MAXIMUM LIKELIHOOD ESTIMATION FOR MULTIVARIATE NORMAL SAMPLES

H.F. Strydom[1]
Department of Statistics, University of Pretoria, Pretoria, 0002, South Africa
E-mail: *nina.strydom@up.ac.za*
and
N.A.S. Crowther
Department of Statistics, University of Pretoria, Pretoria, 0002, South Africa
E-mail: *nico.crowther@up.ac.za*

Key words: Behrens-Fisher; canonical statistic; common mean; exponential class; maximum likelihood estimation; multivariate normal samples; parameter structures; proportional covariance matrices; Toeplitz correlation structure.

Summary: Maximum likelihood estimation of parameter structures in the case of multivariate normal samples is considered. The procedure provides a new statistical methodology for maximum likelihood estimation which does not require derivation and solution of the likelihood equations. It is a flexible procedure for the analysis of specific structures in mean vectors and covariance matrices – including the case where the sample size is small relative to the dimension of the observations. Special cases include different variations of the Behrens-Fisher problem, proportional covariance matrices and proportional mean vectors. Specific structures are illustrated with real data examples.

1. Introduction

Estimation and testing procedures in the case of several multivariate normal populations with unknown and possibly unequal covariance matrices have been studied extensively in statistical literature. Maximum likelihood estimation of parameters and parametric functions in this case is often difficult to perform

[1] Corresponding author
AMS: 62H12, 62H15, 62H10

or often implies that the likelihood function has to be solved numerically. In addition to that, the likelihood equations may have several roots. Estimation of the common mean vector of several multivariate normal populations (Rukhin, Biggerstaff andVangel, 2000), the multivariate Behrens-Fisher problem (Buot, Hosten and Richards, 2007) and estimation of proportional covariance matrices (Jensen and Madsen, 2004) represent such examples.

The MIXED procedure in SAS is used widely to model data which exhibit correlation and nonconstant variability. However, "experience shows that both the ideas behind the techniques and their software implementation are not at all straightforward, and users from applied backgrounds often encounter difficulties in using the methodology effectively" (Verbeke and Molenberghs, 2009). The primary assumptions underlying the analyses performed by PROC MIXED (SAS, 1999) are as follows:

- The data are normally distributed.
- The means (expected values) of the data are linear in terms of a certain set of parameters.
- The variances and covariances of the data are in terms of a different set of parameters, and they exhibit a structure matching one of those available in PROC MIXED.

PROC MIXED does not allow for structural relations between means, variances, and covariances of different groups. Covariance structures are limited. It can not be used when only the sufficient statistics are available. Modelling multivariate observations becomes very cumbersome.

In this paper, a simple unified approach through a maximum likelihood procedure under constraints, where derivation of the likelihood equations is not required, is proposed as a convenient method to obtain the maximum likelihood estimates (mles) in problematic situations such as those mentioned above – also in the case where the sample size is small relative to the dimension of the observations. It provides a method for the direct modelling of the mean vector(s) and covariance structure(s) of random variables belonging to the exponential class. Constraints are imposed on the expected values of the canonical statistics and not specifically on the parameters themselves. This allows for a much wider class of statistical problems to be addressed than only those that imply constraints on parameters. The exact mles are obtained directly without having to derive and solve the likelihood equations. The procedure provides a new statistical methodology for maximum likelihood estimation implying that complicated maximisation problems can be dealt with in a unified way. The elegance of the procedure lies in its simplicity and flexibility. It can easily be adapted to provide for testing and estimation where structural relationships exist amongst mean vectors and/or covariance matrices, e.g. certain variance/covariance structures such as proportional variances and covariances within a population, or proportional (and possibly structured) covariance matrices of several independent populations.

It must be stated clearly that the aim of this paper is to introduce an alternative approach not only to maximum likelihood estimation where the traditional is problematic, but also in standard estimation applications. Apart from solving problematic estimation problems, the procedure provides an attractive alternative to existing procedures. The objective of this paper is not to make a comparison with competing procedures (in the case where there are alternative procedures available). The examples considered in this paper illustrate sufficiently how complicated problems can be solved.

In Section 2 the basic underlying theory is outlined, the procedure is summarised in a step by step algorithm and a very simple introductory example is given to illustrate the basic steps. In Section 3 the theory is given for independent multivariate normal samples and in Section 4 two large normal samples are considered. Also in Section 4, Behrens-Fisher, Behrens-Fisher with proportional covariance matrices, and estimation of the mean vectors and covariance matrices under the assumption of equal correlation coefficients are illustrated. In Section 5 an example is used to illustrate the basic concepts on more than two relatively small multivariate normal samples first published and discussed by Grizzle and Allen (1969).

2. Maximum likelihood estimation under constraints

The random vector $\boldsymbol{t} : k \times 1$ belongs to the canonical exponential family if its probability density function is of the form (Barndorff-Nielsen, 1982 or Brown, 1984)

$$\begin{aligned} p(\boldsymbol{t},\boldsymbol{\theta}) &= a(\boldsymbol{\theta})b(\boldsymbol{t})\exp(\boldsymbol{\theta}'\boldsymbol{t}) \\ &= b(\boldsymbol{t})\exp\left\{\boldsymbol{\theta}'\boldsymbol{t}-\kappa(\boldsymbol{\theta})\right\}, \quad \boldsymbol{t}\in R^k, \quad \boldsymbol{\theta}\in\aleph \end{aligned} \tag{1}$$

where $\boldsymbol{\theta} : k \times 1$ the canonical (natural) parameter, $\boldsymbol{t} : k \times 1$ the canonical statistic and $\aleph$ the natural parameter space for the canonical parameter $\boldsymbol{\theta}$. The function $\kappa(\boldsymbol{\theta}) = -\ell n\, a(\boldsymbol{\theta})$ is referred to as the cumulant generating function or the log Laplace transform. It is important to note that the statistic $\boldsymbol{t}$ in (1) is a canonical statistic. If $\boldsymbol{t}$ is a sufficient statistic in the regular exponential class, it can be transformed to canonical form.

The mean vector and covariance matrix of $\boldsymbol{t}$ are given by $E(\boldsymbol{t}) = \dfrac{\partial}{\partial\boldsymbol{\theta}}\kappa(\boldsymbol{\theta}) = \boldsymbol{m}$ and $Cov(\boldsymbol{t}) = \dfrac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\kappa(\boldsymbol{\theta}) = \boldsymbol{V}$. The mle of $\boldsymbol{m}$ without any constraints is $\widehat{\boldsymbol{m}} = \boldsymbol{t}$.

The mles of $\boldsymbol{m}$ may in general not exist, except under constraints which are implied by a particular model. In this paper, the mles are not obtained from maximising a likelihood function in terms of parameters. The mles

$\widehat{\boldsymbol{m}}$ of $\boldsymbol{m}$ are obtained under constraints imposed on the expected values of the canonical statistics according to the procedure described by Matthews and Crowther (1995, 1998). Consider the constraints $\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{0}$ where $\boldsymbol{g}$ is a vector valued function of $\boldsymbol{m}$. Then the mle of $\boldsymbol{m}$ under the constraints can be obtained from

$$\widehat{\boldsymbol{m}} = \boldsymbol{t} - (\boldsymbol{G}_m \boldsymbol{V})' \left(\boldsymbol{G}_t \boldsymbol{V} \boldsymbol{G}_m'\right)^* \boldsymbol{g}(\boldsymbol{t}) \tag{2}$$

where $\boldsymbol{G}_m = \dfrac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}}$, $\boldsymbol{G}_t = \left.\dfrac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}}\right|_{\boldsymbol{m}=\boldsymbol{t}}$ and $\left(\boldsymbol{G}_t \boldsymbol{V} \boldsymbol{G}_m'\right)^*$ is a generalised inverse of $\left(\boldsymbol{G}_t \boldsymbol{V} \boldsymbol{G}_m'\right)$.

In general, the iterative procedure implies a double iteration over $\boldsymbol{t}$ and $\boldsymbol{m}$. The first iteration stems from the Taylor series linearisation of $\boldsymbol{g}(\boldsymbol{t})$ and the second from the fact that $\boldsymbol{V}$ may be a function of $\boldsymbol{m}$. If $\boldsymbol{g}(\boldsymbol{t})$ is a linear function, no first iteration is necessary. Similarly, if $\boldsymbol{V}$ and $\boldsymbol{G}_m$ are not functions of $\boldsymbol{m}$, no second iteration over $\boldsymbol{m}$ is necessary. The procedure is initialised with the observed canonical statistic as the starting value for both $\boldsymbol{t}$ and $\boldsymbol{m}$. Convergence is attained first over $\boldsymbol{t}$ and then over $\boldsymbol{m}$. The converged value of $\boldsymbol{t}$ is used as the next approximation for $\boldsymbol{m}$, with iteration over $\boldsymbol{m}$ starting at the observed $\boldsymbol{t}$. The covariance matrix $\boldsymbol{V}$ may be a function of $\boldsymbol{m}$, in which case it is recalculated for each new value of $\boldsymbol{m}$ in the iterative procedure. Convergence over $\boldsymbol{m}$ yields $\widehat{\boldsymbol{m}}$, the mle of $\boldsymbol{m}$ under the constraints. Note that in some instances the notation $\boldsymbol{t}$ is used specifically to represent the canonical statistic and in others to represent the converged value of $\boldsymbol{t}$ in the iterative process.

An algorithm for obtaining the mles under constraints are outlined below:

Step 1:	Specify $\boldsymbol{t}_0$, the vector of observed canonical statistics
Step 2:	Let $\boldsymbol{t} = \boldsymbol{t}_0$.
Step 3A:	Let $\boldsymbol{m} = \boldsymbol{t}$, $\boldsymbol{t} = \boldsymbol{t}_0$. Calculate $\boldsymbol{G}_m$, $\boldsymbol{V}$.
Step 3B:	Let $\boldsymbol{t}_p = \boldsymbol{t}$. Calculate $\boldsymbol{g}(\boldsymbol{t})$, $\boldsymbol{G}_t$. Calculate $\boldsymbol{t} = \boldsymbol{t} - (\boldsymbol{G}_m\boldsymbol{V})' \left(\boldsymbol{G}_t\boldsymbol{V}\boldsymbol{G}_m'\right)^* \boldsymbol{g}(\boldsymbol{t})$. If $(\boldsymbol{t} - \boldsymbol{t}_p)'(\boldsymbol{t} - \boldsymbol{t}_p) < \epsilon$ (a small positive number determining the accuracy), then go to Step 3A, else repeat Step 3B.
Step 4	If $(\boldsymbol{m} - \boldsymbol{t})'(\boldsymbol{m} - \boldsymbol{t}) < \epsilon$ then convergence is attained.

The exact mles are given by $\boldsymbol{m}$.

For the mles to exist, the number of independent constraints, $\nu = rank(\boldsymbol{G}_m\boldsymbol{V}\boldsymbol{G}_m')$, should equal the difference between the number of elements of $\boldsymbol{m}$ and the number of unknown parameters of the constrained model. Constraints need not be independent (Matthews and Crowther, 1998). Care should be taken that the constraints are correctly formulated. As with any iterative procedure, choice of starting values may influence convergence or rate of convergence – specifically (and only) where the sample size is small relative to the dimension of the observations.

The asymptotic covariance matrix of $\widehat{\boldsymbol{m}}$ is given by

$$Cov(\widehat{\boldsymbol{m}}) = \boldsymbol{V} - (\boldsymbol{G}_m\boldsymbol{V})' \left(\boldsymbol{G}_m\boldsymbol{V}\boldsymbol{G}_m'\right)^* \boldsymbol{G}_m\boldsymbol{V} \qquad (3)$$

which is estimated by replacing $\boldsymbol{m}$ with $\widehat{\boldsymbol{m}}$.

Estimates of parameters and parameter structures implied by different model(s) are obtained as a function of the elements of $\widehat{\boldsymbol{m}}$, say $f(\widehat{\boldsymbol{m}})$. Using the multivariate delta method, it follows that

$$Cov(f(\widehat{\boldsymbol{m}})) \approxeq \left(\frac{\partial f(\boldsymbol{m})}{\partial \boldsymbol{m}}\right) Cov\,(\widehat{\boldsymbol{m}}) \left(\frac{\partial f(\boldsymbol{m})}{\partial \boldsymbol{m}}\right)'. \tag{4}$$

The standard error of the estimators are given by $\boldsymbol{\sigma}(f(\widehat{\boldsymbol{m}}))$, the square root of the vector of diagonal elements of $Cov(f(\widehat{\boldsymbol{m}}))$.

The Wald statistic $W = \boldsymbol{g}'(\boldsymbol{t})(\boldsymbol{G}_t\boldsymbol{V}\boldsymbol{G}_t')^{*}\boldsymbol{g}(\boldsymbol{t})$ is a measure of goodness of fit under the hypothesis $\boldsymbol{g}(\boldsymbol{m}) = \mathbf{0}$. It is asymptotically χ^2 distributed with ν degrees of freedom.

Different models require different canonical statistics. In general, the appropriate canonical statistics and corresponding covariance matrix are special cases of the following theoretical result and is utilised repeatedly throughout this paper:

Suppose that $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n$ represent n independent observations of a random sample from a $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution with $\overline{\boldsymbol{y}} = \frac{1}{n}\sum\limits_{i=1}^{n}\boldsymbol{y}_i$ the sample mean vector and $\boldsymbol{S} = \frac{1}{n}\sum\limits_{i=1}^{n}\boldsymbol{y}_i\boldsymbol{y}_i'$ the matrix of mean sums of squares and products. The likelihood function is

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \det(2\pi\boldsymbol{\Sigma})^{-n/2} \exp\left\{-\tfrac{1}{2}tr\boldsymbol{\Sigma}^{-1}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})'\right]\right\} \\ &= \exp\left\{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\boldsymbol{y}_i) - \tfrac{n}{2}tr\boldsymbol{\Sigma}^{-1}(\tfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\boldsymbol{y}_i')\right. \\ &\quad \left. -\tfrac{n}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \tfrac{n}{2}\ell n[\det(2\pi\boldsymbol{\Sigma})]\right\} \\ &= \exp\left[\boldsymbol{\theta}'\boldsymbol{t} - \kappa(\boldsymbol{\theta})\right] \end{aligned}$$

with vector of canonical statistics, $\boldsymbol{t}$, and vector of corresponding canonical parameters, $\boldsymbol{\theta}$, given by:

$$\boldsymbol{t} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i \\ vec(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\boldsymbol{y}_i') \end{pmatrix} = \begin{pmatrix} \overline{\boldsymbol{y}} \\ vec(\boldsymbol{S}) \end{pmatrix}, \ \boldsymbol{\theta} = \begin{pmatrix} n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ -\frac{n}{2}vec(\boldsymbol{\Sigma}^{-1}) \end{pmatrix}$$

where the matrix operation vec creates a column vector from the columns of its matrix argument.

Then

$$E(\boldsymbol{t}) = \boldsymbol{m} = \begin{pmatrix} \boldsymbol{\mu} \\ vec(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}') \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \end{pmatrix} \text{ and } Cov(\boldsymbol{t}) = \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix} \tag{5}$$

where

$$\boldsymbol{V}_{11} = \tfrac{1}{n}\boldsymbol{\Sigma}$$

$$\boldsymbol{V}_{21} = \tfrac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma}) \qquad \text{Lemma 1}$$

$$\boldsymbol{V}_{12} = \boldsymbol{V}_{21}'$$

$$\boldsymbol{V}_{22} = \tfrac{1}{n}\left(I_{p^2} + \boldsymbol{K}\right)[\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' \otimes \boldsymbol{\Sigma}]$$

(Muirhead, 1982, page 518)

where the matrix $\boldsymbol{K}$ is given by $\boldsymbol{K} = \sum_{i,j=1}^{p}(\boldsymbol{H}_{ij} \otimes \boldsymbol{H}_{ij}')$ and $\boldsymbol{H}_{ij} : p \times p$ with $h_{ij} = 1$ and all other elements equal to zero (Muirhead, 1982, page 90).

Lemma 1 $\boldsymbol{V}_{21} = \frac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma})$.

Proof. Let $\boldsymbol{z}_i = \boldsymbol{y}_i - \boldsymbol{\mu}$ where $\boldsymbol{y}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$
\begin{aligned}
\boldsymbol{V}_{21} &= Cov[vec(\boldsymbol{S}), \overline{\boldsymbol{y}}'] \\
&= \tfrac{1}{n^2}\textstyle\sum_{i=1}^{n} Cov[vec(\boldsymbol{y}_i\boldsymbol{y}_i'), \boldsymbol{y}_i'] \\
&= \tfrac{1}{n} Cov[(\boldsymbol{y}_i \otimes \boldsymbol{y}_i), \boldsymbol{y}_i'] \\
&= \tfrac{1}{n} Cov[(\boldsymbol{z}_i + \boldsymbol{\mu}) \otimes (\boldsymbol{z}_i + \boldsymbol{\mu}), (\boldsymbol{z}_i + \boldsymbol{\mu})'] \\
&= \tfrac{1}{n}[Cov(\boldsymbol{z}_i \otimes \boldsymbol{\mu}, \boldsymbol{z}_i') + Cov(\boldsymbol{\mu} \otimes \boldsymbol{z}_i, \boldsymbol{z}_i')] \text{ since } Cov(\boldsymbol{z}_i \otimes \boldsymbol{z}_i, \boldsymbol{z}_i') = 0 \\
&= \tfrac{1}{n}[Cov(\boldsymbol{z}_i, \boldsymbol{z}_i') \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes Cov(\boldsymbol{z}_i, \boldsymbol{z}_i')]. \quad \blacksquare
\end{aligned}
$$

2.1 Introductory example

The essence of the procedure can be described by the simple case where we have a sample mean vector $\overline{\boldsymbol{y}} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and sample covariance matrix $\boldsymbol{A} = \begin{pmatrix} 5 & 4 \\ 4 & 8 \end{pmatrix}$ of size $n = 10$ from a $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution where it is assumed that the variance of each variable is equal to its squared mean. This cannot be accomplished by PROC MIXED.

The vector of canonical statistics, $\boldsymbol{t}$, and $Cov(\boldsymbol{t})$ are given by (5):

$$
\boldsymbol{t} = \begin{pmatrix} \overline{\boldsymbol{y}} \\ vec(\boldsymbol{S}) \end{pmatrix} \text{ and } Cov(\boldsymbol{t}) = \boldsymbol{V}
$$

$$
= \begin{pmatrix} \dfrac{1}{n}\boldsymbol{\Sigma} & \dfrac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma})' \\ \dfrac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma}) & \dfrac{1}{n}\left(I_{p^2} + \boldsymbol{K}\right)[\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' \otimes \boldsymbol{\Sigma}] \end{pmatrix}.
$$

Suppose the variances are assumed to be equal to the squared means, i.e. $\sigma_{11} = \mu_1^2$ and $\sigma_{22} = \mu_2^2$. If the i-th element of $vec(\boldsymbol{\Sigma})$ is indicated by $vec(\boldsymbol{\Sigma})[i]$ and the i-th element of $\boldsymbol{\mu}$ by $\boldsymbol{\mu}[i]$, then

$$\boldsymbol{g}(\boldsymbol{m}) = \begin{pmatrix} \boldsymbol{g}_1(\boldsymbol{m}) \\ \\ \boldsymbol{g}_2(\boldsymbol{m}) \end{pmatrix} \quad \text{where}$$

$$g_1(\boldsymbol{m}) = \sigma_{11} - \mu_1^2 = vec(\boldsymbol{\Sigma})[1] - \boldsymbol{\mu}[1]^2 = \{\boldsymbol{m}_2 - vec(\boldsymbol{m}_1'\boldsymbol{m}_1)\}\,[1] - \boldsymbol{m}_1[1]^2,$$

$$g_2(\boldsymbol{m}) = \sigma_{22} - \mu_2^2 = vec(\boldsymbol{\Sigma})[4] - \boldsymbol{\mu}[2]^2 = \{\boldsymbol{m}_2 - vec(\boldsymbol{m}_1\boldsymbol{m}_1')\}\,[4] - \boldsymbol{m}_1[2]^2,$$

with derivative

$$\boldsymbol{G}_m = \begin{pmatrix} \boldsymbol{G}_1(\boldsymbol{m}) \\ \\ \boldsymbol{G}_2(\boldsymbol{m}) \end{pmatrix} = \begin{pmatrix} \dfrac{\partial vec(\boldsymbol{\Sigma})[1]}{\partial vec(\boldsymbol{\Sigma})} \times \dfrac{\partial vec(\boldsymbol{\Sigma})}{\partial \boldsymbol{m}} - 2\boldsymbol{\mu}[1] \times \dfrac{\partial \boldsymbol{\mu}[1]}{\partial \boldsymbol{m}} \\ \\ \dfrac{\partial vec(\boldsymbol{\Sigma})[4]}{\partial vec(\boldsymbol{\Sigma})} \times \dfrac{\partial vec(\boldsymbol{\Sigma})}{\partial \boldsymbol{m}} - 2\boldsymbol{\mu}[2] \times \dfrac{\partial \boldsymbol{\mu}[2]}{\partial \boldsymbol{m}} \end{pmatrix},$$

$$\boldsymbol{G}_t = \left.\frac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}}\right|_{\boldsymbol{m}=\boldsymbol{t}}.$$

The iteration process proceeds as indicated in Table 1.

In each new iteration over $\boldsymbol{m}$, the observed $\boldsymbol{t}$ $(\boldsymbol{t}_0)$ is used as initial value for $\boldsymbol{t}$.

The mle for $\boldsymbol{\Sigma}$ under the restrictions $\sigma_{11} = \mu_1^2$ and $\sigma_{22} = \mu_2^2$ are calculated as a function of the elements of $\widehat{\boldsymbol{m}}$, namely: $\widehat{\boldsymbol{\mu}}_1 = \widehat{\boldsymbol{m}}_1 = \begin{pmatrix} 2.060 \\ 2.820 \end{pmatrix}$ and

$$vec(\widehat{\boldsymbol{\Sigma}}) = \widehat{\boldsymbol{m}}_2 - vec(\widehat{\boldsymbol{m}}_1\widehat{\boldsymbol{m}}_1') = \begin{pmatrix} 4.24 \\ 3.67 \\ 3.67 \\ 7.95 \end{pmatrix}.$$

Table 1 Double iteration process illustrated for the example in Section 2.1

		Iteration over t			
	Iteration over m	Starting Value	1	2	3
1	$\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$	$\begin{pmatrix} 2.047 \\ 2.800 \\ 8.374 \\ 9.374 \\ 9.374 \\ 15.596 \end{pmatrix}$	$\begin{pmatrix} 2.045 \\ 2.789 \\ 8.363 \\ 9.356 \\ 9.356 \\ 15.562 \end{pmatrix}$	$\begin{pmatrix} 2.045 \\ 2.789 \\ 8.363 \\ 9.356 \\ 9.356 \\ 15.562 \end{pmatrix}$
2	$\begin{pmatrix} 2.045 \\ 2.789 \\ 8.363 \\ 9.356 \\ 9.356 \\ 15.562 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$	$\begin{pmatrix} 2.062 \\ 2.828 \\ 8.494 \\ 9.490 \\ 9.490 \\ 15.931 \end{pmatrix}$	$\begin{pmatrix} 2.060 \\ 2.821 \\ 8.487 \\ 9.479 \\ 9.479 \\ 15.914 \end{pmatrix}$	$\begin{pmatrix} 2.060 \\ 2.821 \\ 8.487 \\ 9.479 \\ 9.479 \\ 15.914 \end{pmatrix}$
3	$\begin{pmatrix} 2.060 \\ 2.821 \\ 8.487 \\ 9.479 \\ 9.479 \\ 15.914 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$	$\begin{pmatrix} 2.062 \\ 2.827 \\ 8.496 \\ 9.489 \\ 9.489 \\ 15.924 \end{pmatrix}$	$\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$	$\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$
4	$\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$	$\begin{pmatrix} 2.062 \\ 2.827 \\ 8.496 \\ 9.489 \\ 9.489 \\ 15.924 \end{pmatrix}$	$\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$	$\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$

3. Independent multivariate normal samples

The theory and methodology that follow can be applied for any number, k, of samples which may stem from any experimental design. In view of the examples that follow in the next two sections and for the sake of simplicity, the theoretical results, an extension of the theory given in Section 2 (cf. (5)), are given for $k = 4$ multivariate normal samples. The results for general k follow directly.

Suppose that $\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, \cdots, \boldsymbol{y}_{in_i}$ represent n_i independent observations of 4 independent random samples from $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ distributions $(i = 1, 2, 3, 4)$. Let $\overline{\boldsymbol{y}}_i$ represent the sample mean vector

$$\overline{\boldsymbol{y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij} \tag{6}$$

and $\boldsymbol{S}_i$ the matrix of mean sums of squares and products of the i-th sample

$$\boldsymbol{S}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}\boldsymbol{y}'_{ij}. \tag{7}$$

Note that n_i is not assumed to be larger than $p + 1$. If any of the n_i is less than $p + 1$, the mles in general do not exist unless certain restrictions on parameters are imposed. The likelihood function for the k multivariate samples can be expressed in its canonical exponential form as follows:

$$\begin{aligned} & L\left(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3, \boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3, \boldsymbol{\Sigma}_4\right) \\ & = \textstyle\prod_{i=1}^{4} \det\left(2\pi\boldsymbol{\Sigma}_i\right)^{-n_i/2} \exp\left\{-\frac{1}{2}tr\boldsymbol{\Sigma}_i - 1\left[\sum_{j=1}^{n_i}\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_i\right)\left(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_i\right)'\right]\right\} \\ & = \textstyle\prod_{i=1}^{4} \exp\left\{n_i\boldsymbol{\mu}'_i\boldsymbol{\Sigma}_i^{-1}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}\right) - \frac{n_i}{2}tr\boldsymbol{\Sigma}_i^{-1}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}\boldsymbol{y}'_{ij}\right) \right. \\ & \quad \left. - \frac{n_i}{2}\boldsymbol{\mu}'_i\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{n_i}{2}\ell n[\det(2\pi\boldsymbol{\Sigma}_i)]\right\} \\ & = \exp\left[\boldsymbol{\theta}'\boldsymbol{t} - \kappa\left(\boldsymbol{\theta}\right)\right] \qquad \text{where} \end{aligned} \tag{8}$$

$$\boldsymbol{t}=\begin{pmatrix} \frac{1}{n_1}\sum_{j=1}^{n_1}\boldsymbol{y}_{1j} \\ vec(\frac{1}{n_1}\sum_{j=1}^{n_1}\boldsymbol{y}_{1j}\boldsymbol{y}'_{1j}) \\ \frac{1}{n_2}\sum_{j=1}^{n_2}\boldsymbol{y}_{2j} \\ vec(\frac{1}{n_2}\sum_{j=1}^{n_2}\boldsymbol{y}_{2j}\boldsymbol{y}'_{2j}) \\ \frac{1}{n_3}\sum_{j=1}^{n_3}\boldsymbol{y}_{3j} \\ vec(\frac{1}{n_3}\sum_{j=1}^{n_3}\boldsymbol{y}_{3j}\boldsymbol{y}'_{3j}) \\ \frac{1}{n_4}\sum_{j=1}^{n_4}\boldsymbol{y}_{4j} \\ vec(\frac{1}{n_4}\sum_{j=1}^{n_4}\boldsymbol{y}_{4j}\boldsymbol{y}'_{4j}) \end{pmatrix} = \begin{pmatrix} \overline{\boldsymbol{y}}_1 \\ vec(\boldsymbol{S}_1) \\ \overline{\boldsymbol{y}}_2 \\ vec(\boldsymbol{S}_2) \\ \overline{\boldsymbol{y}}_3 \\ vec(\boldsymbol{S}_3) \\ \overline{\boldsymbol{y}}_4 \\ vec(\boldsymbol{S}_4) \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} n_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 \\ -\frac{n_1}{2}vec(\boldsymbol{\Sigma}_1^{-1}) \\ n_2\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\ -\frac{n_2}{2}vec(\boldsymbol{\Sigma}_2^{-1}) \\ n_3\boldsymbol{\Sigma}_3^{-1}\boldsymbol{\mu}_3 \\ -\frac{n_3}{2}vec(\boldsymbol{\Sigma}_3^{-1}) \\ n_4\boldsymbol{\Sigma}_4^{-1}\boldsymbol{\mu}_4 \\ -\frac{n_4}{2}vec(\boldsymbol{\Sigma}_4^{-1}) \end{pmatrix} \tag{9}$$

and

$$E(\boldsymbol{t}) = \boldsymbol{m} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ vec(\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}'_1) \\ \boldsymbol{\mu}_2 \\ vec(\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}'_2) \\ \boldsymbol{\mu}_3 \\ vec(\boldsymbol{\Sigma}_3 + \boldsymbol{\mu}_3\boldsymbol{\mu}'_3) \\ \boldsymbol{\mu}_4 \\ vec(\boldsymbol{\Sigma}_4 + \boldsymbol{\mu}_4\boldsymbol{\mu}'_4) \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_{11} \\ \boldsymbol{m}_{12} \\ \boldsymbol{m}_{21} \\ \boldsymbol{m}_{22} \\ \boldsymbol{m}_{31} \\ \boldsymbol{m}_{32} \\ \boldsymbol{m}_{41} \\ \boldsymbol{m}_{42} \end{pmatrix}. \tag{10}$$

The covariance matrix of $\boldsymbol{t}$ (cf. (5)) is given by

$$\boldsymbol{V} = Cov(\boldsymbol{t}) = \begin{pmatrix} \boldsymbol{V}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{V}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{V}_{44} \end{pmatrix} \text{ with } \boldsymbol{V}_{ii} = \begin{pmatrix} \boldsymbol{V}_{11}^{i} & \boldsymbol{V}_{12}^{i} \\ \boldsymbol{V}_{21}^{i} & \boldsymbol{V}_{22}^{i} \end{pmatrix} \tag{11}$$

and

$$\left.\begin{aligned} \boldsymbol{V}_{11}^{i} &= \frac{1}{n_i}\boldsymbol{\Sigma}_i \\ \boldsymbol{V}_{21}^{i} &= \frac{1}{n_i}\left(\boldsymbol{\Sigma}_i \otimes \boldsymbol{\mu}_i + \boldsymbol{\mu}_i \otimes \boldsymbol{\Sigma}_i\right) \\ \boldsymbol{V}_{12}^{i} &= \boldsymbol{V}_{21}^{i\,\prime} \end{aligned}\right\} \text{ for } i = 1,2,3,4.$$

$$\boldsymbol{V}_{22}^{i} = \frac{1}{n_i}\left(I_{p^2} + \boldsymbol{K}\right)\left[\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \otimes \boldsymbol{\mu}_i\boldsymbol{\mu}_i' + \boldsymbol{\mu}_i\boldsymbol{\mu}_i' \otimes \boldsymbol{\Sigma}_i\right]. \tag{12}$$

In the procedure (2) for estimating the $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ under restrictions, the vector $\boldsymbol{t}$ of canonical statistics (cf. (9)) with corresponding covariance matrix $\boldsymbol{V}$ (cf. (11)) is used as point of departure for any model considered within this framework. The algorithm given in Section 2 is then used to obtain the mles. In each section the specific expressions for any $\boldsymbol{g}(\boldsymbol{m})$ and $\boldsymbol{G}_m$ are given. Constraints $\boldsymbol{g}(\boldsymbol{m})$ specified will always be functions of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$. Consequently, the corresponding matrix of derivatives $\boldsymbol{G}_m = \dfrac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}}$ may also be expressed as $\boldsymbol{G}_m = \dfrac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \times \dfrac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}}$ where $\boldsymbol{m}_{\boldsymbol{\Sigma}}' = \left(vec(\boldsymbol{\Sigma}_1)', \; vec(\boldsymbol{\Sigma}_2)', \; vec(\boldsymbol{\Sigma}_3)', \; vec(\boldsymbol{\Sigma}_4)' \right)$.

Basic derivatives that will be required throughout, are the following:

$$\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{m}} = (\boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2})$$
$$\frac{\partial \boldsymbol{\mu}_2}{\partial \boldsymbol{m}} = (\boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2})$$
$$\frac{\partial \boldsymbol{\mu}_3}{\partial \boldsymbol{m}} = (\boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2})$$
$$\frac{\partial \boldsymbol{\mu}_4}{\partial \boldsymbol{m}} = (\boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2})$$
$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}} = (\tfrac{1}{4}\boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2}, \quad \tfrac{1}{4}\boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2}, \quad \tfrac{1}{4}\boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2}, \quad \tfrac{1}{4}\boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2}) \tag{13}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 + \boldsymbol{\mu}_3 + \boldsymbol{\mu}_4)/4$.

Since $\dfrac{\partial vec(\boldsymbol{m}'_{i1}\boldsymbol{m}_{i1})}{\partial \boldsymbol{m}_{i1}} = \boldsymbol{I}_p \otimes \boldsymbol{m}_{i1} + \boldsymbol{m}_{i1} \otimes \boldsymbol{I}_p$ and $vec(\boldsymbol{\Sigma}_i) = \boldsymbol{m}_{i2} - vec(\boldsymbol{m}_{i1}\boldsymbol{m}'_{i1})$, it follows that:

$$\frac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}} = \begin{pmatrix} \dfrac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} \\ \dfrac{\partial vec(\boldsymbol{\Sigma}_2)}{\partial \boldsymbol{m}} \\ \dfrac{\partial vec(\boldsymbol{\Sigma}_3)}{\partial \boldsymbol{m}} \\ \dfrac{\partial vec(\boldsymbol{\Sigma}_4)}{\partial \boldsymbol{m}} \end{pmatrix} = \oplus_{i=1}^{4} (-\boldsymbol{I}_p \otimes \boldsymbol{m}_{i1} - \boldsymbol{m}_{i1} \otimes \boldsymbol{I}_p, \quad \boldsymbol{I}_{p^2}), \tag{14}$$

the direct sum (block diagonal structure) of the submatrices $\left(\dfrac{\partial[\boldsymbol{m}_{i2} - vec(\boldsymbol{m}'_{i1}\boldsymbol{m}_{i1})]}{\partial \boldsymbol{m}_{i1}}, \dfrac{\partial[\boldsymbol{m}_{i2} - vec(\boldsymbol{m}_{i1}\boldsymbol{m}'_{i1})]}{\partial \boldsymbol{m}_{i2}}\right)$.

4. Two large independent multivariate normal samples

Johnson and Wichern (1982), Nel and Van der Merwe (1986) considered two measurements of electrical usage during on-peak and off-peak hours, for homeowners with (Population 1) and without air conditioning (Population 2) respectively. Summary statistics, namely the sample mean vectors $\overline{\boldsymbol{y}}_i$, sample covariance matrices $\boldsymbol{A}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i)(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i)'$ and sample correlation coefficients r_i, $i = 1, 2$, for samples of size $n_1 = 45$ and $n_2 = 55$ respectively, are given in Table 2. (The off-peak consumption is higher than the on-peak consumption because there are more off-peak hours in a month). The SAS procedure MIXED can not be used in this case, since the raw data is not available. More importantly, although PROC MIXED provides for a variety of covariance structures, it does not allow for structural relations between covariance matrices. In Table 2 the results for a variety of hypotheses on the two mean vectors and covariance matrices are also presented. These hypotheses imply specific constraints but the procedure does not require the maximisation of a likelihood function.

4.1 $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

The usual hypothesis of equal mean vectors with homogeneous covariance matrices implies the constraints:

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{m}) &= \begin{pmatrix} \boldsymbol{g}_1(\boldsymbol{m}) \\ \\ \boldsymbol{g}_2(\boldsymbol{m}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \\ vec(\boldsymbol{\Sigma}_1) - vec(\boldsymbol{\Sigma}_2) \end{pmatrix} \\ \text{with } \boldsymbol{G}_m &= \begin{pmatrix} \dfrac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{m}} - \dfrac{\partial \boldsymbol{\mu}_2}{\partial \boldsymbol{m}} \\ \\ \dfrac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} - \dfrac{\partial vec(\boldsymbol{\Sigma}_2)}{\partial \boldsymbol{m}} \end{pmatrix} \end{aligned} \quad (15)$$

where

$$\begin{aligned}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{m}} &= \left(\boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2} \right) \\ \frac{\partial \boldsymbol{\mu}_2}{\partial \boldsymbol{m}} &= \left(\boldsymbol{0}_{p\times p}, \quad \boldsymbol{0}_{p\times p^2}, \quad \boldsymbol{I}_p, \quad \boldsymbol{0}_{p\times p^2} \right)\end{aligned} \tag{16}$$

and

$$\begin{aligned}&\frac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} = \frac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \times \frac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}} = (\boldsymbol{I}_{p^2},\ \boldsymbol{0}_{p^2\times p^2}) \\ &\times(-\boldsymbol{I}_p \otimes \boldsymbol{m}_{11} - \boldsymbol{m}_{11} \otimes \boldsymbol{I}_p,\ \boldsymbol{I}_{p^2},\ \boldsymbol{0}_{p^2\times p},\ \boldsymbol{0}_{p^2\times p^2}) \\ &\frac{\partial vec(\boldsymbol{\Sigma}_2)}{\partial \boldsymbol{m}} = \frac{\partial vec(\boldsymbol{\Sigma}_2)}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \times \frac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}} = (\boldsymbol{0}_{p^2\times p^2},\ \boldsymbol{I}_{p^2}) \\ &\times(\boldsymbol{0}_{p^2\times p},\ \boldsymbol{0}_{p^2\times p^2},\ -\boldsymbol{I}_p \otimes \boldsymbol{m}_{21} - \boldsymbol{m}_{21} \otimes \boldsymbol{I}_p,\ \boldsymbol{I}_{p^2})\end{aligned} \tag{17}$$

where $\boldsymbol{m}'_{\boldsymbol{\Sigma}} = \left(vec(\boldsymbol{\Sigma}_1)', \quad vec(\boldsymbol{\Sigma}_2)' \right)$. From Table 2 it follows that this model does not fit well.

4.2 $H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$

Proportional covariance matrices and proportional mean vectors seem to be a plausible model for this particular data set. The difficulties and limitations traditionally encountered with maximum likelihood estimation in the case of proportional covariance matrices are discussed by authors such as Flury (1986), Eriksen (1987), Jensen and Johansen (1987) and Jensen and Madsen (2004). Consider the hypothesis $H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$ where, for example, $c = \frac{\sigma_{11}^2}{\sigma_{11}^1}$ and σ_{11}^i denotes the variance of the first measurement in the i-th population. Using procedure (2), only the following slight adjustment is made to the constraint $\boldsymbol{g}_2(\boldsymbol{m})$ in (15):

$$\boldsymbol{g}_2(\boldsymbol{m}) = vec(\boldsymbol{\Sigma}_2) - c\,vec(\boldsymbol{\Sigma}_1)$$

with derivative

$$\boldsymbol{G}_{2m} = \frac{\partial vec(\boldsymbol{\Sigma}_2)}{\partial \boldsymbol{m}} - c\frac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} - vec(\boldsymbol{\Sigma}_1)\frac{\partial c}{\partial \boldsymbol{m}} \tag{18}$$

and

$$\frac{\partial c}{\partial \boldsymbol{m}} = \frac{\partial c}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \times \frac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}} \text{ where } \frac{\partial c}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} = \left(\frac{\sigma_{11}^2}{(\sigma_{11}^1)^2}\boldsymbol{I}_{p^2}[1,], \ \frac{1}{\sigma_{11}^1}\boldsymbol{I}_{p^2}[1,] \right), \tag{19}$$

$\boldsymbol{I}_{p^2}[1,]$ denotes the first row of $\boldsymbol{I}_{p^2}$ and $\dfrac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}}$ is obtained from (17).

In Table 2, the results for the hypothesis of proportional covariance matrices are given, with no constraints imposed on the mean vectors. This model fits well. Note that the proportional covariances imply equal correlation coefficients. This correlation coefficient $\widehat{\rho}_i$ is also reported.

4.3 $H_0 : \rho_1 = \rho_2$

The correlation coefficient $\widehat{\rho}_i$ which is obtained as a result of covariance matrices being proportional, differs from the estimated common correlation coefficient (cf. Manly and Rayner, 1987) which is obtained when the hypothesis of equal correlation coefficients is imposed as a direct constraint. This constraint, which is less restrictive than proportionality of complete covariance matrices, is given by:

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{m}) &= \rho_1 - \rho_2 = \frac{\sigma_{12}^1}{\sqrt{\sigma_{11}^1 \sigma_{22}^1}} - \frac{\sigma_{12}^2}{\sqrt{\sigma_{11}^2 \sigma_{22}^2}} \\ &= \frac{vec(\boldsymbol{\Sigma}_1)[2]}{\sqrt{vec(\boldsymbol{\Sigma}_1)[1]vec(\boldsymbol{\Sigma}_1)[p+2]}} - \frac{vec(\boldsymbol{\Sigma}_2)[2]}{\sqrt{vec(\boldsymbol{\Sigma}_2)[1]vec(\boldsymbol{\Sigma}_2)[p+2]}}, \end{aligned} \tag{20}$$

where ρ_i and σ_{12}^i the correlation and covariance respectively between on-peak and off-peak electrical usage for population i, and $\sigma_{11}^i, \sigma_{12}^i, \sigma_{22}^i$ the element in row 1, 2, and $p+2$ of $vec(\boldsymbol{\Sigma}_i)$ respectively,

$$
\boldsymbol{G}_m = \frac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \times \frac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}} \text{ where } \frac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} = \frac{\partial(\rho_1 - \rho_2)}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} =
$$

$$
\left(\frac{\boldsymbol{I}_{p^2}[2,]}{\sqrt{vec(\boldsymbol{\Sigma}_1)[1]vec(\boldsymbol{\Sigma}_1)[p+2]}} - \frac{\boldsymbol{I}_{p^2}[1,]vec(\boldsymbol{\Sigma}_1)[2]}{2\sqrt{\{vec(\boldsymbol{\Sigma}_1)[1]\}^3 vec(\boldsymbol{\Sigma}_1)[p+2]}} \right.
$$

$$
\left. - \frac{\boldsymbol{I}_{p^2}[p+2,]vec(\boldsymbol{\Sigma}_1)[2]}{2\sqrt{vec(\boldsymbol{\Sigma}_1)[1]\{vec(\boldsymbol{\Sigma}_1)[p+2]\}^3}},\ \boldsymbol{0}_{1\times p^2} \right)
$$

$$
- \left(\boldsymbol{0}_{1\times p^2}, \frac{\boldsymbol{I}_{p^2}[2,]}{\sqrt{vec(\boldsymbol{\Sigma}_2)[1]vec(\boldsymbol{\Sigma}_2)[p+2]}} - \frac{\boldsymbol{I}_{p^2}[1,]vec(\boldsymbol{\Sigma}_2)[2]}{2\sqrt{\{vec(\boldsymbol{\Sigma}_2)[1]\}^3 vec(\boldsymbol{\Sigma}_2)[p+2]}} \right.
$$

$$
\left. - \frac{\boldsymbol{I}_{p^2}[p+2,]vec(\boldsymbol{\Sigma}_2)[2]}{2\sqrt{vec(\boldsymbol{\Sigma}_2)[1]\{vec(\boldsymbol{\Sigma}_2)[p+2]\}^3}} \right). \tag{21}
$$

The results are also given in Table 2. Since the hypothesis cannot be formulated in terms of the parameters of the means or (co)variances, PROC MIXED cannot be used in this case.

4.4 $H_0 : \boldsymbol{\mu}_2 = \sqrt{c}\boldsymbol{\mu}_1$, $H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$

For the hypothesis of proportional mean vectors, $H_0 : \boldsymbol{\mu}_2 = \sqrt{c}\boldsymbol{\mu}_1$, the following adjustment to $\boldsymbol{g}_1(\boldsymbol{m})$ is required:

$$
\boldsymbol{g}_1(\boldsymbol{m}) = \boldsymbol{\mu}_2 - \sqrt{c}\boldsymbol{\mu}_1 \text{ with derivative } \boldsymbol{G}_{1m} = \frac{\partial \boldsymbol{\mu}_2}{\partial \boldsymbol{m}} - \sqrt{c}\frac{\partial \boldsymbol{\mu}_1}{\partial \boldsymbol{m}} - \frac{1}{2\sqrt{c}} \left(\boldsymbol{\mu}_1 \frac{\partial c}{\partial \boldsymbol{m}} \right). \tag{22}
$$

This model seems to be the proper solution to the problem stated by Johnson and Wichern (1982). The results are given in Table 2.

Table 2 A Behrens-Fisher problem for large multivariate samples

$\overline{\boldsymbol{y}}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$	$\overline{\boldsymbol{y}}_2 = \begin{pmatrix} 130 \\ 355 \end{pmatrix}$	$\boldsymbol{A}_1 = \begin{pmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{pmatrix}$ $r_1 = 0.75$	$\boldsymbol{A}_2 = \begin{pmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{pmatrix}$ $r_2 = 0.89$
$H_0 : \mu_1 = \mu_2, H_0 : \Sigma_1 = \Sigma_2$			
$\widehat{\boldsymbol{\mu}} = \begin{pmatrix} 163.48 \\ 445.72 \end{pmatrix}$		$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 12114.41 & 24787.58 \\ 24787.58 & 72447.12 \end{pmatrix}$	
Wald =27.43, ν=5, p-value $<$0.0001		$\widehat{\rho}_1 = \widehat{\rho}_2$=0.8367	
$H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$			
$\widehat{\boldsymbol{\mu}}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$	$\widehat{\boldsymbol{\mu}}_2 = \begin{pmatrix} 130 \\ 355 \end{pmatrix}$	$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 16066.45 & 33169.95 \\ 33169.95 & 96892.87 \end{pmatrix}$	$\widehat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 7501.54 & 15487.27 \\ 15487.27 & 45239.94 \end{pmatrix}$
Wald =4.01, ν=2, p-value=0.1347		$\widehat{\rho}_1 = \widehat{\rho}_2$=0.8407	
$H_0 : \rho_1 = \rho_2$			
$\widehat{\boldsymbol{\mu}}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$	$\widehat{\boldsymbol{\mu}}_2 = \begin{pmatrix} 130 \\ 355 \end{pmatrix}$	$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 17118.26 & 33130.55 \\ 33130.55 & 90520.39 \end{pmatrix}$	$\widehat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 7230.82 & 15495.79 \\ 15495.79 & 46880.11 \end{pmatrix}$
Wald =4.08, ν=1, p-value =0.0434		$\widehat{\rho}_1 = \widehat{\rho}_2$=0.8416	
$H_0 : \boldsymbol{\mu}_2 = \sqrt{c}\boldsymbol{\mu}_1, H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$			
$\widehat{\boldsymbol{\mu}}_1 = \begin{pmatrix} 199.9 \\ 545.174 \end{pmatrix}$	$\widehat{\boldsymbol{\mu}}_2 = \begin{pmatrix} 132.44 \\ 361.194 \end{pmatrix}$	$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 16718.95 & 34657.21 \\ 34657.21 & 101122.67 \end{pmatrix}$	$\widehat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 7338.72 & 15212.65 \\ 15212.65 & 44387.42 \end{pmatrix}$
Wald =4.19, ν=4, p-value =0.3811		$\widehat{c}$=0.4389, $\widehat{\rho}_1 = \widehat{\rho}_2$=0.8429	

5. More than two relatively small multivariate normal samples

To further illustrate the flexibility of the procedure, we consider the real data example discussed by Grizzle and Allen (1969), Wang (1998), Zhang and Xu (2009). This longitudinal data set consists of coronary sinus potassium concentrations measured on each of 36 dogs. The seven measurements on each dog were taken every 2 minutes from 1 to 13 minutes after occlusion. These 36 dogs were divided into four treatment groups with $n_1 = 9$, $n_2 = 10$, $n_3 = 8$ and $n_4 = 9$ dogs respectively. The sample mean vectors $\overline{\boldsymbol{y}}_i$, sample covariance matrices $\boldsymbol{A}_i$, and sample correlation matrices $\boldsymbol{R}_i$, $i = 1, 2, 3, 4$ are given in Table 3. Sample mean vectors are represented graphically in Figure 1.

To make provision for the dependence of observations on the same individual, Grizzle and Allen (1969) analysed this data set by incorporating into the linear model a set of covariates which are a basis of the set of linear functions of the vectors which span the within individual error space. They assumed a common covariance matrix across individuals. Wang (1998) presented a nested nonparametric mixed effects smoothing spline analysis of variance of the data. Zhang and Xu (2009) transformed this four-sample Behrens-Fisher problem into a one-sample problem by extending Scheffe's transformation method. The resulting sample consisted of eight observations with dimension 21, to which they then applied an L^2 norm based test.

In Sections 5.1, 5.2, 5.3 and 5.4 the method of maximum likelihood under constraints is used to estimate different parameter structures directly. For the sake of illustration only a few of several models considered, are presented in the following sections.

Table 3 Sample mean vectors, covariance and correlation matrices

$\overline{\boldsymbol{y}}_i$	$\boldsymbol{A}_i$
$\begin{pmatrix} 4.111 \\ 4.178 \\ 4.400 \\ 4.656 \\ 5.067 \\ 5.222 \\ 4.722 \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.12 & 0.01 & -0.01 & 0.05 & 0.22 & 0.21 \\ & 0.08 & 0.01 & -0.00 & 0.05 & 0.14 & 0.16 \\ & & 0.22 & 0.24 & 0.17 & 0.07 & 0.07 \\ & & & 0.41 & 0.36 & 0.20 & 0.13 \\ & & & & 0.42 & 0.29 & 0.27 \\ & & & & & 0.51 & 0.47 \\ & & & & & & 0.57 \end{pmatrix}$
$\begin{pmatrix} 3.540 \\ 3.630 \\ 3.620 \\ 3.460 \\ 3.660 \\ 3.500 \\ 3.460 \end{pmatrix}$	$\begin{pmatrix} 0.29 & 0.22 & 0.26 & 0.26 & 0.33 & 0.24 & 0.26 \\ & 0.19 & 0.20 & 0.20 & 0.25 & 0.18 & 0.21 \\ & & 0.24 & 0.22 & 0.27 & 0.21 & 0.23 \\ & & & 0.33 & 0.36 & 0.23 & 0.27 \\ & & & & 0.46 & 0.27 & 0.32 \\ & & & & & 0.23 & 0.23 \\ & & & & & & 0.26 \end{pmatrix}$
$\begin{pmatrix} 3.600 \\ 3.725 \\ 4.200 \\ 4.438 \\ 4.500 \\ 4.325 \\ 4.225 \end{pmatrix}$	$\begin{pmatrix} 0.18 & 0.16 & 0.23 & 0.29 & 0.19 & 0.07 & 0.08 \\ & 0.17 & 0.29 & 0.28 & 0.17 & 0.07 & 0.14 \\ & & 0.79 & 0.63 & 0.34 & 0.31 & 0.48 \\ & & & 0.71 & 0.48 & 0.36 & 0.35 \\ & & & & 0.46 & 0.36 & 0.29 \\ & & & & & 0.51 & 0.49 \\ & & & & & & 0.66 \end{pmatrix}$
$\begin{pmatrix} 3.645 \\ 3.778 \\ 4.011 \\ 4.067 \\ 3.978 \\ 4.056 \\ 4.044 \end{pmatrix}$	$\begin{pmatrix} 0.13 & 0.11 & 0.10 & 0.15 & 0.13 & 0.15 & 0.09 \\ & 0.16 & 0.15 & 0.17 & 0.19 & 0.19 & 0.09 \\ & & 0.26 & 0.24 & 0.17 & 0.11 & 0.00 \\ & & & 0.30 & 0.28 & 0.28 & 0.15 \\ & & & & 0.47 & 0.51 & 0.36 \\ & & & & & 0.65 & 0.48 \\ & & & & & & 0.41 \end{pmatrix}$

Table 3 (Continued)

$\bar{\boldsymbol{y}}_i$	$\boldsymbol{R}_i$
$\begin{pmatrix} 4.111 \\ 4.178 \\ 4.400 \\ 4.656 \\ 5.067 \\ 5.222 \\ 4.722 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.94 & 0.06 & -0.03 & 0.17 & 0.68 & 0.64 \\ & 1 & 0.05 & -0.02 & 0.26 & 0.71 & 0.72 \\ & & 1 & 0.79 & 0.57 & 0.20 & 0.19 \\ & & & 1 & 0.87 & 0.43 & 0.26 \\ & & & & 1 & 0.62 & 0.56 \\ & & & & & 1 & 0.87 \\ & & & & & & 1 \end{pmatrix}$
$\begin{pmatrix} 3.540 \\ 3.630 \\ 3.620 \\ 3.460 \\ 3.660 \\ 3.500 \\ 3.460 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.93 & 0.97 & 0.85 & 0.90 & 0.95 & 0.96 \\ & 1 & 0.94 & 0.81 & 0.83 & 0.87 & 0.92 \\ & & 1 & 0.79 & 0.82 & 0.90 & 0.93 \\ & & & 1 & 0.91 & 0.84 & 0.93 \\ & & & & 1 & 0.85 & 0.92 \\ & & & & & 1 & 0.96 \\ & & & & & & 1 \end{pmatrix}$
$\begin{pmatrix} 3.600 \\ 3.725 \\ 4.200 \\ 4.438 \\ 4.500 \\ 4.325 \\ 4.225 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.92 & 0.60 & 0.82 & 0.67 & 0.23 & 0.22 \\ & 1 & 0.79 & 0.82 & 0.62 & 0.25 & 0.42 \\ & & 1 & 0.84 & 0.57 & 0.49 & 0.66 \\ & & & 1 & 0.84 & 0.59 & 0.52 \\ & & & & 1 & 0.75 & 0.52 \\ & & & & & 1 & 0.84 \\ & & & & & & 1 \end{pmatrix}$
$\begin{pmatrix} 3.645 \\ 3.778 \\ 4.011 \\ 4.067 \\ 3.978 \\ 4.056 \\ 4.044 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.75 & 0.56 & 0.77 & 0.52 & 0.52 & 0.38 \\ & 1 & 0.74 & 0.77 & 0.69 & 0.60 & 0.36 \\ & & 1 & 0.85 & 0.48 & 0.27 & 0.00 \\ & & & 1 & 0.75 & 0.63 & 0.44 \\ & & & & 1 & 0.93 & 0.82 \\ & & & & & 1 & 0.94 \\ & & & & & & 1 \end{pmatrix}$

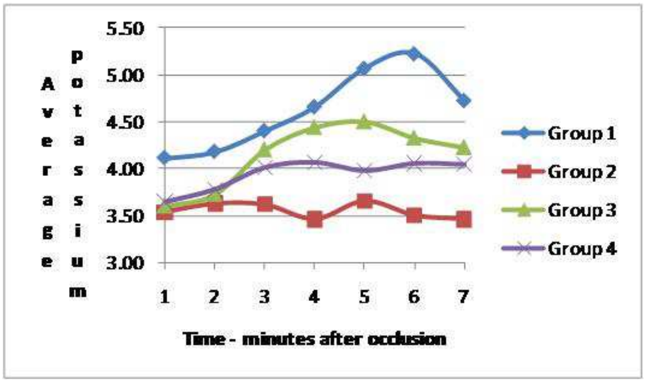

Figure 1. Sample means

5.1 Proportional covariance matrices

Proportional covariance matrices, i.e. $\mathbf{\Sigma}_i = c_i \mathbf{\Sigma}_1, \; i = 2,3,4$ where $\mathbf{\Sigma}_i = \boldsymbol{m}_{i2} - \boldsymbol{\mu}_i \boldsymbol{\mu}_i', \; i = 1,2,3,4$ require the constraints:

$$\boldsymbol{g}_1(\boldsymbol{m}) = \begin{pmatrix} vec(\mathbf{\Sigma}_2 - c_2\mathbf{\Sigma}_1) \\ vec(\mathbf{\Sigma}_3 - c_3\mathbf{\Sigma}_1) \\ vec(\mathbf{\Sigma}_4 - c_4\mathbf{\Sigma}_1) \end{pmatrix} \text{ where } c_i = \frac{\sigma_{11}^i}{\sigma_{11}^1} \tag{23}$$

and σ_{11}^i the variance of the first measurement in the i-th sample. Then

$$\boldsymbol{G}_{1m} = \begin{pmatrix} \frac{\partial vec(\mathbf{\Sigma}_2)}{\partial \boldsymbol{m}} - c_2 \frac{\partial vec(\mathbf{\Sigma}_1)}{\partial \boldsymbol{m}} - vec(\mathbf{\Sigma}_1)\frac{\partial c_2}{\partial \boldsymbol{m}} \\ \frac{\partial vec(\mathbf{\Sigma}_3)}{\partial \boldsymbol{m}} - c_3 \frac{\partial vec(\mathbf{\Sigma}_1)}{\partial \boldsymbol{m}} - vec(\mathbf{\Sigma}_1)\frac{\partial c_3}{\partial \boldsymbol{m}} \\ \frac{\partial vec(\mathbf{\Sigma}_4)}{\partial \boldsymbol{m}} - c_4 \frac{\partial vec(\mathbf{\Sigma}_1)}{\partial \boldsymbol{m}} - vec(\mathbf{\Sigma}_1)\frac{\partial c_4}{\partial \boldsymbol{m}} \end{pmatrix} \tag{24}$$

where $\dfrac{\partial c_i}{\partial \boldsymbol{m}} = \dfrac{\partial c_i}{\partial \boldsymbol{m_\Sigma}} \times \dfrac{\partial \boldsymbol{m_\Sigma}}{\partial \boldsymbol{m}}$, $\dfrac{\partial \boldsymbol{m_\Sigma}}{\partial \boldsymbol{m}}$ and $\dfrac{\partial vec(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{m}}$ are given in (19),

$$\begin{pmatrix} \dfrac{\partial c_2}{\partial \boldsymbol{m_\Sigma}} \\ \dfrac{\partial c_3}{\partial \boldsymbol{m_\Sigma}} \\ \dfrac{\partial c_4}{\partial \boldsymbol{m_\Sigma}} \end{pmatrix} = \begin{pmatrix} \dfrac{\sigma^2_{11}}{(\sigma^1_{11})^2}\boldsymbol{I}_{p^2}[1,] & \dfrac{1}{\sigma^1_{11}}\boldsymbol{I}_{p^2}[1,] & \boldsymbol{0}_{1\times p^2} & \boldsymbol{0}_{1\times p^2} \\ \dfrac{\sigma^3_{11}}{(\sigma^1_{11})^2}\boldsymbol{I}_{p^2}[1,] & \boldsymbol{0}_{1\times p^2} & \dfrac{1}{\sigma^1_{11}}\boldsymbol{I}_{p^2}[1,] & \boldsymbol{0}_{1\times p^2} \\ \dfrac{\sigma^4_{11}}{(\sigma^1_{11})^2}\boldsymbol{I}_{p^2}[1,] & \boldsymbol{0}_{1\times p^2} & \boldsymbol{0}_{1\times p^2} & \dfrac{1}{\sigma^1_{11}}\boldsymbol{I}_{p^2}[1,] \end{pmatrix} \tag{25}$$

and $\boldsymbol{I}_{p^2}[1,]$ denotes the first row of $\boldsymbol{I}_{p^2}$.

In Table 4, the results for the hypothesis of proportional covariance matrices are given, with no constraints imposed on the mean vectors. Note that proportionality of covariance matrices implies equal correlation matrices. One concludes that the proportional model holds for this data.

Table 4 Proportional covariance matrices

$\boldsymbol{\mu}_i$ unrestricted $H_0 : \boldsymbol{\Sigma}_2 = c_2\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_3 = c_3\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_4 = c_4\boldsymbol{\Sigma}_1$
$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 0.25 & 0.19 & 0.19 & 0.21 & 0.23 & 0.22 & 0.21 \\ & 0.18 & 0.19 & 0.19 & 0.21 & 0.19 & 0.18 \\ & & 0.37 & 0.34 & 0.27 & 0.20 & 0.20 \\ & & & 0.46 & 0.41 & 0.29 & 0.26 \\ & & & & 0.53 & 0.41 & 0.36 \\ & & & & & 0.50 & 0.44 \\ & & & & & & 0.49 \end{pmatrix}$
$\widehat{c}_2 = 0.61, \widehat{c}_3 = 1.29, \widehat{c}_4 = 0.87$
$\widehat{\boldsymbol{P}}_i = \begin{pmatrix} 1 & 0.88 & 0.62 & 0.62 & 0.64 & 0.62 & 0.60 \\ & 1 & 0.72 & 0.65 & 0.67 & 0.62 & 0.61 \\ & & 1 & 0.81 & 0.62 & 0.46 & 0.48 \\ & & & 1 & 0.84 & 0.61 & 0.54 \\ & & & & 1 & 0.79 & 0.72 \\ & & & & & 1 & 0.89 \\ & & & & & & 1 \end{pmatrix}$ $i = 1, 2, 3, 4$
Wald $= 49.965, \nu = 91,$ p-value $= 0.9999$

5.2 Behrens-Fisher with proportional covariance matrices

To simultaneously estimate the common mean vector and test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = \boldsymbol{\mu},$ the constraints implied and matrix of derivatives are:

$$\boldsymbol{g}_2(\boldsymbol{m}) = \begin{pmatrix} \boldsymbol{\mu}_2 - \boldsymbol{\mu} \\ \boldsymbol{\mu}_3 - \boldsymbol{\mu} \\ \boldsymbol{\mu}_4 - \boldsymbol{\mu} \end{pmatrix} \quad \boldsymbol{G}_{2m} = \boldsymbol{G}_{2t} = \frac{\partial \boldsymbol{g}_2(\boldsymbol{m})}{\partial \boldsymbol{m}} = \begin{pmatrix} \frac{\partial \boldsymbol{\mu}_2}{\partial \boldsymbol{m}} \\ \frac{\partial \boldsymbol{\mu}_3}{\partial \boldsymbol{m}} \\ \frac{\partial \boldsymbol{\mu}_4}{\partial \boldsymbol{m}} \end{pmatrix} - \begin{pmatrix} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}} \\ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}} \\ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}} \end{pmatrix} \tag{26}$$

where $\dfrac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{m}}$ and $\dfrac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}}$ are given in (13).

The constraints $\boldsymbol{g}_1(\boldsymbol{m})$ and $\boldsymbol{g}_2(\boldsymbol{m})$ can be imposed simultaneously:

$$\boldsymbol{g}(\boldsymbol{m}) = \begin{pmatrix} \boldsymbol{g}_1(\boldsymbol{m}) \\ \boldsymbol{g}_2(\boldsymbol{m}) \end{pmatrix} \quad \text{and} \quad \boldsymbol{G}_m = \begin{pmatrix} \boldsymbol{G}_{1m} \\ \boldsymbol{G}_{2m} \end{pmatrix}. \tag{27}$$

In Table 5 Behrens-Fisher is illustrated under the assumption of proportional covariance matrices. The model does not fit well. For this example, for the sake of interest, the standard error of the estimates, $\widehat{\boldsymbol{\sigma}}(f(\widehat{\boldsymbol{m}}))$ are given in Table 6.

Table 5. Behrens-Fisher and proportional covariance matrices

$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = \boldsymbol{\mu}$ $H_0 : \boldsymbol{\Sigma}_2 = c_2\boldsymbol{\Sigma}_1,\ \boldsymbol{\Sigma}_3 = c_3\boldsymbol{\Sigma}_1,\ \boldsymbol{\Sigma}_4 = c_4\boldsymbol{\Sigma}_1$
$\widehat{\boldsymbol{\mu}}' = (3.691,\ 3.794,\ 3.975,\ 4.02,\ 4.16,\ 4.125,\ 3.993)$
$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 0.32 & 0.26 & 0.27 & 0.33 & 0.38 & 0.40 & 0.33 \\ & 0.25 & 0.27 & 0.31 & 0.35 & 0.35 & 0.30 \\ & & 0.53 & 0.55 & 0.49 & 0.45 & 0.40 \\ & & & 0.79 & 0.75 & 0.68 & 0.57 \\ & & & & 0.93 & 0.86 & 0.70 \\ & & & & & 1.05 & 0.86 \\ & & & & & & 0.82 \end{pmatrix}$
$\widehat{c}_2 = 0.60$, $\widehat{c}_3 = 1.10$, $\widehat{c}_4 = 0.71$
$\widehat{\boldsymbol{P}}_i = \begin{pmatrix} 1 & 0.90 & 0.65 & 0.65 & 0.69 & 0.67 & 0.64 \\ & 1 & 0.75 & 0.69 & 0.72 & 0.69 & 0.67 \\ & & 1 & 0.85 & 0.69 & 0.59 & 0.60 \\ & & & 1 & 0.88 & 0.74 & 0.70 \\ & & & & 1 & 0.87 & 0.80 \\ & & & & & 1 & 0.92 \\ & & & & & & 1 \end{pmatrix}$ $i = 1, 2, 3, 4$
Wald $= 274.178$, $\nu = 108$, p-value $= 0$

Table 6 Asymptotic standard errors of the estimated parameters

$\widehat{\sigma}(\widehat{\boldsymbol{\mu}}_i)$	$\widehat{\sigma}(\widehat{\boldsymbol{\Sigma}}_i)$
$\begin{pmatrix} 0.084 \\ 0.074 \\ 0.108 \\ 0.131 \\ 0.143 \\ 0.152 \\ 0.134 \end{pmatrix}$	$\begin{pmatrix} 0.626 & 0.581 & 0.672 & 0.758 & 0.818 & 0.844 & 0.763 \\ & 0.566 & 0.665 & 0.743 & 0.802 & 0.826 & 0.745 \\ & & 0.872 & 0.935 & 0.952 & 0.954 & 0.875 \\ & & & 1.078 & 1.108 & 1.098 & 0.995 \\ & & & & 1.218 & 1.21 & 1.092 \\ & & & & & 1.289 & 1.163 \\ & & & & & & 1.092 \end{pmatrix}$
$\begin{pmatrix} 0.084 \\ 0.074 \\ 0.108 \\ 0.131 \\ 0.143 \\ 0.152 \\ 0.134 \end{pmatrix}$	$\begin{pmatrix} 0.622 & 0.578 & 0.668 & 0.753 & 0.812 & 0.837 & 0.757 \\ & 0.563 & 0.661 & 0.738 & 0.797 & 0.82 & 0.74 \\ & & 0.863 & 0.926 & 0.943 & 0.945 & 0.868 \\ & & & 1.063 & 1.094 & 1.084 & 0.983 \\ & & & & 1.2 & 1.192 & 1.078 \\ & & & & & 1.267 & 1.145 \\ & & & & & & 1.076 \end{pmatrix}$
$\begin{pmatrix} 0.084 \\ 0.074 \\ 0.108 \\ 0.131 \\ 0.143 \\ 0.152 \\ 0.134 \end{pmatrix}$	$\begin{pmatrix} 0.628 & 0.583 & 0.674 & 0.76 & 0.82 & 0.846 & 0.765 \\ & 0.567 & 0.666 & 0.744 & 0.804 & 0.828 & 0.747 \\ & & 0.875 & 0.939 & 0.955 & 0.957 & 0.878 \\ & & & 1.084 & 1.114 & 1.103 & 0.999 \\ & & & & 1.226 & 1.217 & 1.097 \\ & & & & & 1.298 & 1.17 \\ & & & & & & 1.099 \end{pmatrix}$
$\begin{pmatrix} 0.084 \\ 0.074 \\ 0.108 \\ 0.131 \\ 0.143 \\ 0.152 \\ 0.134 \end{pmatrix}$	$\begin{pmatrix} 0.623 & 0.579 & 0.669 & 0.754 & 0.813 & 0.839 & 0.759 \\ & 0.563 & 0.662 & 0.739 & 0.798 & 0.822 & 0.742 \\ & & 0.866 & 0.928 & 0.945 & 0.947 & 0.869 \\ & & & 1.067 & 1.097 & 1.088 & 0.986 \\ & & & & 1.204 & 1.197 & 1.081 \\ & & & & & 1.272 & 1.15 \\ & & & & & & 1.08 \end{pmatrix}$

5.3 Proportional mean vectors and proportional covariance matrices

In Figure 1, the group means are given. No general trend is observed but the mean vectors may also be modelled as being proportional to an average mean vector, $\boldsymbol{\mu}$, but differing with a constant value, i.e. $\boldsymbol{\mu}_i = d_i \boldsymbol{\mu} + a_i \mathbf{1}$, $i = 1, 2, 3, 4$ where $a_i = \mu_{i1} - d_i \mu_1$, $d_i = \dfrac{\mu_{i1} - \mu_{i2}}{\mu_1 - \mu_2}$, μ_{ij} the j-th element of the i-th mean vector $\boldsymbol{\mu}_i$ and μ_j the j-th element of $\boldsymbol{\mu}$.

This is formulated as $\boldsymbol{g}_1(\boldsymbol{m}) = \begin{pmatrix} \boldsymbol{\mu}_2 - d_2\boldsymbol{\mu} - a_2\mathbf{1} \\ \boldsymbol{\mu}_3 - d_3\boldsymbol{\mu} - a_3\mathbf{1} \\ \boldsymbol{\mu}_4 - d_4\boldsymbol{\mu} - a_4\mathbf{1} \end{pmatrix}$ with matrix of derivatives

$$\boldsymbol{G}_{1m} = \begin{pmatrix} \dfrac{\partial \boldsymbol{\mu}_2}{\partial \boldsymbol{m}} - d_2 \dfrac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}} - \boldsymbol{\mu} \dfrac{\partial d_2}{\partial \boldsymbol{m}} - \mathbf{1} \dfrac{\partial a_2}{\partial \boldsymbol{m}} \\ \dfrac{\partial \boldsymbol{\mu}_3}{\partial \boldsymbol{m}} - d_3 \dfrac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}} - \boldsymbol{\mu} \dfrac{\partial d_3}{\partial \boldsymbol{m}} - \mathbf{1} \dfrac{\partial a_3}{\partial \boldsymbol{m}} \\ \dfrac{\partial \boldsymbol{\mu}_4}{\partial \boldsymbol{m}} - d_4 \dfrac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}} - \boldsymbol{\mu} \dfrac{\partial d_4}{\partial \boldsymbol{m}} - \mathbf{1} \dfrac{\partial a_4}{\partial \boldsymbol{m}} \end{pmatrix} \tag{28}$$

where

$$\begin{pmatrix} \dfrac{\partial d_2}{\partial \boldsymbol{m}} \\ \dfrac{\partial d_3}{\partial \boldsymbol{m}} \\ \dfrac{\partial d_4}{\partial \boldsymbol{m}} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{(\mu_1 - \mu_2)} \dfrac{\partial}{\partial \boldsymbol{m}} (\mu_{21} - \mu_{22}) - \dfrac{(\mu_{21} - \mu_{22})}{(\mu_1 - \mu_2)^2} \dfrac{\partial}{\partial \boldsymbol{m}} (\mu_1 - \mu_2) \\ \dfrac{1}{(\mu_1 - \mu_2)} \dfrac{\partial}{\partial \boldsymbol{m}} (\mu_{31} - \mu_{32}) - \dfrac{(\mu_{31} - \mu_{32})}{(\mu_1 - \mu_2)^2} \dfrac{\partial}{\partial \boldsymbol{m}} (\mu_1 - \mu_2) \\ \dfrac{1}{(\mu_1 - \mu_2)} \dfrac{\partial}{\partial \boldsymbol{m}} (\mu_{41} - \mu_{42}) - \dfrac{(\mu_{41} - \mu_{42})}{(\mu_1 - \mu_2)^2} \dfrac{\partial}{\partial \boldsymbol{m}} (\mu_1 - \mu_2) \end{pmatrix} \tag{29}$$

and

$$\begin{pmatrix} \dfrac{\partial a_2}{\partial \boldsymbol{m}} \\[2ex] \dfrac{\partial a_3}{\partial \boldsymbol{m}} \\[2ex] \dfrac{\partial a_4}{\partial \boldsymbol{m}} \end{pmatrix} = \begin{pmatrix} \mathbf{1}\dfrac{\partial \mu_{21}}{\partial \boldsymbol{m}} - d_2 \dfrac{\partial \mu_1}{\partial \boldsymbol{m}} - \mu_1 \dfrac{\partial d_2}{\partial \boldsymbol{m}} \\[2ex] \mathbf{1}\dfrac{\partial \mu_{31}}{\partial \boldsymbol{m}} - d_3 \dfrac{\partial \mu_1}{\partial \boldsymbol{m}} - \mu_1 \dfrac{\partial d_3}{\partial \boldsymbol{m}} \\[2ex] \mathbf{1}\dfrac{\partial \mu_{41}}{\partial \boldsymbol{m}} - d_4 \dfrac{\partial \mu_1}{\partial \boldsymbol{m}} - \mu_1 \dfrac{\partial d_4}{\partial \boldsymbol{m}} \end{pmatrix} \tag{30}$$

with $\dfrac{\partial \mu_{ij}}{\partial \boldsymbol{m}} = \dfrac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{m}}[j,]$ the j-th row of $\dfrac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{m}}$ (cf. (13)), $\dfrac{\partial \mu_j}{\partial \boldsymbol{m}} = \dfrac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}}[j,]$ the j-th row of $\dfrac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}}$ (cf. (13)).

Results are given in Table 7. The smoothing effect of this model, which fits quite well, is illustrated in Figure 2.

This solution follows the trends observed in the data very well and provides the mles which are obtained by direct modelling of the mean vectors and covariance matrices.

Table 7 Proportional mean vectors and proportional covariance matrices

$H_0 : \boldsymbol{\mu}_2 = d_2\boldsymbol{\mu} + a_2\mathbf{1},\ \boldsymbol{\mu}_3 = d_3\boldsymbol{\mu} + a_3\mathbf{1},\ \boldsymbol{\mu}_4 = d_4\boldsymbol{\mu} + a_4\mathbf{1}$

$H_0 : \boldsymbol{\Sigma}_2 = c_2\boldsymbol{\Sigma}_1,\ \boldsymbol{\Sigma}_3 = c_3\boldsymbol{\Sigma}_1,\ \boldsymbol{\Sigma}_4 = c_4\boldsymbol{\Sigma}_1$

$\widehat{\boldsymbol{\mu}}' = (3.765,\ 3.829,\ 4.012,\ 4.142,\ 4.304,\ 4.333,\ 4.13)$

$\widehat{d}_2 = 0.186, \widehat{d}_3 = 1.294, \widehat{d}_4 = 0.562, \widehat{a}_2 = 2.875,\ \widehat{a}_3 = -1.218,\ \widehat{a}_4 = 1.673$

$$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 0.24 & 0.17 & 0.16 & 0.19 & 0.21 & 0.21 & 0.19 \\ & 0.17 & 0.17 & 0.17 & 0.19 & 0.17 & 0.16 \\ & & 0.36 & 0.33 & 0.26 & 0.18 & 0.19 \\ & & & 0.46 & 0.39 & 0.28 & 0.26 \\ & & & & 0.50 & 0.38 & 0.34 \\ & & & & & 0.49 & 0.43 \\ & & & & & & 0.48 \end{pmatrix}$$

$\widehat{c}_2 = 0.70,\ \widehat{c}_3 = 1.35,\ \widehat{c}_4 = 0.94,\ \widehat{e}_i = 0.71$

$$\widehat{\boldsymbol{P}}_i = \begin{pmatrix} 1.00 & 0.87 & 0.56 & 0.56 & 0.62 & 0.61 & 0.57 \\ & 1.00 & 0.68 & 0.59 & 0.66 & 0.59 & 0.58 \\ & & 1.00 & 0.80 & 0.60 & 0.42 & 0.47 \\ & & & 1.00 & 0.82 & 0.59 & 0.55 \\ & & & & 1.00 & 0.77 & 0.70 \\ & & & & & 1.00 & 0.89 \\ & & & & & & 1.00 \end{pmatrix}$$

$i = 1, 2, 3, 4$

Wald $= 95.961$, $\nu = 105$, p-value $= 0.7245$

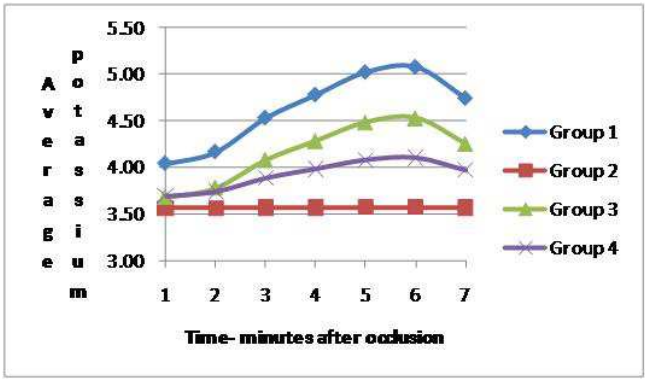

Figure 2. Proportional means assuming proportional covariance matrices.

5.4 Proportional mean vectors and proportional covariance matrices with Toeplitz correlation structure

As a further illustration of the flexibility of the procedure, proportionality of elements along the diagonals of the covariance matrix within each population is imposed additionally. Let the matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_3$ select from $vec(\boldsymbol{\Sigma}_i)$ the elements of $\boldsymbol{\Sigma}_i$ as indicated by the upper and lower triangles ($\boldsymbol{C}_1^*$ and $\boldsymbol{C}_3^*$ respectively) in Figure 3.

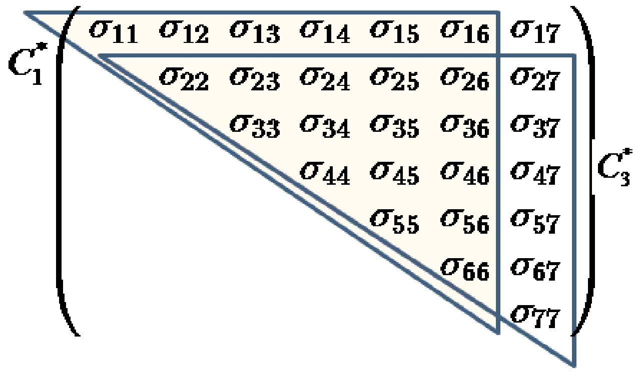

Figure 3. Proportional submatrices.

Impose the additional constraints:

$$\boldsymbol{g}_3(\boldsymbol{m}) = \begin{pmatrix} \boldsymbol{C}_3 vec(\boldsymbol{\Sigma}_1) - e_1 \boldsymbol{C}_1 vec(\boldsymbol{\Sigma}_1) \\ \boldsymbol{C}_3 vec(\boldsymbol{\Sigma}_2) - e_2 \boldsymbol{C}_1 vec(\boldsymbol{\Sigma}_2) \\ \boldsymbol{C}_3 vec(\boldsymbol{\Sigma}_3) - e_3 \boldsymbol{C}_1 vec(\boldsymbol{\Sigma}_3) \\ \boldsymbol{C}_3 vec(\boldsymbol{\Sigma}_4) - e_4 \boldsymbol{C}_1 vec(\boldsymbol{\Sigma}_4) \end{pmatrix} \quad \text{with} \;\; e_i = \frac{\sigma_{22}^i}{\sigma_{11}^i}, \; i = 1, 2, 3, 4. \tag{31}$$

The required matrix of derivatives is

$$\boldsymbol{G}_{3m} = \frac{\partial \boldsymbol{g}_3(\boldsymbol{m})}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \times \frac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}} \text{ where } \frac{\partial \boldsymbol{g}_3(\boldsymbol{m})}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} = \begin{pmatrix} \boldsymbol{C}_3 - e_1\boldsymbol{C}_1 - \boldsymbol{C}_1 vec(\boldsymbol{\Sigma}_1)\dfrac{\partial e_1}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \\ \boldsymbol{C}_3 - e_2\boldsymbol{C}_1 - \boldsymbol{C}_1 vec(\boldsymbol{\Sigma}_2)\dfrac{\partial e_2}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \\ \boldsymbol{C}_3 - e_3\boldsymbol{C}_1 - \boldsymbol{C}_1 vec(\boldsymbol{\Sigma}_3)\dfrac{\partial e_3}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \\ \boldsymbol{C}_3 - e_4\boldsymbol{C}_1 - \boldsymbol{C}_1 vec(\boldsymbol{\Sigma}_4)\dfrac{\partial e_4}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \end{pmatrix} \tag{32}$$

with $\dfrac{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}}{\partial \boldsymbol{m}}$ given in (14) and

$$\begin{pmatrix} \dfrac{\partial e_1}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \\ \dfrac{\partial e_2}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \\ \dfrac{\partial e_3}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \\ \dfrac{\partial e_4}{\partial \boldsymbol{m}_{\boldsymbol{\Sigma}}} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma_{11}^1}\boldsymbol{I}_{p^2}[p+2,] - \dfrac{\sigma_{22}^1}{(\sigma_{11}^1)^2}\boldsymbol{I}_{p^2}[1,] \\ \dfrac{1}{\sigma_{11}^2}\boldsymbol{I}_{p^2}[p+2,] - \dfrac{\sigma_{22}^2}{(\sigma_{11}^2)^2}\boldsymbol{I}_{p^2}[1,] \\ \dfrac{1}{\sigma_{11}^3}\boldsymbol{I}_{p^2}[p+2,] - \dfrac{\sigma_{22}^3}{(\sigma_{11}^3)^2}\boldsymbol{I}_{p^2}[1,] \\ \dfrac{1}{\sigma_{11}^4}\boldsymbol{I}_{p^2}[p+2,] - \dfrac{\sigma_{22}^4}{(\sigma_{11}^4)^2}\boldsymbol{I}_{p^2}[1,] \end{pmatrix}. \tag{33}$$

Note that σ_{22}^i is the element in row $p+2$ of $vec(\boldsymbol{\Sigma}_i)$. This yields a Toeplitz correlation structure common to all four populations, i.e.

$$\boldsymbol{P}_i = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 & \rho_5 & \rho_6 \\ & 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 & \rho_5 \\ & & 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 \\ & & & 1 & \rho_1 & \rho_2 & \rho_3 \\ & & & & 1 & \rho_1 & \rho_2 \\ & & & & & 1 & \rho_1 \\ & & & & & & 1 \end{pmatrix}. \tag{34}$$

Proportionality of these submatrices has the implication that proportionality of the elements on the main diagonal and the first row of each population covariance matrix is now sufficient to specify the full set of required constraints. The constraint $\boldsymbol{g}_1(\boldsymbol{m})$ is adjusted by pre-multiplying each $vec(\boldsymbol{\Sigma}_i)$ with a matrix $\boldsymbol{B}$ which selects only the main diagonal elements σ_{jj} and elements in the first row σ_{1j}, $j = 1, 2, ...p$ from the respective covariance matrices.

Then

$$\boldsymbol{g}_1(\boldsymbol{m}) = \begin{pmatrix} \boldsymbol{B}vec(\boldsymbol{\Sigma}_2) - c_2\boldsymbol{B}vec(\boldsymbol{\Sigma}_1) \\ \boldsymbol{B}vec(\boldsymbol{\Sigma}_3) - c_3\boldsymbol{B}vec(\boldsymbol{\Sigma}_1) \\ \boldsymbol{B}vec(\boldsymbol{\Sigma}_4) - c_4\boldsymbol{B}vec(\boldsymbol{\Sigma}_1) \end{pmatrix},$$

$$\boldsymbol{G}_{1m} = \begin{pmatrix} \boldsymbol{B}\frac{\partial vec(\boldsymbol{\Sigma}_2)}{\partial \boldsymbol{m}} - c_2\boldsymbol{B}\frac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} - \boldsymbol{B}vec(\boldsymbol{\Sigma}_1)\frac{\partial c_2}{\partial \boldsymbol{m}} \\ \boldsymbol{B}\frac{\partial vec(\boldsymbol{\Sigma}_3)}{\partial \boldsymbol{m}} - c_3\boldsymbol{B}\frac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} - \boldsymbol{B}vec(\boldsymbol{\Sigma}_1)\frac{\partial c_3}{\partial \boldsymbol{m}} \\ \boldsymbol{B}\frac{\partial vec(\boldsymbol{\Sigma}_4)}{\partial \boldsymbol{m}} - c_4\boldsymbol{B}\frac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} - \boldsymbol{B}vec(\boldsymbol{\Sigma}_1)\frac{\partial c_4}{\partial \boldsymbol{m}} \end{pmatrix} \quad (35)$$

where $\frac{\partial vec(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{m}}$ and $\frac{\partial c_i}{\partial \boldsymbol{m}}$ are given in (14) and (25) respectively.

The constraints $\boldsymbol{g}(\boldsymbol{m}) = \begin{pmatrix} \boldsymbol{g}_1(\boldsymbol{m}) \\ \boldsymbol{g}_2(\boldsymbol{m}) \\ \boldsymbol{g}_3(\boldsymbol{m}) \end{pmatrix}$ with $\boldsymbol{G}_m = \begin{pmatrix} \boldsymbol{G}_{1m} \\ \boldsymbol{G}_{2m} \\ \boldsymbol{G}_{3m} \end{pmatrix}$ are imposed simultaneously.

The model (cf. Table 8) does not fit very well as is to be expected on inspection of the sample correlation matrices (cf. Table 3). However, it serves to illustrate how more complicated structures can be modelled easily.

Table 8 Proportional mean vectors and proportional covariance matrices with Toeplitz correlation structure

$H_0{:}\boldsymbol{\mu}_2 = d_2\boldsymbol{\mu}+a_2\mathbf{1},\ \boldsymbol{\mu}_3 = d_3\boldsymbol{\mu}+a_3\mathbf{1},\ \boldsymbol{\mu}_4 = d_4\boldsymbol{\mu}+a_4\mathbf{1}$ $H_0 : \boldsymbol{\Sigma}_2 = c_2\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_3 = c_3\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_4 = c_4\boldsymbol{\Sigma}_1$ $H_0 : \boldsymbol{C}_3^{i*} = e_i\boldsymbol{C}_1^{i*}, i = 1, 2, 3, 4$
$\widehat{\boldsymbol{\mu}}' = (3.744,\ 3.813,\ 4.0168,\ 4.155,\ 4.291,\ 4.322,\ 4.134)$ $\widehat{d}_2$=0.0237, $\widehat{d}_3$=1.471, $\widehat{d}_4$=0.709, $\widehat{a}_2$=3.473, $\widehat{a}_3$=−1.828, $\widehat{a}_4$=1.039
$$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 0.28 & 0.24 & 0.20 & 0.19 & 0.22 & 0.25 & 0.23 \\ & 0.32 & 0.27 & 0.23 & 0.22 & 0.25 & 0.28 \\ & & 0.36 & 0.31 & 0.26 & 0.25 & 0.28 \\ & & & 0.41 & 0.35 & 0.29 & 0.29 \\ & & & & 0.47 & 0.40 & 0.33 \\ & & & & & 0.53 & 0.45 \\ & & & & & & 0.61 \end{pmatrix}$$
$\widehat{c}_2 = 0.58, \widehat{c}_3 = 1.29, \widehat{c}_4 = 0.78, \widehat{e}_i = 1.14$
$$\widehat{\boldsymbol{P}}_i = \begin{pmatrix} 1.00 & 0.80 & 0.63 & 0.57 & 0.60 & 0.64 & 0.57 \\ & 1.00 & 0.80 & 0.63 & 0.57 & 0.60 & 0.64 \\ & & 1.00 & 0.80 & 0.63 & 0.57 & 0.60 \\ & & & 1.00 & 0.80 & 0.63 & 0.57 \\ & & & & 1.00 & 0.80 & 0.63 \\ & & & & & 1.00 & 0.80 \\ & & & & & & 1.00 \end{pmatrix}$$ $i = 1, 2, 3, 4$
Wald $= 148.441$, $\nu = 119$, p-value $= 0.0349$

6. Conclusion

Although some problems of this nature can be solved numerically within the mixed model framework, we believe that the approach proposed in this paper provides a natural and elegant way of analysing problems of this kind. A solution is provided to maximum likelihood estimation problems where normally the likelihood equations do not exist or are hard to derive and solve. Mean and covariance structures are modelled directly. A wider class of problems can be addressed in this way. One is not restricted to a fixed set of patterned covariance structures. Proportionality of variances and/or covariance matrices are easily accommodated. The procedure has the potential to provide maximum likelihood estimates in cases where the traditional approach through likelihood equations and maximisation thereof are intractable and complicated. The potential for extension of this procedure to other multivariate structures, e.g. MANOVA, multivariate longitudinal data, dependent multivariate normal populations and other distributions in the exponential class is obvious.

References

BARNDORFF-NIELSEN, O. (1982). *Exponential families. In: Kotz, S., Johnson, N.L. (eds.). Encyclopedia of Statistical Sciences*, **2**, 587–596, New York: John Wiley & Sons.

BROWN, L.D. (1986). Fundamentals of exponential families with applications in statistical decision theory. *Institute of Mathematical Statistics, Lecture Notes-Monograph Series*, **9**.

BUOT, M-L.G., HOSTEN, S. & RICHARDS, D.S.P. (2007). Counting and locating the solutions of polynomial systems of maximum likelihood equations, II: The Behrens-Fisher problem. *Statist. Sinica*, **17**, 1343–1354.

ERIKSEN, P.S. (1987). Proportionality of covariance matrices. *The Annals of Statistics*, **15**(2), 732–748.

FLURY, B.K. (1986). Proportionality of k covariance matrices. *Statistics & Probability Letters* **4**, 29–33.

GRIZZLE, J.E. & ALLEN, D.M. (1969). Analysis of growth and dose response curves. *Biometrics*, **25**(2), 357–381.

JENSEN, S.T. & JOHANSEN, S. (1987). Estimation of proportional covariances. *Statistics & Probability Letters* **6**,83–85.

JENSEN, S.T. & MADSEN, J. (2004). Estimation of proportional covariances in the presence of certain linear restrictions. *The Annals of Statistics*, **32**(1), 219–232.

JOHNSON, R.A. & WICHERN, D.W. (1982). *Applied multivariate statistical analysis*. Englewood Cliffs, N.J.: Prentice Hall.

MANLY, B.F.J. & RAYNER, J.C.W. (1987). The comparison of sample covariance matrices using likelihood ratio tests. *Biometrika*, **74**(4), 841–847.

MATTHEWS, G.B. & CROWTHER, N.A.S. (1995). A maximum likelihood estimation procedure when modelling in terms of constraints. *South African Statist. J.*, **29**, 29–50.

MATTHEWS, G.B. & CROWTHER, N.A.S. (1998). A maximum likelihood estimation procedure for the generalized linear model with restrictions. *South African Statist. J.*, **32**, 119–144.

MUIRHEAD, R.J. (1982). *Aspects of multivariate statistical theory*. New York: John Wiley & Sons, Inc.

NEL, D.G. & VAN DER MERWE, C.A. (1986). A solution to the multivariate Behrens-Fisher problem. *Commun. Statist. – Theory Meth.*, **15**(12), 3719–3735.

RUKHIN, A.L., BIGGERSTAFF, B.J. & VANGEL, M.G. (2000). Restricted maximum likelihood estimation of a common mean and the Mandel-Paule algorithm. *Journal of Statistical Planning and Inference*., **83**, 319–330.

SAS INSTITUTE INC. (1999). *SAS/STAT User's guide, Version 7*, Cary, NC: SAS Institute Inc.

VERBEKE, G. & MOLENBERGHS, G. (2009). *Linear mixed models for longitudinal data.* New York: Springer-Verlag.

WANG, Y. (1998). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **60**(1), 159–174.

ZHANG, J.T. & XU, J.F. (2009). On the k-sample Behrens-Fisher problem for high-dimensional data. *Science in China Series A: Mathematics*, **52**(6), 1285–1304.

Manuscript received, 2010.11.24, revised, 2011.03.26, accepted, 2011.06.12.