Priors, Posterior Odds and Lagrange Multiplier Statistics in Bayesian Analyses of Cointegration

Frank Kleibergen* Richard Paap[†]

Econometric Institute Reports, no.9668/a
November 1996

Abstract

Using the standard linear model as a base, a unified theory of Bayesian Analyses of Cointegration Models is constructed. This is achieved by defining (natural conjugate) priors in the linear model and using the implied priors for the cointegration model. Using these priors, posterior results for the cointegration model are obtained using a Metropolis-Hasting sampler. To compare the cointegration models mutually and with the vector autoregressive model under stationarity, we use two strategies. The first strategy uses the Bayesian interpretation of a Lagrange Multiplier statistic. The second strategy compares the models using prior and posterior odds ratios. The latter enables us to compute prior and posterior distributions over the cointegration rank and shows close resemblance with the posterior information criterion from Phillips and Ploberger (1996). To show the applicability of the derived theory, the constructed procedures are applied to data from Johansen and Juselius (1990) and a few simulated data sets.

---

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

[†]Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, NL–3000 DR Rotterdam, The Netherlands, e-mail: paap@tir.few.eur.nl. Both authors thank the Netherlands Foundation for Scientific Research (N.W.O.) for their financial support. Paper also available at:http://www.cs.few.eur.nl/~kleiberg/publications.html.

1 Introduction

The definition of the concept of cointegration by Engle and Granger (1987) has initiated a rapidly expanding literature on this topic. Although some controversies still exist in the classical statistical analysis of this phenomenon, a largely unified theory of classical statistical analysis of cointegration has emerged, see for example Phillips and Durlauf (1986), Johansen (1991) and Phillips (1991). This is not the case though with respect to the Bayesian statistical analysis of cointegration. The issues discussed in the Bayesian literature are often quite different and it is, therefore, difficult to determine the relationships between parts of the literature.

Topics which are analyzed in a Bayesian setting are, for example, implied moving averages/impulse responses resulting from the Wold decomposition of a time series, Koop (1991), the posterior distributions of the roots of the vector autoregressive models, DeJong (1992), the consequences of local nonidentification and prior specification on the posteriors of the parameters, Kleibergen and van Dijk (1994b), the number of cointegrating vectors implied by the difference between the number of unit roots of the multivariate model and the number of unit roots in the different univariate models, Dorfman (1995), and constructing posterior simulators using the Gibbs sampler, Geweke (1996). These papers typically analyze a specific problem with which one is confronted in a Bayesian cointegration study but do not include a general modelling framework which allows one to start at the outset with a unrestricted linear model and goes through a number of decision problems to end with the posteriors of the parameters of the cointegration model.

The purpose of this paper is to construct such a unified framework. The sections of the paper, which discuss the different steps in this construction, are organized as follows. In section 2 cointegration in a Vector AutoRegressive (VAR) model is defined. We rewrite the VAR model as an unrestricted Error Correction Model (ECM) to obtain a parameter which reflects cointegration, i.e. it is equal to zero when cointegration occurs. In section 3, the implied prior and posterior for the parameters of the unrestricted ECM, using a diffuse prior on the VAR parameters, are constructed. The prior and posterior of the parameters of the cointegration model then equal the conditional prior and conditional posterior of the parameters of the unrestricted ECM given that the parameter reflecting cointegration is equal to zero. This is identical to the classical statistical analysis where the likelihood of the cointegration model equals the conditional likelihood of the parameters of the unrestricted model given that the parameter reflecting cointegration is equal to zero. The, in this way obtained, posterior does not belong to known class of probability densities. In section 4, a Metropolis-Hastings sampler is constructed to generate drawings from the posterior. Using this posterior simulator, in section 5 a Bayesian Lagrange Multiplier statistics to test for the number of cointegrating vectors is constructed. As the computation of a Bayesian Lagrange Multiplier cointegration statistic using a Metropolis-Hastings sampler is not straightforward, we use a linear regression model to show the involved steps. In section 6 the analysis is extended to allow for natural conjugate priors on the VAR parameters, like the informative Minnesota Priors of Doan *et. al.* (1984). To compare the models under different cointegration ranks, in section 7 the construction of Bayes factors and prior and posterior odds ratios is discussed. The relationships between the Bayesian Lagrange Multiplier cointegration statistic and the Likelihood ratio statistic of Johansen (1991), and between the Bayes factor and the Posterior Information Criterium of Phillips and Ploberger (1994,1996), are discussed in section 8. Section 9 shows some applications of the derived procedures by analyzing the Danish dataset from Johansen and Juselius (1990) and a few simulated datasets. The last section concludes and mentions topics for further research.

2 The Cointegration Model

Consider a Vector Autoregressive Model of order p [VAR(p)] for a k-dimensional vector of time series Y_t, for $t = 1, \ldots, T$,

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

where ε_t is a k-dimensional vector normal process with zero mean and variance Ω and where μ and γ are $(k \times 1)$ vectors containing the constant and trend coefficients. The initial values $Y_{-p+1}, \ldots, Y_0$ are fixed. The model in (1) can be rewritten in the error correction form,

$$\Delta Y_t = \mu + \gamma t + \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \varepsilon_t, \tag{2}$$

where $\Pi = \sum_{j=1}^{p} \Phi_j - I_k$ and $\Gamma_i = -\sum_{j=i+1}^{p} \Phi_j$, see Johansen (1991).

The characteristic polynomial of model (1) is equal to $|\Phi(z)| = | I_k - \sum_{i=1}^{p} z^i \Phi_i |$. Since by definition $\Phi(1) = -\Pi$, unit roots enter the model when $\Phi(1)$ $(= -\Pi)$ has a lower rank value. If Π is a zero matrix, the characteristic polynomial has k unit roots. If $k - r$ roots of the polynomial $|\Phi(z)|$ are equal to 1, $0 \leq r \leq k$, the rank of Π equals r and we say that series generated by model (1) are cointegrated. Thus, cointegration implies that we can write the matrix Π as a product of two full rank $(k \times r)$ matrices α' and β,

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

where β contains the cointegrating vectors and α contains the adjustment parameters.

The possible number of unit roots in the analyzed VARs has been the topic of a considerable amount of recent research, for an overview, see Watson (1994). Some discussion still exists about whether one should impose, see Phillips (1991), or test for the number of unit roots, see Johansen (1991). In this paper, we accord with both of these opinions by developing Bayesian estimation, selection and diagnostic testing procedures which are used with imposed numbers of unit roots. The diagnostic testing procedures are Bayesian Lagrange Multipliers statistics and are the analogs of the Lagrange Multiplier Cointegration statistics, developed in Kleibergen and van Dijk (1994a) and Kleibergen (1998). These statistics show close resemblance with the Likelihood Ratio statistics for cointegration provided by Johansen (1991) but differ from these as the model is only analyzed under the hypothesized number of unit roots.

Since the number of parameters in $\alpha'\beta'$, $2kr$ is larger than the number of free parameters in Π, under reduced rank r $(= kr + (k-r)r)$ the α and/or β parameters have to be restricted to become estimable. In this paper we impose the following restriction on the cointegrating vectors β,

$$\beta = (I_r \;\; -\beta_2)', \tag{4}$$

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

To save on notation, it is convenient to write the error correction model (2) with $\Pi = \alpha'\beta'$ in matrix notation,

$$\Delta Y = Y_{-1}\beta\alpha + X\Gamma + \varepsilon, \tag{5}$$

where $\Delta Y = (\Delta y_1 \ldots \Delta y_T)'$, $Y_{-1} = (y_0 \ldots y_{T-1})'$, $\varepsilon = (\varepsilon_1 \ldots \varepsilon_T)'$, $X = (X_1 \ldots X_T)'$, $X_t = (\Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}, 1, t)$, and $\Gamma = (\Gamma_1 \ldots \Gamma_{p-1}\ \mu\ \delta)'$. To save further on notation, in the remainder of this paper we focus on a simple VAR(1) model without deterministic elements,

$$\begin{aligned} \Delta Y &= Y_{-1}\beta\alpha + \varepsilon \\ &= Y_{1,-1}\alpha - Y_{2,-1}\beta_2\alpha + \varepsilon, \end{aligned} \tag{6}$$

where $Y_{1,-1}$ consists of the first r columns of Y_{-1} and $Y_{2,-1}$ consists of the last $k-r$ columns of Y_{-1}. This is not a serious restriction since under a flat prior on D, integrating out the D parameters from the likelihood function leads to analyzing model (6) for the transformed data $M_D \Delta Y$ and $M_D Y_{-1}$, where $M_D = I_T - D(D'D)^{-1}D'$.

3 Prior and Posterior Specification

3.1 Priors

In the Error Correction Cointegration Model (ECCM) with r cointegrating vectors ($k-r$ unit roots) specified by,

$$\Delta Y = Y_{1,-1}\alpha - Y_{2,-1}\beta_2\alpha + \varepsilon, \qquad (7)$$

where ΔY, $Y_{-1} = (Y_{1,-1}\ Y_{2,-1})$, $\varepsilon : T \times k$; $Y_{1,-1} : T \times r$; $Y_{2,-1} : T \times (k-r)$; $\alpha : r \times k$; $\beta_2 : (k-r) \times r$; and $\varepsilon \sim n(0, \Omega \otimes I_T)$, the parameter β_2 is locally nonidentified when $\alpha = 0$, for more discussion on local nonidentification, see Phillips (1989). Consequently, if a diffuse prior is used, such that the joint posterior of the parameters is proportional to likelihood, the conditional posterior of β_2 given α is constant and nonzero when $\alpha = 0$. The integral over this conditional posterior at $\alpha = 0$, which is part of the marginal posterior of α, is, therefore, proportional to the volume of the parameter region of β_2 ($\mathbb{R}^{(k-r)r}$), which is infinity. This leads to a a posteriori favor for locally nonidentified parameter values when diffuse priors are used for the parameters (α, β_2). See also Kleibergen and van Dijk (1994b), (1996) and Kleibergen and Hoek (1996) for a more elaborate discussion of this phenomenon. So, diffuse priors for models which are nonlinear in the parameters, like the ECCM (7), do not lead to posteriors with similar properties as the posteriors using diffuse priors have for models which are linear in the parameters. So, from a posterior perspective, diffuse priors for nonlinear models are not the natural extension of the diffuse priors in linear models. The natural extension of the diffuse prior, for a linear model, for the ECCM (7) results when we analyze the ECCM as a restriction of a unrestricted Error Correction Model (ECM),

$$\Delta Y = Y_{1,-1}\alpha - Y_{2,-1}\beta_2\alpha + Y_{2,-1}\begin{pmatrix} 0 & \lambda \end{pmatrix} + \varepsilon, \qquad (8)$$

where $\lambda : (k-r) \times (k-r)$ and the ECCM corresponds with $\lambda = 0$, see Kleibergen and van Dijk (1994a,b). As this model is observationally equivalent with a multivariate linear model,

$$\Delta Y = (Y_{1,-1}\ Y_{2,-1}) \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} + \varepsilon, \qquad (9)$$

where $\begin{pmatrix}\Pi_{11} & \Pi_{12}\end{pmatrix} = \alpha$, $\begin{pmatrix}\Pi_{21} & \Pi_{22}\end{pmatrix} = \begin{pmatrix}0 & \lambda\end{pmatrix} - \beta_2\alpha$. We can construct the prior for the parameters α, β_2 and λ which is implied by a diffuse (Jeffreys') prior on Π_{11}, Π_{12}, Π_{21} and Π_{22}. These priors are stated in theorem 1.

Theorem 1 *Diffuse (Jeffreys') priors for the parameters (Π_{11}, Π_{21}), (Π_{12}, Π_{22}), and Ω of the multivariate linear model (9), which read,*

$$p_{lin}(\Omega) \propto |\Omega|^{-\frac{1}{2}(k+1)}, \tag{10}$$

$$p_{lin}(\Pi_{11}, \Pi_{12}|\Omega) \propto |\Omega|^{-\frac{1}{2}r}|Y'_{1,-1}M_{Y_{2,-1}}Y_{1,-1}|^{\frac{1}{2}k}, \tag{11}$$

$$p_{lin}(\Pi_{21}, \Pi_{22}|\Pi_{11}, \Pi_{12}, \Omega) \propto |\Omega|^{-\frac{1}{2}(k-r)}|Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}k}, \tag{12}$$

imply the following priors for the parameters α, λ and β_2, of the unrestricted ECM (8),

$$p_{unres}(\Omega) \propto |\Omega|^{-\frac{1}{2}(k+1)}, \tag{13}$$

$$p_{unres}(\alpha|\Omega) \propto |\Omega|^{-\frac{1}{2}r}|Y'_{1,-1}M_{Y_{2,-1}}Y_{1,-1}|^{\frac{1}{2}k}, \tag{14}$$

$$\begin{aligned} p_{unres}(\lambda|\alpha, \Omega) \propto\ & \left|\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r}\end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r}\end{pmatrix}'\right|^{-\frac{1}{2}(k-r)} \\ & |Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}(k-r)}, \end{aligned} \tag{15}$$

$$p_{unres}(\beta_2|\lambda, \alpha, \Omega) \propto |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)}|Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}r}, \tag{16}$$

where unres refers to priors for the unrestricted ECM (8), and lin for the linear model (9), and $\alpha = (\alpha_1\ \alpha_2)$, where $\alpha_1 : r \times r$, $\alpha_2 : (k-r) \times r$.

Proof: see appendix.

The implicit priors for α, λ and β_2 implied by the diffuse (Jeffreys') prior for $\Pi_{11}, \Pi_{21}, \Pi_{12}$ and Π_{22} are constructed such that they obey the sequence, in which the parameter matrices should be analyzed conditional on one another, dictated by the model: the cointegrating vectors β_2 have to be analyzed given λ, α and Ω and λ has to be analyzed given α and Ω. Only this sequence allows for an analytical decomposition of the joint posterior into conditional posteriors. The priors in theorem 1 show the implied priors for the ECCM where $\lambda = 0$. These priors are obtained as follows. The joint posterior of the parameters of the ECCM equals the conditional posterior of the parameters in the unrestricted ECM given $\lambda = 0$. As the posterior is proportional to the product of the prior and likelihood, and the prior does not depend on λ, the joint prior for the parameters of the ECCM equals the joint prior of the parameters of the unrestricted ECM. This prior is stated in lemma 2.

Lemma 2 *The prior for the parameters α, β_2 and Ω of the ECM (4) implied by a diffuse (Jeffreys') prior, on the parameters of the unrestricted linear model (3), reads,*

$$
\begin{aligned}
p_{ecm}(\alpha,\Omega) &\propto |\Omega|^{-\frac{1}{2}(k+n+1)}\left|Z_{1,-1}'M_{Z_{2,-1}}Z_{1,-1}\right|^{\frac{1}{2}k}\left|Z_{2,-1}'Z_{2,-1}\right|^{\frac{1}{2}(k-r)} \quad (17)\\
&\quad \left|\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r}\end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r}\end{pmatrix}'\right|^{-\frac{1}{2}(k-r)},\\
p_{ecm}(\beta_2|\alpha,\Omega) &\propto \left|\alpha\Omega^{-1}\alpha'\right|^{\frac{1}{2}(k-r)}\left|Z_{2,-1}'Z_{2,-1}\right|^{\frac{1}{2}r}, \quad (18)
\end{aligned}
$$

where ecm indicates that the priors are defined for the ECM (4).

Proof: see appendix.

The analog to the fact that the prior for (α,Ω) incorporates parts of the prior of the restricted parameter β, see lemma 2, in the standard linear model is the so-called increase in degrees of freedom. This occurs when certain parameters are set to a priori known values. For example, consider a linear model with two explanatory variables,

$$y = x_1\beta_1 + x_2\beta_2 + \varepsilon, \quad (19)$$

where y, x_1, x_2, ε are $T\times 1$ matrices and $\varepsilon \sim n(0,\sigma^2 I_T)$. A diffuse (Jeffreys') prior, for this model equals,

$$p(\sigma^2,\beta_1,\beta_2) \propto \sigma^{-2}. \quad (20)$$

In case one wants to analyze the posteriors under the assumption that $\beta_1 = 0$, it holds that the joint posterior of the parameters (σ^2,β_2) of the restricted model, equals the conditional posterior of (σ^2,β_2) given $\beta_1 = 0$. So, the resulting conditional posterior of (σ^2,β_2) given $\beta_1 = 0$, which is the marginal posterior of (σ^2,β_2) in the model assuming $\beta_1 = 0$, has so-called more degrees of freedom than the marginal posterior of (σ^2,β_2). In practice, this increase in degrees of freedom is essentially only incorporated when one wants to compare the posteriors under $\beta_1 = 0$ with the unrestricted case. It is often neglected when analyzing the model under the assumption $\beta_1 = 0$, since one does not know in practice how many other parameters one has a priori assumed to be equal to zero and it does not crucially affect the resulting posteriors. So, the additional term appearing in the priors from lemma 2 is comparable with a degrees of freedom factor but has an important difference with the degrees of freedom factor in the standard linear model as it does not only depend on the variance parameters.

2.2 Posteriors

As already stated, only for the specific sequence of the parameters, in which we stated the conditional priors of the parameters in theorem 1, is it possible to derive analytical expressions for the conditional posteriors of the parameters from the unrestricted ECM given the priors from theorem 1. If we follow this specific sequence of the parameters, Π_2 has to be analyzed given (β, α, Ω) and β has to be analyzed given (α, Ω). The conditional posteriors, which obey this sequence, are stated in theorem 2, see also Kleibergen (1996a).

Theorem 2 *The conditional/marginal posteriors of the parameters of the unrestricted ECM (Π, Ω, α, β, and Π_2) using the priors from theorem 1, read,*

$$
\begin{aligned}
p_{unres}(\Omega|Y) &\propto |\Omega|^{-\frac{1}{2}(T+k+1)} \exp[-\frac{1}{2} tr(\Omega^{-1}\Delta Y' M_{X_{-1}} \Delta Y)], && (21)\\
p_{unres}(\alpha|\Omega, Y) &\propto |\Omega|^{-\frac{1}{2}r} |Y_{1,-1}' M_{X_{2,-1}} Y_{1,-1}|^{\frac{1}{2}k} \\
&\quad \exp[-\frac{1}{2} tr(\Omega^{-1}(\alpha-\hat{\alpha})' Y_{1,-1}' M_{X_{2,-1}} Y_{1,-1}(\alpha-\hat{\alpha}))] && (22)\\
p_{unres}(\beta|\alpha, \Omega, Y) &\propto |\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix} \Omega \begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}'|^{-\frac{1}{2}(k-r)} \\
&\quad |Y_{2,-1}' Y_{2,-1}|^{\frac{1}{2}(k-r)} \exp[-\frac{1}{2} tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix} && (23)\\
&\quad \Omega \begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}')^{-1}(\beta-\hat{\beta})' Y_{2,-1}' Y_{2,-1}(\beta-\hat{\beta}))], \\
p_{unres}(\Pi_2|\beta, \alpha, \Omega, Y) &\propto |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)} |Y_{2,-1}' Y_{2,-1}|^{\frac{1}{2}r} && (24)\\
&\quad \exp[-\frac{1}{2} tr(Y_{2,-1}' Y_{2,-1}((\Pi_2-\hat{\Pi}_2)\alpha\Omega^{-1}\alpha'(\Pi_2-\hat{\Pi}_2)'))],
\end{aligned}
$$

where $\alpha = (\alpha_1\ \alpha_2)$, $\alpha_1 : r\times r$, $\alpha_2 : r\times(k-r)$, $\hat{\alpha} = (Y_{1,-1}' M_{X_{2,-1}} Y_{1,-1})^{-1} Y_{1,-1}' M_{X_{2,-1}} \Delta Y$, $\hat{\beta} = (Y_{2,-1}' Y_{2,-1})^{-1} Y_{2,-1}' \Delta Y \begin{pmatrix} -\alpha_1^{-1}\alpha_2 \\ I_{k-r} \end{pmatrix}$, $\hat{\Pi}_2 = -(Y_{2,-1}' Y_{2,-1})^{-1} Y_{2,-1}'(\Delta Y - Y_{1,-1}\alpha - Y_{2,-1}\begin{pmatrix} 0 & \beta \end{pmatrix})\Omega^{-1}\alpha'(\alpha\Omega^{-1}\alpha')^{-1}$.

Proof : see appendix.

All posteriors in theorem 2 belong to a known class of probability density functions, either inverted-Wishart or matrix normal, for a definition of these see Zellner (1971). As shown in lemma 2, the joint prior for the ECM, where it is assumed that $\beta = 0$, is identical to the joint prior for the parameters of the unrestricted ECM. Consequently, the functional form of the conditional posteriors of the parameters which are analyzed conditional on β are identical for the unrestricted ECM and the RECM. So, the conditional posterior of β_2 given (α, Ω) in the RECM is proportional to the conditional posterior of β_2 given $(\alpha, \lambda = 0, \Omega)$, see (34). The posterior of the parameters on which λ is analyzed conditionally, i.e. α and Ω, does change,

$$\begin{aligned} p_{recm}(\alpha, \Omega | Y) \propto\ & |\Omega|^{-\frac{1}{2}(T+2k+1)} |X_{1,-1}' M_{X_{2,-1}} X_{1,-1}|^{\frac{1}{2}k} |X_{2,-1}' X_{2,-1}|^{\frac{1}{2}(k-r)} \\ & |\begin{pmatrix} -\alpha_2' \alpha_1^{-1\prime} \\ I_{k-r} \end{pmatrix}' \Omega \begin{pmatrix} -\alpha_2' \alpha_1^{-1\prime} \\ I_{k-r} \end{pmatrix} |^{-\frac{1}{2}(k-r)} \\ & \exp[-\frac{1}{2}[tr((\begin{pmatrix} -\alpha_2' \alpha_1^{-1\prime} \\ I_{k-r} \end{pmatrix}' \Omega \begin{pmatrix} -\alpha_2' \alpha_1^{-1\prime} \\ I_{k-r} \end{pmatrix})^{-1} \\ & \hat{\lambda}' X_{2,-1}' X_{2,-1} \hat{\lambda}) + tr(\Omega^{-1} Y' M_{X_{-1}} Y) \\ & + tr(\Omega^{-1} (\alpha - \hat{\alpha})' X_{1,-1}' M_{X_{2,-1}} X_{1,-1} (\alpha - \hat{\alpha}))]], \end{aligned} \tag{35}$$

where $\hat{\lambda}$ and $\hat{\alpha}$ have been defined in theorem 3. The posterior in equation (35) does not belong to a known class of probability density functions and, therefore, we construct in the next section a simulation procedure to evaluate the posterior. The simulation procedure is based on the ratio of the joint posterior of α and Ω resulting from the RECM (35), and the marginal posterior of (α, Ω) in the unrestricted ECM, (31) and (33),

$$\begin{aligned} & \frac{p_{recm}(\alpha, \Omega | Y)}{p_{uecm}(\alpha, \Omega | Y)} \\ \propto\ & |\begin{pmatrix} -\alpha_2' \alpha_1^{-1\prime} \\ I_{k-r} \end{pmatrix}' \Omega \begin{pmatrix} -\alpha_2' \alpha_1^{-1\prime} \\ I_{k-r} \end{pmatrix} |^{-\frac{1}{2}(k-r)} |X_{2,-1}' X_{2,-1}|^{\frac{1}{2}(k-r)} \\ & \exp[-\frac{1}{2} tr((\begin{pmatrix} -\alpha_2' \alpha_1^{-1\prime} \\ I_{k-r} \end{pmatrix}' \Omega \begin{pmatrix} -\alpha_2' \alpha_1^{-1\prime} \\ I_{k-r} \end{pmatrix})^{-1} \hat{\lambda}' X_{2,-1}' X_{2,-1} \hat{\lambda})] \\ \propto\ & p_{uecm}(\lambda = 0 | \alpha, \Omega, Y). \end{aligned} \tag{36}$$

Note that this ratio equals the conditional posterior of λ given α and Ω given in (33), evaluated in the hypothesized parameter point, $\lambda = 0$. Not surprisingly this ratio plays an important role in the computation of the posterior odds ratios.

4 Simulating Posterior Distributions

To evaluate the posterior distributions of the RECM (7), we use Markov Chain Monte Carlo techniques. Since not all of the full conditional distributions for our posterior are of a standard type, standard Gibbs sampling is not possible. Therefore, we use a Metropolis-Hastings (M-H) sampler, see Metropolis *et al.* (1953), Hastings (1970) and more recently Smith and Roberts (1993) and Tierney (1994).

The total simulation scheme is based on the following decomposition of the posterior distribution

$$
\begin{aligned}
p_{ecm}(\beta, \alpha, \lambda_2|Y) &\propto w(\beta, \alpha|Y) p_{uecm}(\alpha, \beta|Y) p_{uecm}(\lambda_2|\beta, \alpha, Y) \qquad (37)\\
&\propto w(\beta, \alpha|Y) p_{uecm}(\beta|Y) p_{uecm}(\alpha|\beta, Y) p_{uecm}(\lambda_2|\alpha, \beta, Y),
\end{aligned}
$$

where $p_{uecm}(\beta|Y)$, $p_{uecm}(\alpha|\beta, Y)$ and $p_{uecm}(\lambda_2|\alpha, \beta, Y)$ $(\propto p_{ecm}(\lambda_2|\alpha, \beta, Y))$ are given in (31), (33) and (34), and $w(\beta, \alpha|Y)$ is a bounded weight function, which is given by the ratio of posteriors in (36), $w(\beta, \alpha) = \frac{p_{ecm}(\alpha,\beta|Y)}{p_{uecm}(\alpha,\beta|Y)} \propto p_{uecm}(\lambda = 0|\alpha, \beta, Y)$.

If we ignore the weight function in (37), simulation from the posterior distribution is easy, since the remainder consists of a product of standard densities. Since $w(\beta, \alpha|Y)$ is a bounded function, we can use an acceptance-rejection simulation algorithm. This may, however, lead to large rejection frequencies if the cointegration rank is not correctly specified. Chib and Greenberg (1995) show that in this case a M-H algorithm can speed up the simulation process. Since λ_2 does not enter the weight function w, the M-H step only enters the simulation scheme for the generation of the β and the α parameters. The *candidate-generating* density is $p_{uecm}(\alpha, \beta|Y)$ or $p_{uecm}(\beta|Y)p_{uecm}(\alpha|\beta, Y)$. The acceptance-rejection step simplifies to a ratio of weight functions $w(\beta, \alpha|Y)$. Given the drawings for β and α, we generate a drawing for λ_2 conditional on β and α from a matrix normal distribution.

The four steps to generate from the posterior distribution including the Metropolis-Hasting step can be summarized as follows,

1. Draw β^i from $p_{uecm}(\beta|Y)$
2. Draw α^i from $p_{uecm}(\alpha|\beta^i, Y)$
3. Accept (β^i, α^i) with probability

   $$\min\left\{\frac{g(\beta^i, \alpha^i|Y)w(\alpha^i, \beta^i|Y)g(\beta^{i-1}, \alpha^{i-1}|Y)}{w(\beta^{i-1}, \alpha^{i-1}|Y)g(\beta^{i-1}, \alpha^{i-1}|Y)g(\beta^i, \alpha^i|Y)}, 1\right\} = \min\left\{\frac{w(\beta^i, \alpha^i|Y)}{w(\beta^{i-1}, \alpha^{i-1}|Y)}, 1\right\},$$

   otherwise $(\beta^i, \alpha^i) = (\beta^{i-1}, \alpha^{i-1})$.
4. Draw λ_2^i from $p_{uecm}(\lambda_2|\beta^i, \lambda = 0, \alpha^i, Y)$.

The first three steps of this iterative scheme generate a Markov Chain. When the chain has converged, say after N, iterations, the simulated values $[\beta^i, \alpha^i, i > N]$ can be used as a sample from the joint posterior $p_{ecm}(\beta, \alpha|Y)$, see Tierney (1994) for details.

This simulation scheme has advantages if one wants to analyze the model under every cointegration rank r. Since the sampling of β does not depend on the rank r, one only needs one drawing Ω for every cointegration rank. Furthermore, using the properties of the matrix normal distribution, the sampling of α parameters can be accelerated. Instead of drawing an α matrix for every rank r, one can sample the α matrices at once using a drawing Π from,

$$p(\Pi|\Omega, Y) \propto |\Omega|^{-\frac{1}{2}k} \exp[-\frac{1}{2} tr(\Omega^{-1}(\Pi - \hat{\Pi})'Y_{-1}'Y_{-1}(\Pi - \hat{\Pi}))], \qquad (38)$$

where $\hat{\Pi} = (Y_{-1}'Y_{-1})^{-1}Y_{-1}'\Delta Y$. The α drawings under the cointegration rank r are obtained by taking the the first r rows of the drawing Π for $r = 1, \ldots, k$.

The presented sampling scheme is not unique. It is possible to use a different decomposition than the one proposed in (37). Furthermore, the simulation scheme can be adapted to be applicable for more complicated models, like for instance VAR models with a break in the constant and/or in the cointegration relation or threshold cointegration models. These more complicated models are often analyzed in a Gibbs framework. The sampling of the block $(\Omega, \alpha, \beta_2)$ given the remaining parameters in the model can then be done using the simulation steps presented in this section.

3 Bayesian Diagnostic Cointegration Testing

In the previous section, we assumed for the derivation of the posterior simulators, that the number of cointegrating vectors was known a priori. This is in practice seldom the case such that procedures, which analyze whether the chosen number of cointegrating vectors is plausible, are needed. In classical statistical analysis diagnostic test statistics like Lagrange Multiplier (LM) or score statistics are intended for this purpose. In this section, we will construct the Bayesian analog of these classical LM statistics to test whether the assumed number of cointegrating vectors is plausible. These Bayesian LM statistics can be computed using the M-H simulation procedure proposed in the previous section. The Bayesian LM cointegration statistics are extensions of a LM statistic in a linear regression model discussed in the next subsection.

3.1 Bayesian LM statistics in a Linear Model

Consider again a linear regression model with two explanatory variables,

$$y = x_1\gamma_1 + x_2\gamma_2 + \varepsilon, \qquad (39)$$

where y, ε, x_1, x_2 : $T \times 1$, $\varepsilon \sim N(0, \sigma^2 I_T)$. If we are interested whether the parameter γ_1 is zero, we can test this hypothesis using a Highest Posterior Density (HPD) region, see Box and Tiao (1973). An alternative method to test the hypothesis, $H_0 : \gamma_1 = 0$, is to use a Bayesian analog of a Lagrange Multiplier (LM) statistic, which can be seen as a generalization of a HPD region test. Since in the linear model the marginal posterior distributions are of a known type, it is possible to calculate the LM statistic directly. In the cointegration model, however, the marginal distributions are of a unknown form and we use a M-H sampler to simulate from the posterior distribution. In this subsection, we calculate the LM statistic for the restriction $\gamma_1 = 0$ analytically and using a M-H sampling approach. Of course both calculation methods result in the same outcome. This result can then be generalized to the cointegration model, which will be discussed in the next subsection.

5.1.1 Analytical Approach

Assuming diffuse priors for the different parameters,

$$p(\gamma_1, \gamma_2, \sigma^2) \propto \sigma^{-4}, \tag{30}$$

some conditional and marginal posteriors of the parameters of the two variable linear model read,

$$\begin{aligned}
p(\gamma_2|\gamma_1, \sigma^2, y, X) &\propto \sigma^{-1} \exp[-\frac{1}{2\sigma^2}(\gamma_2 - \hat{\gamma}_2)' x_2' x_2 (\gamma_2 - \hat{\gamma}_2)], & (31)\\
p(\sigma^2|\gamma_1, y, X) &\propto \sigma^{-(T+3)} \exp[-\frac{1}{2\sigma^2}(y - x_1\gamma_1)' M_{x_2}(y - x_1\gamma_1)], & (32)\\
p(\gamma_1|\sigma^2, y, X) &\propto \sigma^{-1} \exp[-\frac{1}{2\sigma^2}(\gamma_1 - \hat{\gamma}_1)' x_1' M_{x_2} x_1 (\gamma_1 - \hat{\gamma}_1)], & (33)
\end{aligned}$$

where $\hat{\gamma}_1 = (x_1' M_{x_2} x_1)^{-1} x_1' M_{x_2} y$, $\hat{\gamma}_2 = (x_2' x_2)^{-1} x_2'(y - x_1\gamma_1)$, $M_{(x_1\ x_2)} = I_T - (x_1\ x_2)((x_1\ x_2)'(x_1\ x_2))^{-1}(x_1\ x_2)'$.

Consider the model for γ_1 given σ^2. To derive the distribution of the Bayesian LM statistic for the hypothesis, $H_0 : \gamma_1 = 0$, under the alternative hypothesis, we use the conditional posterior of γ_1 given σ^2 (33), since

$$\begin{aligned}
(\gamma_1 - \hat{\gamma}_1) &\sim N(0, \sigma^2 (x_1' M_{x_2} x_1)^{-1}) \Leftrightarrow & (34)\\
\sigma^{-1}(x_1' M_{x_2} x_1)^{\frac{1}{2}}(\gamma_1 - \hat{\gamma}_1) &\sim N(0, 1) \Leftrightarrow &\\
\sigma^{-1}(x_1' M_{x_2} x_1)^{-\frac{1}{2}} x_1' M_{x_2} \varepsilon, &\sim N(0, 1), &
\end{aligned}$$

the LM statistic given σ^2 is equal to the square of the last two expressions in (34). The distribution of the LM statistic in this model is, therefore, χ^2

with one degree of freedom. This results holds regardless of the value of σ^2 such that this property is not lost when we go to the marginal result for γ_1 by integrating out σ^2,

$$E_{\sigma^2}(\sigma^{-2}y'M_{x_2}x_1(x_1'M_{x_2}x_1)^{-1}x_1'M_{x_2}y) \sim \chi^2(1). \tag{35}$$

If we substitute $\gamma_1 = 0$ in this expression and use the conditional posterior of σ^2 in (31) with $\gamma_1 = 0$, we obtain the value of this LM statistic under $H_0 : \gamma_1 = 0$,

$$\begin{aligned} t_{BLM}(\gamma_1 &= 0) = E_{\sigma^2}(\sigma^{-2}y'M_{x_2}x_1(x_1'M_{x_2}x_1)^{-1}x_1'M_{x_2}y|\gamma_1 = 0) \\ &= y'M_{x_2}x_1(x_1'M_{x_2}x_1)^{-1}x_1'M_{x_2}y/(y'M_{x_2}y/T) \end{aligned} \tag{36}$$

Now we reject the hypothesis $\gamma_1 = 0$ when the resulting LM statistic (36) lies outside the 95% HPD region of a $\chi^2(1)$ distribution. This can be seen as a generalization of testing whether $\gamma_1 = 0$ using a HPD region for the marginal distribution of γ_1, which is t distributed. In the next theorem we show that it is also possible to obtain the Bayesian LM statistic by adjusting $\sigma^{-2}y'X(X'X)^{-1}X'y$, $X = (x_1\ x_2)$.

Theorem 4 *The Bayesian LM statistic to test $H_0 : \gamma_1 = 0$, in the linear model (29), specified by,*

$$t_{BLM}(\gamma_1 = 0) = E_{\sigma^2}(\sigma^{-2}y'M_{x_2}x_1(x_1'M_{x_2}x_1)^{-1}x_1'M_{x_2}y|\gamma_1 = 0), \tag{37}$$

is equal to

$$\begin{aligned} E_{\gamma_2}E_{\sigma^2}(\sigma^{-2}y'X(X'X)^{-1}X'y|\gamma_1 &= 0) - E(\chi^2(1)) = \\ E_{\gamma_2}E_{\sigma^2}(\sigma^{-2}y'X(X'X)^{-1}X'y|\gamma_1 &= 0) - 1, \end{aligned} \tag{38}$$

where $X = (x_1\ x_2)$.

Proof: see appendix.

5.1.2 The Metropolis-Hastings Approach

Theorem 4 extends also to other kind of hypotheses on γ_1, γ_2, $H_0 : h(\gamma_1, \gamma_2) = 0$, and can be used in any kind of linear model. For certain nonlinear hypotheses on the parameters of a linear model, like the reduced rank restriction for cointegration models, Bayesian LM statistics can only be constructed by using (generalizations of) theorem 4, which explains why the theorem is needed in our case.

To show this latter point, consider the case that we do not construct the Bayesian LM statistic using the marginal and conditional posteriors assuming that $\beta_1 = 0$, but use the marginal posterior of σ^2 and β_2 given σ^2 from the unrestricted model in a M-H sampling approach. So, the marginal/conditional densities from which σ^2 and β_2 are sampled read,

$$
\begin{aligned}
g(\sigma^2|y, X) &\propto \sigma^{-(T+2)} \exp[-\frac{1}{2\sigma^2} y' M_{(x_1\ x_2)} y], && (39)\\
g(\beta_2|\sigma^2, y, X) &\propto \sigma^{-1} \exp[-\frac{1}{2\sigma^2} (\beta_2 - \tilde{\beta}_2)' x_2' M_{x_1} x_2 (\beta_2 - \tilde{\beta}_2)], && (40)
\end{aligned}
$$

where $\tilde{\beta}_2 = (x_2' M_{x_1} x_2)^{-1} x_2' M_{x_1} y$. To correct for not sampling from the true posterior, we have to include a weight function, see (37), which is the ratio of the true posterior and the density from which we sample. This weight function equals,

$$
w(\sigma^2, \beta_2) = \sigma^{-1} \exp[-\frac{1}{2\sigma^2} \tilde{\beta}_1' x_1' x_1 \tilde{\beta}_1], \qquad (41)
$$

where $\tilde{\beta}_1 = (x_1' x_1)^{-1} x_1' (y - x_2 \beta_2)$, i.e. the mean of the conditional posterior of β_1 given β_2. These weights are used to compute acceptance-rejection probabilities.

Using the output of the M-H sampler, we can calculate the LM statistic to test $H_0 : \beta_1 = 0$ in (36). We can also use the result from theorem 4 and calculate the Bayesian LM statistic using $\sigma^{-2} \varepsilon' X (X'X)^{-1} X' \varepsilon$. This latter expression can be decomposed in a part of the kernel of the sampling density of β_2 and part of the weight function (41),

$$
\begin{aligned}
& \sigma^{-2} \varepsilon' X (X'X)^{-1} X' \varepsilon && (42)\\
= \; & \sigma^{-2} (y - x_2 \beta_2)' (M_{x_1} x_2 (x_2' M_{x_1} x_2)^{-1} x_2' M_{x_1} + \\
& x_1 (x_1 x_1)^{-1} x_1') (y - x_2 \beta_2) \\
= \; & \sigma^{-2} [(\beta_2 - \tilde{\beta}_2)' x_2' M_{x_1} x_2 (\beta_2 - \tilde{\beta}_2) + \tilde{\beta}_1' x_1' x_1 \tilde{\beta}_1],
\end{aligned}
$$

Note that the Bayesian LM statistic does not correspond with the expectation of the last part of equation (42), $\sigma^{-2} \tilde{\beta}_1' x_1' x_1 \tilde{\beta}_1$. Since σ^2 and β_2 are sampled using a M-H algorithm, $E_{pt}(\sigma^{-2} (\beta_2 - \tilde{\beta}_2)' x_2' M_{x_1} x_2 (\beta_2 - \tilde{\beta}_2))$ does not have a $\chi^2(1)$ distribution.

For the reduced rank cointegration hypotheses, discussed in the next subsection, the specific dependence of the parameters on one another does only allow for the kind of decompositions as in equation (42). Closed form expressions of the Bayesian LM cointegration statistic, like equation (36), do, therefore, not exist. These Bayesian LM statistics can still be calculated through using the results of (generalizations of) theorem 4.

5.3 Bayesian LM cointegration statistics

The specification of the ECM in equation (7) corresponds with the hypothesis, $H_0 : \lambda = 0$, in the unrestricted ECM (8). Since the marginal posterior of the parameter reflecting cointegration, λ, in the unrestricted ECM, cannot be constructed analytically, Bayesian LM statistics to test for cointegration do not have a closed form analytical expression, as in (36). The marginal posterior of λ can be calculated by sampling from the different marginal and conditional posteriors but as its conditional posterior depends on α_1^{-1}, both in its mean and variance, inference on λ can depend on the ordering of the variables in X_{-1} into $X_{1,-1}$ and $X_{2,-1}$, see Kleibergen and van Dijk (1994a). When we use the model with $\lambda = 0$, as for the construction of the Bayesian LM statistic, to analyze whether $\lambda \neq 0$, this is not the case. We, therefore, prefer to perform the analysis whether $\lambda = 0$, using this restricted model. As analytical expressions for the Bayesian LM statistic to test, $H_0 : \lambda = 0$, do not exist, we use the multivariate extension of theorem 4 to form the Bayesian LM statistic using $tr(\Omega^{-1}Y'X_{-1}(X_{-1}'X_{-1})^{-1}X_{-1}'Y)$. In the unrestricted case, with $\lambda \neq 0$, this expression consists of the kernels of the posteriors in equations (33)-(34) and, therefore, has a $\chi^2(k^2)$ distribution,

$$
\begin{aligned}
& E_{\Omega}[tr(\Omega^{-1}Y'X_{-1}(X_{-1}'X_{-1})^{-1}X_{-1}'Y)] && (43)\\
= \;& E_{\Omega}[tr(\Omega^{-1}(\Pi - \hat{\Pi})'X_{-1}'X_{-1}(\Pi - \hat{\Pi})]\\
= \;& E_{\Omega}[tr(\Omega^{-1}(\alpha - \hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha - \hat{\alpha}))\\
& +tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}')^{-1}\\
& (\lambda - \hat{\lambda})'X_{2,-1}'X_{2,-1}(\lambda - \hat{\lambda}))\\
& +tr(X_{2,-1}'X_{2,-1}(\beta_2 - \hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2 - \hat{\beta}_2)'))] \sim \chi^2(k^2).
\end{aligned}
$$

Under the hypothesis of cointegration, $\lambda = 0$, as outlined in section 4, the marginal posteriors can be calculated using a M-H sampler. In that case, $E_{\alpha,\beta_2,\Omega}[tr(\Omega^{-1}Y'X(X'X)^{-1}X'Y)]$, changes to,

$$
\begin{aligned}
& E_{\alpha,\beta_2,\Omega}[tr(\Omega^{-1}Y'X(X'X)^{-1}X'Y)] && (44)\\
= \;& E_{\alpha,\beta_2,\Omega}[tr(\Omega^{-1}(\alpha - \hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha - \hat{\alpha}))\\
& +tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}')^{-1}\hat{\lambda}'X_{2,-1}'X_{2,-1}\hat{\lambda})\\
& +tr(X_{2,-1}'X_{2,-1}(\beta_2 - \hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2 - \hat{\beta}_2)'))]
\end{aligned}
$$

where $\hat{\beta}_2$ is calculated assuming $\lambda = 0$. Since the same reasoning holds for equation (44) as for equation (43), the Bayesian LM statistic for testing for cointegration, $\lambda = 0$, does not correspond with,

$$
E_{\alpha,\beta_2,\Omega}[tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}')^{-1}\hat{\lambda}'X_{2,-1}'X_{2,-1}\hat{\lambda})].
$$

Therefore, we have to apply theorem 4 to construct the Bayesian LM statistic to test for cointegration, $\lambda = 0$,

$$\begin{aligned} t_{BLM}(\lambda &= 0) = E_{\alpha,\beta_2,\Omega}[tr(\Omega^{-1}y'ZC(C'Z'ZC)^{-1}C'Z'y)] - E(\chi^2(r(2k-r))) \\ &= E_{\alpha,\beta_2,\Omega}[tr(\Omega^{-1}y'ZC(C'Z'ZC)^{-1}C'Z'y)] - r(2k-r). \end{aligned} \tag{45}$$

The resulting Bayesian LM cointegration statistic has to be compared with a χ^2 distribution with $(k-r)^2$ degrees of freedom. If it is not plausible that the calculated statistic has been generated by such a distribution, the hypothesis that $\lambda = 0$ is not considered plausible. Typical extensions of the cointegration hypothesis, $\lambda = 0$, towards hypotheses including parameters of deterministic components, for example to test whether deterministic components lie in the cointegration space, can be dealt with in a straightforward way using the Bayesian LM cointegration statistic.

6 Natural Conjugate Priors

6.1 Priors

In theorem 1 and lemma 2, the (implied) priors are derived for the parameters of the ECMs assuming diffuse (Jeffreys') priors for the parameters of the standard linear model. In the linear model, diffuse priors can be seen as the limiting "noninformative" case of so-called natural conjugate priors, see Zellner (1971). This class of priors can also be used as a base to derive (implied) priors for the ECM. This enables us to specify informative priors on the parameters of the VAR, for instance a Minnesota prior, see Doan *et. al.* (1984) and Litterman (1986). These informative priors then imply a specific kind of prior for the parameters of the ECMs.

Theorem 5 *Natural Conjugate Priors for the parameters of the linear model (9).*

$$p_{Lin}(\Omega) \propto |S|^{\frac{1}{2}h}|\Omega|^{-\frac{1}{2}(h+m+1)}\exp[-\tfrac{1}{2}tr(\Omega^{-1}S)], \tag{46}$$

$$p_{Lin}(\Pi|\Omega) \propto |\Omega|^{-\frac{1}{2}k}|A|^{\frac{1}{2}k}\exp[-\tfrac{1}{2}tr(\Omega^{-1}(\Pi-P)'A(\Pi-P))], \tag{47}$$

where the prior parameters consist of $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$; $A_{11}, P_{11} : r \times r$; $A_{12}, P_{12} : r \times (k-r)$; $A_{21}, P_{21} : (k-r) \times r$; A_{22}, P_{22} :

$(k-r)\times(k-r)$, $S: k\times k$, *and* h, *imply the following kind of priors for the parameters of the unrestricted ECM (4),*

$$
\begin{aligned}
p_{unres}(\Omega) &\propto |S|^{\frac{1}{2}h}|\Omega|^{-\frac{1}{2}(h+k+1)}\exp[-\frac{1}{2}tr(\Omega^{-1}S)], && (48)\\
p_{unres}(\alpha|\Omega) &\propto |\Omega|^{-\frac{1}{2}r}|A_{11.2}|^{\frac{1}{2}k}\exp[-\frac{1}{2}tr(\Omega^{-1}(\alpha-\begin{pmatrix} P_{11} & P_{12}\end{pmatrix})' \\
&\quad A_{11.2}(\alpha-\begin{pmatrix} P_{11} & P_{12}\end{pmatrix}))], && (49)\\
p_{unres}(\lambda|\alpha,\Omega) &\propto |\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r}\end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r}\end{pmatrix}'|^{-\frac{1}{2}(k-r)} \\
&\quad |A_{22}|^{\frac{1}{2}(k-r)}\exp[-\frac{1}{2}tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r}\end{pmatrix}\Omega \\
&\quad \begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r}\end{pmatrix}')^{-1}(\lambda-l)'A_{22}(\lambda-l))], && (50)\\
p_{unres}(\beta_2|\lambda,\alpha,\Omega) &\propto |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)}|A_{22}|^{\frac{1}{2}r} && (51)\\
&\quad \exp[-\frac{1}{2}tr(A_{22}(\beta_2-B)\alpha\Omega^{-1}\alpha'(\beta_2-B)')],
\end{aligned}
$$

where $A_{11.2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$, $\begin{pmatrix} G_{21} & G_{22}\end{pmatrix}=\begin{pmatrix} P_{21} & P_{22}\end{pmatrix}-A_{22}^{-1}A_{21}(\begin{pmatrix} \Pi_{11} & \Pi_{12}\end{pmatrix}-\begin{pmatrix} P_{11} & P_{12}\end{pmatrix})$, $l=\begin{pmatrix} G_{21} & G_{22}\end{pmatrix}\begin{pmatrix} -\alpha_1^{-1}\alpha_2 \\ I_{k-r}\end{pmatrix}$, $B=-\begin{pmatrix} G_{21} & G_{22}-\lambda\end{pmatrix}\Omega^{-1}\alpha'(\alpha\Omega^{-1}\alpha')^{-1}$.

Proof : see Appendix.

Again, continuity of the natural conjugate prior in the parameters of the unrestricted ECM implies a prior for the parameters of the ECM, see lemma 2, which is stated in lemma 6.

Lemma 6 *The Natural Conjugate Priors for the parameters of the linear model ([illegible]) from theorem 5 imply the following kind of priors for the parameters of the ECM (4),*

$$
\begin{aligned}
p_{coint}(\alpha,\Omega) &= p_{unres}(\Omega)p_{unres}(\alpha|\Omega)p_{unres}(\lambda=0|\alpha,\Omega) && (52)\\
p_{coint}(\beta_2|\alpha,\Omega) &= p_{unres}(\beta_2|\lambda=0,\alpha,\Omega) && (53)
\end{aligned}
$$

where the p_{unres}*'s are defined in theorem 5.*

Proof: this results directly from the continuity of the prior of the parameters of the unrestricted ECM in the parameter points where cointegration occurs, $\lambda=0$.

6.3 Posteriors

In a similar way as in theorem 3, it is possible to construct the posterior of the parameters of the unrestricted ECM, using the natural conjugate priors from theorem 5. The resulting posteriors are stated in theorem 7.

Theorem 7 *The Natural Conjugate Priors on the parameters of the linear model (4), as specified in theorem 5, lead to the following expressions for the conditional posteriors of the parameters of the unrestricted ECM (4),*

$$
\begin{aligned}
& p_{unres}(\Omega|Y) && (54)\\
\propto\; & |S + P'NP + \Delta Y'\Delta Y - \hat{\Pi}'(N + Y'_{-1}Y_{-1})\hat{\Pi}|^{\frac{1}{2}(T+k)} \\
& |\Omega|^{-\frac{1}{2}(T+k+m+1)} \exp[-\tfrac{1}{2}tr(\Omega^{-1}(S + P'NP + \Delta Y'\Delta Y \\
& -\hat{\Pi}'(N + Y'_{-1}Y_{-1})\hat{\Pi}))], \\
& p_{unres}(\alpha|\Omega, Y) && (55)\\
\propto\; & |\Omega|^{-\frac{1}{2}m}|(N + Y'_{-1}Y_{-1})_{11.2}|^{\frac{1}{2}k} \exp[-\tfrac{1}{2}tr(\Omega^{-1} \\
& (\alpha - \begin{pmatrix} \hat{\Pi}_{11} & \hat{\Pi}_{12} \end{pmatrix})'(N + Y'_{-1}Y_{-1})_{11.2}(\alpha - \begin{pmatrix} \hat{\Pi}_{11} & \hat{\Pi}_{12} \end{pmatrix}))], \\
& p_{unres}(\lambda|\alpha, \Omega, Y) && (56)\\
\propto\; & |\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-m} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-m} \end{pmatrix}'|^{-\frac{1}{2}(k-m)} \\
& |(N + Y'_{-1}Y_{-1})_{22}|^{\frac{1}{2}(k-m)} \\
& \exp[-\tfrac{1}{2}tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-m} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-m} \end{pmatrix}')^{-1} \\
& (\lambda - \hat{\lambda})'(N + Y'_{-1}Y_{-1})_{22}(\lambda - \hat{\lambda}))], \\
& p_{unres}(\beta_2|\lambda, \alpha, \Omega, Y) && (57)\\
\propto\; & |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-m)}|(N + Y'_{-1}Y_{-1})_{22}|^{\frac{1}{2}m} \\
& \exp[-\tfrac{1}{2}tr(N_{22}(\beta_2 - \hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2 - \hat{\beta}_2)')].
\end{aligned}
$$

where $\hat{\Pi} = (N+Y'_{-1}Y_{-1})^{-1}(NP+Y'_{-1}\Delta Y_{-1}\hat{\Pi}) = \begin{pmatrix} \hat{\Pi}_{11} & \hat{\Pi}_{12} \\ \hat{\Pi}_{21} & \hat{\Pi}_{22} \end{pmatrix}$, $(N+Y'_{-1}Y_{-1}) = \begin{pmatrix} (N + Y'_{-1}Y_{-1})_{11} & (N + Y'_{-1}Y_{-1})_{12} \\ (N + Y'_{-1}Y_{-1})_{21} & (N + Y'_{-1}Y_{-1})_{22} \end{pmatrix}$, $\begin{pmatrix} \Phi_{21} & \Phi_{22} \end{pmatrix} = (\begin{pmatrix} \hat{\Pi}_{21} & \hat{\Pi}_{22} \end{pmatrix} - (\alpha-(N+Y'_{-1}Y_{-1})^{-1}_{22}(N+Y'_{-1}Y_{-1})_{21}\begin{pmatrix} \hat{\Pi}_{11} & \hat{\Pi}_{12} \end{pmatrix}))$, $\hat{\lambda} = \begin{pmatrix} \Phi_{21} & \Phi_{22} \end{pmatrix}\begin{pmatrix} -\alpha_1^{-1}\alpha_2 \\ I_{k-m} \end{pmatrix}$, $\hat{\beta}_2 = -\begin{pmatrix} \Phi_{21} & \Phi_{22} - \lambda \end{pmatrix}\Omega^{-1}\alpha'(\alpha\Omega^{-1}\alpha')^{-1}$.

Proof : see Appendix.

Of course when all prior parameters are equal to zero, the posteriors from theorem 5 equal the posteriors from theorem 3. The joint posterior of (α, β, β_2) in the ECM is again equal to the conditional posterior of (α, β, β_2) in the unrestricted ECM given that $\lambda = 0$,

$$p_{ecm}(\alpha, \beta|Y) \propto p_{uecm}(\beta|Y)p_{uecm}(\alpha|\beta, Y)p_{uecm}(\lambda = 0|\alpha, \beta, Y), \quad (58)$$

$$p_{ecm}(\beta_2|\alpha, \beta, Y) \propto p_{uecm}(\beta_2|\alpha, \lambda = 0, \beta, Y). \quad (59)$$

Similar to the posterior simulation from section 4, a M-H sampler can again be used to sample from this posterior. This sampler generates β and α from $p_{uecm}(\beta|Y)$, $p_{uecm}(\alpha|\beta, Y)$ respectively and uses a weight function proportional to the conditional posterior of λ in the unrestricted ECM evaluated in $\lambda = 0$,

$$w(\alpha, \beta) = p_{uecm}(\lambda = 0|\alpha, \beta, Y). \quad (60)$$

For more details about simulation we refer to section 4.

7 Prior and Posterior Odds Ratios

The in the previous sections developed, procedures for calculating the posteriors of the parameters of the ECM for different numbers of cointegrating vectors r, allow us to construct Prior (PROR) and Posterior Odds Ratios (POR) to compare models with different numbers of cointegrating vectors. As the number of cointegrating vectors can only take $k+1$ discrete values, $r = 0, \ldots, k$, we can calculate the prior, posterior probabilities of the number of cointegrating vectors (r) and implied unit roots ($k - r$).

Initially we only construct a PROR and POR to compare a model with r cointegrating vectors with a model with k cointegrating vectors, the unrestricted ECM. Ratios of these PRORs and PORs then form the PRORs and PORs for comparing the models mutually.

7.1 Proper Priors

POR and PRORs are only defined in case of proper priors. We use the priors from theorem 5 and lemma 6 in the construction of the PRORs and PORs. In the next subsection the limiting case of a natural conjugate prior, a diffuse (implied) prior as used in theorem 1 and lemma 2, is discussed.

The implied prior in lemma 6 is needed in the construction of PRORs,

$$PROR(j, k) = \frac{\Pr(r = j)}{\Pr(r = k)} \quad (61)$$

$$= \frac{p_q \int\int\int p_{exact}(\alpha, \beta_2, \Sigma)\, d\beta_2\, d\alpha\, d\Sigma}{p_k \int\int\int\int p_{approx}(\alpha, \beta_2, \lambda, \Sigma)\, d\beta_2\, d\alpha\, d\Sigma}$$

$$= \frac{p_q}{p_k} \frac{1}{(2\pi)^{\frac{1}{2}(k-r)^2}} \frac{\int\int p_{exact}(\alpha, \Sigma)\, d\alpha\, d\Sigma}{\int\int p_{approx}(\alpha, \Sigma)\, d\alpha\, d\Sigma},$$

where $\Pr(r = q)$ stands for the prior probability that a model has a number of cointegrating vectors, r, equal to q and p_q are prior weights which reflect a prior opinion about the possible number of cointegrating vectors (note that this is also partly incorporated in the specification of the natural conjugate prior). The priors exactly equal the functional expressions from lemma 3 and theorem 2 and only contain the kernels of the priors. Note that the priors are not always proper but the weights p_q are such that the sum of the prior probabilities is equal to one.

We can now define the PBF to compare a model with q cointegrating vectors with a model with k cointegrating vectors,

$$\begin{aligned} PBF(q, k) &= \frac{\Pr(Y|r = q)}{\Pr(Y|r = k)} \\ &= \frac{p_q}{p_k} \frac{1}{(2\pi)^{\frac{1}{2}(k-q)^2}} \frac{\int\int p_{exact}(\alpha, \Sigma|Y)\, d\alpha\, d\Sigma}{\int\int p_{approx}(\alpha, \Sigma|Y)\, d\alpha\, d\Sigma}, \end{aligned} \qquad (62)$$

where $\Pr(Y|r = q)$ stands for the posterior probability that a model with a number of cointegrating vectors, r, equal to q, generated the observed series Y. The parameters β_2 and λ are integrated out analytically and we use the kernels of the conditional posteriors $p_{exact}(\alpha, \Sigma|Y)$, (58), and $p_{approx}(\alpha, \Sigma|Y)$ $(= p_{approx}(\Sigma|Y)\ p_{approx}(\alpha|\Sigma, Y))$, (54) and (55). Note that the kernels in (62) exactly match these functional expressions, such that we do not incorporate all elements of the normalizing constants. The conditional posterior $p_{exact}(\alpha, \Sigma|Y)$ does not belong to a known class of probability density functions and an analytical solution to its integral is not known. We can calculate the ratio of integrals of the conditional posterior efficiently by simulating α and Σ from $p_{approx}(\alpha, \Sigma|Y)$, which is a product of an inverted-Wishart for Σ and a matrix normal for α given Σ. In the generated parameter points we can then calculate the ratio of the two posteriors of α and Σ,

$$w(\alpha^i, \Sigma^i) = \frac{p_{exact}(\alpha^i, \Sigma^i|Y)}{p_{approx}(\alpha^i, \Sigma^i|Y)} = p_{approx}(\lambda = 0|\alpha^i, \Sigma^i, Y), \qquad (63)$$

where i stands for the $i-$th drawing of (α, Σ), see also (56). The average of the generated $w(\alpha^i, \Sigma^i)$ then converges to the ratio of the integrals in (62), see Geweke (1989a,b),

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} w(\alpha^i, \Sigma^i) - \frac{\int\int p_{exact}(\alpha, \Sigma|Y)\, d\alpha\, d\Sigma}{\int\int p_{approx}(\alpha, \Sigma|Y)\, d\alpha\, d\Sigma}\right) \Rightarrow n(0, s), \qquad (64)$$

where n is the number of drawings, $s = var(w(\alpha, \beta))$, $\frac{1}{n}\sum_{i=1}^{n} w(\alpha^i, \beta^i)^2 - (\frac{1}{n}\sum_{i=1}^{n} w(\alpha^i, \beta^i))^2 \Rightarrow s$, and $\Rightarrow$ stands for weak convergence. The PBF (52) can be calculated as follows,

$$PBF(r, k) = \frac{g_r}{g_k}\frac{1}{(2\pi)^{\frac{1}{2}(k-r)^2}}(\frac{1}{n}\sum_{i=1}^{n} w(\alpha^i, \beta^i)) \tag{55}$$

Since the sum of the posterior probabilities is equal to 1,

$$\sum_{r=0}^{k} \Pr(H|r = r) = 1, \tag{56}$$

the posterior probability, that a model with a specific number of cointegrating vectors generated the observed series, equal,

$$\begin{aligned} \Pr(H|r &= r) = \frac{PBF(r, k)}{1 + \sum\limits_{i=0}^{k-1} PBF(i, k)}, \qquad r = 0, .., k-1, \\ \Pr(H|r &= k) = \frac{1}{1 + \sum\limits_{i=0}^{k-1} PBF(i, k)}. \end{aligned} \tag{57}$$

The PBFs can be used by themselves to reflect the support for the different numbers of cointegrating vectors but they can also be compared with the PPBFs to determine up to what extent the posteriors lead to other conclusions than the priors.

Identical to the PBF (52), the PPBF (51) can be calculated by simulating from $g_{post}(\alpha, \beta)$ and attaching a weight to each drawing proportional to the ratio of the priors of (α, β),

$$\frac{g_{prior}(\alpha, \beta)}{g_{post}(\alpha, \beta)} = g_{post}(\lambda = 0|\alpha, \beta). \tag{58}$$

The PPBF (51) can then be approximated by $\frac{g_r}{g_k}(2\pi)^{-\frac{1}{2}(k-r)^2}$ times the average value of the weights, see (54) and (55). The ratio of the PBF and PPBF, i.e. the Bayes factor, can then be used to see up to what extent the data leads to other conclusions than the prior. Using the PPBFs and formulas identical to (57) it is also possible to calculate the prior probabilities.

7.3 Diffuse Priors

In the limiting case where all prior parameters are equal to zero, the results for the diffuse prior case are obtained. The PBF is straightforward to calculate for that case as all posteriors are listed in theorem 3. It is not directly obvious, however, what a PDSS means in case of diffuse priors. By letting the prior parameters converge to zero, the value of the PDSS can be obtained. This value is stated in theorem 8.

Theorem 8 *When all prior parameters in the Natural Conjugate Priors from theorem 3, are equal to zero, then,*

$$PDSS(r,k) = (2\pi)^{-\frac{1}{2}(k-r)^2}, \qquad r = 0,...,k-1, \tag{69}$$

where the PDSS is defined in (68) and $p_r = p_k$, $r = 0,...,k-1$.

Proof: see Appendix.

In case of a diffuse prior, the ratio between the PDS and PDSS shows the support for a specific model given by the data as we correct for any latent prior information by dividing the PDS by the PDSS. We call this ratio of PDS and PDSS, the Bayes Factor (BF), see Zellner (1971),

$$BF(r,k) = \frac{PDS(r,k)}{PDSS(r,k)} = \frac{\int\int p_{ecm}(\alpha,\beta|\Sigma)d\alpha d\beta}{\int\int p_{necm}(\alpha,\beta|\Sigma)d\alpha d\beta}, \tag{70}$$

where $p_{ecm}(\alpha,\beta|\Sigma)$ results from (35) and $p_{necm}(\alpha,\beta|\Sigma)$ from (33) and (31) and no further normalizing constants are included. The BFs can directly be calculated using the average of the simulated weights (36). As further discussed in a later subsection, the BF is closely related to the Posterior Information Criterium of Phillips and Ploberger (1994,1996). Additionaly the BF also allows for the calculation of posterior probabilities like (67).

The applicability of the derived PDSs and posterior probabilities is shown in a later section where we use these methods to compare models with different number of cointegrating vectors for both simulated and real data series.

5 Relationships with Existing Procedures

When the functional expression of the Bayesian LM statistic and/or the PDSs are evaluated in specific parameter points, relationships with other (classical) procedures can be found. Some of these relationships are further investigated in the next subsections.

5.1 LMR cointegration statistic

When the functional expression of the Bayesian LM cointegration statistic (44) is evaluated in the parameters points, $\hat{\beta}$, $\hat{\alpha}$ and the resulting implied

$\hat{\beta}_2$ ($\lambda = 0$), see theorem 3 for expressions of these estimators, it is identical to the Generalized Method of Moments (GMM) cointegration statistic derived in Kleibergen (1998). In a classical statistical analysis, the, in this way, constructed estimator of the cointegrating vectors, $\hat{\beta}_2$, has a so-called mixed normal limiting distribution, see Phillips (1991). Its limiting distribution is also identical to the limiting distribution of the cointegrating vector estimator in the Johansen framework, see Johansen (1991). Furthermore, the limiting distribution of the GMM cointegration statistic is also identical to the limiting distribution of the Johansen cointegration likelihood ratio statistic. The GMM cointegration statistic is, therefore, closely related to the Johansen likelihood ratio statistic and in practice these statistics have similar values. As the parameter points, in which the Bayesian LR cointegration statistic has to be evaluated to obtain the GMM cointegration statistic, are the means of the approximating density from which we simulate in the M-H sampler, we expect the Bayesian LR cointegration statistic to have values which are similar to the values of the Johansen cointegration statistic. The interpretation of the values of the two statistics is entirely different, however. A Bayesian assumes the data as fixed and given, which leads to standard kind of distributions, while a classical analyzes the data as one realization of the data generating process, which in this case leads to limiting distributions of the statistic consisting of Brownian Motion functionals.

8.3 Posterior Information Criterion (PIC)

The BF (70) is closely related to the Posterior Information Criterion (PIC) of Phillips and Ploberger (1994,1996), see also Phillips (1996). When the BF (70) is evaluated in the Maximum Likelihood (ML) parameter points $\hat{a}$ $(= (Y_{1,-1}'M_{Y_{2,-1}}Y_{1,-1})^{-1}Y_{1,-1}'M_{Y_{2,-1}}\Delta Y)$ and $\hat{\Omega}$ $(= \frac{1}{T}\Delta Y'M_{Y_{-1}}\Delta Y)$, (Note that also $\hat{\lambda}$ depends on a, see theorem 3), twice its logarithm would equal the difference between a PIC of the ECM and a PIC of the unrestricted ECM,

$$
\begin{aligned}
& 2\log(BF(r,k)) = PIC(r=r) - PIC(r=k) \qquad (71)\\
= \; & (k-r)[\log(|Y_{2,-1}'Y_{2,-1}|) - \\
& \log(|\begin{pmatrix} -\hat{a}_2'\hat{a}_1^{-1\prime} & I_{k-r} \end{pmatrix}\hat{\Omega}\begin{pmatrix} -\hat{a}_2'\hat{a}_1^{-1\prime} & I_{k-r} \end{pmatrix}'|)] \\
& -tr[(\begin{pmatrix} -\hat{a}_2'\hat{a}_1^{-1\prime} & I_{k-r} \end{pmatrix}\hat{\Omega}\begin{pmatrix} -\hat{a}_2'\hat{a}_1^{-1\prime} & I_{k-r} \end{pmatrix}')^{-1}\hat{\lambda}'Y_{2,-1}'Y_{2,-1}\hat{\lambda}].
\end{aligned}
$$

As its denominator the BF, used for the construction of the PIC (71), has the joint posterior of the parameters in the unrestricted ECM. The BF based PIC of this unrestricted ECM, therefore, equals twice the log of the value of the joint posterior of the parameters in the ML parameter point, see also theorem 3,

$$\begin{aligned}
PIC(r &= k) = \log\left(\left| \begin{pmatrix} I_k \otimes \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}' \\ \begin{pmatrix} \hat{a} & \\ 0 & I_{k-r} \end{pmatrix} \otimes \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}' \end{pmatrix} (\hat{\Omega}^{-1} \otimes Z_{-1}'Z_{-1}) \right.\right. \\
&\left.\left. \begin{pmatrix} I_k \otimes \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}' \\ \begin{pmatrix} \hat{a} & \\ 0 & I_{k-r} \end{pmatrix} \otimes \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}' \end{pmatrix}' \right|\right) - T\log|\hat{\Omega}| \qquad (72) \\
&= [k\log(|Z_{1,-1}'M_{Z_{2,-1}}Z_{1,-1}|) - r\log(|\hat{\Omega}|)] + [r\log(|Z_{2,-1}'Z_{2,-1}|) \\
&+ (k-r)\log(|\hat{a}\hat{\Omega}^{-1}\hat{a}'|)] + (k-r)[\log(|Z_{2,-1}'Z_{2,-1}|) \\
&- \log(|\begin{pmatrix} -\hat{a}_2'\hat{a}_1^{-1\prime} & I_{k-r} \end{pmatrix} \hat{\Omega} \begin{pmatrix} -\hat{a}_2'\hat{a}_1^{-1\prime} & I_{k-r} \end{pmatrix}'|)] - T\log|\hat{\Omega}|,
\end{aligned}$$

where all parameters are evaluated in the ML parameter points and the jacobian of the transformation of the linear model to the unrestricted ECM is incorporated. Note that $PIC(r = k)$, in this setting, depends on the (number of cointegrating vectors of the) ECM with which the unrestricted model is compared. This leads to the PIC of the ECM,

$$\begin{aligned}
PIC(r &= r) = [k\log(|Z_{1,-1}'M_{Z_{2,-1}}Z_{1,-1}|) - r\log(|\hat{\Omega}|)] - T\log(|\hat{\Omega}|) + \\
&[r\log(|Z_{2,-1}'Z_{2,-1}|) + (k-r)\log(|\hat{a}\hat{\Omega}^{-1}\hat{a}'|)] - \qquad (73) \\
&tr[(\begin{pmatrix} -\hat{a}_2'\hat{a}_1^{-1\prime} & I_{k-r} \end{pmatrix} \hat{\Omega} \begin{pmatrix} -\hat{a}_2'\hat{a}_1^{-1\prime} & I_{k-r} \end{pmatrix}')^{-1}\hat{\beta}'Z_{2,-1}'Z_{2,-1}\hat{\beta}].
\end{aligned}$$

In Phillips (1996), the PIC of the ECM reads,

$$\begin{aligned}
PIC(r &= r) = \log\left| \begin{pmatrix} \frac{\partial vec(\Pi)}{\partial vec(\alpha)'} \\ \frac{\partial vec(\Pi)}{\partial vec(\beta_2)'} \end{pmatrix} (\tilde{\Omega}^{-1} \otimes Z_{-1}'Z_{-1}) \begin{pmatrix} \frac{\partial vec(\Pi)}{\partial vec(\alpha)'} \\ \frac{\partial vec(\Pi)}{\partial vec(\beta_2)'} \end{pmatrix}' \right| - T\log|\tilde{\Omega}| \\
&= \log\left| \begin{pmatrix} I_k \otimes \tilde{\beta}' \\ \tilde{\alpha} \otimes \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}' \end{pmatrix} (\tilde{\Omega}^{-1} \otimes Z_{-1}'Z_{-1}) \begin{pmatrix} I_k \otimes \tilde{\beta}' \\ \tilde{\alpha} \otimes \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}' \end{pmatrix}' \right| \\
&\quad - T\log|\tilde{\Omega}| \\
&= -T\log|\tilde{\Omega}| + (k-r)\log(\tilde{\alpha}\tilde{\Omega}^{-1}\tilde{\alpha}') + r\log|Z_{2,-1}'Z_{2,-1}| \\
&\quad + \log|(\tilde{\Omega}^{-1} \otimes \tilde{\beta}'Z_{-1}'Z_{-1}\tilde{\beta}) - \\
&\quad (\tilde{\Omega}^{-1}\tilde{\alpha}'(\tilde{\alpha}\tilde{\Omega}^{-1}\tilde{\alpha}')^{-1}\tilde{\alpha}\tilde{\Omega}^{-1} \otimes \tilde{\beta}'Z_{-1}'Z_{2,-1}(Z_{2,-1}'Z_{2,-1})^{-1}Z_{2,-1}'Z_{-1}\tilde{\beta})|
\end{aligned}$$

$$\approx \quad [(k-r)\log(\hat{\alpha}\hat{\Omega}^{-1}\hat{\alpha}') + r\log|M'_{2,-1}M_{2,-1}|] +$$
$$+[k\log|\hat{\beta}'M'_{-1}M_{-1}\hat{\beta}| - r\log|\hat{\Omega}|] - T\log|\hat{\Omega}| \qquad (74)$$

where $\Pi = \beta\alpha$, $\beta' = (I_r \; -\beta_2')$, $\hat{\alpha}, \hat{\beta}$ and $\hat{\Omega}$ are the ML estimators of α, β and Ω, and we use $\approx$ as we replaced the information matrix by its limiting expression (essentially this only holds for the "true" number of cointegrating vectors). Asymptotically the PIs (73) and (74) are equal for the "true" number of cointegrating vectors, as for that case,

$$\frac{1}{T}M'_{-1}P_{\hat{\beta}_{\perp},-1}M_{-1} \approx \frac{1}{T}\hat{\beta}'M'_{-1}M_{-1}\hat{\beta}, \qquad (75)$$

for a proof see Kleibergen and van Dijk (1994a), and (see also proof of theorem 3),

$$\tilde{\Omega} \approx \hat{\Omega} + \frac{1}{T}(\hat{\beta}_2\hat{\alpha} - \hat{\Pi}_2)'M'_{2,-1}M_{2,-1}(\hat{\beta}_2\hat{\alpha} - \hat{\Pi}_2), \qquad (76)$$

where the estimators are defined in theorem 3 and $d = 0$ in the expression of $\hat{\beta}_2$, such that using a first order Taylor expansion of $\log|\tilde{\Omega}|$ around $\log|\hat{\Omega}|$ and the proof of theorem 3, it follows that,

$$\log(|\tilde{\Omega}|) \quad \approx \quad \log(|\hat{\Omega}|) + \qquad (77)$$
$$\frac{1}{T}tr[((-\hat{\alpha}_2'\hat{\alpha}_1^{-1\prime} \;\; I_{k-r})\hat{\Omega}(-\hat{\alpha}_2'\hat{\alpha}_1^{-1\prime} \;\; I_{k-r})')^{-1}\hat{d}'M'_{2,-1}M_{2,-1}\hat{d}],$$

where $\approx$ implies that the limiting expressions are equal. This shows that the PIs are asymptotically equal to one another. The PIs in the previous expressions only evaluate the joint posteriors in the parameter point of the ML estimator. When the posterior is not well behaved, which is not unlikely as the posterior of the ECM is not analytically tractable, probability statements, as for example the PI, which are not based on the whole posterior can be misleading. We, therefore, prefer to use the PI (71) where the BFs are obtained by integrating out the parameters. When the posteriors are well behaved, these PIs will be similar to the PIs (73) and (74) but these PIs can be bad approximations if the posterior has a lot of probability mass away from the ML parameter point.

The PIs obtained from the BFs, by integrating out the parameters, are a natural Bayesian multivariate generalization of the PIs in Phillips and Ploberger (1994,1996). The limiting results of the PI derived by Phillips and Ploberger (1996), therefore, generalize to these kind of PIs.

8 Illustrative Examples

To illustrate the applicability of the, in the previous sections, constructed methods and procedures for Bayesian cointegration analyses, we analyze sevreal simulated series and the Danish data from Johansen and Juselius (1990).

9.1 Simulated Series

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

$$\text{DGP 1} \;:\; \Delta y_t = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \varepsilon_t \tag{78}$$

$$\text{DGP 2} \;:\; \Delta y_t = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.1 \\ 0.1 \\ 0.1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix} y_{t-1} + \varepsilon_t \tag{79}$$

$$\text{DGP 3} \;:\; \Delta y_t = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.1 & -0.1 \\ 0.1 & -0.1 \\ 0.1 & 0.1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} y_{t-1} + \varepsilon_t \tag{80}$$

$$\text{DGP 4} \;:\; \Delta y_t = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.1 & -0.1 & -0.1 \\ 0.1 & -0.1 & -0.1 \\ 0.1 & 0.1 & -0.1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} y_{t-1} + \varepsilon_t \tag{81}$$

with $\varepsilon_t \sim N(0, I)$. The sample size is 100 observations. The DGPs correspond to 0, 1, 2 and 3 cointegration relations, respectively. The DGP 1 contains three unit roots, the DGP 2 contains 2 unit roots and a root 0.8, the DGP 3 contains the roots 1, 0.8 and 0.8, and the DGP 4 contains the roots 0.9, 0.8 and 0.8.

To analyze the simulated series, we consider a VAR(1) model, which corresponds with the lag order in the DGP. The first step in the Bayesian analysis is to specify a prior on the vector autoregressive parameters Φ and on the covariance matrix Ω. We use a diffuse (Jeffreys') prior on these parameters such that the priors on the parameters from the unrestricted ECM result from theorem 1. We also use equal prior weights, p_r, see (51), for models with different number of cointegrating vectors. Given the priors and prior weights, we can compare a model with reduced rank, the RRECM (7), with the full rank unrestricted ECM (8).

The first column of Table 1 displays the BFs (70) for the four DGPs. A BF exceeding 1 indicates that rank r is preferred above the full rank situation. For DGP 1 every rank reduction is preferred, while for DGP 4 the full rank situation is always preferred. The BFs can be translated into posterior probabilities for the cointegration ranks, see (57). These are displayed in the second column of Table 1. These probabilities put 70% or more weight on the right cointegration rank.

Table 1: Bayes Factors, Posterior Probabilities and Bayesian LR statistics for the four DGPs and the Danish data.

r	$BF(r,4)$	$\Pr[M\|r]$	$t_{trace}(r\|4)$	df	p-value	$LR(r\|3)$[1]
			DGP 0			
0	973.33	0.99	14.55	9	0.10	15.08
1	5.89	0.01	6.90	4	0.14	5.85
2	2.20	0.00	1.16	1	0.28	1.19
3	1.00	0.00	-	-	-	-
			DGP 1			
0	0.01	0.00	30.10	9	0.00	33.73*
1	15.73	0.87	5.73	4	0.22	5.96
2	1.47	0.08	1.78	1	0.18	1.19
3	1.00	0.05	-	-	-	-
			DGP 2			
0	0.00	0.00	39.56	9	0.00	44.33*
1	0.02	0.00	16.44	4	0.00	17.58*
2	3.83	0.79	0.02	1	0.89	0.12
3	1.00	0.21	-	0	-	-
			DGP 3			
0	0.00	0.00	46.93	9	0.00	52.12*
1	0.00	0.00	24.26	4	0.00	25.59*
2	0.08	0.07	6.25	1	0.01	6.43*
3	1.00	0.93	-	0	-	-
			Danish data			
0	0.11	0.00	47.77	20	0.00	49.14+
1	41.37	0.30	24.32	12	0.02	19.06
2	85.62	0.62	9.73	6	0.14	8.69
3	11.21	0.08	2.60	2	0.27	2.35
4	1.00	0.00	-	0	-	-

[1] Johansen trace test, * and + denote significant at 5% and 10% respectively.

The third column of Table 1 contains the Bayesian LM statistics. This statistic tests the null hypothesis of r cointegration vectors against the full rank situation, i.e. 3 cointegration relations, see subsection 5.3. These LM statistics have to be compared with a χ^2 distribution with $(3-r)^2$ degrees of freedom which number is shown in the fourth column. The fifth column of Table 1 shows the p-values for the calculated statistics. For instance, for DGP 4 none of the models with reduced rank is plausible, while for DGP 3 only a model with two cointegration relations is plausible. In general, the Bayesian LM test results point to the right cointegration rank. In the last column we report the results of the Johansen trace tests denoted by $LR(r|3)$. Notice that the value of this statistic is roughly the same as the computed Bayesian LM tests.

9.2 Empirical Data

Johansen and Juselius (1990) analyze the demand function for money $m_t = f(y_t, p_t, i_t)$ for the Danish economy. Money m_t is a function f of y_t real income, p_t price level and i_t the costs of holding money. The costs of holding money can be approximated by a difference between the bank deposit rate i_t^b for interest bearing deposits and the bond rate i_t^d, as M2 is chosen as a proxy for money demand. 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 analyze the same Danish data as in Johansen and Juselius (1990). We have quarterly observed series of m_t, i_t^b, i_t^d and y_t for the period 1974.1–1987.3. The cointegration analysis is performed in the following VAR(2) model,

$$\begin{pmatrix} \Delta m_t \\ \Delta i_t^b \\ \Delta i_t^d \\ \Delta y_t \end{pmatrix} = \sum_{s=1}^{3} \phi_s D_{st} + \Pi \begin{pmatrix} m_{t-1} \\ i_{t-1}^b \\ i_{t-1}^d \\ y_{t-1} \\ 1 \end{pmatrix} + \Gamma_1 \begin{pmatrix} \Delta m_{t-1} \\ \Delta i_{t-1}^b \\ \Delta i_{t-1}^d \\ \Delta y_{t-1} \end{pmatrix} + \varepsilon_t, \qquad (52)$$

where D_{st} represents seasonal dummies with zero mean and ϕ_s is a 4-dimensional parameter vector, $s = 1, \ldots, 3$. Notice that the constant is restricted in the cointegration space. This implies that we have to extend the Π matrix and the Z_{-1} matrix with an extra column.

Table 1 displays the results of a Bayesian cointegration analysis. The results are based on a diffuse (Jeffrey's) prior for the parameters in (52). The first two columns show the BF and the therethrough implied posterior probabilities over the cointegration rank. The BF favor a model with 1, 2 and 3 cointegrating relationships over a full rank model. The posterior probabilities assign about 30% probability to rank 1 and 60% to rank 2 and about 10% to rank 3. The Johansen cointegration trace statistics, which are given in the last column of table 1, indicate one cointegration relation if we test at roughly 10% level of significance.

The remaining columns of table 1 display the results of the Bayesian LM test. Since only the $t_{BALM}(2|4)$ and $t_{BALM}(3|4)$ are inside the 95% HPD interval, the tests indicate two cointegration relations between m_t, i_t^s, i_t^l and y_t. Note that the degrees of freedom are different from $(4-r)^2$ due to the restricted constant. The Bayesian LM tests match the Johansen trace statistics quite well, but indicate only one cointegration relation as in the classical approach the asymptotic distribution is not χ^2 but a functional of Brownian motions.

In summary, although the examples in this section are simple, they show that Bayesian techniques provide useful tools to analyze cointegration. BFs and Bayesian LM tests indicate whether rank reduction is plausible. The former can be used to calculate posterior probabilities for each cointegration rank. Instead of choosing the rank r one can use these probabilities as weights in further analysis, for instance in a forecasting exercise.

10 Conclusions

The paper discusses a Bayesian modelling framework for the analysis of cointegration models. This framework is based on a specification of a unrestricted Error Correction Model which contains a parameter reflecting cointegration, i.e. it is equal to zero when cointegration occurs. Posteriors for parameters in the cointegration model are then proportional to conditional posteriors of the parameters in the Error Correction Model given that the parameter reflecting cointegration is equal to zero. This is identical to the classical analysis where the likelihood of the cointegration model is proportional to the conditional likelihood of the unrestricted Error Correction Model given that the parameter reflecting cointegration is equal to zero. A Metropolis-Hastings sampler is used to calculate the posteriors of the cointegration model. We compare different cointegration models using either a formal testing procedure, Bayesian Lagrange Multiplier testing, or prior and posterior probabilities. The, in the probabilities, involved Bayes factors are related to the posterior information criterium of Phillips and Ploberger (1994,1996). The resulting framework allows for a full Bayesian treatment of all aspects of a cointegration model which gives us the possibility to specify an informative prior. This prior is specified on the parameters of the VAR and therethrough implies the priors for the cointegration models. Therefore, one specification of the prior for the parameters of the VAR only suffices as it implies the functional specification of the priors for the cointegration models. Different prior weights can, however, be given to the cointegration models.

In further research we will extend the framework for Bayesian cointegration analysis to allow for structural breaks and/or MA errors. As discussed in Kleibergen and Hoek (1996), posteriors of the parameters in univariate ARMA models can be calculated using a Metropolis-Hastings sampler. Combining the Metropolis-Hastings sampler used in that paper and the sampler in this paper can lead to a sampler to calculate the posteriors of the parameters of Vector AutoRegressive Moving Average cointegration models. Structural breaks can be incorporated by using the Metropolis-Hastings sampler in a Gibbs sampling environment where one draws the cointegration parameters given the drawn breakpoint and the breakpoint given the cointegration parameters.

Appendix.

Proof of theorem 1.

The Jeffreys' priors of $(\Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22})$ given Ω are proportional to the square root of the determinant of the information matrix. The information matrix of $(\Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22})$ given Ω reads,

$$\begin{aligned}
I(\Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22}|\Omega) &= \begin{pmatrix} \Omega^{-1} \otimes X'_{1,-1}X_{1,-1} & \Omega^{-1} \otimes X'_{1,-1}X_{2,-1} \\ \Omega^{-1} \otimes X'_{2,-1}X_{1,-1} & \Omega^{-1} \otimes X'_{2,-1}X_{2,-1} \end{pmatrix}, \\
|I(\Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22}|\Omega)| &= |\Omega|^{k}|X'_{1,-1}M_{X_{2,-1}}X_{1,-1}|^{k}|X'_{2,-1}X_{2,-1}|^{k}
\end{aligned}$$

which gives the conditional priors for $(\Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22})$ given Ω,

$$\begin{aligned}
p_{Jef}(\Pi_{11}, \Pi_{21}|\Omega) &\propto |\Omega|^{-\frac{1}{2}m}|X'_{1,-1}M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k} \\
&\propto |(\Omega^{-1} \otimes X'_{1,-1}M_{X_{2,-1}}X_{1,-1})|^{\frac{1}{2}} \\
&\propto |I(\Pi_{11}, \Pi_{12}|\Omega)|^{\frac{1}{2}}, \\
p_{Jef}(\Pi_{21}, \Pi_{22}|\Pi_{11}, \Pi_{21}, \Omega) &\propto |\Omega|^{-\frac{1}{2}(k-m)}|X'_{2,-1}X_{2,-1}|^{\frac{1}{2}k} \\
&\propto |(\Omega^{-1} \otimes X'_{2,-1}X_{2,-1})|^{\frac{1}{2}} \\
&\propto |I(\Pi_{21}, \Pi_{22}|\Pi_{11}, \Pi_{12}, \Omega)|^{\frac{1}{2}},
\end{aligned}$$

where $I(\Pi_{11}, \Pi_{12}|\Omega) = (\Omega^{-1} \otimes X'_{1,-1}M_{X_{2,-1}}X_{1,-1})$, $I(\Pi_{21}, \Pi_{22}|\Pi_{11}, \Pi_{12}, \Omega) = (\Omega^{-1} \otimes X'_{2,-1}X_{2,-1})$. The priors for the parameters a, B_2 and λ can now be constructed using the Jacobian of the transformations of $(\Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22})$ to (a, B_2, λ). The ordering is here again important as (Π_{21}, Π_{22}) can only be transformed to (B_2, λ) when a is known.

$$\begin{aligned}
(\Pi_{21}\ \Pi_{22}) &= \begin{pmatrix} I & \lambda \end{pmatrix} - B_2 a \Rightarrow \frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(B_2))'}|a = -(a' \otimes I_{k-m}) \\
(\Pi_{21}\ \Pi_{22}) &= \lambda \begin{pmatrix} I & I_{k-m} \end{pmatrix} - B_2 a \Rightarrow \frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\lambda))'}|a = (\begin{pmatrix} I \\ I_{k-m} \end{pmatrix} \otimes I_{k-m})
\end{aligned}$$

As $a = (\Pi_{11}\ \Pi_{12})$,

$$p_{new}(a|\Omega) = p_{Jef}(\Pi_{11}, \Pi_{12}|\Omega) \propto |\Omega|^{-\frac{1}{2}m}|X'_{1,-1}M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k}.$$

The priors for λ and B_2 are proportional to the square root of the information matrix of (λ, B_2) given a and Ω, which is the quadratic form of the Jacobians with the information matrix, $I(\Pi_{21}, \Pi_{22}|\Pi_{11}, \Pi_{12}, \Omega)$,

$$p_{new}(B_2, \lambda|a, \Omega)$$

$$
\begin{aligned}
&= \left| \begin{pmatrix} \frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\theta_2))'} | \alpha & \frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\lambda))'} | \alpha \end{pmatrix}' \right. \\
&\quad \left. J(\Pi_{21}, \Pi_{22} | \Pi_{11}, \Pi_{12}, \Omega) \begin{pmatrix} \frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\theta_2))'} | \alpha & \frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\lambda))'} | \alpha \end{pmatrix} \right|^{\frac{1}{2}} \\
&= \left| \begin{pmatrix} -(\alpha' \otimes I_{k-r}) & \left(\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \otimes I_{k-r} \right) \end{pmatrix}' (\Omega^{-1} \otimes Z_{2,-1}' Z_{2,-1}) \right. \\
&\quad \left. \begin{pmatrix} -(\alpha' \otimes I_{k-r}) & \left(\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \otimes I_{k-r} \right) \end{pmatrix} \right|^{\frac{1}{2}} \\
&= \left| \begin{pmatrix} \alpha\Omega^{-1}\alpha' \otimes Z_{2,-1}'Z_{2,-1} & \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}' \Omega^{-1}\alpha' \otimes Z_{2,-1}'Z_{2,-1} \\ \alpha\Omega^{-1} \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \otimes Z_{2,-1}'Z_{2,-1} & \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}' \Omega^{-1} \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \otimes Z_{2,-1}'Z_{2,-1} \end{pmatrix} \right|^{\frac{1}{2}} \\
&= \Big(|\alpha\Omega^{-1}\alpha' \otimes Z_{2,-1}'Z_{2,-1}| \\
&\quad \Big| \Big(\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}' (\Omega^{-1} - \Omega^{-1}\alpha(\alpha\Omega^{-1}\alpha')^{-1}\alpha'\Omega^{-1}) \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \Big) \otimes Z_{2,-1}'Z_{2,-1} \Big| \Big)^{\frac{1}{2}} \\
&= |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)} \left| \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}' \alpha_{\perp}(\alpha_{\perp}\Omega\alpha_{\perp}')^{-1}\alpha_{\perp}' \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \right|^{\frac{1}{2}(k-r)} |Z_{2,-1}'Z_{2,-1}|^{\frac{1}{2}k} \\
&= |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)} \left| \begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix} \Omega \begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}' \right|^{-\frac{1}{2}(k-r)} |Z_{2,-1}'Z_{2,-1}|^{\frac{1}{2}k} \\
&= p_{ecm}(\theta_2|\lambda, \alpha, \Omega) p_{ecm}(\lambda|\alpha, \Omega)
\end{aligned}
$$

where

$$
\begin{aligned}
p_{ecm}(\theta_2|\lambda, \alpha, \Omega) &= |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)} |Z_{2,-1}'Z_{2,-1}|^{\frac{1}{2}r}, \\
p_{ecm}(\lambda|\alpha, \Omega) &= \left| \begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix} \Omega \begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}' \right|^{-\frac{1}{2}(k-r)} |Z_{2,-1}'Z_{2,-1}|^{\frac{1}{2}(k-r)},
\end{aligned}
$$

$\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} \Rightarrow \alpha_{\perp} = A \begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}$, $\alpha_1 : r \times r$, $\alpha_2 : (k-r) \times r$, A : $(k-r) \times (k-r)$ unrestricted, and we use that $(\Omega^{-1} - \Omega^{-1}\alpha(\alpha\Omega^{-1}\alpha')^{-1}\alpha'\Omega^{-1})$ equals $\alpha_{\perp}(\alpha_{\perp}\Omega\alpha_{\perp}')^{-1}\alpha_{\perp}'$.

Proof of lemma 2.

The posterior of the ECM equals the conditional posterior of the unrestricted ECM given $\lambda = 0$,

$$p_{ecm}(\Omega, \alpha, \theta_2|Y) = p_{unr}(\Omega, \alpha, \theta_2|\lambda = 0, Y).$$

As the posterior is proportional to the product of the prior and the likelihood and the likelihood is the only component depending on λ, the priors for both models are equal.

Proof of theorem 5.

The joint posterior of $(\alpha, \Omega, \lambda, \beta_2)$ equals the product of the prior and likelihood. As the decomposition of the prior is given in theorem 1, we decompose the likelihood to show its relationship with the conditional posteriors.

$$l(\Omega, \alpha, \lambda, \beta_2|Y) \propto |\Omega|^{-\frac{1}{2}T} \exp[-\frac{1}{2}tr(\Omega^{-1}\varepsilon'\varepsilon)]$$

The elements in the trace operator can be decomposed as follows,

$$\begin{aligned}
& tr(\Omega^{-1}\varepsilon'\varepsilon) \\
= \; & tr(\Omega^{-1}(\Delta Y - Y_{1,-1}\alpha + Y_{2,-1}\beta_2\alpha - Y_{2,-1}(0 \; \lambda))' \\
& (\Delta Y - Y_{1,-1}\alpha + Y_{2,-1}\beta_2\alpha - Y_{2,-1}(0 \; \lambda))) \\
= \; & tr(\Omega^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\alpha - \hat{\alpha})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\alpha - \hat{\alpha}) \\
& + (\beta_2\alpha - (0 \; \lambda) - \hat{B}_2)' Y_{2,-1}' Y_{2,-1}(\beta_2\alpha - (0 \; \lambda) - \hat{B}_2)) \\
= \; & tr(\Omega^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\alpha - \hat{\alpha})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\alpha - \hat{\alpha})) \\
& + tr(Y_{2,-1}' Y_{2,-1}(\beta_2\alpha - (0 \; \lambda) - \hat{B}_2)\Omega^{-1}(\beta_2\alpha - (0 \; \lambda) - \hat{B}_2)') \\
= \; & tr(\Omega^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\alpha - \hat{\alpha})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\alpha - \hat{\alpha})) \\
& + tr(Y_{2,-1}' Y_{2,-1}((\beta_2 - \hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2 - \hat{\beta}_2)' + \\
& + ((0 \; \lambda) - \hat{B}_2)(\Omega^{-1} - \Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1})((0 \; \lambda) - \hat{B}_2)') \\
= \; & tr(\Omega^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\alpha - \hat{\alpha})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\alpha - \hat{\alpha})) \\
& + tr(Y_{2,-1}' Y_{2,-1}(\beta_2 - \hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2 - \hat{\beta}_2)') \\
& + tr(Y_{2,-1}' Y_{2,-1}((0 \; \lambda) - \hat{B}_2)\alpha_\perp'(\alpha_\perp\Omega\alpha_\perp')^{-1}\alpha_\perp((0 \; \lambda) - \hat{B}_2)') \\
= \; & tr(\Omega^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\alpha - \hat{\alpha})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\alpha - \hat{\alpha})) \\
& + tr(Y_{2,-1}' Y_{2,-1}(\beta_2 - \hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2 - \hat{\beta}_2)') \\
& + tr((\alpha_\perp\Omega\alpha_\perp')^{-1}\alpha_\perp((0 \; \lambda) - \hat{B}_2)' Y_{2,-1}' Y_{2,-1}((0 \; \lambda) - \hat{B}_2)\alpha_\perp') \\
= \; & tr(\Omega^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\alpha - \hat{\alpha})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\alpha - \hat{\alpha})) \\
& + tr(Y_{2,-1}' Y_{2,-1}(\beta_2 - \hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2 - \hat{\beta}_2)') \\
& + tr(((\begin{smallmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{smallmatrix})\Omega(\begin{smallmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{smallmatrix})')^{-1} \\
& (\lambda - \hat{\lambda})' Y_{2,-1}' Y_{2,-1}(\lambda - \hat{\lambda}))
\end{aligned}$$

where $\hat{B}_2 = (Y_{2,-1}'Y_{2,-1})^{-1}Y_{2,-1}'(\Delta Y - Y_{1,-1}\alpha)$, $\alpha = (\alpha_1 \; \alpha_2)$, $\alpha_1 : r \times r$, $\alpha_2 : r \times (k-r)$, $\hat{\alpha} = (Y_{1,-1}'M_{Y_{2,-1}}Y_{1,-1})^{-1}Y_{1,-1}'M_{Y_{2,-1}}\Delta Y$, $\hat{\lambda} = (Y_{2,-1}'Y_{2,-1})^{-1}Y_{2,-1}'\Delta Y \begin{pmatrix} -\alpha_1^{-1}\alpha_2 \\ I_{k-r} \end{pmatrix}$,
$\hat{\beta}_2 = -(Y_{2,-1}'Y_{2,-1})^{-1}Y_{2,-1}'(\Delta Y - Y_{1,-1}\alpha - Y_{2,-1}\begin{pmatrix} 0 & \lambda \end{pmatrix})\Omega^{-1}\alpha'(\alpha\Omega^{-1}\alpha')^{-1}$
and we have used that $\alpha_\perp'(\alpha_\perp\Omega\alpha_\perp')^{-1}\alpha_\perp = (\Omega^{-1} - \Omega^{-1}\alpha(\alpha\Omega^{-1}\alpha')^{-1}\alpha'\Omega^{-1})$,

$$\hat{\delta} = (\tilde{X}_{2,-1}'\tilde{X}_{2,-1})^{-1}\tilde{X}_{2,-1}'\tilde{Y}\tilde{X}\beta_{\perp}'\beta_{\perp 2}^{-1\prime} = (\tilde{X}_{2,-1}'\tilde{X}_{2,-1})^{-1}\tilde{X}_{2,-1}'\tilde{Y}\tilde{X}\begin{pmatrix} -\beta_1^{-1}\beta_2 \\ I_{k-r} \end{pmatrix},$$

as

$$\beta = \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix} \Rightarrow \beta_{\perp} = A\begin{pmatrix} -\beta_2'\beta_1^{-1\prime} & I_{k-r} \end{pmatrix} = \begin{pmatrix} \beta_{1\perp} & \beta_{2\perp} \end{pmatrix}$$

where $\beta_1 : r \times r$, $\beta_2 : (k-r) \times r$, $A : (k-r) \times (k-r)$, unrestricted, $\beta_{2\perp} = A$, $\beta_{1\perp} = -A\beta_2'\beta_1^{-1\prime}$, such that,

$$\beta_{\perp}'\beta_{\perp 2}^{-1\prime} = \begin{pmatrix} -\beta_1^{-1}\beta_2 \\ I_{k-r} \end{pmatrix} A'A^{-1\prime} = \begin{pmatrix} -\beta_1^{-1}\beta_2 \\ I_{k-r} \end{pmatrix},$$

which gives the appropriate decomposition of the kernel of the likelihood.

Proof of Theorem 4.

$$\begin{aligned}
& \sigma^{-2}y'Z(Z'Z)^{-1}Z'y \\
= \; & \sigma^{-2}y'(Z_1\ Z_2)((Z_1\ Z_2)'(Z_1\ Z_2))^{-1}(Z_1\ Z_2)'y \\
= \; & \sigma^{-2}y'(Z_1\ Z_2) \\
& \begin{pmatrix} (Z_1'M_{Z_2}Z_1)^{-1} & -(Z_1'M_{Z_2}Z_1)^{-1}Z_1'Z_2(Z_2'Z_2)^{-1} \\ -(Z_2'Z_2)^{-1}Z_2'Z_1(Z_1'M_{Z_2}Z_1)^{-1} & (Z_2'Z_2)^{-1} + (Z_2'Z_2)^{-1}Z_2'Z_1(Z_1'M_{Z_2}Z_1)^{-1}Z_1'Z_2(Z_2'Z_2)^{-1} \end{pmatrix} \\
& (Z_1\ Z_2)'y \\
= \; & \sigma^{-2}y'Z_2(Z_2'Z_2)^{-1}Z_2'y + \sigma^{-2}y'M_{Z_2}Z_1(Z_1'M_{Z_2}Z_1)^{-1}Z_1'M_{Z_2}y \\
= \; & \sigma^{-2}(\gamma_2 - \hat{\gamma}_2)'Z_2Z_2(\gamma_2 - \hat{\gamma}_2) + \sigma^{-2}y'M_{Z_2}Z_1(Z_1'M_{Z_2}Z_1)^{-1}Z_1'M_{Z_2}y
\end{aligned}$$

The expectation of the first expression over γ_2 has a $\chi^2(l)$ distribution when $\gamma_1 = 0$, such that

$$E_{\gamma_2}E_{\sigma^2}(\sigma^{-2}(\gamma_2 - \hat{\gamma}_2)'Z_2Z_2(\gamma_2 - \hat{\gamma}_2)|\gamma_1 = 0) = E(\chi^2(l)) = l,$$

furthermore

$$\begin{aligned}
E_{\gamma_2}E_{\sigma^2}(\sigma^{-2}y'M_{Z_2}Z_1(Z_1'M_{Z_2}Z_1)^{-1}Z_1'M_{Z_2}y|\gamma_1 &= 0) = \\
E_{\sigma^2}(\sigma^{-2}y'M_{Z_2}Z_1(Z_1'M_{Z_2}Z_1)^{-1}Z_1'M_{Z_2}y|\gamma_1 &= 0),
\end{aligned}$$

such that the equality holds.

Proof of theorem 5.

These natural conjugate priors imply a conditional prior for (Π_{11},Π_{12}) given Ω and a conditional prior for (Π_{21},Π_{22}) given $(\Pi_{11},\Pi_{12},\Omega)$,

$$\begin{aligned}
& p_{lin}(\Pi_{11},\Pi_{12}|\Omega) \\
\propto\ & |\Omega|^{-\frac{1}{2}m}|A_{11.2}|^{\frac{1}{2}k}\exp[-\frac{1}{2}tr(\Omega^{-1}(\begin{pmatrix}\Pi_{11} & \Pi_{12}\end{pmatrix}-\begin{pmatrix}P_{11} & P_{12}\end{pmatrix})' \\
& A_{11.2}(\begin{pmatrix}\Pi_{11} & \Pi_{12}\end{pmatrix}-\begin{pmatrix}P_{11} & P_{12}\end{pmatrix}))], \\
& p_{lin}(\Pi_{21},\Pi_{22}|\Pi_{11},\Pi_{12},\Omega) \\
\propto\ & |\Omega|^{-\frac{1}{2}(k-m)}|A_{22}|^{\frac{1}{2}k}\exp[-\frac{1}{2}tr(\Omega^{-1}(\begin{pmatrix}\Pi_{21} & \Pi_{22}\end{pmatrix}-\begin{pmatrix}G_{21} & G_{22}\end{pmatrix})' \\
& A_{22}(\begin{pmatrix}\Pi_{21} & \Pi_{22}\end{pmatrix}-\begin{pmatrix}G_{21} & G_{22}\end{pmatrix}))],
\end{aligned}$$

where $A_{11.2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$, $\begin{pmatrix}G_{21} & G_{22}\end{pmatrix}=\begin{pmatrix}P_{21} & P_{22}\end{pmatrix}-A_{22}^{-1}A_{21}(\begin{pmatrix}\Pi_{11} & \Pi_{12}\end{pmatrix}-\begin{pmatrix}P_{11} & P_{12}\end{pmatrix})$. As $\pi=\begin{pmatrix}\Pi_{11} & \Pi_{12}\end{pmatrix}$, the prior for π equals the first of the two conditional priors stated above. In the construction of the conditional posterior of (λ,β_2), from the latter of the two conditional posterior, we use the following decompositions (see also the proof of theorem 1),

$$\begin{aligned}
& (i).\ |\Omega|^{-\frac{1}{2}(k-m)}|A_{22}|^{\frac{1}{2}k}|J((\lambda,\beta_2),(\Pi_{21},\Pi_{22})|\pi)| \\
=\ & |\begin{pmatrix}\frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\beta_2))'}|\pi & \frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\lambda))'}|\pi\end{pmatrix}'(\Omega^{-1}\otimes A_{22}) \\
& \begin{pmatrix}\frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\beta_2))'}|\pi & \frac{\partial(vec(\Pi_{21}\ \Pi_{22}))}{\partial(vec(\lambda))'}|\pi\end{pmatrix}| \\
=\ & |\pi\Omega^{-1}\pi'|^{\frac{1}{2}(k-m)}|\begin{pmatrix}-\pi_2'\pi_1^{-1\prime} & I_{k-m}\end{pmatrix}\Omega\begin{pmatrix}-\pi_2'\pi_1^{-1\prime} & I_{k-m}\end{pmatrix}'|^{-\frac{1}{2}(k-m)}|A_{22}|^{\frac{1}{2}k} \\
& (ii).\ \exp[-\frac{1}{2}tr(\Omega^{-1}(\begin{pmatrix}\Pi_{21} & \Pi_{22}\end{pmatrix}-\begin{pmatrix}G_{21} & G_{22}\end{pmatrix})' \\
& A_{22}(\begin{pmatrix}\Pi_{21} & \Pi_{22}\end{pmatrix}-\begin{pmatrix}G_{21} & G_{22}\end{pmatrix}))] \\
=\ & \exp[-\frac{1}{2}[tr((\begin{pmatrix}-\pi_2'\pi_1^{-1\prime} & I_{k-m}\end{pmatrix}\Omega\begin{pmatrix}-\pi_2'\pi_1^{-1\prime} & I_{k-m}\end{pmatrix}')^{-1} \\
& (\lambda-l)'A_{22}(\lambda-l))+tr(A_{22}(\beta_2-B)\pi\Omega^{-1}\pi'(\beta_2-B)')]],
\end{aligned}$$

where $l=\begin{pmatrix}G_{21} & G_{22}\end{pmatrix}\begin{pmatrix}-\pi_1^{-1}\pi_2 \\ I_{k-m}\end{pmatrix}$, $B=-\begin{pmatrix}G_{21} & G_{22}-\lambda\end{pmatrix}\Omega^{-1}\pi'(\pi\Omega^{-1}\pi')^{-1}$. Such that the following conditional priors result,

$$\begin{aligned}
p_{nat}(\Omega) &\propto |S|^{\frac{1}{2}k}|\Omega|^{-\frac{1}{2}(k+m+1)}\exp[-\frac{1}{2}tr(\Omega^{-1}S)], \\
p_{nat}(\pi|\Omega) &\propto |\Omega|^{-\frac{1}{2}m}|A_{11.2}|^{\frac{1}{2}k}\exp[-\frac{1}{2}tr(\Omega^{-1}(\pi-\begin{pmatrix}P_{11} & P_{12}\end{pmatrix})'
\end{aligned}$$

$$
\begin{aligned}
& A_{11.2}(\alpha - \begin{pmatrix} P_{11} & P_{12} \end{pmatrix}))], \\
p_{unres}(\lambda|\alpha,\Omega) &\propto |\begin{pmatrix} -\alpha_2'\alpha_1^{-1'} & I_{k-r} \end{pmatrix} \Omega \begin{pmatrix} -\alpha_2'\alpha_1^{-1'} & I_{k-r} \end{pmatrix}'|^{-\frac{1}{2}(k-r)} |A_{22}|^{\frac{1}{2}(k-r)} \\
& \exp[-\frac{1}{2}tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1'} & I_{k-r} \end{pmatrix} \Omega \begin{pmatrix} -\alpha_2'\alpha_1^{-1'} & I_{k-r} \end{pmatrix}')^{-1} \\
& (\lambda - l)'A_{22}(\lambda - l))], \\
p_{unres}(\beta_2|\lambda,\alpha,\Omega) &\propto |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)}|A_{22}|^{\frac{1}{2}r} \\
& \exp[-\frac{1}{2}tr(A_{22}(\beta_2 - B)\alpha\Omega^{-1}\alpha'(\beta_2 - B)')].
\end{aligned}
$$

Proof of theorem 7.

In terms of the parameters of the linear model, the functional form of the natural conjugate prior reads,

$$
\begin{aligned}
p_{lin}(\Omega) &\propto |S|^{\frac{1}{2}h}|\Omega|^{-\frac{1}{2}(h+m+1)}\exp[-\frac{1}{2}tr(\Omega^{-1}S)], \\
p_{lin}(\Pi|\Omega) &\propto |\Omega|^{-\frac{1}{2}k}|A|^{\frac{1}{2}k}\exp[-\frac{1}{2}tr(\Omega^{-1}(\Pi - P)'A(\Pi - P))].
\end{aligned}
$$

The posterior then becomes,

$$
\begin{aligned}
p_{lin}(\Pi,\Omega|Y) &\propto |S|^{\frac{1}{2}h}|A|^{\frac{1}{2}k}|\Omega|^{-\frac{1}{2}(T+h+k+m+1)}\exp[-\frac{1}{2}tr(\Omega^{-1}[S + \\
& (\Pi - P)'A(\Pi - P) + (\Delta Y - Y_{-1}\Pi)'(\Delta Y - Y_{-1}\Pi)])], \\
&\propto |S|^{\frac{1}{2}h}|A|^{\frac{1}{2}k}|\Omega|^{-\frac{1}{2}(T+h+k+m+1)}\exp[-\frac{1}{2}tr(\Omega^{-1}[S + \Delta Y'\Delta Y \\
& P'AP - \tilde{\Pi}'(A + Y_{-1}'Y_{-1})\tilde{\Pi} + \\
& (\Pi - \tilde{\Pi})'(A + Y_{-1}'Y_{-1})(\Pi - \tilde{\Pi})])],
\end{aligned}
$$

where $\tilde{\Pi} = (A+Y_{-1}'Y_{-1})^{-1}(AP+Y_{-1}'Y_{-1}\hat{\Pi}) = \begin{pmatrix} \tilde{\Pi}_{11} & \tilde{\Pi}_{12} \\ \tilde{\Pi}_{21} & \tilde{\Pi}_{22} \end{pmatrix}$, $(A+Y_{-1}'Y_{-1}) =$ $\begin{pmatrix} (A + Y_{-1}'Y_{-1})_{11} & (A + Y_{-1}'Y_{-1})_{12} \\ (A + Y_{-1}'Y_{-1})_{21} & (A + Y_{-1}'Y_{-1})_{22} \end{pmatrix}$. The posteriors of the parameters of the unrestricted ECM then become,

$$
\begin{aligned}
p_{unres}(\Omega|Y) &\propto |S + P'AP + \Delta Y'\Delta Y - \tilde{\Pi}'(A + Y_{-1}'Y_{-1})\tilde{\Pi}|^{\frac{1}{2}(T+h)} \\
& |\Omega|^{-\frac{1}{2}(T+h+m+1)}\exp[-\frac{1}{2}tr(\Omega^{-1}(S + P'AP \\
& +\Delta Y'\Delta Y - \tilde{\Pi}'(A + Y_{-1}'Y_{-1})\tilde{\Pi}))], \\
p_{unres}(\alpha|\Omega,Y) &\propto |\Omega|^{-\frac{1}{2}r}|(A + Y_{-1}'Y_{-1})_{11.2}|^{\frac{1}{2}k}\exp[-\frac{1}{2}tr(\Omega^{-1}(\alpha - \begin{pmatrix} \tilde{\Pi}_{11} & \tilde{\Pi}_{12} \end{pmatrix})'
\end{aligned}
$$

$$
\begin{aligned}
& (S + \Phi_{-1}'\Phi_{-1})_{11.2}(a - \begin{pmatrix} \bar{\Pi}_{11} & \bar{\Pi}_{12} \end{pmatrix}))], \\
p_{ncm}(\lambda|a,\Omega,\Sigma) \;\propto\; & |\begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix} \Omega \begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix}'|^{-\frac{1}{2}(k-m)} \\
& |(S + \Phi_{-1}'\Phi_{-1})_{22}|^{\frac{1}{2}(k-m)} \exp[-\frac{1}{2} tr((\begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix} \Omega \begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix}')^{-1} \\
& (\lambda - \tilde{\lambda})'(S + \Phi_{-1}'\Phi_{-1})_{22}(\lambda - \tilde{\lambda}))], \\
p_{ncm}(\Pi_2|\lambda,a,\Omega,\Sigma) \;\propto\; & |a\Omega^{-1}a'|^{\frac{1}{2}(k-m)} |(S + \Phi_{-1}'\Phi_{-1})_{22}|^{\frac{1}{2}m} \\
& \exp[-\frac{1}{2} tr(S_{22}(\Pi_2 - \tilde{\Pi}_2)a\Omega^{-1}a'(\Pi_2 - \tilde{\Pi}_2)')].
\end{aligned}
$$

where $\begin{pmatrix} \psi_{21} & \psi_{22} \end{pmatrix} = (\begin{pmatrix} \bar{\Pi}_{21} & \bar{\Pi}_{22} \end{pmatrix} - (a - (S + \Phi_{-1}'\Phi_{-1})_{22}^{-1}(S + \Phi_{-1}'\Phi_{-1})_{12} \begin{pmatrix} \bar{\Pi}_{11} & \bar{\Pi}_{12} \end{pmatrix})),$
$\tilde{\lambda} = \begin{pmatrix} \psi_{21} & \psi_{22} \end{pmatrix} \begin{pmatrix} -a_1^{-1}a_2 \\ I_{k-m} \end{pmatrix}$, $\tilde{\Pi}_2 = -\begin{pmatrix} \psi_{21} & \psi_{22} - \lambda \end{pmatrix} \Omega^{-1}a'(a\Omega^{-1}a')^{-1}.$

Proof of Theorem 8.

The Natural Conjugate implied Prior for λ reads,

$$
\begin{aligned}
p_{ncm}(\lambda|a,\Omega) \;\propto\; & |\begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix} \Omega \begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix}'|^{-\frac{1}{2}(k-m)} |S_{22}|^{\frac{1}{2}(k-m)} \\
& \exp[-\frac{1}{2} tr((\begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix} \Omega \begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix}')^{-1} \\
& (\lambda - l)'S_{22}(\lambda - l))],
\end{aligned}
$$

Define:

$$
\begin{aligned}
\eta &= S_{22}^{\frac{1}{2}}\lambda \left[\begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix} \Omega \begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix}' \right]^{-\frac{1}{2}}, \\
\theta &= S_{22}^{\frac{1}{2}} l \left[\begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix} \Omega \begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix}' \right]^{-\frac{1}{2}}.
\end{aligned}
$$

Prior for η then becomes

$$
p_{ncm}(\eta|a,\Omega) \propto \exp[-\frac{1}{2} tr((\eta - \theta)'(\eta - \theta))].
$$

The RCCM corresponds with $\lambda = 0$ which is identical to $\eta = 0$. So, also the PBFs for both parametrisations will be identical. It holds that

$$
\lim_{s_{22}, l \to 0} \exp[-\frac{1}{2} tr(\theta'\theta)] = 1,
$$

as $\theta = S_{22}^{\frac{1}{2}} l \left[\begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix} \Omega \begin{pmatrix} -a_2'a_1^{-1\prime} & I_{k-m} \end{pmatrix}' \right]^{-\frac{1}{2}}$. So,

$$
\lim_{s_{22}, l \to 0} \exp[-\frac{1}{2} tr(\theta'\theta)] = \int (2\pi)^{-\frac{1}{2}(k-m)^2} \exp[-\frac{1}{2} tr((\eta - \theta)'(\eta - \theta))] d\eta = 1.
$$

When we respecify the priors in terms of β instead of B, it holds that,

$$p_{cross}(\alpha, \Omega) = (2\pi)^{-\frac{1}{2}(k-r)^2} \exp[-\frac{1}{2}tr(\beta'\beta)]p_{uncross}(\alpha, \Omega),$$

such that when we use the limit when the prior parameters converge to 0 and equal prior probabilities, we obtain,

$$\begin{aligned}
\lim_{a_{22}, \lambda \to 0} POR(r, k) &= \lim_{a_{22}, \lambda \to 0} \frac{\Pr(r = r)}{\Pr(r = k)} \\
&= \lim_{a_{22}, \lambda \to 0} \frac{p_r \int\int\int p_{cross}(\alpha, \Omega) d\alpha d\Omega}{p_k \int\int p_{uncross}(\alpha, \Omega) d\alpha d\Omega} \\
&= (2\pi)^{-\frac{1}{2}(k-r)^2}.
\end{aligned}$$

References

[1] Box, G.E.P. and G.C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley (Reading, MA), (1973).

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

[3] DeJong, D.N. Co-integration and trend stationarity in macroeconomic time series : Evidence from the likelihood function. *Journal of Econometrics*, 52:347–370, (1992).

[4] Doan, T., R.B. Litterman, and C. Sims. Forecasting and Conditional Projections using Realistic Prior Distributions. *Econometric Reviews*, 3:1–100, (1984).

[5] Dorfman, J.H. A numerical Bayesian test for cointegration of AR processes. *Journal of Econometrics*, 66:289–324, (1995).

[6] Engle, R.F. and C.W.J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55:251–276, (1987).

[7] Geweke, J. Bayesian Inference in Econometric Models using Monte Carlo Integration. *Econometrica*, 57:1317–1339, (1989a).

[8] Geweke, J. Exact Predictive Densities for Linear Models with ARCH Disturbances. *Journal of Econometrics*, 40:63–86, (1989b).

[9] Geweke, J. Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75:121–146, (1996).

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

[11] Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59:1551–1580, (1991).

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

[13] Kleibergen, F. Equality Restricted Random Variables: Densities and Sampling Algorithms. *Econometric Institute Report*, 9662, http://www.cs.few.eur.nl/~kleiberg/publications.html, (1996a).

[14] Kleibergen, F. Reduced Rank Regression using Generalized Method of Moments Estimators. *Center Discussion Paper*, 9621, Tilburg University, (1996b).

[15] Kleibergen, F. and H. Hoek. Bayesian Analysis of ARMA models using Noninformative Priors. *Econometric Institute Report*, 9553, http://www.cs.few.eur.nl/~kleiberg/publications.html, (1995).

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

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

[18] Kleibergen, F. and H.K. van Dijk. Bayesian Simultaneous Equations Analysis using Reduced Rank Structures. (1996). working paper.

[19] Koop, G. Cointegration in Present Value Relationships. *Journal of Econometrics*, 49:105–139, (1991).

[20] Litterman, R.B. Forecasting with Bayesian Vector Autoregressions : Five Years of Experience. *Journal of Business and Economic Statistics*, 4:25–38, (1986).

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

[22] Phillips, P.C.B. Partially Identified Econometric Models. *Econometric Theory*, 5:181–240, (1989).

[23] Phillips, P.C.B. Optimal Inference in Cointegrated Systems. *Econometrica*, 59:283–306, (1991).

[24] Phillips, P.C.B. Econometric Model Determination. *Econometrica*, 64:763–812, (1996).

[25] Phillips, P.C.B. and S.N. Durlauf. Multiple Time Series Regression with Integrated Processes. *Review of Economic Studies*, 53:473–495, (1986).

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

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

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

[29] Tierney, L. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, 22:1701–1762, (1994).

[30] Watson, M.W. Vector Autoregessions and Cointegration. In R.F. Engle and D.L. McFadden, editor, *Handbook of Econometrics, volume 4*, pages 2844–2945. Elsevier Science (Amsterdam), (1994).

[31] Zellner, A. *An Introduction to Bayesian Inference in Econometrics*. Wiley (New York), (1971).