# BAYESIAN INFERENCE OF GENERAL LINEAR RESTRICTIONS ON THE COINTEGRATION SPACE

MATTIAS VILLANI

Abstract. The degree of empirical support of *a priori* plausible structures on the cointegration vectors has a central role in the analysis of cointegration. Villani (2000) and Strachan and van Dijk (2003) have recently proposed finite sample Bayesian procedures to calculate the posterior probability of restrictions on the cointegration space, using the existence of a uniform prior distribution on the cointegration space as the key ingredient. The current paper extends this approach to the empirically important case with different restrictions on the individual cointegration vectors. Prior distributions are proposed and posterior simulation algorithms are developed. Consumers' expenditure data for the US is used to illustrate the robustness of the results to variations in the prior. A simulation study shows that the Bayesian approach performs remarkably well in comparison to other more established methods for testing restrictions on the cointegration vectors.

Keywords: Bayesian inference, cointegration, posterior probability, restrictions.
JEL classification: C11, C12.

## 1. Introduction

The analysis of cointegration (Engle and Granger, 1987) has received an enormous attention both in the theoretical and the more applied economic and econometric literature, see e.g. Johansen (2005) for a recent overview of the econometric analysis. One of the main reasons for the popularity of cointegration models is the separation of the long-run component, where economic theory may provide useful information, from the less understood short-run dynamics. This separation opens up the possibility to determine the degree of empirical support of long run equilibria suggested by economic theory, which is often the centerpiece of the analysis. Traditionally, this has been tackled by classical hypothesis tests of restrictions on the space spanned by the cointegration vectors, the so called *cointegration space*, see Johansen (1995a) for the likelihood ratio test. Opinions on the usefulness of classical tests are very diverse, but it is probably agreed that such tests become much less appealing if more than one 'null' hypothesis is present, which is typically the case in cointegration analysis. In addition, only asymptotic distributions of the statistics are available, and the sample size necessary for them to be useful seems to be larger than the sample size in typical applications (Gredenhoff and Jacobson, 2001), although bootstrapping procedures (Gredenhoff and Jacobson, 2001) and a Bartlett-type correction (Johansen, 2000) have been suggested to alleviate this deficiency.

*Research department, Sveriges Riksbank, SE-103 37 Stockholm, Sweden* and *Department of Statistics, Stockholm University. E-mail: mattias.villani@riksbank.se.* The author would like to thank Sune Karlsson, Rolf Larsson, Rodney Strachan, and Daniel Thorburn for helpful discussions. Financial support from the Swedish Council of Research in Humanities and Social Sciences (HSFR), grant No. F0383/96 and the Swedish Research Council (Vetenskapsrådet), grant no. 412-2002-1007 are gratefully acknowledged. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

The Bayesian analysis of restrictions on the cointegration space focuses on the posterior distribution over the set of cointegration restrictions, or, more generally, models, and is straightforward *in principle*: formulate a prior distribution for the parameters in each model under comparison and compute the posterior distribution over the set of models. The difference between classical hypothesis tests and its Bayesian alternative is well documented in the literature, see *e.g.* Lindley (1957), Edwards, Lindman and Savage (1963), and Berger and Delampady (1987). Previous Bayesian analyses of restrictions on the cointegration vectors include Strachan (2003), where an extension of the embedding approach of Kleibergen and Paap (2002) and Kleibergen and van Dijk (1998) is used, and Martin (2000) and Martin and Martin (2000) where restrictions in bivariate systems are analyzed. Paap and van Dijk (2003) analyze restrictions in a more flexible Markov switching model.

The non-linearity and high dimensionality of the parameter space in error correction (EC) models makes both the prior formulation and the posterior computation a real challenge, however. Even the development of a diffuse, 'non-informative', reference prior has taken more than a decade, see Kleibergen and van Dijk (1994), Bauwens and Lubrano (1996), Geweke (1996), Kleibergen and van Dijk (1998), Martin and Martin (2000), Kleibergen and Paap (2002), Strachan (2003), Strachan and Inder (2004) and Villani (2005a) for different suggestions. The field of Bayesian cointegration is surveyed in Koop, Strachan, van Dijk and Villani (2005).

An additional complication when it comes to comparing models in the Bayesian approach is that the prior distribution must necessarily be proper, with the possible exception of those dimensions of the parameter space which are common to all models under comparison; see O'Hagan (1995) for a discussion. Villani (2000) and Strachan and van Dijk (2003) have pointed out that the EC model belongs to the class of models where there exists a well-defined, non-controversial, *proper* uniform distribution. This comes from the fact that the cointegration vectors are only determined up to arbitrary linear combinations, *i.e.* only the cointegration space is identified (Johansen, 1995a). The parameter space of the cointegration vectors is therefore not Euclidean, but rather the abstract space consisting of all subspaces with a fixed dimension, the *Grassman manifold* (Villani, 2005a,b; Strachan and Inder (2004); Strachan and van Dijk, 2004). This is important as the Grassman manifold is bounded and admits a unique invariant probability measure, which may be used to define the uniform distribution on this space (Mardia and Khatri, 1977). The other parameters of the model, such as adjustment coefficients and short-run dynamics, may conveniently be assigned improper priors as these parameters do not differ across the hypothesized cointegration spaces.

We make a number of contributions in this paper. First, we extend the analysis of cointegration restrictions in Strachan and van Dijk (2003), to the case where the restrictions may differ across cointegration vectors, which is a common situation in applied work. We propose a prior distribution and devise tailored numerical simulation algorithms to compute the posterior distribution over the set of restrictions. We also discuss, and illustrate by an empirical example, the crucial role played by the prior on the adjustment coefficients for the inference on the cointegration restrictions. Finally, we conduct a simulation study to evaluate the performance of the Bayesian approach to restrictions on the cointegration space in comparison to more established methods like the likelihood ratio test (with and without Bartlett correction) and some widely used information criteria.

The analysis of the restrictions on the cointegration space proceeds conditional on a prespecified lag length in the VAR model. Since our focus here is on testing restrictions on the cointegration space we shall also assume the cointegration rank to be known *a priori.* The proposed procedure for posterior analysis of restrictions may be trivially extended to a joint analysis of restrictions and cointegration rank if the proper prior distribution on the adjustment coefficients is used, see Strachan and van Dijk (2003). Alternatively, the posterior distribution of the cointegration rank may be obtained separately using one of the procedures in Kleibergen and Paap (2002), Strachan (2003), Corander and Villani (2004), Strachan and Inder (2004), Strachan and van Dijk (2004) and Villani (2005a).

## 2. The cointegrated vector autoregressive model

The base model used throughout this paper is the $p$-dimensional error correction (EC) model

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t, \tag{2.1}$$

where $x_t$ is a column vector containing the $p$ time series at time $t$, $d_t$ contains observations on $q$ exogenous variables and $\varepsilon_t \sim N_p(0, \Sigma)$ with independence between time periods. The columns of $\beta$ $(p \times r)$ are the *cointegration vectors* such that $\beta' x_t$ represents the stationary departures from the $r$ long run equilibria and $\alpha$ contains the *adjustment coefficients* which control the adjustment back to equilibrium. The space spanned by $\beta$ has been termed the *cointegration space*. The exogenous variables in $d_t$ may be restricted to the cointegration space, as discussed in Johansen (1995a). The analysis presented here can easily accommodate such extensions by a suitable redefinition of $x_{t-1}$ and $d_t$ in exactly the same way as in the maximum likelihood analysis.

The model can be written in the following compact form

$$Y = X\beta\alpha' + Z\Psi + E,$$

where the $t$th row of $Y$, $X$, $Z$ and $E$ is given by $\Delta x_t'$, $x_{t-1}'$, $(\Delta x_{t-1}', ..., \Delta x_{t-k+1}', d_t')$ and $\varepsilon_t'$, respectively, and $\Psi = (\Gamma_1, ..., \Gamma_{k-1}, \Phi)'$.

Any $r \times r$ non-singular matrix $U$ and its inverse can be used to transform $\beta$ and $\alpha$ without affecting their product. *i.e.* $\alpha\beta' = \alpha^*\beta^{*\prime}$, where $\alpha^* = \alpha U'^{-1}$ and $\beta^* = \beta U$. Hence, what the data can determine is the cointegration space, $\operatorname{sp}\beta$, and the adjustment space, $\operatorname{sp}\alpha$, but there is no way to further discriminate between the elements in $\alpha$ and $\beta$. In order to identify $\alpha$ and $\beta$, linear restrictions can be imposed on the columns of $\beta$. Assume that $r_i$ restrictions are imposed on the $i$th cointegration vector

$$R_i'\beta_i = 0, \tag{2.2}$$

where $R_i$ is a $p \times r_i$ full rank restriction matrix and $\beta_i$ is the $i$th column of $\beta$. Over-identifying restrictions are of course represented in the same way by increasing the number of columns of $R_i$. Necessary and sufficient conditions for the restrictions to be identifying are given in Johansen (1995b). Alternatively, restrictions may be imposed on $\alpha$ and analyzed in the same way as $\beta$ (the problem is symmetric in $\alpha$ and $\beta$), provided $\beta$ is left unrestricted. If both $\alpha$ and $\beta$ are restricted, the analysis is more complicated, however, and this case will not be considered here.

It is useful to parametrize the restricted $\beta_i$ in terms of its unrestricted elements (Johansen, 1995a)

$$\beta_i = H_i\varphi_i, \tag{2.3}$$

where $H_i = R_{i\perp}$, an $p\times(p-r_i)$ matrix orthogonal to $R_i$, and $\varphi_i$ is the $(p-r_i)$-dimensional vector of free elements in $\beta_i$ after the restrictions given by $R_i$ have been imposed. For future reference, let $s_i = p - r_i$, and $\varphi = (\varphi_1', ..., \varphi_r')'$. As $H_i$ is of full column rank, each $H_i$ may be made orthonormal by the transformation $H_i \to H_i(H_i'H_i)^{-1/2}$. Since $\operatorname{sp} H_i = \operatorname{sp}[H_i(H_i'H_i)^{-1/2}]$, the restrictions on the cointegration space are unchanged by this transformation and there is no loss in generality in assuming each $H_i$ to be orthonormal. This assumption helps to clarify the form of the prior distribution presented in the next section.

The identifying restrictions only determine each $\beta_i$ up to a constant and a normalization of each $\beta_i$ is therefore necessary. This can be done in many ways, but we choose to normalize each cointegration vector to unit length, thus restricting each $\varphi_i$ to the $(s_i-1)$-dimensional unit sphere in $\mathbb{R}^{s_i}$, which we denote by $\mathbb{S}^{s_i}$[1]. The unit length normalization helps in understanding the type of prior used for $\beta$ and, more importantly, allows us to assume prior independence between $\alpha$ and $\beta$, see the next section. One indeterminacy remains, however; data cannot discriminate between the vectors $\beta_i$ and $-\beta_i$ on opposite poles of $\mathbb{S}^{s_i}$. This could be settled by e.g. restricting the first element of $\varphi_i$ to be positive, thus restricting each $\varphi_i$ to the $(s_i-1)$-dimensional unit hemisphere in $\mathbb{R}^{s_i}$. The numerical algorithms developed in Section 5 are more easily implemented if we do not impose the sign restriction, however. This is possible as long as all considered densities (priors, proposal distributions in MCMC, see Section 5, etc) are antipodally symmetric, i.e. satisfy $f(x)=f(-x)$, where $x$ is a vector of unit length; see also the discussion of the prior in the next section.

## 3. A reference prior for the unit length normalization

Let $\Psi$ be uniformly distributed over $\mathbb{R}^{p\times p(k-1)+q}$ and assume that

$$\Sigma \sim IW(A,v), \tag{3.1}$$

*a priori*, where IW denotes the inverted Wishart distribution (Zellner, 1971).

To motivate the prior used for $\varphi$, let us first consider the case with a single cointegration vector $\beta_1$. The unit length of $\beta_1$ implies that $\beta_1'\beta_1=\varphi_1'H_1'H_1\varphi_1=\varphi_1'\varphi_1=1$. A natural reference prior for $\varphi_1$ is therefore the uniform distribution on $\mathbb{S}^{s_1}$. It easy to see that this prior implies that the line spanned by $\beta_1=H_1\varphi_1$ is uniformly distributed over the set of all lines in $\operatorname{sp} H_1$. Thus, we may say that, every possible one-dimensional cointegration space in $\mathbb{R}^p$ which satisfies the (over-)identifying restrictions receives the same prior probability.

In the case with $r>1$ we would ideally use a prior which assigns the same probability to every possible $r$-dimensional cointegration space in $\mathbb{R}^p$ which satisfies the (over-)identifying restrictions. This prior has been derived in the unrestricted case in Villani (2005a) for the linear normalization ($\beta=(I_r,B')'$) and by Strachan and Inder (2004) in the semi-orthogonal normalization ($\beta'\beta=I_r$). Over-identifying restrictions destroys the mathematically convenient symmetry in the just-identified case, which in turn leads to substantial complications for the prior[2]. We shall instead assume that the columns of $\beta$ are independent *a priori* and that $\varphi_i$ is uniformly distributed over $\mathbb{S}^{s_i}$. That is, the overall prior on $\varphi$ is (Mardia and Jupp, 2000)

$$p(\varphi_1,...,\varphi_r)=\prod_{i=1}^{r}\frac{\Gamma(\frac{s_i}{2})}{2\pi^{s_i/2}}. \tag{3.2}$$

[1]We note that the $(p-1)$-dimensional unit sphere in $\mathbb{R}^p$ is often denoted $\mathbb{S}^{p-1}$ rather than $\mathbb{S}^p$. We have chosen the latter notation for simplicity.

[2]The exception here is when the restrictions are of the form $\beta=(\beta^*,H\varphi)$, where $\beta^*$ is a $p\times m$ matrix of fully specified cointegration vectors, $H$ is a $p\times s$ restriction matrix for the remaining $r-m$ cointegration vectors and $\varphi$ is $s\times(r-m)$. Under such restrictions, the parameter space has the same structure as the unrestricted case (but with smaller dimension) and the invariant prior may be derived as in Strachan and van Dijk (2003). One could of course use a mixed approach where Strachan and van Dijk's (2003) prior is used on those sets of restrictions where it is applicable and use the prior in (3.2) on all other restriction sets.

The assumption of a priori independent columns of $\beta$ is very convenient when we later devise numerical algorithms for computing the marginal likelihood, see Section 5. Note that the density in (3.2) implictly contains the correction factor in Strachan and Inder (2004) between the Grassman manifold (in this case the unit hemisphere) and the Steifel manifold (in this case the unit sphere), which in this special case is $2^r$. Note also that it does not matter which orthonormal version of $H_i$ we use since $\beta_i = H_i\varphi_i = H_iQ_i'Q_i\varphi_i = \bar{H}_i\bar{\varphi}_i$ where $Q_i$ is orthonormal and the uniform prior on $\varphi_i$ is rotationally invariant (Mardia and Jupp, 2000), *i.e.* $p(Q_i\varphi_i) = p(\varphi_i)$ for any orthonormal matrix $Q_i$.

As an illustration, consider a case with a three-element cointegration vector $\beta = (\beta_{11}, \beta_{12}, \beta_{13})'$ of unit length. Say that we want to impose the restriction $\beta_{11} = -\beta_{12}$. The cointegration vector may then be expressed in terms of its unrestricted elements as follows

$$\beta = H\varphi = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}\cos\theta \\ -\frac{1}{\sqrt{2}}\cos\theta \\ \sin\theta \end{pmatrix}, \tag{3.3}$$

where $\varphi = (\varphi_{11}, \varphi_{12})'$ is the unit length vector of unrestricted coefficients and $\theta \in [-\pi, \pi)$ is the angle of $\varphi$ in polar coordinates. As $\theta$ travels from $-\pi$ to $\pi$, equation (3.3) traces out a curve in $\mathbb{R}^3$, which is the parameter space of $\beta$ under the restrictions. The uniform prior on $\varphi$ over the two-dimensional unit sphere implies that $\theta \sim Unif[-\pi, \pi)$ *a priori* (Mardia and Jupp, 2000). The implied prior on $\beta_{11}$ and $\beta_{13}$ (remember $\beta_{12} = -\beta_{11}$ so there is no need to plot all three dimensions) is displayed in Figure 1 (right subgraph).

It is interesting to compare the uniform prior to a normal prior in the commonly used *linear normalization* of $\beta$, where one of the elements of $\beta$ is normalized to unity. Let $\beta = (1, -1, b)'$ denote the restricted cointegration vector in this normalization and assume that $b \sim N(0, \kappa^2)$ *a priori*. Figure 1 depicts the implied prior on $\theta$ (left) and $\beta_{11}$ and $\beta_{13}$ (right) for $\kappa = 2$ and $\kappa = 10$. To aid in the comparison, we have have mirrored the priors in the linear normalization to the opposite side of the ellipse in right hand graph in Figure 1. It is clear that the normal priors in the linear normalization are very informative with respect to the one feature of $\beta$ which is identified, namely the direction of the cointegration vector. In fact, no prior variance $\kappa^2$ produces the uniform distribution on the restricted cointegration space $\operatorname{sp} H$. As shown in Villani (2005a), a standard Cauchy prior on $b$ implies a uniform distribution over the cointegration spaces.

The prior on the adjustment coefficients in $\alpha$ is an extension of the prior used in Strachan and Inder (2004) and Villani (2005a) to the just-identified case

$$\operatorname{vec}\alpha|\Sigma \sim N_p(0, V \otimes \Sigma), \tag{3.4}$$

where $V = \operatorname{diag}(\tau_1^2, ..., \tau_r^2)$. The prior in (3.4) implies that the vector of adjustment coefficients for different cointegrating relations are independent conditional on $\Sigma$, more precisely $\alpha_i|\Sigma \sim N_p(0, \tau_i^2\Sigma)$ independent of $\alpha_j$, $j \neq i$. The possibly differing scales of the time series are taken into account by the use of $\Sigma$ in the conditional variance of $\alpha_i$. The conditional prior is thus of shrinkage type, depending only on a small set of hyperparameters. In many cases it is sufficient to use $\tau_1 = ... = \tau_r = \tau$, and the posterior probabilities of the restrictions on the cointegration space may be plotted as a function of $\tau$, which is an effective way to communicate the results; an example of this procedure is given in Section 6. As explained in the next section, it is even possible to use $\tau_i = \infty$, for all $i$, *i.e.* to assign a flat and improper prior to $\alpha$, and still obtain a well defined posterior distribution over the set of cointegration restrictions.

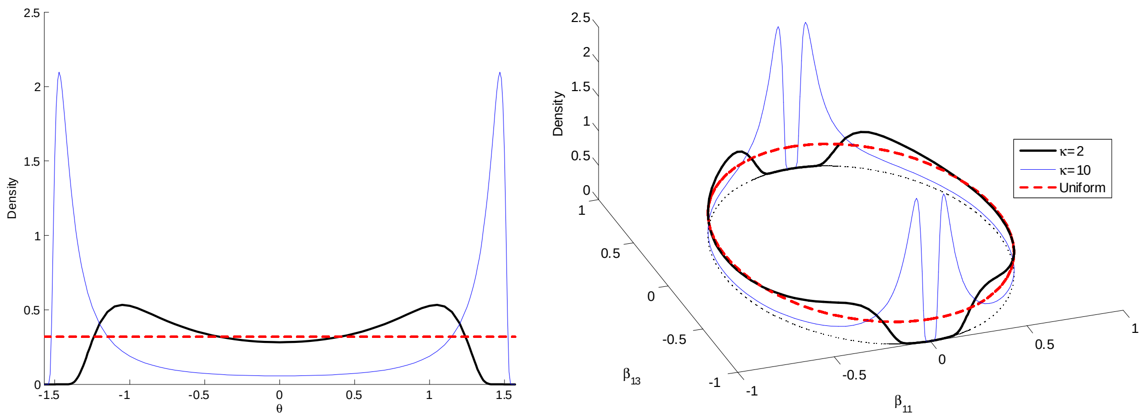


FIGURE 1. Prior distributions on the cointegration vector viewed in $\theta$-space (left) and over the unit circle (right).

In summary, the overall prior for all parameters of the EC model takes the form

$$p(\alpha, \varphi, \Psi, \Sigma) \propto |\Sigma|^{-(p+r+v+1)/2} \operatorname{etr}[\Sigma^{-1}(A + \alpha V^{-1}\alpha')] \prod_{i=1}^{r} \frac{\Gamma(\frac{s_i}{2})}{2\pi^{s_i/2}}, \tag{3.5}$$

where $\operatorname{etr}(X) = \exp[-(1/2) \operatorname{tr} X]$, for any square matrix $X$.

There is an implicit assumption of independence of $\alpha$ and $\beta$ in (3.5), which may considered odd at first sight given that the elements of $\alpha$ are the coefficients in the regression of $\Delta x_t$ on $\beta' x_{t-1}$. The magnitude (scale) of $\alpha$ is therefore inversely related to the magnitude of $\beta$, and the prior distribution of $\alpha$ should therefore be modelled conditional on $\beta$ (see Villani (2005a) for such a construction). The non-identification of $\beta$ opens up a possibility to avoid this prior dependence, however. Since the length of the cointegration vectors are arbitrary, we may use a unit length normalization of $\beta$, thereby pinning down its scale, and we may realistically assume prior independence between $\alpha$ and $\beta$.

To aid in the specification of the prior hyperparameters $\tau_1, ..., \tau_r$ in $V$ it is helpful to derive the marginal posterior distribution of $\alpha$. This marginal prior is obtained by integrating out $\Sigma$, using properties of the inverted Wishart distribution, and reads

$$\alpha \sim t_{p\times r}(0, A, V, v - p + 1), \tag{3.6}$$

where $t_{p\times r}$ denotes the matrix $t$ distribution (Box and Tiao, 1973 and Zellner, 1971). From Box and Tiao (1973, p. 446-447) we have $E(\alpha) = 0$ and $Cov(\operatorname{vec} \alpha) = \frac{1}{v-p-1} V \otimes A$. Thus, in order to specify $\tau_1, ..., \tau_r$ one should first specify $A = (v - p - 1)E(\Sigma)$. It often sufficiently accurate, compared to the data information, to assume a diagonal $A$, whose $p$ elements may reasonably be specified by the user. If one is unwilling to specify $A$ altogether, then a data based prior may be used, as suggested in Villani (2005a), with $v = p + 2$ and $A = E(\Sigma) = \hat{\Sigma}$, the maximum likelihood estimate of $\Sigma$. The choice of $v = p + 2$ makes the prior on $\Sigma$ the least informative prior in the inverted Wishart family subject to a finite expectation, which mitigates the effect of the slightly unorthodox use of sample data in the prior. Yet another approach would be to use $A = 0$, $v = 0$ (which corresponds to the well-known $|\Sigma|^{-(p+1)/2}$ prior for $\Sigma$).

## 4. The posterior probability of over-identifying restrictions on the cointegration space

Suppose that we have $m$ different theories/hypotheses on the cointegration space, $h_1, ..., h_m$, which are represented by different (linear) over-identifying restrictions on $\beta$ and let $\varphi^{(i)}$ denote the remaining free parameters in $\beta$ after the restrictions given by $h_i$ have been imposed. See Johansen (1995a) for a discussion of possible types of restrictions and their interpretation. A Bayesian comparison of these hypotheses is very simple in principle: simply compute the posterior probabilities of the hypotheses under consideration

$$p(h_i|x^{(T)}) = \frac{p(h_i)m_i(\mathbf{x}^{(T)})}{\sum_{j=1}^{m} p(h_j)m_j(\mathbf{x}^{(T)})}, \tag{4.1}$$

where $x^{(T)}$ denotes the data up to time $T$, $p(h_i)$ is the prior probability of $h_i$,

$$m_i(x^{(T)}) = \int_{\varphi^{(i)}}\int_{\alpha}\int_{\Psi}\int_{\Sigma} p(x^{(T)}|\alpha, \varphi^{(i)}, \Psi, \Sigma, h_i)p(\alpha, \varphi^{(i)}, \Psi, \Sigma|h_i)d\Sigma d\Psi d\alpha d\varphi^{(i)}$$

is the *marginal likelihood* of the data under the $i$th hypothesis, $p(x^{(T)}|\alpha, \varphi^{(i)}, \Psi, \Sigma, h_i)$ is the usual (conditional) likelihood function and $p(\alpha, \varphi^{(i)}, \Psi, \Sigma|h_i)$ is the prior distribution of the unrestricted parameters under $h_i$. Note that $m_i(x^{(T)})$ is the actual probability of observing $x^{(T)}$ if $h_i$ is true and the prior is $p(\alpha, \varphi^{(i)}, \Psi, \Sigma|h_i)$. The posterior probabilities in (4.1) can be used in a multitude of ways, *e.g.* to weigh the predictions (and their uncertainty) from the $m$ different models (hypotheses), see Villani (2001) and Strachan and van Dijk (2004).

Since the uniform prior on the cointegration space is proper, the posterior probabilities $p(h_i|x^{(T)})$ are well defined even when the prior distribution of $\alpha, \Psi$ and $\Sigma$ is improper as their normalizing constants (which only exist in a limiting sense, see O'Hagan (1995)) cancel out in (4.1), if the prior on these parameters is the same across all $h_i$. Thus, contrary to what might be expected *prima facie*, Bayesian inference on the restrictions on the cointegration space is still well defined when the prior with $\tau_1 = ... = \tau_r = \infty$ (constant prior on $\alpha$ over $\mathbb{R}^{p\times r}$) is used. Note that this analysis does not require the user to specify any prior hyperparameters at all and may therefore be part of the default statistics reported by software packages.

Suppressing the subscript which denotes a particular model, the marginal likelihood of model (2.1) is proportional to

$$m(x^{(T)}) = \int_{\varphi}\int_{\alpha}\int_{\Psi}\int_{\Sigma} |\Sigma|^{-(T+p+r+v+1)/2} \operatorname{etr}\left[\Sigma^{-1}\left(A + \alpha V^{-1}\alpha' + \sum_{t=1}^{T}\varepsilon_t\varepsilon_t'\right)\right] p(\varphi)d\Sigma d\Psi d\alpha d\varphi,$$

where $\varepsilon_t = \Delta x_t - \alpha\beta' x_{t-1} - \sum_{i=1}^{k-1}\Gamma_i\Delta x_{t-i} - \Phi d_t$.

Using first properties of the inverted Wishart distribution to handle the integral with respect to $\Sigma$ and then properties of the matrix $t$ distribution for the integral with respect to $\alpha$ and $\Psi$ (Zellner, 1971), we obtain

$$m(x^{(T)}) = \int_{\varphi} \frac{\left|V^{-1} + \beta' C_1 \beta\right|^{l_1}}{\left|V^{-1} + \beta' C_2 \beta\right|^{l_2}} p(\varphi)d\varphi, \tag{4.2}$$

where $C_1 = X'Q_1X$, $Q_1 = I - Z(Z'Z)^{-1}Z'$, $C_2 = X'Q_2[I_T - Z(Z'Q_2Z)^{-1}Z'Q_2]X$, $Q_2 = I_T - Y(A+Y'Y)^{-1}Y'$, $l_1 = (T+v-pk-q)/2$ and $l_2 = (T+v-p(k-1)-q)/2$. Note that $\beta$ is a function of $\varphi$. An unimportant multiplicative factor, common to all hypotheses, which will cancel out in all model comparisons, has been left out in (4.2). Expression (4.2) is as far as we get analytically, the integral with respect to $\varphi$ must be computed numerically, see Section 5.

From (4.2) it is seen that the marginal likelihood for $A = 0$, $v = 0$ (which corresponds to the well-known $|\Sigma|^{-(p+1)/2}$ prior) and $\tau_1 = ... = \tau_r = \infty$ is very close to the *a priori* expected value of the likelihood ratio in Johansen (1995a). The only discrepancy is in the powers of the two determinants in (4.2); in the likelihood ratio we have $l_2 = l_1 = T/2$. This difference in 'degrees of freedom' (which diminishes in importance as the length of the time series increases since both $l_1$ and $l_2$ grow with $T$) comes from the treatment of the nuisance parameters (i.e. $\alpha, \Psi$ and $\Sigma$), which are integrated out by their priors in the Bayesian approach but concentrated out with their ML estimates in the likelihood approach (Bauwens and Lubrano, 1996).

## 5. Numerical evaluation of the marginal likelihood

### 5.1. **Monte Carlo integration.**

An obvious suggestion for a numerical evaluation of the integral in (4.2) follows immediately from the observation that $m(x^{(T)})$ is the expectation of

$$Q(\varphi) = \frac{\left|V^{-1} + \beta' C_1 \beta\right|^{l_1}}{\left|V^{-1} + \beta' C_2 \beta\right|^{l_2}},$$

with respect to the prior $p(\varphi)$. A simple numerical approach is therefore to generate samples from $p(\varphi)$ and then estimate $m(x^{(T)})$ with an arithmetic average of the computed $Q(\varphi)$. A draw from the prior of $\varphi_i$ is easily performed using the result that $n/\|n\|$, where $n$ is a $s_i$-dimensional vector of independent normal variates, follows the uniform distribution on $\mathbb{S}^{s_i}$ (see, *e.g.* Muirhead, 1982).

### 5.2. **Importance sampling.**

Importance sampling, introduced to econometricians by Kloek and van Dijk (1978) and further developed by Geweke (1989), is a refinement of simple Monte Carlo integration. Let $g(\varphi)$ be an arbitrary density for $\varphi$, usually called the importance density. The marginal likelihood can then be written

$$m(x^{(T)}) = \int_{\varphi} Q(\varphi) \frac{p(\varphi)}{g(\varphi)} g(\varphi) d\varphi$$

and $m(x^{(T)})$ is therefore the expectation of $Q(\varphi)p(\varphi)/g(\varphi)$ with respect to $g(\varphi)$, which may be estimated by an arithmetic average of $Q(\varphi)p(\varphi)/g(\varphi)$ computed from the $\varphi$ sampled from $g(\varphi)$. Note that by using an importance density which is approximately proportional to the marginal posterior of $\varphi$ (which is $Q(\varphi)p(\varphi)$), we are effectively estimating $m(x^{(T)})$ with an average of terms with small variability, with a resulting precise estimate of $m(x^{(T)})$.

Importance densities are naturally based on the most widely used distribution on $\mathbb{S}^p$: the von Mises distribution (Mardia and Jupp, 2000). The density of the von Mises distribution is of the form

$$p(x) = c_p(\lambda) \exp(\lambda x' \mu) dS_p,$$

where $x$ and $\mu$ are vectors on $\mathbb{S}^p$, $\lambda$ is a positive scalar and $dS_p$ denotes the probability element on $\mathbb{S}^p$. $\mu$ is the mean direction of $x$ and $\lambda$ determines the degree of concentration around the mean. The normalizing constant is

$$c_p^{-1}(\lambda) = (2\pi)^{p/2} \lambda^{-(p-1)/2} I_{(p-1)/2}(\lambda),$$

where $I_q(\lambda)$ is the modified Bessel function of the first kind. Ulrich (1984) describes an efficient algorithm for generating variates from the von Mises distribution. The von Mises distribution is not antipodally symmetric (i.e. $p(x) \neq p(-x)$), however. As discussed in Section 2, this property is necessary here as the unit length cointegration vectors are only unique up to sign switches. The von Mises density is easily modified to be antipodally symmetric by a simple reflection to the opposite side of the sphere, giving the density

$$p(x) = \frac{c_p(\lambda)}{2}[\exp(\lambda x'\mu) + \exp(-\lambda x'\mu)]dS_p. \tag{5.1}$$

We will for simplicity refer to (5.1) as the von Mises density, and denote it by $M_p(\mu, \lambda)$, with the implicit understanding that we are referring to its modified form.

We propose the following importance density based on the von Mises distribution

$$g(\varphi_1, ..., \varphi_r) = g_1(\varphi_1)g_2(\varphi_2|\varphi_1) \cdots g_r(\varphi_r|\varphi_1, ..., \varphi_{r-1}),$$

where

$$\varphi_i|\varphi_1, ..., \varphi_{i-1} \sim M_{s_i}(\hat{\varphi}_i, \lambda_i)$$

and $\hat{\varphi}_i$ is the maximum likelihood estimate of $\varphi_i$ conditional on a cointegration rank equal to $i$ *and* the coefficients in the other $i-1$ cointegration vectors being $\varphi_1, ..., \varphi_{i-1}$ (obtained from e.g. the switching algorithm in Johansen (1995a)). Thus, the mean vector in the von Mises distribution of $\varphi_1$ is the ML estimate in the model with a single cointegration vector and restrictions given by $H_1$, the mean vector in the von Mises distribution of $\varphi_2$ is the ML estimate in the model with a two cointegration vectors where the first vector is fixed to the previously generated $\varphi_1$ and second vector is restricted by $H_2$ and so on. The $\lambda$'s may be used to fine tune the importance function to the problem at hand; the estimated standard errors from the ML estimator of $\varphi$ may be used as a guide. Note that the concentration of $g(\varphi)$ around its modal axis (given by $\pm\mu$) increases with $\lambda_i$.

5.3. **Methods based on posterior sampling.** Several methods approximate the marginal likelihood $m(x^{(T)})$ using a sample from the posterior distribution of the model parameters. Direct sampling from the marginal posterior of $\varphi$ is not feasible. An alternative is to employ the Gibbs sampler to generate variates iteratively from the full conditional posteriors $p(\varphi_i|\varphi_{-i}, \mathcal{D})$, where $\varphi_{-i}$ equals $\varphi$ with the elements in $\varphi_i$ excluded, always conditioning on the most recently updated $\varphi_{-i}$. Since the full conditional posteriors are non-standard distributions on the unit sphere, this procedure is likely to be time-consuming. A preferable approach is to use the Metropolis-Hastings (H-M) algorithm (Metropolis et al, 1953; Hastings, 1970) to sample each full conditional posterior, *i.e.* sampling from $p(\varphi|\mathcal{D})$ is done by the so called *Metropolis-within-Gibbs* algorithm. The M-H algorithm draws (proposals) from a distribution which roughly approximates the target distribution and accepts the draws with a certain probability. Let $q(\varphi_i^{(j+1)}|\varphi_i^{(j)}, \varphi_{-i}^{(j)})$ denote the distribution used to generate a candidate draw of $\varphi_i$ in the $(j+1)$th iteration of the algorithm and let

$$a(\varphi_i^{(j)}, \varphi_i^{(j+1)}, \varphi_{-i}^{(j)}) = \min\left[\frac{q(\varphi_i^{(j+1)}|\varphi_i^{(j)}, \varphi_{-i}^{(j)})p(\varphi_i^{(j+1)}|\varphi_{-i}^{(j)}, \mathcal{D})}{q(\varphi_i^{(j)}|\varphi_i^{(j+1)}, \varphi_{-i}^{(j)})p(\varphi_i^{(j)}|\varphi_{-i}^{(j)}, \mathcal{D})}, 1\right],$$

be the acceptance probability of the transition $\varphi_i^{(j)} \to \varphi_i^{(j+1)}$, where $p(\varphi_i^{(j)}|\varphi_{-i}^{(j)}, \mathcal{D})$ is the full conditional posterior density kernel of $\varphi_i$. The generated sequence of draws are dependent but can be shown to converge in distribution to $p(\varphi|\mathcal{D})$ as $j \to \infty$ (Tierney, 1994). Several proposal distributions are possible, *e.g.*

$$\varphi_i^{(j+1)}|\varphi_i^{(j)}, \varphi_{-i}^{(j)} \sim M_{s_i}(\varphi_i^{(j)}, \lambda_i).$$

Since the von Mises distribution is symmetric in its argument and mean vector, for this specific proposal distribution we have $q(\varphi_i^{(j+1)}|\varphi_i^{(j)}, \varphi_{-i}^{(j)}) = q(\varphi_i^{(j)}|\varphi_i^{(j+1)}, \varphi_{-i}^{(j)})$ and the acceptance probability simplifies to

$$a(\varphi_i^{(j)}, \varphi_i^{(j+1)}, \varphi_{-i}^{(j)}) = \min\left(\frac{p(\varphi_i^{(j+1)}|\varphi_{-i}^{(j)}, \mathcal{D})}{p(\varphi_i^{(j)}|\varphi_{-i}^{(j)}, \mathcal{D})}\right).$$

Another proposal distribution is

$$\varphi_i^{(j+1)}|\varphi_i^{(j)}, \varphi_{-i}^{(j)} \sim M_{s_i}(\tilde{\varphi}_i, \lambda_i)$$

where $\tilde{\varphi}_i$ is either the unconditional ML estimate of $\varphi_i$ in the model with $r$ cointegration relations (independence sampler) or the ML estimate of $\varphi_i$ conditional on the most recent draw of $\varphi_{-i}$.

Once a posterior sample is available, several methods may be used to compute the marginal likelihood $m(x^{(T)})$. Importance sampling with the posterior distribution as importance density may be used. The resulting estimator is the harmonic mean of the likelihoods of the posterior draws; Geweke (1999) describes a modified harmonic estimator along the lines initially suggested by Gelfand and Dey (1994). Yet another use of posterior samples for computing marginal likelihoods has recently been proposed by Chib and Jeliazkov (2001). Their work extends Chib's earlier procedure for computing marginal likelihoods from the Gibbs posterior sample (Chib, 1995) to the Metropolis-Hastings setting. The Chib (1995) procedure was used in Villani (2005a) for computing the posterior distribution of the cointegration rank with encouraging results.

## 6. Empirical illustration

The long-run relationship between consumers' expenditure and income in the US over the time period 1956-1991 is analyzed in Holden and Perman (1994). A VAR model was used with quarterly observations on three variables: real consumers' expenditure, real disposable income and real personal wealth. Three dummy variables and an unrestricted constant are also added to the model. All variables are seasonally adjusted and in natural logarithms. All three series seem to be $I(1)$, with the exception of the wealth variable which may well be $I(0)$. Holden and Perman (1994) conditions the analysis on $k = 2$ and $r = 1$, which we will also use here for comparison. A Bayesian analysis of the cointegration rank may be obtained using the methods suggested by e.g. Kleibergen and Paap (2002), Strachan and Inder (2004) or Villani (2005a). We use the prior with $A = \hat{\Sigma}$ and $v = p + 2$.

Holden and Perman test several restrictions on $\beta$ against the unrestricted alternative, using the likelihood ratio test (LRT) in Johansen (1995a). These restrictions and the restriction that wealth is $I(0)$ (*i.e.* $\beta = (0, 0, 1)'$) are displayed in the first column of Table 1. Note that the first and third restrictions fully specify $\beta$ and no prior needs to be specified on $\varphi$. The second column shows the ML estimates under the restrictions and the remaining columns contains ($-2$ times) the maximum log-likelihood under the hypotheses, the asymptotic $p$-value of the LRT, the bootstrapped $p$-value of the LRT, the posterior probability of the restrictions for $\tau = \infty$ (flat prior on $\alpha$) and the last column contains the first order SBC approximation of the posterior probabilities (Schwarz, 1978). There are several interesting comparisons that may be made from Table 1, but two general remarks are: *i)* although the four methods all favor the restriction $\beta = (1, -1, 0)'$, there is a substantial disagreement regarding the relative weight of evidence of the hypotheses and *ii)* there seems to be more to the inferences than merely comparing the maximum likelihood to the number of imposed restrictions (as is done by the two large sample methods: the asymptotic LRT and SBC); a possible explanation is that although the parameter space is Euclidean asymptotically (locally), it is a curved manifold in finite samples (Villani, 2005b).

| $\beta$ | $\hat{\beta}$ | $-2\ln L$ | As. $p$-value | Boot. $p$-value | $\tau=\infty$ | $SBC$ |
|---|---|---|---|---|---|---|
| $(1,-1,0)'$ | $(1,-1,0)'$ | 1174.57 | 0.103 | 0.241 | 0.950 | 0.693 |
| $(1,-1,a)'$ | $(1,-1,-0.01)'$ | 1174.60 | 0.034 | 0.210 | 0.020 | 0.071 |
| $(0,0,1)'$ | $(0,0,1)'$ | 1167.13 | 0.000 | 0.014 | 0.000 | 0.000 |
| $(1,b,c)'$ | $(1,-0.94,-0.04)'$ | 1176.85 | – | – | 0.002 | 0.071 |
| $(1,d,0)'$ | $(1,-0.96,0)'$ | 1175.43 | 0.092 | 0.192 | 0.028 | 0.172 |

TABLE 1. Restrictions on the cointegration vector in the Holden-Perman data. $\hat{\beta}$ is the ML estimate of $\beta$ under the restrictions, $2\ln L$ is $-2$ times the maximum log-likelihood and the $p$-value refers to the LRT of the restriction against the unrestricted alternative. $\tau=\infty$ is the posterior probability of the restriction using the prior with $A=\hat{\Sigma}$, $v=p+2$ and $\tau=\infty$ and $SBC$ is the asymptotic SBC approximation of the same.

| $\beta$ | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\alpha}_3$ |
|---|---|---|---|
| $(1,-1,0)'$ | 0.183 | 0.399 | 0.078 |
| $(1,-1,a)'$ | 0.184 | 0.414 | 0.077 |
| $(0,0,1)'$ | 0.008 | 0.006 | 0.013 |
| $(1,b,c)'$ | 0.237 | 0.593 | 0.304 |
| $(1,d,0)'$ | 0.205 | 0.424 | 0.167 |

TABLE 2. Maximum likelihood estimates of $\alpha$ under the different restrictions on the cointegration space.

The posterior probabilities of the five hypotheses on the cointegration space are plotted in Figure 2 as a function of the prior hyperparameter $\tau$. The first impression from Figure 2 is that the restriction inference varies considerably with respect to $\tau$ and that this prior hyperparameter must therefore be exactly pinned down by the user. A closer look reveals that this is not the case and that a very rough choice of $\tau$ is actually sufficient. To see this, note that the inferences are essentially the same for all $\tau>10$. Note also that $\text{diag}(A)=\text{diag}(\hat{\Sigma})=(0.010^2, 0.015^2, 0.041^2)$ and the marginal posterior standard deviation of the elements of $\alpha$ are therefore $0.010\tau$, $0.015\tau$ and $0.041\tau$, respectively (see the discussion of the marginal prior of $\alpha$ in (3.6)). A comparison with the conditional ML estimates of $\alpha$ in Table 2 makes it clear that priors with $\tau<10$ are extremely tightly located around the point $\alpha=0$ and grossly in conflict with data. The exception is the restriction $\beta=(0,0,1)'$, where $\hat{\alpha}$ is close to the zero vector; this explains the high posterior support of this restriction for the smallest $\tau$'s. If it is agreed that a reasonable prior has $\tau>10$, then we may conclude that the data information is so over-whelming that the conclusions do not depend the choice of $\tau$ and a subjective consensus has thus been reached.

## 7. SIMULATION EXPERIMENTS

A small simulation study was conducted to compare the Bayesian procedure to established methods of analyzing restrictions on $\beta$. To keep things simple, we restrict attention to the bivariate VAR(1) process with no deterministic variables and one cointegration relation, *i.e.* $p=2$ and $k=r=1$. Johansen (2000) uses the invariance of the likelihood ratio test (LRT) under linear transformation of the time series to find a canonical form of this process suitable

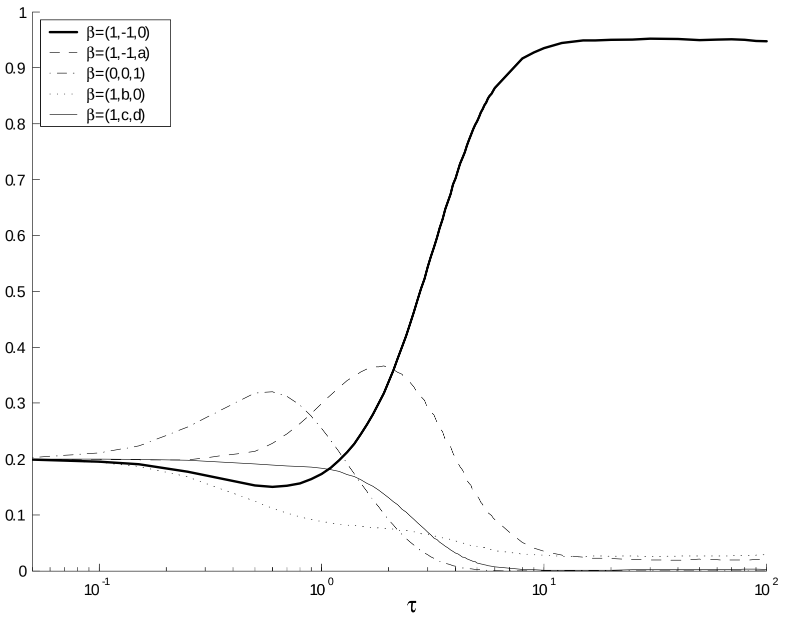


FIGURE 2. Posterior probability of the hypotheses on $\beta$ in the Holden-Perman data as a function of the prior hyperparameter $\tau$.

for simulation studies. The canonical process has the following parameters values

$$\begin{aligned} \beta &= (1,0)' \\ \alpha &= (\eta,\xi)',\ \xi \leq 0 \\ \Sigma &= I_2, \end{aligned}$$

which in turn depend on only two parameters, $\eta$ and $\xi$. The canonical form generates $I(1)$ variables with one cointegrating vector if $-2 < \eta < 0$. $\eta = \xi = 0$ results in an $I(1)$ process with no cointegration relationship (i.e. $r = 0$) whereas $\eta = 0$ and $\xi \neq 0$ determine a $I(2)$ process. Thus, $\eta$ close to zero generates processes where it should be difficult to find support for $\beta = (1,0)'$ and the difficulty increases as $\xi$ also approaches zero. Johansen (2000) investigates $\eta$ ranging from $-1$ to $-0.1$ and, since the LRT is also invariant to the sign of $\xi$, $\xi$ ranging from $-1$ to 0. It will become evident, however, that values of $\eta$ and $\xi$ smaller than $-0.5$ generate data which are too informative to be interesting. The 12 pairs of $\eta$ and $\xi$ considered here are given in Table 3.

Two scenarios for $\beta$ are considered in the simulation study. In Scenario I, the following four hypotheses are used:

$$\begin{aligned} h_1 &: \beta = (1,0)', \\ h_2 &: \beta = (0,1)', \\ h_3 &: \beta = (1,-1)', \\ h_4 &: \beta \text{ is exactly identified but otherwise unrestricted.} \end{aligned}$$

Thus, the true $\beta$ is one of the hypotheses and the data should provide support for $h_1$. Scenario II is the same as the first scenario, with the exception that $\beta = (1,0)'$ is replaced by $\beta = (1,1)'$ under $h_1$ and the true $\beta$ is therefore no longer one of the hypotheses. A good procedure should therefore support $h_4$ strongly.

| $\xi \backslash \eta$ | $-0.1$ | $-0.2$ | $-0.5$ |
|---|---|---|---|
| 0.0 | 1 | 5 | 9 |
| $-0.1$ | 2 | 6 | 10 |
| $-0.2$ | 3 | 7 | 11 |
| $-0.5$ | 4 | 8 | 12 |

TABLE 3. The 12 pairs of $\eta$ and $\xi$ used in the simulations together with their specification numbers used in Figure 2 and 3.

To be able to compare with other model selection procedures, such as selection rules based on the LRT, the SBC criterion (Schwarz, 1978) and the AIC criterion (Akaike, 1974), we focus on the frequency of choice of hypothesis. The Bayes procedure used in the simulations assigns equal prior probability to all hypotheses and then chooses the hypothesis with highest posterior probability. The flat prior on $\alpha$ (i.e. $\tau = \infty$), along with $A = 0$ and $v = 0$, is used for all hypotheses. The following decision rule based on the LRT is used:

> Let $h_*$ denote the $h_i$ ($i = 1, 2, 3$) with largest likelihood. Test $h_*$ against $h_4$ with the LRT at significance level $\delta$. If $h_*$ is rejected, choose $h_4$. If $h_*$ cannot be rejected, choose $h_*$.

The significance level, $\delta$, (which of course is not the significance level of the overall procedure) must be decided upon in the simulation study. A small $\delta$ will obviously work well in Scenario I while a large $\delta$ will be better if Scenario II is at hand. $\delta$ equal to 0.025, 0.05 and 0.1 were used in the simulations, but the results are only presented for $\delta = 0.05$, which gave the best performance of the LRT judged over both scenarios. This search for an 'optimal' $\delta$ biases the presented results in favor of the LRT.

The test statistic of the LRT is asymptotically distributed as a $\chi^2$ variate, but is well known (Gredenhoff and Jacobson, 2001) to be severely oversized even for moderately large sample sizes. To remedy this, Johansen (2000) derived a Bartlett correction to the LRT statistic, see his Corollary 9 for the correction in the special case $k = r = 1$.

In each scenario, $10,000$ processes were generated for each $(\eta, \xi)$-pair and for each simulated process a choice of hypothesis was made by all selection procedures. In addition, two different sample sizes were used: $T = 50$ and $T = 100$. The posterior probabilities were calculated by Monte Carlo integration (see Section 5.1) with $5,000$ draws from the prior, which gave a sufficiently small simulation error.

Consider first the results of Scenario I, where $h_1$ is the true hypothesis. Figure 3 shows the frequency of choice between the four hypotheses (one in each subgraph), for sample size $T = 50$ and different values of $\eta$ and $\xi$ (see Table 3 for the specification numbers of $\alpha$). For example, the upper left subgraph of Figure 3 shows that the Bartlett-corrected LRT chose the true hypothesis ($\beta = (1, 0)'$) in nearly $8,500$ of the $10,000$ simulated processes for the parameter setting $\eta = -0.1$ and $\xi = 0.0$ ($\alpha$-specification number 1).

From Figure 3 it is seen that the Bayes procedure is on average more correct than any other procedure, the exception being the parameter setting $\eta = -0.1$ and $\xi = 0.0$, where it is slightly beaten by the Bartlett corrected LRT. The asymptotic LRT is so close to SBC that the two lines are sometimes completely overlapping.

Not only does Bayes choose the true hypothesis most frequently, but it is also most restrictive in choosing the false hypotheses, $h_2$ and $h_3$. This is important since the consequences of choosing a false null hypothesis are more severe than choosing $h_4$. Note also that all procedures prefer $h_3$ over $h_2$, which is sensible since $(1, -1)'$ is closer than $(0, 1)'$ to the true $\beta$. We do not report detailed results for the sample size $T = 100$ and simply note that the relative

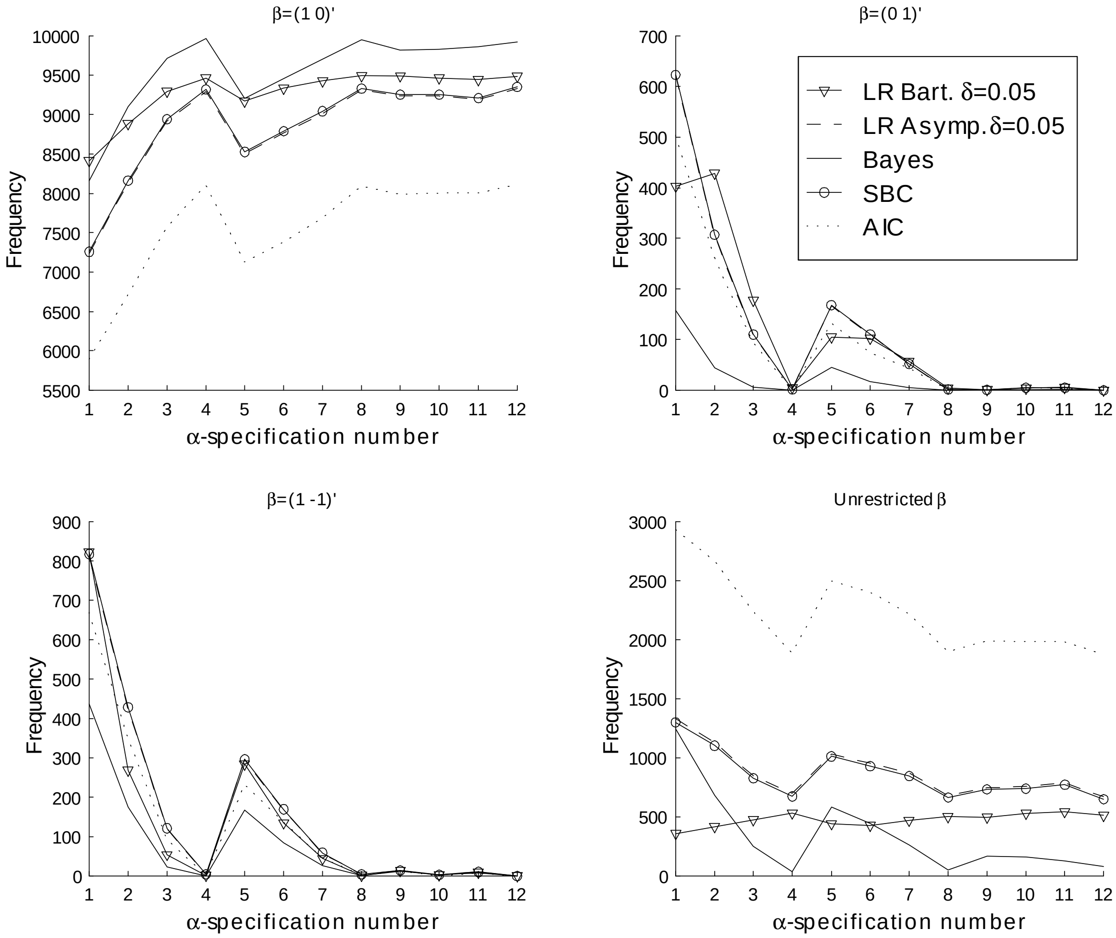


FIGURE 3. Scenario I. Frequency of choice of the four hypotheses for $\beta$ in 10.000 bivariate processes with $T = 50$, $r = k = 1$, $\beta = (1, 0)'$, $\Sigma = I_2$. 12 different $\alpha = (\eta, \xi)'$ are used, where the different pairs of $\eta$ and $\xi$ are given in Table 3 with their specification number.

performance of the five estimators remain about the same. The exception is SBC, which has improved relative to the other procedures and is now uniformly better than asymptotic LRT.

Figure 4 summarizes the results for Scenario II, where the true $\beta$ is not one of the hypotheses. The Bayes procedure chooses $h_4$ more often and any of the other hypotheses less often than any other procedure for all parameter settings. It is even better than AIC, which is facing its ideal situation here (the largest model is the right choice). SBC and both LRT procedures very often (in fact, even more often than not for some parameter settings) choose one of the false hypotheses (i.e. one of $h_1, h_2$ and $h_3$) instead of choosing the unrestricted alternative, an error which should be considered especially grave. This situation improves for $T = 100$ (not shown here), but for processes close to $I(1)$ without cointegration (i.e. for small $\eta$ and $\xi$) SBC and the LR procedures are still too reluctant to choose $h_4$. For $T = 100$, the Bayes procedure still dominates all other procedures for all parameter settings.

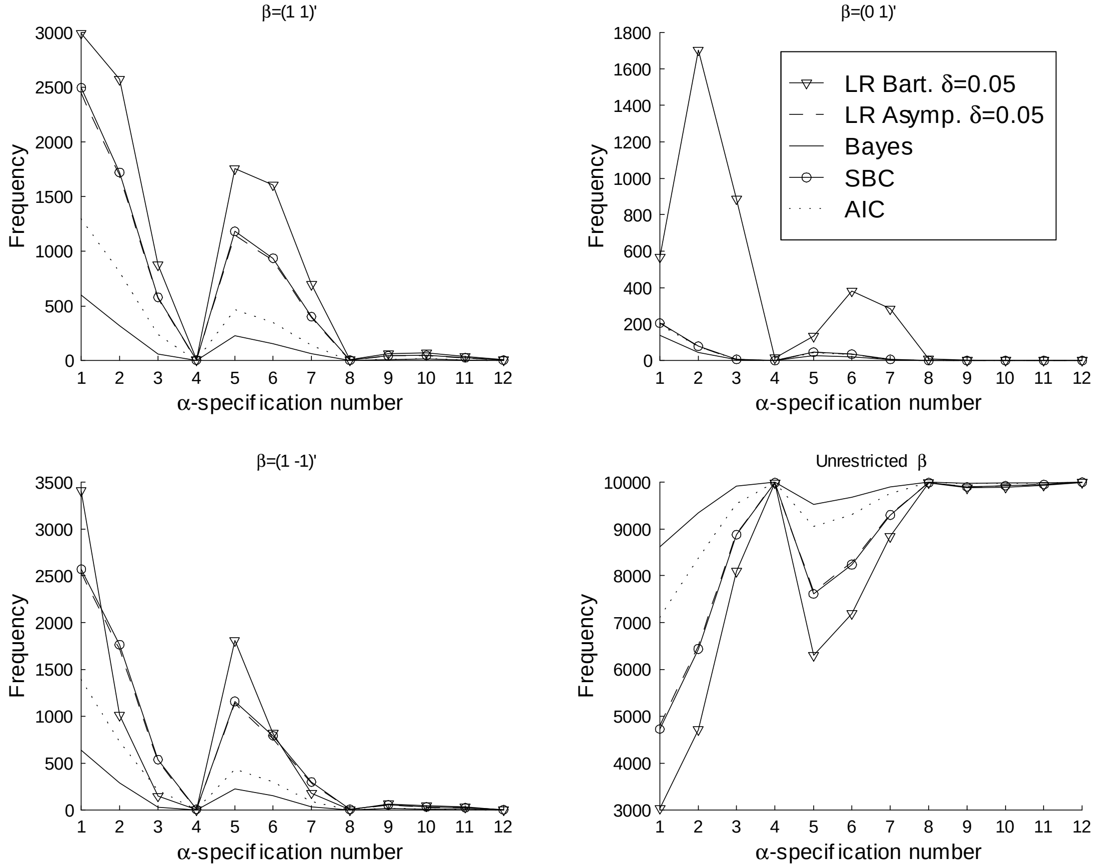


Figure 4. Scenario II. Frequency of choice of the four hypotheses for $\beta$ in 10.000 bivariate processes with $T = 50$, $r = k = 1$, $\beta = (1, 0)'$, $\Sigma = I_2$. 12 different $\alpha = (\eta, \xi)'$ are used, where the different pairs of $\eta$ and $\xi$ are given in Table 3 with their specification number.

## 8. Concluding remarks

We have presented a Bayesian analysis of restrictions on the cointegration space where the restrictions may differ across cointegration vectors. The prior distribution is motivated by the so called cointegration space approach (see Koop et al. (2005) for a discussion), utilizing the fact that the parameter space of the cointegration vectors is compact and therefore admits a proper uniform distribution. Several numerical algorithms for computing the posterior probabilities of the restrictions were proposed. The prior on the adjustment coefficients and its effect on the posterior probabilities of restrictions on the cointegration space were discussed and illustrated in an empirical example. Finally, a simulation study was conducted where the Bayesian approach proved to have remarkably good properties compared to its competitors.

The only drawback of the Bayesian procedure seems to be the need to partly rely on numerical computations. This is an obstacle which should be immaterial in the near future given the current speed of development in computing technology, and already today should not discourage practitioners from using the procedure. The drastic differences between the asymptotic distribution of the LRT and the one obtained with the bootstrap in Section 6 reveals that the use of the LRT is not so straight-forward as is often believed. If the use of the

LRT necessitates a resort to bootstrap methods, then computing time is no longer an item in favor of the LRT compared to the Bayesian procedure proposed here.

The focus in the paper was on the development of a prior which can be a convenient vehicle in inference reporting. It should be clear that the procedure is easily extended to an informative prior on the unrestricted coefficients in the cointegration vectors. A particularly attractive distribution in the unit length normalization is the von Mises distribution on the unit sphere. The non-identification of the cointegration vectors may give the unrestricted coefficients in the cointegration vectors complicated interpretations and such a prior should therefore be very carefully elicitated.

## References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, **19**, 716-727.

[2] Bauwens, L. and Lubrano, M. (1996). Identification restrictions and posterior densities in cointegrated Gaussian VAR systems. In *Advances in Econometrics*, Volume **11**, Part B, JAI Press, 3-28.

[3] Berger, J. O. and Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence (with discussion), *Journal of the American Statistical Association*, **82**, 112-139.

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

[5] Corander, J. and Villani, M. (2004). Bayesian assessment of dimensionality in reduced rank regression, *Statistica Neerlandica, **58**, 255-270.*

[6] Chib, S. (1995). Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**, 1313-21.

[7] Chib, S. and Jeliazkov, I. (2001). Marginal Likelihood from the Metropolis-Hastings Output, *Journal of the American Statistical Association*, **96**, 270-281.

[8] Edwards, W. L., Lindman, H. and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, **70**, 193-242. Reprinted in *The Writings of Leonard Jimmie Savage: a Memorial Collection*. Washington: ASA/IMS, 1981.

[9] Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251-76.

[10] Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, **57**, 1317-40.

[11] Geweke, J. (1996). Bayesian reduced rank regression in econometrics, *Journal of Econometrics*, **75**, 121-46.

[12] Geweke, J. (1999). Using simulation methods for Bayesian econometrics models: inference, development and communication, *Econometric Reviews*, **18** (1), 1-73.

[13] Gredenhoff, M. and Jacobson, T. (2001). Bootstrap testing linear restrictions on the cointegration vectors, *Journal of Business and Economic Statistics*, **19**, 63-72.

[14] Hastings, W. (1970). Monte-Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97-109.

[15] Holden, D. and Perman, R. (1994). Unit roots and cointegration for the economist, in *Cointegration for the Applied Economist*, ed. Bhaskara Rao, B., 47-112. London: Macmillan Press.

[16] Johansen, S. (1995a). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

[17] Johansen, S. (1995b). Identifying restrictions on linear equations: with applications to simultaneous equations and cointegration, *Journal of Econometrics*, **69**, 111-32.

[18] Johansen, S. (2000). A Bartlett correction factor for tests on the cointegrating relations, *Econometric Theory*, **16**, 740-778.

[19] Johansen, S. (2005). Cointegration: an overview, *Palgrave Handbook of Econometrics*: Volume 1 Econometric Theory, forthcoming.

[20] Kleibergen, F. and van Dijk, H. K. (1994). On the shape of the likelihood/posterior in cointegration models, *Econometric Theory*, **10**, 514-51.

[21] Kleibergen, F. and Paap, R. (2002). Priors, posteriors and bayes factors for a Bayesian analysis of cointegration, *Journal of Econometrics*, **111**, 223-249.

[22] Kloek, T. and van Dijk, H. K. (1978). Bayesian estimates of system equation parameters; an application of integration by Monte Carlo, *Econometrica*, **46**, 1-19.

[23] Koop, G., Strachan, R. W., van Dijk, H. K. and Villani, M. (2005). Bayesian Approaches to Cointegration, *Palgrave Handbook of Econometrics*: Volume 1 Econometric Theory, forthcoming.
[24] Lindley, D. V. (1957). A statistical paradox, *Biometrika*, **44**, 187-192.
[25] Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics.* Chichester: Wiley.
[26] Mardia, K. V. and Khatri, C. G. (1977). Uniform distribution on a Steifel manifold, *Journal of Multivariate Analysis*, **7**, 468-473.
[27] Martin, G. M. (2000). US deficit sustainability: a new approach based on multiple endogenous breaks, *Journal of Applied Econometrics*, **15**, 83-105.
[28] Martin, G. M. and Martin, V. L. (2000). Bayesian inference in the triangular cointegration model using a Jeffreys prior, *Communications in Statistics - Theory and Methods*, **29** (8), 1759-1785.
[29] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087-1091.
[30] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: Wiley.
[31] O'Hagan, A. (1995). Fractional Bayes factors for model comparison, *Journal of the Royal Statistical Society B*, **57**, 99-138 (with discussion).
[32] Paap, R. and van Dijk, H. K. (2003): Bayes Estimates of Markov Trends in Possibly Cointegrated Series: An Application to US Consumption and Income, *Journal of Business and Economic Statistics* **21**, 547-563.
[33] Phillips, P. C. B. (1991). Optimal inference in cointegrated systems, *Econometrica*, **59**, 283-306.
[34] Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461-464.
[35] Strachan, R. W. (2003). Valid Bayesian estimation of the cointegrating error correction model, *Journal of Business and Economic Statistics*, **21**, 185-195.
[36] Strachan, R. W. and Inder, B. (2004). Bayesian analysis of the error correction model, *Journal of Econometrics*, **123**, 307-325.
[37] Strachan, R. W. and van Dijk, H. K. (2003). Bayesian model selection with an uninformative prior, *Oxford Bulletin of Economics and Statistics*, **65**, 863-876.
[38] Strachan, R. W. and van Dijk, H. K. 2004, Valuing structure, model uncertainty and model averaging in vector autoregressive processes, Econometric Institute Report EI 2004-23, Erasmus University Rotterdam.
[39] Ulrich, G. (1984). Computer generation of distributions on the $m$-sphere, *Applied Statistics*, **33**, 158-163.
[40] Villani, M. (2001). Bayesian prediction with cointegrated vector autoregressions, *International Journal of Forecasting*, **17**, 585-605.
[41] Villani, M. (2005a). Bayesian reference analysis of cointegration, *Econometric Theory*, **21**, 326-357.
[42] Villani, M. (2005b). Bayesian point estimation of the cointegration space, *Journal of Econometrics*, forthcoming.
[43] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley.