Ann Inst Stat Math (2011) 63:29–42
DOI 10.1007/s10463-008-0213-1

Explicit estimators under *m*-dependence for a multivariate normal distribution

Martin Ohlson · Zhanna Andrushchenko · Dietrich von Rosen

Received: 26 November 2007 / Revised: 20 August 2008 / Published online: 20 December 2008

Abstract The problem of estimating parameters of a multivariate normal p-dimensional random vector is considered for a banded covariance structure reflecting m-dependence. A simple non-iterative estimation procedure is suggested which gives an explicit, unbiased and consistent estimator of the mean and an explicit and consistent estimator of the covariance matrix for arbitrary p and m.

Keywords Banded covariance matrices · Covariance matrix estimation · Explicit estimators · Multivariate normal distribution

1 Introduction

Many testing, estimation and confidence interval procedures discussed in the multivariate statistical literature are based on the assumption that the observation vectors are independent and normally distributed (Muirhead 1982; Srivastava 2002). The main reasons for this are that often sets of multivariate observations are, at least approximately, normally distributed. Moreover, the multivariate normal distribution is mathematically tractable. Normally distributed data can be modelled entirely in terms of their means and variances/covariances. Estimating the mean and the covariance matrix is therefore a problem of great interest in statistics.

Patterned covariance matrices arise from a variety of contexts and have been studied by many authors. Below we give a very brief overview indicating different

M. Ohlson (✉)
Linköping University, 581 83 Linköping, Sweden
e-mail: martin.ohlson@mai.liu.se

Z. Andrushchenko · D. von Rosen
Swedish University of Agricultural Sciences, Box 7032, 75007 Uppsala, Sweden

directions of interest and applications. In a seminal paper, Wilks (1946) considered patterned structures when dealing with measurements on k equivalent psychological tests. This led to a covariance matrix with equal diagonal elements and equal off-diagonal elements. Votaw (1948) extended this model to a set of blocks in which each block had a pattern. Goodman (1963) studied the covariance matrix of the multivariate complex normal distribution, which for example arise in spectral analysis of multiple time series. A direct extension is to study quaternions which has been performed by many authors, e.g., see Andersson et al. (1983). Olkin and Press (1969) considered a circular stationary model, where variables are thought of as being equally spaced around a circle, and the covariance between two variables depends on the distance between the variables. Olkin (1973) studied a multivariate version in which each element was a matrix, and the blocks were patterned. More generally, permutation invariant covariance matrices may be of interest, see for example Nahtman (2006). Browne (1977) reviews patterned correlation matrices arising from multiple psychological measurements. In this context one may mention LISREL models (Jöreskog 1981) or more sophisticated structures within the frame of graphical models (Lauritzen 1996). From linear models with one error term we have natural extensions to mixed linear models and variance component models as well as to patterned covariance matrices in multivariate growth curve models, e.g., see Chinchilli and Carter (1984) and Searle et al. (1992). Block structures in covariance matrices have recently been studied by Naik and Rao (2001), Lu and Zimmerman (2005) and Roy and Khattree (2005), as well as others.

Banded covariance matrices and their inverses arise frequently in biological, economical or technical time series. For example in signal processing applications, including autoregressive or moving average image modelling, covariances of Gauss–Markov random processes (Woods 1972; Moura and Balram 1992), or numerical approximations of partial differential equations based on finite differences. Banded matrices are also used to model the correlation of cyclostationary processes in periodic time series (Chakraborty 1998). There exist many papers on Toeplitz covariance matrices, e.g., see Marin and Dhorne (2002) and Christensen (2007), which all are banded matrices. To have a Toeplitz structure means that certain invariance conditions are fulfilled, e.g., equality of variances. In this paper we will study banded matrices with unequal elements except that certain covariances are zero. The basic idea is that widely separated observations appear often to be uncorrelated and therefore it is reasonable to work with a banded covariance structure where all covariances more than m steps apart equal zero. We will call such a structure an m-dependent structure.

Originally, many estimators of the covariance matrix were obtained from non-iterative least squares methods such as the ANOVA and MINQUE approaches, for example when estimating variance components. When computer sources became stronger iterative methods were introduced such as maximum likelihood, restricted maximum likelihood, generalized estimation equations among others. In a series of papers, Szatrowski (1985) discussed how to obtain maximum likelihood estimators (MLEs) for the elements of a class of patterned covariance matrices. Godolphin and De Gooijer (1982) computed the exact MLEs of the parameters of a Gaussian moving average process. Certainly over the last years the iterative methods have been dominating. However, nowadays one is interested in applying covariance structures, including variance components models, to very large data sets. For example in QTL-analysis in Genetics or time series with densely sampled observations in meteorology or in EEG/EKG-studies in medicine. Therefore, in this paper we will study banded covariance matrices with the goal to obtain reasonable explicit estimators. A simple estimation procedure is suggested which under m-dependence gives unbiased and consistent estimators of the mean and consistent estimators of the banded covariance matrix.

2 Definitions and notation

Throughout this paper matrices will be denoted by capital letters, vectors by bold font, scalars and elements in matrices by ordinary letters if not stated otherwise.

Let X be matrix normally distributed (Kollo and von Rosen 2005) with the same mean for every column and with independent columns, i.e., $X \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}'_n, \Sigma, I_n)$, where the parameter matrix Σ represents the covariance between rows of X, and I_n, the identity matrix of dimension n, indicates that columns in X are independently distributed,

$$\begin{aligned} \boldsymbol{\mu}' &= \left(\mu_1, \mu_2, \ldots, \mu_p\right), \\ \mathbf{1}'_n &= (\underbrace{1, 1, \ldots, 1}_{n-\text{times}}). \end{aligned}$$

Partition X in the following way

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_p \end{pmatrix},$$

where $\mathbf{x}'_i = (x_{i1}, x_{i2}, \ldots, x_{in}) : (1 \times n)$ for $i = 1, \ldots, p$ and $\mathbf{x}'_i$ is the transpose of $\mathbf{x}_i$. If we have i and j such that $1 \leq i < j \leq p$, we will also use the notation $X_{i:j}$ for the matrix including the rows from i to j, i.e.,

$$X'_{i:j} = (\mathbf{x}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_j).$$

For $k = m+1, \ldots, p$ and $\Sigma = \left(\sigma_{ij}\right)$, $i, j = 1, 2, \ldots, p$, define $\Sigma_{(k)}^{(m)}$ as

$$\Sigma_{(k)}^{(m)} = \begin{pmatrix} \sigma_{11} & \ldots & \sigma_{1,m+1} & 0 & 0 & & \ldots & 0 \\ \sigma_{21} & \ldots & \sigma_{2,m+1} & \sigma_{2,m+2} & 0 & & \ldots & 0 \\ \vdots & & & & \ddots & & & \\ \sigma_{m+1,1} & & & & & & & \\ 0 & \ddots & & & & & & \\ \vdots & & & & & & & \\ & & & \ddots & & & & \\ 0 & \ldots & 0 & & \sigma_{k-1,k-(m+1)} & \sigma_{k-1,k-m} & \cdots & \sigma_{k-1,k} \\ 0 & \ldots & 0 & & 0 & \sigma_{k,k-m} & \ldots & \sigma_{kk} \end{pmatrix}. \tag{1}$$

For simplicity the upper index (m) will be omitted in the cases when it is clear from the context. We also define $M^{ji}_{(k)}$ as the matrix obtained when the jth row and ith column have been removed from $\Sigma_{(k)}$.

Moreover, we will often partition the matrix $\Sigma^{(m)}_{(k)}$ as

$$\Sigma^{(m)}_{(k)} = \begin{pmatrix} \Sigma^{(m)}_{(k-1)} & \underline{\boldsymbol{\sigma}}_{1k} \\ \underline{\boldsymbol{\sigma}}'_{k1} & \sigma_{kk} \end{pmatrix}, \tag{2}$$

where

$$\underline{\boldsymbol{\sigma}}'_{k1} = \left(0, \ldots, 0,\ \sigma_{k,k-m}, \ldots, \sigma_{k,k-1}\right).$$

3 Explicit estimator of a banded covariance matrix

In this section, the main result of this paper is presented. We propose explicit estimators of the expectation and the covariance matrix for a multivariate normal distribution when the covariance matrix have an m-dependent structure. First we propose estimators for the general case when $m + 1 < p < n$. Since the estimators are ad hoc we establish some properties such as unbiasedness and consistency. Furthermore, the special case $m = 1$ is considered in some detail since it uncovers the underlying structure of the estimators.

Proposition 1 *Let $X \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}'_n, \Sigma^{(m)}_{(p)}, I_n)$, with arbitrary integer m and $\Sigma^{(m)}_{(p)}$ defined in (1). The estimators of $\boldsymbol{\mu}$ and $\Sigma^{(m)}_{(p)}$ are given by the following two steps.*

(i) *Use the maximum likelihood estimator for $\mu_1, \ldots, \mu_{m+1}$ and $\Sigma^{(m)}_{(m+1)}$.*
(ii) *Calculate the following estimators for $k = m + 2, \ldots, p$ in increasing order where for each k let $i = k - m, \ldots, k - 1$:*

$$\hat{\mu}_k = \frac{1}{n}\mathbf{x}'_k\mathbf{1}_n, \tag{3}$$

$$\hat{\sigma}_{ki} = \hat{\beta}_{ki}\frac{|\hat{\Sigma}_{(k-1)}|}{|\hat{\Sigma}_{(k-2)}|}, \tag{4}$$

$$\hat{\sigma}_{kk} = \frac{1}{n}\mathbf{x}'_k\left(I_n - \hat{X}_{k-1}(\hat{X}'_{k-1}\hat{X}_{k-1})^{-1}\hat{X}'_{k-1}\right)\mathbf{x}_k + \hat{\underline{\boldsymbol{\sigma}}}'_{k1}\hat{\Sigma}^{-1}_{(k-1)}\hat{\underline{\boldsymbol{\sigma}}}_{1k}, \tag{5}$$

where

$$\begin{aligned} \hat{\underline{\boldsymbol{\sigma}}}'_{k1} &= \left(0, \ldots, 0,\ \hat{\sigma}_{k,k-m}, \ldots, \hat{\sigma}_{k,k-1}\right), \\ \hat{\boldsymbol{\beta}}_k &= \left(\hat{\beta}_{k0},\ \hat{\beta}_{k,k-m}, \ldots, \hat{\beta}_{k,k-1}\right)' = (\hat{X}'_{k-1}\hat{X}_{k-1})^{-1}\hat{X}'_{k-1}\mathbf{x}_k, \qquad (6) \\ \hat{X}_{k-1} &= \left(\mathbf{1}_n, \hat{\mathbf{x}}_{k-1,k-m}, \ldots, \hat{\mathbf{x}}_{k-1,k-1}\right) \end{aligned}$$

and

$$\hat{\mathbf{x}}_{k-1,i} = \sum_{j=1}^{k-1} (-1)^{i+j} \frac{|\hat{M}_{(k-1)}^{ji}|}{|\hat{\Sigma}_{(k-2)}|} \mathbf{x}_j,$$

where $\hat{M}_{(k-1)}^{ji}$ *is as* $M_{(k-1)}^{ji}$, *defined in Sect. 2, but* $\Sigma_{(k-1)}$ *is replaced by* $\hat{\Sigma}_{(k-1)}$.

Below follows a motivation for Proposition 1. The estimators are based on the likelihood. However, instead of maximizing the complete likelihood we factor the likelihood and maximize each term. In this way explicit estimators are obtained. By conditioning the probability density equals ($f(\bullet)$ and $f(\bullet|\bullet)$ denote the density and conditional density, respectively)

$$f(X) = f(\mathbf{x}_p' | X_{1:p-1}) \cdots f(\mathbf{x}_{m+2}' | X_{1:m+1}) f(X_{1:m+1}).$$

Hence, for $k = m+2, \ldots, p$ partition the covariance matrix $\Sigma_{(k)}$ as in (2). We have

$$\mathbf{x}_k' | X_{1:k-1} \sim N_{1,n}(\boldsymbol{\mu}_{k|1:k-1}', \sigma_{k|1:k-1}, I_n),$$

where the conditional variance equals

$$\sigma_{k|1:k-1} = \sigma_{kk} - \underline{\boldsymbol{\sigma}}_{k1}' \Sigma_{(k-1)}^{-1} \underline{\boldsymbol{\sigma}}_{1k}$$

and where the conditional expectation equals

$$\begin{aligned} \boldsymbol{\mu}_{k|1:k-1}' &= \mu_k \mathbf{1}_n' + \underline{\boldsymbol{\sigma}}_{k1}' \Sigma_{(k-1)}^{-1} \begin{pmatrix} \boldsymbol{x}_1' - \mu_1 \mathbf{1}_n' \\ \vdots \\ \boldsymbol{x}_{k-1}' - \mu_{k-1} \mathbf{1}_n' \end{pmatrix} \\ &= \beta_{k0} \mathbf{1}_n' + \sum_{i=k-m}^{k-1} \sigma_{ki} \sum_{j=1}^{k-1} \sigma_{(k-1)}^{ij} \boldsymbol{x}_j'. \end{aligned} \tag{7}$$

Here $\sigma_{(k-1)}^{ij}$ are the elements of the matrix

$$\Sigma_{(k-1)}^{-1} = \left(\sigma_{(k-1)}^{ij}\right)_{i,j} = \left((-1)^{i+j} \frac{|M_{(k-1)}^{ji}|}{|\Sigma_{(k-1)}|} \right)_{i,j}.$$

The first regression coefficient equals

$$\begin{aligned} \beta_{k0} &= \mu_k - \sum_{i=k-m}^{k-1} \sigma_{ki} \sum_{j=1}^{k-1} \sigma_{(k-1)}^{ij} \mu_j \\ &= \mu_k - \sum_{i=k-m}^{k-1} \sigma_{ki} \sum_{j=1}^{k-1} (-1)^{i+j} \frac{|M_{(k-1)}^{ji}|}{|\Sigma_{(k-1)}|} \mu_j. \end{aligned}$$

We may rewrite (7) as

$$\begin{aligned}\boldsymbol{\mu}_{k|1:k-1} &= \beta_{k0}\mathbf{1}_n + \sum_{i=k-m}^{k-1} \sigma_{ki} \sum_{j=1}^{k-1} \sigma_{(k-1)}^{ij}\mathbf{x}_j \\ &= \beta_{k0}\mathbf{1}_n + \sum_{i=k-m}^{k-1} \sigma_{ki} \frac{|\Sigma_{(k-2)}|}{|\Sigma_{(k-1)}|} \sum_{j=1}^{k-1}(-1)^{i+j} \frac{|M_{(k-1)}^{ji}|}{|\Sigma_{(k-2)}|}\mathbf{x}_j \\ &= \beta_{k0}\mathbf{1}_n + \sum_{i=k-m}^{k-1} \beta_{ki}\tilde{\mathbf{x}}_{k-1,i} = \tilde{X}_{k-1}\boldsymbol{\beta}_k,\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\beta}_k &= \left(\beta_{k0}, \beta_{k,k-m}, \ldots, \beta_{k,k-1}\right)', \\ \tilde{X}_{k-1} &= \left(\mathbf{1}_n, \tilde{\mathbf{x}}_{k-1,k-m}, \ldots, \tilde{\mathbf{x}}_{k-1,k-1}\right), \\ \beta_{ki} &= \sigma_{ki}\frac{|\Sigma_{(k-2)}|}{|\Sigma_{(k-1)}|}, \quad \text{for } i = k-m, \ldots, k-1\end{aligned}$$

and

$$\tilde{\mathbf{x}}_{k-1,i} = \sum_{j=1}^{k-1}(-1)^{i+j}\frac{|M_{(k-1)}^{ji}|}{|\Sigma_{(k-2)}|}\mathbf{x}_j, \quad \text{for } i = k-m, \ldots, k-1.$$

The proposed estimators for the regression coefficients in the kth step are

$$\hat{\boldsymbol{\beta}}_k = \left(\hat{\beta}_{k0},\ \hat{\beta}_{k,k-m}, \ldots, \hat{\beta}_{k,k-1}\right)' = (\hat{X}_{k-1}'\hat{X}_{k-1})^{-1}\hat{X}_{k-1}'\mathbf{x}_k,$$

where

$$\hat{X}_{k-1} = \left(\mathbf{1}_n, \hat{\mathbf{x}}_{k-1,k-m}, \ldots, \hat{\mathbf{x}}_{k-1,k-1}\right)$$

and

$$\hat{\mathbf{x}}_{k-1,i} = \sum_{j=1}^{k-1}(-1)^{i+j}\frac{|\hat{M}_{(k-1)}^{ji}|}{|\hat{\Sigma}_{(k-2)}|}\mathbf{x}_j, \quad \text{for } i = k-m, \ldots, k-1.$$

Here the estimators from the previous terms $(1, 2, \ldots, k-1)$ are inserted in $\hat{\mathbf{x}}_{k-1,i}$ for all $i = k-m, \ldots, k-1$. The estimator for the conditional variance is given by

$$\hat{\sigma}_{k|1:k-1} = \frac{1}{n}(\mathbf{x}_k - \hat{\boldsymbol{\mu}}_{k|1:k-1})'(\mathbf{x}_k - \hat{\boldsymbol{\mu}}_{k|1:k-1})$$
$$= \frac{1}{n}\mathbf{x}_k'\left(I - \hat{X}_{k-1}(\hat{X}_{k-1}'\hat{X}_{k-1})^{-1}\hat{X}_{k-1}'\right)\mathbf{x}_k.$$

The estimators for the original parameters may be calculated as

$$\hat{\sigma}_{ki} = \hat{\beta}_{ki}\frac{|\hat{\Sigma}_{(k-1)}|}{|\hat{\Sigma}_{(k-2)}|}, \quad \text{for } i = k-m, \ldots, k-1,$$
$$\hat{\mu}_k = \hat{\beta}_{k0} + \sum_{i=k-m}^{k-1}\hat{\sigma}_{ki}\sum_{j=1}^{k-1}(-1)^{i+j}\frac{|\hat{M}_{(k-1)}^{ji}|}{|\hat{\Sigma}_{(k-1)}|}\hat{\mu}_j$$

and

$$\hat{\sigma}_{kk} = \frac{1}{n}\mathbf{x}_k'\left(I_n - \hat{X}_{k-1}(\hat{X}_{k-1}'\hat{X}_{k-1})^{-1}\hat{X}_{k-1}'\right)\mathbf{x}_k + \underline{\hat{\boldsymbol{\sigma}}}_{k1}'\hat{\Sigma}_{(k-1)}^{-1}\underline{\hat{\boldsymbol{\sigma}}}_{1k}.$$

It remains to show that the estimator $\hat{\mu}_k$ is the mean of $\mathbf{x}_k$, i.e., $\hat{\mu}_k = \frac{1}{n}\mathbf{x}_k'\mathbf{1}_n$ for all $k = 1, \ldots, p$ and a proof via induction is now presented.

Base step: For $k = 1, 2, \ldots, m+1$, $\hat{\mu}_k = \frac{1}{n}\mathbf{x}_k'\mathbf{1}_n$ since the estimators are MLEs in a model with a non-structured covariance matrix.

Inductive step: For some $m+1 < k-1$ assume that $\hat{\mu}_j = \frac{1}{n}\mathbf{x}_j'\mathbf{1}_n$, for all $j < k-1$. Then,

$$\hat{\mu}_k = \hat{\beta}_{k0} + \sum_{i=k-m}^{k-1}\hat{\sigma}_{ki}\sum_{j=1}^{k-1}(-1)^{i+j}\frac{|\hat{M}_{(k-1)}^{ji}|}{|\hat{\Sigma}_{(k-1)}|}\hat{\mu}_j$$
$$= \hat{\beta}_{k0} + \sum_{i=k-m}^{k-1}\hat{\beta}_{ki}\sum_{j=1}^{k-1}(-1)^{i+j}\frac{|\hat{M}_{(k-1)}^{ji}|}{|\hat{\Sigma}_{(k-2)}|}\frac{1}{n}\mathbf{x}_j'\mathbf{1}_n$$
$$= \hat{\beta}_{k0} + \sum_{i=k-m}^{k-1}\hat{\beta}_{ki}\frac{1}{n}\hat{\mathbf{x}}_j'\mathbf{1}_n = \frac{1}{n}\mathbf{1}_n'\hat{X}_{k-1}\hat{\boldsymbol{\beta}}_k$$
$$= \frac{1}{n}\mathbf{1}_n'\hat{X}_{k-1}(\hat{X}_{k-1}'\hat{X}_{k-1})^{-1}\hat{X}_{k-1}'\mathbf{x}_k.$$

Since $\hat{X}_{k-1}(\hat{X}_{k-1}'\hat{X}_{k-1})^{-1}\hat{X}_{k-1}'$ is a projection on a space which contains the vector $\mathbf{1}_n$ we have

$$\hat{\mu}_k = \frac{1}{n}\mathbf{1}_n'\hat{X}_{k-1}(\hat{X}_{k-1}'\hat{X}_{k-1})^{-1}\hat{X}_{k-1}'\mathbf{x}_k = \frac{1}{n}\mathbf{1}_n'\mathbf{x}_k.$$

Hence, by induction all the estimators for the expectations are means, i.e.,

$$\hat{\mu}_k = \frac{1}{n}\mathbf{x}_k'\mathbf{1}_n.$$

Although the estimators in Proposition 1 are fairly natural they are ad hoc estimators and it is important to establish some basic properties which are motivating them. We have the following theorem.

Theorem 1 *The estimator* $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_p)'$ *given in Proposition 1 is unbiased and consistent, and the estimator* $\hat{\Sigma}_{(p)}^{(m)} = (\hat{\sigma}_{ij})$ *is consistent.*

Proof The estimators of the expectations are unbiased and consistent, since these are means based on independent and identically distributed observations. The complete proof for the theorem is given by induction.

Base step: The estimator $\hat{\Sigma}_{(m+1)}$ is consistent since it is the MLE of a non-structured covariance matrix.

Inductive step: Assume that $\hat{\Sigma}_{(k-1)}$ is a consistent estimator of $\Sigma_{(k-1)}$. The estimators for the regression coefficients in the kth step are

$$\hat{\boldsymbol{\beta}}_k = (\hat{X}'_{k-1}\hat{X}_{k-1})^{-1}\hat{X}'_{k-1}\mathbf{x}_k = \left(\frac{1}{n}\hat{X}'_{k-1}\hat{X}_{k-1}\right)^{-1}\left(\frac{1}{n}\hat{X}'_{k-1}\mathbf{x}_k\right), \tag{8}$$

where the first part in the right hand side of (8) converges in probability as follows. We have

$$\frac{1}{n}\hat{X}'_{k-1}\hat{X}_{k-1}$$
$$= \begin{pmatrix} 1 & \frac{1}{n}\mathbf{1}'_n\hat{\mathbf{x}}_{k-1,k-m} & \cdots & \frac{1}{n}\mathbf{1}'_n\hat{\mathbf{x}}_{k-1,k-1} \\ \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-m}\mathbf{1}_n & \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-m}\hat{\mathbf{x}}_{k-1,k-m} & \cdots & \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-m}\hat{\mathbf{x}}_{k-1,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-1}\mathbf{1}_n & \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-1}\hat{\mathbf{x}}_{k-1,k-m} & \cdots & \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-1}\hat{\mathbf{x}}_{k-1,k-1} \end{pmatrix}.$$

For $i, l = 1, 2, \ldots, m$

$$\frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-i}\hat{\mathbf{x}}_{k-1,k-l} = |\hat{\Sigma}_{(k-2)}|^{-2}\sum_{j=1,q=1}^{k-1}(-1)^{j-i+q-l}|\hat{M}_{(k-1)}^{j,k-i}||\hat{M}_{(k-1)}^{q,k-l}|\frac{1}{n}\mathbf{x}'_j\mathbf{x}_q$$
$$\xrightarrow{p} |\Sigma_{(k-2)}|^{-2}\sum_{j=1,q=1}^{k-1}(-1)^{j-i+q-l}|M_{(k-1)}^{j,k-i}||M_{(k-1)}^{q,k-l}|(\sigma_{jq} + \mu_j\mu_q) \equiv w_{il}$$

and

$$\frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-i}\mathbf{1}_n = |\hat{\Sigma}_{(k-2)}|^{-1}\sum_{j=1}^{k-1}(-1)^{k-i+j}|\hat{M}_{(k-1)}^{j,k-i}|\frac{1}{n}\mathbf{x}'_j\mathbf{1}_n$$
$$\xrightarrow{p} |\Sigma_{(k-2)}|^{-1}\sum_{j=1}^{k-1}(-1)^{k-i+j}|M_{(k-1)}^{j,k-i}|\mu_j \equiv w_i,$$

since the estimators are assumed to be consistent for the $(k-1)$th step, by the weak law of large numbers and by Cramér–Slutsky's theorem (Cramér 1946). Hence,

$$\frac{1}{n}\hat{X}'_{k-1}\hat{X}_{k-1} \xrightarrow{p} W, \quad \text{as } n \to \infty,$$

where

$$W = \begin{pmatrix} 1 & w_m & \cdots & w_1 \\ w_m & w_{mm} & \cdots & w_{m1} \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_{1m} & \cdots & w_{11} \end{pmatrix}.$$

Turning to the second part in the right hand side of (8)

$$\frac{1}{n}\hat{X}'_{k-1}\mathbf{x}_k = \begin{pmatrix} \frac{1}{n}\mathbf{1}'_n\mathbf{x}_k \\ \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-m}\mathbf{x}_k \\ \vdots \\ \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-1}\mathbf{x}_k \end{pmatrix},$$

where

$$\begin{aligned} \frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-i}\mathbf{x}_k &= |\hat{\Sigma}_{(k-2)}|^{-1}\sum_{j=1}^{k-1}(-1)^{k-i+j}|\hat{M}^{j,k-i}_{(k-1)}|\frac{1}{n}\mathbf{x}'_j\mathbf{x}_k \\ &\xrightarrow{p} |\Sigma_{(k-2)}|^{-1}\sum_{j=1}^{k-1}(-1)^{k-i+j}|M^{j,k-i}_{(k-1)}|(\sigma_{jk}+\mu_j\mu_k) \\ &= |\Sigma_{(k-2)}|^{-1}\left(\sum_{j=k-m}^{k-1}(-1)^{k-i+j}|M^{j,k-i}_{(k-1)}|\sigma_{jk}\right. \\ &\qquad \left. +\sum_{j=1}^{k-1}(-1)^{k-i+j}|M^{j,k-i}_{(k-1)}|\mu_j\mu_k\right) \equiv v_i. \end{aligned}$$

Hence,

$$\frac{1}{n}\hat{X}'_{k-1}\mathbf{x}_k \xrightarrow{p} \mathbf{V}, \quad \text{as } n \to \infty,$$

where $\mathbf{V} = (\mu_k, v_m, \ldots, v_1)'$ and $\hat{\boldsymbol{\beta}}_k \xrightarrow{p} W^{-1}\mathbf{V}$, as $n \to \infty$. Let

$$\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{k,k-m}, \ldots, \beta_{k,k-1})'$$
$$= \begin{pmatrix} \mu_k - |\Sigma_{(k-1)}|^{-1} \sum_{i=k-m}^{k-1} \sigma_{ki} \sum_{j=1}^{k-1} (-1)^{i+j} |M_{(k-1)}^{ji}| \mu_j \\ |\Sigma_{(k-2)}||\Sigma_{(k-1)}|^{-1} \sigma_{k,k-m} \\ \vdots \\ |\Sigma_{(k-2)}||\Sigma_{(k-1)}|^{-1} \sigma_{k,k-1} \end{pmatrix},$$

and it will be shown that $W\boldsymbol{\beta}_k = \mathbf{V}$, i.e.,

$$(1, w_m, \ldots, w_1)\boldsymbol{\beta}_k = \mu_k, \tag{9}$$
$$(w_r,\ w_{rm}, \ldots, w_{r1})\boldsymbol{\beta}_k = w_r \beta_{k0} + \sum_{i=1}^{m} w_{ri} \beta_{k,k-i} = v_r, \quad r = 1, 2, \ldots, m. \tag{10}$$

First, consider (9),

$$\begin{aligned} (1, w_m, \ldots, w_1)\boldsymbol{\beta}_k &= \beta_{k0} + \sum_{i=1}^{m} w_i \beta_{k,k-i} \\ &= \mu_k - |\Sigma_{(k-1)}|^{-1} \sum_{i=k-m}^{k-1} \sigma_{ki} \sum_{j=1}^{k-1} (-1)^{i+j} |M_{(k-1)}^{ji}| \mu_j \\ &\quad + \sum_{i=1}^{m} |\Sigma_{(k-1)}|^{-1} \sigma_{k,k-i} \sum_{j=1}^{k-1} (-1)^{k-i+j} |M_{(k-1)}^{j,k-i}| \mu_j = \mu_k. \end{aligned}$$

Second, consider (10) when $r = m$. The other cases ($r < m$) are verified in the same way. The following chain of calculation exhibits the result:

$$\begin{aligned} w_m \beta_{k0} + \sum_{i=1}^{m} w_{mi} \beta_{k,k-i} &= |\Sigma_{(k-2)}|^{-1} \left\{ \sum_{j=1}^{k-1} (-1)^{k-m+j} |M_{(k-1)}^{j,k-m}| \mu_j \mu_k \right. \\ &\left. + |\Sigma_{(k-1)}|^{-1} \sum_{i=1}^{m} \sigma_{k,k-i} \sum_{j=1}^{k-1} (-1)^{j+k-i} |M_{(k-1)}^{j,k-m}| \underbrace{\sum_{q=1}^{k-1} (-1)^{k-i+q} |M_{(k-1)}^{q,k-i}| \sigma_{jq}}_{=0, \text{when } k-i \neq j} \right\} \end{aligned}$$

$$
\begin{aligned}
&= |\Sigma_{(k-2)}|^{-1}\left\{\sum_{j=1}^{k-1}(-1)^{k-m+j}|M_{(k-1)}^{j,k-m}|\mu_j\mu_k\right.\\
&\quad\left.+\sum_{i=1}^{m}(-1)^{m+i}\sigma_{k,k-i}|\Sigma_{(k-1)}|^{-1}|M_{(k-1)}^{k-i,k-m}|\underbrace{\sum_{q=1}^{k-1}(-1)^{k-i+q}|M_{(k-1)}^{q,k-i}|\sigma_{k-i,q}}_{=|\Sigma_{(k-1)}|}\right\}\\
&= |\Sigma_{(k-2)}|^{-1}\left(\sum_{j=1}^{m}(-1)^{m+j}\sigma_{k,k-j}|M_{(k-1)}^{k-j,k-m}|\right.\\
&\quad\left.+\sum_{j=1}^{k-1}(-1)^{k-m+j}|M_{(k-1)}^{j,k-m}|\mu_j\mu_k\right) = v_m.
\end{aligned}
$$

Thus, it has been shown that $\hat{\boldsymbol{\beta}}_k \xrightarrow{p} \boldsymbol{\beta}_k$, as $n \to \infty$ and we are now able to show consistency for the estimators. By Cramér's theorem and since the estimators are assumed to be consistent for the $(k-1)$th step, we have

$$
\hat{\sigma}_{ki} = \hat{\beta}_{ki}\frac{|\hat{\Sigma}_{(k-1)}|}{|\hat{\Sigma}_{(k-2)}|} \xrightarrow{p} \beta_{ki}\frac{|\Sigma_{(k-1)}|}{|\Sigma_{(k-2)}|} = \sigma_{ki}, \quad \text{for } i = k-m, \ldots, k-1
$$

and

$$
\begin{aligned}
\hat{\sigma}_{kk} &= \frac{1}{n}\mathbf{x}_k'\left(I_n - \hat{X}_{k-1}(\hat{X}_{k-1}'\hat{X}_{k-1})^{-1}\hat{X}_{k-1}'\right)\mathbf{x}_k + \underline{\hat{\boldsymbol{\sigma}}}_{k1}'\hat{\Sigma}_{(k-1)}^{-1}\underline{\hat{\boldsymbol{\sigma}}}_{1k}\\
&\xrightarrow{p} \sigma_{kk} + \mu_k^2 - \mathbf{V}'\boldsymbol{\beta} + \underline{\boldsymbol{\sigma}}_{k1}'\Sigma_{(k-1)}^{-1}\underline{\boldsymbol{\sigma}}_{1k}.
\end{aligned}
$$

However,

$$
\begin{aligned}
\mathbf{V}'\boldsymbol{\beta} &= \mu_k\beta_{k0} + \sum_{i=1}^{m} v_i\beta_{k,k-i}\\
&= \mu_k^2 + |\Sigma_{(k-1)}|^{-1}\left\{-\sum_{i=1}^{m}\sigma_{k,k-i}\sum_{j=1}^{k-1}(-1)^{k-i+j}|M_{(k-1)}^{j,k-i}|\mu_j\mu_k + \sum_{i=1}^{m}\sigma_{k,k-i}\right.\\
&\quad\left.\times\left(\sum_{j=1}^{m}(-1)^{i+j}|M_{(k-1)}^{k-j,k-i}|\sigma_{k-j,k} + \sum_{j=1}^{k-1}(-1)^{k-i+j}|M_{(k-1)}^{j,k-i}|\mu_j\mu_k\right)\right\}\\
&= \mu_k^2 + \sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{k,k-i}(-1)^{i+j}\frac{|M_{(k-1)}^{k-j,k-i}|}{|\Sigma_{(k-1)}|}\sigma_{k-j,k} = \mu_k^2 + \underline{\boldsymbol{\sigma}}_{k1}'\Sigma_{(k-1)}^{-1}\underline{\boldsymbol{\sigma}}_{1k}
\end{aligned}
$$

and hence

$$\hat{\sigma}_{kk} \xrightarrow{p} \sigma_{kk}.$$

Therefore, it has been shown by induction that the estimator $\hat{\Sigma}^{(m)}_{(p)} = (\hat{\sigma}_{ij})$ is consistent. □

We conclude this section by presenting the estimators when $m = 1$, i.e., a banded matrix of order one. In this case it is straightforward to calculate the inverse of the matrix $\hat{X}'_{k-1}\hat{X}_{k-1}$ since it is a 2×2 matrix. Hence, the coefficients $\hat{\beta}_{k,k-1}$ from (6) can be written explicitly. Furthermore, the estimators $\hat{\sigma}_{k,k+1}$ and $\hat{\sigma}_{kk}$ follow from (4) and (5), respectively. The estimators have the same structure as in the general case but now we can better see the underlying structure.

Proposition 2 *Let $X \sim N_{p,n}\left(\boldsymbol{\mu}\mathbf{1}'_n, \Sigma^{(1)}_{(p)}, I_n\right)$. The estimators given in Proposition 1 equal*

$$\hat{\mu}_k = \frac{1}{n}\mathbf{x}'_k\mathbf{1}_n, \quad \hat{\sigma}_{kk} = \frac{1}{n}\mathbf{x}'_k C\mathbf{x}_k, \quad \textit{for } k = 1, \ldots, p,$$
$$\hat{\sigma}_{k,k+1} = \frac{1}{n}\hat{\mathbf{x}}'_k C\mathbf{x}_{k+1}, \quad \textit{for } k = 1, \ldots, p-1,$$

where $\hat{\mathbf{x}}_1 = \mathbf{x}_1$, $\hat{\mathbf{x}}_k = \mathbf{x}_k - \hat{\beta}_{k,k-1}\hat{\mathbf{x}}_{k-1}$ *for* $k = 2, \ldots, p-1$,

$$\hat{\beta}_{k,k-1} = \frac{\hat{\mathbf{x}}'_{k-1}C\mathbf{x}_k}{\hat{\mathbf{x}}'_{k-1}C\hat{\mathbf{x}}_{k-1}}$$

and

$$C = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n.$$

4 Simulation

The examples presented here illustrate the results obtained in the previous sections.

In each simulation a sample of size $n = 100$ observations was randomly generated from a p-variate normal distribution using Release 14 of MATLAB Version 7.0.1 (The Mathworks Inc., Natick, MA, USA). Next, the explicit estimators were calculated in each simulation. Simulations were repeated 500 times and the average values of the obtained estimators were calculated.

Two cases were studied. The first of them correspond to $m = 1$, and the second one considers the case $m = 2$.

Simulations for $p = 5, m = 1$
Data was generated with parameters

$$\mu = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 3 & -2 & 0 & 0 \\ 0 & -2 & 4 & -1 & 0 \\ 0 & 0 & -1 & 5 & 2 \\ 0 & 0 & 0 & 2 & 6 \end{pmatrix}.$$

Based on 500 simulations the average estimators are given by

$$\hat{\mu} = \begin{pmatrix} 0.9937 \\ 1.9923 \\ 3.0016 \\ 4.0083 \\ 4.9901 \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} 1.9642 & 1.0020 & 0 & 0 & 0 \\ 1.0020 & 3.0047 & -1.9968 & 0 & 0 \\ 0 & -1.9968 & 4.0016 & -0.9828 & 0 \\ 0 & 0 & -0.9828 & 4.9589 & 1.9869 \\ 0 & 0 & 0 & 1.9869 & 5.9871 \end{pmatrix}.$$

Simulations for $p = 4$, $m = 2$
Corresponding to the previous case the model is defined through

$$\mu = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 3 & 2 & 1 \\ 1 & 2 & 4 & 1 \\ 0 & 1 & 1 & 5 \end{pmatrix}.$$

After 500 simulations average explicit estimators equal

$$\hat{\mu} = \begin{pmatrix} 1.0013 \\ 2.0015 \\ 3.0005 \\ 4.0073 \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} 1.9875 & 0.9996 & 0.9923 & 0 \\ 0.9996 & 3.0049 & 1.9953 & 0.9741 \\ 0.9923 & 1.9953 & 4.0031 & 1.0021 \\ 0 & 0.9741 & 1.0021 & 4.9911 \end{pmatrix}.$$

We have also compared the explicit estimators derived in our study and the MLEs computed with the statistical software SAS (SAS Institute Inc., Cary, NC, USA). In SAS PROC MIXED the banded covariance structure is one of the options for the covariance structure. The explicit estimators are very close to the MLEs and since both the explicit estimators and the MLEs are consistent, they are asymptotically equivalent. Hence, they should be close to each other.

One conclusion from the above simulations is that the explicit estimators derived in this paper perform very well and indeed are as close to the true values as the iterative MLEs.

Acknowledgments The work of Z. Andrushchenko was supported by Swedish Research Council, VR 621-2002-5578.

References

Andersson, S., Brøns, H., Jensen, S. (1983). Distribution of eigenvalues in multivariate statistical analysis. *The Annals of Statistics, 11*(2), 392–415.

Browne, M. W. (1977). The analysis of patterned correlation matrices by generalized least squares. *British Journal of Mathematical and Statistical Psychology, 30*, 113–124.

Chakraborty, M. (1998). An efficient algorithm for solving general periodic Toeplitz system. *IEEE Transactions on Signal Processing, 46*(3), 784–787.

Chinchilli, V. M., Carter, W. (1984). A likelihood ratio test for a patterned covariance matrix in a multivariate growth-curve model. *Biometrics, 40*(1), 151–156.

Christensen, L. P. B. (2007). An EM-algorithm for band-toeplitz covariance matrix estimation. Submitted to IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Cramér, H. (1946). *Mathematical methods of statistics* (pp. 254–255). Princeton: Princeton University Press.

Godolphin, E. J., De Gooijer, J. G. (1982). On the maximum likelihood estimation of the parameters of a Gaussian moving average process. *Biometrika, 69*, 443–451.

Goodman, N. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics, 34*(1), 152–177.

Jöreskog, K. G. (1981). Analysis of covariance structures. With discussion by Andersen, E. B., Kiiveri, H., Laake, P., Cox, D. B., Schweder, T. With a reply by the author. *Scandinavian Journal of Statistics, 8*(2), 65–92.

Kollo, T., von Rosen, D. (2005). *Advanced multivariate statistics with matrices*. Dordrecht: Springer.

Lauritzen, S. L. (1996). *Graphical models*. New York: The Clarendon Press, Oxford University Press.

Lu, N., Zimmerman, D. (2005). The likelihood ratio test for a separable covariance matrix. *Statistics and Probability Letters, 73*(4), 449–457.

Marin, J., Dhorne, T. (2002). Linear Toeplitz covariance structure models with optimal estimators of variance components. *Linear Algebra and Its Applications, 354*, 195–212.

Moura, J. M. F., Balram, N. (1992). Recursive structure of noncausal Gauss Markov random fields. *IEEE Transactions on Information Theory, 38*(2), 334–354.

Muirhead, R. (1982). *Aspects of multivariate statistical theory*. New York: Wiley.

Nahtman, T. (2006). Marginal permutation invariant covariance matrices with applications to linear models. *Linear Algebra and Its Applications, 417*(1), 183–210.

Naik, D., Rao, S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *Journal of Applied Statistics, 28*(1), 91–105.

Olkin, I. (1973). Testing and estimation for structures which are circularly symmetric in blocks. In D. G. Kabe, R. P. Gupta (Eds.), *Multivariate statistical inference* (pp. 183–195). Amsterdam: North-Holland.

Olkin, I., Press, S. (1969). Testing and estimation for a circular stationary model. *The Annals of Mathematical Statistics, 40*, 1358–1373.

Roy, A., Khattree, R. (2005). On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data. *Statistical Methodology, 2*(4), 297–306.

Searle, S. R., Casella, G., McCulloch, C. E. (1992). *Variance components*. New York: Wiley, Hoboken.

Srivastava, M. S. (2002). *Methods of multivariate statistics*. New York: Wiley-Interscience.

Szatrowski, T. H. (1985). Asymptopic distributions in the testing and estimation of the missing-data multivariate normal linear patterned mean and correlation matrix. *Linear Algebra and its Applications, 67*, 215–231.

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

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

Woods, J. W. (1972). Two-dimensional discrete Markovian fields. *IEEE Transactions on Information Theory, IT-18*(2), 232–240.