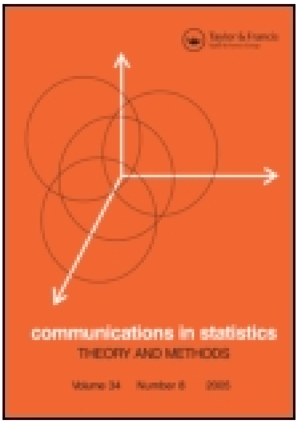

Communications in Statistics - Theory and Methods

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

Cointegration: Bayesian Significance Test

M. Diniz [a] , C. A. B. Pereira [b] & J. M. Stern [c]

[a] Department of Statistics, Universidade Federal de Sao, Carlos, Sao Carlos, Brazil

[b] Department of Statistics, Universidade de Sao Paulo, Sao Paulo, Brazil

[c] Department of Applied Mathematics, Universidade de Sao Paulo, Sao Paulo, Brazil

Communications in Statistics—Theory and Methods, 41: 3562–3574, 2012

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610926.2011.563021

Taylor & Francis
Taylor & Francis Group

Cointegration: Bayesian Significance Test

M. DINIZ[1], C. A. B. PEREIRA[2], AND J. M. STERN[3]

[1]Department of Statistics, Universidade Federal de Sao, Carlos, Sao Carlos, Brazil
[2]Department of Statistics, Universidade de Sao Paulo, Sao Paulo, Brazil
[3]Department of Applied Mathematics, Universidade de Sao Paulo, Sao Paulo, Brazil

To estimate causal relationships, time series econometricians must be aware of spurious correlation, a problem first mentioned by Yule (1926). To deal with this problem, one can work either with differenced series or multivariate models: VAR (VEC or VECM) models. These models usually include at least one cointegration relation. Although the Bayesian literature on VAR/VEC is quite advanced, Bauwens et al. (1999) highlighted that "the topic of selecting the cointegrating rank has not yet given very useful and convincing results".

The present article applies the Full Bayesian Significance Test (FBST), especially designed to deal with sharp hypotheses, to cointegration rank selection tests in VECM time series models. It shows the FBST implementation using both simulated and available (in the literature) data sets. As illustration, standard non informative priors are used.

Keywords Bayesian inference; Cointegration; Hypothesis testing; Reduced rank regression; Time series.

Mathematics Subject Classification Primary 62P20; Secondary 62F03, 62F15.

1. Introduction

One of the main goals of econometricians is to estimate causal relationships. However, since the seminal work of Yule (1926), it is known that spurious regression is a possible problem for time series analysis. This might arise when variables are integrated or have unit roots. One approach is to test the series for unit roots and then, if not rejected, to adopt the procedures to make them stationary, usually by differencing it. Alternatively, one may estimate cointegration relations between them. For the latter procedure, the most used instrument is the parameter estimation of Vector Error Correction Models (VECM).

Received December 8, 2010; Accepted February 4, 2011
Address correspondence to M. Diniz, Department of Statistics, Universidade Federal de Sao Carlos Rod., Washington Luis, KM 235, Sao Carlos, Brazil; E-mail: marcio.alves.diniz@gmail.com

Recently, the authors applied the Full Bayesian Significance Test (FBST) to unit root tests, (Diniz et al., 2007). Following the same path, we apply the FBST to test for cointegration. For more on FBST general aspects and properties, see Pereira and Stern (1999) and Pereira et al. (2008).

The Bayesian literature on cointegration is vast, especially articles concerned with inferences for cointegration space parameters. They usually assume some cointegration rank value and the inference is made conditionally on this value.

On the other hand, the research concerning Bayesian cointegration tests has progressed only recently. Most proposed tests are based on the Bayes Factor. This approach has well-documented shortfalls. As highlighted by Bauwens et al. (1999), using posterior odds approach leads to rather heavy computation and requires specification of proper prior densities; see Kleibergen and Paap (2002), Villani (2005), and Sugita (2002). Since we are dealing with sharp hypotheses, our main contribution here is the implementation of the FBST for cointegration rank selection. Note that we do not have to attach a prior probability to null measure sets (the sharp hypothesis submanifold) and can use a standard non-informative prior: not always possible when working with Bayes Factors.

Section 2 presents alternative Bayesian solutions described in the literature. In Sec. 3, we apply the FBST on cointegration rank selection using both simulated and real data. Finally, crucial remarks are listed in Sec. 4.

2. Alternative Bayesian Tests

This section briefly summarizes the most recent developments made by the Bayesian School for cointegration tests. The discussion made here closely follows Bauwens et al. (1999) and Koop et al. (2006).

The pioneer Bayesian works to approach the VAR models and reduced rank regressions are DeJong (1992), Bauwens and Lubrano (1996), and Geweke (1996). Geweke (1996) discussed Bayesian methods to test for the rank. He proposed the use of the *predictive odd ratios* that do not require a proper prior, but this could be used if desired, and the traditional Bayes factor is a special case of it.

When the main concern of the research is estimating parameters and their posterior distributions usually, the long run matrix rank is assumed known and the calculus made conditional on that. The reference for the Bayesian inference on VECM is Koop et al. (2006).

To justify the inference that assumes known values for the long run matrix rank, Bauwens et al. (1999) argued that an empirical cointegration analysis should be based on some economic theory model that defines equilibrium relations. Taking this into account, cointegration research is "confirmatory" rather than "exploratory." Although the inference conditioned on some given value of the cointegration matrix rank is simple and very useful for small samples, it is important to develop inference procedures to evaluate the cointegration matrix rank, as proposed by Geweke (1996), Kleibergen and Paap (2002), Sugita (2002), Strachan (2003), Strachan and Inder (2004), and Villani (2005).

The Bayesian approach based on Bayes Factors leads to complex calculus and requires the use of proper priors; see Kleibergen and Paap (1997). Bauwens and Lubrano (1996), using an informal approach, calculated the posterior distribution of the ordered eigenvalues of the square of the long-run matrix of VECM, obtained from a VAR model without a cointegration hypothesis. As the long-run matrix has a reduced rank, it has some null eigenvalues. This should be revealed by the fact that the smallest eigenvalues should have a lot of probability mass accumulated on values next to zero. The calculus can be made straightforwardly, simulating values for the long-run matrix from its posterior, which is a matricvariate Student under the non informative prior, also considered in the sequel.

Moreover, it is usual to choose the posterior mode as an estimate for the long-run matrix rank. Conditioned in this value, one proceeds to estimate the matrix itself. Chao and Phillips (1999) used the Posterior Information Criterion (PIC), developed by Phillips (1996), to choose the mode of the long run matrix rank. However, as highlighted by Koop et al. (2006), one of the advantages of the Bayesian approach is the possibility to incorporate the uncertainty about the parameters in the analysis, represented by the posterior distribution of the rank: whatever choice the scientist makes to infer the rank value, he/she must know its posterior distribution.

Kleibergen and Paap (2002) nested the reduced rank model in an unrestricted VAR and used Metropolis-Hastings sampling with the Savage-Dickey density ratio1 to estimate Bayes Factors for the model with incomplete rank up to the model with a full rank. The Bayes Factor derivation requires the estimation of an error correction factor for the incomplete rank. This factor, however, is not defined for an improper prior due to the Bartlett paradox. This problem arises whenever one compares models of different dimensions. It is relevant in the present case because after deriving the rank posterior density, it is possible to understand that we are comparing models of different dimensions. The paradox is enounced informally as: improper priors should be avoided when one calculates Bayes Factors (except for common parameters to both models) as they depend on arbitrary constants (integrals).

Recently, Villani (2005) developed an efficient procedure to obtain the rank posterior density using a uniform proper prior over the cointegration space linearly normalized. The author derived solutions for the posterior probabilities for the null rank and for the full rank. The posterior probabilities for each intermediate rank value are calculated from the posterior samples of the matrices that compose the long run matrix, properly normalized, under each rank and using the method proposed by Chib and Greenberg (1995).

3. FBST

The FBST was introduced by Pereira and Stern (1999). It was created mainly to test sharp hypotheses which is a matter of discussion and controversies. This article assumes that one accept and is interested in testing sharp hypotheses.

Let us now consider general statistical spaces, where the parameter space is $\Theta \subset \mathcal{R}^m$ and the sample space $X^k \subset \mathcal{R}$. A sharp hypothesis H states that θ belongs to a sub-manifold Θ_H of smaller dimension than Θ. The subset Θ_H has null Lebesgue measure whenever H is sharp.

In the FBST construction the posterior probability density on the parameter space is used as an ordering system and all sets of the same nature are dealt with accordingly in the same way. As a consequence, the sets that define sharp hypotheses keep having null probabilities. Instead of changing the nature of H by assigning positive probability to it, we consider the tangential set T of points having posterior density values higher than any θ in Θ_H. We then reject H if the posterior probability of T is large. We will formalize these ideas in the sequel.

Let us consider a standard parametric statistical model: $\theta \in \Theta \subset \mathcal{R}^m$ is the parameter, $g(\cdot)$ a probability prior density over Θ, x is the observation (a scalar or a vector), and $L_x(\cdot)$ is the likelihood given the data x. Posterior to the observation of x, the sole relevant entity for the evaluation of the Bayesian evidence, ev, is the posterior probability density for θ given x, denoted by

$$g_x(\theta) = g(\theta|x) \propto g(\theta)L_x(\theta).$$

We are, of course, restricted to the case where the posterior probability distribution over Θ is absolutely continuous, that is, $g_x(\theta)$ is a density over Θ. For simplicity, we may use H for Θ_H in the sequel. Now, let $r(\theta)$ be a reference density on Θ such that the function $s(\theta) = g_x(\theta)/r(\theta)$ is called the "relative surprise."[1]

Definition 3.1 (Evidence). Consider a sharp hypothesis $H : \theta \in \Theta_H$ and let

$$s^* = \sup_{\theta \in H} s(\theta) \tag{1}$$

and

$$T = \{\theta \in \Theta : s(\theta) > s^*\}. \tag{2}$$

The Bayesian evidence value against H is defined as the posterior probability of the tangential set, i.e., $\overline{ev} = Pr\{\theta \in T|x\} = \int_T g_x(\theta)d\theta$.

Notice that the tangential set T is the highest relative surprise set. It is the set of points $\theta \in \Theta$ with higher relative surprise $s(\theta)$ than any point in H. Therefore, the set is "tangential" to H. This approach does not exclude or avoid the model considered in the hypothesis being tested but just uses the concept of "tangent" to define an evidential measure favouring the hypothesis.

One must also note that the evidence value supporting H, $ev = 1 - \overline{ev}$, is not evidence against A, the alternative hypothesis (which is not sharp anyway). Equivalently, $\overline{ev}$ is not evidence in favor of A, although it is against H.

Definition 3.2 (TEST). The FBST (Full Bayesian Significance Test) is the procedure that rejects H whenever $ev = 1 - \overline{ev}$ is smaller than a critical level, ev_c.

Being a statistic, ev has a sampling distribution. For well-behaved likelihood and posterior densities,[2] Pereira et al. (2008) showed that, asymptotically, the evidence follows a χ^2 distribution with degrees of freedom given by the dimension of the parameter space. This fact gives a way to define, at least asymptotically, a critical level to reject the hypothesis being tested.

A major practical issue for the use of the FBST is the determination of the critical level. Ev being a statistic defined on a zero to one scale does not ease the matter (the same occurs with p-values). The formal identification of the FBST as a Bayes test of hypothesis yields critical values derived from loss functions allowing this identification. In fact, Madruga et al. (2001) showed that there are loss functions

[1]See Good (1983).
[2]See Schervish (1995), p. 436.

the minimization of which makes ev a Bayes estimator of $\phi = I(\theta \in H)$. Hence, the FBST is in fact a Bayes procedure in the formal sense of Wald (1950).

By using a reference density in the definition of the tangential set T, the FBST formulation above presented is explicit invariant under general coordinate transformations of the parameter space.[3] For the FBST application on unit root tests discussed in the sequel, we will use the (improper) uniform density as reference density on Θ. Madruga et al. (2003) remarked that it is possible to generalize the procedure using other reference densities such as neutral or reference priors if one is available.

3.1. *Numerical Calculus*

The calculus of the evidence value supporting H defined in the last section is performed numerically in two steps. The first one involves the optimization of $s(\theta)$ under H and, the second one, the integration of the posterior, $g_x(\theta)$, over T.

The optimization step consists of finding the parameter space point in H that maximizes $s(\theta)$. It is, therefore, a maximization under constraint problem:

$$\theta^* = \arg\max_{\theta \in \Theta_H} s(\theta), \quad s^* = s(\theta^*).$$

To solve this problem we use a numerical optimizer. To calculate the integral, it is possible to use various numerical techniques. We introduce a method based on Laplace approximation that is able to deal with a great number of the problems discussed in the literature, specially when the MCMC sampling is burdensome.

Let θ be the parameter vector and x the observations vector as above. The posterior distribution is given by:

$$g(\theta \mid x) = \frac{g(\theta)L_x(\theta)}{\int_\Theta g(\theta)L_x(\theta)d\theta}.$$

To calculate the e-value we need to integrate the posterior over the tangential set, i.e., $T = \{\theta \in \Theta : s(\theta) \geq s^*\}$:

$$\int_{\{\theta \in \Theta \mid s(\theta) \geq s^*\}} g_x(\theta)d\theta = \frac{\int_T g(\theta)L_x(\theta)d\theta}{\int_\Theta g(\theta)L_x(\theta)d\theta}. \tag{3}$$

One way to approximate integrals like the denominator above is to use the Laplace approximation. Consider the integral:

$$I = \int_\Theta b(\theta)\exp[-Nh(\theta)]d\theta$$

in which N is the sample size, θ is $(k \times 1)$, $\Theta = \mathscr{R}^k$ and $-h(\cdot)$ is a twice differentiable function with only one maximum in $\hat{\theta}$, $\partial h(\theta)/\partial\theta|_{\theta=\hat{\theta}} = 0$, and $H(\theta) = \frac{\partial^2 h(\theta)}{\partial\theta\partial\theta'}$ is positive definite. Furthermore, $b(\cdot)$ is continuous on the neighborhoods of $\hat{\theta}$ with $b(\hat{\theta}) \neq 0$. Expanding $h(\theta)$ as a second order Taylor series for $\hat{\theta}$ we have an approximation

[3]See the Appendix.

of $\exp[-Nh(\theta)]$ proportional to a normal density. By doing the same with $b(\theta)$ we arrive at the following approximation for the above integral:

$$\widehat{I} = (2\pi)^{k/2} b(\hat{\theta}) |NH(\hat{\theta})|^{-1/2} \exp[-Nh(\hat{\theta})]$$

since the $O(N^{-1/2})$ terms from the expansions of $b(\theta)$ and $h(\theta)$ disappear when we integrate.

Now we use the Tierney and Kadane (1986) method to calculate (3). Let us consider $b(\theta) = 1$ and the restriction $\exp[-Nh(\theta)] = g(\theta)L_x(\theta)$. If $h(\theta)$ satisfy the conditons given above, we have that the value of (3) is approximated by:

$$= \frac{\exp[-Nh(\hat{\theta})] \int_{T(\theta)} \exp[-\frac{N}{2}(\theta - \hat{\theta})' H(\hat{\theta})(\theta - \hat{\theta})] d\theta}{\exp[-Nh(\hat{\theta})](2\pi)^{k/2} |NH(\hat{\theta})|^{-1/2}}$$
$$= (2\pi)^{-k/2} |NH(\hat{\theta})|^{1/2} \int_{T(\theta)} \exp\left[-\frac{N}{2}(\theta - \hat{\theta})' H(\hat{\theta})(\theta - \hat{\theta}) \right] d\theta.$$

The last expression is the integral over the tangential set of the θ multivariate normal density with mean $\hat{\theta}$ and variance $(NH(\hat{\theta}))^{-1}$. Therefore, to evaluate the integral we can generate a large number of vectors with this distribution and evaluate the posterior with these vectors. The proportion of them that belongs to T is the approximate value for (3).

4. FBST Cointegration Applications

In this section we present how we implement the FBST to test for cointegration. First, we show some examples with simulated data and close with two real data sets used by the time series literature. We also clarify some aspects about the procedure implementation.

Another important observation should be made about the hypotheses being tested. As mentioned by Diniz et al. (2007), the FBST was designed to deal with sharp hypotheses without the need of attaching probabilites to sets with a null Lebesgue measure. Therefore, its use is appealing especially when the parameter space is continuous. At first sight, the parameter space for the cointegration tests, the rank of matrix Π, is discrete. However, we must remember that this rank is associated with the eigenvalues of the same matrix and they assume continuous values. It is also important to mention that, for the examples here presented, we used $r(\theta)$, as reference density, the improper density proportional to the uniform over the parameter space (non informative prior density).

4.1. *FBST as a Cointegration Test*

Let a vector with n $I(d)$ series that we want to know if are cointegrated, $Y_t = [y_{1t} \dots y_{nt}]'$ and $t = 1, \dots, T$. Consider that its data generating process is a VAR(p), as:

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + E_t. \tag{4}$$

in which $E_t \sim NI_n(0, \Sigma)$. This can be written as an error correction model:

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1}\Delta Y_{t-p+1} + \Pi Y_{t-1} + E_t, \tag{5}$$

where $\Delta Y_t = [\Delta y_{1t} \ldots \Delta y_{nt}]'$, $\Gamma_i = -(\Phi_{i+1} + \cdots \Phi_p)$ for $i = 1, 2, \ldots, p$ and $\Pi = -(I_n - \Phi_1 - \cdots - \Phi_p)$.

Using matrix notation, the error correction model (5) can be written as:

$$\Delta Y = ZB + E, \tag{6}$$

where

$$\Delta Y = \begin{bmatrix} \Delta Y'_{p+1} \\ \vdots \\ \Delta Y'_T \end{bmatrix}, \quad Z = \begin{bmatrix} \Delta Y'_p & \ldots & \Delta Y'_2 & Y'_p \\ \Delta Y'_{p+1} & & \Delta Y'_3 & Y'_{p+1} \\ \vdots & & \vdots & \vdots \\ \Delta Y'_{T-1} & \ldots & \Delta Y'_{T-p+1} & Y'_{T-1} \end{bmatrix}, \quad B = \begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_{p-1} \\ \Pi \end{bmatrix}$$

and the error vector is given by $E \sim MN_{T\times n}(0, \Sigma \otimes I_T)$, denoting the matricvariate normal distribution.[4]

Considering Eq. (6) and a non informative prior

$$p(B, \Sigma) \propto |\Sigma|^{-(n+1)/2},$$

we have that the likelihood and the posterior are, respectively:

$$\begin{aligned} p(\Delta Y|B, \Sigma, Z) &\propto |\Sigma|^{-T/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}(\Delta Y - ZB)'(\Delta Y - ZB)]\right\} \\ p(B, \Sigma|\Delta Y, Z) &\propto |\Sigma|^{-(T+n+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}(\Delta Y - ZB)'(\Delta Y - ZB)]\right\} \\ &= |\Sigma|^{-(T+n+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\{\Sigma^{-1}[S + (B-\widehat{B})'Z'Z(B-\widehat{B})]\}\right\}, \\ &\propto f_{MN}^{n\times k}(B|\widehat{B}, \Sigma \otimes (Z'Z)^{-1}) f_{WI}^{n}(\Sigma|S, T) \end{aligned} \tag{7}$$

where $\widehat{B} = (Z'Z)^{-1}Z'\Delta Y$, $S = \Delta Y'\Delta Y - \Delta Y'Z(Z'Z)^{-1}Z'\Delta Y$ and k the dimension of B, which is $k \times n$. The likelihood and the posterior just presented are for the linear error correction model, which is equivalent to the posterior of a multivariate linear regression model. That is, unlike in the cointegration model, the matrix Π is assumed to have full rank.

It is also important to mention that marginal posterior of B, obtained after integrating out Σ in (7), is a matrix-t distribution. Therefore, the Laplace approximations described in Sec. 3 are not needed since to calculate evidence value, in this case, one simply needs to get draws from the posterior of (B, Σ) and count the proportion of times that the surprise function is higher than s^*.

[4]It is said that the vector $X \sim MN_{p\times q}(\mathrm{vec}M, Q \otimes P)$ if and only if $\mathrm{vec}(X)$ has a multivariate normal distribution, i.e., $\mathrm{vec}(X) \sim N_{pq}(\mathrm{vec}M, Q \otimes P)$.

The posterior (7) is used by the FBST to test the rank of matrix Π, $\rho(\Pi)$. To exemplify, consider a bi-dimensional vector Y_t generated by a VAR(1):

$$Y_t = \Phi_1 Y_{t-1} + E_t, \tag{8}$$

where $E_t \sim NI_2(0, \Sigma)$, and the same model written in the error correction form:

$$\Delta Y_t = \Pi Y_{t-1} + E_t. \tag{9}$$

We want to test, for instance, $H_0 : \rho(\Pi) = 1$. To implement the FBST we have to find the posterior maximum under the space of the hypothesis being tested and then to integrate the posterior over the tangential set. In this case, we assume that Π has a reduced rank and decompose it in two matrices of rank one and dimension 2×1, α and β, in accordance to the Granger representation theorem, implying that $\Pi = \alpha\beta'$. In order to define the restricted posterior (under H_0) in this situation, we write:

$$\Pi = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} \end{bmatrix}$$

and maximize the posterior under this restriction. However, since α and β are not identified in the likelihood, we can use the method proposed by Johansen (1988) for finding the maximum of the posterior under the restriction. We adopted the linear normalization restriction for all the examples below described.

To test $\rho(\Pi) = 0$ is sufficient to test if Π is the null matrix. When the vector Y_t has n components, the approach is automatically extended. In the following examples we compare the results given by the FBST with the maximum eigenvalue test. The reported p-values are asymptotic.

Example 4.1. We simulated a 2-dimension VAR(1) with 50 observations, errors $NI_2(0, \Sigma)$ and the following parameters:

$$\Phi_1 = \begin{bmatrix} 0.8 & 0.1 \\ 1 & 0.5 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}.$$

The Φ_1 matrix has eigenvalues equal to 1 and 0.3. Therefore, there is one cointegration vector. The test for $\rho(\Pi) = 0$ shows an e-value of 0.00428, leading us to reject the hypothesis. By testing $\rho(\Pi) = 1$ we obtain an e-value of 0.99686, and this confirms the existence of one cointegration vector. The maximum eigenvalue test to $H_0 : r = 0$ against $H_1 : r = 1$ reports a p-value close to zero and to $H_0 : r = 1$ against $H_1 : r = 2$, p-value of 0.4746, reaching the same conclusion obtained by the FBST, i.e., that there is one cointegration vector.

Example 4.2. We simulated another 2-dimension VAR(1) with 50 observations and the following parameters:

$$\Phi_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \Sigma \text{ like above.}$$

The matrix Φ_1 has both eigenvalues equal to one. Therefore, the series are $I(1)$ and do not cointegrate. The FBST to test $\rho(\Pi) = 0$ gives an e-value equal to 0.4586, which is substantial evidence to not reject the hypothesis, as expected. The maximum eigenvalue test to $H_0 : r = 0$ against $H_1 : r = 1$ presents a p-value of 0.3889 and, therefore, the null is not rejected.

Example 4.3. Now we present a 3-dimension VAR(1) with 50 observations and the following parameters:

$$\Phi_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 1 \end{bmatrix}.$$

The Φ_1 matrix has eigenvalues equal to 1, 0.5, and 0.3. Therefore, there are two cointegration vectors. The FBST to test $\rho(\Pi) = 0$ gives us an e-value of 0.0151, and it is possible to reject it. After testing $\rho(\Pi) = 1$ we found an e-value of 0.0342, and also rejected the hypothesis. Testing $\rho(\Pi) = 2$ the *e*-value is 0.9991, which confirms the existence of two cointegration vectors.

The maximum eigenvalue test reports a p-value close to zero for $H_0 : r = 0$ against $H_1 : r = 1$ and to $H_0 : r = 1$ against $H_1 : r = 2$. The test of $H_0 : r = 2$ against $H_1 : r = 3$ presents a p-value of 0.11489. Therefore, we do not reject this last hypothesis and conclude that there are two cointegration vectors.

Example 4.4. We simulated a 2-dimension VAR(2) with the following generating process:

$$Y_t = \begin{bmatrix} 0.45 & -0.2 \\ 1.1 & 0.3 \end{bmatrix} Y_{t-1} + \begin{bmatrix} 0.35 & 0.3 \\ -0.1 & 0.2 \end{bmatrix} Y_{t-2} + E_t.$$

where $\Sigma = \left[\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{smallmatrix}\right]$. We know that there is one cointegration vector since the Π matrix has eigenvalues equal to 1 and 0.3.

The FBST reports an e-value of 0.0276 testing $H_0 : r = 0$, leading us to reject it. The maximum eigenvalue test, testing this null against $H_1 : r = 1$ presents a p-value of 0.0003. To test $H_0 : r = 1$, the FBST reaches an e-value of 0.9972 and the maximum eigenvalues test, against $H_1 : r = 2$, a p-value of 0.4392.

Example 4.5 (Johansen and Juselius, 1990)**.** Now we apply the FBST to the Finish data set used by Johansen and Juselius in their seminal work.

The authors used the series in natual logarithms of the $M1$ monetary aggregate, inflation rate, real income, and the primary interest rate set by the Bank of Finland to model the money demand which, in theory, follows a long term relation. The sample has quarterly observations starting at 1958:02 and goes until 1984:03. The chosen model was a VAR(2) with unrestricted constant and seasonal dummies for the first three quarters of the year. Writing the chosen model in the error correction form, we have:

$$\Delta Y_t = \mu + \Psi D_t + \Gamma_1 \Delta Y_{t-1} + \Pi Y_{t-1} + E_t, \tag{10}$$

where $\Pi = \Phi_1 + \Phi_2 - I$, $\Gamma_1 = -\Phi_2$, μ is the constants vector, and D_t has the seasonal dummies. This vector could also contain other deterministic variables. It is assumed that $E_t \sim NI_4(0, \Sigma)$.

Table 1
FBST and maximum eigenvalue test applied to Finish data of Johansen and Juselius (1990)

H_0	FBST	λ_{max}	p-value
$r = 0$	0.0006	38.489	0.0007
$r = 1$	0.0505	26.642	0.0060
$r = 2$	0.9726	7.8924	0.3983

Table 2
FBST and maximum eigenvalue test applied to U.S. data of Lucas (2000)

H_0	FBST	λ_{max}	p-value
$r = 0$	0.0145	25.334	0.0101
$r = 1$	0.9299	4.2507	0.8271

To implementate the FBST in this case, we used the Frisch-Waugh-Lovell theorem. Thus, we run the auxiliary regressions:

$$\Delta Y_t = \mu' + \Psi' D_t + \Gamma_1' \Delta Y_{t-1} + R_{0,t}$$
$$Y_{t-1} = \mu^* + \Psi^* D_t + \Gamma_1^* \Delta Y_{t-1} + R_{1,t}$$

and then we can work only with the residual vectors to study the rank of Π, since the theorem assures that

$$R_{0,t} = \Pi R_{1,t-1} + E_t,$$

being E_t the same of (10). The results are reported in Table 1.

The authors concluded that there is, at least, two cointegration vectors, the same conclusion reached by the FBST.

It is relevant to consider that, for a non informative prior, the value of s^* can be calculated using the techniques proposed by Johansen (1988). He gives analytical formulas for the maximum likelihood estimate in the cointegration model and therefore, there is no need for using the Frisch-Waugh-Lovell theorem. Therefore, it would not be necessary to use a numerical optimizer as claimed in Sec. 3.1.

For the models and examples here presented, the methods in Johansen (1988) would be a much more reliable way to obtain the maximum. Still, it is worthwhile mentioning that a numerical optimizer might be used in the case that an informative prior is used. However, if one is going to use a numerical optimizer it is necessary to use a normalization restriction. We stress again that we used the linear normalization restriction.

Example 4.6 (Lucas, 2000)**.** We apply, as a last example, the FBST to an US data set used by Lucas (2000). The observations have annual periodicity and went from 1900–1985. We tested for cointegration between real national income, $M1$ monetary aggregate deflated by the GDP deflator and the commercial papers return rate.

We adjusted a VAR(1) with an unrestricted constant. The data are used in natural logarithm and the results are reported in Table 2.

5. Concluding Remarks

In the past few decades, the econometric literature introduced statistical tests to identify unit roots and cointegration relations in time series. The Bayesian approach applied to these topics advanced considerably and interesting alternatives were developed. Following Geweke (1996), Kleibergen and Paap (2002), Sugita (2002), Strachan (2003), Strachan and Inder (2004), and Villani (2005), the present work introduces a simple and powerful procedure that performs well for this task of identifying the cointegration rank. It is shown that the FBST works considerably well even when one uses improper priors. Recall that the use of a non-informative prior of this kind may cause problems for the use of Bayes Factors, the standard Bayes hypotheses tests.

To apply classical (frequentist) statistical tests one needs hard simulations to find critical values: no closed analytical forms are available. These simulations depend on restrictions in order to obtain the asymptotic distributions of the statistics used.

If the researcher is working with samples that do not have critical values "tabulated," then either the asymptotical approximation or the closest sample size, for which the critical values were calculated, are considered. This can be a problem especially for small samples: the simulations assume that the data follow given distributions, usually the gaussian.

The FBST does not need any restriction concerning sample sizes and can be calculated assuming any parametric sampling distribution. The FBST can conduct the conduction of the test without restrictions to sample sizes, prior distributions or error sampling distributions; it is in fact an exact test. Comparing with the standard Bayes Factor based tests; the FBST does not have to assume positive probability to any set of nil Lebesgue measure. The most important feature of the FBST is the fact that its use does not violate the likelihood principle. As another contribution, in Sec. 3, the article presents a useful numerical approach to calculate e-values when MCMC sampling is burdensome, even though they were not necessary for the models here presented.

For future work the authors will investigate the power of FBST, based on simulation studies, to compare it with the Johansen procedure. Clearly, the FBST is not restricted to noninformative priors: we shall investigate the effect of the prior choice in the estimates of cointegration relations.

Appendix

The definition of the evidence against some sharp hypothesis H given in Sec. 3 is invariant with respect to a proper reparameterization. For instance, let $\omega = \phi(\theta)$ where $\phi(\cdot)$ is a measurable and integrable function. For the purpose of illustration, assume that it is bijective and continuously differentiable. Under the reparameterization, the Jacobian, surprise, posterior, and e-value (against the hypothesis) are, respectively, $J(\omega)$, $\tilde{r}(\omega)$, $\tilde{g}_x(\omega)$, and $\widetilde{\overline{ev}}(H)$, given by:

$$J(\omega) = \left[\frac{\partial\theta}{\partial\omega}\right] = \left[\frac{\partial\phi^{-1}(\omega)}{\partial\omega}\right] = \begin{bmatrix} \frac{\partial\theta_1}{\partial\omega_1} & \cdots & \frac{\partial\theta_1}{\partial\omega_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial\theta_n}{\partial\omega_1} & \cdots & \frac{\partial\theta_n}{\partial\omega_n} \end{bmatrix}$$

$$\tilde{s}(\omega) = \frac{\tilde{g}_x(\omega)}{\tilde{r}(\omega)} = \frac{g_x(\phi^{-1}(\omega))|J(\omega)|}{r(\phi^{-1}(\omega))|J(\omega)|}.$$

Let $\Omega_H = \phi(\Theta_H)$. It follows that

$$\tilde{s}^* = \sup_{\omega\in\Omega_H} \tilde{s}(\omega) = \sup_{\theta\in\Theta_H} s(\theta) = s^*;$$

hence, the tangential set under the reparameterization is, $T \mapsto \phi(T) = \widetilde{T}$, and

$$\widetilde{\overline{ev}}(H) = \int_{\widetilde{T}} \tilde{g}_x(\omega)d\omega = \int_T g_x(\theta)d\theta = \overline{ev}(H).$$

We remark that the FBST is also invariant with respect to the null hypothesis parameterization. This is not a trivial issue because some statistical procedures do not satisfy this property. The reader interested in a broader discussion of the FBST properties can see Madruga et al. (2003) and Pereira et al. (2008).

References

Bauwens, L., Lubrano, M. (1996). *Advances in Econometrics*. Vol. 11, part B. London: JAI Press, pp. 3–28.

Bauwens, L., Lubrano, M., Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.

Chao, J., Phillips, P. C. (1999). Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* 91:227–271.

Chib, S., Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *Amer. Statistician* 49:327–335.

DeJong, D. (1992). Co-integration and trend-stationary in macroeconomic time series. *J. Econometrics* 52:347–370.

Diniz, M., Pereira, C. A. B., Stern, J. M. (2007). FBST for unit root problems. *AIP Conf. Proc.* 954:260–267.

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

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and its Applications*. Minneapolis: University of Minnesota Press.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *J. Econ. Dyn. Cont.* 12:231–254.

Johansen, S., Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration-with application to the demand for money. *Oxford Bulletin of Economics and Statistics* 52:169–210.

Kleibergen, F., Paap, R. (1997). *Priors, Posterior Odds and Lagrange Multiplier Statistics in Bayesian Analyses of Cointegration*. Discussion paper, Tinbergen Institute.

Kleibergen, F., Paap, R. (2002). Priors, posterior odds and bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics* 111:223–249.

Koop, G., Strachan, R., van Dijk, H. K., Villani, M. (2006). *The Palgrave Handbook of Theoretical Econometrics*. Palgrave McMillan, Ch. 25.

Lucas, R. (2000). Inflation and welfare. *Econometrica* 68:247–274.

Madruga, M. R., Esteves, L. G., Wechsler, S. (2001). On the Bayesianity of Pereira-Stern tests. *Test* 10:291–299.

Madruga, M. R., Pereira, C. A. B., Stern, J. M. (2003). Bayesian evidence test for precise hypotheses. *J. Stat. Plann. Infer.* 117:185–198.

Pereira, C. A. B., Stern, J. M. (1999). Evidence and credibility: full bayesian significance test for precise hypotheses. *Entropy J.* 1:69–80.

Pereira, C. A. B., Stern, J. M., Wechsler, S. (2008). Can a significance test be genuinely bayesian. *Bayes. Anal.* 1:79–100.

Phillips, P. C. (1996). Econometric model determination. *Econometrica* 59:283–306.

Schervish, M. (1995). *Theory of Statistics*. New York: Springer.

Strachan, R. (2003). Valid Bayesian estimation of the cointegrating error correction model. *J. Bus. Econ. Statist.* 21:185–195.

Strachan, R., Inder, B. A. (2004). Bayesian analysis of the error correction model. *J. Econometrics* 123:307–325.

Sugita, K. (2002). Testing for cointegration rank using Bayes factors. The Warwick Economics Research Paper Series, Working Paper no. 654.

Tierney, L., Kadane, J. B. (1986). Accurate approximation for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* 81:82–86.

Verdinelli, I., Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *J. Amer. Statist. Assoc.* 90:614–618.

Villani, M. (2005). Bayesian reference analysis of cointegration. *Econometric Theory* 21:326–357.

Wald, A. (1950). *Statistical Decision Functions*. New York: John Wiley and Sons; London: Chapman and Hall.

Yule, G. U. (1926). Why we do sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series (with discussion). *J. Roy. Statist. Soc.* 89:1–64.