University of Leicester

EFFICIENT POSTERIOR SIMULATION FOR COINTEGRATED MODELS WITH PRIORS ON THE COINTEGRATION SPACE

Gary Koop, University of Leicester, UK

Roberto León-González, University of Leicester, UK

Rodney W. Strachan, University of Leicester, UK

Efficient Posterior Simulation for Cointegrated Models with Priors On the Cointegration Space

Gary Koop*

Roberto León-González

Rodney W. Strachan

April 2006

ABSTRACT: A message coming out of the recent Bayesian literature on cointegration is that it is important to elicit a prior on the space spanned by the cointegrating vectors (as opposed to a particular identified choice for these vectors). In this note, we discuss a sensible way of eliciting such a prior. Furthermore, we develop a collapsed Gibbs sampling algorithm to carry out efficient posterior simulation in cointegration models. The computational advantages of our algorithm are most pronounced with our model, since the form of our prior precludes simple posterior simulation using conventional methods (e.g. a Gibbs sampler involves non-standard posterior conditionals). However, the theory we draw upon implies our algorithm will be more efficient even than the posterior simulation methods which are used with identified versions of cointegration models.

---

*Gary Koop (Gary.Koop@strath.ac.uk) is at the University of Strathclyde. Roberto León-González (rlg19@le.ac.uk) and Rodney Strachan (rws7@le.ac.uk) are at the University of Leicester. Address for correspondence:
 Roberto León-González,
 Department of Economics
 University of Leicester,
 LE1 7RH, Leicester,
 United Kingdom

1 Introduction

Early Bayesian work on cointegration used Vector Autoregressive (VAR) representations (e.g. DeJong (1992), Dorfman (1994), Koop (1994)), for which simple and standard methods of posterior simulation were available. This early work was criticized by subsequent authors for ignoring the reduced rank structure implied by the cointegrating restrictions. Accordingly, the Vector Error Correction Model (VECM), was increasingly adopted for Bayesian work (see, e.g., Bauwens and Lubrano (1996) and Geweke (1996)). For an $n-$vector of unit root variables, y_t, we write the VECM for $t = 1, .., T$ as:

$$\Delta y_t = \Pi y_{t-1} + \sum_{h=1}^{l} \Psi_h \Delta y_{t-h} + \Phi d_t + \varepsilon_t \tag{1}$$

where the $n \times n$ matrix $\Pi = \alpha\beta'$, α and β are $n \times r$ full rank matrices and d_t denotes deterministic terms (see, e.g., Johansen (1995) pages 81-84 for a commonly-used set of choices). The value of r determines the number of cointegrating relationships. ε_t is i.i.d. $N(0, \Sigma)$. Since the crucial issues of identification, prior elicitation and posterior simulation all relate to Π, we will focus on the restricted version of (1):

$$\Delta y_t = \alpha\beta' x_t + \varepsilon_t, \tag{2}$$

where $x_t = y_{t-1}$. Although this paper discusses a cointegrated model, (2) makes clear that the ideas are of relevance for any model with such a reduced rank structure.

Relative to the VAR, Bayesian inference in the ECM is complicated by the fact that $\Pi = \alpha\beta'$ involves a product of parameters. This precludes direct use of analytical or Monte Carlo integration results for the multivariate linear model. However, once we condition on the cointegrating vectors, β, the otherwise nonlinear ECM becomes a linear one. This means that, under suitable informative priors (e.g. Normal priors of the form used in Geweke (1996)), standard Bayesian analysis of the multivariate linear model applies (conditional on β). This suggests posterior simulation can be done in a straightforward fashion if the posterior distribution of β (either a marginal distribution or a distribution conditional on α or on α, Ψ, Φ and Σ where $\Psi = (\Psi_1, .., \Psi_l)$) can be drawn from.

However, the VECM suffers from both a global and local identification problem. A global identification issue can be seen by noting that $\Pi = \alpha\beta'$ and $\Pi = \alpha CC^{-1}\beta'$ are identical for any nonsingular C. This indeterminacy is commonly surmounted by imposing the so-called *linear normalization* where $\beta = \begin{pmatrix} I_r \\ \beta_0 \end{pmatrix}$. Even if global

identification is imposed, a local identification issue occurs at the point $\alpha = \mathbf{0}$ (i.e. at this point β does not enter the model). A large literature (e.g. Kleibergen and van Dijk (1994, 1998), Kleibergen and Paap (2002)) discusses problems arising from local non-identification (e.g. lack of existence of posterior moments and lack of convergence of Gibbs samplers using common noninformative priors) and develops other approaches to prior elicitation which surmount these problems. However, the posterior simulation methods used in these papers become more complicated. Furthermore, they all impose global identification through restrictions analogous to the linear normalization.

A literature has recently emerged which argues that it is only the cointegrating space which is identified (see Strachan (2003), Strachan and Inder (2004), Strachan and van Dijk (2004) and Villani (2005a,b)) and this should be the focus of interest (rather than a particular identified parameter such as β_0). For instance, Strachan and Inder (2004) show how the use of linear identifying restrictions places a restriction on the estimable region of the cointegrating space. Strachan and van Dijk (2004) show that a flat and apparently "noninformative" prior on β_0 in the linear normalization favors regions of the cointegration space near where the linear normalization is invalid. Hence, the linear normalization is used under the assumption that it is valid while at the same time the prior says that the normalization is likely to be invalid.

This recent literature has begun to develop ways of eliciting priors over spaces (as opposed to parameters) and deriving corresponding posterior simulation methods. However, this literature is in its infancy and a good understanding of prior elicitation and efficient posterior simulation have been elusive. A recent survey paper, Koop, Strachan, van Dijk and Villani (2005) describes the development of the Bayesian cointegration literature in detail. The purpose of our paper is to shed further insight on prior elicitation and new methods for posterior computation. In particular, we develop efficient posterior simulation algorithms using a collapsed Gibbs sampler which adapts the algorithm of Liu (1994) to the present context. An empirical illustration demonstrates the large computational gains achieved by this algorithm.

2   Prior Elicitation

Let $\mathfrak{p} = sp\left(\beta\right)$ denote the cointegration space which is an r-dimensional hyperplane in a n-dimensional space. We wish to carry out Bayesian inference relating to this space without imposing identification in such a way as to restrict this space. Furthermore, we wish to develop sensible informative and noninformative priors on this space. To illustrate a key basic idea in a simple case, suppose $n = 2$ and a single cointegrating vector exists. We can parameterize the latter in polar coordinates $\beta = \left(\cos\theta \ \sin\theta\right)'$, where $\theta \in [-\pi/2, \pi/2]$. It is only θ

which determines the cointegration space and, thus, we can restrict the length of β to be unity for identification without restricting the cointegration space. A candidate for a noninformative distribution on $\mathfrak{p}$ is the Uniform distribution on θ and this indeed has sensible properties. These points (and many more) are made in Strachan and Inder (2004) and extended to the case where n and r are of higher dimensions. In this general case, the cointegrating space is an element of the Grassman manifold and, thus, a Uniform prior for the cointegration space is given by the Uniform distribution on the Grassman manifold. An identification restriction which does not restrict the possible cointegration space is:

$$\beta'\beta = I_r. \tag{3}$$

Formally, this restricts the matrix of cointegrating vectors to the Stiefel manifold. These spaces are compact and, hence, a Uniform distribution over them is proper (the integrating constant which ensures propriety is given in Strachan and Inder (2004)). Thus, a noninformative prior for β which does not restrict the cointegrating space is simply proportional to an integrating constant with (3) imposed.

To develop an informative prior, we introduce the following transformation:

$$\beta\alpha' = (\beta\kappa)\left(\alpha\kappa^{-1}\right)' = \left[\beta\left(\alpha'\alpha\right)^{\frac{1}{2}}\right]\left[\alpha\left(\alpha'\alpha\right)^{-\frac{1}{2}}\right]' \equiv BA', \tag{4}$$

where κ is positive definite matrix and $A = \alpha\kappa^{-1}$ is semi-orthogonal. The matrix square root is defined, e.g., in Abadir and Magnus (2005), pages 220-221. This reference also describes a method for its practical calculation.

For future reference, we write various relations between the parameters in (4):

$$\kappa = (\alpha'\alpha)^{\frac{1}{2}} \tag{5}$$
$$\beta = B\left(B'B\right)^{-\frac{1}{2}}$$
$$B'B = \alpha'\alpha. \tag{1}$$

Crucially, in the first of these parameterizations (i.e. involving α and β), β is semi-orthogonal while α is unrestricted, whereas in the second (i.e. involving A and B) it is B which is unrestricted whereas A is semi-orthogonal. Our collapsed Gibbs sampler proceeds by switching between these two parameterizations and, hence, it proves useful to consider prior elicitation in both of them.

We begin by setting out the complete prior written in terms of B, α and Σ, following which we will motivate why they are attractive choices. We will use the standard noninformative prior for Σ:

$$p\left(\Sigma\right) \propto \left|\Sigma\right|^{-(n+1)/2}, \tag{6}$$

although an inverted-Wishart prior can easily be accommodated. The priors for B and α depend on parameters τ and ν (to be explained shortly) and take the forms:

$$B|\tau, \nu \sim MN\left(0, I_r \otimes \nu P_\tau\right), \tag{7}$$

and

$$\alpha|\beta, \tau, \nu \sim MN\left(0, \nu\left(\beta' P_{1/\tau}\beta\right)^{-1} \otimes I_n\right), \tag{8}$$

where P_τ is a function of τ (to be defined shortly) and MN denotes the matricvariate-Normal distribution (see, e.g., Bauwens, Lubrano and Richard (1999), Appendix A).

To motivate the prior for α note that, given (7) and the fact that $\alpha = A\left(B'B\right)^{\frac{1}{2}}$, it follows that (8) arises if we choose a Uniform prior for A (a sensible noninformative choice for similar reasons to those discussed near equation 3). It is proper and, thus, marginal likelihoods (e.g. for calculating posterior probabilities of various cointegrating ranks) can be calculated. In terms of α, (8) has a shrinkage prior interpretation similar to the prior used in Strachan and Inder (2004). ν is a scalar which controls the degrees of informativeness of the prior. ν can either be selected by the researcher subjectively or (8) can be treated as a hierarchical prior with ν being an unknown parameter.

We now turn to the prior in (7). We introduce a semi-orthogonal matrix matrix H for the purposes of eliciting a prior and adopt the standard notation where $H_\perp$ is the orthogonal complement of H. Prior information about the cointegration space can be expressed by eliciting a value for H which spans the space felt, *a priori*, to be most plausible. To obtain H, the researcher will typically first specify a matrix H^g containing desired coefficient values and then use the transformation $H = H^g\left(H^{g'}H^g\right)^{-1/2}$. The matrix H constructed in this way will span the same space as H^g but is semi-orthogonal.

For instance, if y_t contains three interest rates of different maturities, then theories of the term structure suggest pairs of them should be cointegrated and, thus:

$$H^g = \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

H^g is not semi-orthogonal but $H = H^g \left(H^{g\prime} H^g\right)^{-1/2}$ will be (and will span the same space).

Following Strachan and Inder (2004), we can construct a prior for $\mathfrak{p} = sp\left(\beta\right)$ which is a weighted averaged of $\mathfrak{p}^H = sp\left(H\right)$ and $\mathfrak{p}^{H\perp} = sp\left(H_\perp\right)$ where the weights are dependent on the scalar, τ. Details are provided in Strachan and Inder (2004), here we note that, if we define $P_\tau = HH' + \tau H_\perp H'_\perp$, then the resulting prior for β (conditional upon τ) has a matrix angular central Gaussian distribution of Chikuse (1990). More importantly for our purposes, our MCMC algorithm is greatly simplified by the fact that this prior for $\mathfrak{p}$ implies the Normal prior for B given in (7). Prior elicitation is facilitated by noting that if $\tau = 1$, then $P_\tau = I_n$ and we have a flat non-informative prior on β. Hence, it is usually sensible to have $0 \leq \tau \leq 1$, with $\tau = 0$ implying a dogmatic belief that the cointegrating space is $\mathfrak{p}^H$ and $\tau = 1$ being noninformative (formally, the prior for $\mathfrak{p}$ is Uniform in the Stiefel manifold when $\tau = 1$). Should the researcher not wish to subjectively elicit a value for τ, she could instead treat τ as an unknown parameter and specify a prior density for it. In practice, such a prior density would typically allocate most weight to values of τ near zero and be restricted to $[0, 1]$.

Thus, the prior specified by (6) through (8) has many sensible properties. As we shall see in the next section, it also simplifies computation. Another advantage of this formulation is that the standard noninformative prior is found by simply setting $\frac{1}{\nu} = 0$ and, thus, the posterior simulator described in the next section also works with this prior.

3   Posterior Inference

A big advantage of the prior given in (6) and (7) is that it allows for efficient and simple posterior computation through use of a collapsed Gibbs sampler (see Liu (1994) and Liu, Wong and Kong (1994)). Note that standard results imply that MCMC draws for Σ can always be taken from:

$$\Sigma | \alpha, \beta, Data \sim IW\left(\left[\Delta y - X\beta\alpha'\right]' \left[\Delta y - X\beta\alpha'\right], T\right), \tag{10}$$

where Δy and X are $T \times n$ matrices with t^{th} row given by $\Delta y'_t$ and x'_t, respectively. IW denotes the inverted-Wishart distribution (see, e.g., Bauwens, Lubrano and Richard (1999), Appendix A). Alternatively, Σ can be

integrated out (with minor alterations to the algorithm described below). If τ or ν are treated as unknown parameters (as opposed to having values selected for them), then a prior must be selected and an MCMC step which draws τ and ν has to be added. For simplicity, we omit the conditioning arguments Σ, τ and ν in this section and focus on the key issues relating to drawing from α and β.

The basic idea underlying our MCMC algorithm is computational inefficiencies arise from trying to impose the semi-orthogonality restriction on β. Accordingly, our algorithm alternates between the (one-to-one) transformation between (α, β) and (A, B) which only involves drawing from simple distributions.

After choosing an initial value, $\beta^{(0)}$, our MCMC algorithm repeats the following steps for $s = 1, .., S$:

1. Draw $\alpha^{(*)}$ from $p\left(\alpha | \beta, Data\right)$ and transform this to obtain a draw $A^{(*)} = \alpha^{(*)}\left(\alpha^{(*)\prime}\alpha^{(*)}\right)^{-\frac{1}{2}}$.

2. Draw $B^{(s)}$ from $p\left(B | A^{(*)}, Data\right)$ and then transform this to obtain $\beta^{(s)} = B^{(s)}\left(B^{(s)\prime}B^{(s)}\right)^{-\frac{1}{2}}$ and $\alpha^{(s)} = A^{(*)}\left(B^{(s)\prime}B^{(s)}\right)^{\frac{1}{2}}$.

To see why these conditionals define a collapsed Gibbs sampler, note that (A, k) is the unique polar decomposition of α (e.g. Cadet (1996)), and therefore the draw of $\alpha^{(*)}$ in step (1) is a draw of $(A^{(*)}, \kappa^{(*)})$ from the joint density $p\left(A, \kappa | \beta, Data\right)$. Similarly, the draw of $B^{(s)}$ in step (2) is a draw of $\left(\beta^{(s)}, \kappa^{(s)}\right)$ from $p\left(\beta, \kappa | A^{(*)}, Data\right)$. Therefore, $A^{(*)}$ in step (1) is a draw from $p\left(A | \beta, Data\right)$ (i.e. obtained marginally on κ), and $\beta^{(s)}$ in the second step is a draw from $p\left(\beta | A^{(*)}, Data\right)$ (i.e. obtained again marginally on κ). Therefore, our sampling strategy is equivalent to the collapsed Gibbs sampler proposed by Liu (1994) and Liu, Wong and Kong (1994, scheme 1), who show that this algorithm is more efficient than a standard Gibbs sampling algorithm (i.e. one which simply draws sequentially from the conditional posteriors $p\left(\alpha | \beta, Data\right)$ and $p\left(\beta | \alpha, Data\right)$). We stress that the latter algorithm does not exist, so our algorithm is likely to be much faster than the existing choices (e.g. Strachan and van Dijk (2004)).

An advantage of this algorithm is that it only involves drawing from the Normal distribution. In particular, in Step 1 $vec\left(\alpha^*\right)$ is drawn from the Normal with mean $\overline{\alpha}$ and variance $\overline{\Omega}_\alpha$ where

$$\overline{\Omega}_\alpha = \left(\left[\beta' X' X \beta\right] \otimes \Sigma^{-1} + \frac{1}{\nu}\left[\beta' P_\tau^{-1} \beta\right] \otimes I_n\right)^{-1}$$

and

$$\overline{\alpha} = \overline{\Omega}_\alpha \left(\left[\beta' X'\right] \otimes \Sigma^{-1}\right) vec\left(\left[\Delta y\right]'\right).$$

In Step 2 $vec\left(B^{(s)}\right)$ is drawn from the Normal with mean $\overline{B}$ and variance $\overline{\Omega}_B$ where

$$\overline{\Omega}_B = \left(\left[A'\Sigma^{-1}A \right] \otimes \left[X'X \right] + I_r \otimes \frac{1}{\nu}P_\tau^{-1} \right)^{-1}$$

and

$$\overline{B} = \overline{\Omega}_B vec\left(X'\Delta y \Sigma^{-1}A \right).$$

Should the researcher wish to treat τ and ν as unknown parameters, then sensible priors for them would be inverted Gamma-2 $IG_2(s_\tau, n_\tau)$ (see, e.g., Bauwens, Lubrano and Richard (1999), Appendix A) for τ and $IG_2\left(s_\nu, n_\nu\right)$ for ν. These priors result in simple posterior conditionals: $IG_2(tr[\nu^{-1}B'H_\perp H'_\perp B]+s_\tau, (n-r)r+n_\tau)$ for τ and $IG_2\left(s_\nu + tr\left[B'P_{1/\tau}B\right], n_\nu + nr\right)$ for ν. For τ, it will typically make sense to elicit s_τ and n_τ such that the preponderance of prior probability is allocated between 0 and 1.

4   Empirical Evidence on Efficiency of Our Algorithm

In order to simulate data from model (2) we decompose y_t into two components $y_t = \left(y'_{1,t} \ y'_{2,t} \right)'$ with $y_{1,t} : r \times 1$, $y_{2,t} : (n-r) \times 1$ and use the cointegrating system $y_{1,t} = \beta'_0 y_{2,t} + w_{1,t}$, $\Delta y_{2,t} = w_{2,t}$, with the errors generated according to $w_{j,t} = \rho_j w_{j,t-1} + \varepsilon_{j,t}$, $j = 1,2$ where $\varepsilon_{j,t} \backsim iidN\left(0, (1.5)^2 \right)$, $\rho_1 = \rho I_r$, $\rho = 0.3$, $\rho_2 = 0$ and β_0 is a $(n-r) \times r$ matrix of ones. This specification corresponds to (2) with $\alpha = \left[(\rho - 1)I_r \quad 0_{r \times (n-r)} \right]'$, $\beta = \left[\begin{array}{cc} I_r & -\beta'_0 \end{array} \right]'$, and $\varepsilon_t = \left(\varepsilon'_{1,t} + \varepsilon'_{2,t}\beta_0, \quad \varepsilon'_{2,t} \right)'$. We use the noninformative version of the prior.

We compare our algorithm with the following algorithm:

1. Generate $\beta^{(s)}$ from $p\left(\beta | Data \right)$ using a Metropolis-Hasting (MH) algorithm.

2. Draw $\alpha^{(s)}$ from $p\left(\alpha | \beta, Data \right)$.

3. Draw Σ^s from $p\left(\Sigma | \alpha, \beta, Data \right)$.

Following Strachan and van Dijk (2004), we specify a random walk proposal density to generate candidates for β. In particular, they suggest a $MACG(P_{z^2})$ (see Chikuse (1990)) proposal density with parameter $P_{z^2} = \beta^{(s-1)}\beta^{(s-1)\prime} + z^2 \beta_\perp^{(s-1)}\beta_\perp^{(s-1)\prime}$, where $\beta^{(s-1)}$ is the value of β obtained in the previous iteration. This proposal density is therefore of the same type as the prior discussed above. For values of z smaller than 1 this density gives more weight to the space spanned by $\beta^{(s-1)}$, and this weight is greater the closer z is to zero (Strachan and Inder (2004)). The variable z is specified to follow a $N(0, \sigma^2)$ and we adjust σ^2 to obtain acceptance rates

between 20% and 50% (Chib and Greenberg (1995)). Steps 2 and 3 imply drawing from standard distributions (multivariate Student and inverted Wishart, respectively).

We compare the efficiency of these algorithms using the effective sample size (e.g. Brooks (1999)) and the average update distance between iterations (Holmes and Held (2005)). The effective sample size measures the number of independent draws from the posterior that is equivalent to n^* draws from an MCMC algorithm. For $n^* = 1$ the effective sample size is defined by:

$$ESS = \frac{1}{1 + 2\sum_{t=1}^{\infty} \rho(t)}$$

where $\rho(t)$ is the autocorrelation function at lag t. A value near to one means that the algorithm is as efficient as independent sampling, whereas a value near to zero implies that the algorithm is very inefficient (compared to independent sampling). Since the estimated autocorrelations become imprecise as t gets large, we truncate the sum in the denominator using the initial monotone sequence estimator proposed by Geyer (1992). ESS is calculated for $d(\beta, \beta^*)$, where β^* spans the true cointegrating space (specified above) and $d(\beta, \beta^*)$ is the distance measure between cointegrating spaces proposed by Larsson and Villani (2001).

We also calculate the average distance between subspaces sampled in consecutive iterations, defined as:

$$AD = \frac{1}{N-1} \sum_{s=1}^{N-1} d(\beta^s, \beta^{s-1})$$

where N is the number of iterations after the burn-in. Our choice for the burn-in is 300 iterations and we fix N at 15000.

We measure the relative efficiency of our algorithm with respect to the MH algorithm with the ratio of their corresponding ESS and AD values. In particular, we define this ratio as $RESS = ESS_G/ESS_{MH}$, where ESS_G is the ESS value for our algorithm. Similarly, we define the relative gains in update distance as $RAD = AD_G/AD_{MH}$. We compare the algorithms for a range of values of n and r. For each pair (n, r) we generate 100 fictitious samples and run both algorithms for each of these samples. The average values of $RESS$, RAD, ESS_G, ESS_{MH} and their standard deviations together with other summaries are reported in Table 1. When $n = 2$ and $r = 1$, our algorithm is almost as efficient as independent sampling ($ESS = 0.95$), whereas the MH algorithm is about 8.4 times less efficient (in terms of ESS). Moreover, the relative gains increase as n gets larger. For $n = 6$ and $r = 3$ our algorithm is on average 666 times more efficient, and for $n = 9$ and $r = 5$ it is about 1270 times more efficient. Substantial improvements are also observed in terms of the average update distance. Furthermore, our algorithm computes 15000 iterations slightly more quickly than the MH algorithm

(RT), suggesting that the gains adjusted for computation time are even higher.

References

Abadir, K. and Magnus, J., 2005, *Matrix Algebra*. Cambridge: Cambridge University Press.

Bauwens, L. and Lubrano, M., 1996, Identification restrictions and posterior densities in cointegrated Gaussian VAR systems, in *Advances in Econometrics* 11, Part B, JAI Press, pages 3-28.

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

Brooks, S., 1999, Bayesian analysis of animal abundance data via MCMC, in *Bayesian Statistics 6* (ed. J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith). Oxford: Clarendon Press.

Cadet, A., 1996, Polar coordinates in the R^{np}; Application to the computation of the Wishart and Beta Laws, *Sankhya: The Indian Journal of Statistics* 58, 101-114.

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

Chikuse, Y., 1990, The matrix angular central Gaussian distribution, *Journal of Multivariate Analysis* 33, 265-274.

DeJong, D., 1992, Co-integration and trend-stationarity in macroeconomic time series," *Journal of Econometrics* 52, 347-370.

Dorfman, J., 1994, A numerical Bayesian test for cointegration of AR processes, *Journal of Econometrics* 66, 289-324.

Geweke, J., 1996, Bayesian reduced rank regression in econometrics, *Journal of Econometrics* 75, 121-146.

Geyer, C.J., 1992, Practical Markov chain Monte Carlo, *Statistical Science* 7, 473-511.

Holmes, C.C. and Held, K., 2004, Bayesian auxiliary variable models for binary and polychotomous regression, forthcoming in *Bayesian Analysis.*

Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Kleibergen, F. and Paap, R., 2002, Prior, posteriors and bayes factors for a Bayesian analysis of cointegration, *Journal of Econometrics* 111, 223-249.

Kleibergen, F. and van Dijk, H.K., 1994, On the shape of the likelihood/posterior in cointegration models, *Econometric Theory* 10, 514-551.

Kleibergen, F. and van Dijk, H.K., 1998, Bayesian simultaneous equations analysis using reduced rank structures, *Econometric Theory* 14, 701-743.

Koop, G., 1994, An objective Bayesian analysis of common stochastic trends in international stock prices and exchange rates, *Journal of Empirical Finance* 1, 343-364.

Koop, G., Strachan, R., van Dijk, H.K. and Villani, M., 2005, Bayesian approaches to cointegration. To appear in T.C. Mills and K. Patterson (eds.). *Palgrave Handbook of Theoretical Econometrics*, manuscript available at http://www.le.ac.uk/economics/research/RePEc/lec/leecon/dp04-27.pdf.

Larsson, R. and Villani, M., 2001, A distance measure between cointegration spaces, *Economics Letters* 70, 21-27.

Liu, J.S., 1994, The collapsed Gibbs sampler with applications to a gene regulation problem, *Journal of the American Statistical Association* 89, 958-966.

Liu, J.S., Wong, W.H., and Kong, A., 1994, Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes, *Biometrika* 81, 27-40.

Strachan, R., 2003, Valid Bayesian estimation of the cointegrating error correction model, *Journal of Business and Economic Statistics* 21, 185-195.

Strachan, R. and Inder, B., 2004, Bayesian analysis of the error correction model, *Journal of Econometrics* 123, 307-325.

Strachan, R. and van Dijk, H., 2004, Valuing structure, model uncertainty and model averaging in vector autoregressive processes, Econometric Institute Report EI 2004-23, Erasmus University Rotterdam.

Villani, M., 2005a, Bayesian reference analysis of cointegration, *Econometric Theory* 21, 326-357.

Villani, M., 2005b, Bayesian point estimation of the cointegration space, forthcoming in *Journal of Econometrics.*

	$RESS$	RAD	RT	ESS_G	ESS_{MH}	AR	σ
$n=2,\ r=1$	8.4	3.7	0.88	0.95	0.11	0.49	0.025
	(2.00)	(0.271)	(0.100)	(0.046)	(0.016)	(0.108)	
$n=3,r=2$	19.6	6.4	0.97	0.95	0.050	0.37	0.020
	(4.69)	(0.966)	(0.048)	(0.054)	(0.009)	(0.097)	
$n=3,r=1$	28.7	8.25	0.93	0.83	0.035	0.36	0.025
	(15.2)	(3.16)	(0.025)	(0.124)	(0.014)	(0.078)	
$n=4,r=3$	37.0	10.1	0.98	0.928	0.0267	0.30	0.020
	(11.2)	(2.74)	(0.007)	(0.069)	(0.006)	(0.106)	
$n=4,r=2$	125	21.2	0.97	0.763	0.00902	0.30	0.015
	(116)	(12.1)	(0.011)	(0.145)	(0.005)	(0.076)	
$n=4,r=1$	65	14	0.95	0.647	0.0124	0.35	0.015
	(37)	(4.9)	(0.02)	(0.150)	(0.006)	(0.06)	
$n=5,r=3$	303	40	0.98	0.71	0.0043	0.30	0.01
	(278)	(26)	(0.054)	(0.123)	(0.003)	(0.088)	
$n=5,r=2$	343	39.4	0.971	0.592	0.0027	0.35	0.008
	(308)	(17.8)	(0.023)	(0.118)	(0.0017)	(0.07)	
$n=6,r=4$	436	51.7	0.99	0.700	0.0031	0.37	0.006
	(571)	(39.2)	(0.007)	(0.124)	(0.002)	(0.10)	
$n=6,r=3$	666	83.1	0.975	0.565	0.00142	0.275	0.008
	(695)	(48)	(0.057)	(0.128)	(0.001)	(0.068)	
$n=9,r=5$	1270	207	0.995	0.50	0.00054	0.29	0.004
	(1130)	(130)	(0.0041)	(0.095)	(0.0003)	(0.083)	

Table 1: Performance measures of our algorithm and the MH algorithm, averaged over 100 fictitious samples for each value of (n,r). The columns $RESS, RAD, ESS_G, ESS_{MH}$ are defined in the text and standard deviations are in parentheses. RT is the computation time needed for our algorithm to perform 15000 iterations divided by the time needed by the MH algorithm. AR is the acceptance rate in the MH algorithm, which is controlled by σ.