THE MULTIVARIATE SPLIT NORMAL DISTRIBUTION AND ASYMMETRIC PRINCIPAL COMPONENTS ANALYSIS

Mattias Villani
Research Division, Sveriges Riksbank and
Department of Statistics, Stockholm University
SE-103 37 Stockholm, Sweden
mattias.villani@riksbank.se

Rolf Larsson
Department of Information Science, Uppsala University
Box 513, SE-751 20 Uppsala, Sweden
rolf.larsson@dis.uu.se.

Key Words: bayesian inference, statistical distribution, estimation, maximum likelihood, multivariate analysis, skewness.

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, and discusses its role as a population model for asymmetric principal components analysis. Maximum likelihood estimators and a complete Bayesian analysis, including inference on the number of skewed dimensions and their directions, are presented.

1. INTRODUCTION

A natural direction for extending the normal distribution is the introduction of some sort of skewness, and several proposals have indeed emerged, see e.g. Azzalini (1985) for an early contribution and Fernandez and Steel (1998) for a flexible family of skew distributions. Perhaps the most important member of the Fernandez-Steel family 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) and Mudhokar and Hudson (2000). 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, see e.g. Blix and Sellin (1998) and Kadane, Chan and Wolfson (1996).

Several of the univariate skewed distributions have subsequently been generalized to the multivariate setting, or are used as the core component in such a multi-dimensional extension. In particular, the distribution in Azzalini and Dalla Valle (1996) generalizes Azzalini's (1985) distribution to the multivariate setting, and Ferreira and Steel (2004) have recently proposed a general method of transforming a set of univariate distributions to a multivariate family, with special emphasis on the class of skewed distributions in Fernandez and Steel (1998). Bauwens and Laurent (2002) propose a fat tailed skewed multivariate distribution in the analysis of financial data.

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.[1]

The paper is outlined as follows. The next section gives a short review of the univariate split normal distribution, defines the multivariate split normal distribution and derives some of its properties. Statistical inference for the multivariate split normal distribution is discussed in Section 3, where both maximum likelihood estimation and a Bayesian analysis of the multivariate split normal model are presented. The proofs have been collected in an appendix.

2. THE MULTIVARIATE SPLIT NORMAL DISTRIBUTION

In order to describe the multivariate split normal distribution we start out with its univariate special case introduced in John (1982), who derived many of its properties, and further studied in Mudhokar and Hutson (2000). Moreover, with $\mu = 0$ and some reparametrization, it is a special case of the distribution studied in Fernandez and Steel (1998). We use a slightly different parametrization which is more easily generalized to the multivariate

[1] The multivariate split normal distribution is a member of the Ferreira-Steel family. Ferreira and Steel (2004) have independently derived some of the results presented here. Ferreira and Steel (2004) also treat multivariate regression analysis with skewed error terms.

case.

Definition 1 *$x \in R$ follows the univariate split normal distribution, $x \sim SN(\mu, \lambda^2, \tau^2)$, if it has density*

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

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 following result will be useful in the sequel.

Lemma 1 *If $x \sim SN(\mu, \lambda^2, \tau^2)$, then*

$$\begin{aligned} E(x) &= \mu + \sqrt{2/\pi}\lambda(\tau - 1) \\ Var(x) &= b\lambda^2 \\ \beta_1 \overset{def}{=} \frac{E[\{x - E(x)\}^3]}{[Var(x)]^{3/2}} &= b^{-3/2}\sqrt{2/\pi}(\tau-1)[(4/\pi - 1)(\tau-1)^2 + \tau] \\ \beta_2 \overset{def}{=} \frac{E[\{x - E(x)\}^4]}{[Var(x)]^2} &= b^{-2}\left\{3(1+\tau^5)/(1+\tau) - 4\pi^{-2}(1-\tau)^2\left[(3+\pi)(1+\tau^2) + 3(\pi-2)\tau\right]\right\} \\ \phi_x(t) \overset{def}{=} E(e^{tx}) &= \frac{2\lambda\left\{\exp(-\lambda^2 t^2/2)\Phi(-\lambda t) + \tau\exp(-\lambda^2\tau^2 t^2/2)\Phi(-\lambda\tau t)\right\}}{\lambda(1+\tau)\exp(\mu t)} \end{aligned}$$

where $b = \frac{\pi-2}{\pi}(\tau-1)^2 + \tau$, β_1 and β_2 are the usual measures of univariate skewness and kurtosis, and $\phi_x(t)$ is the moment generating function.

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 distributions in Geweke (1989) and one of the members of the Ferreira-Steel family (Ferreira and Steel, 2004).

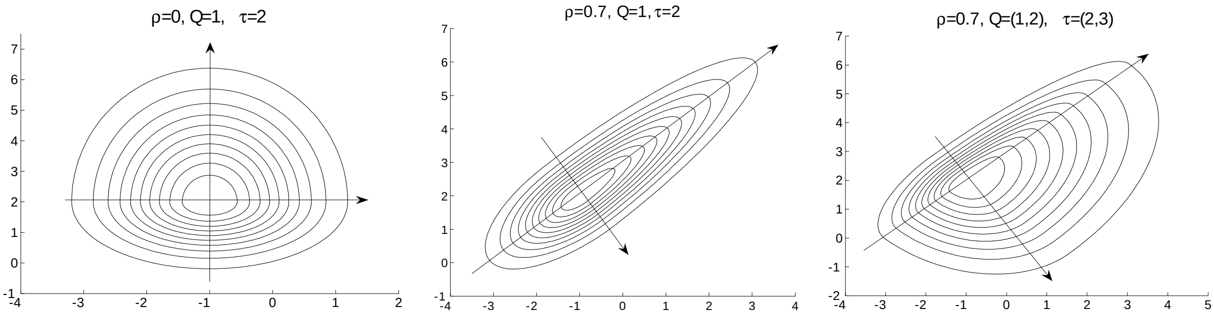

Figure 1: Contours of the bivariate split normal density with $\mu = (-1, 2)$, $\Sigma = (1, \rho; \rho, 1)$.

Definition 2 *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 jth 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

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

where $\hat{\Sigma} = V \hat{\Lambda} V'$, $\hat{\Lambda} = diag(\lambda_1^2, ..., \tau_1^2 \lambda_r^2, ..., \lambda_p^2)$ and $c^{-1} = \frac{1}{2}(2\pi)^{p/2} |\Lambda|^{1/2} (1 + \tau_1)$. This should be compared to the univariate case in Definition 1. Figure 1 illustrates three 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 $j \in \mathcal{Q}$. The normal and split normal distributions in Definition 2 may obviously be replaced by other distributions, *e.g.* the t distribution as in Geweke (1989) or, more generally, the Fernandez-Steel family (Ferreira and Steel, 2004).

Before stating some properties of the $SN_p(\mu, \Sigma, \tau, \mathcal{Q})$ distribution we need to define skewness and kurtosis for a multivariate distribution. The most widely used measures are based on the Mahalanobis distance

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

where x and z are two p-dimensional independent identically distributed random vectors with mean m and covariance matrix S. Mardia (1970) used M_{xz} to define the *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 $\beta_{1,1} = \beta_1^2$. The *multivariate kurtosis* is defined as

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

If $x \sim N_p(\mu, \Sigma)$, then $\beta_{2,p} = p(p+2)$. Note also that $\beta_{2,1} = \beta_2$. We are now ready to generalize Lemma 1 to the multivariate setting.

Theorem 2 *If $x \sim SN(\mu, \Sigma, \tau, \mathcal{Q})$ then*

$$\begin{aligned}
E(x) &= \mu + \sqrt{2/\pi} \sum_{\mathcal{Q}} \lambda_j (\tau_j - 1) v_i \\
Var(x) &= V \tilde{\Lambda} V' \\
\beta_{1,p} &= \sum_{\mathcal{Q}} b_j^{-3} (2/\pi)(\tau_j - 1)^2 [(4/\pi - 1)(\tau_j - 1)^2 + \tau_j]^2 \\
\beta_{2,p} &= p(p+2) + \sum_{\mathcal{Q}} b_j^{-2} r_j - 3q, \\
\phi_x(t) &= \left[\prod_{\mathcal{Q}} \frac{2\lambda_i \left\{ \exp[-(\lambda_j v_j' t)^2/2] \Phi(-\lambda_j v_j' t) + \tau_j \exp[-(\lambda_j \tau_j v_j' t)^2/2] \Phi(-\lambda_j \tau_j v_j' t) \right\}}{\lambda_j (1 + \tau_j) \exp(\mu_j v_j' t)} \right] \\
&\quad \times \exp\left\{ \sum_{\mathcal{Q}^c} [\mu_j v_j' t - \frac{1}{2} (v_j' t)^2 \lambda_j^2] \right\}
\end{aligned}$$

where $b_j = \frac{\pi - 2}{\pi}(\tau_j - 1)^2 + \tau_j$ for $j \in \mathcal{Q}$, $\tilde{\Lambda}$ 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 $j \in \mathcal{Q}$, $r_j = 3(1+\tau_j^5)/(1+\tau_j) - 4\pi^{-2}(1-\tau_j)^2 \left[(3+\pi)(1+\tau_j^2) + 3(\pi - 2)\tau_j\right]$ and $\phi_x(t) = E[\exp(t'x)]$ is the moment generating function for the random vector x.

3. INFERENCE

3.1 NOTATION

The following notation will be used throughout the rest of the paper. Let $x = (x_1, ..., x_n)'$ denote the $n \times p$ matrix containing a random sample from the $SN(\mu, \Sigma, \tau, \mathcal{Q})$ distribution. 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 scores for the whole sample, where ι_n is an n dimensional vector of ones. On any such demeaned principal component 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 indicies 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 $\Lambda_{\mathcal{Q}} = \mathrm{diag}(\lambda_j^2)_{j \in \mathcal{Q}}$ 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 $z_{u,j} = (z_{ij})_{i \in \mathcal{I}_j^c}$.

3.2 MAXIMUM LIKELIHOOD

Maximum likelihood estimation of the parameters in the split normal distribution is treated in John (1982) and Mudhokar and Hudson (2000). Our next result shows that it is possible to maximize the likelihood analytically w.r.t. Λ and τ in the multivariate case.

Theorem 3 *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

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

$s_{1j} = \sum_{\mathcal{I}_j} z_{ij}^2$, $s_{2j} = \sum_{\mathcal{I}_j^c} z_{ij}^2$, and the maximum likelihood estimators of λ_j^2 and τ_j are

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

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

$$V = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad -\frac{\pi}{2} < \theta \leq \frac{\pi}{2}. \tag{1}$$

A similar parametrization of V is available in the general case using generalized Eulerian angles (Khatri and Mardia, 1977). Hence, maximization over V and μ is straightforwardly performed with standard numerical optimization algorithms over a fairly low-dimensional parameter space. 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.

3.3 BAYESIAN INFERENCE

The joint posterior distribution of all parameters may be written as

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

Let us first focus on $p(\mu, V, \lambda, \tau|\mathcal{Q}, x_1, ..., x_n)$ and subsequently turn to the posterior inference of $\mathcal{Q}$.

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, $E(\tau_j^{-2}) = \gamma_j \delta_j^{-1}$.

We will take the prior density for $(\lambda_1^{-2}, ..., \lambda_p^{-2})$ to be 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.

The space of V is the oriented orthogonal group $\mathbb{O}^+(p) = \{V \in \mathbb{R}^{p\times p} : V'V = I_p$ and $v_{1j} > 0$ 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 (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* introduced by Downs (1972) and further studied by Khatri and Mardia (1977). The matrix Fisher density is of the form

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

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 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. The matrix Fisher density has mode equal to V_μ, and K_μ controls the spread around the modal point (Downs, 1972). It will often be sufficient to set $K_\mu = \operatorname{diag}(k_1, ..., k_p)$. In typical applications one would even set some of the k_i to zero (zero precision) to have an informative prior only in certain directions. Note also that since V is orthonormal, one of the columns of V is completely determined by the others and one of the k_i is therefore redundant. As an example, consider the bivariate case in (1). In this case the matrix Fisher density is proportional to $\exp[(k_1 + k_2)\cos(\theta - \mu)]$, where θ and μ are the angles of V and V_μ, respectively. Clearly one of k_1 and k_2 may be set to zero without restricting the prior.

The posterior distribution of μ, V, λ, τ conditional on $\mathcal{Q}$ is intractable. We shall sample from this distribution by iteratively sampling from the posterior distribution of each parameter conditional on the previous draw of all other parameters, the so called *full conditional posteriors*. This simulation technique is usually referred to as Gibbs sampling (Gilks et al. (1996)). The full conditional posteriors are given in the next result.

Proposition 4

- *Full conditional posterior of* λ_j

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

- *Full conditional posterior of* τ_j

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

- *Full conditional posterior of* V

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

- *Full conditional posterior of* μ

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

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 Metropolis Hastings (MH) algorithm (see e.g. Gilks et al. (1996) for a description) will be used to generate variates from these three distributions. The overall algorithm then falls within the class of Metropolis-within-Gibbs algorithms. The MH-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 attraverses $\mathbb{R}^p$ 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 $\mu_{t+1} \sim N_p[\mu_t, h(V\bar{\Lambda}^{-1}V' + \Omega_0^{-1})^{-1}]$, 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.[2] 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 (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. $q(V_{t+1}|V_t) = q(V_t|V_{t+1})$, where $q(\cdot|\cdot)$ is the proposal density. This leads to a simplified version of the Metropolis-Hastings acceptance probability, where only the target posterior density needs to be evaluated (i.e. the ratio of proposal densities cancels in the acceptance probability, see e.g. Gilks et al. (1996)). It is of course possible to rotate along several coordinate planes simultanously by postmultiplying with a product of Givens matrices.

We now turn to the posterior distribution of $\mathcal{Q}$

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

[2] Ferreira and Steel (2004) discuss a similar proposal distribution based on the Householder matrices.

where

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

is the *marginal likelihood* of the model with q skewed principal components given by $\mathcal{Q}$ and $p(\mathcal{Q})$ is the prior distribution over the set of $\mathcal{Q}$'s.

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

4. EMPIRICAL ILLUSTRATION

4.1 DATA AND MODELS

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.

A scatter plot of the raw data is displayed in all of the subgraphs in Figure 2. 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 ($\mathcal{Q} = \{2\}$). We will refer to these three models by the number of skewed principal

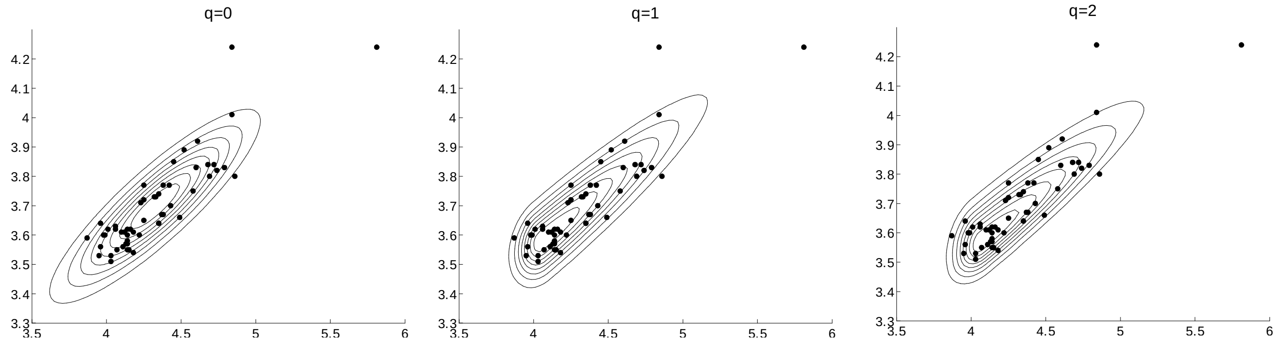

Figure 2: Data on national track records for the 1,500 meter event for women (horizontal axis) and men (vertical axis). Observations are in minutes. Equal probability contours are displayed for the three models with posterior mean estimates of the parameters. The prior with $\delta = 1$ is used.

components, q.

4.2 LIKELIHOOD ANALYSIS

The parameter estimates and maximum log likelihoods ($\log \widehat{L}_q$) of our three tentative models are shown in Table 1.

Table 1. Results from likelihood analysis.

Model	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\theta}$	$\widehat{\lambda}_1^2$	$\widehat{\lambda}_2^2$	$\widehat{\tau}_1$	$\widehat{\tau}_2$	$\log \widehat{L}_q$
$q = 0$	4.33	3.70	0.41	0.0046	0.13	–	–	48.35
$q = 1$	4.01	3.56	0.41	0.0035	0.0046	–	8.88	59.48
$q = 2$	4.02	3.52	0.38	0.0007	0.0035	3.54	8.87	65.68

The likelihood ratio test, i.e. $-2\left(\log \widehat{L}_0 - \log \widehat{L}_1\right)$, 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 $q = 2$.

4.3 BAYESIAN ANALYSIS

To compute the posterior distribution of q we will use a uniform prior on all parameters except the asymmetry parameters in τ. For the sake of presentation we will consider what 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. The second proposal distribution for μ was used. No convergence problems were encountered. The posterior distribution of q is given in Table 1 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 density of the three models with the posterior mean estimates of the parameters in each model (based on the prior with $\delta = 1$). 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 improvement in model fit from the even larger model ($q = 2$) seems to be modest.

One way to investigate the fit of the models more formally is by posterior predictive analysis (Gelman et al., 2004)[3]. The basic idea is that simulated data from a well specified model should not look too different from the actual data. An advantage of the Bayesian approach is that it is not necessary to condition on a specific parameter value when simulating the data; we may simply average across the draws from the posterior distribution. One needs of course to specify the dimensions in which the simulated data should cohere with the actual data, and one natural quantity to analyze here is the sample analogue of Mardia's multivariate skewness measure (Mardia, 1970). The observed skewness in the actual data is 7.599. The probabilities of obtaining an at least as large skewness in the simulated posterior predictive distribution, the so called posterior predictive p-value, are 0, $0.8 \cdot 10^{-4}$ and $1.2 \cdot$

[3] An alternative procedure is to embed the three models under consideration in a larger class of models and compute the posterior distribution over this model class. Such an approach has a more direct Bayesian motivation, but the posterior predictive analysis has the advantage that it can be used to detect model misspecification in specific dimensions (e.g. skewness).

10^{-4} in the models with $q = 0, 1$ and 2, respectively. This is a clear indication of model misspecification. The reason for this lack of fit is clear from Figure 2: none of the models come even close to capturing the two outliers in the north and north-east part of the figure (Cook Islands and Western Samoa). When these two observations are excluded, the posterior predictive p-value increases dramatically to 0.087, 0.423, 0.450 in the models with $q = 0, 1$ and 2, respectively. The skewness in the symmetric model is still on the borderline of acceptance, while the two skewed models now capture the skewness in the data quite well. It is interesting to note that excluding the two outliers does not have drastic effects on the inference. For example, the posterior distribution for q in the smaller data set is $0.006, 0.766$ and 0.228, for $q = 0, 1$ and 2, respectively (based on the prior with $\delta = 1$), which is in line with the results for the full data set. Other aspects of the inferences, e.g. the posterior distribution of the model parameters (see below), were also fairly robust to the exclusion of the two outliers (the main effect was on the posterior distribution of τ_1, which shifted toward a smaller skewness). We analyze the full data set in the following.

We condition the remaining analysis on $q = 1$ and the prior with $\delta = 1$. 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 square-root-inverted gamma family (Bernardo and Smith, 1994). Alternatively, one may look at the multivariate skewness in the upper right subfigure 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 2. The lower left subfigure shows that the proportion of total variance explained by the first principal component is rather close to unity.

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

Model	δ=.1	δ=.5	δ=1	δ=3	δ=5	δ=10	δ=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 3. Ranking of nations based on the principal component with largest eigenvalue.

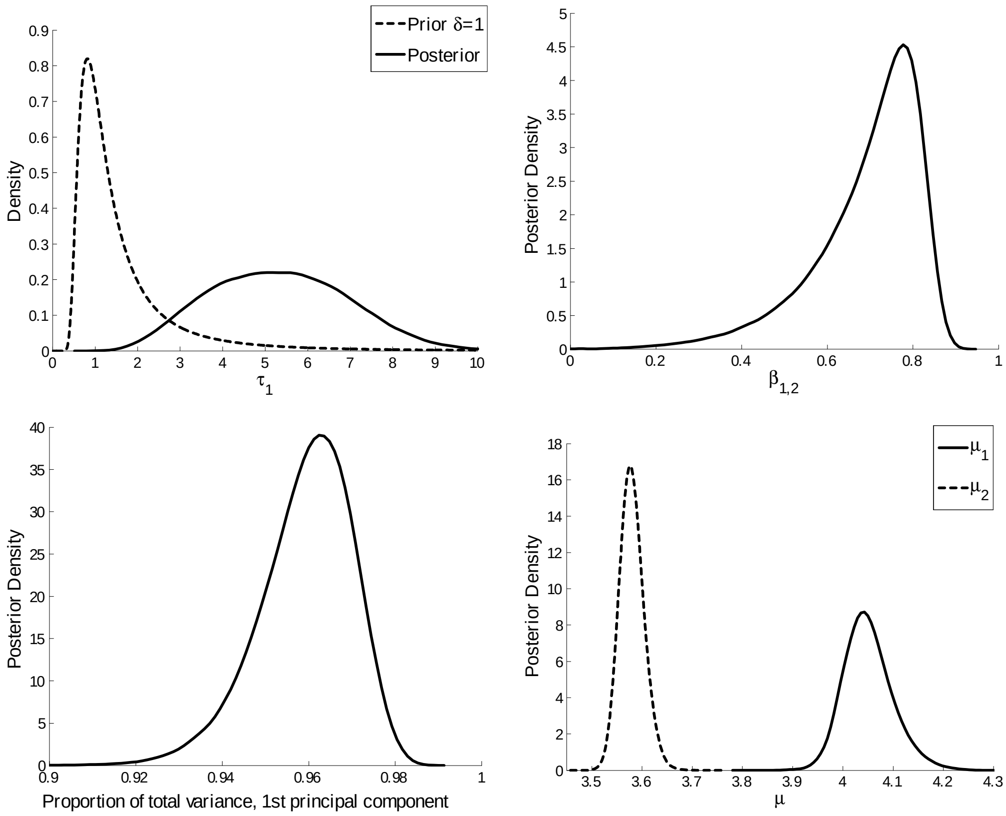

Figure 3: Posterior inferences conditional on $q = 1$. The prior with $\delta = 1$ is used.

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)

Note: 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 first principal component ($q = 1$). The last column displays the posterior distribution of the overall rank based on the first principal component ($q = 1$).

The mean acceptance probabilities for the model with $q = 1$ 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 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 the Euler angle θ is 0.777. The posterior mean of θ is 0.787, which translates into the eigenvector $v_1 = (0.706, 0.708)'$. 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 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 unambigously 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.

APPENDIX A

A.1 PROOF OF LEMMA 1

The expressions for the mean, variance and moment generating functions can be found in John (1982). The skewness and kurtosis of a univariate split normal variable are easily derived from the moments in John (1982), see Larsson and Villani (2003) for details.

A.2 PROOF OF THEOREM 2

Since $x = Vy$, where y is the vector of principal components, we have

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

and

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

using Lemma 1.

The expression for the multivariate skewness may be proved as follows. 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

$$M_{vw}^3 = \sum_{r_1+...+r_p=3} \frac{3!}{r_1! \cdots r_p!} M_{v_1 w_1}^{r_1} \cdots M_{v_p w_p}^{r_p}.$$

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

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

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 1.

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 $Var(y) = \tilde{\Lambda}$,

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

Thus,

$$E(M_{yy}^2) = \Sigma_{j=1}^p E(M_{y_j y_j}^2) + p(p-1).$$

since $E(M_{y_i y_i} M_{y_j y_j}) = E(M_{y_i y_i})E(M_{y_j y_j}) = 1$ 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 1.

The moment generating function can be written

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

where, using Lemma 1,

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

A.3 PROOF OF THEOREM 3

The Jacobian of the transformation from x to $Z = (x - \iota_n\mu')V$ is unity. Thus, by defining

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

the likelihood function can be written

$$L(\mu, \Sigma, \tau, \mathcal{Q}) = c_1^n \left\{\prod_{j\in\mathcal{Q}} \frac{1}{\lambda_j^n (1+\tau_j)^n} \exp\left[-\frac{1}{2\lambda_j^2}\left(z_{l,j}'z_{l,j} + \tau_j^{-2} z_{u,j}'z_{u,j}\right)\right]\right\}$$
$$\times \left\{\prod_{j\in\mathcal{Q}^c} \lambda_j^{-n} \exp\left(-\lambda_j^{-2}\frac{z_j'z_j}{2}\right)\right\} = \prod_{j\in\mathcal{Q}^c} L_{1,j}\left(\lambda_j\right) \prod_{j\in\mathcal{Q}} L_{2,j}\left(\lambda_j, \tau_j\right), \tag{2}$$

where and $L_{1,j}\left(\lambda_j\right)$ is the likelihood arising from observations of n independent $N\left(0, \lambda_j^2\right)$ variables and $L_{2,j}\left(\lambda_j, \tau_j\right)$ is the likelihood arising from observations of n independent univariate $SN\left(0, \lambda_j, \tau_j\right)$ variables. Hence, from the maximum likelihood estimator of the univariate split normal distribution in Mudhokar and Hudson (2000) 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

$$c_2 = \left(\frac{1}{2\pi e}\right)^{(p-q)n/2}\left(\frac{2n}{\pi e}\right)^{qn/2} = \frac{2^{(q-p/2)n} n^{qn/2}}{(\pi e)^{pn/2}},$$

and $\widehat{\lambda}_j^2(\mu, V)$, $\widehat{\tau}_j(\mu, V)$ and $g_j(\mu, V)$ are given in the theorem.

A. 4 PROOF OF PROPOSITION 4

The full conditional posteriors of λ_j and τ_j follow directly from multiplying the likelihood in (2) 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 Proposition 4.

The full conditional posterior of μ reads

$$-\frac{1}{2}\ln p(\mu|\cdot) \propto \sum_{j\in\mathcal{Q}^c} \lambda_j^{-2} z_j' z_j + \sum_{j\in\mathcal{Q}} \lambda_j^{-2} \left(\sum_{i\in\mathcal{I}_j} z_{ij}^2 + \tau_j^{-2} \sum_{i\in\mathcal{I}_j^c} z_{ij}^2 \right) \tag{3}$$

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

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

where $A_j = v_j v_j'$, and

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

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,

$$\sum_{j\in\mathcal{Q}} \lambda_j^{-2} \left(\sum_{i\in\mathcal{I}_j} z_{ij}^2 + \tau_j^{-2} \sum_{i\in\mathcal{I}_j^c} z_{ij}^2 \right) = B_0 + \mu' B \mu - 2\mu' b, \tag{4}$$

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

$$\sum_{j\in\mathcal{Q}^c} \lambda_j^{-2} z_j' z_j = C_0 + \mu' C \mu - 2\mu' c, \tag{5}$$

where $C_0 = \sum_{j\in\mathcal{Q}^c} \lambda_j^{-2} \operatorname{tr}(A_j \sum_{i=1}^n x_i x_i{}') = \operatorname{tr}(x' x V_{\mathcal{Q}^c} \Lambda_{\mathcal{Q}^c}^{-1} V_{\mathcal{Q}^c}')$, $C = n \sum_{j\in\mathcal{Q}^c} \lambda_j^{-2} v_j v_j'$ and $c = n \sum_{j\in\mathcal{Q}^c} \lambda_j^{-2} v_j v_j' \bar{x}$. Inserting (4) and (5) in (3), adding the logarithm of the normal prior density, completing the quadratic form and simplifying proves the result.

BIBLIOGRAPHY

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

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* 12:171-178.

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

Bauwens, L. and Laurent, S. (2002). A new class of multivariate skew densities, with application to GARCH models, Technical report, CORE, Université Catholique de Louvain.

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

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

Dawkins, B. (1989). Multivariate analysis of national track records. *Amer. Statist.* 43:110-115.

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

Edelman, A., Arias, T. A. and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* 20:303-353.

Fernandez, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *J. Amer. Statist. Assoc.* 93:359-371.

Ferreira, J. T. A. S. and Steel, M. F. J. (2004). Bayesian multivariate regression analysis with a new class of skewed distributions. Mimeo, University of Warwick.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis.* London: Chapman & Hall.

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

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

Gibbons, J. F. and Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions. *Appl. Phys. Lett.* 22:568-569.

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

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations* (3rd edition). London:

John Hopkins University Press.

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

John, S. (1982). The two-piece normal family of distributions and its fitting. *Comm. Statist. - Theory Methods* 11:879-885.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. I. New York: Wiley & Sons.

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

Khatri, C. G. and Mardia, K. V. (1977). The von Mises–Fisher matrix distribution in orientation statistics. *J.Roy. Statist. Soc. Ser.* B 39:95-106.

Kimber, A. C. (1985). Methods for the two-piece normal distribution. *Comm. Statist. - Theory Methods* 14:235-245.

Larsson, R. and Villani, M. (2003). Some properties of the multivariate split normal distribution. Research Report 2003:6, Department of Statistics, Stockholm University.

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

Mudhokar, G. S. and Hutson, A. (2000). The epsilon-skew normal distribution for analyzing near-normal data. *J. Statist. Plann. Inference* 83:291-309.

O'Hagan, A. (1995). Fractional Bayes factors for model comparisons (with discussion). *J.Roy. Statist. Soc. Ser.* B 57:99-138.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.