Models with a Kronecker Product Covariance Structure: Estimation and Testing

**Muni S. Srivastava, Tatjana Nahtman,
and Dietrich von Rosen**

Models with a Kronecker Product Covariance Structure: Estimation and Testing

Muni S. Srivastava
Department of Statistics, University of Toronto, Canada

Tatjana Nahtman
Institute of Mathematical Statistics, University of Tartu, Estonia;
Department of Statistics, Stockholm University, Sweden

Dietrich von Rosen[1]
Centre of Biostochastics, Swedish University of Agricultural Sciences, Sweden

Abstract

In this article we consider a pq-dimensional random vector $\underline{x}$ distributed normally with mean vector $\underline{\theta}$ and the covariance matrix Λ, assumed to be positive definite. On the basis of N independent observations on the random vector $\underline{x}$, we wish to estimate parameters and test the hypothesis $H: \Lambda = \Psi \otimes \Sigma$, where $\Psi = (\psi_{ij}) : q \times q$ and $\Sigma = (\sigma_{ij}) : p \times p$, and $\Lambda = (\psi_{ij}\Sigma)$, the Kronecker product of Ψ and Σ. That is instead of $\frac{1}{2}pq(pq+1)$ parameters, it has only $\frac{1}{2}p(p+1) + \frac{1}{2}q(q+1) - 1$ parameters. When this model holds, we test the hypothesis that Ψ is an identity matrix, a diagonal matrix or of intraclass correlation structure. The maximum likelihood estimators (MLE) are obtained under the hypothesis as well as under the alternatives. Using these estimators the likelihood ratio tests (LRT) are obtained. Moreover, it is shown that the estimators are unique.

---

[1]E-mail address to the correspondence author: Dietrich.von.Rosen@bt.slu.se

1 Introduction

When analyzing multivariate data it is often assumed that the m-dimensional random vector $\underline{x}$ is normally distributed with mean vector $\underline{\theta}$ and covariance matrix Λ. In many data analysis, it is often required to assume that Λ has the intraclass correlation structure, that is,

$$\Lambda = \sigma^2[(1 - \rho)I_m + \rho\underline{1}_m\underline{1}_m'],$$

where $-\frac{1}{m-1} < \rho < 1$, $\sigma^2 > 0$, I_m is the $m \times m$ identity matrix, and $\underline{1}_m$ is an m-vector of ones, $\underline{1}_m = (1, \ldots, 1)'$. In other cases it is assumed that Λ has a block compound symmetry structure which, when $m = pq$, can be written as

$$\Lambda = \begin{pmatrix} A & B & \cdots & B \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ B & B & \cdots & A \end{pmatrix},$$

where $A : p \times p$ is a positive definite (written later $A > 0$), and $B = B'$ such that $\Lambda > 0$.

The estimation and testing problems that arise in the intraclass correlation model and compound symmetry models have been considered extensively in the literature, see for example, Wilks (1946), Votaw (1948), Srivastava (1965), Olkin (1973), and Arnold (1973).

However, not much work has been done for a positive definite block covariance matrix Λ, namely when $\Lambda = \Psi \otimes \Sigma$, where $\Psi \otimes \Sigma$ is the Kronecker product of a $q \times q$ matrix $\Psi = (\psi_{ij})$ with a $p \times p$ matrix $\Sigma = (\sigma_{ij})$ given by

$$\Lambda = (\psi_{ij}\Sigma) : pq \times pq, \ m = pq.$$

When Λ is unstructured the model belongs to the exponential family whereas when $\Lambda = \Psi \otimes \Sigma$ it belongs to the curved exponential family. Thus we may expect that estimation and testing are more complicated under the "Kronecker structure" than in the unstructured case.

As an example of a block covariance matrix, consider a p-dimensional random vector $\underline{x}$ representing an observation vector at p time-points on a characteristic of an individual or a subject. If we take observations on q characteristics at p time-points, then these observations can be represented as $\underline{x}_{(1)}, \ldots, \underline{x}_{(q)}$, where $\underline{x}_{(i)}$'s are p-vectors. Since the observations have been taken on the q characteristics of the same individual, $\underline{x}_{(i)}$'s may not be independently distributed. If the mean vector of $\underline{x}_{(i)}$ is $\underline{\mu}_{(i)}$, then we need to define

the parameter

$$cov(\underline{x}_{(i)}, \underline{x}_{(j)}) = E[(\underline{x}_{(i)} - \underline{\mu}_{(i)})(\underline{x}_{(j)} - \underline{\mu}_{(j)})'],$$

$i, j = 1, \ldots, q$, called the covariance between the p-vectors $\underline{x}_{(i)}$ and $\underline{x}_{(j)}$. When we choose

$$cov(\underline{x}_{(i)}, \underline{x}_{(j)}) = (\psi_{ij}\Sigma), \ i, j = 1, \ldots, q \tag{1.1}$$

and assume normality, then the distribution of the pq random vectors

$$(\underline{x}'_{(1)}, \ldots, \underline{x}'_{(q)})' \sim N_{pq}(\underline{\mu}, \Psi \otimes \Sigma), \tag{1.2}$$

where

$$\underline{\mu} = (\underline{\mu}'_1, \ldots, \underline{\mu}'_q)'. \tag{1.3}$$

It may be noted that we have used the standard notation for defining the vectorization of a matrix, namely,

$$vec(\underline{x}_{(1)}, \ldots, \underline{x}_{(q)}) = (\underline{x}'_{(1)}, \ldots, \underline{x}'_{(q)})'.$$

As noted in the literature, see e.g. Galecki (1994) and Naik and Rao (2001), since $(c\Psi) \otimes (c^{-1}\Psi) = \Psi \otimes \Sigma$, all the parameters of Ψ and Σ are not defined uniquely. Thus, without any loss of generality we assume that

$$\psi_{qq} = 1. \tag{1.4}$$

The MLE of Ψ and Σ are not available in the literature. The condition (1.4) or equivalently if we assume that for $\Sigma = (\sigma_{ij}) : p \times p$, $\sigma_{pp} = 1$ instead of $\psi_{qq} = 1$, makes it technically more difficult to obtain the MLE of Ψ and Σ.

To distinguish between different cases we shall write Ψ^* when $\psi_{qq} = 1$ and write Ψ_ρ when $\psi_{ii} = 1$, $i = 1, \ldots, q$. Similarly, we shall write Σ^* when the restriction $\sigma_{pp} = 1$ is imposed, and Ψ remains unrestricted. Naik and Rao (2001) have also considered the problem but did not obtain the MLE of Σ and Ψ^*.

Furthermore, Roy and Khattree (2005) gave many references where $\Psi \otimes \Sigma$ is considered, in particular when Ψ has a compound symmetry structure. In time series analysis (e.g. see Shitan and Brockwell, 1995) one has also considered the "Kronecker structure" but unlike this paper one usually has only 1 observation matrix and hence has to impose various structures on $\Psi \otimes \Sigma$.

In many cases, it is very likely that

$$\psi_{ii} = 1, \ i = 1, \ldots, q, \tag{1.5}$$

2

instead of only $\psi_{qq} = 1$. That is, $\underline{x}_{(1)}, \ldots, \underline{x}_{(q)}$, have the same covariance matrix Σ. When this assumption is made we shall write the covariance matrix of the pq-vector $(\underline{x}'_{(1)}, \ldots, \underline{x}'_{(q)})'$ as

$$\Psi_\rho \otimes \Sigma, \; \Psi_\rho = (\psi_{ij}), \; \text{with } \psi_{ii} = 1, \; i = 1, \ldots, q. \tag{1.6}$$

For estimation and testing we require N iid (independent and identically distributed) observations on the pq-vector $(\underline{x}'_{(1)}, \ldots, \underline{x}'_{(q)})'$. These N observation vectors will be denoted by

$$(\underline{x}'_{(1)j}, \ldots, \underline{x}'_{(q)j})', \; j = 1, \ldots, N. \tag{1.7}$$

Let

$$\begin{aligned}
X_j &= (\underline{x}_{(1)j}, \ldots, \underline{x}_{(q)j}) : p \times q, \; j = 1, \ldots, N, &\tag{1.8} \\
M &= (\underline{\mu}_1, \ldots, \underline{\mu}_q) : p \times q &\tag{1.9}
\end{aligned}$$

and

$$X = (X_1, \ldots, X_N) : p \times qN. \tag{1.10}$$

It has been shown by Srivastava and Khatri (1979, pp.170–171) that

$$vec(X_j) = (\underline{x}'_{(1)j}, \ldots, \underline{x}'_{(q)j})' \sim N_{pq}(vec(M), \Psi^* \otimes \Sigma)$$

if and only if its pdf is given by

$$(2\pi)^{-\frac{1}{2}pq}|\Sigma|^{-\frac{1}{2}q}|\Psi^*|^{-\frac{1}{2}p}etr\{-\frac{1}{2}\Sigma^{-1}(X_j - M)\Psi^{*-1}(X_j - M)'\}, \tag{1.11}$$

where $etr\{A\}$ stands for the exponential of the trace of the matrix A, $tr\{A\} = \sum_{i=1}^p a_{ii}$, $A = (a_{ij})$. Srivastava and Khatri (1979, pp. 54–55, pp. 170–171) used the pdf (1.11) to define the distribution of the random matrix X_j and used the notation

$$X_j \sim N_{p,q}(M, \Sigma, \Psi^*) \tag{1.12}$$

to write the pdf of X_j given in (1.11) which is also the pdf of $vec(X_j)$. We shall follow the same notation.

From the pdf (1.11) it is clear that the role of Σ and Ψ can be interchanged by considering the vectorization of X'_j. For example, from (1.11), the pdf of $vec(X'_j)$ will be given by

$$(2\pi)^{-\frac{1}{2}pq}|\Sigma|^{-\frac{1}{2}q}|\Psi^*|^{-\frac{1}{2}p}etr\{-\frac{1}{2}\Psi^{*-1}(X_j - M)'\Sigma^{-1}(X_j - M)\}, \tag{1.13}$$

and

$$X_j' \sim N_{q,p}(M', \Psi^*, \Sigma). \tag{1.14}$$

Thus, if we wish to test a hypothesis about $\Sigma = (\sigma_{ij})$, we may assume that $\sigma_{pp} = 1$ for the general case and $\sigma_{ii} = 1$, $i = 1, \ldots, p$, for the second case with no restrictions on the elements of Ψ, or, any other representation that may be helpful in estimation and testing problems.

The organization of this paper is as follows. In Section 2, we present a method of estimating M or $\underline{\mu}$, Σ and Ψ_ρ, that is Ψ with diagonal elements equal to one. The maximum likelihood method is rather messy and so we provide a heuristic method in giving consistent estimators of Ψ_ρ and Σ. These estimators, however, may also be useful in taking as the initial values in solving the maximum likelihood equations iteratively for the general case when only $\psi_{qq}=1$, which is done in Section 3. In Section 4, we test the hypothesis that the general $pq \times pq$ covariance matrix $\Lambda = \Psi \otimes \Sigma$ against the alternative that the covariance matrix is not of Kronecker product structure, when $N > pq$. The three other testing problems concerning the Ψ matrix are considered in Sections 5 and 6. The testing problems concerning the means which may follow the growth curve models of Pothoff and Roy (1964) as discussed in the works of Srivastava and Khatri (1979) and Kollo and von Rosen (2005) will be considered in a subsequent communication. In Section 7 a small simulation study is presented and finally in Section 8 fundamental results concerning the uniqueness of the MLEs from sections 3 and 5 are verified.

2 Estimation of M, Σ and Ψ_ρ: heuristic method

Let $X_1, \ldots, X_N$ be iid $N_{p,q}(M, \Sigma, \Psi^*)$, $N > \max(p, q)$. Then the pdf of $X = (X_1, \ldots, X_N)$ is given by

$$(2\pi)^{-\frac{1}{2}pq}|\Sigma|^{-\frac{1}{2}Nq}|\Psi^*|^{-\frac{1}{2}Np}etr\{-\frac{1}{2}\Sigma^{-1}[\sum_{j=1}^{N}(X_j - M)\Psi^{*-1}(X_j - M)']\}. \tag{2.1}$$

Let

$$\bar{X} = \frac{1}{N}\sum_{j=1}^{N} X_j. \tag{2.2}$$

4

Then

$$\sum_{j=1}^{N}(X_j - M)\Psi^{*-1}(X_j - M)' = \sum_{j=1}^{N}(X_j - \bar{X})\Psi^{*-1}(X_j - \bar{X})'$$
$$+ N(\bar{X} - M)\Psi^{*-1}(\bar{X} - M)'. \qquad (2.3)$$

Hence, the maximum likelihood estimator of M is given by

$$\hat{M} = \bar{X}. \qquad (2.4)$$

Next consider the case when the diagonal elements of Ψ are all equal to one, that is, when Ψ is Ψ_ρ given in (1.6). In this case, we can estimate Σ by

$$\tilde{\Sigma} = S = \frac{1}{Nq}\sum_{j=1}^{N}(X_j - \bar{X})(X_j - \bar{X})'. \qquad (2.5)$$

It will be shown later that $(N/n)S$, with $n = N - 1$, is an unbiased and consistent estimator of Σ. Writing

$$\bar{X} = (\bar{\underline{x}}_{(1)}, \ldots, \bar{\underline{x}}_{(q)}),$$

we find that

$$S = \frac{1}{Nq}\sum_{j=1}^{N}\sum_{i=1}^{q}(\underline{x}_{(i)j} - \bar{\underline{x}}_{(i)})(\underline{x}_{(i)j} - \bar{\underline{x}}_{(i)})'. \qquad (2.6)$$

Similarly, we can estimate ψ_{ik}, $i \neq k$, by

$$\tilde{\psi}_{ik} = \frac{1}{Np}\sum_{j=1}^{N}tr\{S^{-1}(\underline{x}_{(i)j} - \bar{\underline{x}}_{(i)})(\underline{x}_{(k)j} - \bar{\underline{x}}_{(k)})'\}. \qquad (2.7)$$

It will be shown that it is a consistent estimator of ψ_{ik}.

To show the unbiasedness of the estimator $(N/n)\tilde{\Sigma}$ and the consistency of $\tilde{\Sigma}$ and $\tilde{\psi}_{ik}$, we proceed as follows. Let Γ' be an $N \times N$ orthogonal matrix with first row as $\underline{1}'_N/\sqrt{N}$ and the ith row given by

$$\underline{g}'_{i-1} = \left(\frac{1}{\sqrt{(i-1)i}}, \ldots, \frac{1}{\sqrt{(i-1)i}}, -\frac{i-1}{\sqrt{(i-1)i}}, 0, \ldots, 0\right), \ i = 2, \ldots, N. (2.8)$$

That is, if

$$G = (\underline{g}_1, \ldots, \underline{g}_n): \ N \times n, \ n = N - 1, \qquad (2.9)$$

5

$$\Gamma = \left(\frac{1_N}{\sqrt{N}}, G\right) \tag{2.10}$$

is an orthogonal matrix of Helmert's type,

$$\Gamma\Gamma' = I_N, \ \underline{1}'_N G = 0, \ G'G = I_{N-1}, \ \text{and} \ GG' = I_N - \frac{1}{N}\underline{1}_N\underline{1}'_N.$$

Since Γ is an orthogonal matrix, it follows that $\Gamma \otimes I_q$ is also an orthogonal matrix. Hence, letting

$$Y \equiv (Y_1, Z) = (X_1, \ldots, X_N)(\Gamma \otimes I_q), \tag{2.11}$$

we find that the transformation from X to Y_1, Z is one-to-one with the Jacobian of the transformation equal to one, where $Y_1 : p \times q$, and $Z : p \times qn$, $n = N - 1$. Furthermore,

$$(X_1, \ldots, X_N)(I_N \otimes \Psi^{-1})(X_1, \ldots, X_N)' = Y(\Gamma' \otimes I_q)(I_N \otimes \Psi_\rho^{-1})(\Gamma \otimes I_q)Y'$$
$$= Y(I_N \otimes \Psi_\rho^{-1})Y' = Y_1 \otimes \Psi^{-1}Y_1' + Z(I_n \otimes \Psi^{-1})Z'. \tag{2.12}$$

It follows from (2.10) that

$$\begin{aligned} Y_1 &= (X_1, \ldots, X_N)(\frac{1_N}{\sqrt{N}} \otimes I_q) = \sqrt{N}\bar{X}, \\ Z &= (Z_1, \ldots, Z_n) = X(G \otimes I_q). \end{aligned} \tag{2.13}$$

Hence,

$$\begin{aligned} \sum_{j=1}^{N}(X_j - \bar{X})\Psi_\rho^{-1}(X_j - \bar{X})' &= \sum_{j=1}^{N}X_j\Psi_\rho^{-1}X_j' - N\bar{X}\Psi_\rho^{-1}\bar{X}' \\ &= X(I_N \otimes \Psi_\rho^{-1})X' - Y_1\Psi_\rho^{-1}Y_1' = Y(I_N \otimes \Psi_\rho^{-1})Y' - Y_1\Psi_\rho^{-1}Y_1' \\ &= Z(I_n \otimes \Psi_\rho^{-1})Z', \end{aligned} \tag{2.14}$$

and

$$N(\bar{X} - M)\Psi_\rho^{-1}(\bar{X} - M)' = (Y_1 - \Theta)\Psi_\rho^{-1}(Y_1 - \Theta)', \ \Theta = \sqrt{N}M. \tag{2.15}$$

Thus,

$$Y_1 \sim N_{p,q}(\Theta, \Sigma, \Psi_\rho)$$

which is independently distributed of $Z_1, \ldots, Z_n$, where

$$Z_j \overset{\text{iid}}{\sim} N_{p,q}(0, \Sigma, \Psi_\rho).$$

6

Let

$$Z_j = (\underline{z}_{(1)j}, \ldots, \underline{z}_{(q)j}) : p \times q, \ j = 1, \ldots, n, \ n = N - 1.$$

Then

$$E[\underline{z}_{(i)j}, \underline{z}'_{(i)j}] = \psi_{\rho ii} \Sigma, \ \text{and} \ E(Z_j Z'_j) = (tr \Psi_\rho) \Sigma.$$

Also, from (2.5)

$$
\begin{aligned}
S &= \frac{1}{Nq} [\sum_{j=1}^{N} X_j X'_j - N \bar{X} \bar{X}'] = \frac{1}{Nq} [XX' - Y_1 Y'_1] \\
&= \frac{1}{Nq} [YY' - Y_1 Y'_1] = \frac{1}{Nq} ZZ' = \frac{1}{Nq} \sum_{j=1}^{n} Z_j Z'_j, \ n = N - 1. \quad (2.16)
\end{aligned}
$$

Hence,

$$\frac{N}{n} E(S) = (\frac{1}{q} tr \Psi_\rho) \Sigma = \Sigma, \ \text{if} \ \psi_{\rho ii} = 1, \ i = 1, \ldots, q.$$

Since $(Z_j Z'_j / q)$ are iid with mean Σ, when $\psi_{\rho ii} = 1$, $i = 1, \ldots, q$, S is a consistent estimator of Σ. Thus, we get the following theorem.

Theorem 2.1. *Let $X_1, \ldots, X_N$ be iid $N_{p,q}(M, \Sigma, \Psi_\rho)$, where $\Psi_\rho = (\psi_{\rho ij})$, $\psi_{\rho ii} = 1$, $i = 1, \ldots, q$. Then,*

$$\frac{N}{n} S = \frac{1}{nq} \sum_{j=1}^{N} (X_j - \bar{X})(X_j - \bar{X})', \ n = N - 1,$$

is an unbiased estimator of Σ as well as consistent, if $N \to \infty$.

Corollary 2.1. *When $\Psi = I$, S is the maximum likelihood estimator of Σ. Similarly, in terms of $\underline{z}_{(i)j}$,*

$$\tilde{\psi}_{\rho ik} = \frac{1}{Np} tr\{S^{-1} \sum_{j=1}^{n} \underline{z}_{(i)j} \underline{z}'_{(k)j}\}, \qquad i \neq k.$$

Since, when $\psi_{\rho ii} = 1$, $i = 1, \ldots, q$, $S \to \Sigma$ in probability and $\Sigma^{-\frac{1}{2}} \underline{z}_{(k)j} \underline{z}'_{(i)j} \Sigma^{-\frac{1}{2}}$, $j = 1, \ldots, N$, are independently distributed with means $\psi_{\rho ik} I_p$, it follows that $\tilde{\psi}_{\rho ik}$ are consistent estimators of ψ_{ik} when all the diagonal elements of Ψ_ρ are all one. Thus, we have the following theorem.

Theorem 2.2. *Let $X_1, \ldots, X_N$ be iid $N_{p,q}(M, \Sigma, \Psi_\rho)$, where Ψ_ρ has all the diagonal elements equal to one. Then, $\tilde{\psi}_{\rho ik}$ defined in (2.7) is a consistent estimator of $\psi_{\rho ik}$, $i \neq k$, $(i, k) = 1, \ldots, q$.*

3 Maximum Likelihood Estimators of M, Σ and Ψ^* $(\psi_{qq} = 1)$

We consider the same model as in Section 2 except that now the $q \times q$ matrix Ψ^* is of the general form. That is, for $\Psi^* = (\psi_{ij})$, we only assume that $\psi_{qq} = 1$. The MLE of M remains the same as in (2.4) and therefore we start with the likelihood

$$(2\pi)^{-\frac{1}{2}pqN}|\Sigma|^{-\frac{1}{2}qN}|\Psi^*|^{-\frac{1}{2}pN}etr\{-\tfrac{1}{2}\sum_{i=1}^{N}\Sigma^{-1}X_{ic}\Psi^{*-1}X_{ic}'\}, \qquad (3.1)$$

where

$$X_{ic} = X_i - \bar{X}, \; i = 1, \ldots, N. \qquad (3.2)$$

Due to the constraint $\psi_{qq} = 1$, it needs special attention. Let

$$\Psi^* = \begin{pmatrix} \Psi_{11} & \underline{\psi}_{1q} \\ \underline{\psi}_{1q}' & \psi_{qq} \end{pmatrix}, \; \Psi_{11} : (q-1) \times (q-1). \qquad (3.3)$$

From Srivastava and Khatri (1979, Corollary 1.4.2 (i), p. 8), it follows since $\psi_{qq} = 1$, that

$$\Psi^{*-1} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} I_{q-1} \\ -\underline{\psi}_{1q}' \end{pmatrix} \Psi_{1\bullet q}^{-1} \left(I_{q-1} : \; -\underline{\psi}_{1q} \right),$$

where

$$\Psi_{1\bullet q} = \Psi_{11} - \underline{\psi}_{1q}\underline{\psi}_{1q}' : \; (q-1) \times (q-1). \qquad (3.4)$$

Moreover,

$$|\Psi^*| = |\Psi_{1\bullet q}| = |\Psi_{11} - \underline{\psi}_{1q}\underline{\psi}_{1q}'|.$$

Thus (3.1) equals

$$(2\pi)^{-\frac{1}{2}pqN}|\Sigma|^{-\frac{1}{2}qN}|\Psi_{1\bullet q}|^{-\frac{1}{2}pN} \times$$

$$etr\{-\tfrac{1}{2}\sum_{i=1}^{N}\Sigma^{-1}(X_{icq}X_{icq}' + X_{ic}\begin{pmatrix} I_{q-1} \\ -\underline{\psi}_{1q}' \end{pmatrix} \Psi_{1\bullet q}^{-1} \left(I_{q-1} : \; -\underline{\psi}_{1q} \right) X_{ic}')\},$$

where $X_{ic} = (X_{ic1} : X_{icq})$: $(p \times (q-1) : p \times 1)$. By differentiation with respect to the upper triangle of Σ^{-1} and $\Psi_{1 \bullet q}^{-1}$ as well as differentiation with respect to $\underline{\psi}_{1q}$ we obtain after some manipulations

$$Nq\widehat{\Sigma} = \sum_{i=1}^{N} (X_{icq}X_{icq}' + X_{ic} \begin{pmatrix} I_{q-1} \\ -\widehat{\underline{\psi}}_{1q}' \end{pmatrix} \widehat{\Psi}_{1\bullet q}^{-1} \begin{pmatrix} I_{q-1} & -\widehat{\underline{\psi}}_{1q} \end{pmatrix} X_{ic}'), \quad (3.5)$$

$$Np\widehat{\Psi}_{1\bullet q} = \sum_{i=1}^{N} \begin{pmatrix} I_{q-1} : & -\widehat{\underline{\psi}}_{1q} \end{pmatrix} X_{ic}' \widehat{\Sigma}^{-1} X_{ic} \begin{pmatrix} I_{q-1} \\ -\widehat{\underline{\psi}}_{1q}' \end{pmatrix} \quad (3.6)$$

and

$$\widehat{\underline{\psi}}_{1q} = \sum_{i=1}^{N} X_{ic1}' \widehat{\Sigma}^{-1} X_{icq} (\sum_{i=1}^{N} X_{icq}' \widehat{\Sigma}^{-1} X_{icq})^{-1}. \quad (3.7)$$

We first show that the scalar quantity in (3.7),

$$\sum_{i=1}^{N} X_{icq}' \widehat{\Sigma}^{-1} X_{icq} = Np. \quad (3.8)$$

To prove (3.8), we post-multiply (3.5) by $\widehat{\Sigma}^{-1}$ and take the trace. This gives

$$
\begin{aligned}
Nqp &= \sum_{i=1}^{N} X_{icq}' \widehat{\Sigma}^{-1} X_{icq} \\
&\quad + tr\{\sum_{i=1}^{N} \widehat{\Psi}_{1\bullet q}^{-1} \begin{pmatrix} I_{q-1} : & -\widehat{\underline{\psi}}_{1q} \end{pmatrix} (X_{ic}' \widehat{\Sigma}^{-1} X_{ic} \begin{pmatrix} I_{q-1} \\ -\widehat{\underline{\psi}}_{1q}' \end{pmatrix})\} \\
&= \sum_{i=1}^{N} X_{icq}' \widehat{\Sigma}^{-1} X_{icq} + Np\, tr\{I_{q-1}\},
\end{aligned}
$$

using (3.6). This proves (3.8). Thus, we get

$$\widehat{\underline{\psi}}_{1q} = \frac{1}{Np} \sum_{i=1}^{N} X_{ic1}' \widehat{\Sigma}^{-1} X_{icq}. \quad (3.9)$$

Next, we simplify (3.6). Using (3.8) we get after some calculations

$$Np\widehat{\Psi}_{1\bullet q} = \sum_{i=1}^{N} X_{ic1}' \widehat{\Sigma}^{-1} X_{ic1} - Np\underline{\widehat{\psi}}_{1q}\underline{\widehat{\psi}}_{1q}'. \quad (3.10)$$

9

Thus,

$$Np\widehat{\Psi}_{11} = \sum_{i=1}^{N} X'_{ic1}\widehat{\Sigma}^{-1}X_{ic1}.$$

Hence, using (3.8), we get

$$
\begin{aligned}
\widehat{\Psi} &= \begin{pmatrix} \widehat{\Psi}_{11} & \widehat{\underline{\psi}}_{1q} \\ \widehat{\underline{\psi}}'_{1q} & 1 \end{pmatrix} \\
&= \frac{1}{Np}\begin{pmatrix} \sum_{i=1}^{N} X'_{ic1}\widehat{\Sigma}^{-1}X_{ic1} & \sum_{i=1}^{N} X'_{ic1}\widehat{\Sigma}^{-1}X_{icq} \\ \sum_{i=1}^{N} X'_{icq}\widehat{\Sigma}^{-1}X_{ic1} & \sum_{i=1}^{N} X'_{icq}\widehat{\Sigma}^{-1}X_{icq} \end{pmatrix} \\
&= \frac{1}{Np}\sum_{i=1}^{N} X'_{ic}\widehat{\Sigma}^{-1}X_{ic}.
\end{aligned}
$$

(3.11)

(3.12)

With $\widehat{\Psi}$ defined in (3.11), we can rewrite (3.5) as

$$\widehat{\Sigma} = \frac{1}{Nq}\sum_{i=1}^{N} X_{ic}\widehat{\Psi}^{-1}X'_{ic}. \tag{3.13}$$

Thus, we solve (3.12) and (3.13) directly subject to the condition (3.8). This we shall call the "flip-flop" algorithm. The starting value of $\widehat{\Sigma}$ can be based on the estimators obtained in Section 2 and satisfying (3.8).

The next theorem which is proven in Section 8 provides us with an important result concerning the flip-flop algorithm and the MLEs. For related works we refer to Lu and Zimmerman (2005) and Dutilleul (1999), where also other references are given.

Theorem 3.1. *Let $\widehat{\Sigma}$ and $\widehat{\Psi}$ satisfy the flip-flop algorithm given in (3.12) and (3.13), satisfying the condition (3.8). If $N > \max(p, q)$, then there is only one solution and the MLEs based on the likelihood given in (3.1) are unique.*

The theorem is interesting because in general uniqueness of MLEs in curved exponential families may not hold. For example, see Drton (2006) where the SUR-model is discussed.

4 Testing that the Kronecker model holds

Let $X_1, \ldots, X_N$ be iid normally distributed $p \times q$ matrices. If the Kronecker model holds, then X_j has the pdf given in (1.11). However, in general the

pq-random vector

$$vec(X_j) \sim N_{pq}(vecM, \Lambda), \qquad (4.1)$$

where Λ is a $pq \times pq$ unstructured covariance matrix. It is assumed that $\Lambda > 0$. To estimate Λ, it is required that $N > pq$. We shall therefore assume in this section that $N > pq$. Thus, we wish to test the hypothesis

$$H : \Lambda = \Psi^* \otimes \Sigma$$

against the alternative $A \neq H$. The maximum likelihood estimators of Ψ^* and Σ have been obtained in Section 3. The MLE of Λ, under the alternative, is given by

$$\widehat{\Lambda} = \frac{1}{N} \sum_{i=1}^{N} (vecX_{cj})(vecX_{cj})', \; N > pq. \qquad (4.2)$$

Thus, the likelihood ratio test (LRT) for H against A is given by

$$\lambda_1 = \frac{|\widehat{\Lambda}|^{\frac{1}{2}N}}{|\widehat{\Psi^*}|^{\frac{1}{2}Np}|\widehat{\Sigma}|^{\frac{1}{2}Nq}}.$$

From asymptotic theory,

$$-2log\lambda_1 \sim \chi^2_{(\frac{1}{2}pq(pq+1)-\frac{1}{2}p(p+1)-\frac{1}{2}q(q+1)+1)}.$$

This result may be compared with the result obtained in Roy and Khattree (2005). Note that if we do not assume $\psi_{qq} = 1$ then we are in a testing situation where not all parameters can be identified under the null distribution, and thus standard asymptotic results for the LRT are not at disposal (e.g. see Ritz and Skovgaard, 2005).

5   Testing that Ψ is of intraclass correlation structure

In order to test that Ψ is of the intraclass correlation structure, we either assume that

$$\Psi = (1 - \rho)I_q + \rho\underline{1}_q\underline{1}'_q, \text{ and } \Sigma = (\sigma_{ij}) > 0, \qquad (5.1)$$

or assume that

$$\Psi = \sigma^2[(1 - \rho)I_q + \rho\underline{1}_q\underline{1}'_q], \text{ and } \sigma_{pp} = 1. \qquad (5.2)$$

We consider the model (5.2) with N iid observation matrices X_j, $j = 1, \ldots, N$, $N > \max(p, q)$, and since $\sigma_{pp} = 1$ we denote Σ by Σ^*. The approach of Section 3 will be adopted with suitable modifications. In particular the uniqueness of the estimators has to be shown in a somewhat different way. The pdf of $X = (X_1, \ldots, X_N)$ is given by (2.1). After maximazing with respect to M, the likelihood function is given by

$$(2\pi)^{-\frac{1}{2}pqN}|\Sigma^*|^{-\frac{1}{2}Nq}|\Psi|^{-\frac{1}{2}Np}etr\{-\frac{1}{2}\Sigma^{*-1}\sum_{i=1}^{N}X_{ic}\Psi^{-1}X_{ic}'\}.$$

Making the transformation

$$U_i = X_{ic}H,$$

where H is a $q \times q$ orthogonal matrix of the Helmert's type used earlier in Section 2. Then

$$H'\Psi^{-1}H = (H'\Psi H)^{-1} = \left(\begin{array}{cc} \tau_1^{-1} & 0 \\ 0 & \tau_2^{-1}I_{q-1} \end{array} \right) \equiv D_\tau^{-1},$$

where

$$\begin{array}{rcl} \tau_1 & = & \sigma^2(1 + (q-1)\rho), \\ \tau_2 & = & \sigma^2(1 - \rho). \end{array}$$

Hence, the likelihood is given by

$$(2\pi)^{-\frac{1}{2}Npq}\tau_1^{-\frac{1}{2}Np}\tau_2^{-\frac{1}{2}Np(q-1)}|\Sigma_{1\bullet p}|^{-\frac{1}{2}Nq}etr\{-\frac{1}{2}\sum_{i=1}^{N}\Sigma^{*-1}U_iD_\tau^{-1}U_i'\}, \quad (5.3)$$

where

$$\Sigma^{*-1} = d_pd_p' + \left(\begin{array}{c} I_{p-1} \\ -\underline{\sigma}_{1q}' \end{array} \right)\Sigma_{1\bullet p}^{-1}\left(\begin{array}{cc} I_{p-1} : & -\underline{\sigma}_{1q} \end{array} \right)$$

and d_p is the pth unit base vector of size p. If e_k denotes the kth unit base vector of size q then

$$tr\{\Sigma^{*-1}U_iD_\tau^{-1}U_i'\} = \tau_1^{-1}tr\{\Sigma^{*-1}U_ie_1e_1'U_i'\}$$

$$+\tau_2^{-1}\sum_{k=2}^{q}tr\{\Sigma^{*-1}U_ie_ke_k'U_i'\} \quad (5.4)$$

12

as well as

$$tr\{\Sigma^{*-1}U_iD_\tau^{-1}U_i'\} = tr\{d_pd_p'U_iD_\tau^{-1}U_i'\}$$

$$+tr\{\begin{pmatrix} I_{p-1} \\ -\underline{\sigma}_{1q}' \end{pmatrix}\Sigma_{1\bullet p}^{-1}\begin{pmatrix} I_{p-1}: & -\underline{\sigma}_{1q} \end{pmatrix}U_iD_\tau^{-1}U_i'\}. \quad (5.5)$$

Put $U = (U_1, U_2, \ldots, U_N)$ and $U_i = (U_{j1}' : U_{ip}')'$. By using (5.4) and (5.5) the likelihood in (5.3) leads after some manipulations to the following ML-equations:

$$\widehat{\tau}_1 = \frac{1}{Np}\sum_{i=1}^{N}tr\{\widehat{\Sigma}^{*-1}U_ie_1e_1'U_i'\} = \frac{1}{Np}tr\{\widehat{\Sigma}^{*-1}U(I_N\otimes e_1e_1')U'\}, \quad (5.6)$$

$$\widehat{\tau}_2 = \frac{1}{Np(q-1)}\sum_{i=1}^{N}\sum_{k=2}^{q}tr\{\widehat{\Sigma}^{*-1}U_ie_ke_k'U_i'\}$$

$$= \frac{1}{Np(q-1)}tr\{\widehat{\Sigma}^{*-1}U(I_N\otimes\sum_{k=2}^{q}e_ke_k')U'\}, \quad (5.7)$$

$$\widehat{\underline{\sigma}}_{1p} = \sum_{i=1}^{N}U_{i1}\widehat{D}_\tau^{-1}U_{ip}'(\sum_{i=1}^{N}U_{ip}\widehat{D}_\tau^{-1}U_{ip}')^{-1}, \quad (5.8)$$

$$\widehat{\Sigma}_{1\bullet p} = \frac{1}{Nq}\sum_{i=1}^{N}\begin{pmatrix} I_{p-1}: & -\underline{\sigma}_{1q}' \end{pmatrix}U_i\widehat{D}_\tau^{-1}U_i'\begin{pmatrix} I_{p-1} \\ -\underline{\sigma}_{1q}' \end{pmatrix}. \quad (5.9)$$

In the next, in correspondence with (3.8), we are going to show that

$$Nq = \sum_{i=1}^{N}U_{ip}\widehat{D}_\tau^{-1}U_{ip}' = \sum_{i=1}^{N}d_p'U_i\widehat{D}_\tau^{-1}U_i'd_p. \quad (5.10)$$

Equation (5.9) implies

$$NqI_{p-1} = \sum_{i=1}^{N}\begin{pmatrix} I_{p-1}: & -\widehat{\underline{\sigma}}_{1q}' \end{pmatrix}U_i\widehat{D}_\tau^{-1}U_i'\begin{pmatrix} I_{p-1} \\ -\widehat{\underline{\sigma}}_{1q} \end{pmatrix}\widehat{\Sigma}_{1\bullet p}^{-1}$$

and taking the trace yields

$$
Nq(p-1) = \sum_{j=1}^{N} tr\{\widehat{\Sigma}_{1\bullet p}^{-1} \begin{pmatrix} I_{p-1} : & -\widehat{\underline{\sigma}}_{1q} \end{pmatrix} U_i \widehat{D}_\tau^{-1} U_i' \begin{pmatrix} I_{p-1} \\ -\widehat{\underline{\sigma}}_{1q}' \end{pmatrix}\}
$$

$$
= Np - \sum_{i=1}^{N} tr\{d_p d_p' U_i \widehat{\tau}_1^{-1} e_1 e_1' U_i'\} + Np(q-1) - \sum_{i=1}^{N}\sum_{k=2}^{q} tr\{d_p d_p' U_i \widehat{\tau}_2^{-1} e_k e_k' U_i'\}
$$

$$
= Npq - \sum_{i=1}^{N} d_p' U_i \widehat{D}_\tau^{-1} U_i' d_p \tag{5.11}
$$

which implies (5.10). In (5.11) we used that $\widehat{\tau}_1$ and $\widehat{\tau}_2$ respectively can be written

$$
Np\widehat{\tau}_1 = \sum_{i=1}^{N} tr\{d_p d_p' U_i e_1 e_1' U_i'\}
$$
$$
+ tr\{\begin{pmatrix} I_{p-1} \\ -\widehat{\underline{\sigma}}_{1q}' \end{pmatrix} \widehat{\Sigma}_{1\bullet p}^{-1} \begin{pmatrix} I_{p-1} : & -\widehat{\underline{\sigma}}_{1q} \end{pmatrix} U_i e_1 e_1' U_i'\},
$$

$$
Np(q-1)\widehat{\tau}_2 = \sum_{i=1}^{N}\sum_{k=2}^{q} tr\{d_p d_p' U_i e_k e_k' U_i'\}
$$
$$
+ tr\{\begin{pmatrix} I_{p-1} \\ -\widehat{\underline{\sigma}}_{1q}' \end{pmatrix} \widehat{\Sigma}_{1\bullet p}^{-1} \begin{pmatrix} I_{p-1} : & -\widehat{\underline{\sigma}}_{1q} \end{pmatrix} U_i e_k e_k' U_i'\}.
$$

Thus,

$$
\widehat{\underline{\sigma}}_{1p} = \frac{1}{Nq} \sum_{i=1}^{N} U_{i1} \widehat{D}_\tau^{-1} U_{i1}'
$$

which implies that

$$
\widehat{\Sigma}^* = \frac{1}{Nq} \sum_{i=1}^{N} U_i \widehat{D}_\tau^{-1} U_i' = \frac{1}{Nq} U(I_N \otimes \widehat{D}_\tau^{-1}) U'. \tag{5.12}
$$

Thus, similarly to Theorem 3.1 we have the following result which also is proven in Section 8:

Theorem 5.1. *Let $\widehat{\Sigma}^*$, $\widehat{\tau}_1$ and $\widehat{\tau}_2$ satisfy (5.12), (5.6) and (5.7), with $\sigma_{pp} = 1$ supposed to hold. If $N > \max(p, q)$, then there is only one solution and the MLEs based on the likelihood given in (5.3) are unique.*

14

Next we obtain the MLE of Ψ with no restriction on the elements of Ψ. On the lines of Section 3, it follows that for given Σ^*, the MLE of Ψ is given by

$$\widehat{\Psi} = \frac{1}{Np} \sum_{i=1}^{N} X'_{ic} \Sigma^{*-1} X_{ic}.$$

Let

$$L = \sum_{i=1}^{N} X_{ic} \widehat{\Psi}^{-1} X'_{ic} = \begin{pmatrix} L_{11} & \underline{l}_{1p} \\ \underline{l}'_{1p} & l_{pp} \end{pmatrix}.$$

Then the MLE of $\Sigma_{1.p}$, Σ_{11}, $\underline{\sigma}_{1p}$ for a given Ψ are given by

$$
\begin{aligned}
\widehat{\Sigma}_{1\bullet p} &= \frac{1}{Nq}[L_{11} - \frac{\underline{l}_{1p}\underline{l}'_{1p}}{Nq(p-1)}], \\
\widehat{\Sigma}_{11} &= \widehat{\Sigma}_{1\bullet p} + \frac{\underline{l}_{1p}\underline{l}'_{1p}}{(Nq(p-1))^2}, \\
\widehat{\underline{\sigma}}_{1p} &= \frac{\underline{l}_{1p}}{Nq(p-1)}, \\
\widehat{\Sigma}^* &= \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\underline{\sigma}}_{1p} \\ \widehat{\underline{\sigma}}_{1p} & 1 \end{pmatrix}.
\end{aligned}
$$

Thus, the maximum of the likelihood function under the alternative hypothesis that Ψ is not of the intraclass correlation model is given by

$$(2\pi)^{-\frac{1}{2}pqN}|\widehat{\Sigma}_{1\bullet p}|^{-\frac{1}{2}qN}|\widehat{\Psi}|^{-\frac{1}{2}pN}e^{-\frac{1}{2}pqN}. \tag{5.13}$$

Similarly, the maximum of the likelihood function under the hypothesis that Ψ is of intraclass correlation model is given by

$$(2\pi)^{-\frac{1}{2}pqN}|\widehat{\Sigma}_{1\bullet p}(H)|^{-\frac{1}{2}qN}(\widehat{\tau}_1)^{-\frac{1}{2}pN}(\widehat{\tau}_2)^{-\frac{1}{2}p(q-1)N}e^{-\frac{1}{2}pqN}, \tag{5.14}$$

where $\widehat{\Sigma}_{1\bullet p}(H)$ stands for the MLE of $\Sigma_{1\bullet p}$ under the null hypthesis. Thus, the likelihood ratio test for testing the hypothesis that Ψ is of the intraclass correlation structure is given by

$$\lambda_2 = [\widehat{\Sigma}_{1\bullet p}\widehat{\Sigma}_{1\bullet p}^{-1}(H)]^{-\frac{1}{2}qN}|\widehat{\Psi}|^{\frac{1}{2}pN}(\widehat{\tau}_1)^{-\frac{1}{2}pN}(\widehat{\tau}_2)^{-\frac{1}{2}p(q-1)N}. \tag{5.15}$$

From asymptotic theory, it follows that $-2log\lambda_2$ is distributed as chisquare with $\frac{1}{2}q(q+1) - 2$ degrees of freedom. These results may be compared with results obtained in Naik and Rao (2001) and Roy and Khattree (2005).

15

6 Testing the hypothesis that $\Psi^* = I$ and $\Psi_\rho = I$

In most practical situations, it would be desirable to test the hypothesis

$$H_1 : \Psi^* = I, \text{ against the alternative } A_1 \neq H_1$$

and

$$H_2 : \Psi_\rho = I, \text{ against the alternative } A_2 \neq H_2.$$

We first consider the hypothesis H_1 vs A_1. For this we use the likelihood ratio procedure. The maximum likelihood estimators $\widehat{\Sigma}$ and $\widehat{\Psi}$ of Σ and Ψ are given in Section 3. From Corollary 2.1, the maximum likelihood estimator of Σ when $\Psi = I$ is S. Hence, the log of the likelihood rates under the hypothesis and under the alternative is given by

$$\log \lambda_3 = \frac{|\widehat{\Psi}^*|^{\frac{1}{2}pN}|\widehat{\Sigma}|^{\frac{1}{2}qN}}{|S|^{\frac{1}{2}qN}}$$

and

$$-2 \log \lambda_3 \sim \chi^2_{\frac{1}{2}q(q+1)-1}$$

asymptotically under the hypothesis H_1.

For testing the hypothesis H_2 against the alternative A_2 when $\psi_{ii} = 1$, $i = 1, \ldots, q$, we consider the estimator $\tilde{\Psi}_{ik}$ of Ψ_{ik} given in Section 2. Under the hypothesis H_2,

$$t_{ik} = \sqrt{np}\ \tilde{\Psi}_{ik},\ i \neq k$$

which are independently normally distributed with mean 0 and variance one. Hence,

$$T = \frac{\sum_{i<j} t_{ij}^2 - \frac{1}{2}q(q-1)}{\sqrt{q(q-1)}}$$

is asymptotically normally distributed with mean zero and variance one.

7 Simulation study

In this section we present a small simulation study indicating that the proposed algorithm in Section 3 is working practically when finding the MLEs. It is

interesting to note that in all simulations the estimators were close to the true value. We present results when the sample sizes equal either 100 or 250, which are relatively small numbers when taking into account the number of parameters. The number of observations equals 250 when either $p = 15$ or $q = 15$. We have also performed a number of simulations (not presented here) with $N = 1000$ and the results agree with those presented in this paper.

In Tables 1-6, the results based on 300 replications are presented, for six different settings of p, q, and N in the likelihood (3.1). Estimators are found according to the proposed algorithm in Section 3. Only the estimators of the unknown diagonal elements of Σ and Ψ are presented together with the true value, the standard deviation (std) as well as the minimum and maximum values of the replications.

Table 1. $p = 5$, $q = 5$ and $N = 100$.

parameter	true value	MLE	std	min	max
σ_{11}	6.02	5.90	0.52	4.65	7.59
σ_{22}	5.42	5.36	0.47	4.14	7.05
σ_{33}	3.11	3.06	0.27	2.40	3.83
σ_{44}	3.63	3.59	0.31	2.73	4.73
σ_{55}	2.05	2.01	0.18	1.60	2.60
ψ_{11}	0.43	0.45	0.04	0.34	0.60
ψ_{22}	0.54	0.55	0.05	0.40	0.69
ψ_{33}	0.28	0.29	0.03	0.21	0.37
ψ_{44}	0.46	0.46	0.04	0.34	0.61

Table 2. $p = 7$, $q = 5$ and $N = 100$.

parameter	true value	MLE	std	min	max
σ_{11}	6.99	6.86	0.52	5.54	8.37
σ_{22}	7.41	7.27	0.55	5.75	8.86
σ_{33}	3.80	3.75	0.30	2.92	4.91
σ_{44}	5.42	5.34	0.43	4.39	6.52
σ_{55}	6.44	6.33	0.47	5.10	8.10
σ_{66}	5.49	5.40	0.40	4.36	6.71
σ_{77}	11.37	11.12	0.87	8.98	13.91
ψ_{11}	0.11	0.11	0.01	0.09	0.14
ψ_{22}	0.21	0.22	0.02	0.17	0.27
ψ_{33}	0.26	0.26	0.02	0.21	0.32
ψ_{44}	0.37	0.37	0.03	0.30	0.44

Table 3. $p = 15$, $q = 5$ and $N = 250$.

parameter	true value	MLE	std	min	max
σ_{11}	14.56	14.53	0.66	12.70	16.33
σ_{22}	10.06	10.20	0.45	8.62	11.36
σ_{33}	12.83	12.80	0.60	10.98	14.70
σ_{44}	12.84	12.77	0.60	11.35	14.82
σ_{55}	10.81	10.77	0.47	9.52	12.40
σ_{66}	7.35	7.31	0.35	6.29	8.45
σ_{77}	13.32	13.32	0.59	11.81	14.87
σ_{88}	17.45	17.44	0.80	15.59	19.70
σ_{99}	10.89	10.83	0.49	9.29	11.94
σ_{1010}	13.17	13.15	0.58	11.71	14.99
σ_{1111}	25.35	25.28	1.15	22.49	29.34
σ_{1212}	17.37	17.27	0.79	14.95	19.38
σ_{1313}	18.72	18.62	0.83	16.31	21.12
σ_{1414}	7.84	7.82	0.37	6.87	8.87
σ_{1515}	12.09	12.03	0.51	10.58	13.36
ψ_{11}	0.73	0.73	0.02	0.66	0.79
ψ_{22}	0.14	0.14	0.004	0.13	0.15
ψ_{33}	0.87	0.87	0.03	0.80	0.94
ψ_{44}	0.23	0.23	0.01	0.21	0.25

Table 4. $p = 15$, $q = 7$ and $N = 250$.

parameter	true value	MLE	std	min	max
σ_{11}	10.95	10.87	0.41	9.45	12.06
σ_{22}	13.00	12.97	0.57	11.56	14.67
σ_{33}	13.14	13.08	0.52	11.60	14.64
σ_{44}	13.90	13.83	0.51	12.53	14.95
σ_{55}	17.50	17.52	0.70	15.56	19.39
σ_{66}	11.34	11.31	0.45	10.06	12.49
σ_{77}	13.83	13.71	0.57	12.28	15.57
σ_{88}	13.25	13.19	0.54	11.60	14.62
σ_{99}	13.43	13.38	0.51	11.83	14.66
σ_{1010}	7.15	7.14	0.31	6.46	7.84
σ_{1111}	10.47	10.46	0.39	9.59	12.18
σ_{1212}	29.49	29.43	1.15	26.35	32.64
σ_{1313}	17.17	17.07	0.69	15.38	18.68
σ_{1414}	19.49	19.37	0.75	17.30	21.13
σ_{1515}	15.35	15.33	0.57	13.71	16.78
ψ_{11}	0.68	0.68	0.02	0.62	0.73
ψ_{22}	0.76	0.76	0.03	0.70	0.82
ψ_{33}	0.93	0.94	0.03	0.86	1.01
ψ_{44}	0.56	0.56	0.02	0.51	0.61
ψ_{55}	0.55	0.55	0.02	0.50	0.60
ψ_{66}	0.92	0.93	0.03	0.86	1.01

Table 5. $p = 5$, $q = 7$ and $N = 100$.

parameter	true value	MLE	std	min	max
σ_{11}	6.02	5.86	0.48	4.40	7.18
σ_{22}	5.42	5.33	0.44	4.35	6.79
σ_{33}	3.11	3.03	0.26	2.16	3.97
σ_{44}	3.64	3.56	0.30	2.79	4.45
σ_{55}	2.05	1.99	0.17	1.45	2.48
ψ_{11}	0.60	0.61	0.06	0.48	0.80
ψ_{22}	0.65	0.66	0.06	0.54	0.89
ψ_{33}	0.49	0.50	0.04	0.40	0.63
ψ_{44}	1.03	1.05	0.10	0.83	1.43
ψ_{55}	0.24	0.24	0.02	0.20	0.32
ψ_{66}	0.42	0.43	0.04	0.34	0.56

Table 6. $p = 5$, $q = 15$ and $N = 250$.

parameter	true value	MLE	std	min	max
σ_{11}	2.45	2.39	0.12	2.07	2.83
σ_{22}	1.46	1.43	0.07	1.25	1.63
σ_{33}	5.66	5.50	0.26	4.81	6.45
σ_{44}	3.01	2.94	0.15	2.56	3.39
σ_{55}	2.94	2.87	0.14	2.46	3.33
ψ_{11}	1.06	1.08	0.06	0.92	1.34
ψ_{22}	0.77	0.72	0.05	0.59	0.87
ψ_{33}	0.73	0.75	0.05	0.63	0.88
ψ_{44}	0.83	0.84	0.04	0.72	0.99
ψ_{55}	0.56	0.58	0.04	0.48	0.68
ψ_{66}	0.60	0.62	0.03	0.54	0.69
ψ_{77}	1.16	1.19	0.07	0.97	1.42
ψ_{88}	0.37	0.38	0.02	0.32	0.45
ψ_{99}	0.52	0.54	0.03	0.45	0.64
ψ_{1010}	0.73	0.75	0.04	0.64	0.89
ψ_{1111}	0.71	0.72	0.04	0.62	0.84
ψ_{1212}	0.84	0.86	0.05	0.72	1.02
ψ_{1313}	1.08	1.11	0.06	0.88	1.30
ψ_{1414}	0.94	0.96	0.05	0.82	1.14

8 Proofs of Theorem 3.1 and Theorem 5.1

Here proofs of the theorems for uniqueness of the MLEs are presented. Firstly it is shown that the flip-flop algorithm given by (3.12) and (3.13) only has one solution, i.e. we obtain unique estimators. The relations in (3.12) and (3.13) may be rewritten as

$$Nq\Sigma = \sum_{i=1}^{N}(e_i' \otimes I)X_c\Psi^{-1}X_c'(e_i \otimes I), \qquad (8.1)$$

$$Np\Psi = X_c'(I \otimes \Sigma^{-1})X_c, \qquad (8.2)$$

where

$$X_c = (X_{1c}' : X_{2c}' : \dots : X_{Nc}')', \qquad (pN \times q),$$

and e_i is a unit base vector of size N, i.e. the ith column of I_N. These equations imply that

$$Nq\Sigma = \sum_{i=1}^{N}(e_i' \otimes I_p)X_c(\frac{1}{Np}X_c'(I_N \otimes \Sigma^{-1})X_c)^{-1}X_c'(e_i \otimes I_p)$$

is to be considered which in turn is equivalent to

$$\frac{q}{p}I = \sum_{i=1}^{N}(e_i' \otimes I_p)X_c(\frac{1}{Np}X_c'(I_N \otimes \Sigma^{-1})X_c)^{-1}X_c'(e_i \otimes I_p)\Sigma^{-1} \qquad (8.3)$$

$$= \sum_{i=1}^{N}(e_i' \otimes I_p)X_c(\frac{1}{Np}X_c'(I_N \otimes \Sigma^{-1})X_c)^{-1}X_c'(I_N \otimes \Sigma^{-1})(e_i \otimes I_p)$$

$$= (vec'I_N \otimes I_p)(I_N \otimes X_c(X_c'(I_N \otimes \Sigma^{-1})X_c)^{-1}X_c'(I_N \otimes \Sigma^{-1}))$$

$$\times (vecI_N \otimes I_p). \qquad (8.4)$$

Put

$$P_\Sigma = X_c(X_c'(I_N \otimes \Sigma^{-1})X_c)^{-1}X_c'(I_N \otimes \Sigma^{-1})$$

which is a projector (idempotent matrix). Suppose now that there are Σ_1 and Σ_2 both satisfying (8.4). Then, by subtraction we obtain

$$0 = (vec'I_N \otimes I_p)(I_N \otimes (P_{\Sigma_1} - P_{\Sigma_2}))(vecI_N \otimes I_p). \qquad (8.5)$$

Our task is to show that the only solution of (8.5) is given by $\Sigma_1 = \Sigma_2$. It may be noted that $\Sigma_1 = c\Sigma_2$, $c \neq 1$ is not a possibility since $\psi_{qq} = 1$ and equation (3.13) has to be satisfied by all the solutions. Let

$$Q_\Sigma = I_{pN} - P_\Sigma = (I_N \otimes \Sigma)X_c^o(X_c^{o'}(I_N \otimes \Sigma)X_c^o)^{-1}X_c^{o'},$$

where X_c^o is defined to be any matrix which under the standard inner product generates the orthogonal complement to the column space generated by X_c. Then,

$$\begin{aligned}
P_{\Sigma_1} - P_{\Sigma_2} &= P_{\Sigma_1}(I_{pN} - P_{\Sigma_2}) = P_{\Sigma_1}Q_{\Sigma_2} \\
&= X_c(X_c'(I_N \otimes \Sigma_1^{-1})X_c)^{-1}X_c'(I_N \otimes \Sigma_1^{-1}) \\
&\quad \times (I_N \otimes \Sigma_2)X_c^o(X_c^{o'}(I_N \otimes \Sigma_2)X_c^o)^{-1}X_c^{o'}.
\end{aligned}$$

Suppose for a while that (8.5) holds if and only if $P_{\Sigma_1} - P_{\Sigma_2} = 0$, i.e. $P_{\Sigma_1}Q_{\Sigma_2} = 0$, which is equivalent to

$$X_c'(I_N \otimes \Sigma_1^{-1}\Sigma_2)X_c^o = 0. \tag{8.6}$$

There are two possibilities for (8.6) to hold. Either $\Sigma_1 = c\Sigma_2$ or the column space, denoted $C(\bullet)$, generated by X_c^o is invariant with respect to $I_N \otimes \Sigma_1^{-1}\Sigma_2$, i.e. the space is generated by the eigenvectors of $I_N \otimes \Sigma_1^{-1}\Sigma_2$. However, since the matrix of eigenvectors is of the form $I_N \otimes \Gamma$ for some Γ it shows, since the column space of X_c^o is a function of the observations, that $I_N \otimes \Gamma$ does not generate the space, unless $N = 1$. Thus, in order for (8.6) to hold $\Sigma_1 = c\Sigma_2$ and we have already noted that in this case $c = 1$, i.e. $\Sigma_1 = \Sigma_2$.

It remains to show that (8.5) is true only if $P_{\Sigma_1} - P_{\Sigma_2} = P_{\Sigma_1}Q_{\Sigma_2} = 0$, i.e.

$$0 = (vec'I_N \otimes I_p)(I_N \otimes P_{\Sigma_1})(I_N \otimes Q_{\Sigma_2})(vecI_N \otimes I_p). \tag{8.7}$$

Since, in (8.6) we have the two projections $(I_N \otimes P_{\Sigma_1})$ and $(I_N \otimes Q_{\Sigma_2})$ we will study the effect of them via column spaces. Using Theorem 1.2.16 in Kollo & von Rosen (2005) gives

$$\begin{aligned}
C((I_N \otimes P_{\Sigma_1}')(vecI_N \otimes I_p)) &= C(I_N \otimes P_{\Sigma_1}') \cap \{C(I_N \otimes P_{\Sigma_1})^\perp + C(\vec{I}_N \otimes I_p)\} \\
&= C(I_N \otimes (I_N \otimes \Sigma_1^{-1})X_c) \cap \{C(I_N \otimes X_c^o) + C(vecI_N \otimes I_p)\}, \tag{8.8}
\end{aligned}$$

$$\begin{aligned}
C((I_N \otimes Q_{\Sigma_2})(vecI_N \otimes I_p)) &= C(I_N \otimes Q_{\Sigma_2}) \cap \{C(I_N \otimes Q_{\Sigma_2}')^\perp + C(\vec{I}_N \otimes I_p)\} \\
&= C(I_N \otimes (I_N \otimes \Sigma_2)X_c^o) \cap \{C(I_N \otimes X_c) + C(vecI_N \otimes I_p)\}. \tag{8.9}
\end{aligned}$$

If (8.7) should hold the spaces presented in (8.8) and (8.9), respectively, must be orthogonal. Thus,

$$\begin{aligned}
&C(I_N \otimes (I_N \otimes \Sigma_1^{-1})X_c) \cap \{C(I_N \otimes X_c^o) + C(vecI_N \otimes I_p)\} \\
&\qquad \subseteq C(I_N \otimes (I_N \otimes \Sigma_2)X_c^o)^\perp + C(I_N \otimes X_c)^\perp \cap C(vecI_N \otimes I_p)^\perp. \tag{8.10}
\end{aligned}$$

22

However, the following trivial facts hold:

$$C(I_N \otimes (I_N \otimes \Sigma_1^{-1})X_c) \cap \{C(I_N \otimes X_c^o) + C(vecI_N \otimes I_p)\}$$
$$\subseteq C(I_N \otimes (I_N \otimes \Sigma_1^{-1})X_c)$$
$$C(I_N \otimes (I_N \otimes \Sigma_2)X_c^o)^\perp$$
$$\subseteq C(I_N \otimes (I_N \otimes \Sigma_2)X_c^o)^\perp + C(I_N \otimes X_c)^\perp \cap C(vecI_N \otimes I_p)^\perp.$$

Hence, if

$$C(I_N \otimes (I_N \otimes \Sigma_1^{-1})X_c) \subseteq C(I_N \otimes (I_N \otimes \Sigma_2)X_c^o)^\perp \qquad (8.11)$$

does not hold (8.10) as well as (8.8) can not be valid and therefore (8.11) must always be true which is the same as stating that $P_{\Sigma_1}Q_{\Sigma_2} = 0$. Thus, the flip-flop algorithm provides us with unique solutions.

Turning to Theorem 5.1 we will show that estimators satisfying

$$\Sigma = \frac{1}{Nq}U(I_N \otimes D_\tau^{-1})U', \qquad (8.12)$$

where

$$D_\tau = \tau_1 e_1 e_1' + \sum_{k=2}^{q} \tau_1 e_1 e_1' \qquad (8.13)$$

with

$$\tau_1 = \frac{1}{Np}tr\{\Sigma^{-1}U(I_N \otimes e_1 e_1')U'\}, \qquad (8.14)$$

$$\tau_2 = \frac{1}{Np(q-1)}tr\{\Sigma^{-1}U(I_N \otimes \sum_{k=2}^{q} e_k e_k')U'\} \qquad (8.15)$$

are unique. The above given equations imply that

$$\Sigma = \frac{p}{q}(tr\{\Sigma^{-1}U(I_N \otimes e_1 e_1')(I_N \otimes e_1 e_1')U'\})^{-1}U(I_N \otimes e_1 e_1')(I_N \otimes e_1 e_1')U'$$

$$+\frac{p(q-1)}{q}(\sum_{k=2}^{q} tr\{\Sigma^{-1}U(I_N \otimes e_k e_k')(I_N \otimes e_k e_k')U'\})^{-1}$$

$$\times \sum_{k=2}^{q} U(I_N \otimes e_k e_k')(I_N \otimes e_k e_k')U'. \qquad (8.16)$$

Postmultiplying (8.16) by Σ^{-1} and putting

$$A = U(I_N \otimes e_1 e_1'), \qquad B = \sum_{k=2}^{q} U(I_N \otimes e_k e_k')$$

we obtain

$$I = \frac{p}{q}(tr\{\Sigma^{-1}AA'\})^{-1}AA'\Sigma^{-1} + \frac{p(q-1)}{q}(tr\{\Sigma^{-1}BB'\})^{-1}BB'\Sigma^{-1}. \quad (8.17)$$

Note that $AB' = 0$ implies that $C(A) \cap C(B) = \{0\}$. Now suppose that there exist Σ_1 and Σ_2, $\Sigma_1 \neq c\Sigma_2$, for any positive constant $c \neq 1$, satisfying (8.17). The case $\Sigma_1 = c\Sigma_2, c \neq 1$ is of no interest since $\sigma_{pp} = 1$ is supposed to hold. Thus, it follows that

$$0 = \frac{p}{q}((tr\{\Sigma_1^{-1}AA'\})^{-1}AA'\Sigma_1^{-1} - (tr\{\Sigma_2^{-1}AA'\})^{-1}AA'\Sigma_2^{-1})$$

$$+ \frac{p(q-1)}{q}((tr\{\Sigma_1^{-1}BB'\})^{-1}BB'\Sigma_1^{-1} - (tr\{\Sigma_2^{-1}BB'\})^{-1}BB'\Sigma_2^{-1}). \quad (8.18)$$

Since $C(AA'\widehat{\Sigma}_i) = C(A)$, $C(BB'\widehat{\Sigma}_i) = C(B)$, $i = 1, 2$, and $C(A) \cap C(B) = \{0\}$, the two terms in (8.18) may be considered separately, i.e.

$$\begin{aligned}
(tr\{\Sigma_1^{-1}AA'\})^{-1}AA'\Sigma_1^{-1} &= (tr\{\Sigma_2^{-1}AA'\})^{-1}AA'\Sigma_2^{-1}, \\
(tr\{\Sigma_1^{-1}BB'\})^{-1}BB'\Sigma_1^{-1} &= (tr\{\Sigma_2^{-1}BB'\})^{-1}BB'\Sigma_2^{-1}.
\end{aligned} \quad (8.19)$$

Because of symmetry it is enough to exploit (8.19) which, since AA' with probability 1 is of full rank p, is identical to

$$\Sigma_2 = \frac{tr\{\Sigma_1^{-1}AA'\}}{tr\{\Sigma_2^{-1}AA'\}}\Sigma_1. \quad (8.20)$$

This implies that $\Sigma_1 = c\Sigma_2$ which according to the assumptions should not hold except when $c = 1$. Hence, there exist only one solution to (8.18).

Acknowledgement

References

[1] Arnold, S. (1973). Applications of the theory of products of problems to certain patterned covariance matrices. *Annals of Statistics*, **1**, 682–699.

[2] Drton, M. (2006). Computing all roots of the likelihood equations of seemingly unrelated regressions. *Journal of Symbolic Computation*, **41**, 245–254.

[3] Dutilleul, P. (1999). The MLE algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation*, **64**, 105–123.

[4] Galecki, A.T. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Comunications in Statistics - Theory and Methods*, **23**, 3105–3119.

[5] Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, Dordrecht: Springer.

[6] Lu, N. and Zimmerman, D.L. (2005). The likelihood ratio test for a separable covariance matrix. *Statistics and Probability Letters*, **73**, 449–457.

[7] Naik, D.N. and Rao, S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *Journal of Applied Statistics*, **28**, 91–105.

[8] Olkin, I. (1973). Testing and estimation for structures which are circularly symmetric in blocks. *Multivariate Statistical Inference*, (Eds. D.G. Kabe, R.P. Gupta), 183–195, North-Holland, Amsterdam.

[9] Potthoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.

[10] Ritz, C. and Skovgaard, I.M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika*, **92**, 507–517.

[11] Roy, A. and Khattree, R. (2005). On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data. *Statistical Methodology*, **2**, 297–306.

[12] Shitan, M. and Brockwell, P.J. (1995). An asymptotic test for separability of a spatial autoregressive model. *Comunications in Statistics - Theory and Methods*, **24**, 2027–2040.

[13] Srivastava, M.S. (1965). Some tests for the intraclass correlation model. *The Annals of Mathematical Statistics*, **36**, 1802–1806.

[14] Srivastava, M.S. and Khatri, C.G. (1979). *An Introduction to Multivariate Statistics*, North Holland, New York.

[15] Votaw, D.F. (1948). Testing compound symmetry in a normal multivariate distribution. *Annals of Mathematical Statistics*, **19**, 447–473.

[16] Wilks, S.S. (1946). Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution. *Annals of Mathematical Statistics*, **17**, 257–281.