*The Annals of Statistics*
1987, Vol. 15, No. 2, 732–748

# PROPORTIONALITY OF COVARIANCE MATRICES

BY P. SVANTE ERIKSEN

*Aalborg University Centre*

The present paper considers inference for the statistical model specifying proportionality between independent Wishart distributions. It is shown that the maximum likelihood estimate of the common covariance matrix and the constants of proportionality is the unique solution to the likelihood equation and an iterative procedure determining the estimator is given. The model is an exponential transformation model, which implies the existence of an exact ancillary (a maximal invariant), and an asymptotic expansion for the distribution of the estimator conditionally on the ancillary is given. In addition, it is shown that the estimator is unbiased to order $O(n^{-3/2})$. Finally, we derive the Bartlett adjustment to the likelihood ratio statistic for the hypothesis of proportionality and subsequently for the hypothesis specifying that all of the constants of proportionality are equal to one. A small simulation study shows that the Bartlett adjustment can be very effective in improving the accuracy of chi-squared approximations.

**1. Introduction.** Suppose that $S_i$, $i = 0, \ldots, k$, are independent $p \times p$ matrices so that $S_i \sim W_p(n_i, n_i^{-1}\Sigma_i)$, i.e., $S_i$ has a $p$-dimensional Wishart distribution with $n_i$ degrees of freedom and expectation $\Sigma_i$. In the sequel we will consider the estimation and testing problem for the hypothesis of proportional covariance matrices, i.e., the hypothesis specified by

$$\Sigma_i = \lambda_i \Sigma_0, \qquad i = 1, \ldots, k. \tag{1.1}$$

My interest in this problem originates from Boldsen (1984), who wanted to improve the estimates of the covariance matrices associated with several bivariate normal populations. Actually, Boldsen ended up pooling the sample covariance matrices, as the present techniques were not available. The problem has been considered by Kim (1971), who proves that, for $k = 1$, the maximum likelihood estimator is the unique solution to the likelihood equation.

In Section 2 it will be shown that the statistical model (1.1) is an exponential transformation model, and this property will be the basis both for proving that the maximum likelihood estimator $(\hat{\Sigma}_0, \hat{\lambda}_1, \ldots, \hat{\lambda}_k)$ of $(\Sigma_0, \lambda_1, \ldots, \lambda_k)$ is the unique solution to the likelihood equation (Section 3) and for the derivation of the distribution of $(\hat{\Sigma}_0, \hat{\lambda}_1, \ldots, \hat{\lambda}_k, W)$, where $W$ is a maximal invariant affine ancillary (Section 4). Furthermore, in Section 4 we give an iterative procedure for determining the maximum likelihood estimate. In Section 5 we present an approximation to the conditional distribution of $(\hat{\Sigma}_0, \hat{\lambda}_1, \ldots, \hat{\lambda}_k)$ given $W$, and this forms the basis for Section 6, where we consider the likelihood ratio test of the hypothesis (1.1) and subsequently the hypothesis

$$\lambda_1 = \cdots = \lambda_k = 1. \tag{1.2}$$

Received August 1985; revised August 1986.

*AMS* 1980 *subject classifications*. Primary 62H15, 62H12; Secondary 62J10.

*Key words and phrases*. Ancillary, asymptotic expansion, Bartlett adjustment, conditional inference, exponential transformation model, Wishart distribution.

The Bartlett adjustment factors for these test statistics are derived, using a main result of Barndorff-Nielsen and Cox (1984). A brief simulation study reported in Section 7 shows that the effect of the Bartlett adjustment for testing the hypothesis of (1.1) can be quite dramatic. We have also found that the unadjusted likelihood ratio test performs well for the hypothesis of (1.2).

**2. Formulating the model as a transformation model.** We start by noting that the model for proportional covariance matrices is an exponential transformation model in the sense of Barndorff-Nielsen, Blæsild, Jensen and Jørgensen (1982). For that purpose, let $T_+(p)$ denote the group of upper triangular $p \times p$ matrices with positive diagonal elements, i.e., $T = \{t_{ij}\} \in T_+(p)$ if $t_{ii} > 0$, $i = 1, \dots, p$, and $t_{ij} = 0$, $1 \le j < i \le p$. Let $P(k) = \{\lambda \in \mathbb{R}^k | \lambda_i > 0, \ i = 1, \dots, k\}$ denote the multiplicative group of positive vectors, i.e., $\lambda\mu = (\lambda_1\mu_1, \dots, \lambda_k\mu_k)$, and let $\mathrm{PD}(p)$ denote the set of positive definite $p \times p$ matrices. Then $G = T_+(p) \times P(k)$ acts on $\mathrm{PD}(p)^{k+1}$ by

$$\begin{aligned} &G \times \mathrm{PD}(p)^{k+1} \to \mathrm{PD}(p)^{k+1}, \\ &((T, \lambda), (S_0, \dots, S_k)) \to (TS_0T^*, \lambda_1 TS_1T^*, \dots, \lambda_k TS_kT^*). \end{aligned} \tag{2.1}$$

The image of $S = (S_0, \dots, S_k)$ under $(T, \lambda) \in G$ will be denoted by $(T, \lambda)S$. Now if $S \sim W_p(n_0, n_0^{-1}\Sigma_0) \otimes \cdots \otimes W_p(n_k, n_k^{-1}\lambda_k\Sigma_0)$, where $\otimes$ denotes product measure, then it is clear that $(T, \alpha)S \sim W_p(n_0, n_0^{-1}\tilde{\Sigma}_0) \otimes \cdots \otimes W_p(n_k, n_k^{-1}\tilde{\lambda}_k\tilde{\Sigma}_0)$, where $\tilde{\Sigma}_0 = T\Sigma_0T^*$, $\tilde{\lambda}_i = \alpha_i\lambda_i$, $i = 1, \dots, k$, i.e., we have a transformation model.

Furthermore, $G$ acts freely [i.e., for all $S$ we have that $(T, \lambda)S = S$ implies $(T, \lambda) = e$, where $e = (I_p, (1, \dots, 1))$ is the identity element of $G$] and, consequently, we may parametrize the model by $G$. The measure

$$d\mu(S) = \prod_{i=0}^{k} |S_i|^{-(p+1)/2}\, dS_i \tag{2.2}$$

on $\mathrm{PD}(p)^{k+1}$ is invariant under the action of $G$ [see, e.g., Eaton (1983), Section 6.3], and if $f$ is the density of $S$ w.r.t. $\mu$, then we have in the parametrization given by $G$,

$$\begin{aligned} f(S; (T, \lambda)) = \omega \prod_{j=1}^{p} t_{jj}^{-n} \prod_{i=1}^{k} \lambda_i^{-n_i p/2} \prod_{i=0}^{k} |S_i|^{n_i/2} \\ \times \exp\left( -\tfrac{1}{2}\mathrm{tr}\left( T^{-1}\left( n_0 S_0 + \sum_{i=1}^{k} \lambda_i^{-1} n_i S_i \right) T^{*-1} \right) \right), \end{aligned} \tag{2.3}$$

where $n = \sum_{i=0}^{k} n_i$ and

$$\begin{aligned} \omega &= \prod_{i=0}^{k} \omega(n_i, p), \\ \omega(f, p) &= \pi^{-p(p-1)/4} (f/2)^{fp/2} \prod_{j=0}^{p-1} \Gamma((f-j)/2)^{-1} \end{aligned} \tag{2.4}$$

is the well-known norming constant.

**3. Estimation.** Since $\{(\Sigma_0, \lambda_1\Sigma_0, \ldots, \lambda_k\Sigma_0)|\lambda \in P(k), \Sigma_0 \in \mathrm{PD}(p)\}$ is a set, which is relatively closed as a subset of $\mathrm{PD}(p)^{k+1}$, it follows that the existence of the maximum likelihood estimator is ensured by Theorem 2.1 of Eriksen (1984a). Let $l((T, \lambda); S)$ denote the log-likelihood function for $(T, \lambda) \in G$ based on the observation $S$ and let $\dot{l}((T, \lambda); S)$ denote the vector of partial derivatives with respect to $(T, \lambda)$. This is well defined since $G$ is an open subset of $\mathbb{R}^{p(p+1)/2+k}$, and we may conclude that the estimator exists as a solution to the likelihood equation.

Let $\mathscr{W} = \{W \in \mathrm{PD}(p)^{k+1} | \dot{l}(e; W) = 0\}$ denote the set of observations for which $\hat{\Sigma}_0 = I_p$, $\hat{\lambda}_i = 1$, $i = 1, \ldots, k$, is a solution to the likelihood equation. If $(\hat{T}, \hat{\lambda})$ is a solution to the likelihood equation corresponding to the observation $S$, then $W = (\hat{T}, \hat{\lambda})^{-1}S \in \mathscr{W}$. Since there is at least one solution it follows that the mapping

$$(3.1) \qquad \begin{aligned} F\colon \quad & G \times \mathscr{W} \to \mathrm{PD}(p)^{k+1}, \\ & ((T, \lambda), W) \to (T, \lambda)W \end{aligned}$$

is surjective. In the sequel we will show that $F$ is injective implying that with $F^{-1}(S) = ((\hat{T}(S), \hat{\lambda}(S)), W(S))$ we have that $(\hat{T}(S), \hat{\lambda}(S))$ is the unique maximum likelihood estimator. Furthermore, $W(S)$ is clearly maximal invariant and, as we will show, $\mathscr{W}$ is an affine subset of $\mathrm{PD}(p)^{k+1}$, i.e., $W(S)$ is an affine ancillary [for a definition, see Barndorff-Nielsen (1980)].

Let us start by characterizing $\mathscr{W}$. From (2.3) it follows that if $V = T^{-1}$ and $S(\lambda) = n_0S_0 + \Sigma n_i\lambda_i^{-1}S_i$, then

$$(3.2) \quad l(S; (T, \lambda)) = -n \sum_{j=1}^{p} \log t_{jj} - \sum_{i=1}^{k} n_i p/2 \log \lambda_i - \tfrac{1}{2}\mathrm{tr}(VS(\lambda)V^*).$$

Let $E_{ij}$, $i \le j$, denote the $p \times p$ matrix with 1 on the $(i, j)$th place and zero otherwise. Then it is obvious that $\partial T/\partial t_{ij} = E_{ij}$ and since $VT = I_p$ it follows that $\partial V/\partial t_{ij} = VE_{ij}V$. From this it is easy to see that

$$(3.3) \qquad \begin{aligned} \frac{\partial l}{\partial t_{ij}} &= -n\delta_{ij}t_{jj}^{-1} + \mathrm{tr}\big(VE_{ij}VS(\lambda)V^*\big), \\ \frac{\partial l}{\partial \lambda_i} &= -n_i p/2\lambda_i^{-1} + n_i/2\lambda_i^{-2}\mathrm{tr}(VS_iV^*). \end{aligned}$$

This means that $W = (W_0, \ldots, W_k) \in \mathscr{W}$ if $\mathrm{tr}(E_{ij}\Sigma_{l=0}^k n_l W_l) = n\delta_{ij}$, $1 \le i \le j \le p$, and $\mathrm{tr}\, W_i = p$, $i = 1, \ldots, k$, which can be expressed as

$$(3.4) \qquad \sum_{i=0}^{k} n_i W_i = nI_p, \qquad \mathrm{tr}\, W_i = p, \qquad i = 0, \ldots, k.$$

Next we show that $F$ given by (3.1) is injective.

So suppose that $(T_1, \lambda_1)W = (T_2, \lambda_2)\tilde{W}$. Setting $(T, \lambda) = (T_2, \lambda_2)^{-1}(T_1, \lambda_1)$ this implies that $\tilde{W} = (T, \lambda)W$, i.e.,

$$\tilde{W}_0 = TW_0T^*, \qquad \tilde{W}_i = \lambda_i TW_iT^*, \qquad i = 1, \ldots, k.$$

We will show that this is fulfilled only if $T = I_P$, $\lambda_i = 1$, $i = 1, \ldots, k$, i.e., the mapping (3.1) is injective.

Now let $T^*T = \theta D \theta^*$, where $\theta$ is an orthogonal matrix and $D = \operatorname{diag}(d_1, \dots, d_p)$ is diagonal with $d_1 \geq \cdots \geq d_p > 0$. First suppose that $d_1 > d_p$. Then we have from (3.4) that

$$p = \operatorname{tr} \tilde{W}_i = \lambda_i \operatorname{tr}(TW_iT^*) = \lambda_i \operatorname{tr}(D\theta^* W_i \theta)$$
$$< \lambda_i d_1 \operatorname{tr}(\theta^* W_i \theta) = \lambda_i d_1 \operatorname{tr}(W_i) = \lambda_i d_1 p,$$

i.e.,

$$\lambda_i > d_1^{-1}, \qquad i = 0, \dots, k,$$

where $\lambda_0 = 1$. From this and (3.4) we obtain

$$(*) \qquad nI_p = \sum n_i \tilde{W}_i = \sum \lambda_i n_i T W_i T^* > d_1^{-1} \sum n_i T W_i T^* = d_1^{-1} n T T^*,$$

where $A > B$ means that $A - B$ is positive definite. Now, since

$$\det(\lambda I - T^*T) = \det(\lambda I - TT^*), \ \lambda \in \mathbb{R},$$

it follows that the eigenvalues of $T^*T$ and $TT^*$ are equal and consequently $TT^* = \mathscr{U} D \mathscr{U}^*$ for some orthogonal matrix $\mathscr{U}$. Combining this with $(*)$ we obtain $I_p > d_1^{-1} D$, which is not fulfilled, i.e., we must have that $d_1 = d_p = d$, so that $T = dI_p$. Since $\tilde{W}_0 = dW_0$ and $\operatorname{tr} \tilde{W}_0 = \operatorname{tr} W_0 = p$, it follows that we must have $d = 1$. Similarly, it is now easy to see that we must have $\lambda_1 = \cdots = \lambda_p = 1$.

These considerations may be summarized in a theorem.

THEOREM 3.1. *The maximum likelihood estimate of* $(\Sigma_0, \lambda)$ *exists uniquely and is given by* $(\hat{T}\hat{T}^*, \hat{\lambda})$ *where* $(\hat{T}, \hat{\lambda})$ *is the unique solution to the likelihood equation determined by* (3.3).

*Furthermore,* $S = (\hat{T}, \hat{\lambda})W$ *where* $W \in \mathscr{W} = \{(W_0, \dots, W_k) | \operatorname{tr} W_i = p, \sum_{i=0}^k n_i W_i = nI\}$ *is a maximal invariant affine ancillary.*

The likelihood equation determined by (3.3) may be rewritten as

$$(3.5) \qquad \hat{\Sigma}_0 = \alpha_0 S_0 + \sum_{i=1}^k \alpha_i \hat{\lambda}_i^{-1} S_i, \qquad \hat{\lambda}_i = \frac{1}{p} \operatorname{tr}(\hat{\Sigma}^{-1} S_i), \quad i = 1, \dots, k,$$

where $\alpha_i = n_i/n$, $i = 0, \dots, k$.

These equations suggest an iterative procedure for determining $(\hat{\Sigma}_0, \hat{\lambda})$ and in fact we have the following:

THEOREM 3.2. *Let* $\lambda_{i1} = \operatorname{tr}(S_i)/\operatorname{tr}(S_0)$, $i = 1, \dots, k$, *and*

$$(3.6) \qquad \Sigma_m = \alpha_0 S_0 + \sum_{i=1}^k \alpha_i \lambda_{im}^{-1} S_i, \qquad m = 1, 2, \dots,$$
$$\lambda_{im} = \frac{1}{p} \operatorname{tr}(\Sigma_{m-1}^{-1} S_i), \qquad i = 1, \dots, k, \ m = 2, 3, \dots.$$

*Then*

$$\lim_{m\to\infty}(\Sigma_m, \lambda_m) = (\hat{\Sigma}_0, \hat{\lambda}).$$

PROOF. Consider the likelihood function

$$L(\Sigma, \lambda) = |\Sigma|^{n/2}\prod_{i=1}^{k}\lambda_i^{n_i p/2}\exp\left(-\tfrac{1}{2}\operatorname{tr}(\Sigma^{-1}S(\lambda))\right). \tag{3.7}$$

When $\lambda \in P(k)$ is fixed, we have that $L$ is uniquely maximized by $\hat{\Sigma}(\lambda) = (1/n)S(\lambda)$. Define the map

$$\nabla_1: \mathrm{PD}(p)\times P(k) \to \mathrm{PD}(p)\times P(k),$$
$$(\Sigma, \lambda) \to (\hat{\Sigma}(\lambda), \lambda).$$

Then $\nabla_1$ is continuous and

$$L(\nabla_1(\Sigma, \lambda)) \geq L(\Sigma, \lambda),$$

with equality if and only if $\nabla_1(\Sigma, \lambda) = (\Sigma, \lambda)$. Similarly, we can define

$$\nabla_2: \mathrm{PD}(p)\times P(k) \to \mathrm{PD}(p)\times P(k),$$
$$(\Sigma, \lambda) \to (\Sigma, \hat{\lambda}(\Sigma)),$$

where $\hat{\lambda}_i(\Sigma) = (1/p)\operatorname{tr}(\Sigma^{-1}S_i)$, $i = 1, \dots, k$.

Then $\nabla_2$ is continuous and since $L$ is uniquely maximized by $\hat{\lambda}(\Sigma)$, when $\Sigma$ is fixed, we obtain

$$L(\nabla_2(\Sigma, \lambda)) \geq L(\Sigma, \lambda),$$

with equality if and only if $\nabla_2(\Sigma, \lambda) = (\Sigma, \lambda)$. It follows that $\nabla = \nabla_2\nabla_1$ is continuous and $(\Sigma_m, \lambda_m) = \nabla^{m-1}(\Sigma_1, \lambda_1)$. By the proof of Theorem 2.1 of Eriksen (1984a), we have that $\mathscr{N} = \{(\Sigma, \lambda) | L(\Sigma, \lambda) \geq L(\Sigma_1, \lambda_1)\}$ is a compact subset of $\mathrm{PD}(p)\times P(k)$. Now, since $\{(\Sigma_m, \lambda_m)\} \subseteq \mathscr{N}$ we only have to show that if $(\Sigma_{m_g}, \lambda_{m_g})$ is a convergent subsequence with limit point $(\Sigma, \lambda)$, then $(\Sigma, \lambda) = (\hat{\Sigma}_0, \hat{\lambda})$.

By the properties of $\nabla$ we have

$$L(\Sigma, \lambda) = \lim_{g\to\infty} L\left(\Sigma_{m_g}, \lambda_{m_g}\right)$$
$$\geq \lim_{g\to\infty} L\left(\nabla\left(\Sigma_{m_{g-1}}, \lambda_{m_{g-1}}\right)\right) = L(\nabla(\Sigma, \lambda)).$$

Hence,

$$L(\Sigma, \lambda) \geq L(\nabla_2\nabla_1(\Sigma, \lambda)) \geq L(\nabla_1(\Sigma, \lambda)) \geq L(\Sigma, \lambda),$$

which at first implies $\nabla_1(\Sigma, \lambda) = (\Sigma, \lambda)$ and secondly $\nabla_2(\Sigma, \lambda) = (\Sigma, \lambda)$. These equalities hold if and only if $(\Sigma, \lambda)$ is a solution to (3.5), i.e., $(\Sigma, \lambda) = (\hat{\Sigma}_0, \hat{\lambda})$. □

REMARK. The starting value of $\lambda$ may, of course, be chosen differently, but the proposed choice seems reasonable and is easily calculated.

**4. Distributional results.** In order to describe the distribution of $(\hat{T}, \hat{\lambda}, W)$ we need to determine the measure $F^{-1}(\mu)$, where $\mu$ is the invariant measure on

$\mathrm{PD}(p)^{k+1}$ given by (2.2) and $F$ is the diffeomorphism (3.1), i.e., $F((T, \lambda), W) = (T, \lambda)W$.

It is well known that [see, e.g., Barndorff-Nielsen, Blæsild, Jensen and Jørgensen (1982), Section 3]

$$F^{-1}(\mu) = \alpha \otimes \mathscr{H},$$

where $\alpha$ is a left-invariant measure on $G$ and $\mathscr{H}$ is the so-called quotient measure on $\mathscr{W}$. Since

$$d\alpha(T, \lambda) = \prod_{i=1}^{k} \lambda_i^{-1} \prod_{j=1}^{p} t_{jj}^{j-p-1}\, dT\, d\lambda$$

(where $dT = \prod_{i \le j} dt_{ij}$, $d\lambda = \prod_i d\lambda_i$) is a left-invariant measure on $G$ [see, e.g., Eaton (1983), Section 6.2], we only need to determine $\mathscr{H}$. Let $dS$ and $dW$ denote the Lebesgue measure on $\mathrm{PD}(p)^{k+1}$ and $\mathscr{W}$, respectively. Then we know that $F^{-1}(dS) = J_F\, dT\, d\lambda\, dW$, where $J_F$ is the Jacobian of $F$. Since as already stated $d\mu(S) = q(S)\, dS$, where $q(S) = \prod_{i=0}^{k} |S_i|^{-(p+1)/2}$, it follows that $dF^{-1}(\mu)(T, \lambda, W) = J_F((T, \lambda), W) q((T, \lambda), W)\, dT\, d\lambda\, dW$. Equating this to $d\alpha(T, \lambda)\, d\mathscr{H}(W)$, it follows that

$$d\mathscr{H}(W) = J_F(e, W) q(W)\, dW.$$

The evaluation of $J_F(e, W)$ is straightforward but very tedious and is for that reason deferred to the Appendix. There it is shown that the calculation of $J_F(e, W)$ can be reduced to evaluating the determinant

$$J(W) = 2^p \alpha_0^{-p(p+1)/2} \left| \left\{ \delta_{ij} p - \operatorname{tr}\left( \sqrt{\alpha_i \alpha_j}\, W_i W_j \right) \right\}_{i, j=1}^{k} \right|. \tag{4.1}$$

It should be noted that $dW = dW_1 \cdots dW_k$, where $W = (W_0, \dots, W_k)$, $W_i = \{w_{ilm}\}_{l, m=1}^{p}$ and $dW_i = \prod_{l=1}^{p-1} dw_{ill} \prod_{l<m} dw_{ilm}$. In particular, in the case of proportionality of two covariance matrices, we have

$$J(W) = 2^p \alpha_0^{-p(p+1)/2} \left( p - \alpha_1 \operatorname{tr}(W_1^2) \right).$$

All things considered we have shown

$$\begin{aligned} d\alpha(T, \lambda) &= \prod_{i=1}^{k} \lambda_i^{-1} \prod_{j=1}^{p} t_{jj}^{j-p-1}\, dT\, d\lambda, \\ d\mathscr{H}(W) &= J(W) \prod_{i=0}^{k} |W_i|^{-(p+1)/2}\, dW, \end{aligned} \tag{4.2}$$

where $J(W)$ is given by (4.1). Denoting the density of $((\hat{T}, \hat{\lambda}), W)$ w.r.t. $\alpha \otimes \mathscr{H}$ by $f$, it is clear from (2.3) that

$$\begin{aligned} &f((\hat{T}, \hat{\lambda}), W; (T, \lambda)) \\ &\quad = \omega \prod_{j=1}^{p} \left( t_{jj}^{-1} \hat{t}_{jj} \right)^{n} \prod_{i=1}^{k} \left( \lambda_i^{-1} \hat{\lambda}_i \right)^{n_i p/2} \\ &\qquad \times \prod_{i=0}^{k} |W_i|^{n_i/2} \exp\left( -n/2 \operatorname{tr}\left( T^{-1} \hat{T} \left( \alpha_0 W_0 + \sum_{i=1}^{k} \lambda_i^{-1} \hat{\lambda}_i \alpha_i W_i \right) \hat{T}^* T^{*-1} \right) \right). \end{aligned} \tag{4.3}$$

In principle it is now possible to determine the conditional distribution of $(\hat{T}, \hat{\lambda})$ given the ancillary statistic $W$, but there seems to be no closed form for the marginal distribution of $W$. This suggests that we try to approximate the relevant distributions.

**5. Asymptotic expansion for the conditional distribution of the maximum likelihood estimator.** Let $f((\hat{T}, \hat{\lambda}); (T_0, \lambda_0)|W)$ denote the density w.r.t. $\alpha$ of the conditional distribution of $(\hat{T}, \hat{\lambda})$ given $W$. Since $f((\hat{T}, \hat{\lambda}); (T_0, \lambda_0)|W) = f((T_0^{-1}\hat{T}, \lambda_0^{-1}\hat{\lambda}); e|W)$ we will concentrate on approximating $f((T, \lambda)|W) = f((T, \lambda); e|W)$. By (4.3) we have that

$$f((T, \lambda)|W) = c(W) \prod_{j=1}^{p} t_{jj}^{n} \prod_{i=1}^{k} \lambda_i^{n_i p/2} \times \exp\left(-n/2 \operatorname{tr}\left(T\left(\alpha_0 W_0 + \sum_{i=1}^{k} \lambda_i \alpha_i W_i\right) T^*\right)\right), \tag{5.1}$$

where $c(W)$ is a normalizing constant.

We will consider an approximation as $n \to \infty$ under the assumption that $\alpha_i = n_i/n$, $i = 0, \dots, k$, is fixed. Instead of $W$ we introduce $U = (U_0, \dots, U_k)$, where $W_i = I_p + n^{-1/2} U_i$, $i = 0, \dots, k$. It is then a standard result that the distribution of $U$ is asymptotically normal on the vector space $\{U | \sum_{i=0}^{k} \alpha_i U_i = 0, \operatorname{tr} U_i = 0\}$.

In order to expand (5.1) we will change from $(T, \lambda)$ to another parametrization, which eases the calculations. For that purpose let $u_1, \dots, u_p \in \mathbb{R}^{p-1}$ be determined so that

$$\begin{pmatrix} p^{-1/2} & \cdots & p^{-1/2} \\ u_1 & & u_p \end{pmatrix}$$

is an orthogonal matrix. Next, for $z \in \mathbb{R}^{p-1}$ we denote by $D(z)$ the diagonal $p \times p$ matrix

$$D(z) = \sum_{j=1}^{p} (u_j^* z) E_{jj}. \tag{5.2}$$

The assumption on $u_1, \dots, u_p$ implies that

$$\operatorname{tr}(D(z)) = 0, \qquad \operatorname{tr}\left(D(z)^2\right) = \|z\|^2. \tag{5.3}$$

Now let $M = \{m_{ij}\}$ denote a $p \times p$ matrix so that $m_{ij} = 0$, $1 \le j \le i \le p$, and let $v^* = (v_0, \dots, v_k) \in \mathbb{R}^{k+1}$.

We will then change to the parametrization $(z, v, M)$ determined by

$$T = \exp\left((2n_0 p)^{-1/2} v_0\right) \exp\left((2n)^{-1/2} D(z)\right)\left(I_p + n^{-1/2} M\right), \qquad \lambda_i = \exp\left((n_i p/2)^{-1/2} v_i - (n_0 p/2)^{-1/2} v_0\right), \quad i = 1, \dots, k, \tag{5.4}$$

where $\exp(A) = \sum_{r=0}^{\infty} A^r / r!$ is the exponential map defined on a $p \times p$ matrix.

By (5.1) and a bit of calculation it now follows that the conditional distribution of $(z, v, M)$ given $U$ has density $f(z, v, M|U)$ with respect to the Lebesgue measure $dz\,dv\,dM$ given by

$$
\begin{aligned}
f(z, v, M|U) = \tilde{c}_0(U)\exp\Bigg((np/2)^{1/2}\gamma^* v - (n/2)\sum_{i=0}^{k}\gamma_i^2\exp\big((n_i p/2)^{-1/2}v_i\big) \\
\times\Big(\operatorname{tr}\Big(\big[I_p + n^{-1/2}M^*\big]\exp\big((n/2)^{-1/2}D(z)\big) \\
\times\big[I_p + n^{-1/2}M\big]\big(I_p + n^{-1/2}U_i\big)\Big)\Big)\Bigg),
\end{aligned}
\tag{5.5}
$$

where $\gamma^* = (\alpha_0^{1/2}, \dots, \alpha_k^{1/2})$ and $\tilde{c}_0(U)$ is a normalizing constant. Expanding $\exp((n/2)^{-1/2}D(z))$ and $\exp((n_i p/2)^{-1/2}v_i)$ then yields

$$
\begin{aligned}
&f(z, v, M|U) \\
&\quad = (2\pi)^{-(k+p(p+1)/2)/2}\tilde{c}(U)\exp\big(-\tfrac{1}{2}(\|z\|^2 + \|v\|^2 + \|M\|^2)\big) \\
&\qquad \times\big[1 - (2n)^{-1/2}A_1 + \tfrac{1}{2}n^{-1}\big(A_1^2/2 - A_2\big) + O(n^{-3/2})\big],
\end{aligned}
\tag{5.6}
$$

where $\|M\|^2 = \operatorname{tr}(MM^*)$, $\tilde{c}(U)$ is a normalizing constant and

$$
\begin{aligned}
A_1 &= \tfrac{1}{3}\operatorname{tr} D(z)^3 + p^{-1/2}\gamma^* v\big(\|M\|^2 + \|z\|^2\big) + \operatorname{tr}(M^* D(z) M) \\
&\quad + \tfrac{1}{3}p^{-1/2}\sum_{i=0}^{k}\gamma_i^{-1}v_i^3 + p^{-1/2}\operatorname{tr}\big(\big[M^* + M + \sqrt{2}\,D(z)\big]X_1(v)\big), \\
X_1(v) &= \sum_{i=0}^{k}\gamma_i v_i U_i,
\end{aligned}
\tag{5.7}
$$

and

$$
\begin{aligned}
A_2 &= \tfrac{1}{6}\operatorname{tr}D(z)^4 + 2p^{-1/2}\gamma^* v\operatorname{tr}\big(M^* D(z) M + \tfrac{1}{3}D(z)^3\big) \\
&\quad + \operatorname{tr}\big(M^* D(z)^2 M\big) + p^{-1}\|v\|^2\big(\|M\|^2 + \|z\|^2\big) + \tfrac{1}{6}p^{-1}\sum\gamma_i^{-2}v_i^4 \\
&\quad + (p/2)^{-1/2}\operatorname{tr}\big(\big[M^*M + \sqrt{2}\,D(z)M + \sqrt{2}\,M^* D(z) + D(z)^2\big]X_1(v)\big) \\
&\quad - p^{-1}\operatorname{tr}\big(\big[M^* + M + \sqrt{2}\,D(z)\big]X_2(v)\big), \\
X_2(v) &= \sum_{i=0}^{k}v_i^2 U_i.
\end{aligned}
\tag{5.8}
$$

Now, denote by $\mathscr{E}(A)$ the quantity

$$
\mathscr{E}(A) = \int A(2\pi)^{-(k+p(p+1)/2)/2}\exp\big(-\tfrac{1}{2}(\|z\|^2 + \|v\|^2 + \|M\|^2)\big)\,dz\,dv\,dM.
$$

Then we have

$$(5.9)\qquad \begin{aligned} \mathscr{E}(A_1) &= 0,\\ \mathscr{E}\left(\left(\operatorname{tr} D(z)^3\right)^2\right) &= 6p + 12p^{-1} - 18,\\ \mathscr{E}\left(D(z)\operatorname{tr} D(z)^3\right) &= 0,\\ \mathscr{E}\left(u_j^* z D(z)\right) &= E_{jj} - p^{-1}I_p, \qquad j = 1,\dots,p. \end{aligned}$$

From the last three equalities and the fact that $\|\gamma\| = 1$, we may deduce

$$\begin{aligned} \mathscr{E}\left(A_1^2\right) = {} & \tfrac{1}{9}(6p + 12p^{-1} - 18) + p^{-1}\left(\tfrac{1}{4}p^2(p+1)^2 - 1\right)\\ & + p^{-1}(k+1)(p(p+1) - 2) + \tfrac{1}{12}(p^3 + 12p^2 - 25p + 12)\\ & + \tfrac{5}{3}p^{-1}\sum_{i=0}^{k}\gamma_i^{-2} + 2p^{-1}\sum_{i=0}^{k}\gamma_i^2 \operatorname{tr}\left(U_i^2\right). \end{aligned}$$

Furthermore, we have

$$\mathscr{E}\left(\operatorname{tr}\left(D(z)^4\right)\right) = 3(p + p^{-1} - 2),$$

and thereby

$$\begin{aligned} \mathscr{E}(A_2) = {} & \tfrac{1}{2}(p + p^{-1} - 2) + (1 - p^{-1})p(p-1)/2\\ & + p^{-1}(k+1)(p(p+1)/2 - 1) + \tfrac{1}{2}p^{-1}\sum_{i=0}^{k}\gamma_i^{-2}. \end{aligned}$$

We may now deduce that

$$(5.10)\qquad \tilde{c}(U)^{-1} = 1 + n^{-1}b(U) + O(n^{-3/2}),$$

where

$$(5.11)\qquad \begin{aligned} b(U) = {} & \tfrac{1}{24}(2p^3 + 3p^2 - p - 4p^{-1})\\ & + \tfrac{1}{6}p^{-1}\sum_{i=0}^{k}\gamma_i^{-2} + \tfrac{1}{2}p^{-1}\sum_{i=0}^{k}\gamma_i^2\operatorname{tr}\left(U_i^2\right). \end{aligned}$$

All things considered we have shown that

$$(5.12)\qquad \begin{aligned} f(z, v, M|U) = {} & \varphi(z, v, M)\big[1 - (2n)^{-1/2}A_1\\ & + \tfrac{1}{4}n^{-1}\left(A_1^2 - 2A_2 - 2b\right) + O(n^{-3/2})\big], \end{aligned}$$

where $\varphi$ is the standard normal density on $\mathbb{R}^{p(p+1)/2+k}$ and $A_1$, $A_2$ and $b$ are given by (5.7), (5.8) and (5.11), respectively.

Next, we will show that, conditionally on $U$, the maximum likelihood estimate $\hat{\lambda}_i\hat{T}\hat{T}^*$, $\hat{\lambda}_0 = 1$, of $\lambda_{i0}\Sigma_0$ ($\lambda_{00} = 1$) is unbiased to order $O(n^{-3/2})$, i.e.,

$$(5.13)\qquad E\left(\hat{\lambda}_i\hat{\Sigma}_0|U\right) = \lambda_{i0}\Sigma_0 + O(n^{-3/2}).$$

Since $\hat{\lambda}_i\hat{\Sigma}_0$ is distributed as $\lambda_{i0}\lambda_i T_0 T T^* T_0^*$, $\lambda_0 = 1$, it is enough to show that

$E(\lambda_i TT^*|U) = I_p + O(n^{-3/2})$. From (5.4) and expansions of $\exp((2n)^{-1/2}D(z))$ and $\exp((n_i p/2)^{-1/2}v_i)$ we obtain

$$\begin{aligned}\lambda_i TT^* = I_p &+ n^{-1/2}\left(M + M^* + \sqrt{2}D(z) + \sqrt{2}p^{-1/2}\gamma_i^{-1}v_i I_p\right)\\ &+ n^{-1}\Big[MM^* + \sqrt{2}p^{-1/2}\gamma_i^{-1}v_i(M + M^*)\\ &\quad + \sqrt{2}/2D(z)(M + M^*) + \sqrt{2}/2(M + M^*)D(z)\\ &\quad + D(z)^2 + \gamma_i^{-2}p^{-1}v_i^2 I_p\Big] + O(n^{-3/2})\end{aligned}$$

and hence

$$\mathscr{E}(\lambda_i TT^*) = I_p + n^{-1}\left[\sum_{j=1}^{p}(p-j)E_{jj} + (1-p^{-1})I_p + \gamma_i^{-2}p^{-1}I_p\right] + O(n^{-3/2}).$$

Now

$$\mathscr{E}(D(z)\operatorname{tr}(M^*D(z)M)) = \sum_{j=1}^{p}(p-j)E_{jj} - (p-1)/2I_p$$

and from this and (5.9) it follows that

$$\begin{aligned}\mathscr{E}(\lambda_i A_1 TT^*) = n^{-1/2}\Bigg[&\sqrt{2}\left(\sum_{j=1}^{p}(p-j)E_{jj} - (p-1)/2I_p\right)\\ &+ \sqrt{2}p^{-1}(p(p+1)/2 - 1)I_p + \sqrt{2}p^{-1}\gamma_i^{-2}I_p\Bigg] + O(n^{-1}).\end{aligned}$$

It is now straightforward to conclude that

$$E(\lambda_i TT^*|U) = I_p + O(n^{-3/2}),$$

i.e., we have shown that $\hat{\lambda}_i\hat{T}\hat{T}^*$ is conditionally and hence unconditionally unbiased to order $O(n^{-3/2})$.

**6. Testing.** This section deals mainly with the determination of the Bartlett adjustments to the likelihood ratio statistics associated with the hypotheses (1.1) and (1.2). The derivation of these quantities follows the directions of Barndorff-Nielsen and Cox (1984).

The likelihood ratio statistic $Q_1$ corresponding to (1.1) is given by $f(S; (\hat{T}, \hat{\lambda}))/f(I_p, \dots, I_p; e)$, where $f$ is the density in (2.3), i.e., we have

$$Q_1 = |\hat{\Sigma}_0|^{-n/2}\prod_{i=1}^{k}\hat{\lambda}_i^{-n_i p/2}\prod_{i=0}^{k}|S_i|^{n_i/2}, \tag{6.1}$$

where $\hat{\Sigma}_0 = \hat{T}\hat{T}^*$.

Similarly, we have for the hypothesis (1.2) that

$$Q_2 = f\big(S; (\hat{T}_0, \lambda_0)\big)/f(S; (\hat{T}, \hat{\lambda})),$$

where $\lambda_0 = (1, \dots, 1)$ and $\hat{T}_0\hat{T}_0^* = \sum_{i=0}^{k}\alpha_i S_i$ is the maximum likelihood estimate of $\Sigma_0$ under (1.2). It follows that

$$Q_2 = \left| \sum_{i=0}^{k} \alpha_i S_i \right|^{-n/2} |\hat{\Sigma}_0|^{n/2} \prod_{i=1}^{k} \hat{\lambda}_i^{n_i p/2}. \tag{6.2}$$

In the sequel we will show the following:

THEOREM 6.1. (i) *The Bartlett adjusted likelihood ratio statistic associated with testing the hypothesis* $\Sigma_i = \lambda_i \Sigma_0$, $i = 1, \dots, k$, *is given by*

$$t_1 = B_1^{-1} n \left( \log|\hat{\Sigma}_0| + \sum_{i=1}^{k} \alpha_i p \log \hat{\lambda}_i - \sum_{i=0}^{k} \alpha_i \log|S_i| \right),$$

*where* $(\hat{\Sigma}_0, \hat{\lambda})$ *is determined by Theorem* 3.2 *and*

$$B_1^{k(p(p+1)/2-1)/2} = h(n) g(np, 0)^{-1} \prod_{i=0}^{k} h(n_i)^{-1} g(n_i p, 0),$$

$$g(n, j) = \sqrt{2\pi} \left( \frac{n}{2} \right)^{(n-j-1)/2} \Gamma\left( \frac{n-j}{2} \right)^{-1}, \tag{6.3}$$

$$h(n) = \prod_{j=0}^{p-1} g(n, j).$$

*The distribution of* $t_1$ *is, to order* $O(n^{-3/2})$, *the chi-squared distribution with* $k(p(p+1)/2 - 1)$ *degrees of freedom.*

(ii) *The Bartlett adjusted likelihood ratio statistic associated with testing the hypothesis* $\lambda_1 = \cdots = \lambda_k = 1$ *under* $\Sigma_i = \lambda_i \Sigma_0$, $i = 1, \dots, k$, *is given by*

$$t_2 = B_2^{-1} n \left( \log\left| \sum_{i=0}^{k} \alpha_i S_i \right| - \log|\hat{\Sigma}_0| - \sum_{i=1}^{k} \alpha_i p \log \hat{\lambda}_i \right),$$

*where*

$$B_2^{k/2} = g(np, 0) \prod_{i=0}^{k} g(n_i p, 0)^{-1}.$$

*The distribution of* $t_2$ *is, to order* $O(n^{-3/2})$, *the chi-squared distribution with k degrees of freedom.*

REMARK. A set of asymptotically equivalent adjustments is obtained by applying the approximation

$$\Gamma(\alpha) = \sqrt{2\pi}\, \alpha^{\alpha - 1/2} \exp(-\alpha) \left( 1 + (12\alpha)^{-1} + O(\alpha^{-2}) \right),$$

which, in particular, yields $B_2 = \tilde{B}_2 + O(n^{-2})$, where

$$\tilde{B}_2 = 1 + (3knp)^{-1} \left( \sum_{i=0}^{k} \alpha_i^{-1} - 1 \right). \tag{6.4}$$

In the case $p = 1$, this is the well-known Bartlett adjustment for the testing of homogeneity of variances.

The method to derive the Bartlett adjustment developed by Barndorff-Nielsen and Cox (1984) is based on the formula for the density of the maximum likelihood estimator $\hat{\theta}$ conditionally on an ancillary statistic $w$, namely

$$p(\hat{\theta};\theta|w) = c|\hat{j}|^{1/2}\frac{L(\theta)}{L(\hat{\theta})}. \tag{6.5}$$

Here $L$ is the likelihood function, $\hat{j} = j(\hat{\theta})$ is minus the matrix of second-order derivatives of $\log L(\theta)$ evaluated at $\hat{\theta}$ and $c$ is a normalizing constant. In general, the validity of (6.5) is to order $O(n^{-3/2})$, but in case of a transformation model the formula is exact.

In order to derive the Bartlett adjustments, we have to determine the normalizing constant in (6.5), when the formula is applied to the full model and to the models given by (1.1) and (1.2). All three models are transformation models, and if $e$ is the identity element of the group generating the model, it follows that

$$c(w) = p(e;e|w)|j(e,w)|^{-1/2}. \tag{6.6}$$

Let us start by considering the full model, where the ancillary is degenerate. Since the model is a product of independent Wishart distributions, the normalizing constant $c_0$ is the product of the constraints associated with each Wishart distribution. So let $p(S,\Sigma)$ denote the density of $S \sim W_p(f^{-1}\Sigma, f)$. Then we have

$$p(I_p, I_p) = e^{-fp/2}\left(\frac{f}{2}\right)^{fp/2}\pi^{-p(p-1)/4}\prod_{j=0}^{p-1}\Gamma\left(\frac{f-j}{2}\right)^{-1}.$$

Furthermore, $j(I_p)^{-1}$ is the covariance matrix of $S$ when $\Sigma = I_p$. All quantities in $S$ are uncorrelated, the diagonal elements have variance $2f^{-1}$ and the off-diagonal elements have variance $f^{-1}$, i.e.,

$$|j(I_p)|^{-1/2} = 2^{p/2}f^{-p(p+1)/4}.$$

We may conclude that the norming constant of a Wishart distribution is

$$c(f) = (2\pi)^{-p(p+1)/4}e^{-fp/2}h(f), \tag{6.7}$$

and hence

$$c_0 = (2\pi)^{-(k+1)p(p+1)/4}e^{-np/2}\prod_{i=0}^{k}h(n_i). \tag{6.8}$$

Next, consider the density (5.1) of $(\hat{T},\hat{\lambda})$ conditionally on $W$. In Eriksen (1984b) it is shown that the normalizing constant $c(W)$ is to the order $O(n^{-3/2})$ independent of $W$, i.e., we may determine the constant $c_1$ associated with (1.1) by

$$c_1 = f(I_p, \lambda_0|W_{00})|j(I_p,\lambda_0,W_{00})|^{-1/2},$$

where $W_{00} = (I_p,\dots,I_p)$ and $\lambda_0 = (1,\dots,1)$. A bit of calculation shows that

$$f(I_p,\lambda_0|W_{00}) = \omega(n,p)2^p e^{-np/2}\Gamma(np/2)\prod_{i=0}^{k}\alpha_i^{n_i p/2}\Gamma(n_i p/2)^{-1}$$

and in the Appendix it is shown that

$$\left|j(I_p, \lambda_0, W_{00})\right| = 2^p n^{p(p+1)/2}(np/2)^k \prod_{i=0}^{k} \alpha_i. \tag{6.9}$$

We may conclude that

$$c_1 = (2\pi)^{-p(p+1)/4-k/2} e^{-np/2} h(n) g(np,0)^{-1} \prod_{i=0}^{k} g(n_i p, 0). \tag{6.10}$$

Finally, we have to consider the conditional distribution of $\hat{T}_0$ given $(\hat{\lambda}, W)$, where $\hat{T}_0 \hat{T}_0^*$ is the estimate of $\Sigma_0$ under (1.2). It is fairly easy to see that under (1.2) we have that the density (2.3) evaluated at $S(\hat{T}_0, \hat{\lambda}, W)$ is the product of a factor depending on $\hat{T}_0$ only and a factor depending on $(\hat{\lambda}, W)$ only. Since $(\hat{\lambda}, W)$ is invariant under the group generating (1.2) we may conclude that $\hat{T}_0$ and $(\hat{\lambda}, W)$ are independent. This means that the normalizing constant $c_2$ associated with (1.2) is determined by the distribution of $\hat{T}_0 \hat{T}_0^*$, which is $W_p((1/n)\Sigma_0, n)$, i.e., we have by (6.7) that

$$c_2 = (2\pi)^{-p(p+1)/4} e^{-np/2} h(n).$$

The Bartlett adjustment $B_1$ to (6.1) is given by $(2\pi)^{-1}(c_1 c_0^{-1})^{2/(kp(p+1)/2-k)}$ and from (6.8) and (6.10) we obtain

$$B_1^{k(p(p+1)/2-1)/2} = h(n) g(np,0)^{-1} \prod_{i=0}^{k} h(n_i)^{-1} g(n_i p, 0).$$

Similarly, the Bartlett adjustment $B_2$ to (6.2) is given by $(2\pi)^{-1}(c_2 c_1^{-1})^{2/k}$, i.e.,

$$B_2^{k/2} = g(np,0) \prod_{i=0}^{k} g(n_i p, 0)^{-1}.$$

This finishes the proof of Theorem 6.1. □

**7. Simulation results.** It turns out that both $B_2$ and $\tilde{B}_2$ are fairly close to one when $p \geq 2$, so that the gain in accuracy of the approximation is modest. In fact, the unadjusted test statistic performs satisfactorily, even for small sample sizes.

It is apparent from Table 1, that it is important to adjust when testing proportionality of the covariance matrices. It seems that only if *all* of the $n_i$-values are "large," it is of no crucial importance to make the adjustment.

**8. Concluding remarks.** The hypothesis of proportional covariance matrices may also be formulated as

$$\Sigma_i = \sigma_i^2 \Sigma, \qquad i = 0, \ldots, k,$$

where $\Sigma$ has determinant one. In this formulation we have that the maximum likelihood estimate of $\sigma_i^2$ is given by

$$\hat{\sigma}_i^2 = \exp\left((n_i p/2)^{-1/2} \hat{v}_i\right), \qquad i = 0, \ldots, k,$$

where $\hat{v}$ is determined from the representation (5.4) of $(\hat{T}, \hat{\lambda})$.

TABLE 1
*Simulated values of* $100P(-2(\text{adj. fac.})^{-1}\log Q_1 \geq \chi^2(d)_{1-\alpha})$, *where* $\chi^2(d)_{1-\alpha}$ *is the* $(1-\alpha)$*-fractile of the chi-squared distribution with d degrees of freedom.*
*Based on* 5000 *samples*

| Type of adjustment | Adj. fac. | α 1% | α 5% | α 10% | Adj. fac. | α 1% | α 5% | α 10% |
|---|---|---|---|---|---|---|---|---|
| $p = 2$ | $n_i$-values: | 2 2 2 2 2 | | $d = 8$ | $n_i$-values: | 2 2 3 3 3 | | $d = 8$ |
| $B_1$ | 2.3166 | 0.68 | 3.50 | 7.68 | 1.8789 | 0.86 | 4.60 | 9.42 |
| None | 1 | 32.90 | 54.08 | 64.52 | 1 | 21.20 | 40.52 | 51.20 |
| $p = 2$ | $n_i$-values: | 3 3 3 3 3 | | $d = 8$ | $n_i$-values: | 3 3 4 4 5 | | $d = 8$ |
| $B_1$ | 1.6316 | 1.02 | 4.72 | 10.18 | 1.4753 | 0.90 | 4.80 | 9.68 |
| None | 1 | 13.56 | 29.56 | 41.00 | 1 | 8.92 | 22.02 | 32.54 |
| $p = 2$ | $n_i$-values: | 2 2 40 40 | | $d = 6$[a] | $n_i$-values: | 9 9 9 9 9 | | $d = 8$ |
| $B_1$ | 1.6080 | 1.40 | 6.36 | 11.28 | 1.1521 | 0.94 | 4.76 | 9.88 |
| None | 1 | 11.80 | 27.36 | 37.36 | 1 | 2.20 | 9.54 | 17.42 |
| $p = 3$ | $n_i$-values: | 3 3 3 | | $d = 10$[b] | $n_i$-values: | 4 4 4 | | $d = 10$[b] |
| $B_1$ | 2.3696 | 1.67 | 6.17 | 10.87 | 1.7513 | 1.10 | 4.95 | 9.82 |
| None | 1 | 48.50 | 67.37 | 76.20 | 1 | 21.02 | 40.05 | 52.67 |
| $p = 3$ | $n_i$-values: | 7 7 7 | | $d = 10$[b] | $n_i$-values: | 4 6 8 | | $d = 10$[b] |
| $B_1$ | 1.3223 | 0.85 | 4.62 | 9.75 | 1.4649 | 1.20 | 5.77 | 10.95 |
| None | 1 | 5.92 | 17.37 | 26.83 | 1 | 11.55 | 27.07 | 38.33 |
| $p = 5$ | $n_i$-values: | 5 5 | | $d = 14$[a] | $n_i$-values: | 10 10 | | $d = 14$[a] |
| $B_1$ | 2.5499 | 3.76 | 12.00 | 19.56 | 1.4080 | 1.20 | 5.20 | 10.88 |
| None | 1 | 73.04 | 85.92 | 90.68 | 1 | 11.84 | 29.00 | 40.12 |

[a] Based on 2500 samples.
[b] Based on 4000 samples.

The likelihood ratio statistic (6.2) for testing $\sigma_0^2 = \cdots = \sigma_k^2$ is then given by

$$\begin{aligned} Q_2 &= \left| \sum_{i=0}^{k} \alpha_i \hat{\sigma}_i^2 W_i \right|^{-n/2} \prod_{i=0}^{k} \left(\hat{\sigma}_i^2\right)^{n_i p/2} \\ &= \left| \sum_{i=0}^{k} \alpha_i \hat{\sigma}_i^2 I_p + n^{-1/2} \sum_{i=0}^{k} \alpha_i \hat{\sigma}_i^2 U_i \right|^{-n/2} \prod_{i=0}^{k} \left(\hat{\sigma}_i^2\right)^{n_i p/2} \\ &= \left( \sum_{i=0}^{k} \alpha_i \hat{\sigma}_i^2 \right)^{-np/2} \prod_{i=0}^{k} \left(\hat{\sigma}_i^2\right)^{n_i p/2} \left(1 + O_p(n^{-1})\right). \end{aligned}$$

This serves to explain the fact that the Bartlett adjustment (6.4) associated with $Q_2$ is the same as the adjustment associated with the comparison of $(k+1)$ variances, where the $i$th observed variance has $n_i p$ degrees of freedom.

Another point of interest might be the conditional density of $\hat{\lambda}$ given $W$. If $\lambda_{i0} = 1$, $i = 1, \dots, k$, we have, as already stated, that $(\hat{\lambda}, W)$ and $\hat{\Sigma}_0 = \sum_{i=0}^{k} \alpha_i S_i$ are independent. This was seen by transforming $((\hat{\lambda}, \hat{T}), W)$ to $((\hat{\lambda}, W), \hat{T}_0)$ where $\hat{T}_0 \hat{T}_0^* = \hat{\Sigma}_0 = \hat{T}(\alpha_0 W_0 + \sum_{i=1}^{k} \alpha_i \hat{\lambda}_i W_i)\hat{T}^*$. Denoting this transformation by $H$, the only problem left is to construct the measure $\rho$ on $P(k) \times \mathscr{W}$ determined by

$$H(\alpha \otimes \mathscr{H}) = \rho \otimes \beta,$$

where $\alpha \otimes \mathscr{H}$ is given by (4.2) and $\beta$ is the left-invariant measure on $T_+(p)$. This may be done in the same way as for the construction of $\mathscr{H}$, i.e., essentially by calculating the Jacobian of $H^{-1}$ (or $H$) evaluated at $((\hat{\lambda}, W), I_p)$.

## APPENDIX

**A.1. Determination of the Jacobian of (3.1).** Let $(S_0, \dots, S_k) = F((T, \lambda), W)$ so that $S_0 = TW_0T^*$ and $S_i = \lambda_i TW_iT^*$, $i = 1, \dots, k$. Since $\sum_{i=0}^k \alpha_i W_i = I_p$ we have that $S_0 = \alpha_0^{-1}T(I_p - \sum_{i=1}^k \alpha_i W_i)T^*$. Let $W_i = \{w_{ilm}\}_{l,m=1}^p$, $i = 1, \dots, k$. The identity $\operatorname{tr} W_i = p$ implies that $w_{ipp} = p - \sum_{l=1}^{p-1} w_{ill}$. In the sequel we determine the derivatives of $F$, all of which are evaluated at $(T, \lambda) = e$.

$$\partial S_i/\partial \lambda_j = \delta_{ij} W_i, \qquad i = 0, \dots, k,\ j = 1, \dots, k,$$

$$\partial S_i/\partial t_{lm} = E_{lm} W_i + W_i E_{lm}^*, \qquad i = 0, \dots, k, 1 \le l \le m \le p,$$

$$\partial S_0/\partial w_{ilm} = -\alpha_0^{-1}\alpha_i E_{lm}, \qquad i = 1, \dots, k, 1 \le l < m \le p,$$

$$\partial S_0/\partial w_{ill} = \alpha_0^{-1}\alpha_i (E_{pp} - E_{ll}), \qquad i = 1, \dots, k,\ l = 1, \dots, p-1,$$

$$\partial S_j/\partial w_{ilm} = \delta_{ij} E_{lm}, \qquad i, j = 1, \dots, k, 1 \le l < m \le p,$$

$$\partial S_j/\partial w_{ill} = \delta_{ij}(E_{ll} - E_{pp}), \qquad i, j = 1, \dots, k,\ l = 1, \dots, p-1.$$

The determinant of the Jacobi matrix is not altered by adding a constant times one row (column) to another row (column).

Now multiplying the rows corresponding to $S_0 = \{s_{0ij}\}_{1 \le i \le j \le p}$ by $\alpha_0$ gives a factor to the Jacobian, which is equal to $\alpha_0^{-p(p+1)/2}$. Next, adding the rows corresponding to $S_j = \{s_{jlm}\}_{1 \le l \le m \le p}$, $j = 1, \dots, k$, multiplied by $\alpha_j$ to the rows of $S_0$ gives that

$$\partial S_0/\partial \lambda_j \approx \alpha_j W_j, \qquad j = 1, \dots, k,$$

$$\partial S_0/\partial w_{ilm} \approx 0, \qquad i = 1, \dots, k, 1 \le l < m \le p, 1 \le l = m < p,$$

$$\partial S_0/\partial t_{lm} \approx E_{lm} + E_{lm}^*, \qquad 1 \le l \le m \le p,$$

where, e.g., $\partial S_0/\partial w_{ilm} \approx 0$ means that currently the rows corresponding to $S_0$ have a zero on the place corresponding to $w_{ilm}$.

Now, the column of $w_{ilm}$ has 1 on the place corresponding to $s_{ilm}$ and zero otherwise, i.e., we may remove the row–column "passing through $(s_{ilm}, w_{ilm})$," $i = 1, \dots, k$, $1 \le l < m \le p$.

Next, we add the row of $s_{jll}$, $l = 1, \dots, p-1$, to the row of $s_{jpp}$ and we obtain

$$\delta s_{jpp}/\partial \lambda_i \approx \delta_{ij} p, \qquad i, j = 1, \dots, k,$$

$$\partial s_{jpp}/\partial t_{lm} \approx 2w_{jlm}, \qquad j = 1, \dots, k, 1 \le l \le m \le p,$$

$$\partial s_{jpp}/\partial w_{ill} \approx 0, \qquad i, j = 1, \dots, k,\ l = 1, \dots, p-1.$$

It follows that we may remove the row–column corresponding to $(s_{jll}, w_{jll})$, $j = 1, \dots, k$, $l = 1, \dots, p-1$. Now, dividing the column of $t_{ii}$, $i = 1, \dots, p$, by 2 gives the factor $2^p$ to the determinant. Multiplying the row of $s_{jpp}$, $j = 1, \dots, k$, by $\alpha_j^{1/2}$ and dividing the column of $\lambda_j$, $j = 1, \dots, k$, by $\alpha_j^{1/2}$ does not alter the determinant and we have

$$\text{(A.1)}\quad \begin{aligned} &\partial s_{0ij}/\partial t_{lm} \approx \delta_{il}\delta_{jm}, && 1 \le i \le j \le p, 1 \le l \le m \le p,\\ &\partial s_{0ij}/\partial \lambda_m \approx \alpha_m^{1/2} w_{mij}, && 1 \le i \le j \le p,\ m = 1, \dots, k,\\ &\partial s_{jpp}/\partial t_{lm} \approx \alpha_j^{1/2}(2 - \delta_{lm}) w_{jlm}, && 1 \le l \le m \le p,\ j = 1, \dots, k,\\ &\partial s_{jpp}/\partial \lambda_i \approx \delta_{ij} p, && i, j = 1, \dots, k. \end{aligned}$$

Now, multiplying the row of $t_{lm}$ by $\alpha_j^{1/2} w_{jlm}$ and subtracting this from the row of $\lambda_j$, $j = 1, \dots, k$, $1 \le l \le m \le p$, gives

$$\begin{aligned} &\partial s_{0ij}/\partial t_{lm} \approx \delta_{il}\delta_{jm}, && 1 \le i \le j \le p, 1 \le l \le m \le p,\\ &\partial s_{0ij}/\partial \lambda_m \approx 0, && 1 \le i \le j \le p,\ m = 1, \dots, k,\\ &\partial s_{jpp}/\partial t_{lm} \approx \alpha_j^{1/2}(2 - \delta_{lm}) w_{jlm}, && 1 \le l \le m \le p,\ j = 1, \dots, k,\\ &\partial s_{jpp}/\partial \lambda_i \approx \delta_{ij} p - (\alpha_i \alpha_j)^{1/2} \operatorname{tr}(W_i W_j), && i, j = 1, \dots, k. \end{aligned}$$

From this it is clear that we may remove the row–column corresponding to $(s_{0ij}, t_{ij})$, $1 \le i \le j \le p$.

In other words, we have shown that the Jacobian of $F$ at $(e, W)$ is given by

$$J(W) = \alpha_0^{-p(p+1)/2} 2^p \left| \left\{ \delta_{ij} p - \sqrt{\alpha_i \alpha_j} \operatorname{tr}(W_i W_j) \right\}_{i,j=1}^{k} \right|.$$

**A.2. Proof of (6.9).** We start by evaluating the determinant of the observed information in the parametrization $(T, \lambda)$, i.e., we will determine minus the second-order derivatives of $l((T, \lambda); W)$, all of which are evaluated at $(T, \lambda) = e$. From (3.3) we obtain

$$\begin{aligned} &-\partial^2 l/\partial t_{ij}^2 = (1 + \delta_{ij}) n, && 1 \le i \le j \le p,\\ &-\partial^2 l/\partial t_{ij}\,\partial t_{rs} = 0, && 1 \le i \le j \le p, 1 \le r \le s \le p, (i, j) \ne (r, s),\\ &-\partial^2 l/\partial \lambda_i\,\partial t_{rs} = n_i w_{irs}, && i = 1, \dots, k, 1 \le r \le s \le p,\\ &-\partial^2 l/\partial \lambda_i\,\partial \lambda_j = \delta_{ij} n_i p/2, && i, j = 1, \dots, k. \end{aligned}$$

Now, multiplying the row of $\lambda_i$ by $2\alpha_i^{-1/2}$, $i = 1, \dots, k$, the column of $\lambda_j$ by $n^{-1}\alpha_j^{-1/2}$, $j = 1, \dots, k$, and dividing the column of $t_{ij}$ by $(1 + \delta_{ij})n$, $1 \le i \le j \le p$, gives the factor $2^{p-k} n^{p(p+1)/2+k} \prod_{i=1}^{k} \alpha_i$ to the determinant. At this stage we have

$$\begin{aligned} &-\partial^2 l/\partial t_{ij}\,\partial t_{rs} \approx \delta_{ir}\delta_{js}, && 1 \le i \le j \le p, 1 \le r \le s \le p,\\ &-\partial^2 l/\partial t_{rs}\,\partial \lambda_i \approx \alpha_i^{1/2} w_{irs}, && i = 1, \dots, k, 1 \le r \le s \le p,\\ &-\partial^2 l/\partial \lambda_i\,\partial t_{rs} \approx \alpha_i^{1/2}(2 - \delta_{rs}) w_{irs}, && i = 1, \dots, k, 1 \le r \le s \le p,\\ &-\partial^2 l/\partial \lambda_i\,\partial \lambda_j \approx \delta_{ij} p, && i, j = 1, \dots, k, \end{aligned}$$

which is the same system as (A.1), i.e.,

$$|\hat{j}(e, W)| = 2^{p-k} n^{p(p+1)/2+k} \prod_{m=1}^{k} \alpha_m \left| \left\{ \delta_{ij} p - \sqrt{\alpha_i \alpha_j} \operatorname{tr}(W_i W_j) \right\}_{i,j=1}^{k} \right|.$$

Let $\gamma^* = (\alpha_1^{1/2}, \dots, \alpha_k^{1/2})$ and suppose $W = W_{00} = (I_p, \dots, I_p)$. Then

$$\begin{aligned} \left| \left\{ \delta_{ij} p - \sqrt{\alpha_i \alpha_j} \operatorname{tr}(W_i W_j) \right\}_{i,j=1}^{k} \right| &= |p(I_k - \gamma\gamma^*)| \\ &= p^k(1 - \|\gamma\|^2) = p^k \alpha_0, \end{aligned}$$

i.e.,

$$|\hat{j}(e, W_{00})| = 2^p n^{p(p+1)/2} (np/2)^k \prod_{m=0}^{k} \alpha_m,$$

which finishes the proof of (6.9).

**Acknowledgments.** Thanks to the referees for useful comments and proposing a simulation study, which led to considering the present version of the Bartlett adjustment.

## REFERENCES

BARNDORFF-NIELSEN, O. E. (1980). Conditionality resolutions. *Biometrika* **67** 293–310.

BARNDORFF-NIELSEN, O. E., BLÆSILD, P., JENSEN, J. L. and JØRGENSEN, B. (1982). Exponential transformation models. *Proc. Roy. Soc. London Ser. A* **379** 41–65.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B* **46** 483–495.

BOLDSEN, J. (1984). A statistical evaluation of the basis for predicting stature from lengths of long bones in European populations. *Amer. J. Phys. Anthrop.* **65** 305–311.

EATON, M. L. (1983). *Multivariate Statistics*. Wiley, Chichester.

ERIKSEN, P. S. (1984a). $(k, 1)$ exponential transformation models. *Scand. J. Statist.* **11** 129–145.

ERIKSEN, P. S. (1984b). Proportionality of covariance matrices. Research Report 116, Dept. of Theoretical Statistics, Univ. of Aarhus.

KIM, D. Y. (1971). Statistical inference for constants of proportionality between covariance matrices. Technical Report 59, Dept. of Statistics, Stanford Univ.

INSTITUTE OF ELECTRONIC SYSTEMS
AALBORG UNIVERSITY CENTRE
STRANDVEJEN 19
DK-9000 AALBORG
DENMARK