Testing for Cointegration Rank Using Bayes Factors

Testing for Cointegration Rank Using Bayes Factors

Katsuhiro Sugita

Abstract

This paper proposes Bayesian methods for estimating the cointegration rank using Bayes factors. We consider natural conjugate priors for computing Bayes factors. First, we estimate the cointegrating vectors for each possible rank. Then, we compute the Bayes factors for each rank against 0 rank. Monte Carlo simulations show that using Bayes factor with conjugate priors produces fairly good results. The methods proposed here are also applied for selecting the appropriate lags and testing for over-identifying restrictions on cointegrating vectors.

Key words: Cointegration; MCMC; Bayes factor

1 Introduction

This paper introduces Bayesian analysis of cointegrated VAR systems. Several researchers have proposed Bayesian approach to cointegrated VAR systems. These include Kleibergen and van Dijk (1994), who proposed using a Jeffrey's prior instead of diffuse prior for the cointegrating vectors since the marginal posteriors may be nonintegrable with reduced rank of cointegrating vector. Geweke(1996) developed general methods for Bayesian inference with noninformative reference priors in the reduced rank regression model. Kleibergen and Paap (1999) use a singular value decomposition of the unrestricted long-run multiplier matrix, Π, for identification of the cointegrating vectors and for Bayesian posterior odds analysis of the rank of Π. Bauwens and Lubrano (1996) reduce the ECM to the form of a multivariate regression model to identify the parameters. The cointegrating rank is assumed to be known a priori, based on a theoretical economic model that defines equilibrium economic relations. If we are interested in identifying the cointegration rank, they suggest checking the plot of the posterior density of the eigenvalues of generated sample $\Pi'\Pi$, which are equal to the square of the singular values of Π. However, this informal visual inspection gives ambiguous results.[1] Bauwens, *et al* (1999) suggest using the trace test of Johansen, since "on the Bayesian side, the topic of selecting the cointegrating rank has not yet given very useful and convincing results"(p.283).

In this paper we propose a simple method of determining of the cointegration rank by Bayes factors. The method is very straightforward. We consider using the conjugate priors for all parameters. If there exist r cointegrating vectors in the system, the adjustment term α has rank r. Applying the Bayes factor to α, which has r rank, against the null of α, which has rank 0, for each rank gives the posterior probabilities for its rank. The procedure for obtaining the posteriors has some similarities with Bauwens and Lubrano method. However, instead of using diffuse priors for all parameters, the conjugate priors are chosen to be able to compute the Bayes factors.

The plan of this paper is as follows. Section 2 presents the prior specifications and derives the posterior densities for estimation of the cointegrated VAR systems. In Section 3 the Bayes factors for cointegration rank is introduced. Section 4 illustrates Monte Carlo simulations with DGPs of 1000 iterations for each rank to compare the performance of the proposed Bayesian methods with the classical Johansen test. In Section 5, an illustrative example of the demand for money in the United Sates is presented. Section 6 concludes. All computations in this paper are performed using code written by the author with Ox v2.20 for Linux (Doornik, 1998).

1 Tsurumi and Wago (1996) use a highest-posterior-density-region (HPDR) test to Π, then derive the posterior pdfs for singular values to see whether 99% highest-posterior-density-interval (HPDI) contains zero.

2 Bayesian Inference in Cointegration Analysis

In this section we present Bayesian analysis of cointegration. Let X_t denote an $I(1)$ vector of n-dimensional time series with r linear cointegrating relations, then unrestricted VECM representation with deterministic trend is:

$$\Delta X_t = \mu + \alpha\gamma z + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \tag{1}$$

where $z = \left[X_{t-1}\, t\right]'$, $\gamma = \left[\beta'\, \delta\right]$, $t = p, p+1, \ldots, T$, p is the number of lags in VAR, and the errors, ε_t, are assumed $N(0, \Sigma)$ and independent over time. $\mu, \varepsilon, \Psi, \Sigma, \alpha$, and β are parameters of dimensions $n \times 1$, $n \times 1$, $n \times n$, $n \times n$, $n \times r$, and $n \times r$, respectively.

Equation (1) can be rewritten in matrix format as:

$$Y = X\Gamma + Z\gamma'\alpha' + E = WB + E \tag{2}$$

where

$$Y = \begin{bmatrix} \Delta X_p' \\ \Delta X_{p-1}' \\ \vdots \\ \Delta X_T' \end{bmatrix}, Z = \begin{bmatrix} X_{p-1}' & p \\ X_p' & p+1 \\ \vdots & \vdots \\ X_{T-1}' & T \end{bmatrix}, E = \begin{bmatrix} \varepsilon_p' \\ \varepsilon_{p+1}' \\ \vdots \\ \varepsilon_T' \end{bmatrix}, \Gamma = \begin{bmatrix} \mu' \\ \Psi_1' \\ \vdots \\ \Psi_{p+1}' \end{bmatrix},$$

$$X = \begin{bmatrix} 1 & \Delta X_{p-1}' & \cdots & \Delta X_{p-1}' \\ 1 & \Delta X_p' & \cdots & \Delta X_2' \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta X_{T-1}' & \cdots & \Delta X_{T-p+1}' \end{bmatrix}, \quad W = [X\ Z\gamma'], \quad B = \begin{bmatrix} \Gamma \\ \alpha' \end{bmatrix}.$$

Let m be the number of columns of Y, so that $m = T - p + 1$, then X is $m \times (1 + n(p-1))$, $\Gamma((1 + n(p-1)) \times n)$, $W(m \times k)$, where $k = 1 + n(p-1) + r$, and B $(k \times n)$. Thus, equation (2) represents the multivariate regression format of (1). This representation is a starting point. We then describe the prior and likelihood specifications in order to derive posteriors.

First, we consider the case of applying the conjugate priors for some parameters. Our strategy to select priors is to choose a conjugate prior only for the parameters that are used in computing Bayes factor. For other parameters except for the cointegrating vector, we consider non-informative priors.

The conjugate prior density for B conditional on covariance Σ follows a matrix-variate normal distribution with covariance matrix $\Sigma \otimes A^{-1}$ of the form

$$p(B|\Sigma) \propto |\Sigma|^{k/2} |A|^{n/2} \exp\left[-\frac{1}{2} \mathrm{tr}\{\Sigma^{-1}(B-P)'A(B-P)\}\right] \tag{3}$$

where A is $(k\times k)$ PDS and $P(k\times n)$, $k=n(p-1)+r+1$ (the number of columns in W). Choosing hyperparameters, P and A, in (3) should be careful since these have direct effect on the value of Bayes factor. Kleibergen and Paap (1996) choose The idea of using data in priors is similar to a g-prior proposed by Zellner (1986).

For the prior density for the covariance Σ in (2), we can assign an inverted Wishart

$$p(\Sigma)\propto|S|^{h/2}|\Sigma|^{-(h+n-1)}\exp\left\{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S)\right\} \tag{4}$$

where h is a degree of freedom, S an $n\times n$ PDS.

The prior for β can be given as a matrix-variate normal

$$\pi(\beta)\propto|\Omega|^{n/2}|H|^{r/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(\beta-\bar{\beta})'H(\beta-\bar{\beta})\right\}\right] \tag{5}$$

where $\bar{\beta}$ is a prior mean of β, Ω is $r\times r$ PDS, H is $n\times n$ PDS. Note that r^2 restrictions for identification are imposed on β, for example, $\beta'=(I_r\ \beta')$,[2] where $\beta_\star$ is $(n-r)\times r$ unrestricted matrix. If we assign r^2 restrictions on β as I_r, then only a part of β, $\beta_\star$, follows a matrix-variate normal.

If we assume that B, Σ and β are mutually independent, then the joint prior of the parameters in (2) is $p(B,\beta,\Sigma)\propto p(B|\Sigma)p(\beta)p(\Sigma)$ and thus can be derived as

$$p(B,\Sigma,\beta)\propto\pi(\beta)|A|^{n/2}|\Sigma|^{-\frac{k+h+n+1}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\{\Sigma^{-1}[S+(B-P)'A(B-P)]\}\right] \tag{6}$$

To derive the conditional posterior distributions, we need to derive the likelihood functions. The likelihood function for B,Σ,and β is given by:

$$L(Y|B,\Sigma,\beta)\propto|\Sigma|^{-t/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left\{\hat{S}+(B-\hat{B})'W'W(B-\hat{B})\right\}\right]\right\} \tag{7}$$

where $\hat{B}=(W'W)^{-1}W'Y$, and $\hat{S}=(Y-W\hat{B})'(Y-W\hat{B})$.

Next we derive the posteriors from the priors and the likelihood function specified above. The joint posterior distribution for the conjugate priors for α is proportional to the joint prior (6) times the likelihood function (7), thus we have

$$\begin{aligned}
&p(B,\Sigma,\beta|Y)\propto p(B,\Sigma,\beta)L(Y|B,\Sigma,\beta)\\
\propto\ &\pi(\beta)|A|^{\frac{n}{2}}|\Sigma|^{-(t+h+k+n+1)/2}\\
&\times\exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[S+(B-P)'A(B-P)+\hat{S}+(B-\hat{B})'W'W(B-\hat{B})\right]\right\}\right]\\
\propto\ &\pi(\beta)|\Sigma|^{-\frac{C}{2}}\exp\left[-\frac{1}{2}\left\{\Sigma^{-1}\left[S+\hat{S}+(P-\hat{B})'[A^{-1}+(W'W)^{-1}]^{-1}(P-\hat{B})\right.\right.\right.\\
&\left.\left.\left.+(B-B_\star)'A_\star(B-B_\star)\right]\right\}\right]\\
=\ &\pi(\beta)|\Sigma|^{-\frac{C}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[S_\star+(B-B_\star)'\ A_\star(B-B_\star)\right]\right\}\right]
\end{aligned} \tag{8}$$

2 The restrictions imposed on β need not to be I_r but can be any r^2 restrictions. See Bauwens and Lubrano (1996, page 14)

where $c=t+k+h+n+1$, $A_\star=A+W'W$, $B_\star=(A+W'W)^{-1}(AP+W'WB)$, and $S_\star=S+\hat{S}+(P-\hat{B})'[A^{-1}+(W'W)^{-1}]^{-1}(P-\hat{B})$.

From (8), the conditional posterior of Σ is derived as an inverted Wishart distribution, and the conditional posterior of B as a matrix-variate normal density with covariance, $\Sigma\otimes A_\star^{-1}$, that is,

$$p(\Sigma|\beta,Y)\propto|S_\star|^{t/2}|\Sigma|^{-(t+h+n+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S_\star)\right] \tag{9}$$

$$p(B|\Sigma,\beta,Y)\propto|A_\star|^{n/2}|\Sigma|^{-k/2}\exp\left[-\frac{1}{2}\mathrm{tr}\{\Sigma^{-1}(B-B_\star)'A_\star(B-B_\star)\}\right] \tag{10}$$

Thus, by multiplying (9) and (10), and integrating with respect to Σ, we obtain the posterior density of B conditional on β, which is a matrix-variate Student-t form,

$$p(B|\beta,Y)\propto|S_\star|^{t/2}|A_\star|^{n/2}|S_\star+(B-B_\star)'A_\star(B-B_\star)|^{-(t+k)/2} \tag{11}$$

The joint posterior of B and β can be derived by integrating (8) with respect to Σ,

$$p(B,\beta|Y)\propto\pi(\beta)|S_\star+(B-B_\star)'A_\star(B-B_\star)|^{-(t+h+k+1)/2} \tag{12}$$

By integrating (12) with respect to B we obtain the posterior density of the cointegrating vector β,

$$p(\beta|Y)\propto\pi(\beta)|S_\star|^{-(t+h+1)/2}|A_\star|^{-n/2} \tag{13}$$

The properties of (13) are not known, so that we have to resort to numerical integration techniques as Bauwens and Lubrano (1996) use importance sampling to compute poly-t posterior results on parameters. Other feasible methods are the Metropolis-Hastings algorithm and the Griddy-Gibbs sampling. The Metropolis-Hastings[3] algorithm requires assignment of a good approximating function, the *candidate-generating function*, to the posterior to draw random numbers, as importance sampling requires the importance function. Since the Griddy-Gibbs sampling method does not require such an approximation, we employ the Griddy-Gibbs sampler for estimation of the cointegrating vector as Bauwens and Giot (1998) use the sampler for estimation of two cointegrating vectors. The Griddy-Gibbs sampler that is proposed by Ritter and Tanner (1992) approximates the true cdf of each conditional distribution by a piecewise linear function and then sample from the approximations. The disadvantage of this sampling method is that we have to assign proper range and the number of the grid. The range should be chosen so that the generated numbers are not truncated.

3 For more details, consult Chen, *et al* (2000), Evans and Swartz (2000). For tutorial for the M-H algorithm, see Chib and Greenberg (1995).

3 Bayes Factors for Cointegration Tests

This section introduces the computation of the Bayes factors for cointegration rank. Subsection 3.1 describes briefiy the basic concept of the Bayes factors and some computation techniques. Subsection 3.2 presents the computation of the Bayes factors for cointegration rank.

3.1 Bayes Factors and Determination of Rank

The Bayes factor, which is defined as the ratio of marginal likelihood of null and alternative hypothesis, has been used for model selection. The Bayes factors can be used to construct posterior probabilities for all models that seemed plausible. In classical hypothesis test, one model represents the truth and the test is based on a pairwise comparison of the alternative. For a detailed discussion of the advantages of Bayesian methods, see Koop and Potter (1999). Kass and Raftery (1995) provide an excellent survey of the Bayes factor.

Suppose, with data Y and the likelihood functions with the parameters Θ, there are two hypotheses H_0 and H_1. The Bayes factor BF_{01} is defined as follows:

$$BF_{01} = \frac{Pr(Y|H_0)}{Pr(Y|H_1)}$$

$$= \frac{\int p(\Theta|H_0)L(Y|\Theta,H_0)d\Theta}{\int p(\Theta|H_1)L(Y|\Theta,H_1)d\Theta} \tag{14}$$

With the prior odds, defined as $\Pr(H_0)/\Pr(H_1)$, we can compute the posterior odds, which are

$$\text{PosteriorOdds}_{01} = \frac{Pr(H_0|Y)}{\Pr(H_1|Y)} = \frac{Pr(Y|H_0)}{Pr(Y|H_1)} \cdot \frac{Pr(H_0)}{Pr(H_1)} \tag{15}$$

When several models are being considered, the posterior odds yield the posterior probabilities. Suppose q models with $H_0, H_1, \ldots, H_{q-1}$ are being considered, and each hypotheses of $H_1, H_2, \ldots, H_{q-1}$ is compared with H_0. Then the posterior probability for model i under Hi is

$$\Pr(\text{Hi}|Y) = \frac{\text{PosteriorOdds}_{i0}}{\sum_{j=0}^{q} \text{PosteriorOdds}_{j0}} \tag{16}$$

where $\text{PosteriorOdds}_{00}$ is defined to be 1. These posterior probabilities are used to select the cointegrating rank, model selection, or as weights for forecasting.

There are several methods to compute the Bayes factors given in (14). For example, the Laplace approximation method (Tierney and Kadane, 1986), or using numerical

integration techniques such as importance sampling (Geweke, 1989) or the Metropolis-Hastings algorithm. See Kass and Raftery (1995) for details. Chib (1995) proposes a simple approach to compute the marginal likelihood from the Gibbs output.

In the case of nested models computation of the Bayes factor can be simplified by using the generalized Savage-Dickey density ratio, proposed by Verdinelli and Wasserman (1995), which is based on Dickey (1971). Suppose we wish to test the null $H_0: \xi = \xi_0$ versus $H_1: \xi \neq \xi_0$, where ξ can be scalar, vector, or matrix. With the condition that $p(\Phi | \xi_0) = p_0(\Phi)$, where $(\Phi, \xi) = \Theta$, then the Bayes factor can be computed with the Savage-Dickey density ratio

$$BF_{01} = \frac{p(\xi_0 | Y)}{p(\xi_0)} \tag{17}$$

The denominator, the marginal prior for ξ evaluated at $\xi = \xi_0$, in (17) is trivial to calculate. The numerator, the marginal posterior for ξ evaluated at $\xi = \xi_0$, can be calculated by integrating out the other parameters, such as:

$$\begin{aligned} p(\xi_0 | Y) &= \int p(\xi_0 | \Phi, Y) p(\Phi | Y) d\Theta \\ &\simeq \frac{1}{N} \sum_{i=1}^{N} p(\xi_0 | \Phi^{(i)}, Y) \end{aligned} \tag{18}$$

where $\Phi^{(i)}, i = 1, \ldots, N$, are sample draws from the posterior.

3.2 Bayes Factor for Cointegration Rank

The Bayes factors are used for model selection, and thus can also be used for rank selection of the cointegration. Kleibergen and Paap (1999) propose a cointegration rank test by Bayes factors using a decomposition derived from the singular value decomposition. In this subsection, we propose a much simpler method for rank test using Bayes factors.

In a cointegrated system with n variables which are $I(1)$, if there are r cointegrating vectors, then the error correction term Π has reduced rank of r. Π can be decomposed as products of α and β', both of which have reduced rank r. Since α is unrestricted and also is given a rank reduction when cointegration exists, we compute the Bayes factor for α that is *against* the null $\alpha = 0$ for determining the number of the rank using inverted form of (17) for each possible rank $(r = 0, 1, \ldots, n)$.

$$\begin{aligned} BF_{r0} &= \frac{\int\int\int\int p(\alpha, \beta, \Gamma, \Sigma) L(Y | \alpha, \beta, \Gamma, \Sigma) d\alpha d\beta d\Gamma d\Sigma}{(1/Cr) \int\int\int\int p(\alpha, \beta, \Gamma, \Sigma)|_{\mathrm{rank}(\alpha)=0} L(Y | \alpha, \beta, \Gamma, \Sigma)|_{\mathrm{rank}(\alpha)=0} d\beta d\Gamma d\Sigma} \\ &= \frac{p(\alpha' = 0_{r \times n} | Y)}{(1/C_r) p(\alpha' = 0_{r \times n} | Y)} \end{aligned} \tag{19}$$

where $C_r = \int\int\int\int p(\alpha, \beta, \Gamma, \Sigma)|_{\mathrm{rank}(\alpha)=0}\, d\beta\, d\Gamma\, d\Sigma$ is the correction factor that is required for reduction of dimension.

If there exists r cointegrating vectors, the Bayes factor for $\alpha'_{(r\times n)}$ in (19) is the most unlikely to be zero and thus should have the highest value in Bayes factors for other possible ranks. Note that the Bayes factor for rank 0 equals to 1. In case of no cointegration, the Bayes factors for $\alpha'_{(r\times n)}$, where $r \neq 0$, are less than 1. If we assign an equal prior probability to each cointegration rank, the posterior probability for each rank can be computed as in (16)

$$\Pr(r|Y) = \frac{BF_{r|0}}{\sum_{j=0}^{n} BF_{j|0}} \tag{20}$$

where $BF_{0|0}$ is defined as 1.

The posterior probabilities given by (20) can be used for solutions of the prediction, decision making and inference problems that take account of model uncertainty. Generally, the hypothesis that has the highest posterior probability can be selected as the 'best' model, only if it dominates the others. Otherwise, analyses will fail to take uncertainty into account.

To compute the Bayes factors using (19), we use (18) with samples from the posteriors. Since α' is a partitioned element of B in (2) and thus the prior for alpha is a matrix-variate Student-t distribution as shown in (3), so the numerator of (19) is:

$$p(\alpha' = 0_{(n\times r)}) = \pi^{-\frac{nr}{2}} |S|^{\frac{h}{2}} |A_{22.1}|^{n/2} \left\{ \prod_{j=1}^{n} \frac{\Gamma\left(\frac{h+r+1-j}{2}\right)}{\Gamma\left(\frac{h+1-j}{2}\right)} \right\} |S|^{-\frac{h+r}{2}} \tag{21}$$

where $A_{22.1} = A_{22} - A_{21}A_{11}^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, A_{11} is $((n(p-1)+1)\times(n(p-1)+1))$, A_{12} $(n(p-1)+1))$, $A_{21}(r\times(n(p-1)+1)$, $A_{22}(r\times r)$.

The posterior for alpha, which is a matrix-variate Student-t from (11), can be estimated by (18) as follows:

$$\begin{aligned} p(\alpha' = 0_{(n\times r)}|Y) &= \int p(\alpha' = 0|\beta, Y)\, p(\beta|Y)\, d\beta \simeq \frac{1}{N}\sum_{i=1}^{N} p(\alpha' = 0|\beta^{(i)}, Y) \\ &= \frac{1}{N}\sum_{i=1}^{N} \pi^{-\frac{nr}{2}} |S_{\star}^{(i)}|^{-\frac{t+h}{2}} |A_{\star 22.1}^{(i)}|^{\frac{n}{2}} \left\{ \prod_{j=1}^{n} \frac{\Gamma\left(\frac{t+h+r+1-j}{2}\right)}{\Gamma\left(\frac{t+h+1-j}{2}\right)} \right\} \\ &\quad \times |S_{\star}^{(i)} + B_{\star 2}^{(i)\prime} A_{\star 22.1}^{(i)} B_{\star 2}^{(i)}|^{-\frac{i+r}{2}} \end{aligned} \tag{22}$$

where $\beta^{(i)}$, $S_{\star}^{(i)}$, and $A_{\star}^{(i)}$ are obtained from the i th iteration of the Gibbs sampler, $A_{\star 22.1} = A_{\star 22} - A_{\star 21}A_{\star 11}^{-1}A_{\star 12}$, $A_{\star} = \begin{bmatrix} A_{\star 11} & A_{\star 12} \\ A_{\star 21} & A_{\star 22} \end{bmatrix}$, A_{11} is $((n(p-1)+1)\times(n(p-1)+1))$, $A_{\star 12}$ $((n(p-1)+1)\times r)A_{\star 21}$ $(r\times(n(p-1)+1))$, $A_{\star 22}(r\times r)$, $B_{\star 2}$ is obtained by partition of $B_{\star}$ as $B_{\star} = \begin{bmatrix} B_{\star 1} \\ B_{\star 2} \end{bmatrix}$, where $B_{\star 1}$ is $((n(p-1)+1)\times n)$, $B_{\star 2}(r\times n)$.

To compute the value of C_r in (19), Chen's method (1992) can be used as $\int\int\int\int p(\alpha,\beta,\Gamma,\Sigma)\, d\alpha d\beta d\Gamma d\Sigma = 1$,

$$
\begin{aligned}
C_r &= \iiint p(\alpha,\beta,\Gamma,\Sigma)|_{\alpha=0}\, d\beta d\Gamma d\Sigma \\
&= \frac{\iiint p(\alpha,\beta,\Gamma,\Sigma)|_{\alpha=0}\, d\beta d\Gamma d\Sigma}{\iiint p(\alpha,\beta,\Gamma,\Sigma) d\alpha d\beta d\Gamma d\Sigma} \\
&= \frac{\iiint p(\alpha,\beta,\Gamma,\Sigma)|_{\alpha=0}(\omega(\alpha|\beta,\Gamma,\Sigma)d\alpha) d\beta d\Gamma d\Sigma}{\iiint p(\alpha,\beta,\Gamma,\Sigma) d\alpha d\beta d\Gamma d\Sigma} \\
&= \frac{\iiint p(\alpha,\beta,\Gamma,\Sigma)|_{\alpha=0}\omega(\alpha|\beta,\Gamma,\Sigma) d\alpha d\beta d\Gamma d\Sigma}{\iiint p(\alpha,\beta,\Gamma,\Sigma) d\alpha d\beta d\Gamma d\Sigma} \\
&\simeq \frac{1}{N}\sum_{i=1}^{N} p(\alpha^{(i)}|\beta^{(i)},\Gamma^{(i)},\Sigma^{(i)}) \cdot \frac{p(\alpha=0,\beta^{(i)},\Gamma^{(i)},\Sigma^{(i)})}{p(\alpha=0,\beta^{(i)},\Gamma^{(i)},\Sigma^{(i)})} \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{p(\alpha=0,\beta^{(i)},\Gamma^{(i)},\Sigma^{(i)})}{p(\beta^{(i)},\Gamma^{(i)},\Sigma^{(i)})} \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{p(\alpha=0,\Gamma^{(i)}|\Sigma^{(i)} p\beta^{(i)} p\Sigma^{(i)})}{p(\Gamma^{(i)}|\Sigma^{(i)}) p(\beta^{(i)}) p(\Sigma^{(i)})} \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{p(\alpha=0,\Gamma^{(i)}|\Sigma^{(i)}}{p(\Gamma^{(i)}|\Sigma^{(i)})}
\end{aligned}
\tag{23}
$$

where $p(\alpha=0,\Gamma|\Sigma)$ and $p(\Gamma|\Sigma)$ in the last line of (23) are derived from (4).

The Bayes factor for alpha can be obtained by dividing (21) by (22) When the posterior probabilities are considered, we assign equal prior probabilities to the possible cointe-gration ranks such that $\Pr(\alpha_{rank=r}) = 1/(n+1)$ for $r = 0, 1, ..., n$. With these $n+1$ Bayes factors, we can compute the posterior probabilities for each rank by using (16).

Note that the method described above has the disadvantage over other methods such as Johansen's and Kelibergen and Paap's methods. Their methods are based on eigenvalues or singular values of the long-run matrix Π, while our method is based on testing for zero restrictions on adjustment parameters. Thus, different choice of the ordering of the variables generates different value of Bayes factors.

4 Monte Carlo Simulation

To illustrate the performance of Bayesian tests for the rank of cointegration described in the previous section, we perform some Monte Carlo simulations. The data generating processes (DGPs) consist of a four-variable VAR with an intercept term having various number of cointegrating vectors (0, 1, 2, 3 and 4) as following:

1. $\Delta y_t = \mu + e_t$

2. $\Delta y_t = \mu + \begin{bmatrix} -0.2 \\ -0.2 \\ -0.2 \\ 0.2 \end{bmatrix} [1 \quad 0 \quad 0 \quad -1] y_{t-1} + e_t$

3. $\Delta y_t = \mu + \begin{bmatrix} -0.2 & -0.2 \\ 0.2 & -0.2 \\ 0.2 & 0.2 \\ -0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix} y_{t-1} + e_t$

4. $\Delta y_t = \mu + \begin{bmatrix} -0.2 & -0.2 & -0.2 \\ 0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & -0.2 \\ 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} y_{t-1} + e_t$

5. $\Delta y_t = \mu + \begin{bmatrix} -0.2 & -0.2 & -0.2 & -0.2 \\ 0.2 & -0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & 0.2 & -0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} y_{t-1} + e_t$

where $\mu = [0.1 \ 0.1 \ 0.1 \ 0.1]'$, and $e_t \sim NID\ (0, I_4)$.
We demonstrate the performance of the test by varying the true rank. Each simulation of DGPs for various ranks is repeated 1000 times. The sample size t is 150, of which the first 50 are used for the first experiment and the rest of 100 are used for the second experiment. We consider a VAR(1) model with a constant term throughout the experiments. As noted in the previous section, since the testing procedure depends upon the chosen ordering of the variables in the VAR, the ordering of the individual series in y is changed randomly during the simulation experiment.

The prior parameter specifications for the natural conjugate prior in (3) are given with $P = 0$ and $A = \lambda (\hat{W}' \hat{W}) / T$, where $\hat{W} = (X \ Z\hat{\beta})$ and $\hat{\beta} = 0_{(n \times r)}$. The specification for A is a g-prior proposed by Zellner (1986) and used by Kleibergen and Paap. In this experiments, we assign $\lambda = 1$ and 0.01 to see how this prior specification affects the results. We also specify the prior parameters for β in (5) with $\hat{\beta}_\star = 0$, $\Omega = I_n$, $H = \tau Z'Z$, where $\tau = 1/T$. As for the prior specification for Σ in (4), we assign $S = \tau X'X$ and $h = n + 1$. For Johansen's LR trace test, the cointegrating rank for each iteration is chosen by p-value with the 5 per cent significance level and then the number of each selected rank is counted over iterations to obtain the rates of

4 The results of the Johansen tests in this paper are obtained by using Pcfiml class of Ox v2.20. The source code of the class was modified and recompiled for the simulations.

selection for each rank[4].

Table 1 summarizes the results of Monte Carlo simulation with the sample size t is 50. The values in the columns are the average posterior probabilities of 1,000 iterations for each true rank. For each iteration, the Griddy-Gibbs sampling is performed with 5,000 draws and the first 1,000 discarded. The column labelled as $\Pr(r|Y)$ with $\lambda = 1$ shows the average posterior probabilities when $\lambda = 1$. As shown in the table, the highest average posterior probability for each rank indicates the correct rank. With the sample size is 50, the overall performances are slightly improved when $\lambda = 0.01$, larger prior variance.

Table 1: Monte Carlo Results: The Average Posterior Probabilities with $T = 50$

DGP	rank r	$\Pr(r\|Y)$ with $\lambda = 1$	$\Pr(r\|Y)$ with $\lambda = 0.01$	Johansen's trace test
True rank $r=0$	0	**0.755**	**0.814**	**0.843**
	1	0.093	0.113	0.128
	2	0.068	0.050	0.029
	3	0.051	0.017	0.000
	4	0.033	0.001	0.000
True rank $r=1$	0	0.034	0.050	0.391
	1	**0.669**	**0.715**	**0.493**
	2	0.160	0.131	0.094
	3	0.117	0.065	0.012
	4	0.020	0.039	0.011
True rank $r=2$	0	0.000	0.000	0.000
	1	0.000	0.000	0.235
	2	**0.818**	**0.852**	**0.655**
	3	0.120	0.092	0.049
	4	0.061	0.057	0.061
True rank $r=3$	0	0.000	0.000	0.000
	1	0.000	0.000	0.123
	2	0.059	0.067	0.558
	3	**0.661**	**0.698**	**0.170**
	4	0.280	0.236	0.149
True rank $r=4$	0	0.000	0.000	0.001
	1	0.000	0.000	0.039
	2	0.003	0.001	0.271
	3	0.037	0.060	0.123
	4	**0.959**	**0.939**	**0.566**

The last column shows the results by Johansen's trace test. The results show that the test apparently suffers from the shortage of samples especially for higher ranks. To improve the finite sample properties for the likelihood ratio test, Johansen (2000) proposed using Bartlett correction for a VAR with small sample.

Increasing the sample size to 100 improves the performances of all tests as shown in Table 2. All highest average posterior probabilities indicate the true rank. Larger the prior variance on α (less value in λ), it tends to choose a lower rank.

Table 2: Monte Carlo Results: The Average Posterior Probabilities with $T=100$

DGP	rank r	Pr(r|Y) with $\lambda=1$	Pr(r|Y) with $\lambda=0.01$	Johansen's trace test
True rank $r=0$	0	**0.922**	**0.931**	**0.891**
	1	0.060	0.061	0.089
	2	0.014	0.007	0.018
	3	0.003	0.00	0.002
	4	0.000	0.000	0.000
True rank $r=1$	0	0.000	0.000	0.110
	1	**0.886**	**0.914**	**0.803**
	2	0.106	0.080	0.087
	3	0.005	0.005	0.000
	4	0.002	0.000	0.000
True rank $r=2$	0	0.000	0.000	0.000
	1	0.000	0.000	0.041
	2	**0.961**	**0.927**	**0.949**
	3	0.038	0.072	0.033
	4	0.001	0.002	0.017
True rank $r=3$	0	0.000	0.000	0.000
	1	0.000	0.000	0.000
	2	0.019	0.012	0.216
	3	**0.965**	**0.946**	**0.605**
	4	0.016	0.042	0.179
True rank $r=4$	0	0.000	0.000	0.000
	1	0.000	0.000	0.000
	2	0.000	0.000	0.001
	3	0.022	0.034	0.011
	4	**0.978**	**0.966**	**0.988**

5 Illustrative Example: The Demand for Money in the US

In this section, we illustrate an example of cointegration analysis using the method that is presented in previous sections. The focus is to show the usefulness of our method with a relatively small sample size and to compare two methods of computing Bayes factors and Johansen's likelihood ratio test. The example is cointegration test for the demand for money in the United States.

There are many papers on the estimation of the money demand function, see Goldfeld and Sichel (1990). A typical money demand equation is

$$m_t - p_t = \mu + \gamma_y y_t + \gamma_R R_t + e_t \tag{24}$$

where m_t is the log of money, p_t is the log of price level, y_t is log of income, R_t is the nominal interest rate.

The data used for our analysis is based on the annual data provided by Lucas (1988). The money stock (m_t) is measured by M1. The income (y_t) is the net national product. The interest rate (R_t) is the annual rate of the six-month commercial paper. We use a part of the range of the original data-from 1933 to 1989. With 60

observations, we test cointegration among three variables $-(m-p)$, y, and R. Before analyzing the application, we briefiy explain Bayesian hypothesis testing for the number of lags in VAR. Since we do not know the actual lag length for the VAR and choice of the appropriate lag length affects the cointegration analysis, we apply our method that explained in Section 3 to select the lag length. Let's consider a VAR model, $Y=XB+E$, where $B=(\mu' \Phi'_1 \Phi' \cdots \Phi'_p)$, and X consists of vectors of lagged Y and 1s in the first column. With conjugate and/or diffuse priors, we have the posteriors which are similar to our posteriors that are given in Section 3. Then compute the Bayes factor for each $\Phi_i=0$ to select the appropriate lag length. Note that for this test we do not assign the correction factor C in Bayes factor.

The Bayes factor selects the appropriate lag length in the VAR is 1 dominantly, though the AIC indicates 2 lags in the VAR. We model a VAR(1) with a constant term for our cointegration analysis.

The prior specifications are the same as in the previous Monte Carlo experiments. We assign equal prior probability to each rank. Table 3 illustrates the posterior probabilities with $\lambda=1$ and 0.01 and p-values of Johansen's trace test. For $\lambda=1$ we see that the posterior probabilities indicate that there is one cointegration relation with 98.5 per cent. For $\lambda=0.01$ our method supports one cointegration relation more strongly with 99.4 per cent. Thus, both results by different prior specification select one cointegration relation, although Johansen's test results in no cointegration.

Table 3: Selection of the Cointegration Rank

rank	$\Pr(r\|Y)\ w/\lambda=1$	$\Pr(r\|Y)\ w/\lambda=0.01$	Johansen's p-value
$r=0$	0.0000	0.0042	0.1045
$r=1$	0.9850	0.9940	0.2234
$r=2$	0.0150	0.0019	0.3190
$r=3$	0.0000	0.0000	——

The posterior means of the cointegration relation among the variables $Y_t=(m_t-p_t,\ y_t,\ R_t)$ is $\beta=(1\quad -0.876\quad 0.109)$. Note that the first element of β is restricted to be 1 for identification.

6 Conclusion

This paper shows simple methods of Bayesian cointegration analysis. The Bayes factors are used for computing the posterior probabilities for each rank. Monte Carlo experiments show that the methods proposed in this paper provide fairly good results. The Bayes factors are also applied to select the appropriate lag length in a VAR

either independently or jointly with a different number of rank.

One of the disadvantages of the method is that the testing procedure depends upon the chosen ordering of the variables in VAR. Our Monte Carlo simulations which allow the ordering to change randomly in each iterations show that the method can select the correct rank.

Another disadvantage of the method is computing time. Computing time depends upon the algorithm we choose for estimating the cointegrating vectors. In this paper the Griddy-Gibbs sampler, which requires heavy computations, is chosen simply because we do not need to assign an approximation function that is needed in Metropolis-Hastings or importance sampling. However, it will not be a problem in the future with much faster computers.

In this paper, a matrix-variate normal density for the cointegrating vector is chosen as a prior. Instead, Jeffrey's or reference priors are also worth considering as Kleibergen and Paap used for their cointegration analysis.

References

[1] Bauwens, L., and Giot, P.(1998)"A Gibbs Sampling Approach to Cointegration", *Computational Statistics,* 13, 339-368.

[2] Bauwens, L., and Lubrano, M. (1996) "Identification Restrictions and Posterior Densities in Cointegrated Gaussian VAR Systems", in T.B. Fomby (ed.), *Advances in Econometrics,* vol. 11B, JAI press, Greenwich, CT.

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

[4] Chib, S. (1995) "Marginal Likelihood From the Gibbs Output", *Journal of the American Statistical Association,* 90, 1313-1321.

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

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

[7] Chen, M-H., Shao, Q-M., and Ibrahim, J-G. (2000) *Monte Carlo Methods in Bayesian Computation,* Springer Series in Statistics, Springer.

[8] Dickey, J., (1971) "The Weighted Likelihood Ratio, Linear Hypothesis on Normal Location Parameters", *The Annals of Mathematical Statistics,* 42, 204-223.

[9] Doornik, J.A. (1998) *OX An Object-Oriented Matrix programming Language,* Timberlake Consultants Press., London.

[10] Evans, M. and Swartz, T. (2000) *Approximating Integrals via Monte Carlo and Deterministic Methods,* Oxford University Press, Oxford.

[11] Geweke, J. (1989) "Bayesian inference in Econometric Models Using Monte Carlo Integration", *Econometrica,* 57, 1317-1340.

[12] Geweke, J. (1996) "Bayesian Reduced Rank Regression in Econometrics", *Journal*

*of Econometrics,*75, 127-146.

[13] Goldfeld, S., and Sichel, D., (1990) "The Demand for Money", Chapter 8 in B. Friedman and F. Hahn (eds.), *Handbook of Monetary Economics,* Vol I, North-Holland, New York.

[14] Johansen, S. (1991) "The Power Function for the Likelihood Ratio Test for Cointegration", in J. Gruber, ed., *Econometric Decision Models: New Methods of Modeling and Applications,* Springer Verlag, New York, NY, 323-335.

[15] Johansen, S. (2000) "A Small Sample Correction of the Test for Cointegrating Rank in the Vector Autoregressive Model", *EUI Working Paper,* ECO No. 2000/15, European University Institute, Italy.

[16] Kass, R.E., and Raftery, A. E. (1995) "Bayes Factors", *Journal of the American Statistical Association,* 90, 773-795.

[17] Kass, R. E., and Wasserman, L. (1996) "The Selection of Prior Distributions by Formal Rules", *Journal of the American Statistical Association,* 91, 1343-1370..

[18] Kleibergen, F., and van Dijk, H.K. (1994) "On the Shape of the Likelihood/Posterior in Cointe-gration Models", *Econometric Theory,* 10, 514-551.

[19] Kleibergen, F., and Paap, R. (1999) "Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration", rewritten version of *Econometric Institute Report 9821/A,* Erasmus University Rotterdam.

[20] Lucas, R. (1988) "Money Demand in the United States: A Quantitative Review", *Carnegie-Rochester Conference Series on Public Policy.* Vol 29, 137-168.

[21] Newton, M.A., and Raftery, A. E. (1994) "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap", *Journal of the Royal Statistical Society,* Ser. B, 56, 3-48.

[22] Phillips, P. C. B. (1996) "Econometric Model Determination", *Econometrica,* 64, 763-812.

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

[24] Ritter. C., and Tanner, M.A. (1992) "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler", *Journal of the American Statistical Association,* 87, 861-868.

[25] Tierney, L., and Kadane, J.B. (1986)"Accurate Approximations for Posterior Moments and Marginal Densities", *Journal of the American Statistical Association,* 81, 82-86

[26] Tsurumi. H., and Wago, H. (1996) "A Bayesian Analysis of Unit Root and Cointegration with an Application to a Yen-Dollar Exchange Rate Model", in T.B. Fomby (ed.), *Advances in Econo-metrics,* vol. 11B, JAI press, Greenwich, CT.

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

[28] Zellner, A. (1986) "On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions", ed. P.K. Goel and A. Zellner, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti,* North-Holland, Amsterdam, 233-243.