle model model also implies there is one stochastic trend, and therefore the hypotheses about r need to be investigated. The calculated likelihood ratio test statistics, twice the natural logarithm of the Bayes factors and the posterior probabilities of the ranks are presented in Table 2. The first column of probabilities in Table 2 come from a Monte Carlo integration scheme developed in earlier work and the second column of probabilities is from the numerical integration method presented in this paper. The classical results suggest acceptance of $r = 1$, however the test statistic is close enough to the critical value to suggest some support for $r = 2$.

Insert Table 2 here.

According to the guidelines suggested by Kass & Raftery (1995) for interpreting twice the log Bayes factor, there is clear evidence that the model has reduced rank, with the greatest support for a rank of two. The classical trace statistic suggests a rank of one. This result implies two stochastic trends in the system instead of one as suggested by the model presented in King *et al.* We have converted these Bayes factor estimates to posterior probabilities for the ranks and these are presented in the final column of Table 2. These probabilities give a clearer indication of the evidence in the data for the rank of two.

In Table 3, the results of the test of hypothesis H_4 conditional on the rank, 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 under all forms of inference and estimation. The bayesian results for the hypothesis H_4 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 classical results suggest support for H_4 at ranks one and two. The ambiguity in these conditional results for different ranks lead us to consider unconditional results. The Bayesian unconditional posterior probabilities of H_4, $P(H_4|y) = \sum_{r=0}^{3} P(H_4|H_r, y)P(H_r|y)$, are included at the bottom of Table 3.

Overall, these results suggest the real-business-cycle model with permanent productivity shocks is valid as there appears to be one stochastic trend, the variables do enter the long run relations through some combination of the terms $(c_t - y_t)$ and $(i_t - y_t)$, regardless of the rank and from these results we have the joint probability of these conditions, $P(H_4 \cap H_2|y) = P(H_4|H_2, y)P(H_2|y)$, between 77% and 94%. While we can compare frequentist and Bayesian conclusions from the conditional tests in Table 3, no classical equivalent to this unconditional inference exists.

7 Conclusion.

In this paper we have demonstrated a method of finding approximations to log Bayes factors, or Bayesian trace statistics, for the rank of a potentially reduced rank regression model. These approximations use the Laplace method of approximating integrals. We also rely heavily on the method of Anderson (1951) and Johansen (1995a) for finding estimates of parameters for the reduced rank regression model at the mode of the posterior modulo the Jacobian. The results for a simple example of a business cycle model suggest that the performance of the Bayesian trace statistic is similar to that of the classical trace statistic.

8 References.

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

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).

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.

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.

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

Kass, R.E. and S.K. Vaidyanathan (1992). "Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions." *Journal of the Royal Statistical Society* B, 54, 129–144.

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. 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.

Lewis, S.M. and A.E. Raftery (1997). "Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator." *Journal of the American Statistical Association*, 92, 648-655.

Lindley, D.V. (1980). "Approximate Bayesian Methods." in: J.M. Bernardo, M.H. Degroot, D.V. Lindley, and A.M.F. Smith, eds., *Bayesian Statistics*, Valencia, Spain: University Press.

Raftery, A.E. (1994). "Approximate Bayes factors and accounting for model uncertainty in generalized linear models." Technical report no. 255, Department of Statistics, University of Washington.

Raftery, A.E. and Richardson S. (1993). "Model selection for generalized linear models via GLIB, with application to epidemiology." draft chapter for *Bayesian Biostatistics,* D.A. Berry and D.K. Strangl, eds..

Strachan, R.W. (1998). "Bayesian Estimation of the Reduced Rank Regression Model without Ordering Restrictions." Working paper, Monash University, Melbourne.

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).

9 Appendix I: Estimates of A, α, Σ and β at the mode of exp$(-\kappa h(\theta))$.

In this Appendix expressions for the modal estimates of A, α, Σ and β at the mode of $\exp(-\kappa h(\theta))$ are derived for the model with rank r. The expressions are derived for the informative prior and the expressions for the diffuse prior are shown to be a special case of these. We drop the caret $\hat{()}$ and tilde $\tilde{()}$ notation which we used to distinguish between scalars, vectors and matrices associated with the different priors in the paper. We do this to ease the notational burden on the reader. We are able to present a form for these expressions using the informative prior and show how to obtain from these the expressions for the diffuse prior with some simple changes.

We begin by concentrating out A, then α and then Σ and β. In deriving the form for the first and second derivatives in the model with rank r, we normally take the derivatives of the model including the term $g(\theta) = S_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}\Sigma$, and then set $\lambda = 0$. However, the derivatives of this term are

$$dg(\theta) = S_{11}^{-1}(d\beta_{\perp})\lambda\alpha_{\perp}\Sigma + S_{11}^{-1}\beta_{\perp}\lambda(d\alpha_{\perp})\Sigma + S_{11}^{-1}\beta_{\perp}\lambda\alpha_{\perp}(d\Sigma)$$

and

$$\begin{aligned} d^2g(\theta) &= S_{11}^{-1}(d\beta_{\perp})\lambda(d\alpha_{\perp})\Sigma + S_{11}^{-1}(d\beta_{\perp})\lambda\alpha_{\perp}(d\Sigma) \\ &\quad + S_{11}^{-1}(d\beta_{\perp})\lambda(d\alpha_{\perp})\Sigma + S_{11}^{-1}\beta_{\perp}\lambda(d\alpha_{\perp})(d\Sigma) \\ &\quad + S_{11}^{-1}(d\beta_{\perp})\lambda\alpha_{\perp}(d\Sigma) + S_{11}^{-1}\beta_{\perp}\lambda(d\alpha_{\perp})(d\Sigma) \end{aligned}$$

which are made up of products of λ. Therefore $dg(\theta) = 0$ and $d^2g(\theta) = 0$ at $\lambda = 0$. So we first set $\lambda = 0$ and then take derivatives as the resultant expressions are identical and this saves considerable notation.

We first remind the reader of the notation used. Let $Z_0 = Y$, $Z_1 = X$, and $Z_2 = Z$, $\kappa M_{ij} = Z_i'Z_j + h_{ij}$, for $(ij) = (1,1), (2,1), (2,2)$, $\kappa M_{20} = Z_2'Z_0 + h_{21}\underline{\Pi} + h_{22}\underline{A}$, and $\kappa M_{10} = Z_1'Z_0 + h_{11}\underline{\Pi} + h_{12}\underline{A}$ and $\kappa M_{00} = Z_0'Z_0 + \underline{B}'\underline{H}\underline{B} + \underline{S}$. Finally let $S_{ij} = M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$, $S_{ij} = S_{ji}'$ and $M_{ij} = M_{ji}'$. Then for the informative prior,

$$\begin{aligned} -\kappa h(\theta) &= -\frac{L}{2}\ln|\underline{H}| - \frac{\kappa}{2}\ln|\Sigma| \\ &\quad -\frac{1}{2}tr\Sigma^{-1}(Z_0 - Z_1\beta\alpha - Z_2A)'(Z_0 - Z_1\beta\alpha - Z_2A) \\ &\quad -\frac{1}{2}tr\Sigma^{-1}\left[(B - \underline{B})'\underline{H}(B - \underline{B}) + \underline{S}\right]. \end{aligned} \tag{14}$$

The expression for $-\kappa h(\theta)$ when we use the diffuse prior is found by first dropping the term $-\frac{L}{2}\ln|\underline{H}|$ and then setting $\underline{H}=0$ and $\underline{S}=0$ in (14). The first order conditions for A are

$$\left(Z_0 - Z_1\beta\alpha - Z_2\overline{A}\right)' Z_2 - (\Pi - \underline{\Pi})' h_{12} - \left(\overline{A} - \underline{A}\right)' h_{22} = 0$$

which may be written as

$$\overline{A}(\beta,\alpha) = M_{22}^{-1}M_{20} - M_{22}^{-1}M_{21}\beta\alpha.$$

Use this expression to rewrite the term

$$\begin{aligned}
& \left(Z_0 - Z_1\beta\alpha - Z_2\overline{A}\right)'\left(Z_0 - Z_1\beta\alpha - Z_2\overline{A}\right) + (B - \underline{B})'\,\underline{H}\,(B - \underline{B}) + \underline{S} \\
= \; & \kappa S_{00} - \kappa S_{01}\beta\alpha - \kappa\alpha'\beta' S_{10} + \kappa\alpha'\beta' S_{11}\beta\alpha
\end{aligned}$$

The first order conditions for α are

$$0 = -\beta' S_{10} + \beta' S_{11}\beta\alpha$$

such that

$$\begin{aligned}
\bar{\alpha}(\beta) &= (\beta' S_{11}\beta)^{-1}\beta' S_{10} \\
&= \beta' S_{10}
\end{aligned}$$

since $\beta' S_{11}\beta = I_r$ and

$$\overline{A}(\beta) = M_{22}^{-1}M_{20} - M_{22}^{-1}M_{21}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}.$$

Next, we concentrate out α.

$$\begin{aligned}
& \left(Z_0 - Z_1\beta\bar{\alpha} - Z_2\overline{A}\right)'\left(Z_0 - Z_1\beta\bar{\alpha} - Z_2\overline{A}\right) + (B - \underline{B})'\,\underline{H}\,(B - \underline{B}) + \underline{S} \\
= \; & \kappa S_{00} - \kappa S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} - \kappa S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} \\
& + \kappa S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} \\
= \; & \kappa S_{00} - \kappa S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} = \kappa S(\beta) = \kappa S
\end{aligned}$$

Therefore, for fixed β, the first order conditions for Σ are

$$\begin{aligned}
0 &= -\frac{\kappa}{2} tr\Sigma^{-1}(d\Sigma) + \frac{\kappa}{2} tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}S \\
&= -\frac{\kappa}{2} tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}[\Sigma - S(\beta)].
\end{aligned}$$

Which implies $\bar{\Sigma}(\beta) = S(\beta)$ and therefore

$$\begin{aligned} -\kappa h\left(\bar{\Sigma}(\beta), \beta, \bar{\alpha}(\beta), \overline{A}(\beta)\right) &= -\frac{L}{2}\ln|\underline{H}| - \frac{\kappa}{2}\ln\left|\bar{\Sigma}(\beta)\right| - \frac{\kappa L}{2} \\ &= -\frac{L}{2}\ln|\underline{H}| - \frac{\kappa}{2}\ln\left|S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10}\right| - \frac{\kappa L}{2}. \end{aligned}$$

As discussed in Johansen (1995, p. 91), we apply the identity

$$\left|\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right| = |\Sigma_{11}|\left|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right| = |\Sigma_{22}|\left|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right|$$

to the matrix

$$\begin{aligned} \left|\begin{array}{cc} S_{00} & S_{01}\beta \\ \beta' S_{10} & \beta' S_{11}\beta \end{array}\right| &= |S_{00}|\left|\beta'\left(S_{11} - S_{10}S_{00}^{-1}S_{01}\right)\beta\right| \\ &= |\beta' S_{11}\beta|\left|S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10}\right| \end{aligned}$$

to show

$$\left|S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10}\right| = |S_{00}|\left|\beta'\left(S_{11} - S_{10}S_{00}^{-1}S_{01}\right)\beta\right|\left|\beta' S_{11}\beta\right|^{-1}.$$

Maximizing $-\kappa h\left(\bar{\Sigma}(\beta), \beta, \bar{\alpha}(\beta), \overline{A}(\beta)\right)$ is equivalent therefore, to minimizing $\left|\bar{\Sigma}(\beta)\right|$. We minimize by setting $\bar{\beta} = S_{11}^{-\frac{1}{2}}U_r$, where U_r are the first r eigenvectors associated with the r largest eigenvalues of $S_{11}^{-\frac{1}{2}}S_{10}S_{00}^{-1}S_{01}S_{11}^{-\frac{1}{2}}$, such that $\bar{\beta}' S_{11}\bar{\beta} = I_r$ and $\bar{\beta}' S_{10}S_{00}^{-1}S_{01}\bar{\beta} = diag\left(\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_r\right)$, where the $\bar{\lambda}_i$ are the eigenvalues of $S_{11}^{-\frac{1}{2}}S_{10}S_{00}^{-1}S_{01}S_{11}^{-\frac{1}{2}}$. The modal estimates become

$$\begin{aligned} \bar{\Sigma} &= S = S_{00} - S_{01}\bar{\beta}\bar{\beta}' S_{10}, \\ \bar{\alpha} &= \bar{\beta}' S_{10}, \\ \overline{A} &= M_{22}^{-1}M_{20} - M_{22}^{-1}M_{21}\bar{\beta}\bar{\beta}' S_{10}. \end{aligned}$$

We can then rewrite $-\kappa h\left(\bar{\Sigma}, \bar{\beta}, \bar{\alpha}, \overline{A}\right) = -\kappa h\left(\bar{\theta}\right)$ as

$$\begin{aligned} -\kappa h\left(\bar{\theta}\right) &= -\frac{L}{2}\ln|\underline{H}| - \frac{\kappa}{2}\ln\left|\bar{\Sigma}\right| - \frac{\kappa L}{2} \\ &= -\frac{L}{2}\ln|\underline{H}| - \frac{\kappa}{2}\ln|S_{00}| - \frac{\kappa}{2}\sum_{i=1}^{r}\left(1 - \bar{\lambda}_i\right) - \frac{\kappa L}{2}, \end{aligned}$$

and so, for the informative prior

$$\begin{aligned}\exp\left\{-\kappa h\left(\bar{\theta}\right)\right\} &= \left|\underline{H}\right|^{-\frac{L}{2}}\left|\bar{\Sigma}\right|^{-\frac{\kappa}{2}}\exp\left\{-\frac{\kappa L}{2}\right\} \\ &= \left|\underline{H}\right|^{-\frac{L}{2}}\left[\left|S_{00}\right|\prod_{i=1}^{r}\left(1-\bar{\lambda}_i\right)\right]^{-\frac{\kappa}{2}}\exp\left\{-\frac{\kappa L}{2}\right\}.\end{aligned}$$

To find the necessary expressions for the diffuse prior for modal values of A, α, β, and Σ and a simplified expression for the function $\exp\left\{-\kappa h\left(\bar{\theta}\right)\right\}$ at its mode, we drop the term $\left|\underline{H}\right|^{-\frac{L}{2}}$ and set $\underline{H}=0$ and $\underline{S}=0$. Using the general notation of this section, the algebraic form of the modal estimates are the same as for the informative prior. Also, we find for the diffuse prior

$$\begin{aligned}\exp\left\{-\kappa h\left(\bar{\theta}\right)\right\} &= \left|\bar{\Sigma}\right|^{-\frac{\kappa}{2}}\exp\left\{-\frac{\kappa L}{2}\right\} \\ &= \left[\left|S_{00}\right|\prod_{i=1}^{r}\left(1-\bar{\lambda}_i\right)\right]^{-\frac{\kappa}{2}}\exp\left\{-\frac{\kappa L}{2}\right\}.\end{aligned}$$

This is a relatively simple function to calculate and has only r parameters to be estimated. That is, λ_i for $i=1,\ldots,r$.

10 Appendix II: The Hessian of $h(\theta)$ and its determinant.

The second term necessary for the Laplace integral is the Hessian of $h(\theta)$. Here we derive and then simplify the Hessian, we then use this expression to simplify the expression for the marginal likelihood. Again the results presented are for the informative prior and the notation from Appendix I is used. We begin by rewriting the expression $-\kappa h(\theta)$ as

$$\begin{aligned}-\kappa h(\theta) &= -\frac{L}{2}\ln\left|\underline{H}\right|-\frac{\kappa}{2}\ln\left|\Sigma\right| \\ &\quad -\frac{1}{2}tr\Sigma^{-1}\left(Z_0-Z_1\beta\alpha-Z_2A\right)'\left(Z_0-Z_1\beta\alpha-Z_2A\right) \\ &\quad -\frac{1}{2}tr\Sigma^{-1}\left[\left(B-\underline{B}\right)'\underline{H}\left(B-\underline{B}\right)+\underline{S}\right]\end{aligned}$$

$$
\begin{aligned}
&= -\frac{L}{2}\ln\left|\underline{H}\right| - \frac{\kappa}{2}\ln|\Sigma| - \frac{\kappa}{2}tr\Sigma^{-1}S \\
&\quad -\frac{1}{2}tr\Sigma^{-1}\left(B-\widetilde{B}\right)'\left(\underline{H}+\underline{X}'\underline{X}\right)\left(B-\widetilde{B}\right)
\end{aligned}
$$

$$
\begin{aligned}
-\kappa dh(\theta) &= -\frac{\kappa}{2}tr\Sigma^{-1}(d\Sigma) \\
&+\frac{1}{2}tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}\left(B-\widetilde{B}\right)'\left(\underline{H}+\underline{X}'\underline{X}\right)\left(B-\widetilde{B}\right) \\
&-\frac{\kappa}{2}tr\Sigma^{-1}[\left(\beta\alpha-\widetilde{\Pi}\right)'M_{11}(d\beta)\alpha+\left(\beta\alpha-\widetilde{\Pi}\right)'M_{11}\beta(d\alpha) \\
&+\alpha'(d\beta)'M_{11}\left(\beta\alpha-\widetilde{\Pi}\right)+(d\alpha)'\beta'M_{11}\left(\beta\alpha-\widetilde{\Pi}\right) \\
&+\left(\beta\alpha-\widetilde{\Pi}\right)'M_{12}(dA)+\alpha'(d\beta)'M_{12}\left(A-\widetilde{A}\right)+(d\alpha)'\beta'M_{12}\left(A-\widetilde{A}\right) \\
&+\left(A-\widetilde{A}\right)'M_{21}(d\beta)\alpha+\left(A-\widetilde{A}\right)'M_{21}\beta(d\alpha)+(dA)'M_{21}\left(\beta\alpha-\widetilde{\Pi}\right) \\
&+\left(A-\widetilde{A}\right)'M_{22}(dA)+(dA)'M_{22}\left(A-\widetilde{A}\right)] \\
&+\frac{\kappa}{2}tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}S
\end{aligned}
$$

$$
\begin{aligned}
-\kappa d^2h(\theta) &= \frac{\kappa}{2}tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}(d\Sigma) \\
&-tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}(d\Sigma)\Sigma^{-1}\left(B-\widetilde{B}\right)'\left(\underline{H}+\underline{X}'\underline{X}\right)\left(B-\widetilde{B}\right) \\
&+\kappa tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}[\left(\beta\alpha-\widetilde{\Pi}\right)'M_{11}(d\beta)\alpha+\left(\beta\alpha-\widetilde{\Pi}\right)'M_{11}\beta(d\alpha) \\
&+\alpha'(d\beta)'M_{11}\left(\beta\alpha-\widetilde{\Pi}\right)+(d\alpha)'\beta'M_{11}\left(\beta\alpha-\widetilde{\Pi}\right) \\
&+\left(\beta\alpha-\widetilde{\Pi}\right)'M_{12}(dA)+\alpha'(d\beta)'M_{12}\left(A-\widetilde{A}\right)+(d\alpha)'\beta'M_{12}\left(A-\widetilde{A}\right) \\
&+\left(A-\widetilde{A}\right)'M_{21}(d\beta)\alpha+\left(A-\widetilde{A}\right)'M_{21}\beta(d\alpha)+(dA)'M_{21}\left(\beta\alpha-\widetilde{\Pi}\right) \\
&+\left(A-\widetilde{A}\right)'M_{22}(dA)+(dA)'M_{22}\left(A-\widetilde{A}\right)] \\
&-\kappa tr\Sigma^{-1}[\alpha'(d\beta)'M_{11}(d\beta)\alpha+(d\alpha)'\beta'M_{11}\beta(d\alpha)+(dA)'M_{22}(dA) \\
&+(d\alpha)'\beta'M_{11}(d\beta)\alpha+\alpha'(d\beta)'M_{11}\beta(d\alpha)+(d\alpha)'\beta'M_{12}(dA)
\end{aligned}
$$

$$
\begin{aligned}
&+\alpha'(d\beta)' M_{12}(dA) + (d\alpha)'(d\beta)' M_{11}\left(\beta\alpha - \widetilde{\Pi}\right) \\
&+\left(\beta\alpha - \widetilde{\Pi}\right)' M_{11}(d\beta)(d\alpha) + (dA)' M_{21}(d\beta)\alpha + (dA)' M_{21}\beta(d\alpha)] \\
&-\frac{\kappa}{2} tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}(d\Sigma)\Sigma^{-1}S - \frac{\kappa}{2} tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}(d\Sigma)\Sigma^{-1}S
\end{aligned}
$$

$$
\begin{aligned}
-\kappa d^2 h\left(\bar{\theta}\right) \;=\; & -\kappa tr\bar{\Sigma}^{-1}[\bar{\alpha}'(d\beta)' M_{11}(d\beta)\bar{\alpha} + (d\alpha)'\bar{\beta}' M_{11}\bar{\beta}(d\alpha) + (dA)' M_{22}(dA) \\
& + (d\alpha)'\bar{\beta}' M_{11}(d\beta)\bar{\alpha} + \bar{\alpha}'(d\beta)' M_{11}\bar{\beta}(d\alpha) + (d\alpha)'\bar{\beta}' M_{12}(dA) \\
& +\bar{\alpha}'(d\beta)' M_{12}(dA) + (dA)' M_{21}(d\beta)\bar{\alpha} + (dA)' M_{21}\beta(d\alpha)] \\
& -\frac{\kappa}{2} tr\bar{\Sigma}^{-1}(d\Sigma)\bar{\Sigma}^{-1}(d\Sigma)
\end{aligned}
$$

$$
\begin{aligned}
d^2 h\left(\bar{\theta}\right) \;=\; & \frac{1}{2}(dvech\Sigma) D_L'\left(\bar{\Sigma}^{-1}\otimes\bar{\Sigma}^{-1}\right) D_L(dvech\Sigma) \\
& + (dvec\beta)'\left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}'\otimes M_{11}\right)(dvec\beta) + (dvec\alpha)'\left(\bar{\Sigma}^{-1}\otimes\bar{\beta}' M_{11}\bar{\beta}\right)(dvec\alpha) \\
& + (dvecA)'\left(\bar{\Sigma}^{-1}\otimes M_{22}\right)(dvecA) \\
& + (dvec\alpha)'\left(\bar{\Sigma}^{-1}\bar{\alpha}'\otimes\bar{\beta}' M_{11}\right)(dvec\beta) + (dvec\beta)'\left(\bar{\alpha}\bar{\Sigma}^{-1}\otimes M_{11}\bar{\beta}\right)(dvec\alpha) \\
& + (dvec\alpha)'\left(\bar{\Sigma}^{-1}\otimes\bar{\beta}' M_{12}\right)(dvecA) + (dvecA)'\left(\bar{\Sigma}^{-1}\otimes M_{21}\beta\right)(dvec\alpha) \\
& + (dvecA)'\left(\bar{\Sigma}^{-1}\bar{\alpha}'\otimes M_{21}\right)(dvec\beta) + (dvec\beta)'\left(\bar{\alpha}\bar{\Sigma}^{-1}\otimes M_{12}\right)(dvecA) \\
\;=\; & \frac{1}{2}(dvech\Sigma) D_L'\left(\bar{\Sigma}^{-1}\otimes\bar{\Sigma}^{-1}\right) D_L(dvech\Sigma) \\
& + (dvec\beta_2)' F'\left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}'\otimes M_{11}\right) F(dvec\beta_2) \\
& + (dvec\alpha)'\left(\bar{\Sigma}^{-1}\otimes\bar{\beta}' M_{11}\bar{\beta}\right)(dvec\alpha) \\
& + (dvecA)'\left(\bar{\Sigma}^{-1}\otimes M_{22}\right)(dvecA) \\
& + (dvec\alpha)'\left(\bar{\Sigma}^{-1}\bar{\alpha}'\otimes\bar{\beta}' M_{11}\right) F(dvec\beta_2) \\
& + (dvec\beta_2)' F'\left(\bar{\alpha}\bar{\Sigma}^{-1}\otimes M_{11}\bar{\beta}\right)(dvec\alpha) \\
& + (dvec\alpha)'\left(\bar{\Sigma}^{-1}\otimes\bar{\beta}' M_{12}\right)(dvecA) \\
& + (dvecA)'\left(\bar{\Sigma}^{-1}\otimes M_{21}\beta\right)(dvec\alpha) \\
& + (dvecA)'\left(\bar{\Sigma}^{-1}\bar{\alpha}'\otimes M_{21}\right) F(dvec\beta_2)
\end{aligned}
$$

$$+ (dvec\beta_2)' F' \left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes M_{12}\right) (dvecA)$$

where

$$\begin{aligned} \frac{(dvec\beta)}{(dvec\beta_2)'} &= F \\ &= \left[I_r \otimes \begin{pmatrix} I_r \\ 0_{(p-r)\times r} \end{pmatrix} \right] \ddot{E} + \left[I_r \otimes \begin{pmatrix} 0_{r\times(p-r)} \\ I_{p-r} \end{pmatrix} \right] \end{aligned}$$

where $\ddot{E}$ is defined in Strachan (1998). This gives determinant of the Hessian as

$$\begin{aligned} \left|H\left(h\left(\bar{\theta}\right)\right)\right| &= \begin{vmatrix} \Upsilon & 0 \\ 0 & \frac{1}{2}D_L'\left(\bar{\Sigma}^{-1} \otimes \bar{\Sigma}^{-1}\right) D_L \end{vmatrix} \\ &= |\Upsilon| \times \left|\frac{1}{2}D_L'\left(\bar{\Sigma}^{-1} \otimes \bar{\Sigma}^{-1}\right) D_L\right| \end{aligned}$$

where

$$\begin{aligned} \Upsilon &= \begin{bmatrix} \left(\bar{\Sigma}^{-1} \otimes \bar{\beta}' M_{11}\bar{\beta}\right) & \left(\bar{\Sigma}^{-1}\bar{\alpha}' \otimes \bar{\beta}' M_{11}\right) F & \left(\bar{\Sigma}^{-1} \otimes \bar{\beta}' M_{12}\right) \\ F'\left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes M_{11}\bar{\beta}\right) & F'\left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}' \otimes M_{11}\right) F & F'\left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes M_{12}\right) \\ \left(\bar{\Sigma}^{-1} \otimes M_{21}\bar{\beta}\right) & \left(\bar{\Sigma}^{-1}\bar{\alpha}' \otimes M_{21}\right) F & \left(\bar{\Sigma}^{-1} \otimes M_{22}\right) \end{bmatrix} \\ &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \end{aligned}$$

$$\begin{aligned} A_{11} &= \begin{matrix} \left(\bar{\Sigma}^{-1} \otimes \bar{\beta}' M_{11}\bar{\beta}\right) & \left(\bar{\Sigma}^{-1}\bar{\alpha}' \otimes \bar{\beta}' M_{11}\right) F \\ F'\left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes M_{11}\bar{\beta}\right) & F'\left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}' \otimes M_{11}\right) F \end{matrix}, \\ A_{12} &= A_{21}' = \begin{bmatrix} \left(\bar{\Sigma}^{-1} \otimes \bar{\beta}' M_{12}\right) \\ F'\left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes M_{12}\right) \end{bmatrix}, \text{ and } A_{22} = \left(\bar{\Sigma}^{-1} \otimes M_{22}\right). \end{aligned}$$

Then

$$\begin{aligned} \Upsilon &= \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = |A_{22}| \left|A_{11} - A_{12}A_{22}^{-1}A_{21}\right| \\ &= \left|\bar{\Sigma}^{-1} \otimes M_{22}\right| \times \end{aligned}$$

$$
\begin{aligned}
& \left| A_{11} - \begin{matrix} \left(\bar{\Sigma}^{-1} \otimes \bar{\beta}' M_{12}\right) \\ F' \left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes M_{12}\right) \end{matrix} \left(\bar{\Sigma} \otimes M_{22}^{-1}\right) \left(\bar{\Sigma}^{-1} \otimes M_{21}\beta\right) \quad \left(\bar{\Sigma}^{-1}\bar{\alpha}' \otimes M_{21}\right) F \right| \\
= & \left|\bar{\Sigma}\right|^{-k} \left|M_{22}\right|^{L} \times \\
& \left| A_{11} - \begin{matrix} \left(I_L \otimes \bar{\beta}' M_{12} M_{22}^{-1}\right) \\ F' \left(\bar{\alpha} \otimes M_{12} M_{22}^{-1}\right) \end{matrix} \left(\bar{\Sigma}^{-1} \otimes M_{21}\beta\right) \quad \left(\bar{\Sigma}^{-1}\bar{\alpha}' \otimes M_{21}\right) F \right| \\
= & \left|\bar{\Sigma}\right|^{-k} \left|M_{22}\right|^{L} \times \\
& \left| A_{11} - \begin{matrix} \left(\bar{\Sigma}^{-1} \otimes \bar{\beta}' M_{12} M_{22}^{-1} M_{21}\beta\right) & \left(\bar{\Sigma}^{-1}\bar{\alpha}' \otimes \bar{\beta}' M_{12} M_{22}^{-1} M_{21}\right) F \\ F' \left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes M_{12} M_{22}^{-1} M_{21}\beta\right) & F' \left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}' \otimes M_{12} M_{22}^{-1} M_{21}\right) F \end{matrix} \right| \\
= & \left|\bar{\Sigma}\right|^{-k} \left|M_{22}\right|^{L} \left| \begin{matrix} \left(\bar{\Sigma}^{-1} \otimes \bar{\beta}' S_{11}\bar{\beta}\right) & \left(\bar{\Sigma}^{-1}\bar{\alpha}' \otimes \bar{\beta}' S_{11}\right) F \\ F' \left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes S_{11}\bar{\beta}\right) & F' \left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}' \otimes S_{11}\right) F \end{matrix} \right| \\
= & \left|\bar{\Sigma}\right|^{-k} \left|M_{22}\right|^{L} \left| \left(\bar{\Sigma}^{-1} \otimes I_r\right) \right| \times \\
& \left| F' \left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}' \otimes S_{11} - \left(\bar{\alpha}\bar{\Sigma}^{-1} \otimes S_{11}\bar{\beta}\right) \left(\bar{\Sigma}^{-1} \otimes I_r\right) \left(\bar{\Sigma}^{-1}\bar{\alpha}' \otimes \bar{\beta}' S_{11}\right)\right) F \right| \\
= & \left|\bar{\Sigma}\right|^{-(k+r)} \left|M_{22}\right|^{L} \left| F' \left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}' \otimes \left(S_{11} - S_{11}\bar{\beta}\bar{\beta}' S_{11}\right)\right) F \right|
\end{aligned}
$$

$$
\begin{aligned}
\left| \frac{1}{2} D_L' \left(\bar{\Sigma}^{-1} \otimes \bar{\Sigma}^{-1}\right) D_L \right| &= \left| 2 D_L^{+} \left(\bar{\Sigma} \otimes \bar{\Sigma}\right) D_L^{+'} \right|^{-1} \\
&= 2^{-L} \left|\Sigma\right|^{-(L+1)}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\left| H\left(h\left(\bar{\theta}\right)\right) \right|^{-\frac{1}{2}} &= 2^{\frac{L}{2}} \left|\bar{\Sigma}\right|^{\frac{(L+k+r+1)}{2}} \left|M_{22}\right|^{-\frac{L}{2}} \times \\
& \left| F' \left(\bar{\alpha}\bar{\Sigma}^{-1}\bar{\alpha}' \otimes \left(S_{11} - S_{11}\bar{\beta}\bar{\beta}' S_{11}\right)\right) F \right|^{-\frac{1}{2}}
\end{aligned}
$$

For the model with full rank, the expression for the Hessian is

$$
\left| H\left(h\left(\bar{\theta}\right)\right) \right|^{-\frac{1}{2}} = 2^{\frac{L}{2}} \left|\bar{\Sigma}\right|^{\frac{(L+p+k+1)}{2}} \left|M_{22}\right|^{-\frac{L}{2}} \left|S_{11}\right|^{-\frac{L}{2}}.
$$

Table 1: *Interpretation of twice the log Bayes factor (Kass & Raftery 1995).*

$2\ lnBF(0\|A)$	$BF(0\|A)$	Evidence against H_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very Strong

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

Rank (r)	$LR(H_r\|H_3)$	5% Critical value	$2\ ln\widetilde{BF}(r\|3)$	$\widetilde{P}(H_r\|y)$	$P(H_r\|y)$
0	35.28	29.68	27.50	0.01	0.00
1	11.12*	15.41	34.50	0.22	0.03
2	2.12*	3.76	37.02	0.77	0.96
3	-	-	-	0.00	0.01

* Accept the null hypothesis $H_r : rank = r$ against the alternative $H_3 : rank = 3$.

Table 3: *LR statistics and posterior probabilities for H_4 given H_r.*

Rank (r)	$LR(H_4\|H_r)$	*p-value*	$2\ ln\widetilde{BF}(4\|r)$	$\widetilde{P}(H_4\|H_r, y)$	$P(H_r\|y)$
0	-	-	0.00	0.50	0.50
1	0.002	0.97	-17.2	0.00	0.00
2	0.286	0.87	1743.2	1.00	0.98
3	-	-	-	0.00	0.00

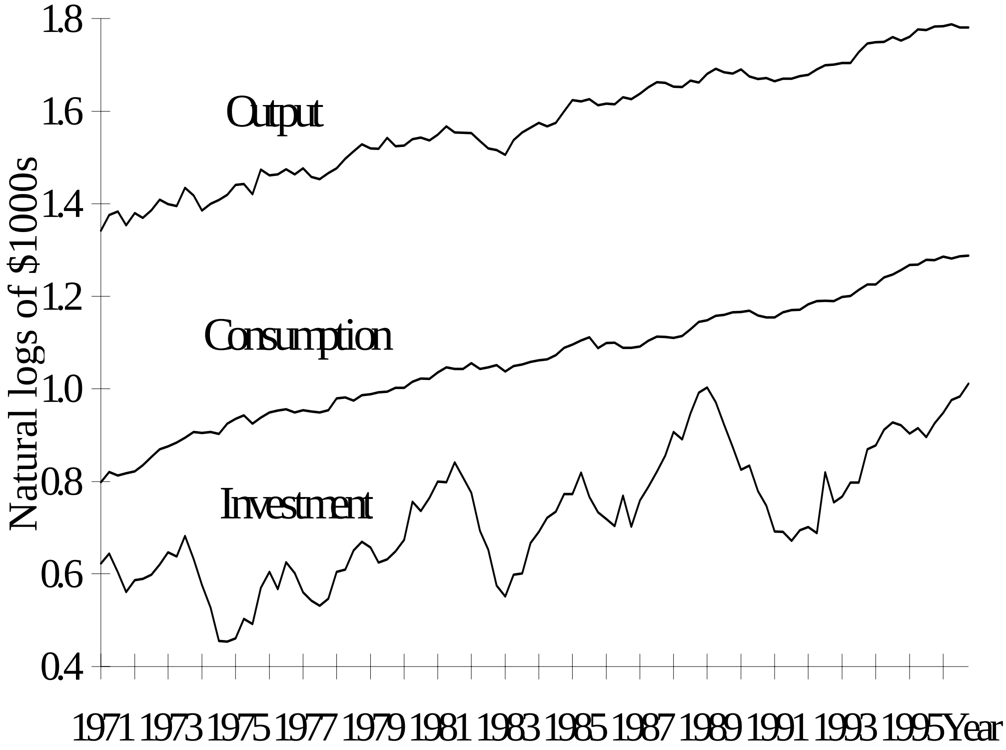

Figure 1: Logs of seasonally adjusted, real private per capita consumption (c_t), investment (i_t) and output (y_t). The data are obtained from the Australian dX database and the series identifiers and construction are detailed in Harris (1997, Figure 1). In this figure, 0.9 has been added to i_t for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