Testing for Cointegration Rank Using Bayes Factors

Katsuhiro Sugita*

October 2001

Department of Economics
University of Warwick, UK
ecrgi@csv.warwick.ac.uk

Abstract

This paper proposes Bayesian methods for estimating the cointegration rank in an automatic way using Bayes factors. First, we consider natural conjugate priors for computing Bayes factors for the adjustment term. Since using conjugate priors requires that we assign the prior parameters of which we often do not have prior information, and testing by Bayes factor is very sensitive to the parameters, we propose in this paper using the maximum likelihood estimators for the prior parameters. Then, we show the case of using non-informative priors. Since normal Bayes factor cannot be computed with non-informative priors, we apply the intrinsic Bayes factor (IBF) proposed by Berger and Pericchi (1996). Monte Carlo simulations show that using Bayes factor with conjugate priors and the IBF with non-informative priors produce fairly good results. The methods proposed here are also applied for selecting the appropriate lags, or other tests for a VAR model.

Key words: Cointegration; MCMC; Bayes factor

JEL classification: C11; C12; C32

1 Introduction

This paper introduces Bayesian analysis of cointegrated VAR systems. For cointegration analysis, Johansen's likelihood ratio test has been the most popular method among applied econometricians. One of the drawbacks is that classical methods such as the Johansen test involves the pre-testing problem: before conducting cointegration tests, one has to check for unit roots for all variables in the systems. The results of the unit roots tests depend upon the significance levels. Also, even though one fails to reject the null of unit root at the 5 per cent significance level, it does not mean the process has a unit root since the process might be stationary as the coefficient is very close to 1. In a Bayesian framework, the posterior probabilities for the unit roots and stationarity are taken account, and thus the pre-testing problem can be avoided.

*The author thanks Luc Bauwens, Mike Clements, Richard Paap and audiences at the workshop "Recent Advances in Bayesian Econometrics (2001)" in Marseilles for their useful comments on an earlier version of this paper.

Several researchers have proposed Bayesian inference with Markov Chain Monte Carlo (MCMC) methods in cointegrated VAR systems. These include Kleibergen and van Dijk (1994), who propose 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) develops 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. Although their method is rather complicated and requires heavy computation, it has proven to be versatile when computing the Bayes factors. 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 simple methods of determining of the cointegration rank by Bayes factors. The methods are very straightforward. First, we consider using the conjugate priors for some 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 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 for both α and β and diffuse priors for others are chosen to be able to compute the Bayes factors. A problem in choosing the conjugate priors is that the Bayes factor is very sensitive in the choice of prior parameters (location, scale, degree of freedom, etc), and also that appropriate subjective choices of prior specifications for those prior parameters for all parameters of all models are often not feasible. Actually, in cointegration rank test, selection of the rank by Bayes factor depends upon what prior parameters we assign. For example, if we choose fairly large prior variance for the speed of convergence (or the cointegrating long-run matrix), it tends to give lower rank by the Bayes factor. Or, if we choose an inverted Wishart distribution as a prior for variance in the system such that $p(\Sigma) \propto |S|^{h/2}|\Sigma|^{-(h+n+1)} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}S\right)\right\}$, the question is choosing the values of S and h (the degrees of freedom) since these values have enormous effects on the Bayes factors. It is not easy to assign the appropriate prior mean and variance and thus one way to assign them is in an ad hoc fashion or manipulate the parameter until the results look nice. To circumvent this problem, we propose that those parameters are derived from the MLE. Although the use of data in priors is not Bayesian in a strict sense, it gives quite satisfying

[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.

results without taking care of choosing the prior parameters.

We then consider using the diffuse priors for the adjustment term α. Since the normal Bayes factor cannot be computed with improper non-informative priors, we apply the intrinsic Bayes factor (IBF) proposed by Berger and Pericchi (1996). The IBF aims to select model in a fully automatic way, and thus can be used as a reference.

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 100 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 'great ratios' 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).

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 the ECM representation is:

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta X_{t-i} + \varepsilon_t \tag{1}$$

where $t = p, p+1, \ldots, T$, p is the number of lags, and the errors, ε_t, are assumed $N(0, \sum)$ and independent over time. The dimensions of the matrices are μ and ε $(n \times 1)$, $\Psi(n \times n)$, Σ $(n \times n$, PDS), $\alpha\,(n \times r)$, $\beta\,(n \times r)$.

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

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

where

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

$$X = \begin{bmatrix} 1 & \Delta X_{p-1}' & \cdots & \Delta X_1' \\ 1 & \Delta X_p' & \cdots & \Delta X_2' \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta X_{T-1}' & \cdots & \Delta X_{T-p+1}' \end{bmatrix}, W = \begin{bmatrix} X & Z\beta \end{bmatrix}, B = \begin{bmatrix} \Gamma \\ \alpha' \end{bmatrix}.$$

Let t be the number of columns of Y, so that $t = T - p + 1$, then X is $t \times (1 + n(p-1))$, Γ $((1 + n(p-1)) \times n)$, W $(t \times k)$, 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

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

where A is $(k \times k)$ PDS and P $(k \times n)$, $k = n(p-1)+r+1$ (the number of columns in W). This conjugate prior is needed for computation of the Bayes factor for α, which is a partitioned matrix in B. 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 prior mean of $\Pi = \alpha\beta'$ to be zero. However, this choice generates the Bayes factors which favors no cointegration, thus the posterior probabilities tend to show less rank in cointegration. One method for choosing these hyperparameters is to use data Y, although using data in priors is not a Bayesian in a strict sense. In this paper we use $P = (\widehat{W}'\widehat{W})^{-1}\widehat{W}'Y$ and $A = (1/t)(\widehat{W}'\widehat{W})$, where $\widehat{W} = \begin{pmatrix} X & Z\widehat{\beta}_{mle} \end{pmatrix}$, and $\widehat{\beta}_{mle}$ is the MLE of β. 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 either an inverted Wishart $p(\Sigma) \propto |S|^{h/2}|\Sigma|^{-(h+n+1)} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}S\right)\right\}$ where h is a degree of freedom, S an $n \times n$ PDS, or a diffuse prior

$$p(\Sigma) \propto \Sigma^{-(n+1)/2} \tag{4}$$

If the inverted Wishart prior is chosen, we have to specify the prior parameters S and h. Those specifications in general do not have an effect on estimations but do have a strong effect on the Bayes factors. To avoid this, we choose a diffuse prior for the covariance Σ unless we have an appropriate subjective prior information.

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

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

where $\overline{\beta} = \begin{pmatrix} I_r & \overline{\beta}_\star \end{pmatrix}$ is a prior mean of β, Q is $r \times r$ PDS, H is $n \times n$ PDS. Note that r^2 restrictions for identification are imposed on β such that $\beta' = \begin{pmatrix} I_r & \beta_\star' \end{pmatrix}$,[2] where $\beta_\star$ is $(n-r)$ $\times$ r unrestricted matrix.

We do not have to worry about the prior mean and variance in (5) since they have little effects on the Bayes factors if we give enough large variance, say $H = 10^{-10}$. However, we do not know whether or not this H is large enough , and it depends on the data. Thus, as in the case of choosing prior parameter for $B|\Sigma$, we use the MLE as $\overline{\beta}_\star = 0$, $Q = \widehat{\alpha}'_{mle}\widehat{\Sigma}^{-1}_{mle}\widehat{\alpha}_{mle}$, and

[2]The restrictions imposed on β need not to be I_r but can be any r^2 restrictions. This flexibility is an advantage over Kleibergen and Paap (1999) method, which restricts the restriction to be I_r. See Bauwens and Lubrano (1996, page 14)

$H = Z'Z/t$. These are derived from the asymptotic distribution of the MLE. Ahn and Reinsel (1990) provide the theorem.

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

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

Next, we consider improper diffuse priors for the speed of adjustment α and covariance Σ. Here, we assume that $p(\beta)$ is independent of $p(\alpha)$, and thus, $p(\beta, B, \Sigma) \propto \pi(\beta) \cdot |\Sigma|^{-(n+1)/2}$, where $\pi(\beta)$ is the prior for β, which can be Jeffreys, normal, or t density prior[3] . Another non-informative prior that should be worth considering is the reference prior proposed originally by Bernardo (1979), and developed by Berger and Bernardo in their series of papers (1989, 1992)[4]. In this paper we do not consider the reference priors since the derivations are not straightforward.

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

$$\begin{aligned}
&L(Y \mid B, \Sigma, \beta) \\
&= (2\pi)^{-\frac{nt}{2}}|\Sigma \otimes I_t|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left\{\mathrm{Vec}(Y-WB)'(\Sigma^{-1}\otimes I_t^{-1})\mathrm{Vec}(Y-WB)\right\}\right] \\
&\propto |\Sigma|^{-t/2}|I_t|^{-n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}(Y-WB)'(Y-WB)\right\}\right] \\
&= |\Sigma|^{-t/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left\{\widehat{S} + (B-\widehat{B})'W'W(B-\widehat{B})\right\}\right]\right\}
\end{aligned} \tag{7}$$

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

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

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

[3]For discussion about using Jeffreys instead of diffuse prior on β in cointegrated VAR systems, see Kleibergen and van Dijk (1994).

[4]For the discussion of the selection of priors, see Kass and Wasserman (1996).

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

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,

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

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

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,

$$\begin{aligned} p(B \mid \beta, Y) &\propto \int p(\Sigma \mid \beta, Y) p(B \mid \Sigma, \beta, Y) d\Sigma \\ &\propto |S_\star|^{t/2} |A_\star|^{n/2} |S_\star + (B - B_\star)' A_\star (B - B_\star)|^{-(t+k)/2} \end{aligned} \tag{11}$$

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

$$\begin{aligned} p(B, \beta \mid Y) &\propto \int p(B, \Sigma, \beta \mid Y) d\Sigma \\ &\propto \int \pi(\beta) |\Sigma|^{-\frac{c}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\left[S_\star + (B - B_\star)' A_\star (B - B_\star)\right]\right\}\right] d\Sigma \\ &\propto \pi(\beta) |S_\star + (B - B_\star)' A_\star (B - B_\star)|^{-(t+k+1)/2} \end{aligned} \tag{12}$$

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

$$\begin{aligned} p(\beta \mid Y) &\propto \int p(B, \beta \mid Y) dB \\ &\propto \int \pi(\beta) \mid S_\star + (B - B_\star)' A_\star (B - B_\star) \mid^{-\frac{t+k+1}{2}} dB \\ &\propto \pi(\beta) \mid S_\star \mid^{-(t+1)/2} \mid A_\star \mid^{-n/2} \end{aligned} \tag{13}$$

For the case of non-informative prior on α, the posteriors are obtained by the same procedure as shown above. The posterior of Σ conditional on B and β is given as an inverted Wishart density, that is,

$$p(\Sigma | B, \beta, Y) \propto \pi(\beta) \cdot |\Sigma|^{-(t+n+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}(Y - WB)'(Y - WB)\right)\right] \tag{14}$$

The posterior of B conditional on β is obtained as a matrix-variate Student-t

$$p(B \mid \beta, Y) \propto |\widehat{S} + (B - \widehat{B})'W'W(B - \widehat{B})|^{-t/2} \tag{15}$$

The posterior of β is given as a poly-t form as Bauwens and Lubrano (1996) derived, that is,

$$p(\beta|Y) \propto \frac{\pi(\beta) \cdot |\beta' Z' M_x Z \beta|^{(t-k-n)/2}}{|\beta' Z' M_x [I_t - Y(Y' M_x Y)^{-1} Y'] M_x Z \beta|^{(t-k)/2}} \tag{16}$$

where $M_x = I_t - X(X'X)^{-1}X'$.

The last line is the posterior kernel of β, which is the product of the prior for β and inverse of the integrating constant of the matrix-variate Student-t density. An interesting point to note is that either posterior for β in (13) or (16) can be generated only from the data and independent of other parameters such as α or Σ.

The properties of (13) and (16) 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[5] 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 algorithm is provided in the Appendix for convenience.

3 Bayes Factors for Cointegration Tests

This section introduces the computation of the Bayes factors for cointegration rank. Subsection 3.1 describes briefly 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 Computation of the Bayes Factors

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:

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

$$
\begin{aligned}
BF_{01} &= \frac{\Pr(Y|H_0)}{\Pr(Y|H_1)} \\
&= \frac{\int p(\Theta_0|H_0)L(Y|\Theta_0,H_0)d\Theta_0}{\int p(\Theta_1|H_1)L(Y|\Theta_1,H_1)d\Theta_1}
\end{aligned} \tag{17}
$$

With the prior odds, defined as $\Pr(H_0)/\Pr(H_1)$, which are often taken to be 1 if we do not know which hypothesis is correct, we can compute the posterior odds, which are

$$
\text{PosteriorOdds}_{01} = \frac{\Pr(H_0|Y)}{\Pr(H_1|Y)} = \frac{\Pr(Y|H_0)}{\Pr(Y|H_1)} \cdot \frac{\Pr(H_0)}{\Pr(H_1)} \tag{18}
$$

In words,

$$
\text{Posterior Odds} = \text{Bayes Factor} \times \text{Prior Odds}.
$$

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 H_i is

$$
\Pr(H_i|Y) = \frac{\text{PosteriorOdds}_{i0}}{\sum_{j=0}^{q} \text{PosteriorOdds}_{j0}} \tag{19}
$$

where $\text{PosteriorOdds}_{00}= 1$. 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 (17). 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 generalised Savage-Dickey density ratio, proposed by Verdinelli and Wasserman (1995). Suppose we wish to test the null $H_0 : \xi_0$ versus $H_1 : \xi \neq \xi_0$, where ξ can be scalar, vector, or matrix. If $p(\Theta|\xi_0) = p_0(\Theta)$, then the Bayes factor can be computed with the Savage-Dickey density ratio

$$
BF_{01} = \frac{p(\xi_0|Y)}{p(\xi_0)} \tag{20}
$$

If $p(\Theta|\xi_0) \neq p_0(\Theta)$, then the Bayes factor is equal to the Savage-Dickey density ratio times a correction factor, that is,

$$
BF_{01} = \frac{p(\xi_0|Y)}{p(\xi_0)} E\left[\frac{p_0(\Theta)}{p(\Theta|\xi_0)}\right] \tag{21}
$$

The equation (21) is the generalized Savage-Dickey density ratio. The denominator, the marginal prior for ξ evaluated at $\xi = \xi_0$, in (20) or (21) 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:

$$\begin{aligned} p(\xi_0|Y) &= \int p(\xi_0|\Theta, Y)p(\Theta|Y)d\Theta \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} p(\xi_0|\Theta_i, Y) \end{aligned} \tag{22}$$

where Θ_i is a sample from the posterior.
If (22) cannot be computed because the conditional posterior density is not available, Verdinelli and Wasserman (1995) suggest that, using Chen's (1992) method with $\omega(\xi|\Theta)$, which is the normal approximation with sample mean and sample covariance of $(\xi_1, \Theta_1), \ldots, (\xi_N, \Theta_N)$, we estimate $p(\xi_0|Y)$ by

$$\widehat{p}(\xi_0|Y) \simeq \frac{1}{N}\sum_{i=1}^{N} \omega(\xi_i|\Theta_i)\frac{\varphi(\xi_0, \Theta_i|Y)}{\varphi(\xi_i, \Theta_i|Y)} \tag{23}$$

where $\varphi(\xi, \Theta|Y)$ is proportional to the joint posterior density $p(\xi, \Theta|Y)$.
Chen (1992) argues that choosing ω as a normal is not always good approximation, although it is a reasonable choice in most cases.

For a rough approximation of Bayes factor, Kass and Raftery (1995) show that the Bayesian information criterion (BIC) can be used as reference since BIC does not require the introduction of the prior densities. The BIC is defined as $BIC(H_i) = -2\log L(Y|\widehat{\Theta}_{iMLE}) + q\log t$, where q denotes the dimension of Θ_i, and then the Bayes factor can be approximated by computing

$$BF_{01} \simeq \exp\left(\frac{1}{2}\left(BIC(H_1) - BIC(H_0)\right)\right) \tag{24}$$

In this paper, we compute the Bayes factor approximated by the BIC evaluated at the posterior mean instead of the maximum likelihood estimator. Phillips (1996) and Phillips and Ploberger (1996) extend this BIC for model selection.

The Bayes factor assumes that the priors are proper. In case of using non-informative improper prior distributions, one cannot compute Bayes factor since Bayes factor depends on an undefined ratio of constant c_0/c_1. Several methods have been proposed to avoid this problem with the idea of using a training sample of data. These include Spiegelhalter and Smith (1982), who considered an imaginary training sample. O'Hagen (1995) proposed fractional Bayes factors that uses a fractional part of the entire likelihood. Berger and Pericchi (1996a) proposed the intrinsic Bayes factor, which uses a minimal training sample to compute the arbitrary ratio c_0/c_1 and remove the ratio by multiplying the Bayes factor by the inverse of ratio. Consider $\varphi_i^N(\Theta_i)$ be any non-informative improper prior under the hypothesis i, then using $\varphi_i^N(\Theta_i)$ in(17) would yield

$$BF_{ji}^N = \frac{m_j^N(Y)}{m_i^N(Y)} = \frac{\int \varphi_j^N(\Theta_j)L_j(Y|\Theta_j)d\Theta_j}{\int \varphi_i^N(\Theta_i)L_i(Y|\Theta_i)d\Theta_i} \tag{25}$$

Next we compute the Bayes factors that uses the minimal training samples, $Y(l)$ for i and j,

by replacing Y in (25). By using the training samples, $Y(l)$, the non-informative prior, $\varphi_i^N(\Theta_i)$, can be converted to proper posterior distributions $\varphi_i^N(\Theta_i|Y(l)) = \varphi_i^N(\Theta_i)L_i(Y(l)|\Theta_i)/m_i^N(Y(l))$, where $m_i^N(Y(l)) = \int \varphi_i^N(\Theta_i)\, L_i(Y(l)|\Theta_i)\, d\Theta_i$. With the remainder of the data, the Bayes factors are computed using the $\varphi_i^N(\Theta_i|Y(l))$ as priors. The minimal training sample size is $\max\{\text{dimension}(\Theta_i)\}$. For multiple hypothesis tests, minimal training samples are defined relative to all the models simultaneously. Then take the average the Bayes factor of the minimal training samples over all possible training samples. Berger and Pericchi (1996) demonstrate two methods of averaging - the arithmetic (AI) and the geometric (GI) intrinsic Bayes factor. Here we use the geometric intrinsic Bayes factor to take an advantage of its symmetric relation, that is, $IBF_{ji} = 1/IBF_{ij}$. The geometric intrinsic Bayes factor is defined by

$$IBF_{ji} = BF_{ji}^N \cdot \left(\prod_{l=1}^{L} BF_{ij}^N(Y(l)) \right)^{1/L} \tag{26}$$

The arithmetic (AI) intrinsic Bayes factor sums the Bayes factor for the minimal training samples over the all ls and then divided by N.

Computing the products over l in (26) can be enormous if the total sample size is large. As Berger and Pericchi (1996) suggest, one method to avoid this problem is to use only over a subset of Y, selecting randomly from Y. In case of small sample sizes, they recommend to use the expected intrinsic Bayes factors.

3.2 Bayes Factor for Cointegration Rank When Prior for α is Conjugate

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 β is restricted to be $\beta' = [\ I_r \quad \beta_\star]$ for identification, we should not compute the Bayes factors for β. Instead, 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 (20) for each possible rank ($r = 0, 1, \ldots, n$).

$$\begin{aligned} BF_{r|0} &= \frac{\int\int\int\int p(\alpha,\beta,\Gamma,\Sigma)\, L(Y \mid \alpha,\beta,\Gamma,\Sigma)\, d\alpha d\beta d\Gamma d\Sigma}{(1/C_r)\cdot \int\int\int\int p(\alpha,\beta,\Gamma,\Sigma)\mid_{\text{rank}(\alpha)=0} L(Y \mid \alpha,\beta,\Gamma,\Sigma)\mid_{\text{rank}(\alpha)=0} d\beta d\Gamma d\Sigma} \\ &= \frac{p(\alpha' = 0_{r\times n})}{(1/C_r)\cdot p(\alpha' = 0_{r\times n}|Y)} \end{aligned} \tag{27}$$

where $C_r = \int\int\int p(\alpha,\beta,\Gamma,\Sigma)\mid_{\text{rank}(\alpha)=0} d\beta d\Gamma d\Sigma$ 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 (27) 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 (19)

$$\Pr(r|Y) = \frac{BF_{r|0}}{\sum_{j=0}^{n} BF_{j|0}} \tag{28}$$

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

The posterior probabilities given by (28) 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 (27), we use (22) 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 normal distribution as shown in (3), so the numerator of (27) is:

$$\begin{aligned} p(\alpha = \alpha_0) &= \int p(\alpha = \alpha_0|\Sigma)p(\Sigma)d\Sigma \simeq \frac{1}{N}\sum_{i=1}^{N} p(\alpha = \alpha_0|\Sigma^i) \\ &= \frac{1}{N}\sum_{i=1}^{N}(2\pi)^{-nr/2}|\Sigma^i|^{-r/2}|A_{22}|^{n/2}\exp\left[-\frac{1}{2}\mathrm{tr}\Sigma^{i-1}(\alpha'_0 - P_2)'A_{22}(\alpha'_0 - P_2)\right] \end{aligned} \tag{29}$$

where $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, A_{11} is $((n(p-1)+1)\times(n(p-1)+1))$, A_{12} $((n(p-1)+1)\times r)$, A_{21} $(r\times(n(p-1)+1))$, A_{22} $(r\times r)$, P_2 is obtained by partition of P as $P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$, where P_1 is $((n(p-1)+1)\times n)$, P_2 $(r\times n)$.

The posterior for alpha, which is a matrix-variate Student-t from (11), can be estimated by (22) as follows:

$$\begin{aligned} p(\alpha' = \alpha'_{0(n\times r)}|Y) &= \int p(\alpha' = \alpha'_0|\beta, Y)p(\beta|Y)d\beta \simeq \frac{1}{N}\sum_{i=1}^{N} p(\alpha' = \alpha'_0|\beta^i, Y) \\ &= \frac{1}{N}\sum_{i=1}^{N}\pi^{-\frac{nr}{2}}|S^i_\star|^{\frac{t}{2}}|A^i_{\star 22,1}|^{\frac{n}{2}}\left\{\prod_{i=1}^{n}\frac{\Gamma\left(\frac{t+r+1-i}{2}\right)}{\Gamma\left(\frac{t+1-i}{2}\right)}\right\} \\ &\quad \times|S^i_\star + \left(\alpha'_0 - B^i_{\star 2}\right)' A^i_{\star 22.1}\left(\alpha'_0 - B^i_{\star 2}\right)|^{-\frac{t+r}{2}} \end{aligned} \tag{30}$$

where β^i, $S^i_\star$, and $A^i_\star$ are obtained from the i th iteration of the Gibbs sampler, $A_{\star 22.1} = A_{\star 22} - A_{\star 21}A^{-1}_{\star 11}A_{\star 12}$, $A_\star = \begin{bmatrix} A_{\star 11} & A_{\star 12} \\ A_{\star 21} & A_{\star 22} \end{bmatrix}$, $A_{\star 11}$ 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 (27), Chen's method in (23) can be used as

$$\begin{aligned} C_r &= \int\int\int p(\alpha, \beta, \Gamma, \Sigma) \mid_{\alpha=0} d\beta d\Gamma d\Sigma \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} p\left(\alpha^i \mid \beta^i, \Gamma^i, \Sigma^i\right) \cdot \frac{p(\alpha = 0, \beta^i, \Gamma^i, \Sigma^i)}{p(\alpha^i, \beta^i, \Gamma^i, \Sigma^i)} \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{p(\alpha = 0, \beta^i, \Gamma^i, \Sigma^i)}{p(\beta^i, \Gamma^i, \Sigma^i)} \end{aligned} \tag{31}$$

where $p(\alpha, \beta, \Gamma, \Sigma)$ and $p(\beta, B, \Sigma)$ are derived from (6).

The Bayes factor for alpha can be obtained by dividing (29) by (30) and (31). When the posterior probabilities are considered, we assign equal prior probabilities to the possible cointegration ranks such that $\Pr(\alpha_{rank=r}) = 1/(n+1)$ for $r = 0, 1, \ldots, n$. With these $n+1$ Bayes factors, we can compute the posterior probabilities for each rank by using (19).

3.3 Bayes Factor for Cointegration Rank When Priors are Improper

As showed in subsection 3.1, the intrinsic Bayes factor can be used for approximation when priors of interest are non-informative. Computation of (25) and (26) with a model (2) can be done by applying the derivation by Denham (1995), which leads to, with default priors for (B_r, Σ_r) of the form $\pi_r^N(B_r, \Sigma_r) = |\Sigma_r|^{-(2+q_r)/2}$, where r denote model with rank rank $= r$,

$$\begin{aligned} BF_{ji}^N &= \frac{\pi^{(k_i-k_j)n/2}}{2^{(q_i-q_j)n/2}} \cdot \left[\prod_{m=1}^{n} \frac{\Gamma\{(t-k_j+q_j+2-n-m)/2\}}{\Gamma\{(t-k_i+q_i+2-n-m)/2\}}\right] \cdot \frac{|\widehat{W}_i'\widehat{W}_i|^{n/2}}{|\widehat{W}_j'\widehat{W}_j|^{n/2}} \\ &\quad \times \frac{\left|(Y-\widehat{W}_i\widehat{B}_i)'(Y-\widehat{W}_i\widehat{B}_i)\right|^{n/2}}{\left|(Y-\widehat{W}_j\widehat{B}_j)'(Y-\widehat{W}_j\widehat{B}_j)\right|^{n/2}} \end{aligned} \tag{32}$$

where $\widehat{W}_i = \begin{pmatrix} X & Z\widehat{\beta} \end{pmatrix}$, $\widehat{\beta}$ is the MLE of β, and $\widehat{B}_i = (\widehat{W}_i'\widehat{W}_i)^{-1}\widehat{W}_i'Y$.

With diffuse priors, the intrinsic Bayes factor is obtained by

$$\begin{aligned} IBF_{ji} &= \left\{\prod_{s=1}^{n} \frac{\Gamma\{(t-k_j+2-s)/2\}}{\Gamma\{(t-k_i+2-s)/2\}}\right\} \frac{\left|\widehat{W}_i'\widehat{W}_i\right|^{n/2}}{\left|\widehat{W}_j'\widehat{W}_j\right|^{n/2}} \cdot \left(\frac{\left|(Y-\widehat{W}_i\widehat{B}_i)'(Y-\widehat{W}_i\widehat{B}_i)\right|}{\left|(Y-\widehat{W}_j\widehat{B}_j)'(Y-\widehat{W}_j\widehat{B}_j)\right|}\right)^{(t-n+1)/2} \\ &\quad \times \left\{\prod_{p=1}^{n} \frac{\Gamma\{(m-k_i+2-p)/2\}}{\Gamma\{(m-k_j+2-p)/2\}}\right\} \\ &\quad \times \left\{\prod_{l=1}^{L} \frac{\left|\widehat{W}_j(l)'\widehat{W}_j(l)\right|^{n/2}}{\left|\widehat{W}_i(l)'\widehat{W}_i(l)\right|^{n/2}} \cdot \left(\frac{\left|(Y-\widehat{W}_j(l)\widehat{B}_j)'(Y-\widehat{W}_j(l)\widehat{B}_j)\right|}{\left|(Y-\widehat{W}_i(l)\widehat{B}_i)'(Y-\widehat{W}_i(l)\widehat{B}_i)\right|}\right)^{(m-n+1)/2}\right\}^{1/L} \end{aligned} \tag{33}$$

where k_j is the number of columns in $\widehat{W}$, and $m = \max\{k_j\} + 1$.
With equal prior probability to each model, posterior probability for i is computed as

$$\Pr(r|Y) = \left(\sum_{j=1}^{q} BF_{jr} \right)^{-1} \tag{34}$$

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 and two lags having various number of cointegrating vectors (0, 1, 2, 3 and 4). We demonstrate the performance of the test by varying the true rank. Each simulation of DGPs for various ranks is repeated 100 times. All coefficients except for the cointegrating vectors are generated by uniform distributions with a range between -0.4 to 0.4. Disturbance terms are generated by a standard normal distribution. The cointegrating vectors are fixed as follows:

$$0, \begin{pmatrix} 1 & 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

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. The number of lags is assumed to be known throughout the experiments.

We consider three methods of Bayesian cointegration test and Johnasen's LR trace test for reference. The first test is computed using the Bayes factor (BF) for α and the natural conjugate prior for B. Then we demonstrate the simulation using the intrinsic Bayes factor with the diffuse priors. The third simulation is computed using the Bayes factor approximated by the BIC. The prior parameter specifications for the natural conjugate prior in (3) are given with $P = (\widehat{W}'\widehat{W})^{-1}\widehat{W}'Y$ and $A = (1/t)(\widehat{W}'\widehat{W})$, where $\widehat{W} = \begin{pmatrix} X & Z\widehat{\beta} \end{pmatrix}$, where $\widehat{\beta}$ is the MLE of β. We also specify the prior parameters for β in (5) with $\overline{\beta}_\star = 0$, $Q = \widehat{\alpha}'_{mle}\widehat{\Sigma}^{-1}_{mle}\widehat{\alpha}_{mle}$, $H = Z'Z/t$. These specifications for β do not have direct effect on the Bayes factor. The intrinsic Bayes factors (IBF) are computed using (33) with conjugate prior for β with the same specification in BF. The column labelled by BIC is the average posterior probabilities using (24) for reference. 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 selection for each rank[6].

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 100 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 table illustrates that the BF provides correct rank with sample size

[6]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.

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

DGP	rank r	Pr(r|Y) by BF	Pr(r|Y) by IBF	Pr(r|Y) by BIC	Johansen's trace test
True rank	0	**0.891**	**0.808**	**0.986**	**0.558**
$r = 0$	1	0.072	0.189	0.013	0.412
	2	0.019	0.003	0.000	0.030
	3	0.018	0.000	0.000	0.000
	4	0.000	0.000	0.000	0.000
True rank	0	0.157	0.093	**0.610**	0.101
$r = 1$	1	**0.836**	**0.854**	0.250	**0.627**
	2	0.005	0.049	0.089	0.221
	3	0.002	0.003	0.003	0.033
	4	0.000	0.000	0.047	0.018
True rank	0	0.000	0.000	0.008	0.003
$r = 2$	1	0.000	0.052	0.081	0.213
	2	**0.869**	**0.931**	**0.563**	**0.617**
	3	0.130	0.017	0.084	0.116
	4	0.001	0.000	0.264	0.051
True rank	0	0.000	0.000	0.098	0.001
$r = 3$	1	0.000	0.053	0.192	0.038
	2	0.052	0.207	0.339	**0.427**
	3	**0.939**	**0.731**	0.013	0.294
	4	0.009	0.009	**0.357**	0.240
True rank	0	0.000	0.000	0.044	0.000
$r = 4$	1	0.000	0.003	0.077	0.004
	2	0.005	0.070	0.148	0.221
	3	0.043	0.289	0.015	0.067
	4	**0.952**	**0.638**	**0.717**	**0.708**

50. The IBF also produces fairly good indication of the evidence except when the true rank is full. Using the BIC to compute the Bayes factor needs more samples for better approximation, however, as Kass and Raftery (1995) suggest, even with very large samples, the BIC does not produce the correct value. The last column shows the results by Johansen's trace test. The results show that the test apparently suffers from the shortage of samples. 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. Even with a sample size of 100, the Bayes factor approximated by BIC often shows the incorrect rank with a higher rank. This is due to the fact that the correction factor for different dimension is not taken account in the BIC.

In sum, the Bayes factor with prior parameters by the MLE is the most robust indicator

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

DGP	rank r	Pr(r|Y) by BF	Pr(r|Y) by IBF	Pr(r|Y) by BIC	Johansen's trace test
True rank	0	**0.993**	**0.856**	**0.998**	**0.824**
$r = 0$	1	0.007	0.141	0.002	0.166
	2	0.000	0.003	0.000	0.010
	3	0.000	0.000	0.000	0.000
	4	0.000	0.000	0.000	0.000
True rank	0	0.000	0.000	0.188	0.007
$r = 1$	1	**0.933**	**0.947**	**0.728**	**0.813**
	2	0.065	0.043	0.082	0.155
	3	0.002	0.009	0.002	0.012
	4	0.000	0.000	0.045	0.013
True rank	0	0.000	0.000	0.000	0.001
$r = 2$	1	0.000	0.000	0.000	0.004
	2	**0.930**	**0.991**	**0.878**	**0.856**
	3	0.069	0.009	0.035	0.101
	4	0.001	0.000	0.087	0.038
True rank	0	0.000	0.000	0.000	0.000
$r = 3$	1	0.000	0.005	0.000	0.007
	2	0.006	0.133	0.444	0.063
	3	**0.988**	**0.852**	**0.523**	**0.704**
	4	0.006	0.000	0.033	0.226
True rank	0	0.000	0.000	0.000	0.000
$r = 4$	1	0.000	0.000	0.000	0.000
	2	0.002	0.066	0.171	0.000
	3	0.006	0.229	0.188	0.012
	4	**0.992**	**0.705**	**0.642**	**0.984**

for the correct rank in cointegration models with small or medium sample size. For the IBFand the BIC, it seems to have less power especially when the true rank is higher than the middle of the numbers of full rank. Since our Bayes factor uses the MLE for automatic method as the IBF, it can be used as a reference.

5 Illustrative Example - Cointegration Test for 'Great Ratios'

In this section, we illustrate an example of cointegration analysis using the methods that are presented in previous sections. The focus is to show the usefulness of our methods 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 'great ratios'.

King, *et al* (1991) (KPSW) examine cointegrating relationships between US output (Y),

consumption (C), investment (I), and three other variables. In this section we investigate a three-variable model containing the real variables, C, I, and Y. The data are quarterly and taken from the KPSW data sets, which are: C (real per capita consumption, in logs), I (investment per capita, in logs), and Y (real private output per capita, in logs). We choose the shorter estimation period of 1968(1)-1988(4), with a sample size of 83. From economic theory, two cointegrating relations are expected to be found among these variables, given by C − Y and I − Y, which are known as the 'great ratios'.

Before analyzing the application, we briefly 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 methods that explained in Section 3 to select the lag length. Let's consider a VAR model, $Y = XB + E$, where $B = \begin{pmatrix} \mu' & \Phi_1' & \cdots & \Phi_p' \end{pmatrix}$, 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. This procedure is similar to classical methods which test lag length first then cointegration rank test as the selected lag is true. It is, however, possible to perform joint tests for lag length and cointegration rank using the Bayes factors with $\Phi_i = 0$ as a null. We choose the joint test for selection of lags and rank in this example.

Table 3 presents the results of cointegration tests by both Bayesian with various lag length and Johansen tests with 2 lags[7] in VAR for the three-dimensional vector of time series $Y_t' = \begin{bmatrix} \mathrm{C}_t & \mathrm{I}_t & \mathrm{Y}_t \end{bmatrix}$ with an intercept term. The prior specifications are given by the MLE as explained in previous sections. We assign equal prior probability to each rank. The Bayesian test shows that the posterior probability when the rank is 1 with lag length $p = 2$ is the highest with 60.2 per cent using the conjugate prior and 50.8 per cent using diffuse prior for α, although there is some evidence to support of no cointegration. The posterior probability for no cointegration is $\Pr(r = 0|Y) = \sum_{i=1}^{4} \Pr(r = 0, p = i|Y) = 0.273$ for the BF and 0.324 for the IBF. From the table we find that the Bayes factor for cointegration rank is very sensitive to the choice of the lags in VAR. There is almost no evidence of $r = 2$ and full rank. Thus, there is a relatively strong evidence to support the cointegration relationship between consumption and income, however, no cointegration between investment and income. On the classical side, Johansen's LR trace statistics cannot reject $r = 0$ at either 5 or 10 per cent significance level. Therefore, Johansen test chooses $r = 0$, while the Bayesian test selects dominantly $r = 1$, that is, consumption-income relation. Both tests do not support $r = 2$ which economic theory suggests.

Table 4 shows the posterior results of $\beta^\star$ and α with the conjugate prior for α. If the rank is 1, we expect one cointegrating vector would be the first 'great ratios', which is the consumption-income relation, that is,

$$\beta' = \begin{bmatrix} 1 & \beta_1^\star & \beta_2^\star \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \tag{35}$$

[7]Both AIC and BIC for the appropriate lag selects the order 2 with the subset of their data.

Table 3: Joint Posterior Probabilities for the cointegration rank with various lag length p using the Bayes Factors (BF), the intrinsic Bayes Factors (IBF), and the Johansen's trace test.

$\Pr(r,p\|Y)$ by BF	$r=0$	$r=1$	$r=2$	$r=3$
$p=1$	0.002	0.108	0.000	0.000
$p=2$	0.050	**0.602**	0.001	0.000
$p=3$	0.085	0.006	0.000	0.000
$p=4$	0.136	0.010	0.000	0.000
$\Pr(r,p\|Y)$ by IBF	$r=0$	$r=1$	$r=2$	$r=3$
$p=1$	0.014	0.156	0.002	0.001
$p=2$	0.012	**0.508**	0.000	0.000
$p=3$	0.195	0.000	0.000	0.000
$p=4$	0.103	0.009	0.000	0.000
Johansen, p-val,w/$p=2$	0.122	0.651	0.804	—

Table 4: Posterior results for $\beta^\star$ and α

	$\beta_1^\star$	$\beta_2^\star$	α_1	α_2	α_3
Mean	-0.0802	-0.9467	0.1350	0.3271	0.2134
S.D.	0.1442	0.1904	0.0310	0.0567	0.0400

Figure 1: Posterior Densities of $\beta^\star$ (the left column) and α (the right column)

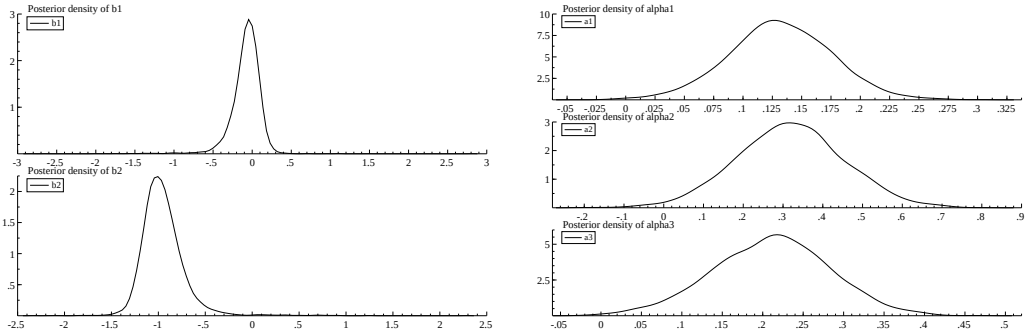
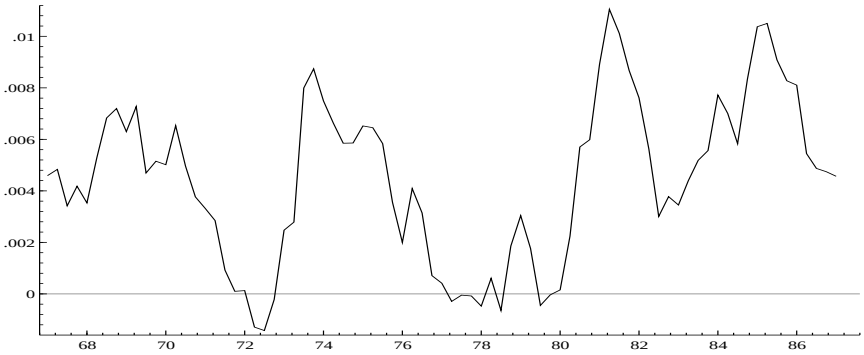


Figure 2: Cointegration Relationship



The posterior means and medians are close to this economic relation. The left column of Figure 1 is the posterior densities of $\beta^\star$, which shows that the expected cointegrating vector, $\beta_1^\star = 0$ and $\beta_2^\star = -1$, lies within the 95 per cent highest posterior density regions. The right column of Figure 1 shows the posterior densities of each element of α, which are skewed and lie far from zero. Figure 2 plots the cointegration relationship, which is slightly upward trending.

6 Conclusion

This paper shows simple methods of Bayesian cointegration analysis, which is robust with a small sample size. 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 Bayes factor is that it is very sensitive to the prior specifications. This issue, however, can be overcome by specifying appropriate priors that can be obtained by, for example, using the MLE if one wishes to use the conjugate priors. Or, the intrinsic Bayes factor can be applied to compute the Bayes factor if one wished to use non-informative priors.

Another disadvantage of Bayesian 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. The expected intrinsic Bayes factor, proposed by Berger and Pericchi (1996a), is also worth considering to improve a performance when the sample size is small.

Appendix : Griddy-Gibbs Sampler

The Griddy-Gibbs sampler is proposed by Ritter and Tanner (1992). This sampler can be implemented when the conditional posterior density is unknown to researcher. The advantage of using this sampler over the importance sampler or the Metropolis-Hastings algorithm is that researcher does not have to provide an approximation of function. The disadvantage is that this sampler demands more computing time. The procedure for implementing the Griddy-Gibbs sampler is as following:

1. Before we begin the chain, we must choose the range of the grid and the number of the grid. The range should be chosen so that the generated numbers are not truncated.

2. Let $\text{vec}(\beta)' = (\beta_1, \beta_2, \ldots, \beta_m)$. With an arbitrary starting value (within the upper and the lower bound of the grid), compute $f(\beta_1|\beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y)$, where i denotes the i-th

loop, over the grid $(\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,U})$, where $\beta_{1,1}$ is the lower bound of the grid of β_1, and $\beta_{1,U}$ is the upper bound of the grid of β_1.

3. Compute the values $G = (0, \Phi_2, \Phi_3, \ldots, \Phi_U)$ where

$$\begin{aligned} \Phi_j &= \int_{\beta_{1,1}}^{\beta_{1,j}} f(\beta_1|\beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y) d\beta_1 \\ & j = 2, \ldots, U \end{aligned}$$

4. Compute the normalized pdf values $G_\zeta = G_j/\Phi_U$ of $\zeta(\beta_1|\beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y)$.

5. Draw the random numbers from the uniform density with the lower bound as zeros and the upper bound as Φ_U and invert cdf G by numerical interpolation to obtain a draw β_1^i from $\zeta(\beta_1|\beta_2^i, \beta_3^i, \ldots, \beta_m^i, Y)$.

6. Repeat steps 2-5 for $\beta_2, \ldots, \beta_m$.

7. Set $i = i + 1$ (increment i by 1) and go to step 2.

Note that integration at step 3 can be done by the deterministic approximation such as the Simpson's rule or the Trapezoidal rule.

References

[1] Ahn, S.K., and Reinsel, G. C. 1990, Estimation for Partially Nonstationary Multivariate Autoregressive Models, *Journal of the American Statistical Association,* 85, 813-856.

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

[3] 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.

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

[5] Berger, J. O., and Bernardo, J. 1989, Estimating a Product of Means: Bayesian Analysis with Reference Priors, *Journal of the American Statistical Association,* 84, 200-207.

[6] Berger, J. O., and Bernardo, J. 1992, On the Development of the Reference Prior Method, *Bayesian Statistics 4,* ed. J.M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Oxford University Press, oxford.

[7] Berger, J. O., and Pericchi, L. R. 1996a, The Intrinsic Bayes Factor for Model Selection and Predicsion, *Journal of the American Statistical Association,* 91, 109-122.

[8] Berger, J. O., and Pericchi, L. R. 1996b, The Intrinsic Bayes Factor for Linear Models, *Bayesian Statistics 5,* ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Oxford University Press, Oxford.

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

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

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

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

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

[14] Dreze, J.H. 1977, Bayesian Regression Analysis Using Poly-t Densities, *Journal of Econometrics,* 6, 329-354.

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

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

[17] Geweke, J. 1996, Bayesian Reduced Rank Regression in Econometrics, *Journal of Econometrics,*75, 127-146.

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

[19] 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.

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

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

[22] King, R.G., Plosser, C.I., Stock, S.H., and Watson, M.W. 1991, Stochastic Trends and Economic Fluctuations, *American Economic Review*, 81, 819-840.

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

[24] 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.

[25] Phillips, P. C. B. 1996, Econometric Model Determination, *Econometrica,* 64, 763-812.

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

[27] 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.

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

[29] 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 Econometrics,* vol. 11B, JAI press, Greenwich, CT.

[30] 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.

[31] Zellner, A. 1986, On Assesing Prior Distrubutions 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.