Explicit estimators of the parameters in a multivariate normal distribution when the covariance matrix is banded of order m

M. Ohlson, Zh. Andrushchenko, and D. von Rosen

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2008:3
ISSN 1651-8543

Explicit estimators of the parameters in a multivariate normal distribution when the covariance matrix is banded of order m

Martin Ohlson
Linköping University, Linköping, Sweden

Zhanna Andrushchenko [1]
Centre of Biostochastics, Swedish University of Agricultural Sciences
P.O. Box 7032, SE-750 07 Uppsala, Sweden

Dietrich von Rosen
Centre of Biostochastics, Swedish University of Agricultural Sciences
P.O. Box 7032, SE-750 07 Uppsala, Sweden

Abstract

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

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

AMS classification: 62H12, 62F12, 62F30

[1] E-mail address to the correspondence author: Zhanna.Andrushchenko@bt.slu.se

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 multivariate observations are often, at least approximately, normally distributed. Moreover, the multivariate normal distribution is mathematically tractable. Since normally distributed data can be modelled entirely in terms of their means and variances/covariances, these parameters specify the complete probability distribution of data. 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 mention some papers. Wilks (1946), is one of the early papers dealing with patterned structures, considered a set of 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 their distance. 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 has 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 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 to partial differential equations based on finite difference. 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 report 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.

In univariate analysis when not assuming any particular distribution one often constructs estimators via a least squares criterion or a generalized least squares criterion. An alternative is to construct estimators so that some property is fulfilled. In multivariate analysis one has tried to copy the univariate approaches by constructions of various criteria which remind on the univariate ones. However, there is a fundamental difference between the univariate and multivariate setups. In the univariate case we have independent observations with a scalar variance. The variance will not effect the estimator of the mean in, for example, the least squares approach. In a multivariate setting we have independent observations with both a covariance structure and a mean structure. When estimating parameters we should often consider both structures simultaneously. For example, we know from studies concerning the growth curve model (Potthoff and Roy, 1964; Kollo and von Rosen, 2005) that estimators of the mean structure and the covariance matrix are connected. 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. However, little attention has been given to a derivation of explicit analytical expressions for estimators of the model with banded covariance structure under m-dependence.

In this report we focus on estimation of parameters of a multivariate random vector (with dimensionality p and sample size n) under the constraint that certain covariances are zero, namely we consider a banded covariance structure under m-dependence. To avoid singularity, we limit ourselves to the case when the matrix dimension p is small compared to the sample size n. The aim in our research is to present explicit estimators for the mean and the covariance matrix under m-dependence. In many applications, e.g., in image analysis, computations are heavy and explicit expressions of estimators are more useful than iterative algorithms obtained for MLEs or restricted MLEs. In this report a simple estimation procedure is suggested which gives unbiased and consistent estimators of the mean and consistent estimators of the covariance matrix.

The report is organized as follows. In Section 2, we present the main definitions and notation used throughout the report. Section 3 provides the algorithm for estimation of the mean and the covariance matrix when $m = 1$. First, the univariate ($p = 1$) and the bivariate ($p = 2$) cases are presented. Estimators are calculated and their properties are analyzed. It is shown that the estimators obtained coincide with the usual MLE. Next, the three-variate case, $p = 3$, is analyzed in detail. Here the algorithm consists of maximizing the likelihood function via inserting the estimated parameters from previous steps. Finally, the general p-variate case is considered. Again, the properties of the estimators are presented. Section 4 provides a further generalization of the approach. Here we consider the case of arbitrary $m > 1$. In the section the main proposition of the report is formulated and motivated, as well as estimators and their properties are presented. Finally, in Section 5 we present some simulations and Section 6 summarizes the report.

2 Definitions and Notation

Throughout this report matrices will be denoted by capital letters, vectors by bold font, scalars and elements in matrices by ordinary letters if nothing else is stated.

Let X be matrix normally distributed (Kollo and von Rosen, 2005), $\mathrm{X} \sim N_{p,n}(\Xi, \Sigma, \mathrm{I}_n)$, where I_n is the identity matrix of dimension n and partition X as

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

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.,

$$X_{i:j} = \begin{pmatrix} \mathbf{x}'_i \\ \mathbf{x}'_{i+1} \\ \vdots \\ \mathbf{x}'_j \end{pmatrix}.$$

In this report we suppose that the expectation is given by

$$\Xi = E(X) = \boldsymbol{\mu}_p \mathbf{1}'_n,$$

where

$$\boldsymbol{\mu}'_p = (\mu_1,\ \mu_2, \ldots, \mu_p)$$

and

$$\mathbf{1}'_n = (1,\ 1, \ldots, 1) : (1 \times n).$$

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

$$\Sigma_{(k)}^{(m)} = \begin{pmatrix} \sigma_{11} & \ldots & \sigma_{1,m+1} & 0 & 0 & \ldots & 0 \\ \sigma_{21} & \ldots & \sigma_{2,m+1} & \sigma_{2,m+2} & 0 & \ldots & 0 \\ \vdots & & & & \ddots & & \\ \sigma_{m+1,1} & & & & & & \\ 0 & \ddots & & & & & \\ \vdots & & & & & & \\ & & & \ddots & & & \\ 0 & \ldots & & 0 & \sigma_{k-1,k-(m+1)} & \sigma_{k-1,k-m} & \ldots & \sigma_{k-1,k} \\ 0 & \ldots & & 0 & 0 & \sigma_{k,k-m} & \ldots & \sigma_{kk} \end{pmatrix}. \tag{2.1}$$

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

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

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

where

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

The likelihood function of the parameters $\boldsymbol{\mu}$ and Σ equals

$$c|\Sigma|^{-n/2}\mathrm{etr}\left\{-\frac{1}{2}\Sigma^{-1}\left(\mathrm{X}-\boldsymbol{\mu}\mathbf{1}'_n\right)\left(\right)'\right\},$$

where c is a normalizing constant and

$$\left(\mathrm{X}-\boldsymbol{\mu}\mathbf{1}'_n\right)\left(\right)' = \left(\mathrm{X}-\boldsymbol{\mu}\mathbf{1}'_n\right)\left(\mathrm{X}-\boldsymbol{\mu}\mathbf{1}'_n\right)'.$$

If we partition $\mathrm{X}, \boldsymbol{\mu}$ and Σ as

$$\mathrm{X} = \begin{pmatrix} \mathrm{X}_1 \\ \mathrm{X}_2 \end{pmatrix} \sim N_{p,n}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}\mathbf{1}'_n, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \mathrm{I}_n\right)$$

we obtain the conditional distribution (see (Kollo and von Rosen, 2005))

$$\mathrm{X}_2|X_1 \sim N_{r,n}\left(\Xi_{2|1}, \Sigma_{2|1}, \mathrm{I}_n\right), \tag{2.2}$$

where

$$\Xi_{2|1} = \boldsymbol{\mu}_2\mathbf{1}'_n + \Sigma_{21}\Sigma_{11}^{-1}\left(\mathrm{X}_1 - \boldsymbol{\mu}_1\mathbf{1}'_n\right), \tag{2.3}$$

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{2.4}$$

and the corresponding conditional likelihood function equals

$$c|\Sigma_{11}|^{-n/2}\mathrm{etr}\left\{-\frac{1}{2}\Sigma_{11}^{-1}\left(\mathrm{X}_1-\boldsymbol{\mu}_1\mathbf{1}'_n\right)\left(\right)'\right\}$$

$$\times|\Sigma_{2|1}|^{-n/2}\mathrm{etr}\left\{-\frac{1}{2}\Sigma_{2|1}^{-1}\left(\mathrm{X}_2-\Xi_{2|1}\right)\left(\right)'\right\}. \tag{2.5}$$

3 m-dependence of order one ($m = 1$)

In this section we present the idea of how to estimate the expectation and the covariance matrix for a multivariate normal distribution when the covariance matrix is banded of order one, i.e., $m = 1$. For the sake of completeness and since our estimators are some kind of recursive estimators we start with the univariate and bivariate cases. In these cases we have a non-structured covariance matrix and the estimators are the maximum likelihood estimators (MLE) but for the three-dimensional case and for higher dimensions there are zeros in the covariance matrix, i.e., the covariance matrix is structured. Hence, in the second half of this section we present our estimators. Since they are ad hoc estimators we conclude the section by establishing some properties such as unbiasness and consistency.

3.1 Univariate and bivariate cases

For the univariate case when $p = 1$ we have the sample observation vector

$$\mathbf{x}_1' \sim N_{1,n}(\mu_1 \mathbf{1}_n', \sigma_{11}, \mathrm{I}_n).$$

The likelihood can be maximized with respect to the two parameters, μ_1 and σ_{11}. The maximum likelihood estimators of these parameters are given by

$$\hat{\mu}_1 = \frac{1}{n}\mathbf{x}_1'\mathbf{1}_n,$$
$$\hat{\sigma}_{11} = \frac{1}{n}(\mathbf{x}_1 - \hat{\mu}_1\mathbf{1}_n)'() = \frac{1}{n}\mathbf{x}_1'\mathrm{C}\mathbf{x}_1,$$

where

$$\mathrm{C} = \mathrm{I}_n - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'. \tag{3.1}$$

For the bivariate case when $p = 2$, there are five unknown parameters, μ_1, μ_2, σ_{11}, σ_{12} and σ_{22}. If we condition $\mathbf{x}_2'$ on $\mathbf{x}_1'$ we have from equation (2.2)

$$\mathbf{x}_1' \sim N_{1,n}(\mu_1\mathbf{1}_n', \sigma_{11}, \mathrm{I}_n),$$
$$\mathbf{x}_2'|\mathbf{x}_1' = \mathbf{x}_{2|1}' \sim N_{1,n}(\boldsymbol{\mu}_{2|1}', \sigma_{2|1}, \mathrm{I}_n),$$

where the conditional expectation, given by (2.3), equals

$$\begin{aligned}\boldsymbol{\mu}_{2|1} &= \mu_2\mathbf{1}_n + \sigma_{21}\sigma_{11}^{-1}(\mathbf{x}_1 - \mu_1\mathbf{1}_n)\\ &= \mu_2\mathbf{1}_n + \beta_{21}(\mathbf{x}_1 - \mu_1\mathbf{1}_n)\\ &= \beta_{20}\mathbf{1}_n + \beta_{21}\mathbf{x}_1\end{aligned} \tag{3.2}$$

and the conditional variance, given by (2.4), equals

$$\sigma_{2|1} = \sigma_{22} - \sigma_{21}\,\sigma_{11}^{-1}\,\sigma_{12} = \sigma_{22} - \beta_{21}^2\,\sigma_{11}, \tag{3.3}$$

where

$$\beta_{20} = \mu_2 - \beta_{21}\mu_1, \tag{3.4}$$
$$\beta_{21} = \frac{\sigma_{21}}{\sigma_{11}}. \tag{3.5}$$

The likelihood function based on all the observations is now given by (2.5) and equals

$$c\sigma_{11}^{-n/2}\exp\left(-\frac{1}{2}\sigma_{11}^{-1}(\mathbf{x}_1 - \mu_1\mathbf{1}_n)'()\right)\sigma_{2|1}^{-n/2}\exp\left(-\frac{1}{2}\sigma_{2|1}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_{2|1})'()\right).$$

Here we see that, as usual, we have a regression

$$\mathbf{x}_{2|1} = \boldsymbol{\mu}_{2|1} + \epsilon = \beta_{20}\mathbf{1}_n + \beta_{21}\mathbf{x}_1 + \epsilon$$

where

$$\epsilon \sim N_{n,1}(0, \mathrm{I}_n, \sigma_{2|1}).$$

The maximum likelihood estimator of $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21})'$ can be written in matrix form as

$$\hat{\boldsymbol{\beta}}_2 = \begin{pmatrix} \hat{\beta}_{20} \\ \hat{\beta}_{21} \end{pmatrix} = (\hat{\mathrm{X}}_1'\hat{\mathrm{X}}_1)^{-1}\hat{\mathrm{X}}_1'\mathbf{x}_2, \tag{3.6}$$

where

$$\hat{\mathrm{X}}_1 = (\mathbf{1}_n : \mathbf{x}_1)$$

and the inverse in equation (3.6) exists since the matrix $\hat{\mathrm{X}}_1'\hat{\mathrm{X}}_1$ has full rank. An estimator of the conditional variance is given by

$$\hat{\sigma}_{2|1} = \frac{1}{n}(\mathbf{x}_2 - \hat{\boldsymbol{\mu}}_{2|1})'(),$$

where

$$\hat{\boldsymbol{\mu}}_{2|1} = \hat{\beta}_{20}\mathbf{1}_n + \hat{\beta}_{21}\mathbf{x}_1.$$

We may now calculate the regression coefficients

$$\hat{\boldsymbol{\beta}}_2 = \frac{1}{n\mathbf{x}_1'\mathrm{C}\mathbf{x}_1} \begin{pmatrix} \mathbf{x}_1'(\mathbf{x}_1\mathbf{1}_n' - \mathbf{1}_n\mathbf{x}_1')\mathbf{x}_2 \\ n\mathbf{x}_1'\mathrm{C}\mathbf{x}_2 \end{pmatrix},$$

where C is as in (3.1). Hence,

$$\hat{\beta}_{20} = \frac{\mathbf{x}_1'(\mathbf{x}_1\mathbf{1}_n' - \mathbf{1}_n\mathbf{x}_1')\mathbf{x}_2}{n\mathbf{x}_1'\mathrm{C}\mathbf{x}_1}, \tag{3.7}$$

$$\hat{\beta}_{21} = \frac{\mathbf{x}_1'\mathrm{C}\mathbf{x}_2}{\mathbf{x}_1'\mathrm{C}\mathbf{x}_1} = \frac{(\mathbf{x}_1 - \hat{\mu}_1\mathbf{1}_n)'(\mathbf{x}_2 - \hat{\mu}_2\mathbf{1}_n)}{(\mathbf{x}_1 - \hat{\mu}_1\mathbf{1}_n)'()}. \tag{3.8}$$

The estimators for the five initial parameters are now easily calculated

$$\begin{aligned}
\hat{\mu}_1 &= \frac{1}{n}\mathbf{x}_1'\mathbf{1}_n,\\
\hat{\sigma}_{11} &= \frac{1}{n}\mathbf{x}_1'\mathrm{C}\mathbf{x}_1,\\
\hat{\mu}_2 &= \hat{\beta}_{20} + \hat{\beta}_{21}\hat{\mu}_1 = \frac{1}{n}\mathbf{x}_2'\mathbf{1}_n,\\
\hat{\sigma}_{22} &= \hat{\sigma}_{2|1} + \hat{\beta}_{21}^2\hat{\sigma}_{11} = \frac{1}{n}\mathbf{x}_2'\mathrm{C}\mathbf{x}_2,\\
\hat{\sigma}_{12} &= \hat{\sigma}_{21} = \hat{\beta}_{21}\hat{\sigma}_{11} = \frac{1}{n}\mathbf{x}_1'\mathrm{C}\mathbf{x}_2.
\end{aligned}$$

These are, of course, the usual maximum likelihood estimator since the covariance matrix in the bivariate case is an ordinary non-structured covariance matrix.

Furthermore, now some calculations are presented which later will be generalized. We can easily calculate the expectations and variances for the β_{21} estimator:

$$\begin{aligned}
E(\hat{\beta}_{21}|\mathbf{x}_1) &= \beta_{21},\\
Var(\hat{\beta}_{21}|\mathbf{x_1}) &= \frac{n\sigma_{2|1}}{n\mathbf{x}_1'\mathbf{x}_1 - (\mathbf{1}_n'\mathbf{x}_1)^2} = \frac{\sigma_{2|1}}{\mathbf{x}_1'\mathrm{C}\mathbf{x}_1}
\end{aligned}$$

so

$$E(\hat{\beta}_{21}) = E(E(\hat{\beta}_{21}|\mathbf{x}_1)) = \beta_{21}$$

and

$$\begin{aligned}
Var(\hat{\beta}_{21}) &= E(Var(\hat{\beta}_{21}|\mathbf{x}_1)) + Var(E(\hat{\beta}_{21}|\mathbf{x}_1))\\
&= \sigma_{2|1}E\left(\frac{1}{\mathbf{x}_1'\mathrm{C}\mathbf{x}_1}\right) = \frac{\sigma_{2|1}}{\sigma_{11}}E\left(\frac{1}{\chi^2_{n-1}}\right) = \frac{\sigma_{2|1}}{\sigma_{11}(n-3)}.
\end{aligned}$$

Since $\hat{\beta}_{21}$ is unbiased and the variance converges zero, as $n \to \infty$, $\hat{\beta}_{21}$ is a consistent estimator.

3.2 Three-dimensional case, $p = 3$

We have shown that for the univariate and bivariate cases the estimators are the MLE since the covariance matrix in each case is non-structured. In the three-dimensional case when the covariance matrix is banded of order one there are two zeros in the covariance matrix, i.e., the variables $\mathbf{x}_1$ and $\mathbf{x}_3$ are independent. We now propose explicit estimators for the expectation and covariance matrix in the three-dimensional case. Some properties for the estimators will also be proven.

Proposition 3.1 *Let* $\mathrm{X} \sim N_{3,n}\left(\boldsymbol{\mu}_3\mathbf{1}_n', \Sigma_{(3)}^{(1)}, \mathrm{I}_n\right)$. *Explicit estimators are given by*

$$\hat{\mu}_1 = \frac{1}{n}\mathbf{x}_1'\mathbf{1}_n, \qquad \hat{\sigma}_{11} = \frac{1}{n}\mathbf{x}_1'\mathrm{C}\mathbf{x}_1, \qquad \hat{\sigma}_{12} = \frac{1}{n}\mathbf{x}_1'\mathrm{C}\mathbf{x}_2,$$
$$\hat{\mu}_2 = \frac{1}{n}\mathbf{x}_2'\mathbf{1}_n, \qquad \hat{\sigma}_{22} = \frac{1}{n}\mathbf{x}_2'\mathrm{C}\mathbf{x}_2, \qquad \hat{\sigma}_{23} = \frac{1}{n}\hat{\mathbf{x}}_2'\mathrm{C}\mathbf{x}_3,$$
$$\hat{\mu}_3 = \frac{1}{n}\mathbf{x}_3'\mathbf{1}_n, \qquad \hat{\sigma}_{33} = \frac{1}{n}\mathbf{x}_3'\mathrm{C}\mathbf{x}_3,$$

where

$$\hat{\mathbf{x}}_2 = \mathbf{x}_2 - \hat{\beta}_{21}\mathbf{x}_1, \qquad \hat{\beta}_{21} = \frac{\mathbf{x}_1'\mathrm{C}\mathbf{x}_2}{\mathbf{x}_1'\mathrm{C}\mathbf{x}_1}$$

and C *is given in (3.1).*

In the next these estimators are motivated. Since $\mathrm{X} \sim N_{3,n}\left(\boldsymbol{\mu}_3\mathbf{1}_n', \Sigma_{(3)}^{(1)}, \mathrm{I}_n\right)$ we have from equation (2.2) that

$$\begin{aligned}\mathbf{x}_1' &\sim N_{1,n}(\mu_1\mathbf{1}_n', \sigma_{11}, \mathrm{I}_n),\\ \mathbf{x}_{2|1}' &\sim N_{1,n}(\boldsymbol{\mu}_{2|1}', \sigma_{2|1}, \mathrm{I}_n),\\ \mathbf{x}_{3|1,2}' &\sim N_{1,n}(\boldsymbol{\mu}_{3|1,2}', \sigma_{3|1,2}, \mathrm{I}_n),\end{aligned}$$

where the conditional expectation and variance, given by (2.3) and (2.4), equals

$$\boldsymbol{\mu}_{3|1,2}' = \mu_3\mathbf{1}_n' + (0,\ \sigma_{32})\,\Sigma_{(2)}^{-1}\begin{pmatrix}\mathbf{x}_1' - \mu_1\mathbf{1}_n'\\ \mathbf{x}_2' - \mu_2\mathbf{1}_n'\end{pmatrix} = \beta_{30}\mathbf{1}_n' + \beta_{32}(\mathbf{x}_2' - \beta_{21}\mathbf{x}_1'),$$
$$\sigma_{3|1,2} = \sigma_{33} - (0,\ \sigma_{32})\ \Sigma_{(2)}^{-1}\begin{pmatrix}0\\ \sigma_{23}\end{pmatrix} = \frac{|\Sigma_{(3)}|}{|\Sigma_{(2)}|} = \sigma_{33} - \beta_{32}^2\ \sigma_{2|1},$$
$$\beta_{30} = \mu_3 - \beta_{32}\beta_{20}, \qquad \beta_{32} = \sigma_{32}\frac{\sigma_{11}}{|\Sigma_{(2)}|}$$

and $\boldsymbol{\mu}_{2|1}$, $\sigma_{2|1}$, β_{20} and β_{21} are given in (3.2)-(3.5).

The likelihood function based on all observations is given by (2.5) and thus proportional to

$$\sigma_{11}^{-n/2}\exp\Big(-\frac{1}{2}\sigma_{11}^{-1}(\mathbf{x}_1-\mu_1\mathbf{1}_n)'()\Big)$$

$$\times\,\sigma_{2|1}^{-n/2}\exp\Big(-\frac{1}{2}\sigma_{2|1}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_{2|1})'()\Big) \tag{3.9}$$

$$\times\,\sigma_{3|1,2}^{-n/2}\exp\Big(-\frac{1}{2}\sigma_{3|1,2}^{-1}(\mathbf{x}_3-\boldsymbol{\mu}_{3|1,2})'()\Big). \tag{3.10}$$

However, β_{21} appears in both (3.9) and (3.10) and therefore the strategy will be to estimate β_{21} via (3.9) and then insert this estimator in (3.10). It means that our estimation strategy is to maximize each part of the likelihood separately, similarly to a REML approach in mixed linear models.

Again, we can use the matrix form of the estimators of $\boldsymbol{\beta}_3=(\beta_{30},\beta_{32})'$, i.e.,

$$\hat{\boldsymbol{\beta}}_3=(\hat{\mathrm{X}}_2'\hat{\mathrm{X}}_2)^{-1}\hat{\mathrm{X}}_2'\mathbf{x}_3,$$

where

$$\hat{\mathrm{X}}_2=(\mathbf{1}_n:\hat{\mathbf{x}}_2)$$

and

$$\hat{\mathbf{x}}_2=\mathbf{x}_2-\hat{\beta}_{21}\mathbf{x}_1.$$

Here we have inserted the estimator $\hat{\beta}_{21}$ which is estimated in the second part of the likelihood function (3.9) given by (3.8). The estimator of the variance $\sigma_{3|1,2}$ equals

$$\hat{\sigma}_{3|1,2}=\frac{1}{n}(\mathbf{x}_3-\hat{\boldsymbol{\mu}}_{3|1,2})'()$$

and the regression coefficients are

$$\hat{\boldsymbol{\beta}}_3=\frac{1}{n\hat{\mathbf{x}}_2'\mathrm{C}\hat{\mathbf{x}}_2}\begin{pmatrix}\hat{\mathbf{x}}_2'(\hat{\mathbf{x}}_2\mathbf{1}_n'-\mathbf{1}_n\hat{\mathbf{x}}_2')\mathbf{x}_3\\ n\hat{\mathbf{x}}_2'\mathrm{C}\mathbf{x}_3\end{pmatrix}.$$

We can now estimate the initial parameters. The parameters $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_{11}$, $\hat{\sigma}_{22}$, $\hat{\sigma}_{12}$ and $\hat{\sigma}_{21}$ are estimated as in the bivariate case and

$$\begin{aligned}
\hat{\mu}_3 &= \hat{\beta}_{30} + \hat{\beta}_{32}\hat{\beta}_{20} = \frac{1}{n}\mathbf{x}_3'\mathbf{1}_n, \\
\hat{\sigma}_{33} &= \hat{\sigma}_{3|1,2} + \hat{\beta}_{32}^2\,\frac{|\hat{\Sigma}_{(2)}|}{\hat{\sigma}_{11}} = \hat{\sigma}_{3|1,2} + \hat{\beta}_{32}^2\;\hat{\sigma}_{2|1}, \\
\hat{\sigma}_{32} &= \hat{\beta}_{32}\,\frac{|\hat{\Sigma}_{(2)}|}{\hat{\sigma}_{11}} = \hat{\beta}_{32}\;\hat{\sigma}_{2|1}
\end{aligned}$$

where $\hat{\beta}_{20}$ is given by (3.7) and

$$\hat{\beta}_{32} = \frac{\hat{\mathbf{x}}_2'\mathrm{C}\mathbf{x}_3}{\hat{\mathbf{x}}_2'\mathrm{C}\hat{\mathbf{x}}_2}.$$

Hence, the proposed estimators

$$\hat{\sigma}_{33} = \frac{1}{n}\mathbf{x}_3'\mathrm{C}\mathbf{x}_3$$

and

$$\hat{\sigma}_{23} = \hat{\sigma}_{32} = \frac{1}{n}\hat{\mathbf{x}}_2'\mathrm{C}\mathbf{x}_3,$$

are obtained.

□

Although the estimators in Proposition 3.1 constitute of explicit expression they are still ad hoc estimators and therefore it is important to establish some basic properties.

Theorem 3.1 *The estimator* $\hat{\boldsymbol{\mu}}_3 = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)'$ *given in Proposition 3.1 is unbiased and consistent, and the estimator* $\hat{\Sigma}_{(3)}^{(1)} = (\hat{\sigma}_{ij})$ *is consistent.*

Proof The conditional expectation and variance of $\hat{\beta}_{32}$ are

$$\begin{aligned}
E(\hat{\beta}_{32}|\mathbf{x}_1, \mathbf{x}_2) &= \beta_{32}, \\
Var(\hat{\beta}_{32}|\mathbf{x}_1, \mathbf{x}_2) &= \frac{\sigma_{3|1,2}}{\hat{\mathbf{x}}_2'\mathrm{C}\hat{\mathbf{x}}_2} = \frac{\sigma_{3|1,2}}{(\mathbf{x}_2 - \hat{\beta}_{21}\mathbf{x}_1)'\mathrm{C}(\mathbf{x}_2 - \hat{\beta}_{21}\mathbf{x}_1)} \\
&= \frac{\sigma_{3|1,2}}{(\mathbf{x}_2 - \hat{\boldsymbol{\mu}}_{2|1})'()} = \frac{\sigma_{3|1,2}}{\mathbf{x}_2'\mathrm{C}(\mathrm{I}_n - \mathbf{x}_1(\mathbf{x}_1'\mathrm{C}\mathbf{x}_1)^{-1}\mathbf{x}_1')\mathrm{C}\mathbf{x}_2}
\end{aligned}$$

and the expectation and variance are

$$E(\hat{\beta}_{32}) = E(E(\hat{\beta}_{32}|\mathbf{x}_1, \mathbf{x}_2)) = \beta_{32}$$

and

$$\begin{aligned}
Var(\hat{\beta}_{32}) &= E(Var(\hat{\beta}_{32}|\mathbf{x}_1, \mathbf{x}_2)) + Var(E(\hat{\beta}_{32}|\mathbf{x}_1, \mathbf{x}_2)) \\
&= \sigma_{3|1,2} E\left(\frac{1}{\mathbf{x}_2' \mathrm{C}(\mathrm{I}_n - \mathbf{x}_1(\mathbf{x}_1'\mathrm{C}\mathbf{x}_1)^{-1}\mathbf{x}_1')\mathrm{C}\mathbf{x}_2}\right) \\
&= \sigma_{3|1,2} E\left(E\left(\frac{1}{\mathbf{x}_2' \mathrm{C}(\mathrm{I}_n - \mathbf{x}_1(\mathbf{x}_1'\mathrm{C}\mathbf{x}_1)^{-1}\mathbf{x}_1')\mathrm{C}\mathbf{x}_2}\Big|\mathbf{x_1}\right)\right) \\
&= \frac{\sigma_{3|1,2}}{\sigma_{2|1}} E\left(\frac{1}{\chi^2_{n-2}}\right) = \frac{\sigma_{3|1,2}}{\sigma_{2|1}(n-4)},
\end{aligned}$$

since the matrix $\mathrm{C}(\mathrm{I}_n - \mathbf{x}_1(\mathbf{x}_1'\mathrm{C}\mathbf{x}_1)^{-1}\mathbf{x}_1')\mathrm{C}$ is idempotent of rank $n-2$ and therefore

$$\frac{\mathbf{x}_2' \mathrm{C}(\mathrm{I}_n - \mathbf{x}_1(\mathbf{x}_1'\mathrm{C}\mathbf{x}_1)^{-1}\mathbf{x}_1')\mathrm{C}\mathbf{x}_2}{\sigma_{2|1}}\Big|\mathbf{x}_1 \sim \chi^2_{n-2}.$$

As before, the estimator $\hat{\beta}_{32}$, is unbiased and consistent. Furthermore, we have

$$\hat{\sigma}_{32} = \hat{\beta}_{32}\frac{|\hat{\Sigma}_{(2)}|}{\hat{\sigma}_{11}} \xrightarrow{p} \beta_{32}\frac{|\Sigma_{(2)}|}{\sigma_{11}} = \sigma_{32}, \quad \text{as } n \to \infty,$$

by Cramér-Slutsky's theorem (Cramér, 1946).

Since the estimator $\hat{\boldsymbol{\mu}}$ is mean based on independent and identically distributed observations the estimator is unbiased and consistent. The estimators $\hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{22}$ and $\hat{\sigma}_{33}$ are the maximum likelihood estimators for a non-structured covariance matrix and hence consistent. □

3.3 General p-variate case

For higher dimensions the estimators are found in the same way as for the three-dimensional case. Here follows the proposed estimators for the general dimension but when the covariance matrix is banded of order one.

Proposition 3.2 *Let* $\mathrm{X} \sim N_{p,n}\left(\boldsymbol{\mu}_p\mathbf{1}_n', \Sigma^{(1)}_{(p)}, \mathrm{I}_n\right)$. *Explicit estimators are given by*

$$\hat{\mu}_i = \frac{1}{n}\mathbf{x}_i'\mathbf{1}_n, \quad \hat{\sigma}_{ii} = \frac{1}{n}\mathbf{x}_i'\mathrm{C}\mathbf{x}_i, \quad \textit{for } i = 1, \ldots, p,$$

$$\hat{\sigma}_{i,i+1} = \frac{1}{n}\hat{\mathbf{x}}_i'\mathrm{C}\mathbf{x}_{i+1}, \quad \textit{for } i = 1, \ldots, p-1,$$

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

$$\hat{\beta}_{i,i-1} = \frac{\hat{\mathbf{x}}'_{i-1}\mathrm{C}\mathbf{x}_i}{\hat{\mathbf{x}}'_{i-1}\mathrm{C}\hat{\mathbf{x}}_{i-1}}.$$

and C *is given in (3.1).*

Here follows the motivation for Proposition 3.2. Since

$$\mathrm{X} \sim N_{p,n}\left(\boldsymbol{\mu}_p \mathbf{1}'_n, \Sigma^{(1)}_{(p)}, \mathrm{I}_n\right)$$

we have from equation (2.2) that

$$\begin{aligned}
\mathbf{x}'_1 &\sim N_{1,n}(\mu_1 \mathbf{1}'_n, \sigma_{11}, \mathrm{I}_n),\\
\mathbf{x}'_{2|1} &\sim N_{1,n}(\boldsymbol{\mu}'_{2|1}, \sigma_{2|1}, \mathrm{I}_n),\\
&\vdots\\
\mathbf{x}'_{p|1:p-1} &\sim N_{1,n}(\boldsymbol{\mu}'_{p|1:p-1}, \sigma_{p|1:p-1}, \mathrm{I}_n),
\end{aligned}$$

and the likelihood function follows from (2.5). Ignoring the normalizing constant the likelihood function is given by

$$\begin{aligned}
&\sigma_{11}^{-n/2}\exp\left(-\frac{1}{2}\sigma_{11}^{-1}(\mathbf{x}_1 - \mu_1\mathbf{1}_n)'()\right)\\
&\times \prod_{k=2}^{p} \sigma_{k|1:k-1}^{-n/2}\exp\left(-\frac{1}{2}\sigma_{k|1:k-1}^{-1}(\mathbf{x}_k - \boldsymbol{\mu}_{k|1:k-1})'()\right),
\end{aligned} \tag{3.11}$$

where, for $k = 2, \ldots p$,

$$\boldsymbol{\mu}'_{k|1:k-1} = \mu_k \mathbf{1}'_n + (0, \ldots,\ 0,\ \sigma_{k,k-1})\ \Sigma^{-1}_{(k-1)} \begin{pmatrix} \mathbf{x}'_1 - \mu_1 \mathbf{1}'_n \\ \vdots \\ \mathbf{x}'_{k-2} - \mu_{k-2}\mathbf{1}'_n \\ \mathbf{x}'_{k-1} - \mu_{k-1}\mathbf{1}'_n \end{pmatrix}.$$

Let

$$\tilde{\mathbf{x}}_k = \mathbf{x}_k - \beta_{k,k-1}\tilde{\mathbf{x}}_{k-1},$$

where

$$\beta_{k,k-1} = \sigma_{k-1,k}\frac{|\Sigma_{(k-2)}|}{|\Sigma_{(k-1)}|}$$

with

$$\tilde{\mathbf{x}}_1 = \mathbf{x}_1.$$

Then

$$\boldsymbol{\mu}_{k|1:k-1} = \beta_{k0}\mathbf{1}_n + \beta_{k,k-1}\tilde{\mathbf{x}}_{k-1}, \quad k \geq 2,$$

where

$$\beta_{k0} = \mu_k - \beta_{k,k-1}\beta_{k-1,0}$$

and

$$\beta_{10} = \mu_1.$$

The conditional variance equals

$$\begin{aligned}\sigma_{k|1:k-1} &= \sigma_{kk} - (0, ..., \ 0, \ \sigma_{k,k-1}) \ \Sigma_{(k-1)}^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sigma_{k-1,k} \end{pmatrix} \\ &= \sigma_{kk} - \frac{\sigma_{k,k-1}\sigma_{k-1,k}|\Sigma_{(k-2)}|}{|\Sigma_{(k-1)}|} = \frac{|\Sigma_{(k)}|}{|\Sigma_{(k-1)}|}.\end{aligned}$$

Each part of the likelihood can now be maximized with respect to the $3p-1$ parameters, $\mu_1, \ \beta_{p0}, \ \beta_{p-1,0}, \ \ldots, \ \beta_{20}, \ \beta_{p,p-1}, \ \beta_{p-1,p-2}, \ \ldots, \ \beta_{21}, \ \sigma_{11}, \ \sigma_{2|1}, \ldots, \sigma_{p|1:p-1}$. We maximize the p parts of the likelihood function (3.11) separately, and each time we insert the estimated parameters from previous steps. Thus, it is just a linear regression in each step and the proposed estimator of the regression coefficients are

$$\hat{\boldsymbol{\beta}}_k = \begin{pmatrix} \hat{\beta}_{k0} \\ \hat{\beta}_{k,k-1} \end{pmatrix} = (\hat{\mathrm{X}}'_{k-1}\hat{\mathrm{X}}_{k-1})^{-1}\hat{\mathrm{X}}'_{k-1}\mathbf{x}_k,$$

where

$$\begin{aligned}\hat{\mathrm{X}}_{k-1} &= (\mathbf{1}_n : \hat{\mathbf{x}}_{k-1}), \\ \hat{\mathbf{x}}_{k-1} &= \mathbf{x}_{k-1} - \hat{\beta}_{k-1,k-2}\hat{\mathbf{x}}_{k-2}\end{aligned}$$

with

$$\hat{\mathbf{x}}_1 = \mathbf{x}_1.$$

Hence, we have the estimators

$$\hat{\boldsymbol{\beta}}_k = \frac{1}{n\hat{\mathbf{x}}'_{k-1}\mathrm{C}\hat{\mathbf{x}}_{k-1}} \begin{pmatrix} \hat{\mathbf{x}}'_{k-1}(\hat{\mathbf{x}}_{k-1}\mathbf{1}'_n - \mathbf{1}_n\hat{\mathbf{x}}'_{k-1})\mathbf{x}_k \\ n\hat{\mathbf{x}}'_{k-1}\mathrm{C}\mathbf{x}_k \end{pmatrix}$$

and the estimator of the conditional variance is given by

$$\hat{\sigma}_{k|1:k-1} = \frac{1}{n}(\mathbf{x}_k - \hat{\boldsymbol{\mu}}_{k|1:k-1})'(),$$

where

$$\hat{\boldsymbol{\mu}}_{k|1:k-1} = \hat{\beta}_{k0}\mathbf{1}_n + \hat{\beta}_{k,k-1}\hat{\mathbf{x}}_{k-1}$$

for $k = 2, \ldots, p$. The estimators of the initial parameters, $\mu_1, \mu_2, \ldots, \mu_p$, $\sigma_{11}, \sigma_{12}, \sigma_{22}, \ldots, \sigma_{p-1,p}, \sigma_{pp}$, can be recovered from here as

$$\begin{aligned} \hat{\mu}_1 &= \frac{1}{n}\mathbf{x}'_1\mathbf{1}_n, \\ \hat{\sigma}_{11} &= \frac{1}{n}\mathbf{x}'_1\mathrm{C}\mathbf{x}_1 \end{aligned}$$

and for $k = 2, \ldots p$

$$\begin{aligned} \hat{\mu}_k &= \hat{\beta}_{k0} + \hat{\beta}_{k,k-1}\hat{\beta}_{k-1,0} = \frac{1}{n}\mathbf{x}'_k\mathbf{1}_n, \\ \hat{\sigma}_{kk} &= \hat{\sigma}_{k|1:k-1} + \hat{\beta}^2_{k,k-1}\,\frac{|\hat{\Sigma}_{(k-1)}|}{|\hat{\Sigma}_{(k-2)}|} \\ &= \hat{\sigma}_{k|1:k-1} + \hat{\beta}^2_{k,k-1}\;\hat{\sigma}_{k-1|1:k-2} = \frac{1}{n}\mathbf{x}'_k\mathrm{C}\mathbf{x}_k, \\ \hat{\sigma}_{k-1,k} &= \hat{\beta}_{k,k-1}\,\frac{|\hat{\Sigma}_{(k-1)}|}{|\hat{\Sigma}_{(k-2)}|} = \hat{\beta}_{k,k-1}\;\hat{\sigma}_{k-1|1:k-2} = \frac{1}{n}\hat{\mathbf{x}}'_{k-1}\mathrm{C}\mathbf{x}_k \\ &= \frac{1}{n}\mathbf{x}'_{k-1}\mathrm{C}\mathbf{x}_k - \hat{\beta}_{k-1,1}\frac{1}{n}(\hat{\mathbf{x}}_{k-2} - \hat{\beta}_{k-2,0}\mathbf{1}_n)'(\mathbf{x}_k - \hat{\mu}_k\mathbf{1}_n). \end{aligned}$$

□

Theorem 3.2 *The estimator* $\hat{\boldsymbol{\mu}}_p = (\hat{\mu}_1, \ldots, \hat{\mu}_p)'$ *given in Proposition 3.2 is unbiased and consistent, and the estimator* $\hat{\Sigma}^{(1)}_{(p)} = (\hat{\sigma}_{ij})$ *is consistent.*

Proof Again we can calculate the variance of β_{k1}:

$$Var(\hat{\beta}_{k,k-1}) = E(Var(\hat{\beta}_{k,k-1}|\mathrm{X}'_{1:k-1})) + Var(E(\hat{\beta}_{k,k-1}|\mathrm{X}'_{1:k-1}))$$
$$= \sigma_{k|1:k-1}E\left(\frac{1}{\hat{\mathbf{x}}'_{k-1}\mathrm{C}\hat{\mathbf{x}}_{k-1}}\right)$$
$$= \sigma_{k|1:k-1}E\left(\frac{1}{\mathbf{x}'_{k-1}\mathrm{C}(\mathrm{I}_n - \hat{\mathbf{x}}_{k-2}(\hat{\mathbf{x}}'_{k-2}\mathrm{C}\hat{\mathbf{x}}_{k-2})^{-1}\hat{\mathbf{x}}'_{k-2})\mathrm{C}\mathbf{x}_{k-1}}\right)$$
$$= \sigma_{k|1:k-1}E\left(E\left(\frac{1}{\mathbf{x}'_{k-1}\mathrm{C}(\mathrm{I}_n - \hat{\mathbf{x}}_{k-2}(\hat{\mathbf{x}}'_{k-2}\mathrm{C}\hat{\mathbf{x}}_{k-2})^{-1}\hat{\mathbf{x}}'_{k-2})\mathrm{C}\mathbf{x}_{k-1}}\Big|\mathrm{X}'_{1:k-2}\right)\right)$$
$$= \frac{\sigma_{k|1:k-1}}{\sigma_{k-1|1:k-2}}E\left(\frac{1}{\chi^2_{n-2}}\right) = \frac{\sigma_{k|1:k-1}}{\sigma_{k-1|1:k-2}(n-4)}$$

since

$$\frac{\mathbf{x}'_{k-1}\mathrm{C}(\mathrm{I}_n - \hat{\mathbf{x}}_{k-2}(\hat{\mathbf{x}}'_{k-2}\mathrm{C}\hat{\mathbf{x}}_{k-2})^{-1}\hat{\mathbf{x}}'_{k-2})\mathrm{C}\mathbf{x}_{k-1}}{\sigma_{k-1|1:k-2}}|\mathrm{X}'_{1:k-2} \sim \chi^2_{n-2}.$$

Thus, all estimated parameters, $\hat{\beta}_{k1}$, are unbiased and consistent and for $k = 2, \ldots, p$

$$\hat{\sigma}_{k,k-1} = \hat{\beta}_{k,k-1}\frac{|\hat{\Sigma}_{(k-1)}|}{|\hat{\Sigma}_{(k-2)}|} \xrightarrow{p} \beta_{k,k-1}\frac{|\Sigma_{(k-1)}|}{|\Sigma_{(k-2)}|} = \sigma_{k,k-1}, \quad \text{as } n \to \infty,$$

by Cramér-Slutsky's theorem.

Furthermore, the estimators $\hat{\mu}_i$, for $i = 1, \ldots, p$ are means based on independent and identically distributed observations, hence these are unbiased and consistent. The estimators $\hat{\sigma}_{ii}$, for $i = 1, \ldots, p$ are equal to the maximum likelihood estimators for a non-structured covariance matrix which are consistent. □

4 m-dependence of greater order than one ($m > 1$)

In Section 3 m-dependence of order one was investigated. Here in this section the same technique will be used to find estimators for the case when the order is greater than one, i.e., when $m > 1$. We start with a special case, when the order is two and the dimension is four, i.e., $m = 2$ and $p = 4$, just to see how things work out. At the end of this section we propose estimators for the general case when $m + 1 < p < n$. Furthermore, in line with previous section some properties of the estimators are established.

4.1 Special case; $m = 2$ and $p = 4$

Suppose that $\mathrm{X} \sim N_{4,n}(\boldsymbol{\mu}_4\mathbf{1}_n', \Sigma_{(4)}^{(2)}, \mathrm{I}_n)$, where we can write

$$\Sigma_{(4)}^{(2)} = \begin{pmatrix} \Sigma_{(3)} & \boldsymbol{\sigma}_{14} \\ \boldsymbol{\sigma}_{41}' & \sigma_{44} \end{pmatrix}.$$

If we use a chain rule factorization of the density we have

$$f(\mathrm{X}) = f(\mathbf{x}_4'|\mathrm{X}_{1:3})f(\mathrm{X}_{1:3}),$$

where

$$\mathbf{x}_4'|\mathrm{X}_{1:3} \sim N_{1,n}(\boldsymbol{\mu}_{4|1:3}', \sigma_{4|1:3}, \mathrm{I}_n)$$

and

$$\mathrm{X}_{1:3} \sim N_{3,n}(\boldsymbol{\mu}_3\mathbf{1}_n', \Sigma_{(3)}, \mathrm{I}_n), \tag{4.1}$$

where the conditional expectation, given by (2.3), equals

$$\begin{aligned}
\boldsymbol{\mu}_{4|1:3}' &= \mu_4\mathbf{1}_n' + \boldsymbol{\sigma}_{41}'\Sigma_{(3)}^{-1}\begin{pmatrix} \mathbf{x}_1' - \mu_1\mathbf{1}_n' \\ \mathbf{x}_2' - \mu_2\mathbf{1}_n' \\ \mathbf{x}_3' - \mu_3\mathbf{1}_n' \end{pmatrix} \\
&= \mu_4\mathbf{1}_n' + \begin{pmatrix} 0 & \sigma_{42} & \sigma_{43} \end{pmatrix}\begin{pmatrix} \sigma^{11} & \sigma^{12} & \sigma^{13} \\ \sigma^{21} & \sigma^{22} & \sigma^{23} \\ \sigma^{31} & \sigma^{32} & \sigma^{33} \end{pmatrix}\begin{pmatrix} \mathbf{x}_1' - \mu_1\mathbf{1}_n' \\ \mathbf{x}_2' - \mu_2\mathbf{1}_n' \\ \mathbf{x}_3' - \mu_3\mathbf{1}_n' \end{pmatrix} \\
&= \mu_4\mathbf{1}_n' + (\sigma_{42}\sigma^{21} + \sigma_{43}\sigma^{31})(\mathbf{x}_1' - \mu_1\mathbf{1}_n') \\
&\quad + (\sigma_{42}\sigma^{22} + \sigma_{43}\sigma^{32})(\mathbf{x}_2' - \mu_2\mathbf{1}_n') + (\sigma_{42}\sigma^{23} + \sigma_{43}\sigma^{33})(\mathbf{x}_3' - \mu_3\mathbf{1}_n') \\
&= \mu_4\mathbf{1}_n' + \sigma_{42}\big[\sigma^{21}(\mathbf{x}_1' - \mu_1\mathbf{1}_n') + \sigma^{22}(\mathbf{x}_2' - \mu_2\mathbf{1}_n') + \sigma^{23}(\mathbf{x}_3' - \mu_3\mathbf{1}_n')\big] \\
&\quad + \sigma_{43}\big[\sigma^{31}(\mathbf{x}_1' - \mu_1\mathbf{1}_n') + \sigma^{32}(\mathbf{x}_2' - \mu_2\mathbf{1}_n') + \sigma^{33}(\mathbf{x}_3' - \mu_3\mathbf{1}_n')\big].
\end{aligned}$$

Moreover,

$$\Sigma_{(3)}^{-1} = \big(\sigma^{ij}\big)_{ij} = \left((-1)^{i+j}\frac{|\mathrm{M}_{(3)}^{ji}|}{|\Sigma_{(3)}|}\right)_{ij}$$

and we can now write

$$
\begin{aligned}
\boldsymbol{\mu}'_{4|1:3} &= \beta_{40}\mathbf{1}'_n + \sigma_{42}\left(-\frac{|\mathrm{M}^{12}_{(3)}|}{|\Sigma_{(3)}|}\mathbf{x}'_1 + \frac{|\mathrm{M}^{22}_{(3)}|}{|\Sigma_{(3)}|}\mathbf{x}'_2 - \frac{|\mathrm{M}^{32}_{(3)}|}{|\Sigma_{(3)}|}\mathbf{x}'_3\right) \\
&\qquad + \sigma_{43}\left(\frac{|\mathrm{M}^{13}_{(3)}|}{|\Sigma_{(3)}|}\mathbf{x}'_1 - \frac{|\mathrm{M}^{23}_{(3)}|}{|\Sigma_{(3)}|}\mathbf{x}'_2 + \frac{|\mathrm{M}^{33}_{(3)}|}{|\Sigma_{(3)}|}\mathbf{x}'_3\right) \\
&= \beta_{40}\mathbf{1}'_n + \sigma_{42}\frac{|\Sigma_{(2)}|}{|\Sigma_{(3)}|}\left(-\frac{|\mathrm{M}^{12}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}'_1 + \frac{|\mathrm{M}^{22}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}'_2 - \frac{|\mathrm{M}^{32}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}'_3\right) \\
&\qquad + \sigma_{43}\frac{|\Sigma_{(2)}|}{|\Sigma_{(3)}|}\left(\frac{|\mathrm{M}^{13}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}'_1 - \frac{|\mathrm{M}^{23}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}'_2 + \mathbf{x}'_3\right)
\end{aligned}
$$

Hence,

$$\boldsymbol{\mu}_{4|1:3} = \tilde{\mathrm{X}}_3\boldsymbol{\beta}_4,$$

where the regression coefficients are

$$
\begin{aligned}
\boldsymbol{\beta}_4 &= (\beta_{40}, \beta_{42}, \beta_{43})', \\
\beta_{42} &= \sigma_{42}\frac{|\Sigma_{(2)}|}{|\Sigma_{(3)}|}, && (4.2) \\
\beta_{43} &= \sigma_{43}\frac{|\Sigma_{(2)}|}{|\Sigma_{(3)}|}, && (4.3) \\
\beta_{40} &= \mu_4 - \beta_{42}\left(-\frac{|\mathrm{M}^{12}_{(3)}|}{|\Sigma_{(2)}|}\mu_1 + \frac{|\mathrm{M}^{22}_{(3)}|}{|\Sigma_{(2)}|}\mu_2 - \frac{|\mathrm{M}^{32}_{(3)}|}{|\Sigma_{(2)}|}\mu_3\right) \\
&\qquad - \beta_{43}\left(\frac{|\mathrm{M}^{13}_{(3)}|}{|\Sigma_{(2)}|}\mu_1 - \frac{|\mathrm{M}^{23}_{(3)}|}{|\Sigma_{(2)}|}\mu_2 + \mu_3\right) && (4.4)
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{\mathrm{X}}_3 &= \left(\mathbf{1}_n : \tilde{\mathbf{x}}_{32} : \tilde{\mathbf{x}}_{33}\right), \\
\tilde{\mathbf{x}}_{32} &= -\frac{|\mathrm{M}^{12}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}_1 + \frac{|\mathrm{M}^{22}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}_2 - \frac{|\mathrm{M}^{32}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}_3, \\
\tilde{\mathbf{x}}_{33} &= \frac{|\mathrm{M}^{13}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}_1 - \frac{|\mathrm{M}^{23}_{(3)}|}{|\Sigma_{(2)}|}\mathbf{x}_2 + \mathbf{x}_3.
\end{aligned}
$$

The conditional variance is given by (2.4) and equals

$$\sigma_{4|1:3} = \sigma_{44} - \boldsymbol{\sigma}'_{41}\Sigma_{(3)}^{-1}\boldsymbol{\sigma}_{14} = \frac{|\Sigma_{(4)}|}{|\Sigma_{(3)}|}.$$

For estimating all parameters we start by estimating $\boldsymbol{\mu}_3$ and $\Sigma_{(3)}$ using the likelihood function for the model given in (4.1). The estimators are the usual maximum likelihood estimators since the covariance matrix is an ordinary non-structured covariance matrix. Furthermore, for estimating the rest of the parameters μ_4, σ_{42}, σ_{43} and σ_{44}, we use the estimators for the regression coefficients (4.2)-(4.4) with $\hat{\boldsymbol{\mu}}_3$ and $\hat{\Sigma}_{(3)}$ inserted, instead of $\boldsymbol{\mu}_3$ and $\Sigma_{(3)}$. The estimators are

$$\hat{\boldsymbol{\beta}}_4 = \begin{pmatrix} \hat{\beta}_{40} \\ \hat{\beta}_{42} \\ \hat{\beta}_{43} \end{pmatrix} = (\hat{\mathrm{X}}'_3\hat{\mathrm{X}}_3)^{-1}\hat{\mathrm{X}}'_3\mathbf{x}_4,$$

$$\hat{\boldsymbol{\mu}}_{4|1:3} = \hat{\mathrm{X}}_3\hat{\boldsymbol{\beta}}_4$$

and

$$\hat{\sigma}_{4|1:3} = \frac{1}{n}(\mathbf{x}_4 - \hat{\boldsymbol{\mu}}_{4|1:3})'() = \frac{1}{n}\mathbf{x}'_4\left(\mathrm{I}_n - \hat{\mathrm{X}}_3(\hat{\mathrm{X}}'_3\hat{\mathrm{X}}_3)^{-1}\hat{\mathrm{X}}'_3\right)\mathbf{x}_4,$$

where

$$\hat{\mathrm{X}}_3 = (\mathbf{1}_n : \hat{\mathbf{x}}_{32} : \hat{\mathbf{x}}_{33})$$

and

$$\hat{\mathbf{x}}_{32} = -\frac{|\hat{\mathrm{M}}_{(3)}^{12}|}{|\hat{\Sigma}_{(2)}|}\mathbf{x}_1 + \frac{|\hat{\mathrm{M}}_{(3)}^{22}|}{|\hat{\Sigma}_{(2)}|}\mathbf{x}_2 - \frac{|\hat{\mathrm{M}}_{(3)}^{32}|}{|\hat{\Sigma}_{(2)}|}\mathbf{x}_3,$$

$$\hat{\mathbf{x}}_{33} = \frac{|\hat{\mathrm{M}}_{(3)}^{13}|}{|\hat{\Sigma}_{(2)}|}\mathbf{x}_1 - \frac{|\hat{\mathrm{M}}_{(3)}^{23}|}{|\hat{\Sigma}_{(2)}|}\mathbf{x}_2 + \mathbf{x}_3,$$

where $\hat{\mathrm{M}}_{(k-1)}^{ji}$ is as $\mathrm{M}_{(k-1)}^{ji}$ but $\Sigma_{(k-1)}$ is replaced by $\hat{\Sigma}_{(k-1)}$.

Hence, the following estimators of the parameters are obtained

$$
\begin{aligned}
\hat{\sigma}_{42} &= \hat{\beta}_{42}\frac{|\hat{\Sigma}_{(3)}|}{|\hat{\Sigma}_{(2)}|},\\
\hat{\sigma}_{43} &= \hat{\beta}_{43}\frac{|\hat{\Sigma}_{(3)}|}{|\hat{\Sigma}_{(2)}|},\\
\hat{\sigma}_{44} &= \hat{\sigma}_{4|1:3} + \hat{\boldsymbol{\sigma}}'_{41}\hat{\Sigma}^{-1}_{(3)}\hat{\boldsymbol{\sigma}}_{14},\\
\hat{\mu}_4 &= \hat{\beta}_{40} + \hat{\beta}_{42}\left(-\frac{|\hat{\mathrm{M}}^{12}_{(3)}|}{|\hat{\Sigma}_{(2)}|}\hat{\mu}_1 + \frac{|\hat{\mathrm{M}}^{22}_{(3)}|}{|\hat{\Sigma}_{(2)}|}\hat{\mu}_2 - \frac{|\hat{\mathrm{M}}^{32}_{(3)}|}{|\hat{\Sigma}_{(2)}|}\hat{\mu}_3\right)\\
&\quad - \hat{\beta}_{43}\left(\frac{|\hat{\mathrm{M}}^{13}_{(3)}|}{|\hat{\Sigma}_{(2)}|}\hat{\mu}_1 - \frac{|\hat{\mathrm{M}}^{23}_{(3)}|}{|\hat{\Sigma}_{(2)}|}\hat{\mu}_2 + \hat{\mu}_3\right).
\end{aligned}
$$

In the next section we will propose estimators for arbitrary m and p. Unbiasedness and consistency of above estimators follow from the more general result given in that section.

4.2 Arbitrary order m and arbitrary dimension p

Using the same technique as in the previous section we can find estimators for arbitrary band covariance matrices. Here follows the main result of this report.

Proposition 4.1 *Let* $\mathrm{X} \sim N_{p,n}(\boldsymbol{\mu}_p\mathbf{1}'_n, \Sigma^{(m)}_{(p)}, \mathrm{I}_n)$*, with arbitrary integer* m *and* $\Sigma^{(m)}_{(p)}$ *defined in (2.1). The estimators of* $\boldsymbol{\mu}_p$ *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* $i = k-m, \ldots, k-1$*:*

$$\hat{\mu}_k = \frac{1}{n}\mathbf{x}'_k\mathbf{1}_n, \tag{4.5}$$

$$\hat{\sigma}_{ki} = \hat{\beta}_{ki}\frac{|\hat{\Sigma}_{(k-1)}|}{|\hat{\Sigma}_{(k-2)}|}, \tag{4.6}$$

$$\hat{\sigma}_{kk} = \frac{1}{n}\mathbf{x}'_k\left(\mathrm{I}_n - \hat{\mathrm{X}}_{k-1}(\hat{\mathrm{X}}'_{k-1}\hat{\mathrm{X}}_{k-1})^{-1}\hat{\mathrm{X}}'_{k-1}\right)\mathbf{x}_k + \hat{\boldsymbol{\sigma}}'_{k1}\hat{\Sigma}^{-1}_{(k-1)}\hat{\boldsymbol{\sigma}}_{1k}, \tag{4.7}$$

where

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

and

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

Motivation of Proposition 4.1. By conditioning we obtain the probability density

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

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

$$\Sigma_{(k)} = \begin{pmatrix} \Sigma_{(k-1)} & \boldsymbol{\sigma}_{1k} \\ \boldsymbol{\sigma}'_{k1} & \sigma_{kk} \end{pmatrix},$$

where

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

We have

$$\mathbf{x}'_k|\mathrm{X}_{1:k-1} \sim N_{1,n}(\boldsymbol{\mu}'_{k|1:k-1},\sigma_{k|1:k-1},\mathrm{I}_n),$$

where the conditional variance, given by (2.4), equals

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

and where the conditional expectation, given by (2.3), equals

$$\begin{aligned}
\boldsymbol{\mu}'_{k|1:k-1} &= \mu_k\mathbf{1}'_n + \boldsymbol{\sigma}'_{k1}\Sigma^{-1}_{(k-1)}\begin{pmatrix} \boldsymbol{x}'_1 - \mu_1\mathbf{1}'_n \\ \vdots \\ \boldsymbol{x}'_{k-1} - \mu_{k-1}\mathbf{1}'_n \end{pmatrix}\\
&= \beta_{k0}\mathbf{1}'_n + \sum_{i=k-m}^{k-1}\sigma_{ki}\sum_{j=1}^{k-1}\sigma^{ij}_{(k-1)}\boldsymbol{x}'_j. \qquad (4.8)
\end{aligned}$$

Here σ^{ij} are the elements of the inverse matrix

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

and the first regression coefficient equals

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

We may rewrite equation (4.8) as

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

where

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

and

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

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

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

where

$$\hat{\mathrm{X}}_{k-1} = (\mathbf{1}_n : \hat{\mathbf{x}}_{k-1,k-m} : \cdots : \hat{\mathbf{x}}_{k-1,k-1})$$

and

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

Here the estimators from the previous steps $(1, \ldots, k-1)$ 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

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

The estimators for the original parameters can be calculated as

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

and

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

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 $k = 1, \ldots, p$, we use induction.

Base step: For $k = 1, \ldots, m+1$, $\mathbf{x}_k$, i.e., $\hat{\mu}_k = \frac{1}{n}\mathbf{x}'_k\mathbf{1}_n$ since the estimators are the MLE for 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 we have

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

But $\hat{\mathrm{X}}_{k-1}(\hat{\mathrm{X}}_{k-1}'\hat{\mathrm{X}}_{k-1})^{-1}\hat{\mathrm{X}}_{k-1}'$ is the projection on a space which contains $\mathbf{1}_n$ and therefore

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

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

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

□

Since the estimators in Proposition 4.1 are ad hoc estimators it is important to establish some properties which are motivating them. We have the following theorem.

Theorem 4.1 *The estimator $\hat{\boldsymbol{\mu}}_p = (\hat{\mu}_1, \ldots, \hat{\mu}_p)'$ given in Proposition 4.1 is unbiased and consistent. Furthermore, the estimator $\hat{\Sigma}_{(p)}^{(m)} = (\hat{\sigma}_{ij})$ is consistent.*

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

Base step: The estimator $\hat{\Sigma}_{(m+1)}$ is consistent since it is the maximum likelihood estimator for 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

$$\hat{\boldsymbol{\beta}}_k = (\hat{\mathrm{X}}_{k-1}'\hat{\mathrm{X}}_{k-1})^{-1}\hat{\mathrm{X}}_{k-1}'\mathbf{x}_k = \left(\frac{1}{n}\hat{\mathrm{X}}_{k-1}'\hat{\mathrm{X}}_{k-1}\right)^{-1}\left(\frac{1}{n}\hat{\mathrm{X}}_{k-1}'\mathbf{x}_k\right), \tag{4.9}$$

where the first part in (4.9) converges in probability as follows. We have

$$\frac{1}{n}\hat{\mathrm{X}}'_{k-1}\hat{\mathrm{X}}_{k-1}$$

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

For $i, l = 1, .., m$

$$\frac{1}{n}\hat{\mathbf{x}}'_{k-1,k-i}\hat{\mathbf{x}}_{k-1,k-l} = |\hat{\Sigma}_{(k-2)}|^{-2}\sum_{j=1,q=1}^{k-1}(-1)^{j-i+q-l}|\hat{\mathrm{M}}^{j,k-i}_{(k-1)}||\hat{\mathrm{M}}^{q,k-l}_{(k-1)}|\frac{1}{n}\mathbf{x}'_j\mathbf{x}_q$$

$$\xrightarrow{p} |\Sigma_{(k-2)}|^{-2}\sum_{j=1,q=1}^{k-1}(-1)^{j-i+q-l}|\mathrm{M}^{j,k-i}_{(k-1)}||\mathrm{M}^{q,k-l}_{(k-1)}|(\sigma_{jq} + \mu_j\mu_q) \equiv w_{il}$$

and

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

$$\xrightarrow{p} |\Sigma_{(k-2)}|^{-1}\sum_{j=1}^{k-1}(-1)^{k-i+j}|\mathrm{M}^{j,k-i}_{(k-1)}|\mu_j \equiv w_i$$

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. Hence,

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

where

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

The second part in (4.9) converges also in probability. We have

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

where

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

Hence,

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

where

$$\mathbf{V} = \begin{pmatrix}\mu_k \\ v_m \\ \vdots \\ v_1\end{pmatrix}.$$

Now we show that $\hat{\boldsymbol{\beta}}_k \xrightarrow{p} \boldsymbol{\beta}_k$, as $n\to\infty$, where

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

by showing that $\mathrm{W}\boldsymbol{\beta}_k = \mathbf{V}$. Hence, we must show that

$$(1,w_m,...,w_1)\boldsymbol{\beta}_k = \mu_k \tag{4.10}$$

and

$$(w_r,w_{rm},...,w_{r1})\boldsymbol{\beta}_k = w_r\beta_{k0}+\sum_{i=1}^{m}w_{ri}\beta_{k,k-i} = v_r \tag{4.11}$$

for $r = 1, ..., m$.

First, consider equation (4.10)

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

Next, let us show equation (4.11) for $r = m$. The other cases are verified in the same way.

$$
\begin{aligned}
& w_m \beta_{k0} + \sum_{i=1}^{m} w_{mi} \beta_{k,k-i} = \\
&= \left(|\Sigma_{(k-2)}|^{-1} \sum_{j=1}^{k-1} (-1)^{k-m+j} |\mathrm{M}^{j,k-m}_{(k-1)}| \mu_j \right) \\
&\quad \times \left(\mu_k - |\Sigma_{(k-1)}|^{-1} \sum_{i=k-m}^{k-1} \sigma_{ki} \sum_{j=1}^{k-1} (-1)^{i+j} |\mathrm{M}^{ji}_{(k-1)}| \mu_j \right) \\
&+ \sum_{i=1}^{m} \left(|\Sigma_{(k-2)}|^{-1} |\Sigma_{(k-1)}|^{-1} \sigma_{k,k-i} \right. \\
&\quad \left. \times \sum_{j=1,q=1}^{k-1} (-1)^{j-m+q-i} |\mathrm{M}^{j,k-m}_{(k-1)}| |\mathrm{M}^{q,k-i}_{(k-1)}| (\sigma_{jq} + \mu_j \mu_q) \right) \\
&= |\Sigma_{(k-2)}|^{-1} \left\{ \sum_{j=1}^{k-1} (-1)^{k-m+j} |\mathrm{M}^{j,k-m}_{(k-1)}| \mu_j \mu_k \right. \\
&+ \Sigma_{(k-1)}|^{-1} \left[\sum_{i=1}^{m} \sigma_{k,k-i} \sum_{j=1,q=1}^{k-1} (-1)^{j-m+q-i} |\mathrm{M}^{j,k-m}_{(k-1)}| |\mathrm{M}^{q,k-i}_{(k-1)}| (\sigma_{jq} + \mu_j \mu_q) \right. \\
&\left. \left. - \left(\sum_{j=1}^{k-1} (-1)^{k-m+j} |\mathrm{M}^{j,k-m}_{(k-1)}| \mu_j \right) \left(\sum_{i=k-m}^{k-1} \sigma_{ki} \sum_{j=1}^{k-1} (-1)^{i+j} |\mathrm{M}^{ji}_{(k-1)}| \mu_j \right) \right] \right\}
\end{aligned}
$$

$$= |\Sigma_{(k-2)}|^{-1}\left\{ \sum_{j=1}^{k-1}(-1)^{k-m+j}|\mathrm{M}_{(k-1)}^{j,k-m}|\mu_j\mu_k \right.$$

$$+ |\Sigma_{(k-1)}|^{-1}\left(\sum_{i=1}^{m}\sigma_{k,k-i}\sum_{j=1,q=1}^{k-1}(-1)^{j-m+q-i}|\mathrm{M}_{(k-1)}^{j,k-m}||\mathrm{M}_{(k-1)}^{q,k-i}|(\sigma_{jq}+\mu_j\mu_q)\right.$$

$$\left.\left. - \sum_{i=1}^{m}\sigma_{k,k-i}\sum_{j=1,q=1}^{k-1}(-1)^{j-m+q-i}|\mathrm{M}_{(k-1)}^{j,k-i}||\mathrm{M}_{(k-1)}^{q,k-m}|\mu_j\mu_q \right)\right\}$$

$$= |\Sigma_{(k-2)}|^{-1}\left\{ \sum_{j=1}^{k-1}(-1)^{k-m+j}|\mathrm{M}_{(k-1)}^{j,k-m}|\mu_j\mu_k \right.$$

$$\left. + |\Sigma_{(k-1)}|^{-1}\sum_{i=1}^{m}\sigma_{k,k-i}\sum_{j=1}^{k-1}(-1)^{j+k-i}|\mathrm{M}_{(k-1)}^{j,k-m}|\underbrace{\sum_{q=1}^{k-1}(-1)^{k-i+q}|\mathrm{M}_{(k-1)}^{q,k-i}|\sigma_{jq}}_{=0,\text{ when } k-i\neq j}\right\}$$

$$= |\Sigma_{(k-2)}|^{-1}\left\{ \sum_{j=1}^{k-1}(-1)^{k-m+j}|\mathrm{M}_{(k-1)}^{j,k-m}|\mu_j\mu_k \right.$$

$$\left. + \sum_{i=1}^{m}(-1)^{m+i}\sigma_{k,k-i}|\Sigma_{(k-1)}|^{-1}|\mathrm{M}_{(k-1)}^{k-i,k-m}|\underbrace{\sum_{q=1}^{k-1}(-1)^{k-i+q}|\mathrm{M}_{(k-1)}^{q,k-i}|\sigma_{k-i,q}}_{=|\Sigma_{(k-1)}|}\right\}$$

$$= |\Sigma_{(k-2)}|^{-1}\left(\sum_{j=1}^{m}(-1)^{m+j}\sigma_{k,k-j}|\mathrm{M}_{(k-1)}^{k-j,k-m}| \right.$$

$$\left. + \sum_{j=1}^{k-1}(-1)^{k-m+j}|\mathrm{M}_{(k-1)}^{j,k-m}|\mu_j\mu_k \right) = v_m.$$

We have 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-Slutsky's theorem and since the estimators are assumed to be consistent for the $(k-1)$th step, we have

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

and

$$\hat{\sigma}_{kk} = \frac{1}{n}\mathbf{x}'_k \left(\mathrm{I}_n - \hat{\mathrm{X}}_{k-1}(\hat{\mathrm{X}}'_{k-1}\hat{\mathrm{X}}_{k-1})^{-1}\hat{\mathrm{X}}'_{k-1} \right) \mathbf{x}_k + \hat{\boldsymbol{\sigma}}'_{k1}\hat{\Sigma}^{-1}_{(k-1)}\hat{\boldsymbol{\sigma}}_{1k}$$
$$\xrightarrow{p} \sigma_{kk} + \mu_k^2 - \mathbf{V}'\boldsymbol{\beta} + \boldsymbol{\sigma}'_{k1}\Sigma^{-1}_{(k-1)}\boldsymbol{\sigma}_{1k}.$$

But

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

and hence

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

By induction, the estimator $\hat{\Sigma}^{(m)}_{(pp)} = (\hat{\sigma}_{ij})$ is consistent. □

5 Simulation

The examples presented here illustrate the results obtained in the previous sections. We will compare the explicit estimators derived in our study and the maximum likelihood estimators for the expectation and covariance matrix. It should be noted however that one cannot use ordinary MLE (see, for example, Muirhead (1982) and Srivastava (2002)) for estimation of the covariance matrix due to the fact that covariance matrix is structured (certain covariances are zero). Here the maximum likelihood estimators for the covariance matrix are obtained by maximizing the likelihood function.

In each simulation a sample with $n = 100$ observations was randomly generated from p-variate normal distributions $N_{p,n}$ using Release 14 of MATLAB Version 7.0.1 (The Mathworks Inc., Natick, MA, USA). Next, both the explicit and ML estimators were calculated in each simulation for the same set of observations. Simulations were repeated 100 (500) times, and the average values of estimators were calculated.

Four cases were studied, first three of them correspond to $m = 1$, and the forth one describes the case with $m = 2$. The results of the simulation study are presented in subsections below.

5.1 $p = 3, m = 1$

Here:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ 0 & \sigma_{32} & \sigma_{33} \end{pmatrix}.$$

We start with the following case:

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

After 100 simulations explicit estimators (EE) versus MLEs are given by:

$$\hat{\boldsymbol{\mu}}_{EE} = \begin{pmatrix} 0.9939 \\ 1.9956 \\ 3.0203 \end{pmatrix}, \qquad \hat{\Sigma}_{EE} = \begin{pmatrix} 1.9722 & 0.9999 & 0 \\ 0.9999 & 3.0371 & 1.9795 \\ 0 & 1.9795 & 3.9490 \end{pmatrix},$$

and

$$\hat{\boldsymbol{\mu}}_{MLE} = \begin{pmatrix} 0.9939 \\ 1.9956 \\ 3.0203 \end{pmatrix}, \qquad \hat{\Sigma}_{MLE} = \begin{pmatrix} 1.8942 & 1.0415 & 0 \\ 1.0415 & 7.0177 & 1.8341 \\ 0 & 1.8341 & 3.8635 \end{pmatrix}.$$

After 500 simulations they are:

$$\hat{\boldsymbol{\mu}}_{EE} = \begin{pmatrix} 0.9896 \\ 1.9996 \\ 2.9983 \end{pmatrix}, \qquad \hat{\Sigma}_{EE} = \begin{pmatrix} 1.9710 & 0.9996 & 0 \\ 0.9996 & 2.9834 & 1.9804 \\ 0 & 1.9804 & 3.9884 \end{pmatrix},$$

and

$$\hat{\boldsymbol{\mu}}_{MLE} = \begin{pmatrix} 0.9896 \\ 1.9996 \\ 2.9983 \end{pmatrix}, \qquad \hat{\Sigma}_{MLE} = \begin{pmatrix} 2.1213 & 1.0821 & 0 \\ 1.0821 & 6.5308 & 2.3738 \\ 0 & 2.3738 & 4.2793 \end{pmatrix}.$$

5.2 $p = 4, m = 1$

Here:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & 0 \\ 0 & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ 0 & 0 & \sigma_{43} & \sigma_{44} \end{pmatrix}.$$

We start with the following case:

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

After 100 simulations explicit estimators (EE) versus MLEs are given by:

$$\hat{\boldsymbol{\mu}}_{EE} = \begin{pmatrix} 0.9841 \\ 1.9833 \\ 2.9822 \\ 4.0106 \end{pmatrix}, \quad \hat{\Sigma}_{EE} = \begin{pmatrix} 2.0556 & 1.0065 & 0 & 0 \\ 1.0065 & 2.9586 & 2.0197 & 0 \\ 0 & 2.0197 & 4.0224 & 1.0174 \\ 0 & 0 & 1.0174 & 4.9714 \end{pmatrix},$$

and

$$\hat{\boldsymbol{\mu}}_{MLE} = \begin{pmatrix} 0.9841 \\ 1.9833 \\ 2.9822 \\ 4.0106 \end{pmatrix}, \quad \hat{\Sigma}_{MLE} = \begin{pmatrix} 1.9541 & 1.0109 & 0 & 0 \\ 1.0109 & 3.3732 & 2.2100 & 0 \\ 0 & 2.2100 & 7.2378 & 0.9673 \\ 0 & 0 & 0.9673 & 5.3763 \end{pmatrix}.$$

After 500 simulations they are:

$$\hat{\boldsymbol{\mu}}_{EE} = \begin{pmatrix} 0.9910 \\ 1.9886 \\ 2.9869 \\ 3.9884 \end{pmatrix}, \quad \hat{\Sigma}_{EE} = \begin{pmatrix} 1.9779 & 0.9872 & 0 & 0 \\ 0.9872 & 2.9513 & 1.9405 & 0 \\ 0 & 1.9405 & 3.9067 & 0.9518 \\ 0 & 0 & 0.9518 & 4.9052 \end{pmatrix},$$

and

$$\hat{\boldsymbol{\mu}}_{MLE} = \begin{pmatrix} 0.9910 \\ 1.9886 \\ 2.9869 \\ 3.9884 \end{pmatrix}, \quad \hat{\Sigma}_{MLE} = \begin{pmatrix} 2.0789 & 1.1632 & 0 & 0 \\ 1.1632 & 5.0500 & 2.4115 & 0 \\ 0 & 2.4115 & 6.1002 & 1.1951 \\ 0 & 0 & 1.1951 & 4.9881 \end{pmatrix}.$$

5.3 $p = 5, m = 1$

Here:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & 0 & 0 \\ 0 & \sigma_{32} & \sigma_{33} & \sigma_{34} & 0 \\ 0 & 0 & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ 0 & 0 & 0 & \sigma_{54} & \sigma_{55} \end{pmatrix}.$$

We start with the following case:

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

After 100 simulations explicit estimators (EE) versus MLEs are given by:

$$\hat{\boldsymbol{\mu}}_{EE} = \begin{pmatrix} 0.9892 \\ 1.9871 \\ 2.9989 \\ 4.0428 \\ 5.0148 \end{pmatrix}, \hat{\Sigma}_{EE} = \begin{pmatrix} 1.9833 & 0.9840 & 0 & 0 & 0 \\ 0.9840 & 2.9564 & 2.0027 & 0 & 0 \\ 0 & 2.0027 & 3.9817 & 0.9962 & 0 \\ 0 & 0 & 0.9962 & 4.9571 & 1.9668 \\ 0 & 0 & 0 & 1.9668 & 6.0371 \end{pmatrix},$$

and

$$\hat{\boldsymbol{\mu}}_{MLE} = \begin{pmatrix} 0.9892 \\ 1.9871 \\ 2.9989 \\ 4.0428 \\ 5.0148 \end{pmatrix}, \hat{\Sigma}_{MLE} = \begin{pmatrix} 2.0818 & 0.9666 & 0 & 0 & 0 \\ 0.9666 & 3.8183 & 2.5715 & 0 & 0 \\ 0 & 2.5715 & 6.4569 & 1.5107 & 0 \\ 0 & 0 & 1.5107 & 7.4205 & 1.7723 \\ 0 & 0 & 0 & 1.7723 & 5.9781 \end{pmatrix}.$$

After 500 simulations they are:

$$\hat{\boldsymbol{\mu}}_{EE} = \begin{pmatrix} 1.0066 \\ 2.0068 \\ 2.9995 \\ 4.0036 \\ 4.9993 \end{pmatrix}, \hat{\Sigma}_{EE} = \begin{pmatrix} 1.9648 & 0.9925 & 0 & 0 & 0 \\ 0.9925 & 2.9871 & 1.9655 & 0 & 0 \\ 0 & 1.9655 & 3.9889 & 0.9833 & 0 \\ 0 & 0 & 0.9833 & 4.9279 & 1.9691 \\ 0 & 0 & 0 & 1.9691 & 5.9461 \end{pmatrix},$$

and

$$\hat{\boldsymbol{\mu}}_{MLE} = \begin{pmatrix} 1.0066 \\ 2.0068 \\ 2.9995 \\ 4.0036 \\ 4.9993 \end{pmatrix}, \hat{\Sigma}_{MLE} = \begin{pmatrix} 2.1918 & 1.0661 & 0 & 0 & 0 \\ 1.0661 & 7.7813 & 2.7181 & 0 & 0 \\ 0 & 2.7181 & 6.3799 & 1.4754 & 0 \\ 0 & 0 & 1.4754 & 10.3522 & 2.3042 \\ 0 & 0 & 0 & 2.3042 & 6.2207 \end{pmatrix}.$$

5.4 $p = 4, m = 2$

Here:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ 0 & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix}.$$

We start with the following case:

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

After 100 simulations explicit estimators (*EE*) versus *MLE*s are given by:

$$\hat{\boldsymbol{\mu}}_{EE} = \begin{pmatrix} 1.0255 \\ 2.0238 \\ 3.0220 \\ 4.0134 \end{pmatrix}, \quad \hat{\Sigma}_{EE} = \begin{pmatrix} 2.0221 & 1.0143 & 1.0251 & 0 \\ 1.0143 & 3.0350 & 2.0227 & 1.0755 \\ 1.0251 & 2.0227 & 4.0505 & 1.0827 \\ 0 & 1.0755 & 1.0827 & 5.0561 \end{pmatrix},$$

and

$$\hat{\boldsymbol{\mu}}_{MLE} = \begin{pmatrix} 1.0255 \\ 2.0238 \\ 3.0220 \\ 4.0134 \end{pmatrix}, \quad \hat{\Sigma}_{MLE} = \begin{pmatrix} 2.1071 & 0.9380 & 0.8142 & 0 \\ 0.9380 & 4.6819 & 2.6835 & 1.6190 \\ 0.8142 & 2.6835 & 4.9478 & 0.8905 \\ 0 & 1.6190 & 0.8905 & 6.3692 \end{pmatrix}.$$

After 500 simulations they are:

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

and

$$\hat{\boldsymbol{\mu}}_{MLE} = \begin{pmatrix} 1.0013 \\ 2.0015 \\ 3.0005 \\ 4.0073 \end{pmatrix}, \quad \hat{\Sigma}_{MLE} = \begin{pmatrix} 2.2747 & 1.1883 & 1.3723 & 0 \\ 1.1883 & 8.0570 & 5.4203 & 1.3056 \\ 1.3723 & 5.4203 & 8.5184 & 0.7398 \\ 0 & 1.3056 & 0.7398 & 5.8338 \end{pmatrix}.$$

5.5 Discussion

Numerical examples presented above have shown that our explicit estimators for the covariance matrix describe the covariance structure better than the MLEs. Even the averages of 100 simulations resemble the initial covariance matrix. However, in general, results after 500 observations are a bit better.

6 Conclusion

In this report, we have proposed explicit estimators for the expectations and for a banded covariance matrix in a multivariate normal distribution. Since the covariance matrix is banded the covariances outside a diagonally bordered band are equal to zero. In many applications, e.g., in image analysis, computations are heavy and explicit expressions of estimators are more useful than iterative algorithms obtained for MLEs or restricted MLEs. Furthermore, our proposed explicit estimators are shown to be consistent and through a small simulation study they seem to be at least as good as the maximum likelihood estimators.

References

Andersson, S. A., Brøns, H. K., and Jensen, S. T. (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. and 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 University Press, Princeton.

Godolphin, E. J. and 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. R. (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 E. B. Andersen, H. Kiiveri, P. Laake, D. B. Cox and T. Schweder. With a reply by the author. *Scandinavian Journal of Statistics*, 8(2):65–92.

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

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

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

Marin, J. M. and 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. and 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.* John Wiley & Sons, New York, USA.

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

Naik, D. N. and Rao, S. 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 and R.P. Gupta, eds, 'Multivariate Statistical Inference', pp. 183-195, North-Holland, Amsterdam.

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

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.

Roy, A. and 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., and McCulloch, C. E. (1992). *Variance components.* John Wiley & Sons, Hoboken, NJ, USA.

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

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*, 17:257–28.

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