Research Report

Department of Statistics

No. 2001:1

Bayesian Reference Analysis of Cointegration

Mattias Villani

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

BAYESIAN REFERENCE ANALYSIS OF COINTEGRATION

MATTIAS VILLANI

Abstract. A Bayesian reference analysis of the cointegrated vector autoregression is presented based on a new prior distribution. Among other properties, it is shown that this prior distribution distributes its probability mass uniformly over all cointegration spaces for a given cointegration rank. A simple procedure based on the Gibbs sampler is used to obtain the posterior distribution of both the number of cointegrating relations and the form of those relations, together with accompanying error correcting dynamics. Simulated data are used for illustration and for discussing the well-known issue of local non-identification.

1. Introduction

Many macroeconomic time series behave in a random walk-like fashion and tend to move around wildly. Typically, such variables move around together, striving to fulfill one or several economic laws, or long run equilibria, which tie them together. A random walk is often referred to as an *integrated* process, and integrated processes that move around together have therefore been termed *cointegrated*(Engle and Granger, 1987).

The present work is concerned with estimation of both the *number* of equilibria, the so called cointegration rank, and the *form* of the equilibria conditional on the rank. Inferences regarding the error correcting coefficients and other short run dynamics are also treated.

Several non-Bayesian statistical treatments of cointegration have been presented during the last two decades, see e.g. Ahn and Reinsel (1990), Phillips (1991) and Stock and Watson (1988). The likelihood-based approach of Søren Johansen is the most widely used procedure, however; see e.g. Johansen (1991) or Johansen (1995) for a textbook treatment.

More recently, a handful Bayesian analyses of cointegration have been presented, see Bauwens and Giot (1998), Bauwens and Lubrano (1996), Geweke (1996), Kleibergen and Paap (1998), Kleibergen and van Dijk (1994), Strachan (2000) and Villani (2000). Philosophical issues aside, a Bayesian approach is advantageous for many reasons, e.g. it produces whole probability distributions for each unknown parameter which are valid for any sample size, it affords a straight-forward handling of the inferences for the cointegration rank and tests of restrictions on the model parameters (Geweke, 1996; Kleibergen and Paap, 1998; Strachan, 2000; Villani, 2000), and it makes a satisfactory treatment of the prediction problem possible (Villani, 2001b).

The crucial step in a Bayesian analysis is the choice of prior distribution and in each of the above mentioned papers a new prior distribution has been introduced. The degree of motivation of the priors has varied, but the authors seem to have been more or less

---

Key words and phrases. Cointegration, Bayesian inference, Estimation, Rank.

The author thanks Luc Bauwens, Daniel Thorburn and Herman van Dijk for helpful discussions. Financial support from the Swedish Council of Research in Humanities and Social Sciences (HSFR) is gratefully acknowledged.

focused on vague priors which add only a small amount of information into the analysis, i.e. priors largely dominated by data.

This paper will be less concerned with whether or not the prior is 'non-informative'. The aim here is to propose a Bayesian analysis based on a sound prior which appeals to practitioners. Such a prior must consider several partially conflicting aspects of actual econometric practise: Firstly, the number of parameters in cointegration models is usually very large and it is not realistic to demand a detailed subjective specification of priors on such high-dimensional spaces, at least not at the current state of elicitation techniques for multivariate distributions. A prior with relatively few hyperparameters, each with a clear interpretation, is thus mandatory. Secondly, priors will not, or at least should not, be used by practitioners unless they are transparent in the sense that one can easily understand the kind of information they convey. Thirdly, the prior must lead to straight-forward posterior calculations which can be performed on a routine basis without necessary fine tuning in each new application. Finally, the posterior distribution of the cointegration rank can only be obtained if some parameter matrices are given proper, integrable, priors. A prior which fulfills these objectives will probably not coincide with the investigators actual prior beliefs, but should nevertheless be useful as point of reference, or an agreed standard, and is called a *reference* prior accordingly.

2. The model

Let $\{x_t\}_{t=1}^{T}$ be a vector-valued process with p components. The most widely used model for non-stationary time series is the Error Correction (EC) model

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t, \tag{2.1}$$

where $\Delta x_{t-i} = x_{t-i} - x_{t-i-1}$, Π $(p \times p)$ is the matrix of long-run multipliers and Γ_i $(p \times p)$, for $i = 1, ..., k-1$, govern the short run dynamics of the process. d_t $(w \times 1)$ is a vector of trend, seasonal dummies or other exogenous variables with coefficient matrix Φ $(p \times w)$. ε_t $(p \times 1)$ contains the unexplained errors at time t which are assumed to follow the $N_p(0, \Sigma)$ distribution with independence across time periods. The number of lagged differences, $k-1$, will be assumed known or determined before the analysis, see Villani (2001a) for a Bayesian approach.

The stochastic part of a process is integrated of order zero, denoted $x_t \sim I(0)$, if it is stationary while its cumulative sum is not (see Johansen, 1995, p.35 for a more precise definition). Furthermore, $x_t \sim I(h)$ if $\Delta^h x_t \sim I(0)$, i. e. x_t is integrated of order h if the hth difference of the process is integrated of order zero.

Even if all p time series in x_t are $I(1)$, it can be shown that the condition rank$(\Pi) = r < p$ implies that r linear combinations of the time series are stationary, see Engle and Granger (1987) and Johansen (1995, p.49); the time series are then *cointegrated*. Cointegration is thus equivalent to the existence of r long run equilibria between otherwise drifting series. If rank$(\Pi) = r$, then Π can be written as a product of two $p \times r$ full rank matrices α and β, and the model in (2.1) can be written

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t, \tag{2.2}$$

where $\beta' x_{t-1} - E(\beta' x_{t-1})$ are the deviations from the r equilibria and α contains the adjustment coefficients governing the dynamics back to equilibrium after a disturbance. The ith column of β contains the coefficients of the ith equilibrium and is called a cointegration vector accordingly.

It is well known that only the space spanned by the cointegration vectors (sp β), the *cointegration space*, is identified, i.e. β is only determined up to arbitrary linear combinations of its columns. We will follow the traditional route in Bayesian analyses of cointegration by using a linear normalization

$$\beta = \begin{pmatrix} I_r \\ B \end{pmatrix} \tag{2.3}$$

to settle this indeterminacy, where B is a $(p-r)\times r$ matrix of fully identified parameters. When β is used as an argument in density functions it must remembered that some of its elements are known with probability one as a result of the normalization.

The linear normalization is very convenient for computational reasons (see Sections 4 and 5) but has the limitation that the existence of B depends on the invertibility of the matrix formed by the r first rows in β. Put differently, the linear normalization implicitly assumes that each of the r first components of x_t enter at least one of the r cointegrating relations, see Strachan (2000) for a thorough discussion.

The linear normalization may blur the interpretation of the estimated cointegration vectors, but this is easily remedied by a simple transformation of the posterior results, see Section 4.

3. The prior distribution

The prior distribution is conveniently decomposed as

$$p(\alpha, \beta, \Gamma, \Sigma, r) = p(\alpha, \beta, \Gamma, \Sigma | r) p(r),$$

where $p(r)$ is any probability distribution over the possible cointegration ranks, $r = 0,1,...,p$.

The essential conceptual difficulty in a Bayesian approach to cointegration is the prior distribution of α and β. Early contributions include Kleibergen and van Dijk (1994) who suggested the Jeffreys prior, Bauwens and Lubrano (1996) with a uniform prior on α and student t priors on the cointegration vectors and Geweke's (1996) normal shrinkage priors. The Jeffreys prior is complicated, student t priors are used for convenience in the posterior calculation and the normal shrinkage prior is used to assure the convergence of the algorithm for evaluating the posterior distribution.

Recently, Kleibergen and Paap (1998) proposed a reference prior on α and β which is essentially a prior on Π in the full rank EC model projected down to the subspace where $\text{rank}(\Pi) = r$; Strachan (2000) extended this idea to more general identifying restrictions. It is an approach which is rather common and well understood in linear models, but its implications in non-linear models, such as the EC model with reduced rank in (2.2), are not as transparent, see also Section 6.

The approach taken here differs from the above mentioned works by focusing directly on the structure of the parameter space of β, which is non-standard as a result of the identification problem pointed out in Section 2. We introduce the proposed reference prior now and spend the rest of this section motivating its particular form. Let $\Gamma = (\Gamma_1, ..., \Gamma_{k-1}, \Phi)'$ and $\text{etr}(H) = \exp(-\frac{1}{2}\,\text{tr}\, H)$, for any square matrix H. The prior can then be written

$$p(\alpha, \beta, \Gamma, \Sigma | r) = c_r \left|\Sigma\right|^{-(p+r+q+1)/2} \operatorname{etr}[\Sigma^{-1}(A + v\alpha\beta'\beta\alpha')], \tag{3.1}$$

where $v > 0$, $q \geq p$ and A, an $p \times p$ positive definite matrix, are the three hyperparameters to be specified by the investigator. The normalizing constant is

$$c_r = |A|^{q/2} \frac{\Gamma_r(p)}{\Gamma_p(q)\Gamma_r(r)} \frac{2^{-qp/2}\pi^{-p(p-1)/4}}{(2\pi/v)^{pr/2}\pi^{(p-r)r/2}},$$

where $\Gamma_b(a) = \prod_{i=0}^{b-1} \Gamma[(a-i)/2]$, for positive integers a and b satisfying $a \geq b-1$.

Note that Γ is uniformly distributed over $\mathbb{R}^{(p-1)k+w}$, which makes the overall prior $p(\alpha, \beta, \Gamma, \Sigma | r)$ improper. The prior on α, β and Σ conditional on Γ is proper, however. The uniform prior for Γ is used here for simplicity, but a general multivariate normal prior on $\operatorname{vec}\Gamma$ (e.g. a structured shrinkage prior as in Litterman, 1986) leads to essentially the same posterior computations.

Implicit in (3.1) is the assumption of common A, q and v for all r; the ensuing analysis proceeds in the same manner in the general case with varying A, q and v.

3.1. **Marginal and conditional distributions.** Throughout this section, we will assume that $k = 1$ and $w = 0$, for notational convenience. The results will still be valid for $k > 1$ and $w > 0$ as long as prior independence between $(\Gamma_1, ..., \Gamma_{k-1}, \Phi)$ and the other parameter matrices is assumed. All probability distributions in this section will be conditional on a given cointegration rank, though this will not be written out explicitly.

In order to derive the relevant marginal and conditional distributions a few definitions and lemmas are needed.

Definition 3.1. The *Grassman manifold*, $\mathcal{G}_{p,r}$, is the space of all r-dimensional subspaces of $\mathbb{R}^p$.

As only the space spanned by the columns of β is unique, $\beta \in \mathcal{G}_{p,r}$. The usual way to define the uniform distribution over $\mathcal{G}_{p,r}$ is given in the next definition.

Definition 3.2. The uniform distribution on $\mathcal{G}_{p,r}$ is the Haar invariant distribution under the group of orthonormal transformations.

It can be shown that this uniform distribution on $\mathcal{G}_{p,r}$ is unique, see James (1954) for a proof and more details regarding distributions on the Grassman manifold.

Definition 3.3. A $m \times s$ matrix A follows the matrix normal distribution, $A \sim N_{m \times s}(\mu, \Omega_1, \Omega_2)$, if and only if $\operatorname{vec} A \sim N_{ms}(\operatorname{vec}\mu, \Omega_1 \otimes \Omega_2)$, where $\operatorname{vec} A$ is the ms-dimensional vector obtained by stacking the s columns below each other and $\otimes$ is the Kronecker product.

Definition 3.4. A $m \times s$ matrix D follows the matrix t distribution, $D \sim t_{m \times s}(\mu, \Upsilon, \Theta, g)$, if its density is given by

$$\frac{\Gamma_s(g+m+s-1)}{\Gamma_s(g+s-1)\pi^{ms/2}\left|\Upsilon\right|^{s/2}\left|\Theta\right|^{m/2}} \left|I_s + \Theta^{-1}(D-\mu)'\Upsilon^{-1}(D-\mu)\right|^{-(g+m+s-1)/2}.$$

See Box and Tiao (1973) and Bauwens et al (1999) for properties of the matrix t distribution.

Lemma 3.5. *If* $\beta = (I_r \; B')'$ *and* $B \sim t_{(p-r)\times r}(0, I_{p-r}, I_r, 1)$, *then* β *is uniformly distributed over* $\mathcal{G}_{p,r}$.

Proof. See Villani (2000). □

With the preceding definitions and lemmas out of the way, we are now prepared to state an important characterization of the distribution in (3.1).

Theorem 3.6. *β is marginally uniformly distributed over $\mathcal{G}_{p,r}$.*

Proof. See the appendix. □

Thus, the prior in (3.1) assigns equal probability to every possible cointegration space of dimension r. This should be a sensible reference prior given that only the cointegration space is identified.

Theorem 3.7. *The marginal prior of Σ is*

$$\Sigma \sim IW(A, q),$$

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

Proof. Follows directly from the proof of Theorem 3.6. □

It is easily shown that (see the proof of Theorem 3.8 below)

$$\alpha|\beta, \Sigma \sim N_{p\times r}[0, (\beta'\beta)^{-1}, v^{-1}\Sigma]. \tag{3.2}$$

The linear normalization of β makes α difficult to interpret, however, and the conditional prior in (3.2) may not shed much light on the prior in (3.1). Consider instead the prior of α conditional on β and Σ when β is orthonormal. Restricting β to be orthonormal is not sufficient to identify the model, however, as any orthonormal version of β can be rotated to a new one by postmultiplying it with an $r \times r$ orthonormal matrix. This need not concerns us here as β only enters $p(\alpha|\beta, \Sigma)$ in the form $\beta'\beta$ and $p(\alpha|\beta, \Sigma)$ is therefore invariant under these rotations. Define $\tilde{\beta} = \beta(\beta'\beta)^{-1/2}$ and note that $\tilde{\beta}$ is orthonormal. In order for $\Pi = \alpha\beta'$ to remain unchanged by the transformation $\beta \to \tilde{\beta}$, we must make the corresponding transformation of the adjustment matrix from α to $\tilde{\alpha} = \alpha(\beta'\beta)^{1/2}$. In the following theorem, let $\tilde{\alpha}_i$ denote the ith column of $\tilde{\alpha}$ and note that $\tilde{\alpha}_i$ describes how the p response variables are affected by the ith cointegrating relation under the orthonormal normalization.

Theorem 3.8. $\tilde{\alpha}_i|\Sigma \overset{iid}{\sim} N_p(0, v^{-1}\Sigma),\ i = 1, 2, ..., r.$

Proof. See the appendix. □

The rather restrictive form of the prior in Theorem 3.8 must be motivated. First, the restriction to conditional normal priors on α (and thereby also on $\tilde{\alpha}$) is necessary for an efficient numerical evaluation of the posterior based on Gibbs sampling, see Section 4. Second, non-identical priors on the columns of $\tilde{\alpha}$ do not make sense unless over-identifying restrictions on the columns of β are used to give a unique meaning to each cointegration vector, i.e. it is not reasonable to express different beliefs about the columns of $\tilde{\alpha}$ without knowing which cointegration vector each $\tilde{\alpha}_i$ refers to. Another way to see this is that within the class of matrix normal priors $\tilde{\alpha}|\tilde{\beta}, \Sigma \sim N_{p\times r}(\mu, \Omega_1, \Omega_2)$, only the priors with $\mu = 0$, $\Omega_1 = I_r$ are invariant to rotations of $\tilde{\beta}$. Third, the reason for centering the conditional prior over zero is motivated by the invariance requirement just stated. It has the effect of centering the prior over $\Pi = 0$, which is often a good starting point in an analysis, see the discussion of the 'sum of coefficients' prior in Doan et al (1984) and the section on prior stability below. Finally, the scale matrix in the conditional prior could be any positive definite matrix, the posterior computations remain almost exactly the same. By making the conditional covariance matrix proportional to Σ we incorporate the belief that the uncertainty of a variables error-correcting coefficients is proportional to size of the unexplained component of that variable.

Further clarification of the hyperparameters A, q and v is obtained from the marginal prior of $\tilde{\alpha}$. By multiplying $p(\tilde{\alpha}|\Sigma)$ with the marginal inverted Wishart prior of Σ and integrating with respect to Σ we obtain

$$\tilde{\alpha} \sim t_{p\times r}(0, v^{-1}A, I_r, q-p+1).$$

Results in Box and Tiao (1973, p. 446-7) then give

$$E(\tilde{\alpha}) = 0 \text{ and } Cov(\text{vec}\,\tilde{\alpha}) = I_r \otimes v^{-1}E(\Sigma),$$

where $E(\Sigma) = A/(q-p-1)$ is the expected value of Σ *a priori*, see e.g. Bauwens et al (1999, p. 306).

A is determined from $E(\Sigma)$ and q and the investigator thus faces subjective specification of: 1) the expected value of Σ, 2) the degree of certainty regarding Σ (large values of q imply large certainty) and 3) the tightness around the point zero for $\tilde{\alpha}$ (large values of v give high concentration of probability mass around zero). Note that whether a value for v is large or not depends on $E(\Sigma)$, which should therefore be specified before v.

The main difficulty for the investigator is likely to be the specification of $E(\Sigma)$. If interest only centers on the posterior of $\alpha, \beta, \Gamma, \Sigma$ conditional on a given cointegration rank, then A may be set equal to the zero matrix and $q = 0$. This corresponds to using the usual improper prior $p(\Sigma) \propto |\Sigma|^{-(p+1)/2}$. If we also aim at analyzing the cointegration rank, but are either unable or unwilling to state our beliefs about Σ, then $A = \hat{\Sigma}$ and $q = p + 2$ may be used, where $\hat{\Sigma}$ is the ML estimate of Σ in the full rank model; note that this implies that $E(\Sigma) = \hat{\Sigma}$. This suggestion is of course not a proper Bayesian solution as the prior then becomes dependent on the observed data. The consequences of this side-step are minimized by the choice of the smallest possible q (maximum uncertainty) subject to a finite expected value of Σ.

3.2. **Invariance.** The choice of normalizing variables may be somewhat arbitrary and it is therefore desirable to have a posterior distribution which is invariant to this choice. The likelihood function is well-known to have this property and the posterior distribution is therefore invariant if and only if the prior distribution is. Let $\mathcal{N}_1 = \{i_1, ..., i_r\}$ denote the set of indices for the r variables used to normalize β. Consider the change in normalization $\mathcal{N}_1 \to \mathcal{N}_2$, where $\mathcal{N}_2$ equals $\mathcal{N}_1$ with jth variable in the normalized set replaced by the kth variable in the non-normalized set. This change in normalizing variables is accomplished by the transformation $(\alpha, \beta, \Sigma) \to (\bar{\alpha}, \bar{\beta}, \Sigma)$, where $\bar{\alpha} = \alpha U'$, $\bar{\beta} = \beta U^{-1}$ and U is an $r \times r$ transformation matrix whose elements are functions of the kth row of B. The exact form of U need not concern us for the moment, it is sufficient to note that such a matrix always exist if the variables in the new normalizing set enter at least one of the cointegration vectors (see Section 2) and that U then is unique. It is important to note that $\Pi = \alpha\beta'$ is unchanged by the transformation.

The next definition formalizes the idea that the prior distribution (and therefore also the posterior distribution) should be the same whether we i) work directly with $\mathcal{N}_1$ or ii) start with $\mathcal{N}_2$ and then transform to $\mathcal{N}_1$.

Definition 3.9. A distribution on α, β and Σ is *normalization invariant* if, for any U corresponding to a change in normalizing variables,

$$p(\alpha, \beta, \Sigma) = p(\bar{\alpha}, \bar{\beta}, \Sigma) J(\bar{\alpha}, \bar{\beta}, \Sigma \to \alpha, \beta, \Sigma),$$

where $\bar{\alpha} = \alpha U'$, $\bar{\beta} = \beta U^{-1}$ and $J(\bar{\alpha}, \bar{\beta}, \Sigma \to \alpha, \beta, \Sigma)$ is the Jacobian of the transformation from $\bar{\alpha}, \bar{\beta}, \Sigma$ to α, β, Σ.

Lemma 3.10. $J(\bar{\alpha}, \bar{\beta}, \Sigma \to \alpha, \beta, \Sigma) = 1$.

Proof. See the appendix. □

Theorem 3.11. *The prior in (3.1) is normalization invariant.*

Proof.

$$\begin{aligned} p(\bar{\alpha}, \bar{\beta}, \Sigma) &= c_r \, |\Sigma|^{-(p+r+q+1)/2} \operatorname{etr}[\Sigma^{-1}(A + v\alpha U' U^{-1\prime} \beta' \beta U^{-1} U \alpha')] \\ &= p(\alpha, \beta, \Sigma), \end{aligned}$$

which, using Lemma 3.10, proves the theorem. □

3.3. **Prior stability.** Define

$$\Pi_C = \begin{pmatrix} I_p + \alpha\beta' + \Gamma_1 & \Gamma_2 - \Gamma_1 & \cdots & \Gamma_{k-1} - \Gamma_{k-2} & -\Gamma_{k-1} \\ I_p & 0 & \cdots & 0 & 0 \\ 0 & I_p & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_p & 0 \end{pmatrix}$$

The assumption of $rank(\Pi) = r$ implies that r of the eigenvalues of Π_C are equal to one. A cointegrated process is stable if all the remaining eigenvalues of Π_C are smaller than one in modulus. It is clearly of interest to know what prior probability is implicitly being placed on the set of stable processes if the prior in (3.1) is used. This could be investigated either by analytical approximation or by simulation methods for different models, i.e. by varying p, r and k. We shall here be content with simulating the special case $p = 2$ and $r = k = 1$. Table I displays the prior probability that the process is stable for $A = I_2$ as a function of q and $\sigma = v^{-1/2}$ (note that σ is on a standard deviation scale). Experiments with other choices of A with strong positive and negative correlation structure did not have a large impact on the probability. Note also that it is unnecessary to increase the magnitude of the diagonal elements in A as this has the same effect as increasing σ.

Table I

Implied Prior Probability that the Process is Stable

	σ									
	0.01	0.1	0.25	0.5	0.75	1	5	10	50	100
$q = 2$	0.48	0.46	0.40	0.35	0.30	0.26	0.08	0.04	0.01	0.00
$q = 4$	0.49	0.49	0.47	0.46	0.42	0.40	0.15	0.08	0.02	0.01
$q = 10$	0.50	0.50	0.49	0.48	0.47	0.45	0.26	0.14	0.03	0.02
$q = 20$	0.50	0.49	0.49	0.48	0.47	0.47	0.33	0.19	0.05	0.02

Note: $A = I_2$.

INSERT FIGURE HERE

FIGURE 1. Implied prior distribution on the unrestricted eigenvalue. $A = I_2$ and $q = 4$.

Densities of the unrestricted eigenvalue (λ) are displayed in Figure 1 for different values of σ. The densities are symmetric around the modal value $\lambda = 1$. A non-symmetric density for λ which places more mass to the left of $\lambda = 1$ than to the right of this point would perhaps better represent actual beliefs. The gain from a non-symmetric prior is probably less than the loss in computational efficiency in the posterior calculations, however.

A crude way to obtain a non-symmetric prior is to simply exclude explosive processes *a priori* (or 'too explosive' processes, e.g. with eigenvalues larger than 1.1 in modulus) by restricting the domain of the prior in (3.1) to the space of α, β and Γ where the process is stable. This is neatly handled in the posterior calculations for a given cointegration rank by simply rejecting the draws from the posterior corresponding to non-stable processes, see Section 4. Note that the latter region will be small if the process actually is stable and data informative, and most draws will then be accepted. The posterior distribution of the rank will require heavier numerical computations, however.

4. The posterior distribution conditional on the rank

4.1. **Preliminaries.** Let us write the cointegrated EC model (2.2) in the following compact form

$$Y = X\beta\alpha' + Z\Gamma + E, \tag{4.1}$$

where $Y = (\Delta x_T, ..., \Delta x_1)'$, $X = (x_{T-1}, ..., x_0)'$, $Z = (\Delta X_{-1}, ..., \Delta X_{-k+1}, D)$, $\Delta X_{-i} = (\Delta x_{T-i}, ..., \Delta x_{-i+1})'$, $D = (d_T, ..., d_1)'$, $\Gamma = (\Gamma_1, ..., \Gamma_{k-1}, \Phi)'$ and $E = (\varepsilon_T, ..., \varepsilon_1)'$. $\mathcal{D} = \{Y, X, Z\}$ will be used as a short hand for the available data and $d = (p-1)k + w$ denotes the number of columns in Z. Further, define

$$M_H = I_m - H(H'H)^{-1}H', \tag{4.2}$$

for any $m \times s$ matrix H.

The following two lemmas will be used in some of the proofs later on.

Lemma 4.1.

$$p(\alpha, \beta|\mathcal{D}, r) \propto \left|(Y - X\beta\alpha')'M_Z(Y - X\beta\alpha') + A + v\alpha\beta'\beta\alpha'\right|^{-(T+q+r-d)/2}.$$

Proof. See the appendix. □

Lemma 4.2. *For any non-singular $m \times s$ matrix H*

$$\int_{-\infty}^{\infty} \left|A + (H-M)'B(H-M)\right|^{-h/2} dH = \pi^{ms/2}\frac{\Gamma_s(h-m)}{\Gamma_s(h)} |A|^{-(h-m)/2} |B|^{-s/2},$$

where A $(s \times s)$ and B $(m \times m)$ are positive definite matrices, and M $(m \times s)$ is of full rank.

Proof. The integrand is the kernel of a matrix t and the lemma is proved by using the fact that a density integrates to one. □

4.2. **Marginal posterior distribution of β.** The next result gives the marginal posterior of the cointegration vectors.

Theorem 4.3. *The marginal posterior distribution of β is*

$$p(\beta|\mathcal{D}, r) \propto \frac{|\beta' C_1 \beta|^{(T+q-d-p)/2}}{|\beta' C_2 \beta|^{(T+q-d)/2}},$$

where $C_1 = X'M_Z X + vI_p$, $C_2 = vI_p + X'Q[I_T - Z(Z'QZ)^{-1}Z'Q]X$ and $Q = I_T - Y(A + Y'Y)^{-1}Y'$.

Proof. See the appendix. □

$p(\beta|\mathcal{D}, r)$ in Theorem 4.3 is thus a ratio of two matrix t kernels and could therefore be termed a matrix 1-1 poly-t density, see Bauwens et al (1999) for a discussion of multivariate poly-t densities. Contrary to the multivariate case, no properties have been derived for matrix poly-t densities and Theorem 4.3 is therefore of little practical interest, in general.

In the special case $r = 1$, the posterior distribution of B is a (vector-valued) 1-1 poly-t density and it can be been shown (Bauwens and Lubrano, 1996) that $p(B|\mathcal{D}, r)$ is integrable but has no finite moments. The non-existence of integer moments is not a consequence of the prior distribution in (3.1), but rather of the linear normalization of β, where each element of B is a ratio with the first element of β in the denominator.

4.3. **Conditional posteriors and Gibbs sampling.** An alternative route to calculate the posterior distribution of α, B, Γ and Σ using Gibbs sampling (see e.g. Smith and Roberts, 1993, for a good introduction) was pioneered by Geweke (1996).

The Gibbs sampler is an easily implemented iterative method for generating observations from complex multidimensional densities by sampling from the so called *full conditional posterior distributions*. The full conditional posterior distribution of a subset of parameters in a model is the posterior distribution of the subset conditional on all other parameters. A frequently encountered situation is that the overall posterior distribution is too complex to generate samples from while the full conditional posteriors are all easily sampled. The Gibbs sampler exploits this fact and produces a sample from the posterior by iteratively generating parameter values from the full conditional posteriors.

The sampled parameter values are not independent, but following can be shown to hold under certain conditions which are satisfied here (Geweke, 1996)

$$\begin{aligned} \theta^{(i)} &\overset{d}{\to} p(\theta|\mathcal{D}), \\ N^{-1}\sum_{i=1}^{N} f(\theta^{(i)}) &\overset{a.s.}{\to} E_{\theta|\mathcal{D}}[f(\theta)], \end{aligned}$$

where θ is a parameter vector of interest, $\theta^{(i)}$ is the sampled value of θ at the ith Gibbs iteration, $\overset{d}{\to}$ and $\overset{a.s.}{\to}$ denote convergence in distribution and convergence almost surely, respectively, and $p(\theta|\mathcal{D})$ is the posterior distribution of θ conditional on data $\mathcal{D}$. $f(\cdot)$ is any well-behaved function with finite posterior expectation and $E_{\theta|\mathcal{D}}(\cdot)$ denotes the posterior expectation.

Initial values for all parameters are needed to start up the Gibbs sampler. The maximum likelihood estimates in Johansen (1995) are natural candidates.

Theorem 4.4.

- *The full conditional posterior of* Σ
$$\Sigma|\alpha, \beta, \Gamma, \mathcal{D}, r \sim IW_p(E'E + A + v\alpha\beta'\beta\alpha', T + q + r),$$
where $E = Y - X\beta\alpha' - Z\Gamma$.
- *The full conditional posterior of* Γ
$$\Gamma|\alpha, \beta, \Sigma, \mathcal{D}, r \sim N_{d\times p}[\mu_\Gamma, \Sigma, (Z'Z)^{-1}],$$
where $\mu_\Gamma = (Z'Z)^{-1}Z'(Y - X\beta\alpha')$.
- *The full conditional posterior of* α
$$\alpha|\beta, \Gamma, \Sigma, \mathcal{D}, r \sim N_{p\times r}\left\{\mu_\alpha, [\beta'(X'X + vI_p)\beta]^{-1}, \Sigma\right\},$$
- *The full conditional posterior of* B
$$B|\alpha, \Gamma, \Sigma, \mathcal{D}, r \sim N_{(p-r)\times r}[\mu_B, (\alpha'\Sigma^{-1}\alpha)^{-1}, (X_2'X_2 + vI_{p-r})^{-1}],$$
where $\mu_B = (X_2'X_2 + vI_{p-r})^{-1}X_2'(Y - X_1\alpha' - Z\Gamma)\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}$, X_1 *denotes the* r *first columns of* X *and* X_2 *the* $p - r$ *last ones.*

Proof. See the appendix. □

It is often the case that α and β is the focus of attention, whereas Γ and Σ are considered nuisance parameters, or at least parameters of lesser interest. Furthermore, most of the model parameters are located in Γ and Σ and the Gibbs updating steps for these two matrices usually dominates the total computing time. The time to convergence of the Gibbs sampler also increases as the dimensions of Γ and Σ grow. The next theorem gives the conditional posteriors necessary to perform a (marginal) Gibbs sampler to generate samples directly from $p(\alpha, B|\mathcal{D}, r)$. The marginal Gibbs sampler in the following theorem is also used in Section 5 to calculate the posterior distribution of the rank.

Theorem 4.5.

- *The posterior of* α *conditional on* β *and* r
$$\alpha|\beta, \mathcal{D}, r \sim t_{p\times r}[\hat{\alpha}, A + Y'M_Z(Y - X\beta\hat{\alpha}'), (\beta'C_1\beta)^{-1}, T + q - (d + p) + 1]$$
where $\hat{\alpha} = Y'M_ZX\beta(\beta'C_1\beta)^{-1}$.
- *The posterior of* B *conditional on* α *and* r
$$B|\alpha, \mathcal{D}, r \sim t_{(p-r)\times r}[\hat{B}, G_3 - G_2'G_1^{-1}G_2, C_3, T + q + r - (d + p) + 1],$$
where $\hat{\beta} = \hat{\Pi}'S^{-1}\alpha(\alpha'S^{-1}\alpha)^{-1}$, $\hat{\Pi} = Y'M_ZXC_1^{-1}$, $S = A + Y'M_ZY - \hat{\Pi}C_1\hat{\Pi}'$, $\hat{\beta}_1$ *contains the* r *first rows of* $\hat{\beta}$ *and* $\hat{\beta}_2$ *the* $p - r$ *remaining ones, and*
$$C_1^{-1} + \hat{\Pi}'S^{-1}\hat{\Pi} - \hat{\beta}\alpha'S^{-1}\alpha\hat{\beta}' = \begin{pmatrix} \underset{r\times r}{G_1} & \underset{r\times(p-r)}{G_2} \\ \underset{(p-r)\times r}{G_2'} & \underset{(p-r)\times(p-r)}{G_3} \end{pmatrix}$$
is decomposed conformably, $C_3 = (I_r - \hat{\beta}_1)'G_1^{-1}(I_r - \hat{\beta}_1) + (\alpha'S^{-1}\alpha)^{-1}$ *and* $\hat{B} = \hat{\beta}_2 + G_2'G_1^{-1}(I_r - \hat{\beta}_1)$.

Proof. See the appendix. □

The marginal analysis in Theorem 4.5 provides no inferences for Γ and Σ, but the posterior mean and covariance matrix of Γ can be obtained from the marginal Gibbs sampler as follows. Note first that

$$\text{E}(\Gamma|\mathcal{D},r) = \text{E}_{\alpha,\beta|\mathcal{D},r}[\text{E}(\Gamma|\alpha,\beta,\mathcal{D},r)] \tag{4.3}$$

$$\text{C}(\Gamma|\mathcal{D},r) = \text{E}_{\alpha,\beta|\mathcal{D},r}[\text{C}(\Gamma|\alpha,\beta,\mathcal{D},r)] + \text{C}_{\alpha,\beta|\mathcal{D},r}[\text{E}(\Gamma|\alpha,\beta,\mathcal{D},r)], \tag{4.4}$$

where $\text{E}_{\alpha,\beta|\mathcal{D},r}$ and $\text{C}_{\alpha,\beta|\mathcal{D},r}$ denotes the expectation and covariance with respect to the posterior distribution $p(\alpha, B|\mathcal{D}, r)$. By integrating the full posterior distribution with respect to Σ, it is easily shown that

$$\Gamma|\alpha,\beta,\mathcal{D},r \sim t_{d\times p}[\mu_\Gamma, (Z'Z)^{-1}, W'M_ZW + A + v\alpha\beta'\beta\alpha', T + q + r - (d+p) + 1],$$

with μ_Γ as defined in Theorem 4.4. Thus, using results in Box and Tiao (1973, p. 446-7), we have

$$\text{E}(\Gamma|\alpha,\beta,\mathcal{D},r) = \mu_\Gamma \tag{4.5}$$

$$\text{C}(\Gamma|\alpha,\beta,\mathcal{D},r) = \frac{(W'M_ZW + A + v\alpha\beta'\beta\alpha') \otimes (Z'Z)^{-1}}{T + q + r - (d + p + 1)}. \tag{4.6}$$

The expectation and covariance with respect to $p(\alpha, B|\mathcal{D}, r)$ in (4.3) and (4.4) are easily computed by averaging the conditional expectation and covariance in (4.5) and (4.6) over the Gibbs samples obtained from the marginal Gibbs sampler in Theorem 4.5. Note that the possibly high-dimensional inverse, $(Z'Z)^{-1}$, is fixed in each step of the marginal Gibbs sampler.

It is also straight-forward to show that $\Sigma|\alpha,\beta,\mathcal{D},r \sim IW(W'M_ZW+A+v\alpha\beta'\beta\alpha', T+q+r-d)$. The marginal moments of Σ can be obtained in the same way as for Γ, using the expressions for the mean and covariance matrix of the inverted Wishart distribution.

It should be noted, however, that the marginal posteriors of Γ and Σ under reasonably vague priors seem to be well approximated by the asymptotic distributions of the maximum likelihood estimators of Γ and Σ in Johansen (1995).

We conclude this section with a note on the linear normalization of β. The main advantages of this simple normalization are that the prior which assigns the same probability to every cointegration space is of rather simple form and that an easily implemented Gibbs sampler can be used to compute the posterior results. Note also that we are free to transform the posterior distribution of α and β as long as the space spanned by the columns of β and the matrix of long run multipliers $\Pi = \alpha\beta'$ remain unchanged, i.e. the class of allowable transformations are $(\alpha, \beta) \to (\alpha U', \beta U^{-1})$, for any invertible $r \times r$ matrix U. This transformation is conveniently performed directly on the sampled α's and β's after the final iteration of the Gibbs sampler. Thus, the restriction to a linear normalization is in this sense no restriction at all as the final results may be transformed to any desired normalization. The question of existence mentioned in Section 2 is of course still relevant.

5. The posterior distribution of the cointegration rank

The posterior distribution of the cointegration rank is

$$p(r|\mathcal{D}) = \frac{p(\mathcal{D}|r)p(r)}{\sum_{r=0}^{p} p(\mathcal{D}|r)p(r)}, \tag{5.1}$$

where $p(r)$ is the prior probability of r cointegrating relations and

$$p(\mathcal{D}|r) = \int\int\int\int p(\mathcal{D}|\alpha,\beta,\Gamma,\Sigma,r)p(\alpha,\beta,\Gamma,\Sigma|r)d\Sigma d\Gamma d\alpha d\beta \tag{5.2}$$

is the marginal likelihood of the data given rank$(\Pi) = r$.

At least one of the integrals in (5.2) must be computed numerically, perhaps by importance sampling (Geweke, 1989; Kloek and van Dijk, 1978). This approach is likely to be imprecise, at least when p and r are large, and will not be pursued here.

Chib (1995) proposed a simple way to compute marginal likelihoods from the output of the Gibbs sampler in a general setting. The ideal setting for Chib's approach is when the Gibbs sampler consist of only two updating steps, i.e. when the full set of parameters in a given model can be partitioned into two blocks with easily sampled full conditional posteriors. This is exactly our case here as the parameter dependent on the rank can be sampled by the two-block marginal Gibbs sampler in Theorem 4.5.

Assume that a Gibbs sample of size n has been generated from $p(\alpha, B|\mathcal{D}, r)$ using the two conditional posteriors $p(\alpha|B,\mathcal{D},r)$ and $p(B|\alpha,\mathcal{D},r)$ in Theorem 4.5. By a slight rearrangement of Bayes' theorem we obtain

$$p(\mathcal{D}|r) = \frac{p(\mathcal{D}|\alpha,B,r)p(\alpha,B|r)}{p(\alpha,B|\mathcal{D},r)} = \frac{p(\mathcal{D}|\alpha,B,r)p(\alpha,B|r)}{p(B|\alpha,\mathcal{D},r)p(\alpha|\mathcal{D},r)}. \tag{5.3}$$

The expression for $p(\mathcal{D}|r)$ in (5.3) clearly holds for any α and B. Let $(\alpha, B) = (\tilde{\alpha}, \tilde{B})$ be the point where $p(\mathcal{D}|r)$ is evaluated. As we will see below $(\tilde{\alpha}, \tilde{B})$ should be a point of high posterior density; natural candidates are the posterior mode or median (the posterior mean does not exist, at least not when $r = 1$, see Section 4). The next result gives the expression for the numerator of (5.3) up to a multiplicative constant which does not depend on r.

Theorem 5.1.

$$p(\mathcal{D}|\alpha,B,r)p(\alpha,B|r) \propto \frac{\Gamma_p(T+q+r-d)\Gamma_r(p)}{\Gamma_r(r)\pi^{(2pr-r^2)/2}v^{-pr/2}} \left|A + v\alpha\beta'\beta\alpha' + W'M_Z W\right|^{-(T+q+r-d)/2},$$

where $W = Y - X\beta\alpha'$.

Proof. See the appendix. □

We still need to calculate $p(B|\alpha,\mathcal{D},r)$ and $p(\alpha|\mathcal{D},r)$ in (5.3). $p(B|\alpha,\mathcal{D},r)$ is given in the second part of Theorem 4.5. The final term $p(\alpha|\mathcal{D},r)$ is not available in closed form, but its value in the point $\alpha = \tilde{\alpha}$, which is all we need, can be computed from the Gibbs sample by

$$\hat{p}(\tilde{\alpha}\|\mathcal{D},r) = \frac{1}{n}\sum_{i=1}^{n} p(\tilde{\alpha}|B^{(i)},\mathcal{D},r),$$

where $B^{(i)}$ is the generated B in the ith Gibbs cycle and $p(\alpha|B\,\mathcal{D},r)$ is given in the first part of Theorem 4.5.

In order to calculate the marginal likelihoods for $r = 0$ and $r = p$, we must choose a prior for the relevant parameter matrices. It is possible to use the prior in (3.1) even in these extreme cases. This distribution agrees with our earlier prior in the reduced rank case and has the added benefit of giving tractable integrals. In addition, no new prior hyperparameters needs to be specified by the investigator. If $r = 0$, then $\alpha = \beta = 0$ and the prior in (3.1) becomes

$$p(\Gamma, \Sigma | r) = c_0 \, |\Sigma|^{-(p+q+1)/2} \, \mathrm{etr}(\Sigma^{-1} A), \tag{5.4}$$

which is a $IW(A, q)$ prior on Σ and $p(\Gamma) \propto c$. For $r = p$, $\Pi = \alpha\beta'$ is of full rank and

$$p(\Pi, \Gamma, \Sigma | r) = c_p \, |\Sigma|^{-(2p+q+1)/2} \, \mathrm{etr}[\Sigma^{-1}(A + v\Pi\Pi')], \tag{5.5}$$

which implies $\Sigma \sim IW(A, q)$, $\mathrm{vec}\,\Pi | \Sigma \sim \mathrm{N}_{p^2}(0, I_p \otimes v^{-1}\Sigma)$. The Kronecker structure on the prior covariance matrix of Π may be too restrictive for some applications, but it is possible to use a general normal-Wishart prior on Π. Some of the integrals in $p(\mathcal{D}|r = p)$ then become intractable, but can be computed from the Gibbs output via Chib's device, see e.g. Kadiyala and Karlsson (1997) for a simple Gibbs sampler.

The marginal likelihoods for $r = 0$ and $r = p$ are given in the next theorem.

Theorem 5.2. *For the priors in (5.4) and (5.5)*

$$\begin{aligned} p(\mathcal{D}|r = 0) \quad &\propto \quad \Gamma_p(T + q - d) \left| A + Y' M_Z Y \right|^{-(T+q-d)/2} \\ p(\mathcal{D}|r = p) \quad &\propto \quad \Gamma_p(T + q - d) v^{p^2/2} \, |S|^{-(T+q-d)/2} \, |C_1|^{-p/2}, \end{aligned}$$

where S is defined in Theorem 4.5 and C_1 is given in Theorem 4.3.

Proof. See the appendix. □

In summary, a complete analysis consist of the following steps:

- Generate a sample from each of the $p - 1$ marginal posteriors $p(\alpha, B|\mathcal{D}, r)$, $r = 1, 2, ..., p - 1$, using the marginal Gibbs sampler in Theorem 4.5 and compute $\tilde{\alpha}$ and $\tilde{B}$ in each sample.
- Compute $p(\mathcal{D}|r = 0)$ and $p(\mathcal{D}|r = p)$ using the formulas in Theorem 5.2 and $p(\mathcal{D}|r)$ for $r = 1, 2, ..., p - 1$ from the previously generated Gibbs samples as described above. Use (5.1) to compute the posterior distribution of the cointegration rank.
- Select the cointegration ranks of interest and summaries the posterior distribution of α and B (possibly transformed to a more interpretable normalization) conditional on these ranks using the already available Gibbs samples.

If inferences about Σ and Γ are wanted for the selected cointegration ranks, and the ML estimates with standard errors cannot be used for approximate inference, the full Gibbs sampler in Theorem 4.4 should be used. If the first two posterior moments of Σ and Γ are sufficient, the approach based on the marginal Gibbs sampler described in Section 4 is preferred.

Note that $p - 1$ is typically a rather small number (ranging from 1 to 4 in many applications) and that the marginal Gibbs sampler in Theorem 4.5 operates on rather small-dimensional spaces, at least compared to the dimension of the whole parameter space.

6. An numerical illustration

A single data set of length $T = 100$ was simulated from a bivariate model, without short-run dynamics and constant term, with parameters $\alpha = (0, 0.1)$, $\beta = (1, -1)$ and $\Sigma = I_2$. Note that α is close to the zero vector and the model is thus close to the zero rank model. This difficult setup has been chosen in order to accentuate some features of the posterior distribution in cointegration models which were initially raised by Kleibergen and van Dijk (1994). The simulated time series are displayed in panel a) of Figure 2.

Table II shows the posterior distribution of the cointegration rank for three different values of $\sigma = v^{-1/2}$, together with the likelihood trace and maximal eigenvalue tests (Johansen, 1995). A uniform distribution on the ranks was used *a priori.* q was set to 4 and the ML estimate

$$\hat{\Sigma} = \begin{pmatrix} 0.83 & \text{-0.10} \\ \text{-0.10} & 1.02, \end{pmatrix}$$

was used for A; other choices of A with larger positive and negative off-diagonal elements had only minor effects on the results. Note that as $\hat{\Sigma} \approx I_2$, σ corresponds roughly to the standard deviation of the adjustment coefficients for an orthonormal β, conditional on Σ, see Theorem 3.8.

The results in Table II are based on 25.000 iterations of the Gibbs sampler (after 1.000 burn-in iterations); convergence was reached at a much smaller number of iterations, however. From Table II it is seen that the trace test favors the full rank model whereas the maximal eigenvalue test prefers $r = 0$. For the two larger σ, $r = 0$ has the largest posterior probability. With $\sigma = 0.25$, the unit rank model is most probable *a posteriori.* The inconclusive evidence regarding the cointegration rank is of course expected as we purposely simulated data from a very difficult parametric setup. Panel b) of Figure 2 displays an estimate of the cointegrating relation which does not appear stationary.

It should be noted that the large uncertainty in the rank inference can, and should, be incorporated in the subsequent analysis by averaging these final inferences over the posterior distribution of r.

Table II

Inferences on the cointegration rank

Hypothesis	LR_{trace}	LR_{max}	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$
$r = 0$	17.41∗	12.43	0.263	0.622	0.874
$r = 1$	4.98∗	4.98*	0.504	0.337	0.122
$r = 2$	−	−	0.233	0.041	0.004

Note: ** and * denotes significant at the 1 and 5 percent level, respectively.

To discuss the issue of local non-identification the simulated data set is analyzed conditional on $r = 1$. Table III and the first three panels of Figure 3 display the inferences for α_1, α_2 and B. The striking point in Table III is how insensitive the inferences are to changes in σ; the choice of loss function (which determines the measure of posterior centrality) is more important.

INSERT FIGURE HERE.

FIGURE 2. a) The two simulated time series. b) Estimated cointegration relation, $\beta' x_t$, with $\beta = (1, -1.15)'$.

INSERT FIGURE HERE

FIGURE 3. The posterior distribution of α and β for $\sigma = 0.5$ conditional on $r = 1$. $\theta = \arctan(B)$ is the angle of the cointegration vector in the orthonormal normalization. 2% of the draws from each tails of the posterior distribution of B were excluded in the histogram construction.

INSERT FIGURE HERE

FIGURE 4. Sample points from the joint posterior distribution of α and β for $\sigma = 0.5$ conditional on $r = 1$.

TABLE III

POSTERIOR DISTRIBUTION OF α AND β CONDITIONAL ON $r = 1$

		Posterior mode			Posterior median		
	ML	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$
α_1	0.016	0.015	0.012	0.015	0.010	0.010	0.011
α_2	0.096	0.079	0.089	0.091	0.080	0.083	0.083
B	−1.171	−1.063	−1.061	−1.031	−1.151	−1.137	−1.150

Note: All estimates are based on 25.000 Gibbs iterations (excluding burn-in).

The extra local mode in the marginal posterior of $\alpha_2 = 0$ in Figure 3 is an effect of the local non-identification discussed in Kleibergen and van Dijk (1994). They pointed out that when $\alpha = (0, 0)'$, β drops out of the likelihood function and the likelihood is then constant along the B-axis (which has infinite length) and all values for B are observationally equivalent; B is said to be *locally non-identified* when $\alpha = (0, 0)'$. The posterior distribution based on the prior in (3.1) has the same property as it is flat in the direction of B when α is the zero vector. This is illustrated in Figure 4, where the sample points from $p(\alpha_1, B|\mathcal{D}, r = 1)$ and $p(\alpha_2, B|\mathcal{D}, r = 1)$ are displayed. Note how the conditional variance of B grows as $\alpha \to 0$. The posterior variance of B given $\alpha = 0$ is actually infinite as can be seen from the second part of Theorem 4.5. This of course as it should be: if the processes do not react at all to past deviations from the equilibrium, then the data are necessarily uninformative regarding the cointegration vector.

Kleibergen and van Dijk (1994) argue that this local non-identification causes problems for a Bayesian analysis with uniform, improper, priors on α and B. Their argument is as follows: the marginal posterior of α is obtained by integrating the posterior $p(\alpha, B|\mathcal{D})$ with respect to B. As the posterior under a uniform prior is flat along the B-axis when $\alpha = (0, 0)'$, the marginal posterior density of α in the point $\alpha = (0, 0)'$ is proportional to the integral of a constant over an unbounded region ($-\infty < B < \infty$), i.e. infinity. The marginal posterior of α is thus expected to have an asymptote in the point $(0, 0)'$ which is entirely created by the local non-identification. This conclusion is given without formal proof, but a numerical example is used to support the argument.

Kleibergen and van Dijk suggests the Jeffreys prior to counter-attack the unwanted asymptote as this prior is zero in the locally non-identified points. The prior in Kleibergen and Paap (1998) has the same property. In the case of unit rank their prior reduces to

$$p(\alpha, \beta) \propto (\|\alpha\| \, \|\beta\|)^{(p-1)/2},$$

where $\|\cdot\|$ is the usual Euclidean length.

Our view on the local non-identification problem is best illustrated by transforming the posterior results so that β is restricted to a half-circle with unit radius, i.e.

$$\tilde{\beta} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \qquad -\frac{\pi}{2} \leq \theta < \frac{\pi}{2}, \tag{6.1}$$

where θ is the angle of the cointegration vector. This change in normalization is accomplished by the transformation $\theta = \arctan B$ and $\tilde{\alpha} = \alpha(1+B^2)^{1/2}$; note that the product $\alpha\beta'$ is unchanged. The last three panels of Figure 3 display the marginal posteriors in the new normalization. Note that the local mode in the posterior of $\tilde{\alpha}_2$ has disappeared, or at least is very small, after the transformation.

To explain this effect, note that B is a ratio of the two elements of β and that the tails in the marginal posterior of B are therefore heavy. Heavy tails in $p(B|\mathcal{D}, r)$ correspond to very small values for α, in the sense that a large β must be matched by a small α to keep the product $\Pi = \alpha\beta'$ at a reasonable magnitude. When we transform to the more natural orthonormal normalization we are multiplying α with $(1 + B^2)^{1/2}$, which is large if B is drawn far out in the tails of $p(B|\mathcal{D}, r)$ and has the effect of spreading out the extra mode at $\alpha = (0, 0)'$ and thereby producing a more well-behaved surface.

It should be noted that the argument of Kleibergen and van Dijk (1994) breaks down if the space of the free parameters in β is bounded. This exactly the case if the normalization of β in (6.1) is used as θ is bounded. More generally, the Grassman manifold is bounded.

Theorems 3.6 and 3.8 show that the prior on $\tilde{\beta}$, the orthonormal matrix of cointegration vectors, is uniformly distributed over the Grassman manifold independently of $\tilde{\alpha}$. This means that the prior on $\tilde{\beta}$ conditional on $\tilde{\alpha} = 0$ is still uniform over the Grassman manifold. Thus, given the information that $\tilde{\alpha} = 0$, the prior in (3.1) represents the belief that every possible cointegration space of dimension r has the same probability *a priori*. This seems sensible.

7. Concluding remarks

This paper has introduced a practicable Bayesian analysis of cointegration based on prior which is convenient both in elicitation and computation and could serve as a standard for inference reporting. The posterior distribution of both the cointegration rank and the model parameters conditional on the rank are obtained from the same Gibbs sampler.

Interesting topics for future research within the proposed framework include the testing of restrictions on the model parameters, extensions to more general structures on the model disturbances and regime switching parameters.

References

Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially non-stationary multivariate autoregressive processes, *Journal of the American Statistical Association*, **85**, 813-23.

Bauwens, L. and Giot, P. (1998). A Gibbs sampler approach to cointegration, *Computational Statistics*, **13**, 339-68.

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

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

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

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

Doan, T., Litterman, R. B. and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions, *Econometrics Reviews*, **3**, 1-100.

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

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

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

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. New York: Springer-Verlag.

James, A. T. (1954). Normal multivariate analysis and the orthogonal group, *Ann. Math. Statist.,* **25**, 40-74.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551-80.

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

Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models, *Journal of Applied Econometrics*, **12**, 99-132.

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

Kleibergen, F. and Paap, R. (1998). Priors, posterior odds and Lagrange multiplier statistics in Bayesian analysis of cointegration. *Econometric Institute Report*, no. **9821/A**, Econometric Institute, Erasmus University Rotterdam.

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

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - Five years of experience, *Journal of Business and Economic Statistics*. **4**, 25-38.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems, *Econometrica*, **59**, 283-306.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends, *Journal of the American Statistical Association*, **83**, 1097-107.

Strachan, R. W. (2000). Bayesian analysis of the cointegrating error correction model: with extension to general reduced rank regression models. PhD thesis, Monash University, Australia.

Villani, M. (2000). Aspects of Bayesian Cointegration. PhD thesis, Stockholm University, Sweden.

Villani, M. (2001a). Fractional Bayesian lag length inference in multivariate autoregressive processes, *Journal of Time Series Analysis*, **22**, 67-86.

Villani, M. (2001b). Bayesian Prediction with Cointegrated Vector Autoregressions, *International Journal of Forecasting*, forthcoming.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Appendix A. Proofs

A.1. **Proof of Theorem 3.6.** To obtain the marginal distribution of β, we first derive the marginal distribution of B. The joint prior of B and Σ is

$$\begin{aligned} p(B,\Sigma) &= \int p(\alpha,B,\Sigma)d\alpha = c_r\,|\Sigma|^{-(p+r+q+1)/2}\int \operatorname{etr}[\Sigma^{-1}(A+v\alpha\beta'\beta\alpha')]d\alpha \\ &= c_r\,|\Sigma|^{-(p+r+q+1)/2}\operatorname{etr}(\Sigma^{-1}A)\int \operatorname{etr}(\Sigma^{-1}v\alpha\beta'\beta\alpha')d\alpha. \end{aligned}$$

By Theorem 16.2.2 in Harville

$$\operatorname{tr}(\Sigma^{-1}v\alpha\beta'\beta\alpha') = \operatorname{vec}(\alpha)'(\beta'\beta\otimes v\Sigma^{-1})\operatorname{vec}(\alpha).$$

Thus, using properties of the normal distribution,

$$\begin{aligned} p(B,\Sigma) &= c_r\,|\Sigma|^{-(p+r+q+1)/2}\operatorname{etr}(\Sigma^{-1}A)\int \exp\left\{-\frac{1}{2}\operatorname{vec}(\alpha)'[\beta'\beta\otimes v\Sigma^{-1}]\operatorname{vec}(\alpha)\right\}d\operatorname{vec}\alpha \\ &= c_r\,|\Sigma|^{-(p+r+q+1)/2}\operatorname{etr}(\Sigma^{-1}A)(2\pi)^{pr/2}\left|\beta'\beta\otimes v\Sigma^{-1}\right|^{-1/2} \\ &= c_r(2\pi/v)^{pr/2}\,|\Sigma|^{-(p+q+1)/2}\operatorname{etr}(\Sigma^{-1}A)\left|I_r+B'B\right|^{-p/2}. \end{aligned}$$

This shows that B and Σ are independent and marginally $B\sim t_{(p-r)\times r}(0,I_{p-r},I_r,1)$. Thus, using Lemma 3.5, β is uniformly distributed over $\mathcal{G}_{p,r}$.

A.2. **Proof of Theorem 3.8.** First we derive the distribution of α conditional on β and Σ

$$\begin{aligned} p(\alpha|\beta,\Sigma) &= \frac{p(\alpha,\beta,\Sigma)}{p(\beta,\Sigma)} = \frac{c_r\,|\Sigma|^{-(p+r+q+1)/2}\operatorname{etr}[\Sigma^{-1}(A+v\alpha\beta'\beta\alpha')]}{c_r(2\pi/v)^{pr/2}\,|\Sigma|^{-(p+q+1)/2}\operatorname{etr}(\Sigma^{-1}A)\,|I_r+B'B|^{-p/2}} \\ &= (2\pi)^{-pr/2}\left|(\beta'\beta)^{-1}\otimes v^{-1}\Sigma\right|^{-1/2}\exp\left\{-\frac{1}{2}(\operatorname{vec}\alpha)'(\beta'\beta\otimes v\Sigma^{-1})(\operatorname{vec}\alpha)\right\}. \end{aligned}$$

Thus,

$$\alpha|\beta,\Sigma\sim N_{p\times r}[0,(\beta'\beta)^{-1},v^{-1}\Sigma].$$

As $\tilde{\alpha}=\alpha(\beta'\beta)^{1/2}$ we have (see e.g. Bauwens et al 1999, p. 302)

$$\tilde{\alpha}|\tilde{\beta},\Sigma\sim N_{p\times r}(0,I_r,v^{-1}\Sigma).$$

The density $p(\tilde{\alpha}|\tilde{\beta},\Sigma)$ is not a function of $\tilde{\beta}$ and we may write $\tilde{\alpha}|\Sigma\sim N_{p\times r}(0,I_r,v^{-1}\Sigma)$. The statement of the theorem now follows from the usual independence property of the multivariate normal distribution.

A.3. **Proof of Lemma 3.10.** Let $\mathcal{N}_1$ denote that β is normalized on the r first variables and $\mathcal{N}_2$ that β is normalized on variables $1,2,...,r-1$ and $r+1$, i.e. the change in normalizing variables from $\mathcal{N}_1$ to $\mathcal{N}_2$ is accomplished by replacing the last variable of the normalizing set with the first variable in the non-normalizing set. It will be evident that the lemma holds generally under any change in normalizing variables. Let

$$B=\begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,r} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p-r,1} & b_{p-r,2} & \cdots & b_{p-r,r} \end{pmatrix},$$

denote the matrix of free coefficients in β under $\mathcal{N}_1$. The transformation matrix in this case is

$$U=\begin{pmatrix} J \\ b_{1,1},b_{1,2},...,b_{1,r} \end{pmatrix},\ \text{if } r>1 \text{ and } U=b_{1,1} \text{ if } r=1,$$

where J denotes the $r-1$ first rows of I_r. To see that U actually produces the intended change in normalization note that

$$\bar{\beta} = \begin{pmatrix} I_r \\ B \end{pmatrix} U^{-1} = \begin{pmatrix} & J & & \\ -b_{1,1}b_{1,r}^{-1} & -b_{1,2}b_{1,r}^{-1} & \cdots & b_{1,r}^{-1} \\ 0 & 0 & \cdots & 1 \\ b_{2,1} - b_{2,r}b_{1,1}b_{1,r}^{-1} & b_{2,2} - b_{2,r}b_{1,2}b_{1,r}^{-1} & \cdots & b_{2,r}b_{1,r}^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p-r,1} - b_{p-r,r}b_{1,1}b_{1,r}^{-1} & b_{p-r,2} - b_{p-r,2}b_{1,2}b_{1,r}^{-1} & \cdots & b_{p-r,r}b_{1,r}^{-1} \end{pmatrix}$$

as can be verified by direct calculation. It is easy to see that $|U| = b_{1,r}$ and

$$U^{-1} = \begin{pmatrix} J \\ -b_{1,1}b_{1,r}^{-1}, -b_{1,2}b_{1,r}^{-1}, ..., b_{1,r}^{-1} \end{pmatrix}, \text{ if } r > 1 \text{ and } U^{-1} = b_{1,1}^{-1} \text{ if } r = 1,$$

The matrix of free coefficients under $\mathcal{N}_2$ is

$$\bar{B} = \begin{pmatrix} -b_{1,1}b_{1,r}^{-1} & -b_{1,2}b_{1,r}^{-1} & \cdots & b_{1,r}^{-1} \\ b_{2,1} - b_{2,r}b_{1,1}b_{1,r}^{-1} & b_{2,2} - b_{2,r}b_{1,2}b_{1,r}^{-1} & \cdots & b_{2,r}b_{1,r}^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p-r,1} - b_{p-r,r}b_{1,1}b_{1,r}^{-1} & b_{p-r,2} - b_{p-r,r}b_{1,2}b_{1,r}^{-1} & \cdots & b_{p-r,r}b_{1,r}^{-1} \end{pmatrix}. \tag{A.1}$$

The change in normalization from $\mathcal{N}_2$ to $\mathcal{N}_1$ is thus given by the transformation $\bar{\alpha}, \bar{B}, \Sigma \to \alpha, B, \Sigma$, where $\bar{\alpha} = \alpha U'$. The Jacobian of this transformation is

$$J(\bar{\alpha}, \bar{\beta}, \Sigma \to \alpha, \beta, \Sigma) = \begin{vmatrix} \frac{d\,\mathrm{vec}(\bar{\alpha})}{d\,\mathrm{vec}(\alpha)'} & \frac{d\,\mathrm{vec}(\bar{\alpha})}{d\,\mathrm{vec}(B)'} \\ \frac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(\alpha)'} & \frac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(B)'} \end{vmatrix} = |U|^p \left| \frac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(B)'} \right|, \tag{A.2}$$

as Σ is unaffected by the transformation, $\frac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(\alpha)'} = 0$ and $\frac{d\,\mathrm{vec}(\bar{\alpha})}{d\,\mathrm{vec}(\alpha)'} = U \otimes I_p$. From the definition of the vec-operator,

$$\left| \frac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(B)'} \right| = \begin{vmatrix} \frac{d\bar{b}_1}{db_1} & \frac{d\bar{b}_1}{db_2} & \cdots & \frac{d\bar{b}_1}{db_r} \\ \frac{d\bar{b}_2}{db_1} & \frac{d\bar{b}_2}{db_2} & \cdots & \frac{d\bar{b}_2}{db_r} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d\bar{b}_r}{db_1} & \frac{d\bar{b}_r}{db_2} & \cdots & \frac{d\bar{b}_r}{db_r} \end{vmatrix},$$

where b_i and $\bar{b}_i$ are the ith column of B and $\bar{B}$, respectively. It is easily seen from (A.1) that $\frac{d\bar{b}_i}{db_j} = 0$ for $i > j$, and thus

$$\left| \frac{d\,\mathrm{vec}(\bar{B})}{d\,\mathrm{vec}(B)'} \right| = \left| \frac{d\bar{b}_1}{db_1} \right| \left| \frac{d\bar{b}_2}{db_2} \right| \cdots \left| \frac{d\bar{b}_r}{db_r} \right|, \tag{A.3}$$

where

$$\frac{d\bar{b}_i}{db_i} = \begin{pmatrix} -b_{1,r}^{-1} & 0 \\ \cdot & I_{p-r-1} \end{pmatrix}, \text{ for } i = 1, ..., r-1, \text{ and } \frac{d\bar{b}_r}{db_r} = \frac{d\bar{b}_1}{db_1} b_{1,r}^{-1}, \tag{A.4}$$

and the dot replaces an expression which is unnecessary to calculate. Thus, from (A.2), (A.3) and (A.4)

$$J(\bar{\alpha}, \bar{\beta}, \Sigma \to \alpha, \beta, \Sigma) = |U|^p \left| \frac{d\bar{b}_1}{db_1} \right| \left| \frac{d\bar{b}_2}{db_2} \right| \cdots \left| \frac{d\bar{b}_r}{db_r} \right| = b_{1,r}^p (b_{1,r}^{-1})^r (b_{1,r}^{-1})^{p-r} = 1.$$

A.4. **Proof of Lemma 4.1.** The joint posterior of $\alpha, \beta, \Gamma, \Sigma$ is

$$\begin{aligned} p(\alpha, \beta, \Gamma, \Sigma | \mathcal{D}, r) &\propto p(\mathcal{D} | \alpha, \beta, \Gamma, \Sigma, r) p(\alpha, \beta, \Gamma, \Sigma | r) \\ &\propto |\Sigma|^{-(T+p+r+q+1)/2} \operatorname{etr}[\Sigma^{-1}(E'E + A + v\alpha\beta'\beta\alpha')], \end{aligned}$$

where $E = Y - X\beta\alpha' - Z\Gamma$. Using properties of the inverted Wishart density (Zellner, 1971),

$$p(\alpha, \beta, \Gamma | \mathcal{D}, r) \propto \left| E'E + A + v\alpha\beta'\beta\alpha' \right|^{-(T+q+r)/2}.$$

Let $W = Y - X\beta\alpha'$. We have

$$E'E = (W - Z\Gamma)'(W - Z\Gamma) = W'M_Z W + (\Gamma - \hat{\Gamma})'Z'Z(\Gamma - \hat{\Gamma}),$$

where $\hat{\Gamma} = (Z'Z)^{-1}Z'W$. Thus, by Lemma 4.2

$$\begin{aligned} p(\alpha, \beta|\mathcal{D}, r) &\propto \int \left|W'M_Z W + A + v\alpha\beta'\beta\alpha' + (\Gamma - \hat{\Gamma})'Z'Z(\Gamma - \hat{\Gamma})\right|^{-(T+q+r)/2} d\Gamma \\ &\propto \left|(Y - X\beta\alpha')'M_Z(Y - X\beta\alpha') + A + v\alpha\beta'\beta\alpha'\right|^{-(T+q+r-d)/2}. \end{aligned}$$

A.5. **Proof of Theorem 4.3.** The following proof follows closely the proof of Theorem 3.1 in Bauwens and Lubrano (1996). From the proof of Lemma 4.1

$$p(\alpha, \beta, \Gamma|\mathcal{D}) \propto \left|E'E + A + \Delta' F \Delta\right|^{-(T+q+r)/2} \tag{A.5}$$

where

$$\underset{(r+d)\times p}{\Delta} = \begin{pmatrix} \alpha' \\ \Gamma \end{pmatrix}, F = \begin{pmatrix} v\beta'\beta & 0 \\ 0 & 0 \end{pmatrix}$$

$E = Y - W\Delta$ and $W = (X\beta, Z)$. The equality

$$E'E + A + \Delta' F \Delta = C + (\Delta - \tilde{\Delta})'(W'W + F)(\Delta - \tilde{\Delta}), \tag{A.6}$$

where $\tilde{\Delta} = (W'W + F)^{-1}W'Y$ and $C = A + Y'Y - Y'W(W'W + F)^{-1}W'Y$, can be verified by direct calculation. Inserting (A.6) in (A.5) and integrating out Δ using Lemma 4.2 gives

$$\begin{aligned} p(\beta|\mathcal{D}) &\propto \int \left|C + (\Delta - \tilde{\Delta})'(W'W + F)(\Delta - \tilde{\Delta})\right|^{-(T+q+r)/2} d\Delta \\ &\propto |C|^{-(T+q-d)/2} \left|W'W + F\right|^{-p/2}. \end{aligned}$$

Using the result

$$\left|D_4 - D_3 D_1^{-1} D_2\right| = |D_1|^{-1} |D_4| \left|D_1 - D_2 D_4^{-1} D_3\right|,$$

which holds for any conformable matrices D_1, D_2, D_3 and D_4, we can write

$$|C| = \left|A + Y'Y - Y'W(W'W + F)^{-1}W'Y\right| = \left|W'W + F\right|^{-1} \left|A + Y'Y\right| \left|F + W'QW\right|,$$

where $Q = I_T - Y(A + Y'Y)^{-1}Y'$. Thus,

$$p(\beta|\mathcal{D}) \propto \left|F + W'QW\right|^{-(T+q-d)/2} \left|W'W + F\right|^{(T+q-d-p)/2}, \tag{A.7}$$

where

$$\begin{aligned} \left|W'W + F\right| &= \begin{vmatrix} \beta'(X'X + vI_p)\beta & \beta'X'Z \\ Z'X\beta & Z'Z \end{vmatrix} \\ &= \left|Z'Z\right| \left|\beta' \left\{vI_p + X'[I_T - Z(Z'Z)^{-1}Z']X\right\} \beta\right|, \end{aligned} \tag{A.8}$$

and

$$\begin{aligned} \left|F + W'QW\right| &= \begin{vmatrix} \beta'(X'QX + vI_p)\beta & \beta'X'QZ \\ Z'QX\beta & Z'QZ \end{vmatrix} \\ &= \left|Z'QZ\right| \left|\beta' \left\{vI_p + X'Q[I_T - Z(Z'QZ)^{-1}Z'Q]X\right\} \beta\right|. \end{aligned} \tag{A.9}$$

Inserting (A.8) and (A.9) in (A.7) proves the result.

A.6. **Proof of Theorem 4.4.** All full conditional posteriors are proportional to the likelihood function multiplied with the prior in (3.1), i.e. proportional to

$$|\Sigma|^{-(T+p+r+q+1)/2} \operatorname{etr}[\Sigma^{-1}(E'E + A + v\alpha\beta'\beta\alpha')], \tag{A.10}$$

where $E = Y - X\beta\alpha' - Z\Gamma$.

It follows directly from (A.10) that the full conditional posterior of Σ is the $IW_p(E'E + A + v\alpha\beta'\beta\alpha', T + q + r)$ density.

The full conditional posterior of Γ follows from the treatment of the multivariate regression in Zellner (1971); see also Geweke (1996).

To obtain the full conditional posterior of B, let $X = (X_1, X_2)$, where X_1 contains the r first columns of X and X_2 contains the $p-r$ remaining ones, and $W = Y - X_1\alpha' - Z\Gamma$. The full conditional likelihood of B is then

$$\begin{aligned}p(\mathcal{D}|\alpha,\beta,\Gamma,\Sigma,r) &\propto \text{etr}[\Sigma^{-1}(W-X_2B\alpha')'(W-X_2B\alpha')]\\ &= \text{etr}[(W\Sigma^{-1/2}-X_2B\alpha'\Sigma^{-1/2})'(W\Sigma^{-1/2}-X_2B\alpha'\Sigma^{-1/2})]\\ &= \exp\left\{-\frac{1}{2}\left[\text{vec}(W\Sigma^{-1/2})-H\,\text{vec}\,B\right]'\left[\text{vec}(W\Sigma^{-1/2})-H\,\text{vec}\,B\right]\right\},\end{aligned}$$

where $H=(\Sigma^{-1/2}\alpha\otimes X_2)$. Thus,

$$p(\mathcal{D}|\alpha,\beta,\Gamma,\Sigma,r)\propto\exp\left\{-\frac{1}{2}(\text{vec}\,B-\text{vec}\,\hat{B})'(\alpha'\Sigma^{-1}\alpha\otimes X_2'X_2)(\text{vec}\,B-\text{vec}\,\hat{B})\right\}, \tag{A.11}$$

where

$$\begin{aligned}\text{vec}\,\hat{B} &= \left[(\alpha'\Sigma^{-1}\alpha)^{-1}\otimes(X_2'X_2)^{-1}\right](\alpha'\Sigma^{-1/2}\otimes X_2')\,\text{vec}(W\Sigma^{-1/2})\\ &= \text{vec}\left[(X_2'X_2)^{-1}X_2'W\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\right].\end{aligned}$$

The prior in (3.1) can be rewritten as

$$\begin{aligned}p(\alpha,B,\Sigma,\Gamma|r) &\propto \text{etr}(\Sigma^{-1}v\alpha\alpha')\,\text{etr}(\Sigma^{-1}v\alpha B'B\alpha')\\ &\propto \exp\left\{-\frac{1}{2}(\text{vec}\,B)'(\alpha'\Sigma^{-1}\alpha\otimes vI_{p-r})(\text{vec}\,B)\right\}. \end{aligned}\tag{A.12}$$

By multiplying (A.11) by $p(\alpha,B,\Sigma,\Gamma|r)$ in (A.12) and completing the square in the exponential (see Lemma 1 in Box and Tiao, 1973, p. 418), it is seen that

$$p(B|\alpha,\Gamma,\Sigma,\mathcal{D},r)\propto\exp\left\{-\frac{1}{2}(\text{vec}\,B-\text{vec}\,\mu_B)'\Omega_B^{-1}(\text{vec}\,B-\text{vec}\,\mu_B)\right\},$$

where $\Omega_B^{-1}=\alpha'\Sigma^{-1}\alpha\otimes(X_2'X_2+vI_{p-r})$ and

$$\text{vec}\,\mu_B=\Omega_B(\alpha'\Sigma^{-1}\alpha\otimes X_2'X_2)\,\text{vec}\,\hat{B}=\text{vec}[(X_2'X_2+vI_{p-r})^{-1}X_2'W\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}].$$

Thus, $B|\alpha,\Gamma,\Sigma,D\sim N_{(p-r)\times r}[\mu_B,(\alpha'\Sigma^{-1}\alpha)^{-1},(X_2'X_2+vI_{p-r})^{-1}]$.

To derive the full conditional posterior of α, let $Q=Y-Z\Gamma$. The full conditional likelihood of α is

$$\begin{aligned}p(\mathcal{D}|\alpha,\beta,\Gamma,\Sigma,r) &\propto \text{etr}[\Sigma^{-1}(Q-X\beta\alpha')'(Q-X\beta\alpha')]\\ &= \text{etr}[(Q\Sigma^{-1/2}-X\beta\alpha'\Sigma^{-1/2})'(Q\Sigma^{-1/2}-X\beta\alpha'\Sigma^{-1/2})]\\ &\propto \exp\left\{-\frac{1}{2}\left[\text{vec}(\alpha')-\text{vec}(\hat{\alpha}')\right]'\left(\Sigma^{-1}\otimes\beta'X'X\beta\right)\left[\text{vec}(\alpha')-\text{vec}(\hat{\alpha}')\right]\right\}\\ &= \exp\left\{-\frac{1}{2}\left(\text{vec}\,\alpha-\text{vec}\,\hat{\alpha}\right)'\left(\beta'X'X\beta\otimes\Sigma^{-1}\right)\left(\text{vec}\,\alpha-\text{vec}\,\hat{\alpha}\right)\right\}\end{aligned}$$

where

$$\begin{aligned}\text{vec}(\hat{\alpha}') &= \left[\Sigma\otimes(\beta'X'X\beta)^{-1}\right](\Sigma^{-1/2}\otimes\beta'X')\,\text{vec}(Q\Sigma^{-1/2})\\ &= \text{vec}\left[(\beta'X'X\beta)^{-1}\beta'X'Q\right].\end{aligned}$$

Thus,

$$\hat{\alpha}=Q'X\beta(\beta'X'X\beta)^{-1}.$$

By multiplying the conditional likelihood of α with

$$p(\alpha,\beta,\Sigma,\Gamma|r)\propto\exp\left\{-\frac{1}{2}(\text{vec}\,\alpha)'(\beta'\beta\otimes v\Sigma^{-1})(\text{vec}\,\alpha)\right\}$$

we obtain (see Lemma 1 in Box and Tiao, 1973, p. 418)

$$p(\alpha|\beta,\Sigma,\Gamma,\mathcal{D},r)\propto\exp\left\{-\frac{1}{2}\left(\text{vec}\,\alpha-\text{vec}\,\mu_\alpha\right)'\Omega_\alpha^{-1}\left(\text{vec}\,\alpha-\text{vec}\,\mu_\alpha\right)\right\},$$

where $\Omega_\alpha^{-1}=\beta'(X'X+vI_p)\beta\otimes\Sigma^{-1}$ and

$$\text{vec}\,\mu_\alpha=\Omega_\alpha(\beta'X'X\beta\otimes\Sigma^{-1})\,\text{vec}\,\hat{\alpha}=\text{vec}\left(Q'X\beta[\beta'(X'X+vI_p)\beta]^{-1}\right).$$

A.7. **Proof of Theorem 4.5.** From Lemma 4.1, we have

$$\begin{aligned} p(\alpha|\beta,\mathcal{D},r) &\propto \left|(Y-X\beta\alpha')'M_Z(Y-X\beta\alpha')+A+v\alpha\beta'\beta\alpha'\right|^{-(T+q+r-d)/2} \\ &= \left|A+Y'M_Z(Y-X\beta\hat{\alpha}')+(\alpha'-\hat{\alpha}')'(\beta'C_1\beta)(\alpha'-\hat{\alpha}')\right|^{-(T+q+r-d)/2}. \end{aligned}$$

where $\hat{\alpha}'=(\beta'C_1\beta)^{-1}\beta'X'M_ZY$. Thus, $\alpha'\sim t_{r\times p}[\hat{\alpha}',(\beta'C_1\beta)^{-1},A+Y'M_Z(Y-X\beta\hat{\alpha}'),T+q-(d+p)+1]$. From Box and Tiao (1973, p. 442), $\alpha\sim t_{p\times r}[\hat{\alpha},A+Y'M_Z(Y-X\beta\hat{\alpha}'),(\beta'C_1\beta)^{-1},T+q-(d+p)+1]$.

Since $\Pi=\alpha\beta'$, the posterior of β conditional on α can be written

$$\begin{aligned} p(\beta|\alpha,\mathcal{D},r) &\propto \left|(Y-X\Pi')'M_Z(Y-X\Pi')+A+v\Pi\Pi'\right|^{-(T+q+r-d)/2} \\ &= \left|S+(\Pi-\hat{\Pi})C_1(\Pi-\hat{\Pi})'\right|^{-(T+q+r-d)/2}, \end{aligned}$$

where $S=A+Y'M_ZY-\hat{\Pi}C_1\hat{\Pi}'$ and $\hat{\Pi}=Y'M_ZXC_1^{-1}$. Thus,

$$\begin{aligned} p(\beta|\alpha,\mathcal{D},r) &\propto \left|C_1^{-1}+(\alpha\beta'-\hat{\Pi})'S^{-1}(\alpha\beta'-\hat{\Pi})\right|^{-(T+q+r-d)/2} \\ &= \left|R+(\beta-\hat{\beta})(\alpha'S^{-1}\alpha)(\beta-\hat{\beta})'\right|^{-(T+q+r-d)/2} \\ &\propto \left|(\alpha'S^{-1}\alpha)^{-1}+(\beta-\hat{\beta})'R^{-1}(\beta-\hat{\beta})\right|^{-(T+q+r-d)/2} \end{aligned} \tag{A.13}$$

where $\hat{\beta}=\hat{\Pi}'S^{-1}\alpha(\alpha'S^{-1}\alpha)^{-1}$ and $R=C_1^{-1}+\hat{\Pi}'S^{-1}\hat{\Pi}-\hat{\beta}(\alpha'S^{-1}\alpha)\hat{\beta}'$. Let $\hat{\beta}=(\hat{\beta}_1',\hat{\beta}_2')'$, where $\hat{\beta}_1$ contains the r first rows of $\hat{\beta}$ and $\hat{\beta}_2$ the $p-r$ remaining ones, and R is conformably decomposed as

$$R=\begin{pmatrix} \underset{r\times r}{G_1} & \underset{r\times(p-r)}{G_2} \\ \underset{(p-r)\times r}{G_2'} & \underset{(p-r)\times(p-r)}{G_3} \end{pmatrix}$$

By using the result (see e.g. Harville, 1997)

$$R^{-1}=\begin{pmatrix} (G_1-G_2G_3^{-1}G_2')^{-1} & -(G_1-G_2G_3^{-1}G_2')^{-1}G_2G_3^{-1} \\ -G_3^{-1}G_2'(G_1-G_2G_3^{-1}G_2')^{-1} & (G_3-G_2'G_1^{-1}G_2)^{-1} \end{pmatrix}$$

it is straight-forward to show that

$$(\beta-\hat{\beta})'R^{-1}(\beta-\hat{\beta})=(I_r-\hat{\beta}_1)'G_1^{-1}(I_r-\hat{\beta}_1)+(B-\hat{B})'(G_3-G_2'G_1^{-1}G_2)^{-1}(B-\hat{B}),$$

where $\hat{B}=\hat{\beta}_2+G_2'G_1(I_r-\hat{\beta}_1)$. From (A.13)

$$p(B|\alpha,\mathcal{D},r)\propto\left|C_3+(B-\hat{B})'(G_3-G_2'G_1^{-1}G_2)^{-1}(B-\hat{B})\right|^{-(T+q+r-d)/2},$$

where $C_3=(I_r-\hat{\beta}_1)'G_1^{-1}(I_r-\hat{\beta}_1)+(\alpha'S^{-1}\alpha)^{-1}$. This is proportional to the matrix t density in Theorem 4.5.

A.8. **Proof of Theorem 5.1.** Using properties of the inverted Wishart density,

$$p(\alpha,B|r)=\int p(\alpha,B,\Sigma|r)d\Sigma=c_r2^{p(q+r)/2}\pi^{p(p-1)/4}\Gamma_p(q+r)\left|A+v\alpha\beta'\beta\alpha'\right|^{-(q+r)/2}. \tag{A.14}$$

$$p(\mathcal{D}|\alpha,B,r)=\iint p(\mathcal{D}|\alpha,B,\Gamma,\Sigma,r)p(\Gamma,\Sigma|\alpha,B,r)d\Sigma d\Gamma, \tag{A.15}$$

where

$$p(\Gamma,\Sigma|\alpha,B,r)=\frac{p(\alpha,B,\Gamma,\Sigma|r)}{p(\alpha,B|r)}=\frac{|\Sigma|^{-(p+r+q+1)/2}\operatorname{etr}[\Sigma^{-1}(A+v\alpha\beta'\beta\alpha')]}{2^{p(q+r)/2}\pi^{p(p-1)/4}\Gamma_p(q+r)\left|A+v\alpha\beta'\beta\alpha'\right|^{-(q+r)/2}}.$$

Thus, from (A.15)

$$\begin{aligned} p(\mathcal{D}|\alpha,B,r) &= \iint(2\pi)^{-Tp/2}|\Sigma|^{-T/2}\operatorname{etr}(\Sigma^{-1}E'E)\frac{|\Sigma|^{-(p+r+q+1)/2}\operatorname{etr}[\Sigma^{-1}(A+v\alpha\beta'\beta\alpha')]}{2^{p(q+r)/2}\pi^{p(p-1)/4}\Gamma_p(q+r)\left|A+v\alpha\beta'\beta\alpha'\right|^{-(q+r)/2}}d\Sigma d\Gamma \\ &= k_1\int\left|E'E+A+v\alpha\beta'\beta\alpha'\right|^{-(T+q+r)/2}d\Gamma, \end{aligned}$$

where $k_1=\pi^{-Tp/2}\left|A+v\alpha\beta'\beta\alpha'\right|^{(q+r)/2}\Gamma_p(T+q+r)/\Gamma_p(q+r)$. Let $W=Y-X\beta\alpha'$. We have

$$E'E=(W-Z\Gamma)'(W-Z\Gamma)=W'M_ZW+(\Gamma-\hat{\Gamma})'Z'Z(\Gamma-\hat{\Gamma}),$$

where $\hat{\Gamma} = (Z'Z)^{-1}Z'W$. Thus, using Lemma 4.2

$$\begin{aligned} p(\mathcal{D}|\alpha, B, r) &= k_1 \int \left|A + v\alpha\beta'\beta\alpha' + W'M_Z W + (\Gamma - \hat{\Gamma})'Z'Z(\Gamma - \hat{\Gamma})\right|^{-(T+q+r)/2} d\Gamma \\ &= k_1 \pi^{pd/2} \frac{\Gamma_p(T+q+r-d)}{\Gamma_p(T+q+r)} \left|A + v\alpha\beta'\beta\alpha' + W'M_Z W\right|^{-(T+q+r-d)/2} \left|Z'Z\right|^{-p/2}. \end{aligned}$$

Multiplying $p(\mathcal{D}|\alpha, B, r)$ with $p(\alpha, B|r)$ and simplifying gives

$$p(D|\alpha, B, r)p(\alpha, B|r) = k_2 \frac{\Gamma_r(p)\Gamma_p(T+q+r-d)}{\Gamma_r(r)\pi^{(2pr-r^2)/2}v^{-pr/2}} \left|A + v\alpha\beta'\beta\alpha' + W'M_Z W\right|^{-(T+q+r-d)/2},$$

where $k_2 = \frac{\left|Z'Z\right|^{-p/2}|A|^{q/2}}{\pi^{p(T-d)/2}\Gamma_p(q)}$ also appears in the marginal likelihood of $r = 0$ and $r = p$ (see the proof of Theorem 5.2) and can therefore be discarded.

A.9. **Proof of Theorem 5.2.** If $r = 0$, $\Pi = 0$ and, using the prior in (5.4) on Γ and Σ, we have

$$\begin{aligned} p(\mathcal{D}|r=0) &= \iint p(\mathcal{D}|\Gamma, \Sigma)p(\Gamma, \Sigma) d\Sigma d\Gamma \\ &= (2\pi)^{-Tp/2} c_0 \int\int |\Sigma|^{-(T+p+q+1)/2} \operatorname{etr}\left\{\Sigma^{-1}[A + (Y - Z\Gamma)'(Y - Z\Gamma)]\right\} d\Sigma d\Gamma \\ &= \pi^{-Tp/2} |A|^{q/2} \frac{\Gamma_p(T+q)}{\Gamma_p(q)} \int \left|A + (Y - Z\Gamma)'(Y - Z\Gamma)\right|^{-(T+q)/2} d\Gamma \\ &= \pi^{-Tp/2} |A|^{q/2} \frac{\Gamma_p(T+q)}{\Gamma_p(q)} \int \left|A + Y'M_Z Y + (\Gamma - \hat{\Gamma})'Z'Z(\Gamma - \hat{\Gamma})\right|^{-(T+q)/2} d\Gamma \\ &= \pi^{-p(T-d)/2} |A|^{q/2} \frac{\Gamma_p(T+q-d)}{\Gamma_p(q)} \left|A + Y'M_Z Y\right|^{-(T+q-d)/2} \left|Z'Z\right|^{-p/2}, \end{aligned}$$

where $\hat{\Gamma} = (Z'Z)^{-1}Z'Y$.

If $r = p$, Π has full rank and, using the prior in (5.5), we have

$$\begin{aligned} p(\mathcal{D}|r=p) &= \iiint p(\mathcal{D}|\Pi, \Gamma, \Sigma)p(\Pi, \Gamma, \Sigma) d\Sigma d\Gamma d\Pi \\ &= (2\pi)^{-Tp/2} c_p \iiint |\Sigma|^{-(T+2p+q+1)/2} \exp\left\{-\frac{1}{2} \operatorname{tr} \Sigma^{-1}(W'W + A + v\Pi\Pi')\right\} d\Sigma d\Gamma d\Pi \\ &= k \iint \left|W'W + A + v\Pi\Pi'\right|^{-(T+p+q)/2} d\Gamma d\Pi \\ &= k \iint \left|A + v\Pi\Pi' + (Y - X\Pi')'M_Z(Y - X\Pi') + (\Gamma - \hat{\Gamma})'Z'Z(\Gamma - \hat{\Gamma})\right|^{-(T+p+q)/2} d\Gamma d\Pi, \end{aligned}$$

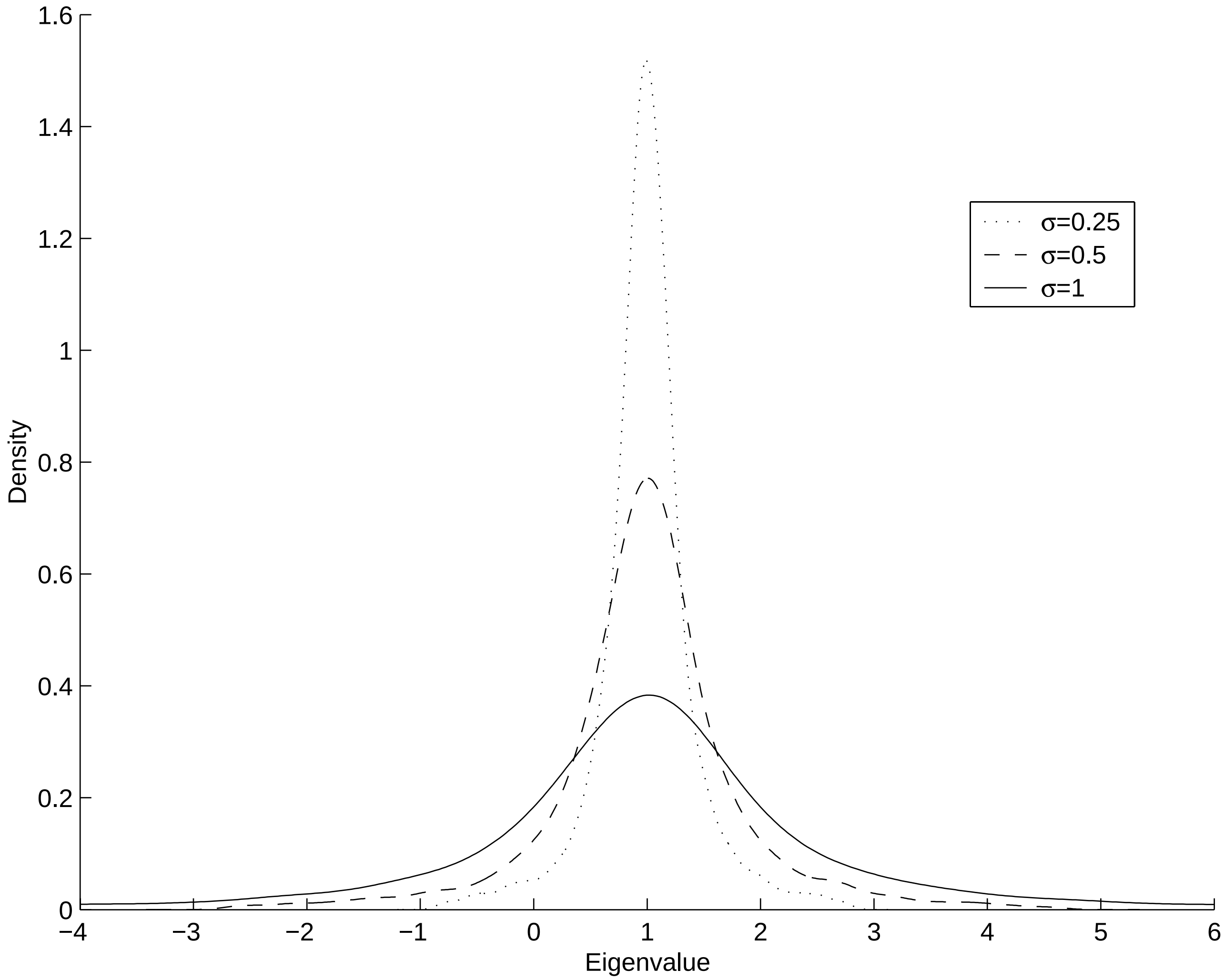
where $W = Y - X\Pi' - Z\Gamma$, $\hat{\Gamma} = (Z'Z)^{-1}Z'(Y - X\Pi')$ and $k = c_p \pi^{-Tp/2} 2^{p(p+q)/2} \pi^{p(p-1)/4} \Gamma_p(T+p+q)$. Thus, using Lemma 4.2 twice,

$$\begin{aligned} p(\mathcal{D}|r=p) &= k\pi^{pd/2} \frac{\Gamma_p(T+p+q-d)}{\Gamma_p(T+p+q)} \left|Z'Z\right|^{-p/2} \int \left|A + v\Pi\Pi' + (Y - X\Pi')'M_Z(Y - X\Pi')\right|^{-(T+p+q-d)/2} d\Pi \\ &= k\pi^{pd/2} \frac{\Gamma_p(T+p+q-d)}{\Gamma_p(T+p+q)} \left|Z'Z\right|^{-p/2} \int \left|S + (\Pi' - \hat{\Pi}')'C_1(\Pi' - \hat{\Pi}')\right|^{-(T+p+q-d)/2} d\Pi \\ &= k\pi^{pd/2} \pi^{p^2/2} \frac{\Gamma_p(T+q-d)}{\Gamma_p(T+q+p)} \left|Z'Z\right|^{-p/2} |S|^{-(T+q-d)/2} |C_1|^{-p/2} \\ &= |A|^{q/2} v^{p^2/2} \pi^{-p(T-d)/2} \frac{\Gamma_p(T+q-d)}{\Gamma_p(q)} \left|Z'Z\right|^{-p/2} |S|^{-(T+q-d)/2} |C_1|^{-p/2}, \end{aligned}$$

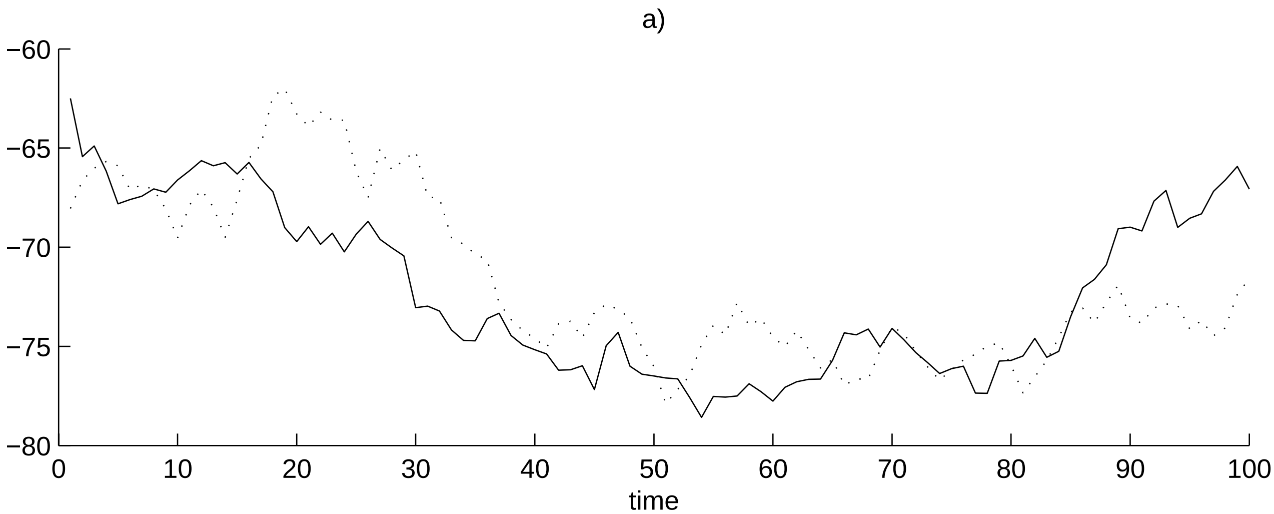
where $\hat{\Pi}$ and S are defined in Theorem 4.5 and C_1 in Theorem 4.3.

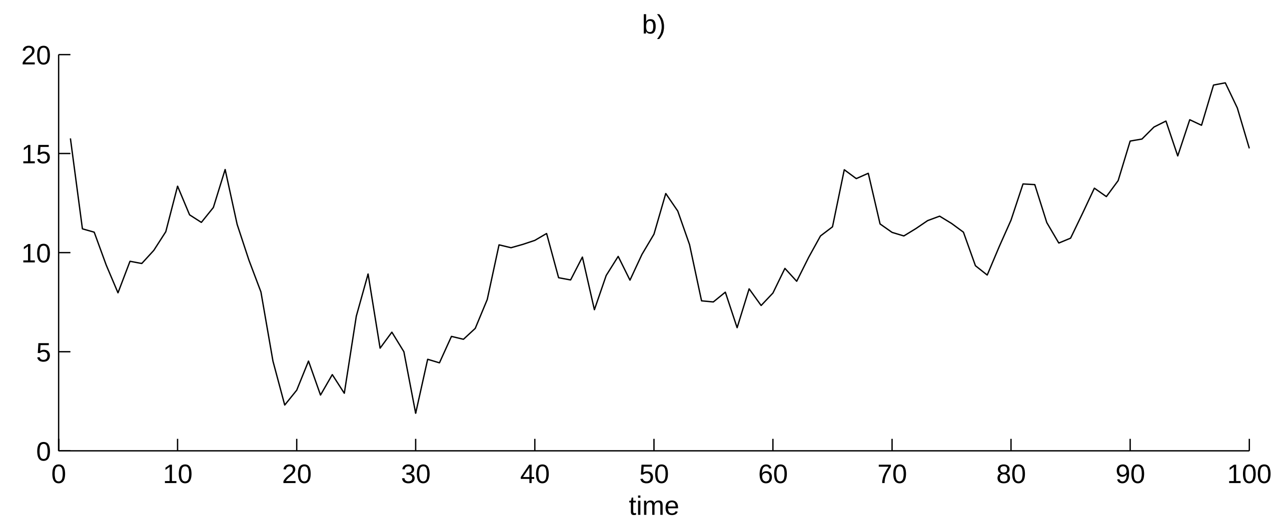
The multiplicative constant $\frac{\left|Z'Z\right|^{-p/2}|A|^{q/2}}{\pi^{p(T-d)/2}\Gamma_p(q)}$ which appears in both $p(\mathcal{D}|r=0)$ and $p(\mathcal{D}|r=p)$ also appears in the marginal likelihood of $r = 1, ..., p-1$ (see the proof of Theorem 5.1) and can therefore be discarded.

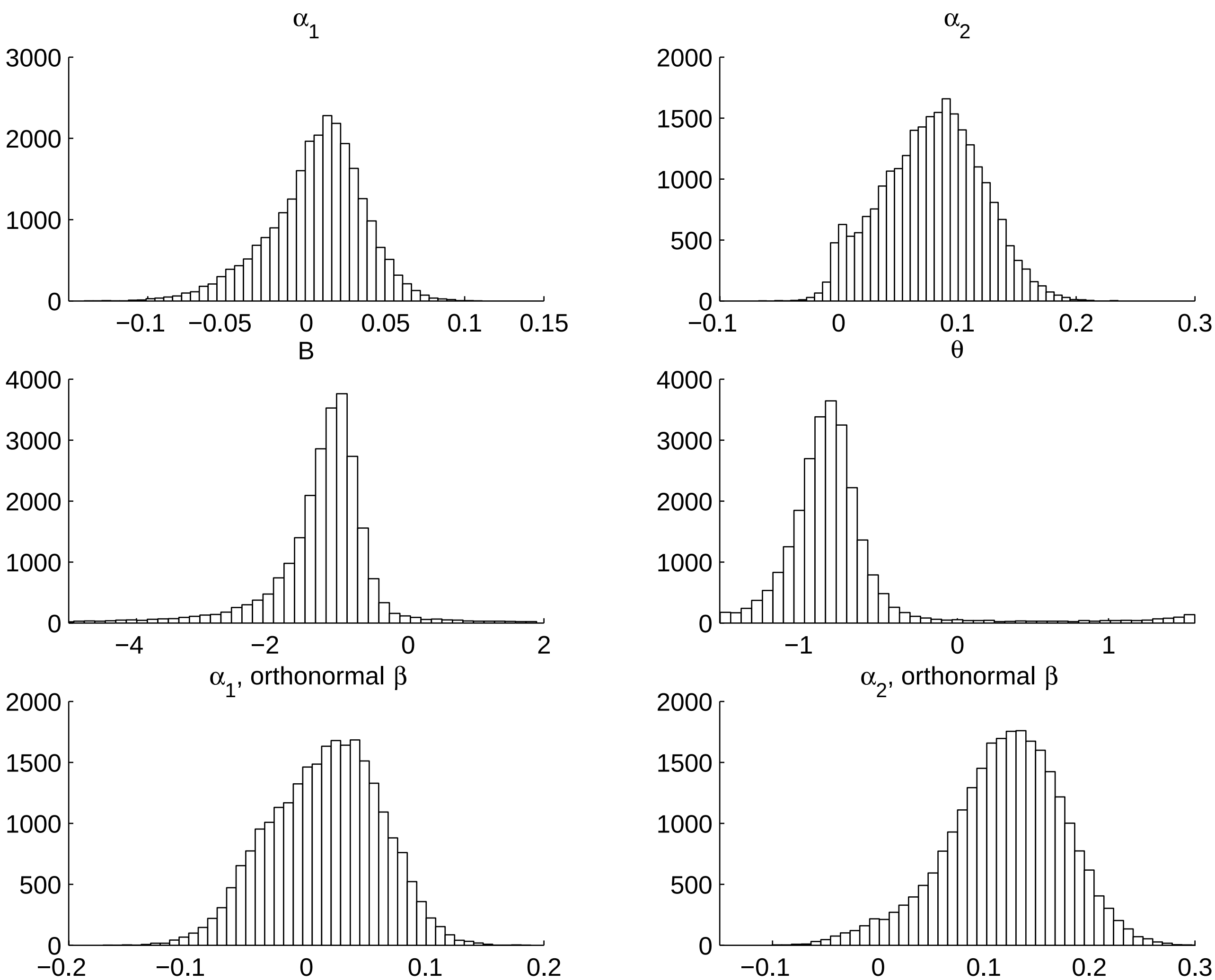
Department of Statistics, Stockholm University, S-106 91 Stockholm, Sweden
E-mail address: mattias.villani@stat.su.se



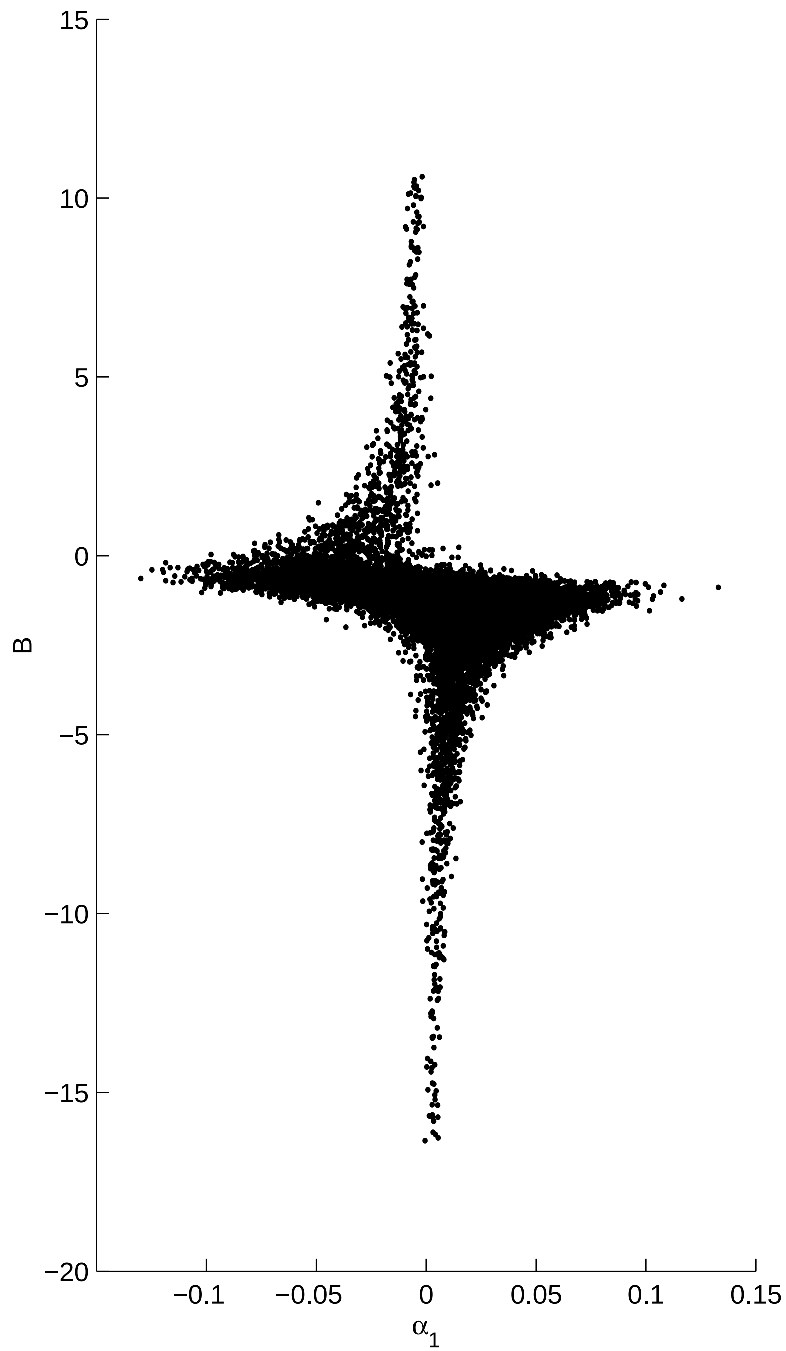
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Density
−4
−3
−2
−1
0
1
2
3
4
5
6
Eigenvalue
σ=0.25
σ=0.5
σ=1

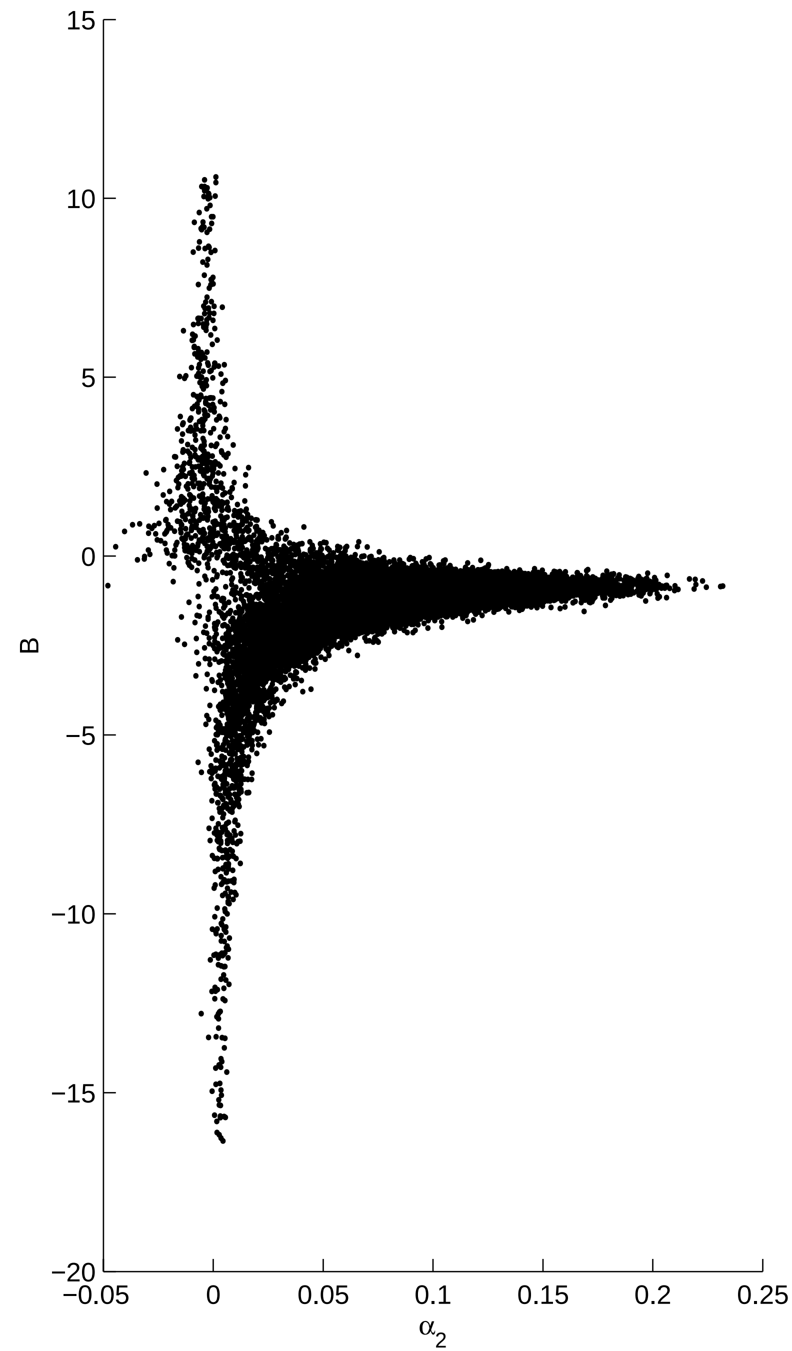

a)
−60
−65
−70
−75
−80
0
10
20
30
40
50
60
70
80
90
100
time


b)
20
15
10
5
0
0
10
20
30
40
50
60
70
80
90
100
time



α_1
3000
2000
1000
0
−0.1
−0.05
0
0.05
0.1
0.15
α_2
2000
1500
1000
500
0
−0.1
0
0.1
0.2
0.3
B
4000
3000
2000
1000
0
−4
−2
0
2
θ
4000
3000
2000
1000
0
−1
0
1
α_1, orthonormal β
2000
1500
1000
500
0
−0.2
−0.1
0
0.1
0.2
α_2, orthonormal β
2000
1500
1000
500
0
−0.1
0
0.1
0.2
0.3


B
15
10
5
0
−5
−10
−15
−20
−0.1
−0.05
0
0.05
0.1
0.15
α_1


B
15
10
5
0
−5
−10
−15
−20
−0.05
0
0.05
0.1
0.15
0.2
0.25
α_2