Bayesian Testing in Cointegration Models using the Jeffreys' Prior

Preliminary version. Not to be quoted

Frank Kleibergen[*] Richard Kleijn[†]

Abstract

We develop a Bayesian cointegration test statistic that can be used under a Jeffreys' prior. The test statistic is equal to the posterior expectation of the classical score statistic. Under the assumption of a full rank value of the long run multiplier the test statistic is a random variable with a chi-squared distribution. We evaluate whether the value of the test statistic under the restriction of cointegration is a plausible realization from its distribution under the encompassing, full rank model. We provide the posterior simulator that is needed to compute the test statistic. The simulator utilizes the invariance properties of the Jeffreys' prior such that the parameter drawings from a suitably rescaled model can be used. The test statistic can straightforwardly be extended to a more general model setting. For example, we show that structural breaks in the constant or trend and general mixtures of normal disturbances can be modelled, because conditional on some latent parameters all derivations still hold. We apply the Bayesian cointegration statistic to the Danish dataset of Johansen and Juselius (1990) and to four artificial examples to illustrate the use of the statistic as a diagnostic tool.

1 Introduction

In Bayesian statistics models are typically compared using the Bayes Factor or Posterior Odds Ratios. They cannot however not routinely be applied in models containing improper priors such as the commonly used Jeffreys' prior. We develop a Bayesian statistic that can be used to compare a cointegration model to an encompassing full rank model using a Jeffreys' prior.

We specify the full rank model in such a way that setting a particular, uniquely defined parameter to zero corresponds to imposing the cointegration restriction of rank reduction on the long-run multiplier parameter matrix. When we use a Jeffreys' prior both on the parameters of the cointegration model and the unrestricted error correction model, we show that the posterior of the

[*]Dept. of Quantitative Economics, University of Amsterdam, Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands, e-mail: kleiberg@fee.uva.nl

[†]Tinbergen Institute and Econometric Institute, Erasmus University Rotterdam, P.O.Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: rkleijn@few.eur.nl

cointegration model is equal to a conditional posterior of the parameters of the unrestricted error correction model given that the parameter associated with the cointegration restriction equals zero.

Hence, when we can construct the marginal posterior of this specific parameter from the unrestricted error correction model, it indicates whether the cointegration restriction is a plausible parameter realization. The analytical expression of the marginal posterior of the parameter that reflects cointegration is unknown. Instead of using the marginal posterior, we therefore construct a statistic that is a random variable with a standardized distribution when we use the posterior of the parameters of the unrestricted error correction model. This statistic is equal to the posterior expectation, with respect to the parameters that are present in the cointegration model, of the classical score statistic. We can then construct the expression of this statistic using the posterior that results under the nested model. We consider the resulting value as a realization from the standardized distribution under the unrestricted error correction model and we can determine whether it is a realization from the tail or the area with the bulk of the probability mass. This then shows whether the cointegration assumption is a plausible restriction on the unrestricted error correction model.

The paper is organized as follows. In section 2, we define the cointegration model. We show that the Jeffreys' prior on the parameters of the cointegration model can also be considered to result as a conditional prior of a Jeffreys' prior on the parameters of the encompassing unrestricted error correction model. This prior is conditional on a particular parameter being equal to zero. The parameter results from a singular value decomposition and reflects rank reduction of the long run multiplier. We construct a posterior simulator for the posterior parameters of the cointegration model using a Jeffreys' prior. In section 3, we construct the Bayesian score test statistic for cointegration. This statistic also results from the singular value decomposition of the long run multiplier. We construct it in such a way that it has a χ^2 distribution under the unrestricted model.

Furthermore, we apply the resulting Bayesian cointegration analysis to four artificial examples. We find that the statistic accurately detects misspecification of the cointegration restriction. We also apply the method to the Danish dataset from Johansen and Juselius (1990). In section 4, we indicate some model extensions that are allowed for by our Bayesian cointegration testing procedure. The fifth section concludes.

2 Cointegration model with $N(0, \Sigma)$ errors

2.1 Model and likelihood

The vector autoregressive model of order k (VAR(k)) reads

$$Y_t = \mu + \tau t + \sum_{i=1}^{k} \Phi_i Y_{t-i} + \epsilon_t, \tag{2.1}$$

where Y_t, $t = 1, 2, \ldots, T$, is a n-dimensional process and ϵ_t follows a n-dimensional white noise process. In this section we consider the case of multivariate NID ϵ_t, with mean 0 and covariance matrix Σ.

We refer to the error correction form of the VAR(k) model as the linear error correction model (**LEC**),

$$\Delta Y_t = \mu + \tau t + \Pi' Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \epsilon_t, \tag{2.2}$$

where $\Pi' = \sum_{i=1}^{k} \Phi_i - I_n$ and $\Gamma_i = -\sum_{j=i+1}^{k} \Phi_j$.

The likelihood function of the **LEC** model is

$$\begin{aligned} L_{lec}(Y|\Pi, \Sigma) &\propto \prod_{t=2}^{T} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\epsilon_t' \Sigma^{-1} \epsilon_t\right] \\ &= |\Sigma|^{-\frac{1}{2}(T-1)} \exp\left[-\frac{1}{2} \operatorname{tr}\left(\Sigma^{-1}\epsilon'\epsilon\right)\right] \\ &= |\Sigma|^{-\frac{1}{2}(T-1)} \exp\left[-\frac{1}{2} \operatorname{tr}\left(\Sigma^{-1}(\Delta Y - Y_{-1}\Pi)'(\Delta Y - Y_{-1}\Pi)\right)\right]. \end{aligned} \tag{2.3}$$

If Π has reduced rank r, then it can be specified as the product of two (full rank) $n \times r$ matrices α' and β:

$$\Pi = \beta\alpha. \tag{2.4}$$

A straightforward way of identifying the elements of α and β is by normalizing β as

$$\beta = \begin{pmatrix} I_r \\ -\beta_2 \end{pmatrix}. \tag{2.5}$$

We shall refer to the resulting model as the error correction cointegration (**ECC**) model. In matrix notation the **ECC** model can be specified as

$$\Delta Y = Y_{-1}\beta\alpha + X\Phi + \epsilon, \tag{2.6}$$

where $\Delta Y = (\Delta Y_{k+1} \ldots \Delta Y_T)'$, $Y_{-1} = (Y_k \ldots Y_{T-1})'$, $\epsilon = (\epsilon_{k+1} \ldots \epsilon_T)'$, $X = (X_{k+1}' \ldots X_T')'$, $X_t = (\Delta Y_{t-1}' \ldots \Delta Y_{t-k}'\ 1\ t)$ and $\Phi = (\Gamma_1 \ldots \Gamma_{k-1}\ \mu\ \tau)'$.

In the following sections we confine our attention to the case of an **ECC** model of order 1 without deterministic components, *i.e*

$$\Delta Y = Y_{-1}\beta\alpha + \epsilon \tag{2.7}$$

whose likelihood function is

$$\begin{aligned} L_{ecc}(Y|\alpha, \beta_2, \Sigma) &\propto |\Sigma|^{-\frac{1}{2}(T-1)} \exp\left[-\frac{1}{2} \operatorname{tr}\left(\Sigma^{-1}(\Delta Y - Y_{-1}\beta\alpha)'(\Delta Y - Y_{-1}\beta\alpha)\right)\right] \\ &= L_{lec}(Y|\Pi, \Sigma)|_{\Pi=\beta\alpha}. \end{aligned} \tag{2.8}$$

2.2 Jeffreys' prior and posterior

Jeffreys' prior on α and β

We assume a Jeffreys' prior on the parameters of the **ECC** model. The Jeffreys' prior is proportional to the square root of the determinant of the information matrix of and it can be specified as:

$$p_{ecc}(\Sigma) \propto |\Sigma|^{-(n+1)/2} \tag{2.9}$$

and

$$\begin{aligned} p_{ecc}(\alpha, \beta_2|\Sigma) \propto & |\mathcal{I}(\alpha, \beta_2|\Sigma)|^{\frac{1}{2}} \\ & = \left| \left(\frac{\partial \operatorname{vec}(\Pi)}{\partial(\operatorname{vec}(\alpha)' \ \ \operatorname{vec}(\beta_2)')} \right)' \mathcal{I}(\Pi|\Sigma) \left(\frac{\partial \operatorname{vec}(\Pi)}{\partial(\operatorname{vec}(\alpha)' \ \ \operatorname{vec}(\beta_2)')} \right) \right|^{\frac{1}{2}} \\ & = \left| \left(I_n \otimes \beta \quad \alpha' \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \right)' (\Sigma^{-1} \otimes Y_{-1}' Y_{-1}) \right. \\ & \qquad \left. \left(I_n \otimes \beta \quad \alpha' \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \right) \right|^{\frac{1}{2}}, \end{aligned} \tag{2.10}$$

where $\mathcal{I}(\alpha, \beta_2|\Sigma)$ and $\mathcal{I}(\Pi|\Sigma)$ denote the conditional information matrices.

Jeffreys' prior invariant under transformations

The Jeffreys' prior is invariant under data and parameter transformations, see *e.g* Box and Tiao (1973). Therefore, a large part of the analysis of the cointegration model can also be done in a suitably transformed model provided that also on its parameters a Jeffreys' prior is specified.

Jeffreys' prior on α^* and β_2^* in a rescaled model

The **LEC** model of order 1 without deterministic components is rescaled by postmultiplying both sides of the equation by the inverse of the matrix square root [1] of Σ such that the covariance matrix of the resultant disturbances $\epsilon_t \Sigma^{-\frac{1}{2}}$ becomes the identity matrix. Note that implicitly we have conditioned on Σ. Moreover, the parameter matrix Π is premultiplied by the matrix square root of $Y_{-1}' Y_{-1}$ and the data matrix Y_{-1} is postmultiplied by the inverse of this expression, such that

$$\Delta Y \Sigma^{-\frac{1}{2}} = Y_{-1}(Y_{-1}' Y_{-1})^{-\frac{1}{2}} \Pi^* + \eta, \tag{2.11}$$

[1] A matrix square root of a matrix A is any matrix $A^{\frac{1}{2}}$ that satisfies $A = A^{\frac{1}{2}} A^{\frac{1}{2}}$. For a positive semidefinite matrix A the matrix square root is calculated as $C' \Lambda^{\frac{1}{2}} C$, where $\Lambda^{\frac{1}{2}}$ is a diagonal matrix with the square roots of the eigenvalues of A on its diagonal and C contains the corresponding orthonormal eigenvectors of A in its columns.

with $\eta \sim N(0, I_n \otimes I_T)$ and $\Pi^* = (Y'_{-1}Y_{-1})^{\frac{1}{2}}\Pi\Sigma^{-\frac{1}{2}}$. Rank reduction of the parameter matrix Π^* implies that similar to (2.4) it can be specified as the product of two matrices α^* and β^* of lower dimension such that

$$\Pi^* = \beta^*\alpha^* = (Y'_{-1}Y_{-1})^{\frac{1}{2}}\beta\alpha\Sigma^{-\frac{1}{2}} \tag{2.12}$$

We normalize β^* similarly to β as

$$\beta^* = \begin{pmatrix} I_r \\ -\beta_2^* \end{pmatrix}. \tag{2.13}$$

The Jeffreys' prior on the parameters α^* and β_2^* has the same structure as the Jeffreys' prior on α and β in the unscaled model but as a result of the scaling the expression corresponding to $\Sigma^{-1} \otimes Y'_{-1}Y_{-1}$ in the prior (2.9) of the unscaled model now simplifies to $I_n \otimes I_n$ such that

$$\begin{aligned} p_{ecc}(\alpha^*, \beta_2^*|\Sigma) &\propto \left| \left(I_n \otimes \beta^* \quad \alpha^{*\prime} \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \right)' \left(I_n \otimes \beta^* \quad \alpha^{*\prime} \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \right) \right|^{\frac{1}{2}} \\ &= \begin{vmatrix} I_n \otimes \beta^{*\prime}\beta^* & \alpha^{*\prime} \otimes \beta_2^{*\prime} \\ \alpha^* \otimes \beta_2^* & \alpha^*\alpha^{*\prime} \otimes I_{n-r} \end{vmatrix}^{\frac{1}{2}} \end{aligned} \tag{2.14}$$

The Jeffreys' prior on α and β_2 (or α^* and β_2^*) can also be obtained using another way of reasoning. Kleibergen and Paap (1998) provide a framework to construct priors on α and β that are consistent with a natural conjugate or diffuse prior on the unrestricted matrix Π, conditional on the rank restriction. This nesting principle is then used to derive the posterior for the parameters and an algorithm to simulate from it is described. We also show that the Jeffreys' prior on α and β is equivalent to a diffuse prior on the parameter matrix of some rescaled unrestricted VAR model conditional on the restriction of rank reduction.

Specification of the rank reduction restriction using the singular value decomposition

The rank restriction on Π and hence on Π^* can be expressed explicitly using the following decomposition that is based on the singular value decomposition (see Kleibergen and Paap (1998))

$$\Pi^* = \beta^*\alpha^* + \beta_\perp^*\lambda\alpha_\perp^*, \tag{2.15}$$

where $\beta^* = \begin{pmatrix} I_r & -\beta_2^{*\prime} \end{pmatrix}'$ in which β_2^* is a $(n-r) \times r$ matrix; $\beta_\perp^*$ is a $n \times (n-r)$ matrix such that $\beta'\beta_\perp^* = 0$ and $\beta_\perp^{*\prime}\beta_\perp^* = I_{n-r}$; $\alpha_\perp^*$ is a $(n-r) \times n$ matrix such that $\alpha_\perp^*\alpha^{*\prime} = 0$ and $\alpha_\perp^*\alpha_\perp^{*\prime} = I_{n-r}$; λ is a $(n-r) \times (n-r)$ matrix. Rank reduction occurs when $\lambda = 0$ because (2.15) then reduces to $\Pi^* = \beta^*\alpha^*$.

Explicit expressions for $\alpha_\perp^*$ and $\beta_\perp^*$ that obey these requirements can be constructed. We will use $\alpha_\perp^* = (I_{n-r} + \alpha_2^{*\prime}\alpha_1^{*-1\prime}\alpha_1^{*-1}\alpha_2^*)^{-\frac{1}{2}} \begin{pmatrix} -\alpha_2^{*\prime}\alpha_1^{*-1\prime} & I_{n-r} \end{pmatrix}$ and $\beta_\perp^* = \begin{pmatrix} \beta_2^* & I_{n-r} \end{pmatrix}' (I_{n-r} + \beta_2^*\beta_2^{*\prime})^{-\frac{1}{2}}$.

All matrices involved in this decomposition can readily be computed from Π^* using the singular value decomposition

$$\Pi^* = USV', \tag{2.16}$$

with U and V orthogonal $n \times n$ matrices such that $U'U = I_n$ and $V'V = I_n$, and S a $n \times n$ diagonal matrix having the singular values of Π^* on its diagonal in decreasing order. The singular values of a matrix Π^* are the eigenvalues of the square symmetric matrix $\Pi^{*\prime}\Pi^*$ and are therefore always real and non-negative. As the rank of a matrix equals the number of non-zero singular values, the singular value decomposition seems a natural way of dealing with the restriction of rank reduction.

The matrices U, S and V are partitioned,

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, \; S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}, \text{ and } V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}', \tag{2.17}$$

with U_{11}, S_1, and V_{11} $r \times r$ matrices, U_{22}, S_2, and V_{22} $(n-r)\times(n-r)$ matrices, U_{21} and V_{21} $(n-r) \times r$ matrices, and U_{12} and V_{12} $r \times (n-r)$ matrices.

The matrices in decomposition (2.15) can all be written in terms of the blocks of U, S and V.

$$\alpha^* = U_{11}S_1 \begin{pmatrix} V_{11}' & V_{21}' \end{pmatrix}, \tag{2.18}$$

$$\beta_2^* = -U_{21}U_{11}^{-1}, \tag{2.19}$$

$$\lambda = (U_{22}U_{22}')^{-\frac{1}{2}} U_{22}S_2V_{22}'(V_{22}V_{22}')^{-\frac{1}{2}}, \tag{2.20}$$

$$\alpha_\perp^* = (V_{22}V_{22}')^{\frac{1}{2}} V_{22}^{-1\prime} \begin{pmatrix} V_{12}' & V_{22}' \end{pmatrix}, \tag{2.21}$$

$$\beta_\perp^* = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1}(U_{22}U_{22}')^{\frac{1}{2}}. \tag{2.22}$$

Jeffreys' prior is implied by diffuse prior on Π^*

We specify a diffuse prior on Π^*, that is $p_{lec}(\Pi^*|\Sigma) \propto 1$. The prior of $\alpha^*, \beta_2^*, \lambda|\Sigma$ results from $\Pi^*|\Sigma$ by the transformation of random variables defined by (2.15) such that

$$\begin{aligned} p_{uec}(\alpha^*, \beta_2^*, \lambda|\Sigma) =& p_{lec}(\Pi^*|\Sigma)|_{\Pi^*=\beta^*\alpha^*+\beta_\perp^*\lambda\alpha_\perp^*} |J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))| \\ \propto& |J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))|, \end{aligned} \tag{2.23}$$

where $J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))$ denotes the Jacobian of the transformation from Π^* to $(\alpha^*, \beta_2^*, \lambda)$. The conditional prior $p_{uec}(\alpha^*, \beta_2^*|\lambda, \Sigma)$ is fully determined by the joint prior $p_{uec}(\alpha^*, \beta_2^*, \lambda|\Sigma)$. We evaluate the conditional prior $p_{uec}(\alpha^*, \beta_2^*|\lambda, \Sigma)$ in $\lambda = 0$ to obtain a prior $p_{ecc}(\alpha^*, \beta_2^*|\Sigma)$ for the **ECC** model. Constructing the prior for the **ECC** model in this way ensures that it is consistent with the prior knowledge from the encompassing **LEC** model, *i.e* $p_{lec}(\Pi^*|\Sigma) \propto 1$. The extra information in the **ECC** model, namely the restriction $\lambda = 0$ which represents the imposition of rank reduction, is taken into account by conditioning on it.

We now have,

$$
\begin{aligned}
p_{ecc}(\alpha^*, \beta_2^*|\Sigma) =& p_{uec}(\alpha^*, \beta_2^*|\lambda = 0, \Sigma) \\
=& \frac{p_{uec}(\alpha^*, \beta_2^*, \lambda|\Sigma)|_{\lambda=0}}{p_{uec}(\lambda|\Sigma)|_{\lambda=0}} \\
\propto& |J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))||_{\lambda=0}.
\end{aligned}
\tag{2.24}
$$

To calculate the Jacobian we first note that the total differential of (2.15) evaluated in $\lambda = 0$ reads,

$$
\begin{aligned}
d\Pi^* &= d(\beta^*\alpha^*) + d(\beta_\perp^*\lambda\alpha_\perp^*) \\
&= \beta^*(d\alpha^*) + (d\beta^*)\alpha^* + \beta_\perp^*\lambda(d\alpha_\perp^*) + \beta_\perp^*(d\lambda)\alpha_\perp^* + (d\beta_\perp^*)\lambda\alpha_\perp^* \\
&\overset{\lambda=0}{=} \beta^*(d\alpha^*) + (d\beta^*)\alpha^* + \beta_\perp^*(d\lambda)\alpha_\perp^* \\
&= \beta^*(d\alpha^*) + \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} (d\beta_2^*)\alpha^* + \beta_\perp^*(d\lambda)\alpha_\perp^*,
\end{aligned}
\tag{2.25}
$$

such that for the partial derivatives it holds that

$$
\left.\frac{\partial\,\mathrm{vec}(\Pi^*)}{\partial\,\mathrm{vec}(\alpha^*)'}\right|_{\lambda=0} = I_n \otimes \beta^*, \tag{2.26}
$$

$$
\left.\frac{\partial\,\mathrm{vec}(\Pi^*)}{\partial\,\mathrm{vec}(\beta_2^*)'}\right|_{\lambda=0} = \alpha^{*\prime} \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix}, \tag{2.27}
$$

$$
\text{and } \frac{\partial\,\mathrm{vec}(\Pi^*)}{\partial\,\mathrm{vec}(\lambda)'} = \alpha_\perp^{*\prime} \otimes \beta_\perp^*, \tag{2.28}
$$

see *e.g.* Magnus and Neudecker (1999).

Therefore, the determinant of the Jacobian reads,

$$
\begin{aligned}
|J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))||_{\lambda=0} &= \left| \left.\frac{\partial\,\mathrm{vec}(\Pi^*)}{\partial\,\mathrm{vec}(\alpha^*)'}\right|_{\lambda=0} \quad \left.\frac{\partial\,\mathrm{vec}(\Pi^*)}{\partial\,\mathrm{vec}(\beta_2^*)'}\right|_{\lambda=0} \quad \left.\frac{\partial\,\mathrm{vec}(\Pi^*)}{\partial\,\mathrm{vec}(\lambda)'}\right|_{\lambda=0} \right| \\
&= \left| I_n \otimes \beta^* \quad \alpha^{*\prime} \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \quad \alpha_\perp^{*\prime} \otimes \beta_\perp^* \right| \\
&= \left| \left(I_n \otimes \beta^* \quad \alpha^{*\prime} \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \quad \alpha_\perp^{*\prime} \otimes \beta_\perp^* \right)' \right. \\
&\qquad \left. \left(I_n \otimes \beta^* \quad \alpha^{*\prime} \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \quad \alpha_\perp^{*\prime} \otimes \beta_\perp^* \right) \right|^{\frac{1}{2}} \\
&= \begin{vmatrix} I_n \otimes \beta^{*\prime}\beta^* & \alpha^{*\prime} \otimes \beta_2' & 0 \\ \alpha^* \otimes \beta_2^* & \alpha^*\alpha^{*\prime} \otimes I_{n-r} & 0 \\ 0 & 0 & I_{n-r} \otimes I_{n-r} \end{vmatrix}^{\frac{1}{2}} \\
&= \begin{vmatrix} I_n \otimes \beta^{*\prime}\beta^* & \alpha^{*\prime} \otimes \beta_2^{*\prime} \\ \alpha^* \otimes \beta_2^* & \alpha^*\alpha^{*\prime} \otimes I_{n-r} \end{vmatrix}^{\frac{1}{2}},
\end{aligned}
\tag{2.29}
$$

where we have used the orthogonality properties of $\beta_\perp^*$ and $\alpha_\perp^*$.

The last expression from (2.29) equals the Jeffreys' prior from (2.14). We have thus shown that in the rescaled model both Jeffreys' rule and the conditional densities approach yield the same prior.

Posterior under a diffuse prior on Π^*

Specifying a diffuse prior on Π^*, that is $p_{lec}(\Pi^*|\Sigma) \propto 1$ the posterior conditional on Σ becomes

$$\Pi^*|\Sigma, Y \sim N(\hat{\Pi}^*, I_n \otimes I_n), \tag{2.30}$$

where $\hat{\Pi}^* = (Y_{-1}'Y_{-1})^{-\frac{1}{2}}Y_{-1}'\Delta Y\Sigma^{-\frac{1}{2}} = (Y_{-1}'Y_{-1})^{\frac{1}{2}}\hat{\Pi}\Sigma^{-\frac{1}{2}}$ when we define $\hat{\Pi} = (Y_{-1}'Y_{-1})^{-1}Y_{-1}'\Delta Y$. The marginal posterior of Σ in the unscaled model is

$$\begin{aligned} p(\Sigma|Y) \propto &|\Delta Y'\Delta Y - \hat{\Pi}'Y_{-1}'Y_{-1}\hat{\Pi}|^{(T-n)/2}|\Sigma|^{T/2} \\ &\exp(-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}\Delta Y'\Delta Y - \hat{\Pi}'Y_{-1}'Y_{-1}\hat{\Pi})), \end{aligned} \tag{2.31}$$

which is an inverted Wishart distribution with $T-1$ degrees of freedom and scale parameter matrix $\Delta Y'\Delta Y - \hat{\Pi}'Y_{-1}'Y_{-1}\hat{\Pi}$.

Not only the prior but also the posterior of $\alpha^*, \beta_2^*, \lambda|\Sigma, Y$ in the **UEC** model satisfies the transformation of random variables defined by (2.15) such that

$$p_{uec}(\alpha^*, \beta_2^*, \lambda|\Sigma, Y) = p_{lec}(\Pi^*|\Sigma, Y)|_{\Pi^*=\beta^*\alpha^*+\beta_\perp^*\lambda\alpha_\perp^*}|J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))|. \tag{2.32}$$

The conditional posterior $\alpha^*, \beta_2^*|\lambda, \Sigma, Y$ which is proportional to it can be evaluated in $\lambda = 0$ to obtain the posterior of $\alpha^*, \beta_2^*|\Sigma, Y$ in the **ECC** model.

$$\begin{aligned} p_{ecc}(\alpha^*, \beta_2^*|\Sigma, Y) &= p_{uec}(\alpha^*, \beta_2^*|\lambda, \Sigma, Y)|_{\lambda=0} \\ &\propto p_{uec}(\alpha^*, \beta_2^*, \lambda|\Sigma, Y)|_{\lambda=0} \\ &= p_{lec}(\Pi^*|\Sigma, Y)|_{\Pi^*=\beta^*\alpha^*}|J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))||_{\lambda=0}. \end{aligned} \tag{2.33}$$

It is identical to the posterior that is obtained using a Jeffreys' prior. Hence, we can consider the **ECC** model as a parameter realization $\lambda = 0$ in the **LEC**.

2.3 Simulating from the posterior

We employ the equivalence of the Jeffreys' prior and the prior based on the idea of a rescaled encompassing model to simulate from the posterior of the parameters of the cointegration model. First, parameter values are sampled from the rescaled model and then these are transformed back to obtain a sample from the posterior of the parameters of the (unscaled) **ECC** model under a Jeffreys' prior.

Sampling algorithm based on diffuse prior on Π^* and nesting

Straightforwardly simulating α^* and β_2^* using a Gibbs sampler is not possible due to the difficult dependence structure of the full conditional posterior densities of α^* and β_2^*, see Kleibergen en Van Dijk (1994). Π^* from the **LEC** model has *a posteriori* a normal distribution. The decomposition in (2.15) allows us to obtain a drawing of α^* and β_2^* (and λ) for the **UEC** model from a drawing of Π^*. The dependency of α^* and β_2^* is now avoided by determining α^* and β_2^* simultaneously.

This poses the problem that our posterior of interest, $p_{ecc}(\alpha^*, \beta_2^*|\Sigma, Y)$, does not involve λ while it is sampled. Chen (1994) suggests the following approach. For simulating from the posterior $p_{ecc}(\alpha^*, \beta_2^*|\Sigma, Y)$ it is first extended with an artificial extra parameter λ with density $g(\lambda|\alpha^*, \beta_2^*, \Sigma, Y)$. We use a Metropolis-Hastings (M-H) sampling algorithm, see *e.g.* Chib and Greenberg (1994), for simulating from the joint density

$$p_g(\alpha^*, \beta_2^*, \lambda, \Sigma|Y) = g(\lambda|\alpha^*, \beta_2^*, \Sigma, Y)p_{ecc}(\alpha^*, \beta_2^*, \Sigma|Y). \tag{2.34}$$

. The posterior $p_{uec}(\alpha^*, \beta_2^*, \lambda|\Sigma, Y)$ from (2.32) is used as the candidate generating density. When $p_g(\alpha^*, \beta_2^*, \lambda, \Sigma, Y)$ is marginalized with respect to λ in order to remove the artificial parameter λ, the resultant distribution is $p_{ecc}(\alpha^*, \beta_2^*, \Sigma|Y)$. The simulated values of $\alpha^*, \beta_2^*, \Sigma$ (discarding λ) therefore are a sample from $p_{ecc}(\alpha^*, \beta_2^*, \Sigma|Y)$.

The choice of $g(\lambda|\alpha^*, \beta_2^*, \Sigma, Y)$ leads to the weight function $w(\alpha^*, \beta_2^*, \lambda, \Sigma)$ for use in the M-H algorithm. The acceptance probability in the M-H depends on a weight function which is the ratio of the target density (2.34) and the candidate generating density (2.32),

$$\begin{aligned} w(\alpha^*, \beta_2^*, \lambda, \Sigma) =& \frac{p_g(\alpha^*, \beta_2^*, \lambda, \Sigma|Y)}{p_{uec}(\alpha^*, \beta_2^*, \lambda, \Sigma|Y)} \\ =& \frac{g(\lambda|\alpha^*, \beta_2^*, \Sigma, Y)p_{ecc}(\alpha^*, \beta_2^*|\Sigma, Y)}{p_{uec}(\alpha^*, \beta_2^*, \lambda, \Sigma|Y)} \\ =& \frac{g(\lambda|\alpha^*, \beta_2^*, \Sigma, Y)\exp(-\frac{1}{2}(\mathrm{tr}(\beta^*\alpha^* - \hat{\Pi}^*)'(\beta^*\alpha^* - \hat{\Pi}^*))}{\exp(-\frac{1}{2}(\mathrm{tr}(\beta^*\alpha^* + \beta_\perp^*\lambda\alpha_\perp^* - \hat{\Pi}^*)'(\beta^*\alpha^* + \beta_\perp^*\lambda\alpha_\perp^* - \hat{\Pi}^*))} \frac{|J||_{\lambda=0}}{|J|}. \end{aligned} \tag{2.35}$$

The exponentiated trace expressions in numerator and denominator are related to each other by

$$\begin{aligned} &\mathrm{tr}((\beta^*\alpha^* + \beta_\perp^*\lambda\alpha_\perp^* - \hat{\Pi}^*)'(\beta^*\alpha^* + \beta_\perp^*\lambda\alpha_\perp^* - \hat{\Pi}^*)) \\ =&\,\mathrm{tr}((\beta^*\alpha^* - \hat{\Pi}^*)'(\beta^*\alpha^* - \hat{\Pi}^*)) + \mathrm{tr}((\lambda - \beta_\perp^{*\prime}\hat{\Pi}^*\alpha_\perp^{*\prime})'(\lambda - \beta_\perp^{*\prime}\hat{\Pi}^*\alpha_\perp^{*\prime})) \\ &+ \mathrm{tr}((\beta_\perp^{*\prime}\hat{\Pi}^*\alpha_\perp^{*\prime})'\beta_\perp^{*\prime}\hat{\Pi}^*\alpha_\perp^{*\prime}) \\ =&\,\mathrm{tr}((\beta^*\alpha^* - \hat{\Pi}^*)'(\beta^*\alpha^* - \hat{\Pi}^*)) + \mathrm{tr}((\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda})) + \mathrm{tr}(\tilde{\lambda}'\tilde{\lambda}) \end{aligned} \tag{2.36}$$

where $\tilde{\lambda} = \beta_\perp^{*\prime}\hat{\Pi}^*\alpha_\perp^{*\prime}$. A sensible choice for the density function $g(\lambda|\alpha^*, \beta_2^*, \Sigma, Y)$ thus turns out to be

$$g(\lambda|\alpha^*, \beta_2^*, \Sigma, Y) \propto \exp(-\frac{1}{2}\,\mathrm{tr}((\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda}))). \tag{2.37}$$

Using this choice of $g(\lambda|\alpha^*, \beta_2^*, \Sigma, Y)$ the weight function reduces to

$$w(\alpha^*, \beta_2^*, \lambda, \Sigma) = (2\pi)^{-\frac{1}{2}(n-r)^2}\exp(-\frac{1}{2}\,\mathrm{tr}(\tilde{\lambda}'\tilde{\lambda}))\frac{|J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))||_{\lambda=0}}{|J(\Pi^*, (\alpha^*, \beta_2^*, \lambda))|}. \tag{2.38}$$

The steps required in the sampling algorithm are,

1. Draw Σ^{i+1} from $p_{ecc}(\Sigma|Y)$

2. Draw Π^{*i+1} from (2.30)

3. Compute $\alpha^{*i+1}, \beta_2^{*i+1}, \lambda^{i+1}$ from Π^{*i+1} using the singular value decomposition

4. Accept Σ^{i+1}, α^{*i+1} and β^{*i+1} with probability $\min\left(\frac{w(\alpha^{*i+1},\beta_2^{*i+1},\lambda^{i+1},\Sigma^{i+1})}{w(\alpha^{*i},\beta_2^{*i},\lambda^{i},\Sigma^{i})},1\right)$

Sampling algorithm can also be used for unscaled model

The idea behind simulation from the posterior of $\alpha, \beta_2|\Sigma$ is that α^* and β_2^* from the transformed model are sampled using the procedure from Kleibergen and Paap (1998). Subsequently, these values are transformed back to the original model. This is possible because both in the scaled and in the unscaled model the Jeffreys' prior is used. The invariance to scaling of the Jeffreys' prior ensures that the prior information is consistent among both models and that a posterior sampler in either model can be used. Obviously the posterior sampler of the well-behaved scaled model is preferred. The procedure can thus be extended with one extra step to obtain a drawing from the posterior of the unscaled model:

5 Compute $(\beta\alpha)^{i+1} = (Y_{-1}'Y_{-1})^{-\frac{1}{2}}\beta^{*i+1}\alpha^{*i+1}\Sigma^{\frac{1}{2}}$ and solve for α^{i+1} and β_2^{i+1} using the singular value decomposition.

3 Bayesian test for cointegration

We showed that the posterior of the parameters of the cointegration model using a Jeffreys' prior is equal to a conditional posterior of the parameters of the unrestricted error correction model, given that the parameter λ associated with the cointegration restriction is equal to zero. Hence, we can consider the cointegration model as a realization from the posterior of the parameters of the unrestricted error correction model. If we can determine whether this realization is from the tail or the area with the bulk of the probability mass, then we can use this information as a diagnostic device, because it reflects the plausibility of the cointegration restriction. It enables us to detect misspecification of the cointegration rank. An analytical expression of the marginal posterior of the parameter on which we need to condition is, however, unknown. We can then not use the marginal posterior to determine whether the realization resulted from the tail.

We therefore construct a statistic that is equal to the expectation of a function of λ and the remaining model parameters, such that it has a standardized distribution. The expectation is taken with respect to the posterior distribution from the unrestricted error correction model of the remaining parameters. We show that the function that we take the expectation of can be interpreted as the classical score statistic. A Bayesian test of the cointegration restrictions now consists of comparing the value of this statistic under the cointegration model to a χ^2 distribution.

3.1 Derivation

Distribution of λ

In (2.32) we showed how the joint posterior $\alpha^*, \beta_2^*, \lambda|\Sigma, Y$ is derived from $\Pi^*|\Sigma, Y$, which has a normal distribution, using the decomposition $\Pi^* = \beta^*\alpha^* + \beta_\perp^*\lambda\alpha_\perp^*$. Conversely, it holds that $\beta^*\alpha^* + \beta_\perp^*\lambda\alpha_\perp^*$, which is a function of $\alpha^*, \beta_2^*, \lambda$, also has a normal distribution

$$\beta^*\alpha^* + \beta_\perp^*\lambda\alpha_\perp^*|\Sigma, Y \sim N(\hat{\Pi}^*, I_n \otimes I_n). \tag{3.1}$$

Premultiplying this expression by $\beta_\perp^{*\prime}$, postmultiplying it by $\alpha_\perp^{*\prime}$ the left hand side reduces to λ because of the orthogonality properties of $\alpha_\perp^*$ and $\beta_\perp^*$. Conditioning on α^* and β_2^* and using properties of the matrix normal distribution we obtain that λ is matrix normal with mean $\beta_\perp^{*\prime}\hat{\Pi}^*\alpha_\perp^{*\prime} = \tilde{\lambda}$ and covariance matrix $\alpha_\perp\alpha_\perp' \otimes \beta_\perp'\beta_\perp = I_{n-r} \otimes I_{n-r}$, such that

$$\lambda|\alpha^*, \beta_2^*, \Sigma, Y \sim N(\tilde{\lambda}, I_{n-r} \otimes I_{n-r}). \tag{3.2}$$

Subtracting its conditional mean from λ it follows that $\lambda - \tilde{\lambda}|\alpha^*, \beta_2^*, \Sigma, Y$ has a matrix standard normal distribution. Note that $\lambda - \tilde{\lambda}$ depends on the conditioning parameters α^* and β_2^* but that its conditional distribution is matrix standard normal and consequently independent of the conditioning parameters.

The trace of the inner product of a matrix normal random variable has a χ^2 distribution and in this case

$$\mathrm{tr}((\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda}))|\alpha^*, \beta_2^*, \Sigma, Y \sim \chi^2_{(n-r)^2}. \tag{3.3}$$

The left hand side of this equation, however still depends on α^*, β_2^* and Σ. In order to remove the dependence we take the expectation of (3.3) with respect to α^*, β_2^* and Σ. Since the distribution of $\mathrm{tr}((\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda}))$ does not depend on α^*, β_2^* or Σ the expectation remains χ^2 distributed,

$$E_{\alpha^*, \beta_2^*, \Sigma}[\mathrm{tr}((\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda}))|\alpha^*, \beta^*, \Sigma, Y] \sim \chi^2_{(n-r)^2}. \tag{3.4}$$

So far, we have constructed a random quantity (3.4) that is a function of the parameters α^*, β_2^*, λ and Σ. The posterior distribution of the parameters determine the posterior distribution of the random quantity. We have constructed the quantity (3.4) such that it has a known distribution, namely the familiar χ^2 distribution.

Hypothesis of rank reduction: $\lambda = 0$

In the context of the cointegration model we are interested in testing the hypothesis of rank reduction of the long run multiplier parameter matrix Π^*. We have established in the previous section that the hypothesis can be represented by the restriction $H_0 : \lambda = 0$ when the decomposition of Π^* into α^*, β_2^* and λ is used. The random quantity (3.4) is derived under the **UEC** model or, stated differently, under the alternative hypothesis $H_1 : \lambda \neq 0$.

In Bayesian econometrics the plausibility of $H_0 : \lambda = 0$ can be assessed by calculating the posterior probability that the value of (3.4) evaluated under H_0 lies in the bulk of the probability mass of a $\chi^2_{(n-r)^2}$ distribution which it follows in the unrestricted case.

We substitute $\lambda = 0$ in (3.4) to calculate its value under H_0. Moreover, instead of taking the expectation with respect to the posterior of α^* and β^* from the **UEC** model we use the posterior of α^* and β^* under the null hypothesis, that is from the **ECC** model. Note that the posterior density $p_{ecc}(\alpha^*, \beta^*|\Sigma, Y)$ from the restricted model equals $p_{uec}(\alpha^*, \beta^*|\lambda, \Sigma, Y)|_{\lambda=0}$ from the unrestricted model,

$$E_{\alpha^*,\beta_2^*,\Sigma}[\mathrm{tr}((\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda}))]\Big|_{H_0} = E_{\Sigma} E_{\alpha^*,\beta_2^*|\lambda=0,\Sigma}[\mathrm{tr}(\tilde{\lambda}'\tilde{\lambda})]. \tag{3.5}$$

Remember that the parameters and data of the scaled model and the original, unscaled model are related by

$$\beta^*\alpha^* = (Y_{-1}'Y_{-1})^{\frac{1}{2}}\beta\alpha\Sigma^{-\frac{1}{2}}. \tag{3.6}$$

such that $\beta_\perp^* = (Y_{-1}'Y_{-1})^{-\frac{1}{2}}\beta_\perp$ and $\alpha_\perp^* = \alpha_\perp\Sigma^{\frac{1}{2}}$. The trace expression $\mathrm{tr}(\tilde{\lambda}'\tilde{\lambda})$ can thus be rewritten in terms of data and parameters of the unscaled model as

$$\begin{aligned}
&\mathrm{tr}((\beta_\perp^{*\prime}\hat{\Pi}^*\alpha_\perp^{*\prime})'(\beta_\perp^{*\prime}\hat{\Pi}^*\alpha_\perp^{*\prime}))\\
=\,&\mathrm{tr}(\alpha_\perp^{*\prime}\alpha_\perp^*\hat{\Pi}^{*\prime}\beta_\perp^*\beta_\perp^{*\prime}\hat{\Pi}^*)\\
=\,&\mathrm{tr}(\alpha_\perp'\alpha_\perp\Delta Y'Y_{-1}(Y_{-1}'Y_{-1})^{-1}\beta_\perp\beta_\perp'(Y_{-1}'Y_{-1})^{-1}Y_{-1}'\Delta Y)\\
=\,&\mathrm{tr}(\Sigma^{-\frac{1}{2}}M_{\Sigma^{-\frac{1}{2}}\alpha'}\Sigma^{-\frac{1}{2}}\Delta Y'(M_{Y_{-1}\beta} - M_{Y_{-1}})\Delta Y),
\end{aligned} \tag{3.7}$$

where $M_X = I - X(X'X)^{-1}X'$ and we have used that $\alpha_\perp'\alpha_\perp = \alpha_\perp'(\alpha_\perp\Sigma\alpha_\perp')^{-1}\alpha_\perp$ because $\alpha_\perp\Sigma\alpha_\perp' = \alpha_\perp^*\alpha_\perp^{*\prime} = I_{n-r}$ by definition. We also use that

$$\alpha_\perp'(\alpha_\perp\Sigma\alpha_\perp')^{-1}\alpha_\perp = \Sigma^{-1} - \Sigma^{-1}\alpha'(\alpha\Sigma^{-1}\alpha')^{-1}\alpha\Sigma^{-1} = \Sigma^{-\frac{1}{2}}M_{\Sigma^{-\frac{1}{2}}\alpha'}\Sigma^{-\frac{1}{2}}, \tag{3.8}$$

which is a well known result in cointegration analysis, see e.g. Johansen (1995). Similarly it holds that $\beta_\perp\beta_\perp' = \beta_\perp(\beta_\perp'(Y_{-1}'Y_{-1})^{-1}\beta_\perp)^{-1}\beta_\perp'$ since by definition $\beta_\perp'(Y_{-1}'Y_{-1})^{-1}\beta_\perp = \beta_\perp^{*\prime}\beta_\perp^* = I_{n-r}$ and that

$$\beta_\perp(\beta_\perp'(Y_{-1}'Y_{-1})^{-1}\beta_\perp)^{-1}\beta_\perp' = Y_{-1}'Y_{-1} - Y_{-1}'Y_{-1}\beta(\beta'Y_{-1}'Y_{-1}\beta)^{-1}\beta'Y_{-1}'Y_{-1}. \tag{3.9}$$

The quadratic form in the score of λ, $\frac{\partial \log L}{\partial\,\mathrm{vec}(\lambda)'}$, evaluated in $\lambda = 0$ also equals $\mathrm{tr}(\tilde{\lambda}'\tilde{\lambda})$. Hence, the test statistic can be interpreted as the posterior expectation of the classical score statistic. In our Bayesian procedure all calculations are conditional on the data, *i.e.* we treat the data as given. This is contrary to classical econometrics, where the data is considered random. In that case the nonstationary behaviour of the data in cointegration analysis influences the properties of the test statistics and more complicated distributions involving Brownian motions do occur.

3.2 Computation

The test statistic can be computed straightforwardly using the sampling scheme from the previous section. Let $\{\alpha^{(s)}, \beta_2^{(s)}, \Sigma_2^{(s)}\}_{s=1}^S$ be a sample of size S from the posterior of the **ECC** model. Then, for each drawing $(\alpha^{(s)}, \beta_2^{(s)}, \Sigma_2^{(s)})$ the corresponding value of $\tilde{\lambda}$ can be calculated, such that $\{\tilde{\lambda}^{(s)}\}_{s=1}^S$ is a sample of $\tilde{\lambda}$ from the **ECC** model.

The sample equivalent of the statistic under the null hypothesis can now be written as

$$E_{\Sigma}^{ecc} E_{\alpha^*,\beta_2^*|\Sigma}^{ecc}[\mathrm{tr}(\tilde{\lambda}'\tilde{\lambda})] \approx \frac{1}{S}\sum_{s=1}^{S} \mathrm{tr}(\tilde{\lambda}^{(s)\prime}\tilde{\lambda}^{(s)}). \tag{3.10}$$

3.3 Simulation results

In this section we discuss the results of a Monte Carlo study which is carried out to illustrate the performance of the sampling algorithm and the test statistic. We generate data using a 3-dimensional VAR(1) model in which the number of cointegration relations is varied from 0 to 3. It has an identity covariance matrix for the disturbances and the number of observations is $T = 250$. Throughout, a constant $\mu = (0.1\ 0.1\ 0.1)'$ is used. The long run multiplier parameter matrices of the processes read

I. $\Pi = \beta\alpha = I_r \begin{pmatrix} -0.2 & 0.2 & 0.2 \\ -0.2 & -0.2 & 0.2 \\ 0.2 & -0.2 & -0.6 \end{pmatrix},$

II. $\Pi = \beta\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -0.2 & 0.2 & 0.2 \\ 0.2 & -0.2 & 0.2 \end{pmatrix} = \begin{pmatrix} -0.2 & 0.2 & 0.2 \\ 0.2 & -0.2 & 0.2 \\ 0 & 0 & -0.4 \end{pmatrix},$

III. $\Pi = \beta\alpha = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} -0.2 & 0.2 & 0.2 \end{pmatrix} = \begin{pmatrix} -0.2 & 0.2 & 0.2 \\ 0.2 & -0.2 & -0.2 \\ 0 & 0 & 0 \end{pmatrix},$

IV. and $\Pi = 0_{3\times 3}$.

DGP I is a process with roots 0.8, 0.6 and 0.6, DGP II contains a unit root and two roots of 0.6, DGP III has two unit roots and a root 0.6 and DGP IV has three unit roots.

We use a VAR(1) model with a constant term to analyse the data and the analysis is carried out for 0 to 3 cointegration relations for each data set. We expect the statistic to detect too restricted specifications, because the parameter λ reflecting the rank reduction of the long-run multiplier parameter matrix then is forced to be zero, whereas it is not zero in the data generating process. In that case the statistic should have a relatively low value. In the 'true' model the statistic should obviously not indicate misspecification. In models which are less restricted than the DGP the statistic is not likely to detect this as the less restricted model is nested in the DGP. This phenomenon of nested

Table 1: Values of the Bayesian test statistics and p-values for the four DGP's

	Bayesian			classical	
r	statistic	df	tail prob.	LR	95%
			DGP I (rank(Π)=3)		
0	33.45	9	0.000	204.67	29.38
1	50.94	4	0.000	77.68	15.34
2	18.77	1	0.000	19.55	3.84
			DGP II (rank(Π)=2)		
0	32.05	9	0.000	151.78	29.38
1	19.01	4	0.001	66.33	15.34
2	1.01	1	0.315	1.01	3.84
			DGP III (rank(Π)=1)		
0	20.15	9	0.017	106.18	29.38
1	3.49	4	0.480	10.55	15.34
2	1.94	1	0.933	1.91	3.84
			DGP IV (rank(Π)=0)		
0	1.94	9	0.992	16.78	29.38
1	3.89	4	0.421	8.05	15.34
2	1.35	1	0.245	1.23	3.84

Note: the column marked 95% contains the 95% quantile of the likelihood ratio test.

hypotheses also occurs in the classical trace test for cointegration. The results of the Bayesian cointegration test statistics are summarized in Table 1. We also report a tail probability which is the probability that a $\chi^2_{(n-r)^2}$ distributed random variable is larger than the computed statistic. It can be given the interpretation of the Bayesian equivalent of the classical p-value. Evidently, the test statistic cannot be computed for DGP I, which involves a full rank matrix Π, and therefore no parameter λ associated with rank reduction is present there. For the other three DGP's there is indeed little indication of a misspecified cointegration rank for the 'true' model. The reported statistics are all within the (100-31.5)% quantile of its associated χ^2 distribution. In classical econometrics this would be interpreted as a p-value of 31.5%, which exceeds the usual significance levels. For all DGP's, the models that are more restricted than the DGP have Bayesian cointegration statistics that are far away from zero. This results in tail probabilities of 0.001 or less for DGP's I (at $r = 0, 1$ and 2) and II (at $r = 0$ and 1). Under DGP III the zero rank alternative ($r = 0$) has a tail probability of 1.7%, also giving strong evidence of a misspecified cointegration model.

3.4 Application

In this subsection we apply the Bayesian cointegration statistic to the Danish dataset of Johansen and Juselius (1990). They analyse the demand for money in Denmark using quarterly time series of real money M_2 (denoted by m_t), real income y_t, and the opportunity cost of holding money, which is approximated by the difference between the long term interest rate i_t^b (bond rate) and the short term interest rate i_t^d (deposit rate). The sample period is 1974.1-1987.3 and all variables are in logs.

Table 2: Values of the Bayesian test statistics and p-values for the Danish data from Johansen and Juselius (1990).

	Bayesian			classical	
r	statistic	df	tail prob.	LR	p-value
0	10.47	16	0.841	45.67	0.08
1	6.46	9	0.693	17.07	0.63
2	1.42	4	0.840	6.71	0.61
3	0.76	1	0.384	0.38	0.54

Note: The values of the classical trace test and the corresponding p-value are taken from Johansen and Juselius (1990).

They use a VAR(2) model with a constant μ and with (centered) seasonal dummies. They also propose to restrict the constant by $\alpha'_\perp \mu = 0$. In our analysis we leave the constant μ unrestricted in order to allow at least real income to have a linear trend. We present the posterior results of our method for $r = 1$ which is the number of cointegrating relations that Johansen and Juselius (1990) found to be appropriate using the cointegration trace test. The test statistics for this data set are summarized in Table 2. The statistic does not indicate a misspecified rank for $r = 1$ or larger. Unexpectedly, also no misspecification at $r = 0$ was detected. The small number of observations ($T = 54$) is probably responsible for this effect.

The posterior means of the components of α and β_2 are

$$\begin{array}{rccccl} & m_t & y_t & i_t^b & i_t^d & \\ \alpha' = (& -0.187 & 0.154 & 0.025 & 0.045 &), \\ & (0.084) & (0.074) & (0.024) & (0.02) & \end{array} \tag{3.11}$$

and

$$\begin{array}{rccccl} & m_t & y_t & i_t^b & i_t^d & \\ \beta = (& 1 & -1.07 & 5.14 & -4.47 &), \\ & (-) & (0.33) & (1.45) & (2.74) & \end{array} \tag{3.12}$$

and the values in parentheses are the posterior sample standard errors. These values correspond reasonably well to the maximum likelihood estimates of Johansen and Juselius (1990). The posterior sample mean and the sampling standard errors of β_2 should however be interpreted with caution. The marginal

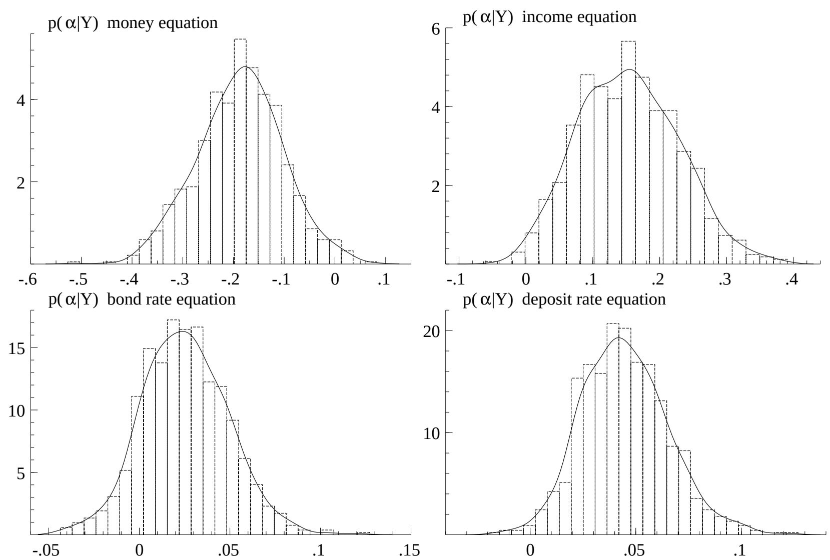

Figure 1: Marginal posteriors of the elements of α for the Danish data set using 1 cointegration relation.

posterior of β_2 has actually Cauchy type tails, see Kleibergen and Van Dijk (1998), such that the first moment does not exist. Therefore one might look at the sample median of β_2, but it is not reported here as it differs hardly from the sample mean.

The marginal posterior densities of α and β_2 are shown in Figures 1 and 2. The tails of β_2 are not shown because that would obscure the behaviour of the density near the center of the distribution. From Figure 2 it can be seen that the second and third component of β_2 are evidently skewed towards zero.

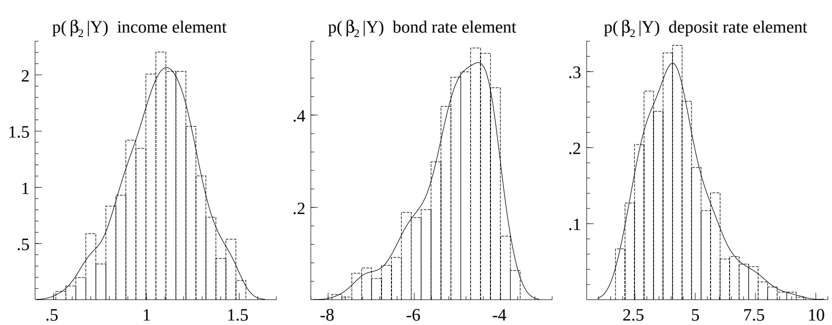

Figure 2: Marginal posteriors of the elements of β_2 for the Danish data set using 1 cointegration relation.

We asserted in the previous section that in the unrestricted model $\lambda - \tilde{\lambda}$ follows a matrix standard normal distribution. As a by-product of the sampling scheme, values of λ from the **UEC** model are sampled, because the posterior of the unrestricted model is used as the candidate generating density. The posterior density plots of the elements of $\lambda - \tilde{\lambda}$ in Figure 3 are all very similar to a standard normal density function. Also a more formal test like the Jarque-Bera normality test applied to each of the elements of $\lambda - \tilde{\lambda}$ confirms this finding.

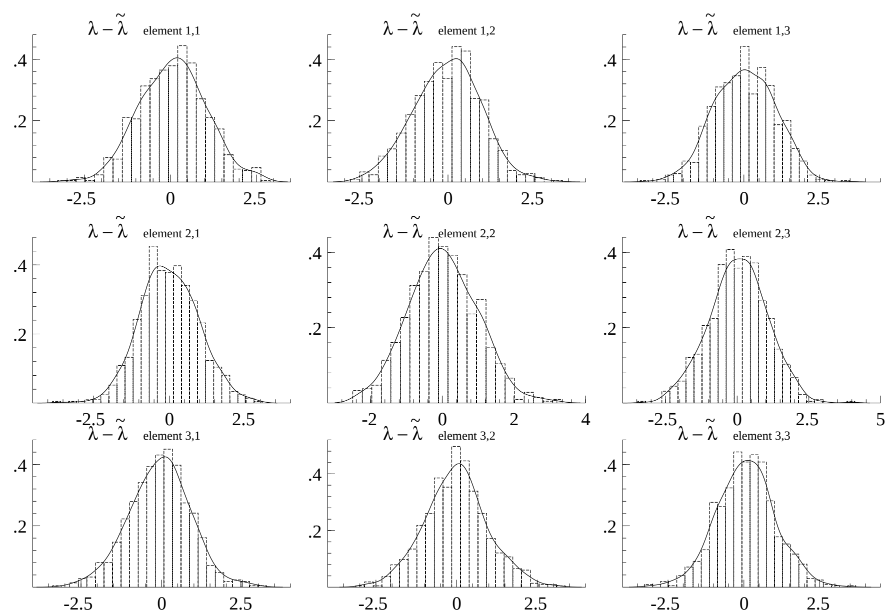

Figure 3: Marginal posteriors of the elements of $\lambda - \tilde{\lambda}$ for the Danish data set using 1 cointegration relation.

4 Model extensions

The Bayesian cointegration testing procedure that was developed in the previous sections uses a Jeffreys' prior. This Jeffreys' prior is conditional on the covariance matrix and the data. Hence, we can allow the covariance matrix or the matrices ΔY or Y_{-1} to be dependent on additional parameters,

$$\Delta Y(\theta) = Y(\theta)_{-1}\beta\alpha + \epsilon \tag{4.1}$$

with $\epsilon_t \sim N(0, \Sigma_t(\theta))$. After conditioning on the parameter vector θ the testing procedure and all other results from the previous sections can be applied. Obviously, we then need to specify priors on these additional parameters that are independent from the cointegration parameters. The resulting expression of the parameters in the singular value decomposition remains unchanged but the expression for Π^* changes however.

This extension enables us to conduct the cointegration testing procedure for models with mixtures of normal disturbances, such that time-varying heteroscedasticity can be included or Student-t distributed disturbances can be allowed for, the latter enrichment being a multivariate extension of *e.g.* Geweke (1993). In that case we let $\Sigma_t(\theta)$ be a random matrix from an inverted Wishart distribution with some degrees of freedom parameter and scale parameters, which are all contained in the parameter vector θ. Similarly, the matrices ΔY

or Y_{-1} can be a function of latent variables θ such that data with measurement errors or with structural breaks in the constant or trend can be modelled. Both topics are issues for further research.

5 Conclusions

In this paper we develop a Bayesian cointegration test statistic that can be used under a Jeffreys' prior. In order to do so, we start with constructing a posterior simulator to generate posterior drawings from the cointegration model. The simulator is based on the invariance properties of the Jeffreys' prior. We show that it is allowed to use the parameter drawings from a suitably rescaled model. For this rescaled model we use the posterior simulator from Kleibergen and Paap (1998). It is based on a transformation of the long-run multiplier matrix. The decomposition is related to the singular value decomposition and it specifies the full rank model in such a way that setting a particular, uniquely defined parameter to zero corresponds to imposing the cointegration restriction on the long-run multiplier. The final step in the posterior simulator consists of transforming back the drawings from the scaled model to the original, unscaled model.

The decomposition forms the basis of the derivation of the test statistic. We construct a function of the parameter that is associated with the cointegration restriction and the other model parameters, such that its posterior expectation has a χ^2 distribution under the assumption of a full rank value of the long run multiplier. The test statistic is equal to the posterior expectation of the classical score statistic and as both the Jeffreys' prior and the classical score statistic are invariant, the resulting test statistic is also invariant. We then evaluate whether the value of the test statistic under the restriction of cointegration is a plausible realization from its distribution under the encompassing, full rank model. The Bayesian cointegration statistic is a convenient diagnostic as it is evaluated under the cointegration model only. This then shows whether the cointegration assumption is a plausible restriction on the parameters of the unrestricted error correction model.

We apply the Bayesian cointegration statistic to four artificial examples to illustrate the use of the statistic as a diagnostic tool. According to the results of this Monte Carlo experiment, the test statistic is able to detect misspecified cointegration ranks when the rank reduction is too restrictive. The test is also applied to the Danish dataset of Johansen and Juselius (1990). Finally, we show that the test statistic can straightforwardly be extended to a more general model setting. For example, we show that structural breaks in the constant or trend and a general kind of disturbance distribution can be modelled, because conditional on some latent parameters all derivations still hold.

References

[1] Chen, M-H., *Importance-Weighted Marginal Bayesian Posterior Density Estimation*, Journal of the American Statistical Association **89** (1994), 818–824.

[2] Chib, S. and E. Greenberg, *Understanding the Metropolis-Hastings Algorithm*, The American Statistician **49** (1995), 327–335.

[3] Geweke, J., *Bayesian Treatment of the Independent Student-t Linear Model*, Journal of Applied Econometrics **8** (1993), 519–540.

[4] Johansen, S., *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford, 1995.

[5] Johansen, S. and K. Juselius, *Maximum Likelihood Estimation and Inference on Cointegration - with Applications to the Demand for Money*, Oxford Bulletin of Economics and Statistics **52** (1990), 169–210.

[6] Kleibergen, F. and H.K. van Dijk, *On the Shape of the Likelihood/Posterior in Cointegration Models*, Econometric Theory **10** (1994), 514–551.

[7] ______, *Bayesian Simultaneous Equation Analysis using Reduced Rank Structures*, Econometric Theory **14** (1998), 701–743.

[8] Kleibergen, F. and R. Paap, *Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration*, Econometric Institute Report 9821/A, Econometric Institute, Erasmus University Rotterdam, 1998.

[9] Magnus, J.R. and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition*, Wiley (Chichester), 1999.