SVERIGES RIKSBANK
WORKING PAPER SERIES

175

SVERIGES RIKSBANK

The Multivariate Split Normal Distribution and Asymmetric Principal Components Analysis

Mattias Villani and Rolf Larsson

DECEMBER 2004

WORKING PAPERS ARE OBTAINABLE FROM

Sveriges Riksbank • Information Riksbank • SE-103 37 Stockholm
Fax international: +46 8 787 05 26
Telephone international: +46 8 787 01 00
E-mail: info@riksbank.se

The Multivariate Split Normal Distribution and Asymmetric Principal Components Analysis*

Mattias Villani† and Rolf Larsson‡

Sveriges Riksbank Working Paper Series
No. 175
December 2004

Abstract

The multivariate split normal distribution extends the usual multivariate normal distribution by a set of parameters which allows for skewness in the form of contraction/dilation along a subset of the principal axes. This paper derives some properties for this distribution, including its moment generating function, multivariate skewness and kurtosis. Maximum likelihood estimation is discussed and a complete Bayesian analysis of the multivariate split normal model is developed.

Keywords: Bayesian inference, Elicitation, Estimation, Maximum likelihood, Multivariate analysis, Skewness.

JEL classification: C11, C16.

---

*The first author gratefully acknowledges financial support from the Swedish Council of Research in Humanities and Social Sciences (HSFR), grant no. F0582/1999 and the Swedish Research Council (Vetenskapsrådet) grant no. 412-2002-1007. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

†*Sveriges Riksbank, Research department, SE-103 37 Stockholm, Sweden*, and *Department of Statistics, Stockholm University. E-mail*: mattias.villani@riksbank.se.

‡*Department of Information Science, Uppsala University*, Box 513, SE-751 20 Uppsala, Sweden. E-mail: rolf.larsson@dis.uu.se.

1. Introduction

A natural direction for extending the normal distribution is the introduction of some sort of skewness, and several proposals have indeed emerged, both in the univariate and multivariate setting, see *e.g.* Azzalini (1985) and Azzalini and Dalla Valle (1996). Among those is the split normal distribution, or the two-piece normal, originally introduced by Gibbons and Mylroie (1973), with most of its known properties derived by John (1982); see also Kimber (1985) for some additional results. Johnson, Kotz and Balakrishnan (1994) contains references to papers where the split normal distribution is used as a statistical model. The easily interpreted form of the split normal distribution has merited its use as a convenient vehicle for elicitation of subjective beliefs (Blix and Sellin (1998); Kadane, Chan and Wolfson (1996)) which in turn has motivated extensions to the multivariate case; see *e.g.* the bivariate translation approach in Blix and Sellin (2000) and the discussion of Kadane *et al.* (1996) in Bauwens, Polasek and van Dijk (1996).

In an influential paper on Monte Carlo integration, Geweke (1989) suggested a multivariate generalization of the split normal distribution to be used in the construction of an importance function. The density was only given up to a constant and no distributional properties were presented. This paper derives some properties of this distribution and develops a complete Bayesian inference procedure for this model.

The paper is outlined as follows. The next section gives a short review of the univariate split normal distribution. Section 3 defines the multivariate split normal distribution and derives some of its properties. Maximum likelihood estimation is discussed in Section 4. The fifth section develops a Bayesian analysis of the multivariate split normal model, which may be seen as a model based principal components analysis where a subset of the principal components are allowed to have skewed distributions. The proposed inference procedures are illustrated in Section 6 on national track records for the $1,500$ meters event. The proofs have been collected in an appendix.

2. The univariate split normal distribution

The following definition is a reparametrization of the univariate split normal distribution in John (1982).

Definition 2.1. *$x \in R$ follows the univariate split normal distribution, $x \sim SN(\mu, \lambda^2, \tau^2)$, if it has density*

$$f(x) = \begin{cases} c \cdot \exp\left[-\frac{1}{2\lambda^2}(x-\mu)^2\right] & \text{if } x \le \mu \\ c \cdot \exp\left[-\frac{1}{2\tau^2\lambda^2}(x-\mu)^2\right] & \text{if } x > \mu, \end{cases}$$

where $c = \sqrt{2/\pi}\lambda^{-1}(1+\tau)^{-1}$.

The density of the $SN(\mu, \lambda^2, \tau^2)$-distribution is thus proportional to the density of the $N(\mu, \lambda^2)$-distribution to the left of the mode, μ, whereas to the right of the mode it is proportional to the density of the $N(\mu, \tau^2\lambda^2)$-distribution. For $\tau < 1$ the distribution is skewed to the left, for $\tau > 1$ it is skewed to the right and for $\tau = 1$ it reduces to the usual symmetric normal distribution. The fact that the SN distribution behaves like the well-known symmetric normal distribution on both sides of its mode has been considered to be a very useful property in practical work.

John (1982) derived several properties of the univariate split normal distribution. The following result will be useful in the sequel.

Lemma 2.1. *If* $x \sim SN(\mu, \lambda^2, \tau^2)$, *then*

$$\begin{aligned} E(x) &= \mu + \sqrt{2/\pi}\lambda(\tau - 1) \\ Var(x) &= b\lambda^2 \end{aligned}$$

where $b = \frac{\pi-2}{\pi}(\tau - 1)^2 + \tau$.

The next lemma gives the univariate skewness

$$\beta_1 = \frac{E[\{x - E(x)\}^3]}{[Var(x)]^{3/2}}$$

and univariate kurtosis

$$\beta_2 = \frac{E[\{x - E(x)\}^4]}{[Var(x)]^2}$$

of a $SN(\mu, \lambda^2, \tau^2)$ variable.

Lemma 2.2. *If* $x \sim SN(\mu, \lambda^2, \tau^2)$ *then*

$$\beta_1 = b^{-3/2}\sqrt{2/\pi}(\tau - 1)[(4/\pi - 1)(\tau - 1)^2 + \tau]$$

and

$$\beta_2 = b^{-2}q,$$

where $q = 3(1 + \tau^5)/(1 + \tau) - 4\pi^{-2}(1 - \tau)^2 \left[(3 + \pi)(1 + \tau^2) + 3(\pi - 2)\tau\right]$.

The next lemma gives the moment generating function $\phi_x(t) = E(e^{tx})$ of a univariate split normal variable as derived by John (1982).

Lemma 2.3. *If* $x \sim SN(\mu, \lambda^2, \tau^2)$, *then*

$$\phi_x(t) = \frac{2\lambda\left\{\exp(-\lambda^2 t^2/2)\Phi(-\lambda t) + \tau\exp(-\lambda^2\tau^2 t^2/2)\Phi(-\lambda\tau t)\right\}}{\lambda(1 + \tau)\exp(\mu t)}.$$

3. The multivariate split normal distribution

The following definition is a natural generalization of the univariate split normal distribution in John (1982) to the multivariate setting and is a reparametrization of the multivariate split normal distribution in Geweke (1989).

Definition 3.1. *A vector* $x \in R^p$ *follows the q-split normal distribution,* $x \sim SN_p(\mu, \Sigma, \tau, \mathcal{Q})$, *if its principal components are independently distributed as*

$$v_j'x \sim \begin{cases} SN(v_j'\mu, \lambda_j^2, \tau_j^2) & \text{if } j \in \mathcal{Q} \\ N(v_j'\mu, \lambda_j^2) & \text{if } j \in \mathcal{Q}^c, \end{cases}$$

where $\mathcal{Q} \subseteq \{1, ..., p\}$ *of size* q, $\mathcal{Q}^c = \{1, 2, ..., p\}\backslash\mathcal{Q}$ *is the complement of* $\mathcal{Q}$, v_j *is the eigenvector corresponding to the* j*th largest eigenvalue in the spectral decomposition of* $\Sigma = V\Lambda V'$, $\Lambda = \mathrm{diag}(\lambda_1^2, ..., \lambda_p^2)$ *and* $\tau = (\tau_j)_{j\in\mathcal{Q}}$ *is a* q*-dimensional vector of contraction/dilation parameters.*

Consider the case $\mathcal{Q} = \{r\}$ for illustration, *i.e.* where only the rth principal component has a skewed distribution. It is then easy to see that the density of x is

$$f(x) = \begin{cases} c \cdot \exp\left\{-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\right\} & \text{if } v_r'(x - \mu) \leq 0 \\ c \cdot \exp\left\{-\frac{1}{2}(x - \mu)'\hat{\Sigma}^{-1}(x - \mu)\right\} & \text{if } v_r'(x - \mu) > 0, \end{cases}$$

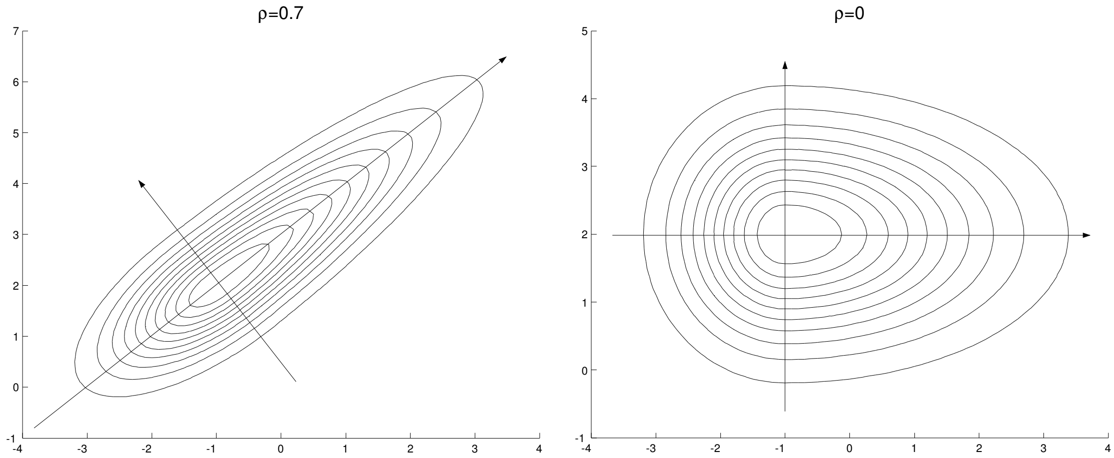

FIGURE 1. Contour plots of bivariate 1-split normal density functions. $\mu = (-1, 2)$, $\Sigma = (1, \rho; \rho, 1)$, $\mathcal{Q} = 1$ and $\tau = 2$.

where $\hat{\Sigma} = V\hat{\Lambda}V'$, $\hat{\Lambda} = diag(\lambda_1^2, ..., \tau_1^2\lambda_r^2, ..., \lambda_p^2)$ and $c^{-1} = \frac{1}{2}(2\pi)^{p/2}\left|\Lambda\right|^{1/2}(1+\tau_1)$. This should be compared to the univariate case in Definition 1. Figure 1 illustrates two possible shapes of the $SN_2(\mu, \Sigma, \tau, \mathcal{Q})$-distribution.

The general $SN_p(\mu, \Sigma, \tau, \mathcal{Q})$-distribution amounts to using different multivariate normal distributions, all with mode μ, over 2^q regions of R^p separated by the q hyperplanes $v_j'(x - \mu) = 0$, for $j \in \mathcal{Q}$. Other forms of the separating hyperplanes, or more general changes in covariance structure between the 2^q regions, produce ill-behaved densities with sharp ridges. The normal and split normal distributions in Definition 3.1 may obviously be replaced by other distributions, *e.g.* the t distribution (Geweke, 1989).

The following results give some properties of the $SN_p(\mu, \Sigma, \tau, \mathcal{Q})$ distribution. In the following theorems, let $b_j = \frac{\pi-2}{\pi}(\tau_j - 1)^2 + \tau_j$ for $j \in \mathcal{Q}$. Our first result generalizes Lemma 2.1 to the multivariate setting.

Theorem 3.1. *If* $x \sim SN_p(\mu, \Sigma, \tau, \mathcal{Q})$, *then*

$$\begin{aligned} E(x) &= \mu + \sqrt{2/\pi}\sum_{\mathcal{Q}}\lambda_j(\tau_j - 1)v_i \\ Var(x) &= V\tilde{\Lambda}V' \end{aligned}$$

where $\tilde{\Lambda}$ *is a diagonal matrix with ith element equal to* λ_j^2 *if* $j \in \mathcal{Q}^c$ *or* $b_j\lambda_j^2$ *if* $j \in \mathcal{Q}$.

Let

$$M_{xz} = (x - m)'S^{-1}(z - m),$$

be the Mahalanobis distance between two p-dimensional independent identically distributed random vectors x and z, where m and S are the common mean and covariance matrix, respectively. Mardia (1970) used M_{xz} to define a widely used measure of *multivariate skewness*

$$\beta_{1,p} = E(M_{xz}^3).$$

Note that if $x \sim N_p(\mu, \Sigma)$, then $\beta_{1,p} = 0$. $\beta_{1,p}$ is related to the univariate skewness through the equality $\beta_{1,1} = \beta_1^2$. The Mahalanobis distance may also be used to define *multivariate kurtosis* (Mardia, 1970)

$$\beta_{2,p} = E(M_{xx}^2).$$

If $x \sim N_p(\mu, \Sigma)$, then $\beta_{2,p} = p(p+2)$. Note also that $\beta_{2,1} = \beta_2$. The following result is the multivariate extension of Lemma 2.2.

Theorem 3.2. *If $x \sim SN(\mu, \Sigma, \tau, \mathcal{Q})$ then*

$$\begin{aligned}\beta_{1,p} &= \sum_{\mathcal{Q}} b_j^{-3}(2/\pi)(\tau_j - 1)^2[(4/\pi - 1)(\tau_j - 1)^2 + \tau_j]^2 \\ \beta_{2,p} &= p(p+2) + \sum_{\mathcal{Q}} b_j^{-2} q_j - 3q,\end{aligned}$$

where $q_j = 3(1+\tau_j^5)/(1+\tau_j) - 4\pi^{-2}(1-\tau_j)^2\left[(3+\pi)(1+\tau_j^2) + 3(\pi-2)\tau_j\right]$.

The moment generating function $\phi_x(t) = E[\exp(t'x)]$ of a $SN_p(\mu, \Sigma, \tau, \mathcal{Q})$ variable is given in the next result.

Theorem 3.3. *If $x \sim SN_p(\mu, \Sigma, \tau, \mathcal{Q})$, then*

$$\begin{aligned}\phi_x(t) &= \left[\prod_{\mathcal{Q}} \frac{2\lambda_i\left\{\exp[-(\lambda_j v_j' t)^2/2]\Phi(-\lambda_j v_j' t) + \tau_j \exp[-(\lambda_j \tau_j v_j' t)^2/2]\Phi(-\lambda_j \tau_j v_j' t)\right\}}{\lambda_j(1+\tau_j)\exp(\mu_j v_j' t)}\right] \\ &\quad \times \exp\left\{\sum_{\mathcal{Q}^c}[\mu_j v_j' t - \frac{1}{2}(v_j' t)^2 \lambda_j^2]\right\}.\end{aligned}$$

4. Maximum likelihood estimation

Before embarking on the multivariate case we give a useful lemma concerning maximum likelihood estimation in the univariate setting.

Lemma 4.1. *Given a random sample $x_1, ..., x_n$ from $SN(\mu, \lambda^2, \tau^2)$, the likelihood, maximized over λ and τ, is*

$$\widehat{L}(\mu) = \left(\frac{2n}{\pi e}\right)^{n/2} g(\mu)^{-3n/2},$$

where

$$\begin{aligned}g(\mu) &= s_1^{1/3} + s_2^{1/3}, \\ s_1 &= \sum_{\mathcal{I}} (x_i - \mu)^2, \\ s_2 &= \sum_{\mathcal{I}^c} (x_i - \mu)^2,\end{aligned}$$

where $\mathcal{I} = \{i = 1, ..., n : (x_i - \mu) \le 0\}$ and $\mathcal{I}^c = \{i = 1, ..., n : (x_i - \mu) > 0\}$. Moreover, the maximum likelihood estimators of λ^2 and τ are

$$\begin{aligned}\widehat{\lambda}^2 &= \frac{s_1^{2/3} g(\mu)}{n}, \\ \widehat{\tau} &= \frac{s_2^{1/3}}{s_1^{1/3}}.\end{aligned}$$

We now turn to maximum likelihood estimation in the multivariate case. It is possible to maximize the likelihood analytically w.r.t. Λ and τ. The result is given in the following theorem.

Theorem 4.1. *Given a random sample of vectors* $x_1, ..., x_n$ *from* $SN(\mu, \Sigma, \tau, \mathcal{Q})$*, where* $\Sigma = V \Lambda V'$*, the likelihood, maximized w.r.t.* Λ *and* τ *is*

$$\widehat{L}(\mu, V, \mathcal{Q}) = \frac{2^{(q-p/2)n} n^{qn/2}}{(\pi e)^{pn/2}} \prod_{j \in \mathcal{Q}^c} \widehat{\lambda}_j (\mu, V)^{-n} \prod_{j \in \mathcal{Q}} g_j (\mu, V)^{-3n/2},$$

where

$$g_j(\mu, V) = s_{1j}^{1/3} + s_{2j}^{1/3},$$

where $s_{1j} = \sum_{\mathcal{I}_j} [v_j'(x_i - \mu)]^2$, $\mathcal{I}_j = \{i = 1, ..., n : v_j'(x_i - \mu) \leq 0\}$, $s_{2j} = \sum_{\mathcal{I}_j^c} [v_j'(x_i - \mu)]^2$, $\mathcal{I}_j^c = \{i = 1, ..., n : v_j'(x_i - \mu) > 0\}$, *and the maximum likelihood estimators of* λ_j^2 *and* τ_j *are*

$$\widehat{\lambda}_j^2 (\mu, V) = \begin{cases} \frac{1}{n} s_{1j}^{2/3} g_j(\mu, V) & \text{if } j \in \mathcal{Q}, \\ \frac{1}{n} \sum_{i=1}^n [v_j'(x_i - \mu)]^2 & \text{if } j \in \mathcal{Q}^c, \end{cases}$$

and

$$\widehat{\tau}_j(\mu, V) = \left(\frac{s_{2j}}{s_{1j}} \right)^{1/3}.$$

We may use this theorem for numerical maximization of the likelihood w.r.t. μ and V, for a given $\mathcal{Q}$. In the two-dimensional case, V may be explicitly parametrized as

$$V = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad -\frac{\pi}{2} < \theta \leq \frac{\pi}{2}. \tag{4.1}$$

A similar parametrization of V is available in the general case using Eulerian angles (Khatri and Mardia, 1977). Hence, maximization over V and μ is straightforwardly performed with standard numerical optimization algorithms. Alternatively, Edelman, Arias and Smith (1998) have developed optimization algorithms on the Stiefel manifold (the set of orthonormal matrices) which avoid an explicit parametrization of V.

5. Bayesian inference

Let $x_1, ..., x_n$ be a random sample from $SN_p(\mu, \Sigma, \tau, \mathcal{Q})$. The joint posterior distribution of all parameters may be written as

$$p(\mu, V, \lambda, \tau, \mathcal{Q}, q | x_1, ..., x_n) = p(\mu, V, \lambda, \tau | \mathcal{Q}, q, x_1, ..., x_n) p(\mathcal{Q}, q | x_1, ..., x_n).$$

Let us first focus on $p(\mu, V, \lambda, \tau | \mathcal{Q}, q, x_1, ..., x_n)$ and subsequently turn to the posterior inferences of $\mathcal{Q}$ and q. Given a prior distribution $p(\mu, V, \lambda, \tau | \mathcal{Q}, q)$, the posterior distribution is obtained by Bayes' theorem as

$$p(\mu, V, \lambda, \tau | \mathcal{Q}, q, x_1, ..., x_n) \propto p(x_1, ..., x_n | \mu, V, \lambda, \tau, \mathcal{Q}, q) p(\mu, V, \lambda, \tau | \mathcal{Q}, q),$$

where $p(x_1, ..., x_n | \mu, V, \lambda, \tau, \mathcal{Q}, q)$ is the likelihood function.

We note that our Bayesian procedure is applicable also in the univariate case in John (1982), who derives estimators using the method of moments and maximum likelihood.

5.1. Prior distribution.

We will assume independence between μ, V, λ and τ *a priori*. The following priors will be used for μ and the τ's

$$\begin{aligned} \mu &\sim N_p(\mu_0, \Omega_0), \\ \tau_j^{-2} &\sim Ga(\gamma_j, \delta_j), \quad j = 1, ..., q, \end{aligned}$$

with independence between the τ's *a priori*. All gamma distributions are parametrized so that, for example, $E(\tau_j^{-2}) = \gamma_j \delta_j^{-1}$.

One suggestion for a prior on the λ's is

$$\lambda_j^{-2} = \lambda_{j-1}^{-2} + \varepsilon_j, \ \ j = 1, ..., p,$$

where $\lambda_0^2 = 0$ and $\varepsilon_1, ..., \varepsilon_p$ are independently $Ga(\alpha_j, \beta_j)$ distributed. This prior satisfies the order restriction $\lambda_1 \geq \cdots \geq \lambda_p$ with probability one. Note that if $\beta_j = \beta$ for all $j = 1, ..., p$, then $\lambda_j^{-2} \sim Ga(\sum_{i=1}^{j} \alpha_i, \beta)$ which may be useful for elicitation, but is perhaps too restrictive for some applications. An alternative prior density for $(\lambda_1^{-2}, ..., \lambda_p^{-2})$ is proportional to the product of $Ga(\alpha_j, \beta_j)$ densities, $j = 1, ..., p$, except on the subset of $\mathbb{R}^p$ where the order restriction is violated where the prior density is defined to be zero. This is a less appealing prior from a substantive point of view, but has the advantage of simplifying the posterior computations. We shall for simplicity present the posterior algorithm for the latter prior.

The space of V is the oriented orthogonal group $\mathbb{O}^+(p) = \{V \in \mathbb{R}^{p \times p} : V'V = I_p$ and $v_{jj} > 0$ for $j = 1, ..., p\}$. The usual definition of a uniform distribution on $\mathbb{O}^+(p)$ is the conditional Haar invariant distribution (Anderson, 1958). To illustrate this distribution, consider the bivariate case where V may be explicitly parametrized as in (4.1). In this case, the conditional Haar invariant distribution reduces to a uniform distribution on the angle θ (James, 1954). Other more informative priors may be defined with respect to this uniform measure, for example the *matrix Fisher (MF) distribution* first proposed by Downs (1972). The matrix Fisher density is of the form

$$p(V) = \left[{}_0F_1\left(\frac{p}{2}, \frac{1}{4}F'F\right)\right]^{-1} \exp(\mathrm{tr}\, F'V)[dV],$$

where $F = V_\mu \Sigma_\mu$ is the polar decomposition of F, with $V_\mu \in \mathbb{O}(p)$ and Σ_μ is positive definite, and $[dV]$ is the probability element of V on $\mathbb{O}(p)$. The hypergeometric function ${}_0F_1\left(\frac{p}{2}, \frac{1}{4}F'F\right)$ will cancel in all posterior computations (see the Metropolis-Hastings algorithm below) and hence need not be evaluated. The matrix Fisher density has mode equal to V_μ and Σ_μ controls the spread around the modal point. The fact that the matrix of eigenvectors is restricted to have positive first element in each column only affects the matrix Fisher density by a constant and may be disregarded for the purposes here. We will assume that $V \sim MF(V_\mu, \Sigma_\mu)$ *a priori.*

5.2. **Posterior distribution of μ, V, λ, τ conditional on $\mathcal{Q}$ and q.** The posterior distribution of μ, V, λ, τ conditional on $\mathcal{Q}$ and q is intractable. We shall devise a numerical algorithm to sample from this distribution using the posterior distribution of each parameter conditional on all other parameters, the so called *full conditional posteriors.* As it turns out, some of these full conditional posteriors are non-standard and direct sampling is inefficient. We therefore sample from these distributions using the Metropolis-Hastings (MH) algorithm (see *e.g.* Gilks, Richardson and Spiegelhalter, 1996, for an introduction). At the tth step of the algorithm, a *candidate* draw is generated from a *proposal density,* $q(y_{t+1}|y_t)$ in general, which may be of essentially any form, but for efficiency reasons should furnish a reasonably accurate approximation to the target density, $p(y_{t+1})$. Note that the proposal density is allowed to depend on the most recent draw, y_t. The candidate draw is then accepted with Markov transition probability

$$\pi(y_t \rightarrow y_{t+1}) = \min\left[1, \frac{p(y_{t+1})q(y_t|y_{t+1})}{p(y_t)q(y_{t+1}|y_t)}\right]. \tag{5.1}$$

If the transition from $y_t \rightarrow y_{t+1}$ is rejected, then the Markov chain does not move, i.e. $y_{t+1} = y_t$.

The following notation will be used. $x = (x_1, ..., x_n)'$ is the $n \times p$ matrix of sample observations. Let $z_{ij} = v_j'(x_i - \mu)$ be the demeaned score of x_i on the jth principal component and $Z = (z_{ij}) = (x - \iota_n \mu')V$ the $n \times p$ matrix of demeaned principal component (PC) scores for the whole sample, where ι_n is an n dimensional vector of ones. On any such demeaned PC score matrix Z we define the sets $\mathcal{I}_j(Z) = \{i \in \{1, ..., n\} : z_{ij} \leq 0\}$, $j = 1, ..., p$. $\mathcal{I}_j$ thus contains the indices of the observations with a non-positive score on the jth demeaned principal component. Furthermore, let $n_j = |\mathcal{I}_j|$ and $\mathcal{I}_j^c = \{1, ..., n\} \backslash \mathcal{I}_j$. Let $V_\mathcal{Q} = (v_j)_{j \in \mathcal{Q}}$ and $\Lambda_\mathcal{Q} = \text{diag}(\lambda_j^2)_{j \in \mathcal{Q}}$ denote the matrix of eigenvectors and diagonal matrix of eigenvalues corresponding to the principal axes defined by $\mathcal{Q}$. $V_{\mathcal{Q}^c}$ and $\Lambda_{\mathcal{Q}^c}$ are defined analogously. Furthermore, $Z_\mathcal{Q} = (z_j)_{j \in \mathcal{Q}} = (x - \iota_n \mu')V_\mathcal{Q}$, where z_j denotes the jth column of Z, and $Z_{\mathcal{Q}^c}$ is defined correspondingly. Finally, $z_{l,j} = (z_{ij})_{i \in \mathcal{I}_j}$ and $z_{u,j} = (z_{ij})_{i \in \mathcal{I}_j^c}$.

Theorem 5.1.

- *Full conditional posterior of* λ_j

$$\lambda_j^{-2} | \mu, V, \lambda_{-j}, \tau, \mathcal{Q}, q, x \sim \begin{cases} Ga\left(\alpha_j + \frac{n}{2}, \beta_j + \frac{z_j' z_j}{2}\right) & \text{if } j \in \mathcal{Q}^c \\ Ga\left(\alpha_j + \frac{n}{2}, \beta_j + \frac{z_{l,j}' z_{l,j} + \tau_j^{-2} z_{u,j}' z_{u,j}}{2}\right) & \text{if } j \in \mathcal{Q} \end{cases}.$$

- *Full conditional posterior of* τ_j

$$p(\tau_j^{-2} | \mu, V, \lambda, \tau_{-j}, \mathcal{Q}, q, x) \propto (\tau_j^{-2})^{\gamma_j - 1}(1 + \tau_j)^{-n} \exp\left[-\tau_j^{-2}\left(\delta_j + \frac{\lambda_j^{-2} z_{u,j}' z_{u,j}}{2}\right)\right].$$

- *Full conditional posterior of* V

$$p(V | \mu, \lambda, \tau, \mathcal{Q}, q, x) \propto \exp\left\{-\frac{1}{2}\left[\Lambda_{\mathcal{Q}^c}^{-1} Z_{\mathcal{Q}^c}' Z_{\mathcal{Q}^c} + \sum_{j \in \mathcal{Q}} \lambda_j^{-2}\left(z_{l,j}' z_{l,j} + \tau_j^{-2} z_{u,j}' z_{u,j}\right)\right]\right\}.$$

- *Full conditional posterior of* μ

$$p(\mu | V, \lambda, \tau, \mathcal{Q}, q, x) \propto \exp\left\{-\frac{1}{2}\left[a(\mu) + (\mu - \bar{\mu})'(V\bar{\Lambda}^{-1}V' + \Omega_0^{-1})(\mu - \bar{\mu})\right]\right\},$$

where $\bar{\Lambda} = diag(l_1, ..., l_p)$, $l_j = n^{-1}\lambda_j^2$ *for* $j \in \mathcal{Q}^c$ *and* $l_j = [n_j + \tau_j^{-2}(n - n_j)]^{-1}\lambda_j^2$ *for* $j \in \mathcal{Q}^c$,

$$\bar{\mu} = (V\bar{\Lambda}^{-1}V' + \Omega_0^{-1})^{-1}\left(\mu_0 + \sum_{j=1}^{p} \lambda_j^{-2} v_j v_j' w_j\right),$$

where $w_j = n\bar{x}$ *if* $j \in \mathcal{Q}^c$ *and* $w_j = n_j \bar{x}_j + \tau_j^{-2}(n - n_j)\bar{x}_j^c$ *if* $j \in \mathcal{Q}$, *and*

$$a(\mu) = \sum_{j \in \mathcal{Q}} \lambda_j^{-2} \operatorname{tr} v_j v_j' \left(\sum_{i \in \mathcal{I}_j} x_i x_i' + \tau_j^{-2} \sum_{i \in \mathcal{I}_j^c} x_i x_i'\right) - \bar{\mu}'(V\bar{\Lambda}^{-1}V' + \Omega_0^{-1})\bar{\mu}.$$

The full conditional posterior of $\lambda_1, ..., \lambda_p$ is easily sampled using a standard generator of Gamma variates. A draw which violates the order restriction $\lambda_1 \geq \cdots \geq \lambda_p$ is simply rejected with probability one.

The full conditional posteriors of τ_j, V and μ are non-standard, and the MH algorithm will be used to generate variates from these distributions as described above. The proposals for

τ_j^{-2} will be sampled from $\tau_j^{-2} \sim Ga(\rho_j, \eta_j)$ with $\eta_j = \delta_j + 2^{-1}(\lambda_j^{-2} z'_{u,j} z_{u,j})$, and ρ_j chosen so that the mode of the proposal density matches that of the full conditional posterior of τ_j^{-2}:

$$\rho_j = 1 + \frac{\eta_j}{9}\left(\frac{d_j}{\eta_j} + \frac{3(\gamma_j - 1) + \eta_j}{d_j} - 1\right)^2,$$

where

$$d_j = \eta_j^{2/3}\left\{9(\gamma_j - 1) + \frac{27n}{4} - \eta_j + 3^{3/2}\eta_j^{-1/2}\left((1-\gamma_j)^3 + g_j\eta_j + (1 - \gamma_j - \frac{n}{2})\eta_j^2\right)^{1/2}\right\}^{1/3}$$

and

$$g_j = 2 - 4\gamma_j + 2\gamma_j^2 - \frac{9(1-\gamma_j)n}{2} + \frac{27n^2}{16}.$$

We now turn to the proposal for μ. Note that when μ traverses $\mathbb{R}^p$ the index sets $\mathcal{I}_1, ..., \mathcal{I}_p$ change in a discrete fashion, which in turn brings forth changes in $a(\mu)$ in the full conditional posterior of μ. The full conditional posterior of μ is therefore *not* a multivariate normal distribution, but is locally proportional to the $N_p[\bar{\mu}, (V\bar{\Lambda}^{-1}V' + \Omega_0^{-1})^{-1}]$ density on the subsets of $\mathbb{R}^p$ where $a(\mu)$ is constant. This suggests the following two reasonable proposal densities: $\mu_{t+1} \sim N_p[\bar{\mu}, h(V\bar{\Lambda}^{-1}V' + \Omega_0^{-1})^{-1}]$ or $\mu_{t+1}|\mu_t \sim N_p[\mu_t, h(V\bar{\Lambda}^{-1}V' + \Omega_0^{-1})^{-1}]$, where μ_t is the candidate draw at iteration t and $h > 0$ is a scaling factor to fine tune the algorithm.

A proposal for V is constructed by applying a random *Givens rotation* (see *e.g.* Golub and Van Loan, 1996) to the columns of the current V. The *Givens matrix* for the (i, j) coordinate plane in $\mathbb{R}^p$, denoted by $G_{ij}(\theta_{ij})$, is $p \times p$ with unities on the diagonal except in the (i, i) and (j, j) positions which are equal to $\cos\theta_{ij}$, and all off-diagonal elements are zero except in the (i, j) and (j, i) positions which contain $\sin\theta_{ij}$ and $-\sin\theta_{ij}$, respectively for $j > i$. For example, the matrix in (4.1) is the only Givens matrix in $\mathbb{R}^2$. Note that postmultiplication of V by $G_{ij}(\theta_{ij})$ amounts to a counterclockwise rotation of θ_{ij} radians of the coordinate plan spanned by the ith and the jth column of V. The coordinate plane (i, j) may be chosen randomly from the set of $p(p-1)/2$ possible coordinate planes with equal probability on all planes and the angle θ_{ij} generated from a generalized $Beta(\xi, \xi)$ density taking values in the interval $[-\pi/2, \pi/2)$. The fact that θ_{ij} is distributed symmetrically around zero makes the proposal density symmetric, i.e. $q(V_{t+1}|V_t) = q(V_t|V_{t+1})$. This leads to a simplified version of the acceptance probability (5.1) where only the target density needs to be evaluated

$$\pi(V_t \to V_{t+1}) = \min\left[1, \frac{p(V_{t+1})}{p(V_t)}\right].$$

It is of course possible to rotate along several coordinate planes simultaneously by postmultiplying with a product of Givens matrices.

An alternative approach to sample V is to use a Matrix Fisher distribution as a proposal with a mean matrix equal to the maximum likelihood estimate of V and covariance matrix modeled on the asymptotic covariance matrix of the ML estimate. An efficient algorithm for generating variates from the Matrix Fisher distribution is still to be developed, however.

Initial values for the model parameters are needed to start up the algorithm. The maximum likelihood estimate is a natural candidate. The ML estimate must be obtained by numerical optimization, however (see Section 4). A rough initial value may be obtained as follows. An estimate of the data mode may be used as initial value for μ, perhaps using a kernel density estimator (Silverman, 1986). By Theorem 3.1, V may be initialized by the matrix of eigenvectors of the sample covariance matrix and λ_j^2, $j \in \mathcal{Q}_c$, by the eigenvalues corresponding to $\mathcal{Q}_c$. The remaining eigenvalues λ_j^2, $j \in \mathcal{Q}$ may be estimated by $\lambda_j^2 = Var[(z'_{l,j}, -z'_{l,j})']$, $j \in \mathcal{Q}$, *i.e.* the variance of the jth principal component's non-positive scores and the same values reflected to the positive part of the axis; similarly we may use $\tau_j^2 = \lambda_j^{-2} Var[(z'_{u,j}, -z'_{u,j})']$, $j \in \mathcal{Q}$, as initial values for τ_j^2.

5.3. **Posterior distribution of $\mathcal{Q}$ and q.** By Bayes' theorem, the joint posterior distribution of $\mathcal{Q}$ and q is

$$p(\mathcal{Q}, q|, x_1, ..., x_n) \propto p(x_1, ..., x_n|\mathcal{Q}, q)p(\mathcal{Q}|q)p(q),$$

where

$$p(x_1, ..., x_n|\mathcal{Q}, q) = \int\int\int\int p(x_1, ..., x_n|\mu, V, \lambda, \tau, \mathcal{Q}, q)p(\mu, V, \lambda, \tau|\mathcal{Q}, q)d\mu dV d\lambda d\tau,$$

is the *marginal likelihood* of the model with q skewed principal components given by $\mathcal{Q}$. Note that for a fixed q, the set of possible $\mathcal{Q}$'s is the set of $\frac{p!}{(p-q)!}$ possible sequences of size q from the integers $\{1, ..., p\}$. Normally, some of the $(\mathcal{Q}, q)$ pairs will be assigned zero probability *a priori* and therefore excluded from the analysis.

The marginal likelihood, $p(x_1, ..., x_n|\mathcal{Q}, q)$, is not tractable but may be computed from the posterior sample from $p(\mu, V, \lambda, \tau|\mathcal{Q}, q, x_1, ..., x_n)$ using, for example, the modified harmonic mean estimator (Geweke, 1999). It should be remarked that while it is possible to use improper priors on all model parameters in the computation of $p(\mu, V, \lambda, \tau|\mathcal{Q}, q, x_1, ..., x_n)$, this is no longer an option if also q is analyzed as this will produce indeterminate marginal likelihoods (O'Hagan (1995)). It is sufficient, however, to use a proper prior on τ; μ, V and λ may still be assigned improper priors as the dimension of their spaces does not vary with q.

6. Empirical illustration

We illustrate the proposed inferential procedures on a data set on track records for 55 nations. The data are taken from the *IAAF/ATFS Track and Field Statistics Handbook for the 1984 Los Angeles Olympics*. Dawkins (1989) uses this data set to analyze eight different track events ranging from 100 meters to the marathon. Separate analyses are made for men and women. Here we restrict the analysis to the 1,500 meters event, but analyze men and women jointly. The observations are measured in minutes and are first analyzed in unstandardized form.

A scatter plot of the raw data with an overlaid kernel estimate of the density is displayed in Figure 2 (upper left corner). The skewness in the distribution is clearly visible. We investigate this formally by comparing the following three models:

(1) Symmetric model. $\mathcal{Q} = \varnothing$, $q = 0$.
(2) Skewness in the first principal component. $\mathcal{Q} = \{1\}$, $q = 1$.
(3) Skewness in both principal components. $\mathcal{Q} = \{1, 2\}$, $q = 2$.

Note that we have excluded the model with skewness in only the second principal component ($\mathcal{Q} = \{2\}$).

Our first task is to make inferences on q. The ML estimates and maximum log likelihoods ($\log \widehat{L}_q$) of our three tentative models are shown in Table 1. The likelihood ratio test, i.e. $-2(\log \widehat{L}_0 - \log \widehat{L}_1)$, of $q = 0$ against $q = 1$ gives a test statistic of 22.3, which has a p-value of $2 \cdot 10^{-6}$ with respect to its approximate χ^2 distribution with one degree of freedom. The test of $q = 1$ against $q = 2$ gives a test statistic of 12.4, which corresponds to a p-value of 0.0004. Thus, the likelihood analysis suggests quite strongly that $q = 2$.

To compute the posterior distribution of q we will use a uniform prior on all parameters with the exception of the asymmetry parameters in τ. For the sake of presentation we consider what

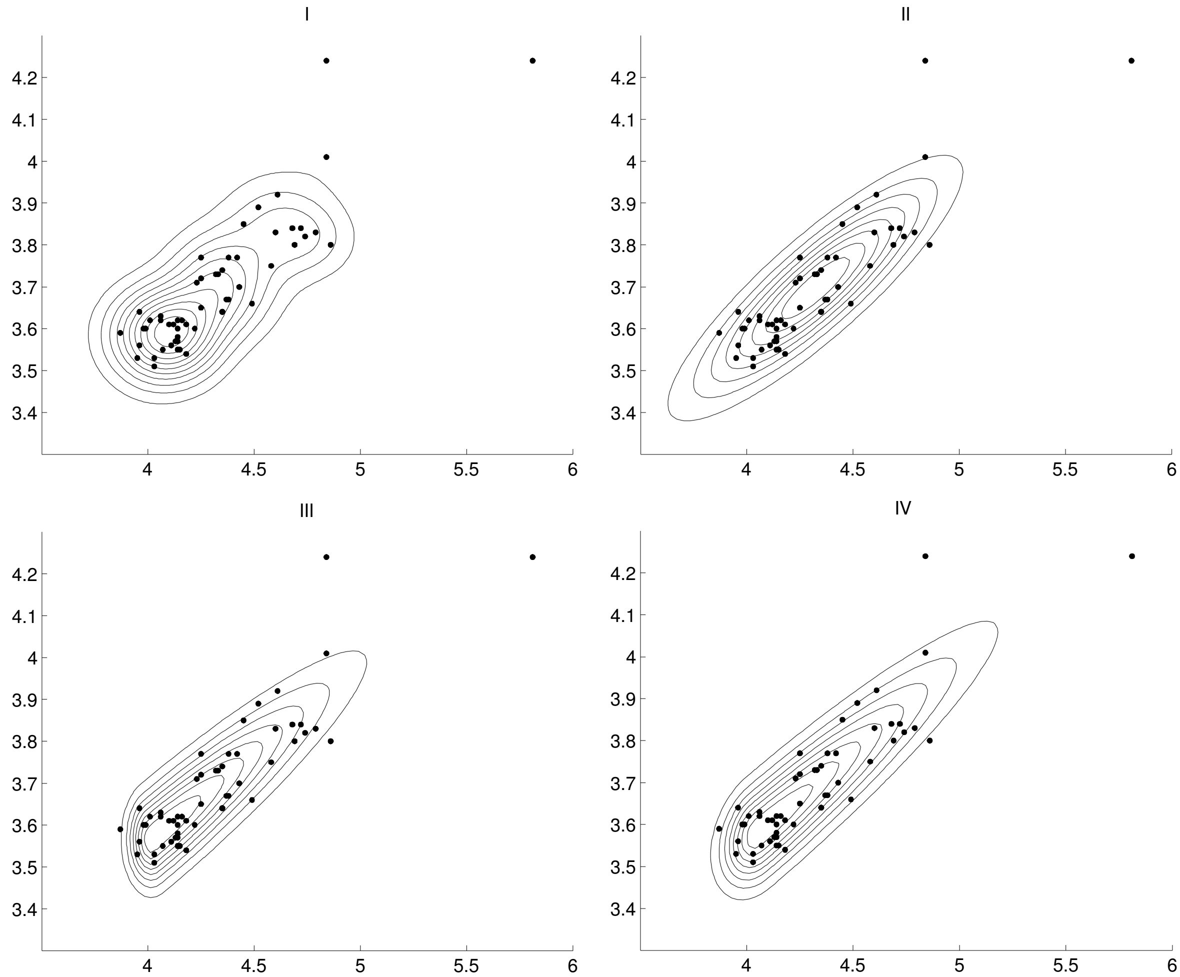

FIGURE 2. Kernel density estimate of the data (I). Estimated split normal density based on the maximum likelihood estimate for $q = 0$ (II) and $q = 1$ (III). Estimated split normal density based on the posterior mean for $q = 1$ (IV). The prior with $\delta = 1$ is used.

may be called a sceptics prior for τ which centers over the symmetric model, *i.e.* $E(\tau_1^{-2}) = E(\tau_2^{-2}) = 1$. Specifically, we assume that

$$\begin{aligned} \tau_1^{-2} &\sim Ga(\delta, \delta), \\ \tau_2^{-2} &\sim Ga(2\delta, 2\delta), \end{aligned}$$

where δ may be used to adjust the precision of the prior around the mean of unity. Note that the prior becomes tighter around the symmetric model as δ increases and that the prior variance of τ_2^{-2} is one half that of τ_1^{-2}, reflecting the judgement that the second principal component is more likely to be symmetric than the first.

All presented Bayesian analyses are based on 100,000 draws from the posterior. No convergence problems were encountered. The posterior distribution of q is given in Table 2 for several different values for δ. The Bayesian analysis is in favor of $q = 1$, unless the prior is very tightly concentrated around the symmetric model ($\delta = 50$), but there is also a relatively large posterior probability on $q = 2$. Figure 2 shows the fit of the models $q = 0$ and $q = 1$ for

	Model	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\theta}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\tau}_1$	$\widehat{\tau}_2$	$\log \widehat{L}_q$
	$q = 0$	4.33	3.70	0.41	0.361	0.068	–	–	48.35
ML	$q = 1$	4.01	3.56	0.41	0.068	0.059	8.88	–	59.48
	$q = 2$	4.02	3.52	0.38	0.059	0.027	8.87	3.54	65.68
	$q = 0$	4.326	3.698	0.406	0.365	0.070	–	–	–
Posterior mean	$q = 1$	4.051	3.580	0.407	0.109	0.068	5.339	–	–
	$q = 2$	4.065	3.566	0.398	0.114	0.077	4.948	1.423	–

Table 1. Point estimates of the models parameters. The posterior mean is computed for the prior $\delta = 1$.

Model	$\delta = .1$	$\delta = .5$	$\delta = 1$	$\delta = 3$	$\delta = 5$	$\delta = 10$	$\delta = 50$
$q = 0$	0.000	0.000	0.000	0.002	0.018	0.133	0.468
$q = 1$	0.845	0.703	0.686	0.680	0.624	0.560	0.361
$q = 2$	0.155	0.297	0.314	0.318	0.358	0.307	0.172

Table 2. Posterior distribution of q for different values on the prior hyperparameter δ.

different point estimates of the model parameters (see below). The symmetric model ($q = 0$) appears to fit the data poorly with too few observations near the center of the density. The model with one skewed principal component does a much better job.

The ML estimate and the posterior mean estimate ($\delta = 1$) of the models' parameters are shown in Table 1. Since we are using a uniform prior on μ, λ and θ, the ML and Bayes estimates are very close for those parameters. The informative sceptic's prior on τ has the effect of moving the ML estimates toward the point of symmetry ($\tau_1 = \tau_2 = 1$). This, in turn, produces larger Bayes estimates of λ_1 and λ_2 compared to the ML estimates, since the λ's must be increased when the τ's decrease in order to match the variance in the data.

We condition the remaining analysis on $q = 1$. The marginal posterior distributions of the models parameters are displayed in Figure 3; the eigenvectors in V will be analyzed below. The upper right subfigure shows that the posterior uncertainty of τ_1 is rather large but that the point of symmetry, $\tau_1 = 1$, does not belong to any reasonably sized probability interval. Note that the gamma prior on τ_1^{-2} has been converted into a prior for τ_1, which belongs to the square-root-inverted gamma family (Bernardo and Smith, 1994). Alternatively, one may look at the multivariate skewness in the upper right graph of Figure 3. The posterior distribution of the skewness is computed by inserting the posterior draws of τ_1 into the expression for the multivariate skewness in Theorem 3.2. The lower left graph shows that the proportion of total variance explained by the first principal component is rather close to unity.

The mean acceptance probabilities in the Metropolis-Hastings algorithm for $q = 1$ and the prior with $\delta = 1$ were 0.581, 0.759 and 0.975, for V, μ and τ, respectively. The very large acceptance probability of τ is a result of the Gamma proposal density being a very accurate approximation of the full conditional posterior, so that the Metropolis-Hastings τ-step is essentially a Gibbs step.

The principal components are not invariant to the scale of the original variables. To analyze the principal components in some more detail, we scale both variables to have unit variance. The ML estimate of θ is 0.777. The posterior mean of θ is 0.787 which translates into $v_1 = (0.706, 0.708)'$. The first principal component may therefore be interpreted as an overall measure of performance on $1,500$ meters for both women and men. The posterior distribution

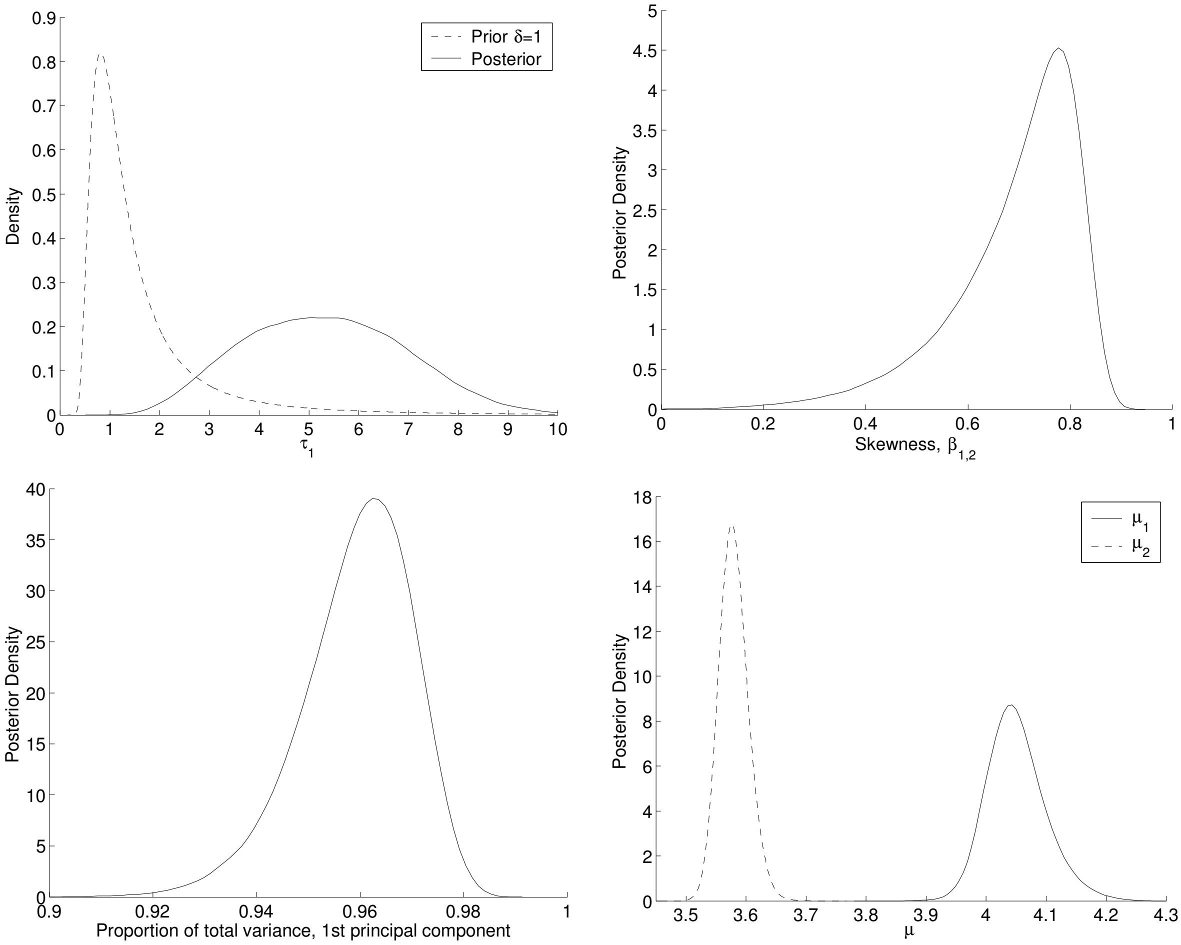

FIGURE 3. Posterior inferences conditional on $q = 1$. The prior with $\delta = 1$ is used.

of the first principal component xv_1 thus gives us the posterior distribution of the overall rank. Table 3 displays the posterior distribution of the overall rank for the top five nations. The good performance of the U.S. for both men and women places them unambiguously first in the ranking. It is also possible to compute the posterior probability that *e.g.* Norway (rank 15 according to the posterior mean of the 1st PC) is better than Kenya (rank 11 according to the posterior mean of the 1st PC), which is 0.173.

Nation	Women	Men	Post. mean	Post. distr. of overall rank (1st PC)
U.S.	2nd	3rd	1st	1st (1.00)
G. Britain	9th	1st	2nd	2nd (0.82), 3rd (0.17)
U.S.S.R.	1st	13th	3rd	2nd (0.17), 3rd (0.82), 4th (0.01)
F. Rep. of Ger.	8th	2nd	5th	3rd (0.01), 4th (0.10), 5th (0.89)
German D. Rep.	3rd	8th	4th	4th (0.89), 5th (0.16)

Table 3. Ranking of nations. Second and third columns give the nation's rank for men and women separately. The fourth column contains the ranking based on the posterior mean of the first principal component (q=1). The last column displays the posterior distribution of the overall rank based on the first principal component (q=1).

References

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York.

[2] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171-178.

[3] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715-726.

[4] Bauwens, L., Polasek, W. and van Dijk, H. K. (1996). Editors' introduction, *Journal of Econometrics*, **75**, 1-5.

[5] Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, Wiley & Sons, New York.

[6] Blix, M. and Sellin, P. (1998). Uncertainty bands for inflation forecasts, Sveriges Riksbank Working Paper Series No. **65**.

[7] Blix, M. and Sellin, P. (2000). A bivariate distribution for inflation and output forecasts, Sveriges Riksbank Working Paper Series No. **102**.

[8] Dawkins, B. (1989). Multivariate analysis of national track records, *American Statistician*, **43**, 110-115.

[9] Downs, T. D. (1972). Orientation statistics, *Biometrika*, **59**, 665-676.

[10] Edelman, A., Arias, T. A. and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints, *SIAM Journal of Matrix Analysis and Applications*, **20** (2), 303-353.

[11] Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, **57** (6), 1317-1339.

[12] Geweke, J. (1999). Using simulation methods for Bayesian econometrics models: inference, development and communication, *Econometric Reviews*, **18** (1), 1-73.

[13] Gibbons, J. F. and Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions, *Applied Physics Letters*, **22**, 568-569.

[14] Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

[15] Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations* (3rd edition), John Hopkins University Press, London.

[16] James, A. T. (1954). Normal multivariate analysis and the orthogonal group, *Annals of Mathematical Statistics*, **25**, 40-74.

[17] John, S. (1982). The two-piece normal family of distributions and its fitting, *Communications in Statistics - Theory and Methods*, **11** (8), 879-885.

[18] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. I.

[19] Kadane, J. B., Chan, N. H. and Wolfson, L. J. (1996). Priors for unit root models, *Journal of Econometrics*, **75**, 99-111.

[20] Khatri, C. G. and Mardia, K. V. (1977). The von Mises–Fisher matrix distribution in orientation statistics. *Journal of the Royal Statistical Society*, B **39**, 95-106.

[21] Kimber, A. C. (1985). Methods for the two-piece normal distribution, *Communications in Statistics - Theory and Methods*, **14** (1), 235-245.

[22] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519-530.

[23] O'Hagan, A. (1995). Fractional Bayes factors for model comparisons. *Journal of the Royal Statistical Society,* **B 57**, 99-138.
[24] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Appendix A. Proofs

A.1. **Proof of Lemma 2.2.** From John (1982) we have

$$E[\{x-E(x)\}^3]=\sqrt{2/\pi}\lambda(\tau-1)[(4/\pi-1)(\lambda(\tau-1))^2+\lambda^2\tau].$$

Make the transformation $y=(x-\mu)/\lambda$. It is easy to see that $y\sim SN(0,1,\tau^2)$ with $Var(y)=\lambda^{-2}Var(x)=b$. Since skewness is invariant to linear transformations we have

$$\beta_1(x)=\beta_1(y)=\frac{E[\{y-E(y)\}^3]}{[Var(y)]^{3/2}}=b^{-3/2}\sqrt{2/\pi}(\tau-1)[(4/\pi-1)(\tau-1)^2+\tau].$$

Similarly, since kurtosis is invariant to linear transformations

$$\beta_2(x)=\beta_2(y)=\frac{E[\{y-E(y)\}^4]}{[Var(y)]^2}=b^{-2}E[\{y-E(y)\}^4],$$

where

$$E[\{y-E(y)\}^4]=E(y^4)-4E(y^3)E(y)+6E(y^2)[E(y)]^2-3[E(y)]^4. \tag{A.1}$$

and

$$\begin{aligned} E(y) &= \sqrt{2/\pi}(\tau-1)\\ E(y^2) &= (1-\tau)^2+\tau\\ E(y^3) &= 2\sqrt{2/\pi}(\tau^4-1)/(1+\tau)\\ E(y^4) &= 3(1+\tau^5)/(1+\tau). \end{aligned}$$

Inserting these moments into (A.1) and simplifying yields

$$E[\{y-E(y)\}^4]=3(1+\tau^5)/(1+\tau)-4\pi^{-2}(1-\tau)^2[(3+\pi)(1+\tau^2)+3(\pi-2)\tau)],$$

which proves the result.

A.2. **Proof of Theorem 3.1.** Since $x=Vy$, where y is the vector of principal components, we have

$$\begin{aligned} E(x) &= VE(y)=\sum_{\mathcal{Q}}v_jE(y_j)+\sum_{\mathcal{Q}^c}v_jE(y_j)=\sum_{\mathcal{Q}}v_j[v_j'\mu+\sqrt{2/\pi}\lambda_j(\tau_j-1)]+\sum_{\mathcal{Q}^c}v_jv_j'\mu\\ &= \mu+\sqrt{2/\pi}\sum_{\mathcal{Q}}\lambda_j(\tau_j-1)v_j, \end{aligned}$$

by Lemma 2.1.

The covariance matrix can be written

$$Var(x)=V\cdot Var(y)\cdot V'=\sum_{i=1}^{p}Var(y_j)v_jv_j'=\sum_{\mathcal{Q}}b_j\lambda_j^2v_jv_j'+\sum_{\mathcal{Q}^c}\lambda_j^2v_jv_j'=V\Lambda_{\mathcal{Q}}V',$$

again using Lemma 2.1.

A.3. **Proof of Theorem 3.2.** We first prove the expression for the multivariate skewness. Since $x = Vy$, where y are the principal components of x,

$$\beta_{1,p}(x) = \beta_{1,p}(Vy) = \beta_{1,p}(y),$$

by the invariance of $\beta_{1,p}$ under linear transformations (Mardia, 1970). Let v and w be independent random vectors from the same distribution of y, $m = (m_1, ..., m_p)' = E(y)$ and $Var(y) = \tilde{\Lambda} = Diag(\sigma_1^2, ..., \sigma_p^2)$, where $\sigma_j^2 = \lambda_j^2$ if $j \in \mathcal{Q}^c$ and $\sigma_j^2 = b_j\lambda_j^2$ if $j \in \mathcal{Q}$. By definition, $\beta_{1,p}(y) = E(M_{vw}^3)$, where M_{vw} may be decomposed as

$$M_{vw} = (v-m)'\tilde{\Lambda}^{-1}(w-m) = \Sigma_{j=1}^p \sigma_j^{-2}(v_j - m_j)(w_j - m_j) = \Sigma_{j=1}^p M_{v_j w_j},$$

and therefore

$$M_{vw}^3 = \sum_{r_1+...+r_p=3} \frac{3!}{r_1! \cdots r_p!} M_{v_1w_1}^{r_1} \cdots M_{v_pw_p}^{r_p}.$$

Since $E(M_{v_jw_j}) = 0$ for $j = 1, ..., p$, by the independence of the elements of v and w, we have

$$E(M_{vw}^3) = \Sigma_{j=1}^p E(M_{v_jw_j}^3),$$

which proves that

$$\beta_{1,p}(y) = \Sigma_{j=1}^p \beta_{1,1}(y_j).$$

Since $\beta_{1,1} = \beta_1^2$ and $\beta_j = 0$ for $j \in \mathcal{Q}^c$, the result now follows from Lemma 2.2.

To derive the multivariate kurtosis, note that

$$\beta_{2,p}(x) = \beta_{2,p}(Vy) = \beta_{2,p}(y),$$

by the invariance of $\beta_{2,p}$ under linear transformations (Mardia, 1970). Now, by the diagonality of $Var(y) = \tilde{\Lambda}$,

$$M_{yy}^2 = \Sigma_{j=1}^p M_{y_jy_j}^2 + 2\sum_{i<j} M_{y_iy_i} M_{y_jy_j}.$$

Thus,

$$\begin{aligned} E(M_{yy}^2) &= \Sigma_{j=1}^p E(M_{y_jy_j}^2) + 2\sum_{i<j} E(M_{y_iy_i})E(M_{y_jy_j}) \\ &= \Sigma_{j=1}^p E(M_{y_jy_j}^2) + p(p-1). \end{aligned}$$

since $E(M_{y_jy_j}) = 1$ for $j = 1, ..., p$. Thus

$$\beta_{2,p}(y) = \Sigma_{j=1}^p \beta_{2,p}(y_j) + p(p-1) = \sum_{\mathcal{Q}} \beta_2(y_j) - 3q + p(p+2).$$

since $\beta_{2,p}(y_j) = \beta_2(y_j)$ and $\beta_2(y_j) = 3$ for $j \in \mathcal{Q}^c$. The result now follows from Lemma 2.2.

A.4. **Proof of Theorem 3.3.** Since $x = Vy$,

$$\begin{aligned} \phi_x(t) &= E[\exp(t'x)] = E[\exp(t'Vy)] = \prod_{\mathcal{Q}^c} E[\exp(t'v_jy_j)] \prod_{\mathcal{Q}} E[\exp(t'v_jy_j)] \\ &= \prod_{\mathcal{Q}^c} \phi_{y_j}(v_j't) \prod_{\mathcal{Q}} \phi_{y_j}(v_j't) = \exp\left\{\sum_{\mathcal{Q}^c} [\mu_j v_j't - \frac{1}{2}(v_j't)^2\lambda_j^2]\right\} \prod_{\mathcal{Q}} \phi_{y_j}(v_j't), \end{aligned}$$

where, using Lemma 2.3,

$$\phi_{y_j}(v_j't) = \frac{2\lambda_j \left\{ \exp[-(\lambda_j v_j't)^2/2]\Phi(-\lambda_j v_j't) + \tau_j \exp[-(\lambda_j \tau_j v_j't)^2/2]\Phi(-\lambda_j \tau_j v_j't) \right\}}{\lambda_j(1+\tau_j)\exp(\mu_j v_j't)} \quad \text{for } j \in \mathcal{Q}.$$

A.5. **Proof of Lemma 4.1.** It follows from John (1982) that the log likelihood is

$$l(\mu, \lambda, \tau) = \frac{n}{2}\log\left(\frac{2}{\pi}\right) - n\log\lambda - n\log(1+\tau) - \frac{1}{2\lambda^2}\left(s_1 + \frac{s_2}{\tau^2}\right),$$

where

$$\begin{aligned} s_1 &= \sum_{\mathcal{I}} (x_i - \mu)^2, \\ s_2 &= \sum_{\mathcal{I}^c} (x_i - \mu)^2. \end{aligned}$$

Hence, the first derivatives are

$$\begin{aligned} \frac{\partial l}{\partial \lambda} &= -\frac{n}{\lambda} + \frac{1}{\lambda^3}\left(s_1 + \frac{s_2}{\tau^2}\right), \\ \frac{\partial l}{\partial \tau} &= -\frac{n}{1+\tau} + \frac{s_2}{\lambda^2\tau^3}, \end{aligned}$$

implying

$$\begin{aligned} \widehat{\tau} &= \frac{s_2^{1/3}}{s_1^{1/3}}, \\ \widehat{\lambda}^2 &= \frac{\left(s_1^{1/3} + s_2^{1/3}\right) s_1^{2/3}}{n}. \end{aligned}$$

Insertion of these into the log likelihood yields

$$l\left(\mu, \widehat{\lambda}, \widehat{\tau}\right) = \frac{n}{2}\log\left(\frac{2n}{\pi e}\right) - \frac{3}{2}n\log\{g(\mu)\},$$

and the result follows. It is straightforward to check that we indeed have a maximum.

A.6. **Proof of Theorem 4.1.** Write $z_i = V'(x_i - \mu)$, for observation i, where $i = 1, 2, ..., n$. The probability density of z_i is

$$f(z_i) = c_1 \left(\prod_{j=1}^{p} \lambda_j\right)^{-1} \left\{\prod_{j\in\mathcal{Q}} (1+\tau_j)\right\}^{-1} \exp\left(-\frac{1}{2}y_i'\widetilde{\Lambda}^{-1}y_i\right),$$

where

$$c_1 = \left(\frac{1}{\sqrt{2\pi}}\right)^{p-q} \left(\sqrt{\frac{2}{\pi}}\right)^{q},$$

$\widetilde{\Lambda}$ is diagonal with ith element $\lambda_j^2\tau_j^2$ if $j \in \mathcal{Q}$ and $z_{ij} \geq 0$, and λ_j^2 otherwise. Hence, the likelihood is

$$L(\mu, \Sigma, \tau, \mathcal{Q}) = c_1^n \left(\prod_{j=1}^{p} \lambda_j\right)^{-n} \left\{\prod_{j\in\mathcal{Q}} (1+\tau_j)\right\}^{-n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} z_i'\widetilde{\Lambda}^{-1}z_i\right),$$

where, denoting the components of z_i by z_{ij}, $j = 1, ..., p$,

$$z_i'\widetilde{\Lambda}^{-1}z_i = \sum_{j\in\mathcal{Q}^c}\lambda_j^{-2}z_{ij}^2 + \sum_{j\in\mathcal{Q}}\lambda_j^{-2}z_{ij}^2\left(1_{\{z_{ij}<0\}} + \tau_j^{-2}1_{\{z_{ij}\geq 0\}}\right),$$

where 1_A is the indicator function of the event A. Thus,

$$\begin{aligned} L(\mu,\Sigma,\tau,\mathcal{Q}) &= c_1^n\prod_{j\in\mathcal{Q}^c}\left\{\lambda_j^{-n}\exp\left(-\frac{1}{2}\sum_{i=1}^n\lambda_j^{-2}z_{ij}^2\right)\right\} \\ &\times\prod_{j\in\mathcal{Q}}\lambda_j^{-n}(1+\tau_j)^{-n}\exp\left\{-\frac{1}{2}\sum_{i=1}^n\lambda_j^{-2}z_{ij}^2\left(1_{\{z_{ij}<0\}}+\tau_j^{-2}1_{\{z_{ij}\geq 0\}}\right)\right\} \\ &= \prod_{j\in\mathcal{Q}^c}L_{1,j}(\lambda_i)\prod_{j\in\mathcal{Q}}L_{2,j}(\lambda_j,\tau_j), \end{aligned}$$

where $L_{1,j}(\lambda_j)$ is the likelihood arising from observations of n independent $N\left(0,\lambda_j^2\right)$ variables and $L_{2,j}(\lambda_j,\tau_j)$ is the likelihood arising from observations of n independent univariate $SN(0,\lambda_j,\tau_j)$ variables. Hence, from Lemma 4.1 and standard theory, the likelihood maximized w.r.t. Λ and τ is

$$\widehat{L} = c_2\prod_{j\in\mathcal{Q}^c}\widehat{\lambda}_j^{-n}\prod_{j\in\mathcal{Q}}g_j^{-3n/2},$$

where

$$\begin{aligned} c_2 &= \left(\frac{1}{2\pi e}\right)^{(p-q)n/2}\left(\frac{2n}{\pi e}\right)^{qn/2} = \frac{2^{(q-p/2)n}n^{qn/2}}{(\pi e)^{pn/2}}, \\ \widehat{\lambda}_j^2(\mu,V) &= \begin{cases}\frac{1}{n}s_{1j}^{2/3}g_j \text{ if } j\in\mathcal{Q}, \\ \frac{1}{n}\sum_{i=1}^n z_{ij}{}^2 \text{ if } j\in\mathcal{Q}^c,\end{cases} \\ \widehat{\tau}_j(\mu,V) &= \left(\frac{s_{2j}}{s_{1j}}\right)^{1/3}, \end{aligned}$$

and

$$g_j = \left(\sum_{i=1}^n z_{ij}^2 1_{\{z_{ij}<0\}}\right)^{1/3} + \left(\sum_{i=i}^n z_{ij}^2 1_{\{z_{ij}\geq 0\}}\right)^{1/3}.$$

Inserting $z_{ij} = v_j'(x_i-\mu)$ yields the result.

A.7. **Proof of Theorem 5.1.** From the proof of Theorem 4.1, the likelihood function can be written

$$\begin{aligned} p(x_1,...,x_n|\mu,V,\lambda,\tau,\mathcal{Q},q) &= \frac{\pi^{-np/2}}{2^{n(p-2q)/2}}\left\{\prod_{j\in\mathcal{Q}^c}\lambda_j^{-n}\exp\left(-\lambda_j^{-2}\frac{z_j'z_j}{2}\right)\right\} \\ &\times\left\{\prod_{j\in\mathcal{Q}}\frac{1}{\lambda_j^n(1+\tau_j)^n}\exp\left[-\frac{1}{2\lambda_j^2}\left(z_{l,j}'z_{l,j}+\tau_j^{-2}z_{u,j}'z_{u,j}\right)\right]\right\}, \end{aligned}$$

where z_j denotes the jth column of $Z = (x-\iota_n\mu')V$, $z_{l,j} = (z_{ij})_{i\in\mathcal{I}_j}$ and $z_{u,j} = (z_{ij})_{i\in\mathcal{I}_j^c}$.

The full conditional posteriors of λ_j and τ_j follow directly from multiplying the likelihood with the priors $\lambda_j^{-2}\sim Ga(\alpha_j,\beta_j)$ and $\tau_j^{-2}\sim Ga(\gamma_j,\delta_j)$, respectively.

To obtain the full conditional posterior of V we rewrite the first factor of the likelihood as follows

$$\prod_{j\in\mathcal{Q}^c}\lambda_j^{-n}\exp\left(-\lambda_j^{-2}\frac{z_j'z_j}{2}\right)\propto\exp\left(-\frac{1}{2}\sum_{j\in\mathcal{Q}^c}\lambda_j^{-2}z_j'z_j\right)=\exp\left(-\frac{1}{2}\operatorname{tr}\Lambda_{\mathcal{Q}^c}^{-1}Z_{\mathcal{Q}^c}'Z_{\mathcal{Q}^c}\right).$$

Combining this factor with the factor of the likelihood function corresponding to $\mathcal{Q}$ we obtain the result in Theorem 5.1.

The conditional likelihood of μ reads

$$-\frac{1}{2}\ln p(\mu|\cdot)\propto\sum_{j\in\mathcal{Q}^c}\lambda_j^{-2}z_j'z_j+\sum_{j\in\mathcal{Q}}\lambda_j^{-2}\left(\sum_{i\in\mathcal{I}_j}z_{ij}^2+\tau_j^{-2}\sum_{i\in\mathcal{I}_j^c}z_{ij}^2\right)\tag{A.2}$$

Let us rewrite this expression to show explicitly its dependence on μ. All terms which do not involve μ will be discarded. Note first that

$$\sum_{i\in\mathcal{I}_j}z_{ij}^2=\sum_{i\in\mathcal{I}_j}[v_j'(x_i-\mu)]^2=\operatorname{tr}A_j\sum_{i\in\mathcal{I}_j}x_ix_i{}'+n_j(\mu'A_j\mu-2\mu'A_j\bar{x}_j),$$

where $A_j=v_jv_j'$, and

$$\tau_j^{-2}\sum_{i\in\mathcal{I}_j^c}z_{ij}^2=\operatorname{tr}A_j\tau_j^{-2}\sum_{i\in\mathcal{I}_j^c}x_ix_i{}'+\tau_j^{-2}(n-n_j)(\mu'A_j\mu-2\mu'A_j\bar{x}_j^c),$$

where $\bar{x}_j=n_j^{-1}\sum_{i\in\mathcal{I}_j}x_i$ and $\bar{x}_j^c=(n-n_j)^{-1}\sum_{i\in\mathcal{I}_j^c}x_i$. Thus,

$$\sum_{j\in\mathcal{Q}}\lambda_j^{-2}\left(\sum_{i\in\mathcal{I}_j}z_{ij}^2+\tau_j^{-2}\sum_{i\in\mathcal{I}_j^c}z_{ij}^2\right)=B_0+\mu'B\mu-2\mu'b,\tag{A.3}$$

where $B_0=\sum_{j\in\mathcal{Q}}\lambda_j^{-2}\operatorname{tr}A_j\left(\sum_{i\in\mathcal{I}_j}x_ix_i{}'+\tau_j^{-2}\sum_{i\in\mathcal{I}_j^c}x_ix_i{}'\right)$, $B=\sum_{j\in\mathcal{Q}}\lambda_j^{-2}v_jv_j'[n_j+\tau_j^{-2}(n-n_j)]$ and $b=\sum_{j\in\mathcal{Q}}\lambda_j^{-2}v_jv_j'[n_j\bar{x}_j+\tau_j^{-2}(n-n_j)\bar{x}_j^c]$. Correspondingly,

$$\sum_{j\in\mathcal{Q}^c}\lambda_j^{-2}z_j'z_j=C_0+\mu'C\mu-2\mu'c,\tag{A.4}$$

where $C_0=\sum_{j\in\mathcal{Q}^c}\lambda_j^{-2}\operatorname{tr}(A_j\sum_{i=1}^n x_ix_i{}')=\operatorname{tr}(x'xV_{\mathcal{Q}^c}\Lambda_{\mathcal{Q}^c}^{-1}V_{\mathcal{Q}^c}')$, $C=n\sum_{j\in\mathcal{Q}^c}\lambda_j^{-2}v_jv_j'$ and $c=n\sum_{j\in\mathcal{Q}^c}\lambda_j^{-2}v_jv_j'\bar{x}$. Thus, inserting (A.3) and (A.4) in (A.2) and multiplying the likelihood with the $N_p(\mu_0,\Omega_0)$ prior we obtain

$$\begin{aligned}-\frac{1}{2}\ln p(\mu|\cdot)&\propto C_0+\mu'C\mu-2\mu'c+B_0+\mu'B\mu-2\mu'b+\mu\Omega_0^{-1}\mu-2\mu'\Omega_0^{-1}\mu_0\\&=B_0+C_0-\bar{\mu}'(B+C+\Omega_0^{-1})\bar{\mu}+(\mu-\bar{\mu})'(B+C+\Omega_0^{-1})(\mu-\bar{\mu}),\end{aligned}$$

where $\bar{\mu}=(B+C+\Omega_0^{-1})^{-1}(b+c+\mu_0)$,

$$B+C=\sum_{j\in\mathcal{Q}}\lambda_j^{-2}[n_j+\tau_j^{-2}(n-n_j)]v_jv_j'+n\sum_{j\in\mathcal{Q}^c}\lambda_j^{-2}v_jv_j'=V\bar{\Lambda}^{-1}V',$$

$\bar{\Lambda} = diag(l_1, ..., l_p)$, $l_j = n^{-1}\lambda_j^2$ for $j \in \mathcal{Q}^c$ and $l_j = [n_j + \tau_j^{-2}(n - n_j)]^{-1}\lambda_j^2$ for $j \in \mathcal{Q}^c$. Finally,

$$b + c = \sum_{j \in \mathcal{Q}} \lambda_j^{-2} v_j v_j' [\bar{x}_j n_j + \bar{x}_j^c \tau_j^{-2}(n - n_j)] + \sum_{j \in \mathcal{Q}^c} \lambda_j^{-2} v_j v_j' \sum_{i=1}^{n} x_i = \sum_{j=1}^{p} \lambda_j^{-2} v_j v_j' w_j$$

where $w_i = n\bar{x}$ for $j \in \mathcal{Q}^c$ and $w_i = n_j \bar{x}_j + (n - n_j)\tau_j^{-2}\bar{x}_j^c$ for $j \in \mathcal{Q}$.

Earlier Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se

ISSN 1402-9103 Riksbank Printing Office Stockholm 2004