Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration

Frank Kleibergen* Richard Paap†

Econometric Institute Report 9821/A

Abstract

Cointegration occurs when the long run multiplier of a vector autoregressive model exhibits rank reduction. Priors and posteriors of the parameters of the cointegration model are therefore proportional to priors and posteriors of the long run multiplier given that it has reduced rank. Rank reduction of the long run multiplier is modelled using a decomposition resulting from its singular value decomposition. It specifies the long run multiplier matrix as the sum of a matrix that equals the product of the adjustment parameters and the cointegrating vectors, *i.e.* the cointegration specification, and a matrix that models the deviation from cointegration. Priors and posteriors for the parameters of the cointegration model are obtained by restricting the latter matrix to zero in the prior and posterior of the unrestricted long run multiplier. The special decomposition of the long run multiplier results in unique posterior densities. This theory leads to a complete Bayesian framework for cointegration analysis. It includes prior specification, simulation schemes for obtaining posterior distributions and determination of the cointegration rank via Bayes factors. We illustrate the analysis with several simulated series, the UK data of Hendry and Doornik (1994) and the Danish data of Johansen and Juselius (1990).

Key words: cointegration, Bayesian analysis, Bayes factors.

1 Introduction

The introduction of the concept of cointegration by Engle and Granger (1987) has introduced a rapidly expanding literature on multivariate analysis of stochastic trends. This

*Corresponding author: Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL–3000 DR Rotterdam, The Netherlands, e-mail: kleibergen@few.eur.nl.

†Rotterdam Institute for Business Economic Studies, Erasmus University Rotterdam, P.O. Box 1738, NL–3000 DR Rotterdam, The Netherlands, e-mail: paap@few.eur.nl. Both authors thank the Netherlands Foundation for Scientific Research (N.W.O.) for their financial support. Comments from Peter Boswijk and participants of the 1997 ESEM meeting in Toulouse, the 1997 EC^2 meeting in Amsterdam, the 1996 world meeting of ISBA in Cape Town and seminars at CORE and the Tinbergen Institute, are gratefully acknowledged.

has lead to a largely unified theory of classical statistical analysis of cointegration, see among others Johansen (1991) and Phillips (1991). However, there does not exist a complete framework for Bayesian analysis of cointegration, like in the classical literature. The main contributions to Bayesian analysis of cointegration are: Koop (1991) analyses implied moving averages/impulse responses resulting from the Wold decomposition of a time series, DeJong (1992) considers the posterior distributions of the roots of vector autoregressive models, Kleibergen and van Dijk (1994b) analyse the consequences of local non-identification and prior specification on the posteriors of the parameters, Dorfman (1995) tests for the number of cointegrating vectors by analysing the difference between the number of unit roots in the different univariate models and the number of unit roots in the multivariate model, and Geweke (1996) proposes posterior simulators using the Gibbs sampler.

This paper differs from the previous papers in several ways. First, the above mentioned papers usually focus on one aspect of Bayesian cointegration analysis in vector autoregressive [VAR] models, for instance the computation of posterior distributions. In the present paper we propose a complete framework for cointegration analysis, including prior specification, simulation schemes for posterior distributions and cointegration rank determination based on posterior odds. Second, we introduce a new decomposition of the long run multiplier of the VAR model in error correction form. This decomposition of the long run multiplier allows us to condition on rank reduction of the long run multiplier in a unambiguous way. It specifies the long run multiplier as the sum of two matrices. The first matrix is the product of the adjustment parameter matrix and the cointegration vectors, *i.e.* the cointegration specification. The second matrix models the deviation from the cointegration specification. Cointegration occurs if the second matrix is zero. The decomposition results from a singular value decomposition as singular values are a natural way to represent rank reduction, see Golub and van Loan (1989). Priors and posteriors of the parameters of the cointegration model are now obtained by putting the matrix that models the deviation from the cointegration specification in the priors and posteriors of the parameters of the unrestricted model equal to zero. This way of construction priors allows us to extend the classes of priors for the parameters of linear models, for instance conjugate priors, to the parameters of the cointegration model. Furthermore, Bayes factors/posterior odds ratios to analyse rank reduction can be computed using a Savage-Dickey density ratio of Dickey (1971), see also Verdinelli and Wasserman (1995).

The outline of the paper is as follows. In Section 2 we briefly discuss cointegration and provide the new decomposition of the long run multiplier. In Section 3 we define prior distributions for the model parameters of the cointegration model. Section 4 discusses posterior distributions and contains simulation schemes to obtain posterior results. To compare cointegration models with different number of cointegrating relations we consider in Section 5 posterior odds to determine the cointegration rank. Section 6 illustrates the proposed cointegration analysis with several simulated series. Additionally, we consider the UK data analysed in Hendry and Doornik (1994) and the Danish data analysed in Johansen and Juselius (1990). Finally, Section 7 concludes.

2 The Cointegration Model

Consider a vector autoregressive model of order k [VAR(k)] for a n-dimensional vector of time series $\{Y_t\}_{t=1}^T$

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

where ε_t is an independent n-dimensional vector normal process with zero mean and $(n \times n)$ positive definite symmetric covariance matrix Σ. The $(n \times 1)$ vectors μ and τ contain the constant and trend coefficients and Φ_i, $i = 1, \ldots, k$ are $(n \times n)$ matrices with autoregressive coefficients. The initial values $Y_1, \ldots, Y_k$ are fixed. The VAR model in (1) can be written in the error correction form

$$\Delta Y_t = \mu + \tau t + \Pi\, Y_{t-1} + \sum_{i=1}^{k-1} \bar{\Phi}_i \Delta Y_{t-i} + \varepsilon_t, \tag{2}$$

where $\Pi = \sum_{j=1}^{k} \Phi_j - \mathbf{I}_n$ is the $(n \times n)$ long run multiplier matrix $\bar{\Phi}_i = -\sum_{j=i+1}^{k} \Phi_j$ are $(n \times n)$ matrices, $i = 1, \ldots, k-1$, see *e.g.* Johansen (1991).

The characteristic polynomial of model (1) is equal to $|\Phi(z)| = |\mathbf{I}_n z^k - \sum_{i=1}^{k} \Phi_i z^{k-i}|$. Since by definition $\Phi(1) = -\Pi$, unit roots enter the model when $\Phi(1)$ has a lower rank value. If Π is a zero matrix, the characteristic polynomial has n unit roots, which corresponds to n stochastic trends. Common stochastic trends appear if $(n - r)$ roots of the polynomial $|\Phi(z)|$ are equal to one, $0 < r < n$, see Johansen (1991). In that case the rank of Π equals r and we say that series generated by model (1) are cointegrated. Hence, cointegration implies that we can write the matrix Π as a product of two full rank $(n \times r)$ matrices α and β

$$\Pi = \alpha\beta'. \tag{3}$$

The matrix β contains the cointegrating vectors, which reflect the stationary long term relations (or equilibria) between the univariate series in Y_t. The α matrix contains the adjustment parameters, which indicate the speed of adjustment to the equilibria $\beta' Y_t$.

Since the number of parameters in $\alpha\beta'$, $2nr$ is larger than the number of free parameters in Π, under reduced rank $(= nr + (n - r)r)$ the α and/or β parameters have to be restricted to become estimable. A common restriction to identify β is $\beta'\beta = \mathbf{I}_r$. Since the results in this paper are not sensitive to the way we restrict β we choose, for notational convenience, for the specification

$$\beta' = (\mathbf{I}_r \;\; -\beta_2'), \tag{4}$$

where β_2 is a $((n - r) \times r)$ matrix. Note that due to this normalization the β matrix has always full rank.

To save on notation we write the error correction model (2) in matrix notation,

$$\Delta Y = Y_{-1}\Pi' + X\Phi + \varepsilon, \tag{5}$$

where $\Delta Y = (\Delta Y_{k+1} \ldots \Delta Y_T)'$, $Y_{-1} = (Y_k \ldots Y_{T-1})'$, $\varepsilon = (\varepsilon_{k+1} \ldots \varepsilon_T)'$, $X = (X'_{k+1} \ldots X'_T)'$, $X_t = (\Delta Y'_{t-1} \ldots \Delta Y'_{t-k+1}\ 1\ t))$, and $\Phi = (\bar{\Phi}_1 \ldots \bar{\Phi}_{k-1}\ \mu\ \tau)'$. To save even further on notation, we focus in the remainder of this paper on a simple VAR(1) model without deterministic elements. This is not a serious restriction since under a diffuse prior specification on Φ, integrating out the Φ parameters from the likelihood function leads to analysing a VAR(1) model for the transformed data $M_X \Delta Y$ and $M_X Y_{-1}$, where $M_X = \mathbf{I}_{T-k} - X(X'X)^{-1}X'$. We refer to this VAR(1) model in error correction form as a linear error correction [lec] model

$$\Delta Y = Y_{-1}\mathbf{\Pi} + \varepsilon, \tag{6}$$

where we define $\mathbf{\Pi} = \Pi'$ for notational convenience. Under the restriction $\Pi = \alpha\beta'$ this model simplifies to an error correction cointegration [ecc] model

$$\Delta Y = Y_{-1}\beta\boldsymbol{\alpha} + \varepsilon, \tag{7}$$

where $\boldsymbol{\alpha} = \alpha'$. To clarify the restriction that leads to rank reduction in $\mathbf{\Pi}$ we consider the following decomposition of $\mathbf{\Pi}$

$$\mathbf{\Pi} = \beta\boldsymbol{\alpha} + \beta_\perp\lambda\boldsymbol{\alpha}_\perp = (\beta\ \beta_\perp)\begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \lambda \end{pmatrix}\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}_\perp \end{pmatrix}, \tag{8}$$

where $\beta = (\mathbf{I}_r\ -\beta'_2)'$ and $\boldsymbol{\alpha}_\perp$ and $\beta_\perp$ are specified such that $\boldsymbol{\alpha}_\perp\boldsymbol{\alpha}' = \mathbf{0}$ with $\boldsymbol{\alpha}_\perp\boldsymbol{\alpha}'_\perp = \mathbf{I}_{n-r}$ and $\beta'\beta_\perp = \mathbf{0}$ with $\beta'_\perp\beta_\perp = \mathbf{I}_{n-r}$ and $\beta_\perp = (\beta_2\ \mathbf{I}_{n-r})'(\mathbf{I}_{n-r} + \beta'_2\beta_2)^{-\frac{1}{2}}$, $\boldsymbol{\alpha}_\perp = (\mathbf{I}_{n-r} + \boldsymbol{\alpha}'_1\boldsymbol{\alpha}_2^{-1'}\boldsymbol{\alpha}_2^{-1}\boldsymbol{\alpha}_1)^{-\frac{1}{2}}(\mathbf{I}_{n-r}\ -\boldsymbol{\alpha}'_1\boldsymbol{\alpha}_2^{-1'})$ where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2)$ with $\boldsymbol{\alpha}_1$ a $(r \times (n-r))$ matrix and $\boldsymbol{\alpha}_2$ a $(r \times r)$ matrix. We use that for a positive (semi)-definite real symmetric matrix M, $M^{\frac{1}{2}} = C\Lambda^{\frac{1}{2}}C'$ where Λ is a diagonal matrix containing the eigenvalues of M and C contains the orthonormal eigenvectors of M and $M^{-\frac{1}{2}} = C\Lambda^{-\frac{1}{2}}C'$, see *e.g.* Johansen (1995, p. 222). Using this decomposition we reparametrise the linear error correction model (6) in an unrestricted error correction [uec] model

$$\Delta Y = Y_{-1}\beta\boldsymbol{\alpha} + Y_{-1}\beta_\perp\lambda\boldsymbol{\alpha}_\perp + \varepsilon. \tag{9}$$

When $\lambda = \mathbf{0}$, the long run multiplier $\mathbf{\Pi}$ in (8) displays rank reduction and cointegration occurs. Note that the row- and columnspace of the matrix $(\beta_\perp\lambda\boldsymbol{\alpha}_\perp)$, which models the deviation from the cointegration specification $\beta\boldsymbol{\alpha}$, are spanned by the orthogonal complements of the cointegrating vectors β and the matrix of adjustment parameters $\boldsymbol{\alpha}$, respectively.

The decomposition of $\mathbf{\Pi}$ in (8) results from the singular value decomposition of $\mathbf{\Pi}$,

$$\mathbf{\Pi} = U\,S\,V', \tag{10}$$

where U and V are $(n \times n)$ orthonormal matrices, *i.e.* $U'U = \mathbf{I}_n$, $V'V = \mathbf{I}_n$ and S is an $(n \times n)$ diagonal matrix containing the positive singular values of $\mathbf{\Pi}$ (in decreasing order), see among others Golub and van Loan (1989) and Magnus and Neudecker (1988). If we write

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

with U_{11}, S_{11}, V_{11} $(r \times r)$, U_{22}, S_{22}, V_{22} $((n-r) \times (n-r))$, U_{21}, V_{12} $((n-r) \times r)$ and U_{12}, V_{21} $(r \times (n-r))$ matrices, we obtain the following expressions for $\boldsymbol{\alpha}$, λ and β_2

$$\begin{aligned} \boldsymbol{\alpha} &= U_{11} S_{11} (V_{11}' \; V_{21}') \\ \lambda &= (U_{22} U_{22}')^{-\frac{1}{2}} U_{22} S_{22} V_{22}' (V_{22} V_{22}')^{-\frac{1}{2}} \\ \beta_2 &= -U_{21} U_{11}^{-1}, \end{aligned} \tag{12}$$

where the square roots of the matrices are defined below (8), see Kleibergen (1998). From (12) if follows that λ consists of an orthogonal transformation of the $(n-r)$ smallest singular values of $\mathbf{\Pi}$, which end up in S_{22}. The singular values are the square root of the eigenvalues of the symmetric matrix $\mathbf{\Pi}'\mathbf{\Pi}$ and hence are always real. The number of non-zero singular values determines the rank of a matrix, see Golub and van Loan (1989). Since λ is an orthogonal transformation of the smallest singular values, restricting λ to zero is equivalent to restricting the smallest singular values to zero and is thus an unambiguous way of restricting the rank of the long run multiplier. The decomposition (8) does therefore not suffer from dependence of the order of the variables in the VAR model like in Kleibergen and van Dijk (1994b).

The Likelihood Function

The likelihood function of the unrestricted error correction model (9) conditional on the initial observations Y_1 is given by

$$\begin{aligned} \mathcal{L}_{lec}(Y|\Sigma, \mathbf{\Pi}) &= (\sqrt{2\pi})^{-(T-1)n} |\Sigma \otimes \mathbf{I}_n|^{-\frac{1}{2}} \exp(-\frac{1}{2}(\text{vec}(\varepsilon)'(\Sigma^{-1} \otimes \mathbf{I}_n^{-1})\text{vec}(\varepsilon))) \\ &= (\sqrt{2\pi})^{-(T-1)n} |\Sigma|^{-\frac{1}{2}(T-1)} |\mathbf{I}_{T-1}|^{-\frac{1}{2}n} \exp(-\frac{1}{2}\text{tr}(\Sigma^{-1}\varepsilon'\varepsilon)), \end{aligned} \tag{13}$$

where ε is given in (6). In addition, the conditional likelihood function of the unrestricted error correction model (9) and the cointegration model (7) are

$$\mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2) = \mathcal{L}_{lec}(Y|\Sigma, \mathbf{\Pi})|_{\mathbf{\Pi} = \beta\boldsymbol{\alpha} + \beta_\perp \lambda \boldsymbol{\alpha}_\perp} \tag{14}$$

$$\mathcal{L}_{ecc}(Y|\Sigma, \boldsymbol{\alpha}, \beta_2) = \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda = \mathbf{0}}, \tag{15}$$

respectively.

In the next section we propose a prior framework to analyse the unrestricted error correction models (9) and the cointegration models (7).

3 Prior Specification

Traditional Bayesian analysis of the cointegration model starts directly with specifying priors on the parameters Σ, $\boldsymbol{\alpha}$ and β_2 in the cointegration model (7). The cointegration model (7) is non-linear in the parameters $\boldsymbol{\alpha}$ and β_2. It is easy to see that the parameter β_2 is not identified when $\boldsymbol{\alpha} = \mathbf{0}$ (or when $\boldsymbol{\alpha}$ is of reduced rank), see Phillips (1989) for more discussion on local non-identification. Consequently, if a diffuse prior is used, such that the joint posterior of the parameters is proportional to the likelihood, the conditional posterior of β_2 given $\boldsymbol{\alpha}$ is constant and non-zero when $\boldsymbol{\alpha} = \mathbf{0}$. The integral over this conditional posterior at $\boldsymbol{\alpha} = \mathbf{0}$, which is part of the marginal posterior of $\boldsymbol{\alpha}$, is therefore proportional to the volume of the parameter region of β_2 ($\mathbb{R}^{(n-r)r}$), which is infinity. This leads to a *a posteriori* favour for locally non-identified parameter values when diffuse priors are used for the parameters $(\boldsymbol{\alpha}, \beta_2)$, see Kleibergen and van Dijk (1994b) for a more elaborate discussion of this phenomenon. This problem with local non-identification disappears if one specifies proper priors. However, if under a proper specification $\boldsymbol{\alpha} = \mathbf{0}$, the posterior of β_2 is completely determined by the prior on β_2 and Bayes factors to compare models with different number of cointegration relations may in such case be very sensitive to prior specification.

The approach in this paper is based on the idea that the error correction cointegration model (7) is nested in the linear error correction model (6) and results from imposing a reduced rank restriction ($\lambda = \mathbf{0}$) on the long run multiplier $\boldsymbol{\Pi}$. Since this is also true for the likelihood functions of the models, we also want this to hold for prior and therefore posterior distributions. The parametrisation of rank reduction of $\boldsymbol{\Pi}$ using $(\boldsymbol{\alpha}, \lambda, \beta_2)$ in (8) is such that starting from a density $p(\boldsymbol{\Pi})$ for unrestricted values for $\boldsymbol{\Pi}$ it leads to a unique expression of the conditional density $p_r(\boldsymbol{\Pi})$ for reduced rank values of $\boldsymbol{\Pi}$

$$\begin{aligned} p_r(\boldsymbol{\Pi}) &\propto p(\boldsymbol{\Pi})|_{\text{rank}(\boldsymbol{\Pi})=r} \\ &\propto p(\boldsymbol{\Pi}(\boldsymbol{\alpha}, \beta_2, \lambda))|_{\lambda=\mathbf{0}} |J(\boldsymbol{\alpha}, \beta_2, \lambda)|_{\lambda=0}|, \end{aligned} \tag{16}$$

where $|J(\boldsymbol{\alpha}, \beta_2, \lambda)|$ represents the Jacobian of the transformation from $\boldsymbol{\Pi}$ to $(\boldsymbol{\alpha}, \beta_2, \lambda)$, $|_{\lambda=\mathbf{0}}$ stands for evaluated in $\lambda = \mathbf{0}$ and the subscript r denotes restricted, see Kleibergen (1998). The conditional density (16) is unique since it is constructed by restricting an orthogonal transformation of the smallest singular values of $\boldsymbol{\Pi}$ to zero, see also below. The density $p(\boldsymbol{\Pi})$ can be a posterior or a prior of $\boldsymbol{\Pi}$ and may depend on hyperparameters. Given a prior or posterior on $\boldsymbol{\Pi}$ we can use (16) to construct the prior or posterior on the parameters of $\boldsymbol{\Pi}$ under rank reduction, *i.e.* the cointegration model (7).

For instance, (diffuse) priors for the Σ and $\boldsymbol{\Pi}$ parameters in the linear error correction model (6) implies via the Jacobian transformation $J(\boldsymbol{\alpha}, \lambda, \beta_2)|$ priors for the Σ, $\boldsymbol{\alpha}$, λ and β_2 parameters in the unrestricted error correction model (9), which leads to well-behaved posterior distributions. The joint prior of the parameters of the cointegration model (7) is now proportional to the joint prior of the parameters of the unrestricted error correction model (9) evaluated in $\lambda = \mathbf{0}$. Since the likelihood of the cointegration model (15) as a function of the parameters results from the likelihood of the unrestricted error correction

model (14) evaluated in $\lambda = \mathbf{0}$, the posterior of the parameters of the cointegration model results from restricting $\lambda = \mathbf{0}$ in the posterior of the parameters of the unrestricted error correction model (9), *i.e.* the posterior of $\mathbf{\Pi}$ under rank reduction.

Normally, the construction of posteriors and priors using conditional densities like (16) suffers from the Borel-Kolmogorov paradox, see Kolmogorov (1950) and Billingsley (1986). Consequently, the posterior or prior then depends on the way we restrict the $\mathbf{\Pi}$ matrix to obtain rank reduction and therefore on the cointegration specification. In other words it depends on the way λ is defined in the restriction $\lambda = \mathbf{0}$. For instance, restricting $\lambda = \mathbf{0}$ in the decomposition of $\mathbf{\Pi}$ in $(\boldsymbol{\alpha}, \lambda, \beta_2)$ in Kleibergen and van Dijk (1994a) may lead to a different posterior than the decomposition in (8). Also the restriction $\nu = \lambda/(1-\lambda) = 0$, which also implies rank reduction may lead to a different posterior. Although this Borel-Kolmogorov paradox cannot be avoided, Kleibergen (1998) shows that it is possible to obtain unique prior and posterior densities if the transformation from $\mathbf{\Pi}$ to $(\boldsymbol{\alpha}, \lambda, \beta_2)$ allows us to uniquely obtain $\mathbf{\Pi}$ from every value of $(\boldsymbol{\alpha}, \lambda, \beta_2)$ given that $(\boldsymbol{\alpha}, \beta_2)$ allow for it. Only specifications of λ which satisfy this condition allow us to condition on the restriction of rank reduction and on nothing else. Note that this condition is stronger than just invertibility. It is not allowed that different values of λ lead to the same value of $\mathbf{\Pi}$. The decomposition of $\mathbf{\Pi}$ in (8) satisfies this condition, since λ it is an orthogonal transformation of the singular values. However, the decomposition of $\mathbf{\Pi}$ in Kleibergen and van Dijk (1994a) does not satisfy this condition since conditioning on $\lambda = \mathbf{0}$ in that case does not always corresponds with the desired kind of rank reduction. Also conditioning on $\nu = \lambda/(1-\lambda) = 0$ does not satisfy this condition since we cannot retain the value of $(\boldsymbol{\alpha}, \lambda, \beta_2)$ from $(\boldsymbol{\alpha}, \nu, \beta_2)$ when $\lambda = 1$ and thus also not the value of $\mathbf{\Pi}$. Our prior framework therefore leads to unique posteriors within the limited class of restrictions that lead to rank reduction, see Kleibergen (1998) for details.

Finally, note that the above proposed construction of prior distributions also allows us to construct conjugate priors for the parameters of the cointegration model. If we specify a conjugate prior for the parameters of the linear error correction model (6), we have specified a conjugate prior for the parameters of the unrestricted error correction model (9). Evaluating this prior in $\lambda = \mathbf{0}$ then gives us the conjugate prior for the parameters of the error correction cointegration model (7). Furthermore, if the prior of the parameters of the error correction cointegration model is conjugate, this is also true for the posterior of the parameters of the error correction cointegration model and hence the prior and posterior of $(\boldsymbol{\alpha}, \beta_2)$ are of the same type.

Linear Error Correction Model

A conjugate prior for Σ and $\mathbf{\Pi}$ in the linear error correction model (6) consists of an inverted Wishart prior for Σ and a matrix normal prior for $\mathbf{\Pi}$ given Σ

$$\begin{aligned} p_{lec}(\Sigma) &\propto |S|^{\frac{1}{2}h}|\Sigma|^{-\frac{1}{2}(h+n+1)}\exp(-\frac{1}{2}tr(\Sigma^{-1}S)), \\ p_{lec}(\mathbf{\Pi}|\Sigma) &\propto |\Sigma|^{-\frac{1}{2}n}|A|^{\frac{1}{2}n}\exp(-\frac{1}{2}tr(\Sigma^{-1}(\mathbf{\Pi}-P)'A(\mathbf{\Pi}-P))), \end{aligned} \tag{17}$$

where h and the positive definite symmetric [PDS] $(n \times n)$ matrix S are prior parameters for the inverted Wishart and the PDS $(n \times n)$ matrix A and the $(n \times n)$ matrix P are prior parameters for the matrix normal prior. Therefore, the marginal prior for $\mathbf{\Pi}$ is a matrix t density. A diffuse (non-informative) prior specification for Σ and $\mathbf{\Pi}$ given Σ is given by $p_{lec}(\Sigma) \propto |\Sigma|^{-\frac{1}{2}(n+1)}$ and $p_{lec}(\mathbf{\Pi}|\Sigma) \propto 1$.

Unrestricted Error Correction Model

The joint prior of the parameters of the unrestricted error correction model (9) results directly from the joint prior of the linear error correction model $p_{lec}(\Sigma, \mathbf{\Pi})$ (17)

$$p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2) \propto p_{lec}(\Sigma, \mathbf{\Pi})|_{\mathbf{\Pi}=\beta\boldsymbol{\alpha}+\beta_\perp\lambda\boldsymbol{\alpha}_\perp}|J(\boldsymbol{\alpha}, \lambda, \beta_2)|, \tag{18}$$

where $|J(\boldsymbol{\alpha}, \lambda, \beta_2)|$ is the Jacobian of the transformation from $\mathbf{\Pi}$ to $(\boldsymbol{\alpha}, \lambda, \beta_2)$. The derivation of this Jacobian and its functional form are given in Appendix A. It is not possible to write the joint prior as a product of marginal and/or conditional priors of a known type. Note that the marginal prior of Σ is still an inverted Wishart density. The marginal priors of the remaining parameters can be obtained through simulation, see Section 4.

Cointegration Model

The joint prior of the parameters of the error correction cointegration model (7) is proportional to the joint prior of the unrestricted model (18) evaluated in $\lambda = \mathbf{0}$,

$$\begin{aligned} p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2) &= \frac{1}{c_r}p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}} \\ &= \frac{1}{c_r}p_{uec}(\Sigma, \mathbf{\Pi})|_{\mathbf{\Pi}=\beta\boldsymbol{\alpha}}|J(\boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}|, \end{aligned} \tag{19}$$

where c_r is a correction factor which corrects the integrating constant and

$$\begin{aligned} |J(\boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}| &= |(\mathbf{I}_n \otimes \beta) \ (\boldsymbol{\alpha}' \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}) \ (\boldsymbol{\alpha}'_\perp \otimes \beta_\perp)| \\ &= \left| \begin{pmatrix} \mathbf{I}_n \otimes \beta'\beta & \boldsymbol{\alpha}' \otimes \beta'_2 \\ \boldsymbol{\alpha} \otimes \beta_2 & \boldsymbol{\alpha}\boldsymbol{\alpha}' \otimes \mathbf{I}_{n-r} \end{pmatrix} \right|^{\frac{1}{2}}, \end{aligned} \tag{20}$$

see Appendix A. The correction factor c_r equals

$$c_r = \iiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2. \tag{21}$$

Note that for the diffuse prior specification c_r is not defined. For determining the posterior distributions the factor c_r does not matter since it is a constant. However, for the cointegration rank determination via posterior odds analysis the value of this factor is needed. Section 5.1 shows how to calculate c_r through simulation. Furthermore, Section 4 shows how to simulate from the marginal prior distributions $p_{ecc}(\Sigma)$, $p_{ecc}(\boldsymbol{\alpha})$ and $p_{ecc}(\beta_2)$ which are of unknown analytical form while the conditional prior of Σ given (α, β) is an inverted Wishart density.

3.1 Specification of the Prior in Practise

When constructing priors for the parameters of cointegration models, we typically do not have prior information regarding the unrestricted long run multiplier Π but we have some prior beliefs about the cointegrating vectors β, their loading factors α and the cointegration rank r. The prior framework based on a prior for the long run multiplier is based on the first property though. Given the prior on the long run multiplier, for example a Minnesota prior, see Doan *et al.* (1984), it gives the implied prior on the parameters of the error correction cointegration model. We can however also specify the implied prior on α and β_2 directly, such that it corresponds in functional form to the implied prior which would result from a conjugate prior on the long run multiplier. We then substitute the information we have regarding the cointegrating vector and loading vectors in this prior which implies a conjugate prior on the parameter of the long run multiplier.

The prior for $(\Sigma, \boldsymbol{\alpha}, \beta_2)$ in the cointegration model (7) implied by the conjugate prior specification (17) reads

$$\begin{aligned} p_{ecc}(\Sigma, \alpha, \beta_2) \propto & \left| \begin{pmatrix} \mathbf{I}_n \otimes \beta'\beta & \boldsymbol{\alpha}' \otimes \beta_2' \\ \boldsymbol{\alpha} \otimes \beta_2 & \boldsymbol{\alpha}\boldsymbol{\alpha}' \otimes \mathbf{I}_{n-r} \end{pmatrix} \right|^{\frac{1}{2}} \\ & |\Sigma|^{-\frac{1}{2}(h+2n+1)} \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(S + (\beta\boldsymbol{\alpha} - ba)'A(\beta\boldsymbol{\alpha} - ba)))), \end{aligned} \tag{22}$$

where $ba = P$ with b a $(n \times r)$ and a an $(r \times n)$ matrix and S and A are the remaining prior parameters. By specifying P as ba we can now reflect the information we have with respect to the location of the cointegrating vector and their loading vectors in the specification of b and a respectively. Note that the dimensions of b and a also reflect our prior beliefs about the cointegration rank. The scale matrices S and A allow us to reflect our prior ideas about the correlation and variances of β_2 and $\boldsymbol{\alpha}$, where it is such that the prior variance of both β_2 and $\boldsymbol{\alpha}$ are decreasing in A and the prior variance of $\boldsymbol{\alpha}$ is increasing in S. The prior (22) results from the conjugate prior (17) with $P = ba$ and shows the functional form of a conjugate prior on the parameters of the cointegration model. Note that the prior (22) incorporates the local non-identification of β_2 for lower

rank values of α since it is flat and zero at these parameter values which shows that the prior variance of β_2 is infinite at lower rank values of $\boldsymbol{\alpha}$.

Finally, we emphasize the importance of using priors on the parameters of error correction cointegration models which are implied by priors on the long run multiplier, whose properties are well understood. The linear error correction model is linear in the long run multiplier $\boldsymbol{\Pi}$ and it is therefore well known how prior information is updated to posterior information and that all properties of its prior are thus reflected in its posterior. The cointegration model is non-linear in $\boldsymbol{\alpha}$ and β_2 and it is not directly clear how the prior information is updated to posterior information, if one specifies a prior directly on its parameters. To verify the plausibility of a specified prior on the parameters of the cointegration model, one has to construct the implied prior on the long run multiplier. Since the linear error correction model is linear in the long run multiplier, the properties of the implied prior are reflected in the posterior of the long run multiplier and hence also in the posterior of the parameters of the cointegration model, see Kleibergen (1998). Using this approach Kleibergen (1998) shows that specifying independent normal priors directly on $\boldsymbol{\alpha}$ and β_2 on models with different number of cointegration relations can lead to incoherent Bayes factors.

4 Posterior Distributions

In this section we discuss the posterior distributions of the parameters of the linear error correction model (6), the unrestricted error correction model (9) and the cointegration model (7). We focus on the posteriors of the parameters under the conjugate prior specification. The posteriors under diffuse prior specification follow from the conjugate specification by putting $h = -n$, $S = \mathbf{0}$ and $A = \mathbf{0}$.

Linear Error Correction Model

The posterior of the parameters of the linear error correction model (6) is proportional to the prior (17) times the likelihood (13). Since it can be seen as a multivariate linear regression model, the marginal posterior of Σ and the conditional posterior of $\boldsymbol{\Pi}$ given Σ are just

$$\begin{aligned} p_{lec}(\Sigma|Y) \propto\ & |S + P'AP + \Delta Y'\Delta Y - \tilde{\boldsymbol{\Pi}}'(A + Y_{-1}'Y_{-1})\tilde{\boldsymbol{\Pi}}|^{\frac{1}{2}(T+h)}|\Sigma|^{-\frac{1}{2}(T+h+n)} \\ & \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(S + P'AP + \Delta Y'\Delta Y - \tilde{\boldsymbol{\Pi}}'(A + Y_{-1}'Y_{-1})\tilde{\boldsymbol{\Pi}}))), \\ p_{lec}(\boldsymbol{\Pi}|\Sigma, Y) \propto\ & |\Sigma|^{-\frac{1}{2}n}|(A + Y_{-1}'Y_{-1})|^{\frac{1}{2}n} \\ & \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(\boldsymbol{\Pi} - \tilde{\boldsymbol{\Pi}})'(A + Y_{-1}'Y_{-1})(\boldsymbol{\Pi} - \tilde{\boldsymbol{\Pi}}))), \end{aligned} \tag{23}$$

with

$$\tilde{\boldsymbol{\Pi}} = (A + Y_{-1}'Y_{-1})^{-1}(AP + Y_{-1}'\Delta Y), \tag{24}$$

see *e.g.* Zellner (1971). The marginal posterior of Σ is just like the prior an inverted Wishart distribution. The conditional posterior of $\mathbf{\Pi}$ given Σ is a matrix normal distribution. Hence the marginal posterior of $\mathbf{\Pi}$ is a matrix t density, see Zellner (1971).

Unrestricted Error Correction Model

The posterior of the parameters of the unrestricted error correction model is proportional to the prior (18) times the likelihood $\mathcal{L}_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)$ defined in (14).

$$p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y) \propto p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2). \tag{25}$$

This posterior density cannot be decomposed into a product of conditional and/or marginal densities belonging to a known class of probability density functions. We can however simulate from this posterior since we can easily simulate from the posterior of the linear error correction model (23) and it holds that

$$p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y) = p_{lec}(\Sigma, \mathbf{\Pi}|Y)|_{\mathbf{\Pi}=\beta\boldsymbol{\alpha}+\beta_\perp\lambda\boldsymbol{\alpha}_\perp} |J(\boldsymbol{\alpha}, \lambda, \beta_2)|. \tag{26}$$

The simulation scheme is as follows

Step 1: Draw Σ^i from $p_{lec}(\Sigma|Y)$ given in (23).

Draw $\mathbf{\Pi}^i$ from $p_{lec}(\mathbf{\Pi}|\Sigma^i, Y)$ given in (23).

Step 2: Perform a singular value decomposition of $\mathbf{\Pi}^i = U^i\, S^i\, V^{i\prime}$.

Step 3: Compute $\boldsymbol{\alpha}^i$, λ^i and β_2^i using (12).

The simulated values $\boldsymbol{\alpha}^i$, λ^i and β_2^i can be used to determine marginal results. Likewise, we can use this simulation scheme to obtain marginal prior results for $\boldsymbol{\alpha}$, λ and β_2.

Cointegration Model

The posterior of the parameters of the error correction cointegration model (7) is proportional to the prior (19) times the likelihood (15)

$$p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2|Y) \propto p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2)\, \mathcal{L}_{ecc}(Y|\Sigma, \boldsymbol{\alpha}, \beta_2). \tag{27}$$

Again this joint posterior cannot be decomposed in marginal and/or conditional posterior densities of a known type. Since the prior and the likelihood function of the parameters of the cointegration model (7) equals the prior and likelihood function of the parameters of the unrestricted error correction model (9) evaluated in $\lambda = \mathbf{0}$ it holds that

$$\begin{aligned} p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2|Y) &\propto p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}} \\ &\propto p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)|_{\lambda=\mathbf{0}}. \end{aligned} \tag{28}$$

We use this relation to set up a Markov Chain Monte Carlo simulation algorithm to obtain posterior results. Since the full conditional posterior distributions are of a unknown type, standard Gibbs sampling is not possible. Therefore, we apply the Metropolis-Hastings sampler of Metropolis *et al.* (1953) and Hastings (1970).

To describe the Metropolis-Hastings [M-H] sampling algorithm, let ψ be a random variable with density function $f(\psi)$. Let $g(\zeta|\psi)$ be a *candidate-generating* density function in ζ. The simulation algorithm to sample from the density $f(\psi)$ works as follows:

Step 1: Specify starting values ψ^0 and set $i = 0$.

Step 2: Simulate ζ from $g(\zeta|\psi^i)$.

$$\text{Define } a(\zeta, \psi^i) = \begin{cases} \min\left(\frac{f(\zeta)g(\psi^i|\zeta)}{f(\psi^i)g(\zeta|\psi^i)}, 1\right) & f(\psi^i)g(\zeta|\psi^i) > 0 \\ 1 & f(\psi^i)g(\zeta|\psi^i) = 0. \end{cases}$$

Choose $\psi^{i+1} = \zeta$ with probability $a(\zeta, \psi^i)$

and $\psi^{i+1} = \psi^i$ with probability $(1 - a(\zeta, \psi^i))$.

Step 3: Set $i = i + 1$ and go to step 2.

The described iterative scheme generates a Markov chain. After the chain has converged, say at H iterations, the simulated values $\{\psi^i, i \geq H\}$ can be used as a sample from the distribution of ψ to compute means, variances, *etc.* Different choices for the *candidate-generating* function result in different specific forms of the algorithm. For example, if $g(\psi^i|\zeta) = g(\zeta|\psi^i)$ the acceptance probability simplifies to $a(\zeta, \psi^i) = \min(f(\zeta)/f(\psi^i), 1)$. This describes the original Metropolis algorithm. If $g(\zeta, \psi^i) = g(\zeta)$, we get $a(\psi^i, \zeta) = \min(w(\zeta)/w(\psi^i), 1)$, where $w(\zeta) = f(\zeta)/g(\zeta)$, which can be interpreted as importance weights. For details we refer to Smith and Roberts (1993) and Tierney (1994).

If we opt for a M-H algorithm, we can take the posterior of the parameters of the unrestricted error correction model (25) as *candidate-generating* density function, since we have already shown how to sample from this distribution. However, in this case we also sample λ which does not show up in the posterior of the parameters of the cointegration model (27). To circumvent this problem we extend the posterior of the parameters of the cointegration model (27) with a proper conditional density $g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)$[1]

$$g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)\, p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2|Y) \propto g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)\, p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)|_{\lambda=\mathbf{0}}, \tag{29}$$

and sample from this distribution using the M-H approach. Since g is a proper density, the draws Σ, $\boldsymbol{\alpha}$ and β_2 can be seen as draws from the posterior (27). The acceptance-rejection step depends on the ratio of the extended posterior of the parameters of the cointegration

[1]This solution is based on the ideas in Chen (1994).

model (29) and the posterior of the parameters of the unrestricted error correction model (25)

$$w(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2) = \frac{g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)\, p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)|_{\lambda=\mathbf{0}}}{p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)}. \tag{30}$$

Implementing this in a M-H sampler results in

Step 1: Draw $(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}, \lambda^{i+1}, \beta_2^{i+1})$ from (25).

Step 2: Accept $(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}, \lambda^{i+1}, \beta_2^{i+1})$ with probability $\min\left(\frac{w(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}, \lambda^{i+1}, \beta_2^{i+1})}{w(\Sigma^{i}, \boldsymbol{\alpha}^{i}, \lambda^{i}, \beta_2^{i})}, 1\right)$

otherwise $(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}, \lambda^{i+1}, \beta_2^{i+1}) = (\Sigma^{i}, \boldsymbol{\alpha}^{i}, \lambda^{i}, \beta_2^{i})$.

Since the *candidate-generating* density function has to approximate the density from which one wants to sample, it is necessary to take for $g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)$ a density function which is close to the conditional posterior of λ. Therefore, the choice of g depends on the functional form of the prior for $(\Sigma, \mathbf{\Pi})$. The decomposition of the trace in the posterior under a conjugate prior specification given in Appendix B shows that a good choice for g is

$$g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2, Y) = (2\pi)^{-\frac{1}{2}(n-r)^2} |\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp'|^{\frac{1}{2}(n-r)} |\beta_\perp'(A + Y_{-1}'Y_{-1})\beta_\perp|^{\frac{1}{2}(n-r)}$$
$$\exp(-\frac{1}{2}\mathrm{tr}(\beta_\perp'(A + Y_{-1}'Y_{-1})\beta_\perp(\lambda - \tilde{\lambda})\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp'(\lambda - \tilde{\lambda})')), \tag{31}$$

with $\tilde{\lambda} = (\beta_\perp'(A + Y_{-1}'Y_{-1})\beta_\perp)^{-1}\beta_\perp'(A(P - \beta\boldsymbol{\alpha}) + Y_{-1}'(\Delta Y - \beta\boldsymbol{\alpha}))\Sigma^{-1}\boldsymbol{\alpha}_\perp'(\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp')^{-1}$. This results in the following expression for the weight function w

$$w(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2) = (2\pi)^{-\frac{1}{2}(n-r)^2} \frac{|J(\boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}|}{|J(\boldsymbol{\alpha}, \lambda, \beta_2)|} |\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp'|^{\frac{1}{2}(n-r)}$$
$$|\beta_\perp'(A + Y_{-1}'Y_{-1})\beta_\perp|^{\frac{1}{2}(n-r)} \exp(-\frac{1}{2}\mathrm{tr}((\beta_\perp'(A + Y_{-1}'Y_{-1})\beta_\perp)\tilde{\lambda}(\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp')\tilde{\lambda}')). \tag{32}$$

Remember that the functional form of the Jacobian $|J(\boldsymbol{\alpha}, \lambda, \beta_2)|$ is given in Appendix A. Using the same strategy as above we can also obtain drawings from the marginal priors $p_{ecc}(\boldsymbol{\alpha})$ and $p_{ecc}(\beta_2)$, see also the end of Section 5.1.

The Metropolis-Hastings sampler presented in this section may sometimes lead to high rejection frequencies and therefore slow convergence. An alternative approach is importance sampling, see Kloek and van Dijk (1978) and Geweke (1989). The weight function w in (32), evaluated in the draws, represents in that case important weights, see also Chen (1994) and (Verdinelli and Wasserman 1995, p. 615). The M-H sampling approach, however, has the advantage that it can be implemented in a Gibbs sampler, see Chib and Greenberg (1995). This makes it possible to analyse more complicated VAR models, for instance a VAR model with an endogenous break in the constant. These models are usually analysed using a Gibbs framework. The sampling of the block $(\Sigma, \boldsymbol{\alpha}, \beta_2)$ given

the remaining parameters in the model can then be done using the simulation steps in this section. In the next subsection we show how we can use importance sampling techniques to compute posterior odds to compare cointegration models with different number of cointegrating relations.

5 Posterior Odds Ratios

The in the previous sections developed procedures for calculating the posteriors of the parameters of the cointegration model for different number of cointegrating vectors r, allow us to compare models with different cointegration ranks using posterior odds analysis. Since the number of cointegrating vectors r can only take $n+1$ different discrete values, we can consider prior and posterior probabilities of the cointegration rank r and the implied number of unit roots $(n-r)$, $r = 0, \ldots, n$.

First we assign prior probabilities to every cointegration rank r,

$$\Pr[\text{rank} = r] \qquad r = 0, \ldots, n. \tag{33}$$

These prior probabilities imply prior odds ratios [PROR] to compare *a priori* the cointegration models with different number of cointegrating relations. Since every cointegration model (7) is nested in the full rank model (9) it is convenient to consider

$$\text{PROR}(r|n) = \frac{\Pr[\text{rank} = r]}{\Pr[\text{rank} = n]}, \qquad r = 0, \ldots, n. \tag{34}$$

The Bayes factor [BF] which compares the cointegration model (7) with the unrestricted error correction model (9) is given by

$$\begin{aligned} \text{BF}(r|n) &= \frac{\Pr[Y|\text{rank} = r]}{\Pr[Y|\text{rank} = n]}, \qquad r = 0, \ldots, n \\ &= \frac{\iiint p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2)\, \mathcal{L}_{ecc}(Y|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2}, \end{aligned} \tag{35}$$

see *e.g.* Leamer (1978) for a formal discussion about Bayes factors. Now we can define the posterior odds ratios [POR] to compare *a posteriori* a cointegration model with r cointegrating vectors with a model with n cointegrating vectors

$$\text{POR}(r|n) = \text{PROR}(r|n) \times \text{BF}(r|n), \qquad r = 0, \ldots, n. \tag{36}$$

These posterior odds ratios imply posterior probabilities for every cointegration rank. The posterior probability for a cointegration model with rank r equals

$$\Pr[\text{rank} = r|Y] = \frac{\text{POR}(r|n)}{\sum_{i=0}^{n} \text{POR}(i|n)}, \qquad r = 0, \ldots, n. \tag{37}$$

The posterior probabilities can be used to choose the cointegration rank, or as weights in further analyses, like forecasting exercises. In the next subsection we show how the Bayes factors can be computed in case of conjugate priors.

5.1 Computation of Bayes Factors

The computation of the Bayes factor (35) requires the evaluation of two integrals. Since it is not possible to integrate out $\boldsymbol{\alpha}$, λ or β analytically, we need again simulation techniques to calculate Bayes factors.

For the computation of the Bayes factor we use that the prior and the likelihood of the error correction cointegration model (7) equals the prior and the likelihood of the unrestricted error correction model (9) evaluated in $\lambda = 0$ times $1/c_r$. The Bayes factor (35) simplifies to

$$\begin{aligned}
\mathrm{BF}(r|n) &= \frac{\iiint p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2)\, \mathcal{L}_{ecc}(Y|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2} \\
&= \frac{\iiint \frac{1}{c_r} p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2} \\
&= \frac{\iiint \frac{1}{c_r} p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2}.
\end{aligned} \tag{38}$$

Dickey (1971) shows that under certain regularity conditions this Bayes factor can be simplified as the ratio of the marginal posterior density of λ, $p_{uec}(\lambda|Y)$, and the marginal prior density of λ, $p_{uec}(\lambda)$, both evaluated in $\lambda = \mathbf{0}$. This ratio is known as the Savage-Dickey density ratio, see also Verdinelli and Wasserman (1995). Unfortunately, the marginal prior and posterior distribution are of an unknown type and can only be obtained through simulation. The height of the marginal posterior density in $\lambda = \mathbf{0}$ can be computed using a kernel estimator, see Silverman (1986). However, this strategy suffers from the curse of dimensionality in kernel estimation since $(n-r)^2$ can become large. The same reasoning holds for the computation of c_r, which is in fact the height of the marginal prior of λ evaluated in $\lambda = \mathbf{0}$, see (21). If one wants to avoid the nonparametric estimation, one can also use the approach of Chen (1994), see also (Verdinelli and Wasserman 1995, p. 615). We extend the numerator with the integral $\int g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2)d\lambda$, where g is a proper density function

$$\begin{aligned}
\mathrm{BF}(r|n) &= \frac{\iiint \frac{1}{c_r} p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)|_{\lambda=\mathbf{0}} \left(\int g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\lambda\right) d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2} \\
&= \frac{\frac{1}{c_r} \iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)|_{\lambda=\mathbf{0}}\, g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2}.
\end{aligned} \tag{39}$$

An appropriate candidate for $g(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2)$ is a density function which is close to the conditional posterior of λ, see Chen (1994) for details. Therefore, the conditional density function (31) is again a good candidate.

We can calculate the ratio of integrals in (39) by simulating Σ, $\boldsymbol{\alpha}$, λ and β_2 from $p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)$, see Section 4. For the simulated parameters $(\Sigma^i, \boldsymbol{\alpha}^i, \lambda^i, \beta_2^i)$, $i = 1, \ldots, N$, we calculate the ratio of the integrands $w(\Sigma^i, \boldsymbol{\alpha}^i, \lambda^i, \beta_2^i)$ defined in (32). The average of the simulated weights $w(\Sigma^i, \boldsymbol{\alpha}^i, \lambda^i, \beta_2^i)$, which can be seen as importance weights,

then converges to the ratio of the integrals (39) times c_r

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} w(\Sigma^i, \boldsymbol{\alpha}^i, \lambda^i, \beta_2^i) - c_r \mathrm{BF}(r|n)\right) \Rightarrow N(0, v), \tag{40}$$

where N is the number of draws, $v = \mathrm{var}(w(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2))$, $\Rightarrow$ stands for weak convergence, and $\left(\frac{1}{N}\sum_{i=1}^{N} w(\Sigma^i, \boldsymbol{\alpha}^i, \lambda^i, \beta_2^i)^2 - (\frac{1}{N}\sum_{i=1}^{N} w(\Sigma^i, \boldsymbol{\alpha}^i, \lambda^i, \beta_2^i))^2\right) \Rightarrow v$, see Geweke (1989) and Chen (1994).

To compute the Bayes factors we also need the value of c_r. We can use the same simulation technique of Chen (1994) to calculate c_r. Since $\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda d\beta_2 = 1$ we can write

$$\begin{aligned}
c_r &= \iiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2 \\
&= \frac{\iiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda d\beta_2} \\
&= \frac{\iiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}} \left(\int h(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\lambda\right) d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda d\beta_2} \\
&= \frac{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, h(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda d\beta_2},
\end{aligned} \tag{41}$$

where $h(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2)$ is a proper conditional density function. Simulate from the prior $p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)$ (18) and compute the ratio of the integrands of the numerator and denominator. The average of these simulated ratios then converges to c_r. An appropriate density function h for the prior specification (17) is a density function which is close to the conditional prior of λ

$$\begin{aligned}
h(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2) = (2\pi)^{-\frac{1}{2}(n-r)^2} &|\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp'|^{\frac{1}{2}(n-r)} |\beta_\perp' A \beta_\perp|^{\frac{1}{2}(n-r)} \\
&\exp(-\frac{1}{2}\mathrm{tr}(\beta_\perp' A \beta_\perp (\lambda - l)\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp' (\lambda - l)')),
\end{aligned} \tag{42}$$

with $l = (\beta_\perp' A \beta_\perp)^{-1} \beta_\perp' A (P - \beta\boldsymbol{\alpha}) \Sigma^{-1} \boldsymbol{\alpha}_\perp' (\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp')^{-1}$. The weight function belonging to this density equals

$$\frac{|J(\boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}|}{|J(\boldsymbol{\alpha}, \lambda, \beta_2)|}\, h(\lambda|\Sigma, \boldsymbol{\alpha}, \beta_2)|_{\lambda=\mathbf{0}}. \tag{43}$$

In case of a diffuse prior specification c_r is not defined and Bayes factors for rank reduction cannot be interpreted. Under diffuse prior specifications the height of the marginal prior, in our case c_r, is often replaced by a penalty function depending on the number of restricting parameters. For instance, if we put c_r equal to $T^{-\frac{1}{2}(n-r)^2}$ with T the number of observations we have the Bayesian information criterion [BIC] of Schwartz (1978), see also Kass and Raferty (1995). We can also use a Bayesian generalisation of the posterior information criterion [PIC] of Phillips and Ploberger (1994, 1996). In that case the penalty function equals $c_r = (2\pi)^{-\frac{1}{2}(n-r)^2}$. We note that the resulting Bayes factor is not equal to the PIC for cointegration models constructed in Phillips (1996).

6 Application

To illustrate the applicability of the, in the previous sections, constructed methods and procedures for Bayesian cointegration analyses, we consider four simulated time series, the UK data analysed in Hendry and Doornik (1994) and the Danish data analysed in Johansen and Juselius (1990). Since we want to illustrate the performance of our Bayesian analysis without the risk of prior dominance, we choose in this section for an approach with proper conjugate priors, which do not contain much information. More informative priors can easily be incorporated in the analysis.

6.1 Simulated Series

We consider the following four data generating processes [DGPs],

$$
\begin{aligned}
\text{I} \quad &: \quad \Delta Y_t = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \varepsilon_t \\
\text{II} \quad &: \quad \Delta Y_t = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 \\ 0.2 \\ 0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix} Y_{t-1} + \varepsilon_t \\
\text{III} \quad &: \quad \Delta Y_t = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.2 \\ 0.2 & -0.2 \\ 0.2 & 0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} Y_{t-1} + \varepsilon_t \\
\text{IV} \quad &: \quad \Delta Y_t = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.2 & -0.2 \\ 0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & -0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} Y_{t-1} + \varepsilon_t,
\end{aligned}
\tag{44}
$$

where $\varepsilon_t \sim \text{NID}(\mathbf{0}, \mathbf{I}_3)$ and the sample size T is 100 observations. The four DGPs contain 0, 1, 2 and 3 cointegrating relations, respectively. DGP I contains three unit roots, DGP II contains 2 unit roots and a root 0.6, DGP III contains the roots 1, 0.6 and 0.6, and DGP IV contains the roots 0.8, 0.6 and 0.6.

To analyse the simulated series, we consider a VAR(1) model with a constant term, which corresponds to the specification in the DGP. The first step in the Bayesian analysis is to specify a prior on the vector autoregressive parameters $\mathbf{\Pi}$ and on the covariance matrix Σ. The prior specification is given in (17). For the Σ prior we take $h = 4$ and $S = \mathbf{I}_3$. For the mean of the prior for $\mathbf{\Pi}$ given Σ we take $P = \mathbf{0}$, favouring the hypothesis of three unit roots. Note that the variance A of this prior distribution also reflects our prior beliefs about the number of cointegrating relations, since it corresponds to the term $(Y'_{-1}Y_{-1})$ in the likelihood, see (23). Note that the series $\beta' Y_{t-1}$ is stationary, while $\beta'_\perp Y_{t-1}$ is not. If we want to be uninformative about the rank of $\mathbf{\Pi}$ it is difficult to propose a prior value for A. To circumvent this problem we take a g-prior for $\mathbf{\Pi}$, see Zellner (1986). In our case this would imply that the variance of this prior is a fraction of the matrix $(Y'_{-1}Y_{-1})$. Since we are dealing with non-stationary time series, we divide this matrix by

Table 1: Bayes factors and posterior probabilities and LR statistics for the four DGPs.

	$\theta = 1$		$\theta = 0.1$		$\theta = 0.01$		diffuse		classical	
r	ln(BF)	Pr$[r\|Y]$	ln(BF)	Pr$[r\|Y]$	ln(BF)	Pr$[r\|Y]$	ln(PIC)	Pr$[r\|Y]$	LR	p-val.
					DGP I					
0	10.27	1.00	20.53	1.00	30.89	1.00	25.06	1.00	10.97	0.96
1	4.36	0.00	8.92	0.00	13.52	0.00	12.50	0.00	5.10	0.80
2	1.18	0.00	2.32	0.00	3.48	0.00	3.58	0.00	1.38	0.24
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
					DGP II					
0	−7.91	0.00	2.16	0.00	12.50	0.47	6.82	0.01	44.98	0.00
1	3.43	0.88	7.99	1.00	12.59	0.53	12.02	0.99	6.39	0.64
2	1.15	0.09	2.29	0.00	3.44	0.00	3.78	0.00	1.09	0.29
3	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00		
					DGP III					
0	−32.81	0.00	−23.05	0.00	−12.75	0.00	−23.88	0.00	94.86	0.00
1	−9.83	0.00	−5.42	0.00	−0.84	0.03	−4.13	0.00	33.05	0.00
2	0.13	0.53	1.26	0.78	2.41	0.89	2.01	0.88	2.95	0.09
3	0.00	0.47	0.00	0.22	0.00	0.08	0.00	0.12		
					DGP IV					
0	−31.35	0.00	−21.54	0.00	−11.23	0.00	−23.96	0.00	90.93	0.00
1	−13.42	0.00	−9.03	0.00	−4.45	0.01	−8.62	0.00	39.68	0.00
2	−4.35	0.01	−3.26	0.04	−2.12	0.11	−2.70	0.05	12.32	0.00
3	0.00	0.99	0.00	0.96	0.00	0.88	0.00	0.95		

[1] A log Bayes factor ln(BF)> 0 denotes that a cointegration model with r cointegration relations is more likely than a model with 3 cointegration relations.

[2] Posterior probability of the cointegration rank Pr$[r|Y]$ (37) is based on equal prior probabilities (33) for every rank r.

[3] The prior is given in (17) with $h = 4$, $S = \mathbf{I}_3$, $P = \mathbf{0}$ and $A = \theta(Y'_{-1}Y_{-1})/T$.

[4] LR denotes the Johansen likelihood ratio trace statistic for cointegration (p-values based on classical distribution theory in the final column), see Johansen (1991).

T, where T equals the number of observations. The prior variance for $\mathbf{\Pi}$ is now given by $A = \theta(Y'_{-1}Y_{-1})/T$. A smaller value of θ implies less prior information. Finally, we give each cointegration rank the same prior probability $\Pr[\text{rank} = r] = \frac{1}{4}$, $r = 0, \ldots, 3$, see (33).

Given the priors and prior probabilities, we can compare models with reduced rank (cointegration models) with the full rank unrestricted error correction model. Table 1 shows the Bayes factors and posterior probabilities for the four DGPs for three values of θ, $\theta = 1, 0.1, 0.001$. A Bayes factor exceeding one (or $\ln(\text{BF}(r|3))$ exceeding zero) indicates that rank r is preferred above the full rank situation. For instance, for DGP I every rank reduction is preferred, while for DGP IV the full rank situation is always preferred. The Bayes factors can be translated into posterior probabilities for the cointegration ranks, see (37). In all cases the posterior probabilities of the correct cointegration rank exceeds the posterior probabilities of the other ranks. Note that unit roots become more likely if we increase the prior variance on $\mathbf{\Pi}$ by decreasing θ. This results since our prior is centered at $\mathbf{\Pi} = 0$. When we increase the prior variance, the prior height in $\mathbf{\Pi} = 0$ decreases. The posterior height in $\mathbf{\Pi} = 0$ remains almost the same since the value of θ is such that the prior only minorly affects the posterior. From Section 5.1 we have seen that the Bayes factor for $\mathbf{\Pi} = 0$ equals the ratio of the posterior and prior height in $\mathbf{\Pi} = 0$. Hence the Bayes factor increases when we increase the prior variance. The same reasoning holds for the Bayes factors for $\lambda = \mathbf{0}$. This shows that one needs to be careful in specifying prior variances and avoid priors with a too large prior variance. The columns labelled 'diffuse' show the Bayes factors and posterior probabilities in case of a diffuse prior specification. These Bayes factors are constructed using the $c_r = (2\pi)^{-\frac{1}{2}(n-r)^2}$ penalty function, see the end of Section 5.1 for a discussion. We see that the PIC based Bayes factors give quite satisfactory results and we therefore prefer them above a proper informative prior with a very large prior variance.

The final two columns of Table 1 show the classical likelihood ratio [LR] statistics for rank reduction of Johansen (1991) with p-values based on classical asymptotic theory. The trace statistics also indicate the right cointegration rank.

6.2 Small Monetary Model for the UK

Hendry and Doornik (1994) construct a small linear dynamic monetary model for the United Kingdom. The model consists of the variables nominal M_1, denoted by m_t, total final expenditure y_t, the total final expenditure deflator p_t, and the differential between the three-month local authority interest rate and the M_1 retail sight-deposit interest rate denoted by r_t. The latter represents the opportunity cost of holding M_1. All variables are in logs except for the interest rate r_t.

In this section we analyse the same UK data as in Hendry and Doornik (1994). We have the same quarterly observed series of m_t, y_t, p_t and r_t for the period 1963.I–1989.II. The data are seasonally adjusted. The first step in the modelling strategy is to specify an unrestricted VAR model. Hendry and Doornik (1994) propose a VAR(2) model for the

four-dimensional vector of time series $Y_t = (m_t - p_t, y_t, \Delta p_t, r_t)'$

$$\Delta Y_t = \mu + \tau t + \Pi Y_{t-1} + \bar{\Phi}_1 \Delta Y_{t-1} + \xi_1 \mathrm{DOIL}_t + \xi_2 \mathrm{DOUT}_t + \varepsilon_t \tag{45}$$

where $\varepsilon_t \sim \mathrm{NID}(\mathbf{0}, \Sigma)$, ξ_1 and ξ_2 are (4×1) parameter vectors and DOIL_t and DOUT_t are dummy variables to capture outlying observations caused by the Heath-Barber boom and the first effects of the Thatcher government, and the two oil crises respectively, see Hendry and Doornik (1994) for details.[2]

The trend t and the dummy variable DOUT_t are restricted such that they lie in the cointegration space, *i.e.* $\alpha_\perp' \tau = \mathbf{0}$ and $\alpha_\perp' \xi_2 = \mathbf{0}$, which means that the vector $(t \ \mathrm{DOUT}_t)'$ is added to the Y_{t-1} vector and that Π becomes a (6×4) matrix. The long run multiplier is now no longer a square matrix as in the decomposition in (10). However, the decomposition can be directly extended to allow for the nonsquare long run multiplier by adjusting the sizes of the involved matrices in the appropriate manner. In general when $\mathbf{\Pi}$ is a $(m \times n)$ matrix with $m > n$, the singular value decomposition (10) is such that U is an $(m \times m)$ matrix with $U'U = \mathbf{I}_m$, V a $(n \times n)$ matrix with $V'V = \mathbf{I}_n$ and S a $(m \times n)$ matrix consisting of a $(n \times n)$ diagonal matrix with the n singular values on its main diagonal on top of a $((m-n) \times n)$ matrix of zeros, see Golub and van Loan (1989). We can write U, S and V as in (11) but now with U_{11} a $(r \times r)$ matrix, U_{12} a $(r \times (m-r))$, U_{21} a $((m-r) \times r)$, U_{22} a $((m-r) \times (m-r))$ and S_{22} a $((m-r) \times (n-r))$ matrix. The values for $\boldsymbol{\alpha}$, β_2 and λ follow now directly from (12). Furthermore, the specification of conditional densities, derivation of the Jacobian, *etc.* remain unaltered, see Kleibergen and van Dijk (1998).

The first part of Table 2 reports the results of a Bayesian cointegration analysis for the model (45). The first row displays the results for a model without the dummy variables DOIL_t and DOUT_t. The results are based on the conjugate prior specification (17) with $h = 5$, $S = 0.0001 \times \mathbf{I}_4$, $P = \mathbf{0}$ and $A = \theta(Y_{-1}'Y_{-1})/T$ for different values of θ. For the remaining parameters we take flat priors. Again, we assume equal prior probabilities $\Pr[\mathrm{rank} = r] = \frac{1}{5}$, $r = 0, \ldots, 4$. For $\theta = 1$ we see that the posterior probabilities indicate that two cointegration relations are most likely. For the other two values of θ only one cointegration relation is more likely. The posterior probabilities resulting from these Bayes factors also indicate one cointegration relation. The final two columns shows that the Johansen trace statistics indicate one cointegration if we test at a 10% level of significance.

The results change if we include the dummy variables DOUT_t and DOIL_t like in Hendry and Doornik (1994), see second row of Table 2. The posterior probabilities for $\theta = 1$ and $\theta = 0.1$ now indicate two cointegration relations between the series in Y_t, while for $\theta = 0.01$ we have only one cointegration relation. The PIC based Bayes factors also indicate two cointegration relations. Hence, this specification provides more posterior evidence for two cointegration relations. The Bayesian results correspond with the classical results in Hendry and Doornik (1994), who also find two cointegration relations using

[2] DOUT_t is zero except for unity in 1972.IV, 1973.I and 1979.II and DOIL_t is zero except in 1973.III, 1973.IV and 1979.III.

Table 2: Bayes factors, posterior probabilities and LR statistics for the UK and Danish data.

	$\theta = 1$		$\theta = 0.1$		$\theta = 0.01$		diffuse		classical	
r	ln(BF)	Pr$[r\|Y]$	ln(BF)	Pr$[r\|Y]$	ln(BF)	Pr$[r\|Y]$	ln(PIC)	Pr$[r\|Y]$	LR	p-val.
					UK data					
					no dummies and restricted trend					
0	−35.68	0.00	−13.49	0.00	9.44	0.00	24.58	0.03	119.38	0.00
1	3.00	0.00	21.05	0.73	39.43	1.00	28.08	0.94	40.89	0.08
2	9.83	0.98	20.08	0.27	30.43	0.00	24.83	0.03	12.12	0.80
3	5.88	0.02	10.44	0.00	15.04	0.00	13.95	0.00	4.48	0.68
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
					dummies and restricted trend					
0	−48.10	0.00	−21.56	0.00	5.95	0.00	25.71	0.15	152.85	0.00
1	7.28	0.00	35.59	0.02	64.10	1.00	13.78	0.00	71.42	0.00
2	21.10	1.00	39.31	0.98	57.71	0.00	27.43	0.85	19.65	0.24
3	15.58	0.00	25.87	0.00	36.22	0.00	19.61	0.00	6.43	0.40
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
					Danish data					
					unrestricted constant					
0	16.40	0.89	34.29	1.00	52.66	1.00	20.65	0.21	45.67	0.08
1	14.34	0.11	24.51	0.00	34.85	0.00	21.97	0.79	17.07	0.63
2	7.26	0.00	11.77	0.00	16.37	0.00	13.35	0.00	6.71	0.61
3	2.39	0.00	3.53	0.00	4.68	0.00	4.81	0.00	0.38	0.54
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
					restricted constant					
0	22.65	0.01	45.07	0.33	68.03	0.98	44.36	1.00	49.14	0.11
1	27.63	0.99	45.78	0.67	64.18	0.02	25.83	0.00	19.06	0.79
2	16.40	0.00	26.62	0.00	36.96	0.00	18.42	0.00	8.69	0.77
3	7.84	0.00	12.40	0.00	17.03	0.00	10.10	0.00	2.35	0.70
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		

[1] A log Bayes factor ln(BF)>0 denotes that a cointegration model with r cointegration relations is more likely than a model with 4 cointegration relations.

[2] Posterior probability of the cointegration rank Pr$[r|Y]$ (37) is based on equal prior probabilities (33) for every rank r.

[3] The prior is given in (17) with $h = 5$, $P = \mathbf{0}$, $A = \theta(Y'_{-1}Y_{-1})/T$ with $S = 0.0001 \times \mathbf{I_4}$ for the UK series and $S = 0.001 \times \mathbf{I_4}$ for the Danish series.

[4] LR denotes the Johansen likelihood ratio trace statistic for cointegration (p-values based on classical distribution theory in the final column), see Johansen (1991).

the two dummy variables, see the final two columns of Table 2. Note that the favour of the PIC based Bayes factor for the absence of cointegration above one cointegration relation results from the restricted trend in the cointegration space. The model without cointegration is therefore more flexible than the model with one cointegration relation. Because none of these models are good models in terms of Bayes factors, the most flexible one is preferred.

6.3 Danish Money Demand

Johansen and Juselius (1990) analyse the demand function for money for the Danish economy using a VAR model. Their model consist of M_2 denoted by m_t, real income y_t, price level p_t and the costs of holding money. The costs of holding money is approximated by a difference between the bank deposit rate i_t^d for interest bearing deposits and the bond rate i_t^b. All variables are in logs. Since the inflation rate Δp_t does not alter the cointegration analysis significantly, this variable is not considered in the Johansen and Juselius study.

In this subsection we analyse the same Danish data as in Johansen and Juselius (1990). We have quarterly observed series of m_t, i_t^d, i_t^b and y_t for the period 1974.1–1987.3. The cointegration analysis is performed in the following VAR(2) model,

$$\begin{pmatrix} \Delta m_t \\ \Delta y_t \\ \Delta i_t^b \\ \Delta i_t^d \end{pmatrix} = \mu + \sum_{s=1}^{3} \bar{\delta}_s \bar{D}_{s,t} + \Pi \begin{pmatrix} m_{t-1} \\ y_{t-1} \\ i_{t-1}^b \\ i_{t-1}^d \end{pmatrix} + \bar{\Phi}_1 \begin{pmatrix} \Delta m_{t-1} \\ \Delta y_{t-1} \\ \Delta i_{t-1}^b \\ \Delta i_{t-1}^d \end{pmatrix} + \varepsilon_t, \tag{46}$$

where $\bar{D}_{s,t}$ represents seasonal dummies with zero mean and $\bar{\delta}_s$ is a four-dimensional parameter vector, $s = 1, \ldots, 3$. Although it is not likely that real income does not have a linear trend, Johansen and Juselius restrict the constant in the cointegrating space, *i.e.* $\alpha_\perp' \mu = \mathbf{0}$, see *e.g.* Johansen and Juselius (1990). Hence, the Π matrix is extended with an extra row and the Y_{-1} matrix with an extra column, see also above.

The second part of Table 2 displays the results of a Bayesian cointegration analysis for the Danish data. The results are based on a conjugate prior specification for Σ and $\mathbf{\Pi}$ (17) with $h = 5$, $S = 0.001 \times \mathbf{I}_4$, $P = \mathbf{0}$ and $A = \theta(Y_{-1}' Y_{-1})/T$ for different values of θ. For the remaining parameters we take a flat prior. We assume again equal probabilities (33) $\Pr[\text{rank} = r] = \frac{1}{5}$, $r = 0, \ldots, n$.

First, we consider a model where the constant is not restricted in the cointegrating space ($\alpha_\perp' \mu \neq \mathbf{0}$). The third row of Table 2 shows the Bayes factors and posterior probabilities for three values of θ. The Bayes factors favour every rank reduction over the full rank model, resulting in no posterior indication for cointegration. Hence the model with $\mathbf{\Pi} = \mathbf{0}$ is more likely than any cointegration specification. The PIC based Bayes factors indicate that a model with one cointegration relation is preferred.

In case we restrict the constant in the cointegrating space, we see that the Bayes factors again favour every rank reduction over a full rank model, see the final row of Table 2.

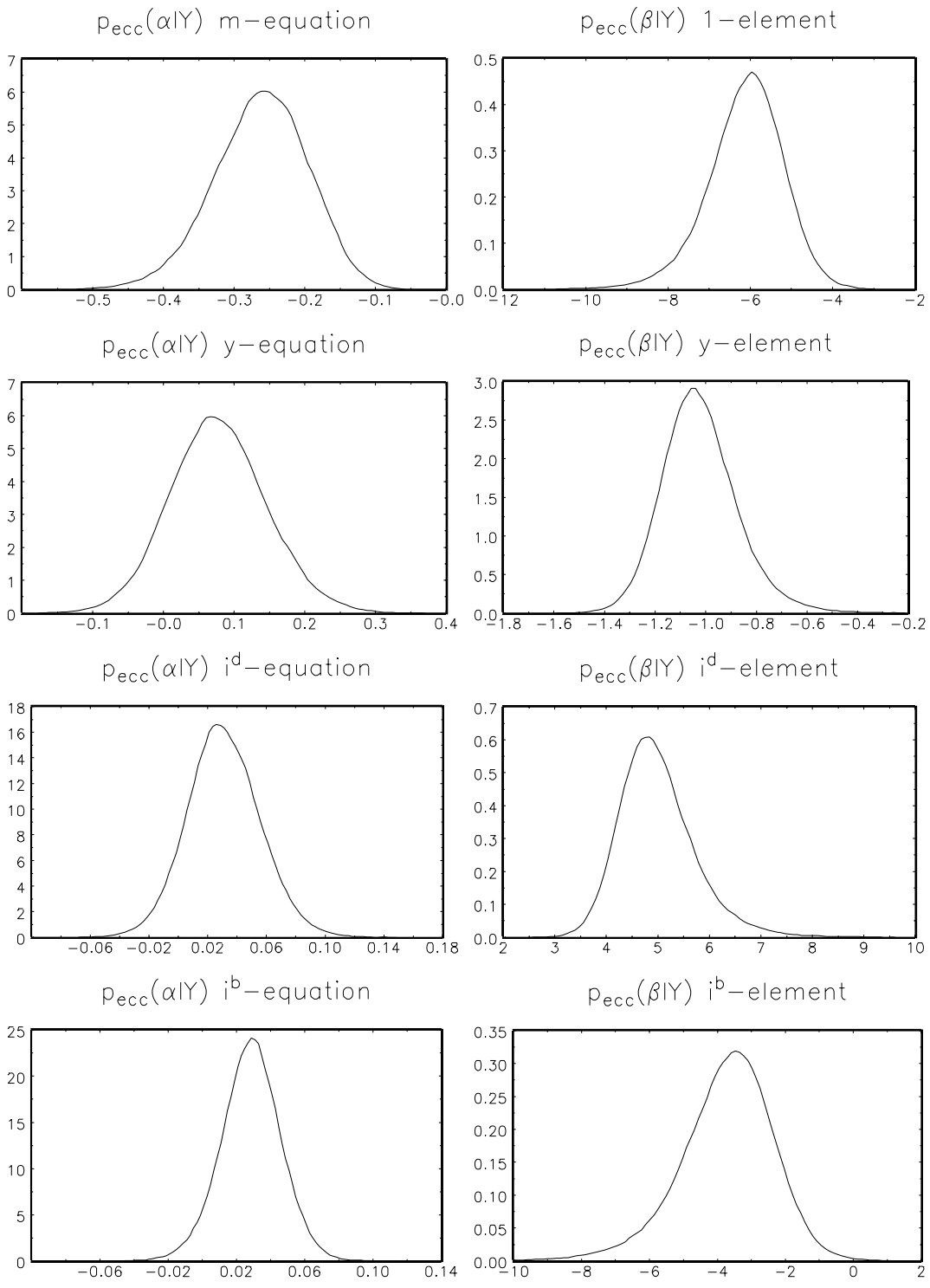

Figure 1: Marginal posterior densities of the adjustment parameters α and the cointegrating vector β for the UK series.

However, for $\theta = 1$ and $\theta = 0.1$ the posterior probabilities assign most probability to a model with rank one. For $\theta = 0.01$ there is again no evidence of cointegration. Note again that similar to the UK data, we find less cointegration when we decrease θ. As explained before this results from the decrease of the prior height in the point where the tested hypothesis lies. PIC based Bayes factors indicate no cointegration. The final two columns show again the LR test statistics. The classical results show that we need to test at a 11% level of significance to find cointegration.

To compare Bayesian posterior results with classical maximum likelihood estimates we compute posterior results for the cointegration model with one cointegrating relationship and the constant restricted in the cointegrating space ($\alpha'_\perp \mu = \mathbf{0}$). We use a diffuse prior specification to compare the posterior outcomes directly with the maximum likelihood results of Johansen and Juselius (1990). The posterior means of the α and β parameters are

$$\begin{array}{cccccc} & m_t & y_t & i_t^b & i_t^d & \\ \alpha' = (& -0.26 & 0.08 & 0.03 & 0.03 &) \\ & (0.07) & (0.07) & (0.02) & (0.02) & \end{array} \text{ and } \begin{array}{cccccc} & m_t & y_t & i_t^b & i_t^d & 1 & \\ \beta = (& 1 & -1.02 & 4.99 & -3.76 & -6.23 &), \\ & (-) & (0.15) & (0.76) & (1.42) & (0.95) & \end{array}$$

where the posterior standard deviations are between parentheses. Note that the posterior means correspond reasonably well to the maximum likelihood estimates in Table 2 of Johansen and Juselius (1990) and that the posterior standard deviations of the elements of the cointegrating vector are essentially infinite since the marginal posterior has Cauchy type tails, see Kleibergen (1998) and Kleibergen and van Dijk (1998). However, since the marginal posterior of the first element of α has almost no probability mass in zero, the posterior standard deviations do not show their infinite value. Figure 1 shows the marginal posterior densities of the adjustment parameters α and the cointegrating vector β. The first column shows the marginal posteriors of the α parameters. The marginal posterior of the adjustment parameter for the money equation is situated far away from zero. This is not the case for the other adjustment parameters, where zero lies within the 95% one-sided highest posterior density regions of the marginal posteriors. The second column of Figure 1 shows the marginal posteriors of the β parameters. These marginal posterior distributions are more skewed and have fatter tails which are even of the Cauchy type, see Kleibergen (1998). The posterior masses of the marginal posteriors of the cointegration parameters are situated far away from zero except for the i^b-element.

In summary, although the examples in this section are simple, they show that Bayesian techniques provide useful tools to analyse cointegration. The Bayes factors indicate whether rank reduction is plausible. Bayes factors can be used to calculate posterior probabilities for each cointegration rank, to show the best model. If there is no clear preference for one of the cointegration ranks, it is also possible to use the posterior probabilities as weights in a forecasting exercise.

7 Concluding Remarks

In this paper we have proposed a new Bayesian approach for cointegration analysis. This approach is based on the idea that the cointegration model is nested in a linear error correction model where the parameter modelling the error correction, *i.e.* the long run multiplier, has full rank. Cointegration then occurs when the long run multiplier exhibits rank reduction. We therefore explicitly model the priors and posteriors of the parameters of the cointegration model as proportional to the priors and posteriors of the long run multiplier given that it has reduced rank. This is achieved by specifying the long run multiplier as the sum of two matrices. The first matrix is the product of the adjustment parameters and the cointegrating vectors, *i.e.* the cointegration specification. The second matrix models the deviation from the cointegration specification and we can therefore restrict it to zero to obtain cointegration.

The Bayesian analysis starts with defining a prior for the full rank parameter matrix modelling error correction in the linear error correction model. This prior leads via a Jacobian transformation to the joint prior for the adjustments parameters, the cointegration vectors and the matrix modelling the deviation from the cointegration specification. The prior and posterior of the parameters of the cointegration model are obtained by putting the matrix modelling the deviation in the prior and posterior of the unrestricted model equal to zero. To obtain marginal posterior results we propose a Metropolis-Hastings simulation algorithm. Bayes factors to determine the cointegration rank are obtained using the Savage-Dickey density ratio.

The Bayesian cointegration analysis proposed in this paper is flexible and is therefore also applicable in more complicated models. We can for instance allow for structural breaks in the means and deterministic trend or we can consider non-linear cointegration models, like Markov Switching cointegration and threshold cointegration. Also we may change the assumption of normal distributed errors and extend the analysis to t-distributed errors or vector moving average errors. All of these issues are subjects of our future research.

A Jacobian Transformation

For the derivation of the Jacobian transformation, it is convenient to split up the transformation from $\mathbf{\Pi}$ to $\boldsymbol{\alpha}$, λ and β_2 in two steps, firstly from $\mathbf{\Pi}$ to $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda, \beta_2)$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1 \ \boldsymbol{\alpha}_2)$ and $\vartheta_2 = -\boldsymbol{\alpha}_1^{-1}\boldsymbol{\alpha}_2$, and secondly from $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda, \beta_2)$ to $(\boldsymbol{\alpha}, \lambda, \beta_2)$. In the following we construct the Jacobians for the two transformations. We can denote $\mathbf{\Pi}$ as a function of $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda, \beta_2)$

$$\mathbf{\Pi} = \begin{pmatrix} \beta & \beta_\perp \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 & \mathbf{0} \\ \mathbf{0} & \lambda \end{pmatrix} \begin{pmatrix} \vartheta \\ \vartheta_\perp \end{pmatrix}$$

where $\vartheta = (\mathbf{I}_r \ -\vartheta_2)$ with $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1\vartheta$, $\beta = (\mathbf{I}_r \ -\beta_2')'$, $\vartheta_\perp = (\mathbf{I}_{n-r} + \vartheta_2'\vartheta_2)^{-\frac{1}{2}}(\vartheta_2' \ \mathbf{I}_{n-r})$ and $\beta_\perp = (\beta_2 \ \mathbf{I}_{n-r})'(\mathbf{I}_{n-r} + \beta_2\beta_2')^{-\frac{1}{2}}$ so that $\vartheta_\perp\vartheta_\perp' = \mathbf{I}_{n-r}$ and $\beta_\perp'\beta_\perp = \mathbf{I}_{n-r}$.[3] The derivatives of $\mathbf{\Pi}$ with respect to $\boldsymbol{\alpha}_1$, ϑ_2, λ and β_2 read

$$\begin{aligned}
J_1 &= \frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1))'} = (\vartheta' \otimes \beta) \\
J_2 &= \frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\vartheta_2))'} = -\Big(\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix} \otimes \beta\boldsymbol{\alpha}_1\Big) + (\mathbf{I}_n \otimes \beta_\perp\lambda)\frac{\partial \operatorname{vec}(\vartheta_\perp)}{\partial(\operatorname{vec}(\vartheta_2))'} \\
J_3 &= \frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\lambda))'} = (\vartheta_\perp' \otimes \beta_\perp) \\
J_4 &= \frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\beta_2))'} = -\Big(\vartheta'\boldsymbol{\alpha}_1' \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}\Big) + (\vartheta_\perp'\lambda' \otimes \mathbf{I}_n)\frac{\partial \operatorname{vec}(\beta_\perp)}{\partial(\operatorname{vec}(\beta_2))'},
\end{aligned}$$

with

$$\begin{aligned}
\frac{\partial \operatorname{vec}(\vartheta_\perp)}{\partial(\operatorname{vec}(\vartheta_2))'} &= (\mathbf{I}_n \otimes \vartheta_n^{-\frac{1}{2}})\frac{\partial \operatorname{vec}(\vartheta_2' \ \mathbf{I}_{n-r})}{\partial(\operatorname{vec}(\vartheta_2))'} \\
&\quad + ((\vartheta_2' \ \mathbf{I}_{n-r})' \otimes \mathbf{I}_{n-r})\frac{\partial \operatorname{vec}(\vartheta_n^{-\frac{1}{2}})}{\partial\left(\operatorname{vec}(\vartheta_n^{\frac{1}{2}})\right)'}\frac{\partial \operatorname{vec}(\vartheta_n^{\frac{1}{2}})}{\partial(\operatorname{vec}(\vartheta_n))'}\frac{\partial \operatorname{vec}(\vartheta_n)}{\partial(\operatorname{vec}(\vartheta_2))'} \\
\frac{\partial \operatorname{vec}(\beta_\perp)}{\partial(\operatorname{vec}(\beta_2))'} &= (\beta_n^{-\frac{1}{2}\prime} \otimes \mathbf{I}_n)\frac{\partial \operatorname{vec}((\beta_2 \ \mathbf{I}_{n-r})')}{\partial(\operatorname{vec}(\beta_2))'} \\
&\quad + (\mathbf{I}_{n-r} \otimes (\beta_2 \ \mathbf{I}_{n-r})')\frac{\partial \operatorname{vec}(\beta_n^{-\frac{1}{2}})}{\partial\left(\operatorname{vec}(\beta_n^{\frac{1}{2}})\right)'}\frac{\partial \operatorname{vec}(\beta_n^{\frac{1}{2}})}{\partial(\operatorname{vec}(\beta_n))'}\frac{\partial \operatorname{vec}(\beta_n)}{\partial(\operatorname{vec}(\beta_2))'}
\end{aligned}$$

[3] If M is a positive definite real symmetric matrix, then $M^{\frac{1}{2}} = C\Lambda^{\frac{1}{2}}C'$ where Λ is a diagonal matrix containing the eigenvalues of M and C contains the orthonormal eigenvectors of M and $M^{-\frac{1}{2}} = C\Lambda^{-\frac{1}{2}}C'$, see *e.g.* (Johansen 1995, p. 222).

where we define for notational convenience $\vartheta_n = (\mathbf{I}_{n-r} + \vartheta_2'\vartheta_2)$ and $\beta_n = (\mathbf{I}_{n-r} + \beta_2\beta_2')$ so that $\vartheta_\perp = \vartheta_n^{-\frac{1}{2}}(\vartheta_2' \ \ \mathbf{I}_{n-r})$, $\beta_\perp = (\beta_2 \ \ \mathbf{I}_{n-r})'\beta_n^{-\frac{1}{2}}$ with $\vartheta_n^{\frac{1}{2}}\vartheta_n^{\frac{1}{2}} = \vartheta_n$, $\beta_n^{\frac{1}{2}}\beta_n^{\frac{1}{2}} = \beta_n$,

$$\begin{aligned}
\frac{\partial \operatorname{vec}(\vartheta_2' \ \ \mathbf{I}_{n-r})}{\partial(\operatorname{vec}(\vartheta_2))'} &= \left(\begin{pmatrix} \mathbf{I}_r \\ \mathbf{0} \end{pmatrix} \otimes \mathbf{I}_{n-r}\right)\mathbb{K}_{n-r,r} \\
\frac{\partial \operatorname{vec}(\vartheta_n^{-\frac{1}{2}})}{\partial\left(\operatorname{vec}(\vartheta_n^{\frac{1}{2}})\right)'} &= -(\vartheta_n^{-\frac{1}{2}\prime} \otimes \vartheta_n^{-\frac{1}{2}}) \\
\frac{\partial \operatorname{vec}(\vartheta_n^{\frac{1}{2}})}{\partial(\operatorname{vec}(\vartheta_n))'} &= ((\vartheta_n^{\frac{1}{2}\prime} \otimes \mathbf{I}_{n-r}) + (\mathbf{I}_{n-r} \otimes \vartheta_n^{\frac{1}{2}}))^{-1} \\
\frac{\partial \operatorname{vec}(\vartheta_n)}{\partial(\operatorname{vec}(\vartheta_2))'} &= (\mathbf{I}_{n-r} \otimes \vartheta_2') + (\vartheta_2' \otimes \mathbf{I}_{n-r})\mathbb{K}_{n-r,r}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \operatorname{vec}((\beta_2 \ \ \mathbf{I}_{n-r})')}{\partial(\operatorname{vec}(\beta_2))'} &= \left(\mathbf{I}_{n-r} \otimes \begin{pmatrix} \mathbf{I}_r \\ \mathbf{0} \end{pmatrix}\right)\mathbb{K}_{n-r,r} \\
\frac{\partial \operatorname{vec}(\beta_n^{-\frac{1}{2}})}{\partial\left(\operatorname{vec}(\beta_n^{\frac{1}{2}})\right)'} &= -(\beta_n^{-\frac{1}{2}\prime} \otimes \beta_n^{-\frac{1}{2}}) \\
\frac{\partial \operatorname{vec}(\beta_n^{\frac{1}{2}})}{\partial(\operatorname{vec}(\beta_n))'} &= ((\beta_n^{\frac{1}{2}\prime} \otimes \mathbf{I}_{n-r}) + (\mathbf{I}_{n-r} \otimes \beta_n^{\frac{1}{2}}))^{-1} \\
\frac{\partial \operatorname{vec}(\beta_n)}{\partial(\operatorname{vec}(\beta_2))'} &= (\beta_2 \otimes \mathbf{I}_{n-r}) + (\mathbf{I}_{n-r} \otimes \beta_2)\mathbb{K}_{n-r,r}
\end{aligned}$$

where $\mathbb{K}_{i,j}$ are so-called commutation matrices. For any $(i \times j)$ matrix W, $\operatorname{vec}(W) = \mathbb{K}_{i,j}\operatorname{vec}(W')$, $\operatorname{vec}(W') = \mathbb{K}_{j,i}\operatorname{vec}(W)$, and $\mathbb{K}_{i,j}' = \mathbb{K}_{j,i}$, see (Lütkepohl 1993, p. 466). The Jacobian of the transformation from $\mathbf{\Pi}$ to $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda, \beta_2)$ becomes

$$\left|\frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1)' \ \operatorname{vec}(\vartheta_2)' \ \operatorname{vec}(\lambda)' \ \operatorname{vec}(\beta_2)')}\right| = |(J_1 \ J_2 \ J_3 \ J_4)|.$$

Since $\vartheta_2 = -\boldsymbol{\alpha}_1^{-1}\boldsymbol{\alpha}_2$ the derivatives of $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda, \beta_2)$ with respect to $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, λ and β_2 are respectively

$$G_1 = \frac{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1)' \ \operatorname{vec}(\vartheta_2)' \ \operatorname{vec}(\lambda)' \ \operatorname{vec}(\beta_2)')'}{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1))'} = \begin{pmatrix} \mathbf{I}_r \otimes \mathbf{I}_r \\ (\boldsymbol{\alpha}_1^{-1}\boldsymbol{\alpha}_2)' \otimes \boldsymbol{\alpha}_1^{-1} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

$$G_2 = \frac{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1)' \ \operatorname{vec}(\vartheta_2)' \ \operatorname{vec}(\lambda)' \ \operatorname{vec}(\beta_2)')'}{\partial(\operatorname{vec}(\boldsymbol{\alpha}_2))'} = \begin{pmatrix} \mathbf{0} \\ -\mathbf{I}_{n-r} \otimes \boldsymbol{\alpha}_1^{-1} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

$$G_3 = \frac{\partial(\mathrm{vec}(\boldsymbol{\alpha}_1)'\ \mathrm{vec}(\vartheta_2)'\ \mathrm{vec}(\lambda)'\ \mathrm{vec}(\beta_2)')'}{\partial(\mathrm{vec}(\lambda))'} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{I}_{n-r} \otimes \mathbf{I}_{n-r} \\ \mathbf{0} \end{pmatrix}$$

$$G_4 = \frac{\partial(\mathrm{vec}(\boldsymbol{\alpha}_1)'\ \mathrm{vec}(\vartheta_2)'\ \mathrm{vec}(\lambda)'\ \mathrm{vec}(\beta_2)')'}{\partial(\mathrm{vec}(\beta_2))'} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{I}_r \otimes \mathbf{I}_{n-r} \end{pmatrix}$$

The Jacobians of the two transformations determine the Jacobian of the total transformation from $\boldsymbol{\Pi}$ to $(\boldsymbol{\alpha}, \lambda, \beta_2)$

$$\begin{aligned} &|J(\boldsymbol{\alpha}, \lambda, \beta_2)| \\ &\quad = \left| \frac{\partial\, \mathrm{vec}(\boldsymbol{\Pi})}{\partial(\mathrm{vec}(\boldsymbol{\alpha})'\ \mathrm{vec}(\lambda)'\ \mathrm{vec}(\beta_2)')} \right| \\ &\quad = \left| \frac{\partial\, \mathrm{vec}(\boldsymbol{\Pi})}{\partial(\mathrm{vec}(\boldsymbol{\alpha}_1)'\ \mathrm{vec}(\vartheta_2)'\ \mathrm{vec}(\lambda)'\ \mathrm{vec}(\beta_2)')} \right| \left| \frac{\partial(\mathrm{vec}(\boldsymbol{\alpha}_1)'\ \mathrm{vec}(\vartheta_2)'\ \mathrm{vec}(\lambda)'\ \mathrm{vec}(\beta_2)')'}{\partial(\mathrm{vec}(\boldsymbol{\alpha})'\ \mathrm{vec}(\lambda)'\ \mathrm{vec}(\beta_2)')'} \right| \\ &\quad = |(J_1\ J_2\ J_3\ J_4)|\ |(G_1\ G_2\ G_3\ G_4)|. \end{aligned}$$

Straightforward algebra shows that the Jacobian evaluated in $\lambda = \mathbf{0}$ equals

$$|J(\boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}| = |(\mathbf{I}_n \otimes \beta)\ \ (\boldsymbol{\alpha}' \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix})\ \ (\boldsymbol{\alpha}'_\perp \otimes \beta_\perp)|.$$

B Decomposition

The trace in the posterior density of the linear error correction model (23) can be decomposed as follows

$$\begin{aligned} &\mathrm{tr}(\Sigma^{-1}(\boldsymbol{\Pi} - \tilde{\boldsymbol{\Pi}})'(A + Y'_{-1}Y_{-1})(\boldsymbol{\Pi} - \tilde{\boldsymbol{\Pi}})) \\ &\qquad = \mathrm{tr}(\Sigma^{-1}(\beta\boldsymbol{\alpha} + \beta_\perp\lambda\boldsymbol{\alpha}_\perp - \tilde{\boldsymbol{\Pi}})'(A + Y'_{-1}Y_{-1})(\beta\boldsymbol{\alpha} + \beta_\perp\lambda\boldsymbol{\alpha}_\perp - \tilde{\boldsymbol{\Pi}}))) \\ &\qquad = \mathrm{tr}(\Sigma^{-1}((\beta\boldsymbol{\alpha} - \tilde{\boldsymbol{\Pi}})'(A + Y'_{-1}Y_{-1})(\beta\boldsymbol{\alpha} - \tilde{\boldsymbol{\Pi}}))) \\ &\qquad\quad + \mathrm{tr}((\boldsymbol{\alpha}_\perp\Sigma^{-1}\boldsymbol{\alpha}'_\perp)(\lambda - \tilde{\lambda})'\beta'_\perp(A + Y'_{-1}Y_{-1})\beta_\perp(\lambda - \tilde{\lambda})) \\ &\qquad\qquad - \mathrm{tr}(\Sigma^{-1}\boldsymbol{\alpha}'_\perp(\boldsymbol{\alpha}_\perp\Sigma^{-1}\boldsymbol{\alpha}'_\perp)^{-1}\boldsymbol{\alpha}_\perp\Sigma^{-1}(\beta\boldsymbol{\alpha} - \tilde{\boldsymbol{\Pi}})'(A + Y'_{-1}Y_{-1})\beta_\perp \\ &\qquad\qquad\quad (\beta'_\perp(A + Y'_{-1}Y_{-1})\beta_\perp)^{-1}\beta'_\perp(A + Y'_{-1}Y_{-1})(\beta\boldsymbol{\alpha} - \tilde{\boldsymbol{\Pi}})) \end{aligned}$$

with

$$\begin{aligned} \tilde{\boldsymbol{\Pi}} &= (A + Y'_{-1}Y_{-1})^{-1}(AP + Y'_{-1}\Delta Y) \\ \tilde{\lambda} &= (\beta'_\perp(A + Y'_{-1}Y_{-1})\beta_\perp)^{-1}\beta'_\perp(A(P - \beta\alpha) + Y'_{-1}(\Delta Y - \beta\alpha))\Sigma^{-1}\boldsymbol{\alpha}'_\perp(\boldsymbol{\alpha}_\perp\Sigma^{-1}\boldsymbol{\alpha}'_\perp)^{-1}. \end{aligned}$$

References

Billingsley, P., 1986, *Probability and Measure*, Wiley, New York.

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

Chib, S. and E. Greenberg, 1995, Understanding the Metropolis-Hastings Algorithm, *American Statistician* **49**, 327–335.

DeJong, D.N., 1992, Co-Integration and Trend Stationarity in Macroeconomic Time Series: Evidence from the Likelihood Function, *Journal of Econometrics* **52**, 347–370.

Dickey, J., 1971, The Weighted Likelihood Ratio, Linear Hypothesis on Normal Location Parameters, *The Annals of Mathematical Statistics* **42**, 240–223.

Doan, T., R.B. Litterman and C. Sims, 1984, Forecasting and Conditional Projections using Realistic Prior Distributions, *Econometric Reviews* **3**, 1–144.

Dorfman, J.H., 1995, A Numerical Bayesian Test for Cointegration of AR processes, *Journal of Econometrics* **66**, 289–234.

Engle, R.F. and C.W.J. Granger, 1987, Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica* **55**, 251–276.

Geweke, J., 1989, Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica* **57**, 1317–1339.

Geweke, J., 1996, Bayesian Reduced Rank Regression in Econometrics, *Journal of Econometrics* **75**, 121–146.

Golub, G.H. and C.F. van Loan, 1989, *Matrix Computations*, The John Hopkins University, Baltimore.

Hastings, W.K., 1970, Monte Carlo Sampling using Markov Chains and their Applications, *Biometrika* **57**, 97–10.

Hendry, D.F. and J.A. Doornik, 1994, Modelling Linear Dynamic Econometrics Systems, *Scottish Journal of Political Economy* **41**, 1–33.

Johansen, S., 1991, Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica* **59**, 1551–1580.

Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

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

Kass, R.E. and A.E. Raferty, 1995, Bayes Factors, *Journal of the American Statistical Association* **90**, 773–795.

Kleibergen, F., 1998, Conditional Densities in Econometrics, Unpublished working paper, available at http://www.eur.nl/few/ei/faculty/kleibergen.

Kleibergen, F. and H.K. van Dijk, 1994a, Direct Cointegration Testing in Error Correction Models, *Journal of Econometrics* **63**, 61–103.

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

Kleibergen, F. and H.K. van Dijk, 1998, Bayesian Simultaneous Equation Analysis using Reduced Rank Structures, *Econometric Theory* . forthcoming.

Kloek, T. and H.K. van Dijk, 1978, Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte-Carlo, *Econometrica* **44**, 345–351.

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

Koop, G., 1991, Cointegration in Present Value Relationships, *Journal of Econometrics* **49**, 105–139.

Leamer, E.E., 1978, *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, Wiley, New York.

Lütkepohl, H., 1993, *Introduction to Multiple Time Series Analysis*, 2nd edition, Springer-Verlag, Berlin.

Magnus, J.R. and H. Neudecker, 1988, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley & Sons, Chichester.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, 1953, Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics* **21**, 1087–1092.

Phillips, P.C.B., 1989, Partially Identified Econometric Models, *Econometric Theory* **5**, 181–240.

Phillips, P.C.B., 1991, Optimal Inference in Cointegrated Systems, *Econometrica* **59**, 283–306.

Phillips, P.C.B., 1996, Econometric Model Determination, *Econometrica* **64**, 763–812.

Phillips, P.C.B. and W. Ploberger, 1994, Posterior Odds Testing for a Unit Root with Data-based Model Selection, *Econometric Theory* **10**, 774–808.

Phillips, P.C.B. and W. Ploberger, 1996, An Asymptotic Theory of Bayesian Inference for Time Series, *Econometrica* **64**, 381–412.

Schwartz, G., 1978, Estimating the Dimension of a Model, *The Annals of Statistics* **6**, 461–464.

Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Smith, A.F.M. and G.O. Roberts, 1993, Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society B* **55**, 3–23.

Tierney, L., 1994, Markov Chains for Exploring Posterior Distributions, *The Annals of Statistics* **22**, 1701–1762.

Verdinelli, I. and L. Wasserman, 1995, Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio, *Journal of the American Statistical Association* **90**, 614–618.

Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Zellner, A., 1986, On Assesing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions, *in* P.K Goel and A. Zellner (editors), *Bayesian Inference and Decesion Techniques: Essays in Honor of Bruno de Finetti*, North-Holland, Amsterdam, pp. 233–243.