Estimation of banded covariance matrices in a multivariate normal distribution

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

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2008:2
ISSN 1651-8543

Estimation of banded covariance matrices in a multivariate normal distribution

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

Martin Ohlson
Linköping University, Linköping, Sweden

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

Abstract

The estimation of parameters of a multivariate p-dimensional random vector is considered for a banded covariance structure under the constrain that the covariances $\sigma_{ij} = 0$ for $|i - j| > 1$. Explicit analytical estimators for the mean and the covariance matrix are presented. The estimators are unbiased and consistent for the mean and consistent for the covariance matrix. Likelihood based tests which are asymptotically equivalent to likelihood ratio tests are presented and hypotheses for covariance matrices are tested.

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

AMS classification: 62H12, 62F12, 62F30

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

1 Introduction

The multivariate normal distribution plays an important role in multivariate statistical analysis. Most of the testing, estimation, and confidence interval procedures discussed by statistical researches are based on the assumption that the observation vectors are independent and normally distributed (Anderson, 1984; Srivastava, 2002; Muirhead, 2005). It is true that in practice the multivariate normal assumption does not always hold, but in many cases the normal model will still be very useful, even though the data are not normally distributed. Two main reasons for using multivariate normality are that it is often the case that multivariate observations are, at least approximately, normally distributed, and that the multivariate normal distribution is mathematically tractable.

Since normally distributed data can be modeled entirely in terms of their means and variances/covariances, these parameters actually specify the complete probability distribution of data. Estimating the mean and the covariance matrix is therefore a problem of great interest within the statistical science.

There is a lot of literature on estimating the mean and the covariance matrix in the multivariate normal distribution. The majority is based on the idea of maximizing the likelihood. The basic idea of using the likelihood function as the foundation for statistical inference is due to Fisher (1922), who also introduced maximum likelihood (ML) estimation.

However, estimation of a covariance matrix can be difficult, especially when the size of the covariance matrix, $p \times p$, is large. The two main difficulties are that the number of unknown elements in the covariance matrix increases quadratically with p, and that it is difficult to deal directly with individual elements of the covariance matrix because it is necessary to keep the estimated matrix positive definite. Unless the number of observations, n, is very large, estimation is often inefficient, and models with many parameters are, in general, difficult to interpret.

A number of approaches have been suggested for estimating a covariance matrix efficiently. The earliest works are probably by Wishart (1928), who studied the probability distribution of the maximum likelihood estimator (MLE) of the covariance matrix of a multivariate normal distribution, and Hotelling (1931), who presented a generalization of Student's t-statistic to multivariate hypothesis testing. James and Stein (1961) showed that the estimator of the mean of a multivariate normal distribution with the identity as covariance matrix is inadmissible, and presented the general problem of admissibility of estimators for problems with quadratic loss. Efron and Morris (1976) showed domination of the MLE for the mean of a multivariate normal distribution. Haff (1980) presented empirical Bayes estimation of the multivariate normal covariance matrix. Sinha and Ghosh (1987) studied the best equivariant estimators of the variance-covariance matrix under entropy loss. Pal (1993) dealt with the estimation of a normal precision matrix. Yang and Berger (1994) used a Bayesian approach based on a spectral decomposition of the covariance matrix. Leonard and Hsu (1992) and Chiu et al. (1996) modeled the matrix logarithm of the covariance matrix. Pourahmadi (1999, 2000) estimated the covariance matrix by parameterizing the Cholesky decomposition of its inverse. A more comprehensive review on approaches using an unconstrained parametrization and on general work in variance-covariance estimation is provided in Diggle et al. (2003). Another important aspect of statistical inference in the multivariate analysis is the testing hypotheses about the covariance matrix. In the literature, these tests have been studied extensively under the assumption of normality, see, for example, Anderson (1984); Srivastava (2002); Muirhead (2005). It has been found that the most robust test within the elliptical family is likelihood ratio test (LRT), that was first introduced by Wilks (1935) and Neyman and Pearson (1938).

In many practical situations there will be some inter-relationships among several variables that will impose a structure on the covariance matrix.

For example, patterned covariance matrices arise from a variety of contexts and were studied by a number of authors. Wilks (1946), in one of the early papers with patterned structure, 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) looked at the covariance matrix of multivariate complex normals, which arise in spectral analysis of multiple time series. Browne (1977) reviews patterned correlation matrices arising from multiple psychological measurements. Chinchilli and Carter (1984) considered a patterned covariance arising from a multivariate growth-curve model.

Covariance matrices with banded structure arise frequently in signal processing applications, including autoregressive or moving average image modeling, covariances of Gauss-Markov random processes (Woods, 1972; Moura and Balram, 1992), or with finite difference numerical approximations to partial differential equations. Banded matrices are also used to model the correlation of cyclostationary processes in periodic time series (Chakraborty, 1998). For autoregressive and moving average models that have been largely used for the analysis of time series, the autocovariance function is known to have a symmetric Toeplitz covariance structure (see for example, Anderson (1971);

Brockwell and Davis (1991); Brown and Prescott (1999); Fuller (1996); Marin and Dhorne (2002)).

One special class of the banded covariance matrices of a multivariate random vector is that under the constraint that some variables are conditionally independent given other remaining variables. For the multivariate normal distribution, this will correspond to some zeros among the entries of the inverse covariance matrix (Dempster, 1972; Whittaker, 1990). Bayesian model selection of detecting zeros in the inverse of covariance matrix can be found in Wong et al. (2003). To assume such a covariance matrix is in many cases a natural assumption and usually much more natural than to suppose independence between observations which is frequently applied.

If widely separated observations appear to be uncorrelated, it is reasonably to create a banded covariance structure by setting all covariances more than m steps apart equal to zero. We will call such a structure as *banded covariance structure of order* m.

It should be noted that such a situation differs from standard time series cases in a way that one allows unequal variances and correlation coefficients. That results in more parameters to be estimated than in traditional time series.

In univariate analysis when not assuming any particular distribution one often constructs 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 of the univariate ones. The basic maximum-likelihood problem for the multivariate case was first studied in detail by Dempster (1972), who used the name covariance selection. For the special case without conditional independence between variables the problem reduces to traditional maximum-likelihood estimation of the covariances for multivariate Gaussian random variables.

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. Moreover, m-dependence implies a specific covariance structure and REML as well as ML estimators are inconvenient to obtain since they imply an iterative treatment.

In this paper we focus on estimation of parameters of a multivariate random vector under the constraint that certain covariances are zero, namely we consider a banded covariance of order $m = 1$. The case with $m > 1$ will be analyzed in a forthcoming paper. 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 and to test some hypotheses for covariance matrices.

In this paper a simple estimation procedure is suggested which gives unbiased and consistent estimators of the mean and consistent estimators of the covariance matrix.

The paper is organized as follows. In Section 2, we present the main definitions and the notations used through the paper. Section 3 provides the algorithm for estimation the mean and the covariance matrix when $m = 1$. In the univariate ($p = 1$) and the bivariate ($p = 2$) cases the estimators obtained coincide with the usual ML estimators. The three-variate case, $p = 3$, is analyzed in details. Here a proposed algorithm consists of maximizing the likelihood function via inserting the estimated parameters from previous steps. The properties of estimators are presented as well. Furthermore, the general p-variate case is considered. In section 4 the likelihood ratio test is presented and likelihood based tests for banded covariance matrices are derived. Finally, Section 5 presents some simulation results and Section 6 summarizes the paper.

2 Basic notations

Unless otherwise stated, matrices will be denoted throughout the article by bold capital letters, vectors by bold font, scalars and matrix elements by ordinary letters.

Let a p-dimensional random vector $\mathbf{x}$ have a multivariate normal distribution with mean $\boldsymbol{\mu} : p \times 1$ and dispersion matrix $\boldsymbol{\Sigma}_p : p \times p$, $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_p)$, and suppose that we have n independent samples on $\mathbf{x}$. Then the observation matrix $\mathbf{X} : p \times n$,

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

belongs to the p-variate normal distribution, $\mathbf{X} \sim N_{p,n}(\mathbf{M}, \boldsymbol{\Sigma}_p, \mathbf{I}_n)$, with mean

$$\mathbf{M} = \boldsymbol{\mu}\mathbf{1}_n', \quad \boldsymbol{\mu} = (\mu_1,\ \mu_2,\ \dots,\ \mu_p)'. \tag{2}$$

The covariance (between rows) is $\boldsymbol{\Sigma}_p$, and $\mathbf{I}_n : n \times n$ indicates that columns are independent. Here, $\mathbf{1}_n : n \times 1$ is the unity vector and $\mathbf{I}_n$ the identity matrix, and we assume that the covariance matrix $\boldsymbol{\Sigma}_p$ is positive definite.

Let $\mathbf{X},\ \mathbf{M},\ \mathbf{\Sigma}_p$ be partitioned as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} : \begin{pmatrix} r \times n \\ (p-r) \times n \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{pmatrix} : \begin{pmatrix} r \times n \\ (p-r) \times n \end{pmatrix},$$

$$\mathbf{\Sigma}_p = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix} : \begin{pmatrix} r \times r & r \times (p-r) \\ (p-r) \times r & (p-r) \times (p-r) \end{pmatrix}. \tag{3}$$

Then we have the conditional distribution

$$\mathbf{X}_2 | \mathbf{X}_1 \sim N_{r,n} \left(\mathbf{M}_{2|1}, \mathbf{\Sigma}_{2|1}, \mathbf{I}_n \right), \tag{4}$$

where we adopt the notation of Srivastava and Khatri (1979, p.47) and Kollo and von Rosen (2005, p.195):

$$\mathbf{M}_{2|1} = \mathbf{M}_2 + \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \left(\mathbf{X}_1 - \mathbf{M}_1 \right), \ \ \mathbf{\Sigma}_{2|1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12}. \tag{5}$$

The joint probability density function of $\mathbf{X}$ is

$$f(\mathbf{X}) = (2\pi)^{-np/2} \left| \mathbf{\Sigma}_p \right|^{-n/2} \exp\Big(-\frac{1}{2} \left(\mathbf{X} - \mathbf{M} \right) \mathbf{\Sigma}_p^{-1} \left(\mathbf{X} - \mathbf{M} \right)' \Big), \tag{6}$$

and it can be written as the product of conditional and marginal distributions:

$$\begin{aligned} f(\mathbf{X}) &= f(\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_p) = f(\mathbf{x}_1)\ f(\mathbf{x}_2,\ \ldots,\ \mathbf{x}_p \mid \mathbf{x}_1) \\ &= f(\mathbf{x}_1)\ f(\mathbf{x}_2 \mid \mathbf{x}_1)\ \ldots\ f(\mathbf{x}_p \mid \mathbf{x}_1,\ \ldots,\ \mathbf{x}_{p-1}). \end{aligned} \tag{7}$$

For convenience, the following notation will be used

$$\mathbf{x}_2 \mid \mathbf{x}_1 = x_{2|1i}, \quad \ldots, \quad \mathbf{x}_p \mid \mathbf{x}_1, \ldots, \mathbf{x}_{p-1} = x_{p|1,\ \ldots,\ p-1;i} \tag{8}$$

The probability density function (6) considered as a function of the parameters $\mathbf{M}$ and $\mathbf{\Sigma}_p$ (for fixed observed $\mathbf{X}$) will serve as the likelihood function.

In this paper we consider the estimation of the mean and of the covariance matrix provided that a specific assumption holds, i.e., the covariance matrix $\mathbf{\Sigma}_p$ of the underlying normal distribution is patterned as follows

$$\mathbf{\Sigma}_p = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 & \ldots & 0 & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & 0 & \ldots & 0 & 0 \\ 0 & \sigma_{32} & \sigma_{33} & \sigma_{34} & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{p-2,p-3} & \sigma_{p-2,p-2} & \sigma_{p-2,p-1} & 0 \\ 0 & 0 & \ldots & 0 & \sigma_{p-1,p-2} & \sigma_{p-1,p-1} & \sigma_{p-1,p} \\ 0 & 0 & \ldots & 0 & 0 & \sigma_{p,p-1} & \sigma_{p,p} \end{pmatrix}. \tag{9}$$

The main aim is to find reasonable explicit estimators.

We denote by $\mathbf{\Sigma}^{-1}$ the inverse of $\mathbf{\Sigma}$, and by $\mathbf{\Sigma}_k$, $k < p$, the $k \times k$ submatrix of $\mathbf{\Sigma}_p$ located on the top left corner of $\mathbf{\Sigma}_p$. $|\mathbf{\Sigma}_k|$ stands for the determinant of corresponding matrix.

3 Estimation of parameters in a multivariate normal distribution

Statistical inference can take several forms. Point estimation is one of the most common forms, where one replaces the value of an unknown parameter θ with an appropriate function of the sample. If one replaces the sample values by the corresponding random variables, one gets an estimator. A good estimator must fulfill some criteria, such as unbiasedness, consistency etc. Another kind of statistical inference is to test whether the parameter θ equals a given value θ_0. One puts up a hypothesis $H_0 : \theta = \theta_0$ and based on some criteria one decides whether H_0 is true or not.

We focus now on the estimation of $\mathbf{M}$ and $\mathbf{\Sigma}$ under the assumptions given in (2) and (9).

3.1 Univariate case, $p = 1$

For the univariate case when $p = 1$

$$\boldsymbol{\mu} = \mu_1, \quad \mathbf{\Sigma}_1 = \sigma_{11}, \tag{10}$$

and

$$x_{1i} \sim N_1(\mu_1, \sigma_{11}). \tag{11}$$

The likelihood function based on all the observations can be written as

$$L_1 = \Big(2\pi\sigma_{11}\Big)^{-n/2} \exp\Big(- \frac{1}{2\sigma_{11}} \sum_{i=1}^{n} (x_{1i} - \mu_1)^2 \Big). \tag{12}$$

It can be maximized with respect to two parameters, μ_1 and σ_{11}, and the estimators of these parameters are given as follows:

$$\begin{aligned} \hat{\mu}_1 &= \frac{1}{n} \sum_{i=1}^{n} x_{1i}, \\ \hat{\sigma}_{11} &= \frac{1}{n} \sum_{i=1}^{n} (x_{1i} - \hat{\mu}_1)^2. \end{aligned} \tag{13}$$

It is well known that the estimator of the mean is unbiased and consistent; the estimator of the variance is not unbiased, but consistent.

3.2 Bivariate case, $p = 2$

When $p = 2$ one has again a standard (unstructured) case,

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}. \tag{14}$$

Therefore, the traditional ML approach should work, and there are five parameters to be estimated, μ_1, μ_2, σ_{11}, σ_{22}, σ_{12}.

Under the assumption of bivariate normality of (x_{1i}, x_{2i}), the distributions of interest are

$$\begin{aligned} x_{1i} &\sim N_1(\mu_1, \sigma_{11}), \\ x_{2|1i} &\sim N_1(\mu_{2|1i}, \sigma_{2|1}). \end{aligned} \tag{15}$$

The likelihood function based on all the observations is therefore given by

$$\begin{aligned} L_2 = & \Big(2\pi\sigma_{11}\Big)^{-n/2} \exp\Big(- \frac{1}{2\sigma_{11}} \sum_{i=1}^{n} (x_{1i} - \mu_1)^2 \Big) \\ & \times \Big(2\pi\sigma_{2|1}\Big)^{-n/2} \exp\Big(- \frac{1}{2\sigma_{2|1}} \sum_{i=1}^{n} (x_{2i} - \mu_{2|1i})^2 \Big). \end{aligned} \tag{16}$$

Here

$$\begin{aligned} \mu_{2|1i} &= \mu_2 + \sigma_{21}\, \sigma_{11}^{-1}\, (x_{1i} - \mu_1) = \beta_{10} + \beta_1 x_{1i}, \\ \sigma_{2|1} &= \sigma_{22} - \sigma_{21}\, \sigma_{11}^{-1}\, \sigma_{12} = \sigma_{22} - \beta_1^2\, \sigma_{11}, \end{aligned} \tag{17}$$

where

$$\beta_{10} = \mu_2 - \beta_1 \mu_1, \quad \beta_1 = \frac{\sigma_{21}}{\sigma_{11}}. \tag{18}$$

The likelihood can now be maximized with respect to five parameters, (μ_1, σ_{11}, β_{10}, β_1, $\sigma_{2|1}$), and each part (each line) of the likelihood in (16) can be maximized separately. Then the estimators of the initial parameters, (μ_1, μ_2, σ_{11}, σ_{22}, σ_{12}), can be carried out. One should note that the estimators of μ_1

and σ_{11} coincide with those given by (13), whereas the estimators of μ_2, σ_{22}, σ_{12} are given by

$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} x_{2i},$$

$$\hat{\sigma}_{22} = \hat{\sigma}_{2|1} + \hat{\beta}_1^2 \, \hat{\sigma}_{11} = \frac{1}{n} \sum_{i=1}^{n} (x_{2i} - \hat{\mu}_2)^2,$$

$$\hat{\sigma}_{12} = \hat{\beta}_1 \, \hat{\sigma}_{11} = \frac{1}{n} \sum_{i=1}^{n} (x_{2i} - \hat{\mu}_2)\, (x_{1i} - \hat{\mu}_1). \tag{19}$$

Here

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_{2i} - \hat{\mu}_2)\, (x_{1i} - \hat{\mu}_1)}{\sum_{i=1}^{n} (x_{1i} - \hat{\mu}_1)^2},$$

$$\hat{\sigma}_{2|1} = \frac{1}{n} \sum_{i=1}^{n} (x_{2i} - \hat{\mu}_{2|1i})^2 = \frac{1}{n} \sum_{i=1}^{n} \left((x_{2i} - \hat{\mu}_2)^2 - \hat{\beta}_1^2 \, (x_{1i} - \hat{\mu}_1)^2 \right),$$

$$\hat{\mu}_{2|1i} = \hat{\beta}_{10} + \hat{\beta}_1 x_{1i}, \quad \hat{\beta}_{10} = \hat{\mu}_2 - \hat{\beta}_1 \hat{\mu}_2. \tag{20}$$

The estimators are of course the usual MLEs. Thus, the estimator of the mean is unbiased and consistent and the estimator of the covariance matrix is not unbiased, but consistent.

3.3 Three-dimensional case, $p = 3$

The banded covariance matrix of order $m = 1$ in the case $p = 3$ provides us with a situation different from the one considered in Subsections 3.1 and 3.2, because now there are zero elements in the covariance matrix,

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \quad \boldsymbol{\Sigma}_3 = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ 0 & \sigma_{32} & \sigma_{33} \end{pmatrix}. \tag{21}$$

There are eight parameters to be estimated, (μ_1, μ_2, μ_3, σ_{11}, σ_{22}, σ_{12}, σ_{33}, σ_{23}), and under the assumption of normality of (x_{1i}, x_{2i}, x_{3i}) we observe

$$x_{1i} \sim N_1(\mu_1, \sigma_{11}),$$

$$x_{2|1i} \sim N_1(\mu_{2|1i}, \sigma_{2|1}),$$

$$x_{3|1,2i} \sim N_1(\mu_{3|1,2i}, \sigma_{3|1,2}). \tag{22}$$

The likelihood function based on all the observations is now given by

$$\begin{aligned} L_3 &= \Big(2\pi\sigma_{11}\Big)^{-n/2} \exp\Big(-\frac{1}{2\sigma_{11}}\sum_{i=1}^{n}(x_{1i}-\mu_1)^2\Big) \\ &\times \Big(2\pi\sigma_{2|1}\Big)^{-n/2} \exp\Big(-\frac{1}{2\sigma_{2|1}}\sum_{i=1}^{n}(x_{2i}-\mu_{2|1i})^2\Big) \\ &\times \Big(2\pi\sigma_{3|1,2}\Big)^{-n/2} \exp\Big(-\frac{1}{2\sigma_{3|1,2}}\sum_{i=1}^{n}(x_{3i}-\mu_{3|1,2i})^2\Big). \end{aligned} \tag{23}$$

Here $\mu_{2|1i}$ and $\sigma_{2|1}$ are the same as in the previous subsection, see (17), and

$$\begin{aligned} \mu_{3|1,2i} &= \mu_3 + (0 \;\; \sigma_{32}) \; \mathbf{\Sigma}_2^{-1} \begin{pmatrix} x_{1i}-\mu_1 \\ x_{2i}-\mu_2 \end{pmatrix}, \\ \sigma_{3|1,2} &= \sigma_{33} - (0 \;\; \sigma_{32}) \; \mathbf{\Sigma}_2^{-1} \begin{pmatrix} 0 \\ \sigma_{23} \end{pmatrix}. \end{aligned} \tag{24}$$

It is easy to see that

$$\begin{aligned} \mu_{3|1,2i} &= \mu_3 + \beta_2(x_{2i}-\mu_{2|1i}) = \mu_3 + \beta_2(x_{2i}-\mu_2-\beta_1(x_{1i}-\mu_1)) \\ &= \beta_{20} + \beta_2(x_{2i}-\beta_1 x_{1i}), \\ \sigma_{3|1,2} &= \sigma_{33} - \beta_2^2 \; \sigma_{2|1} = \frac{|\mathbf{\Sigma}_3|}{|\mathbf{\Sigma}_2|}, \end{aligned} \tag{25}$$

where

$$\beta_{20} = \mu_3 - \beta_2(\mu_2 - \beta_1\mu_1), \quad \beta_2 = \sigma_{32}\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_2|}. \tag{26}$$

Now β_1 appears in both the expression for $\mu_{2|1i}$, and for $\mu_{3|1,2i}$, which causes some problems. Therefore the likelihood cannot be maximized directly as we did before, i.e. by maximizing each part of the likelihood separately. We propose a sequential approach: the likelihood is factored according to (7) and (23). The estimation starts with the first factor. Parameters are estimated and those parameters which also appear in the second factor are replaced by the estimators from the previous factor. The estimation proceeds in a same manner until the parameters of the last factor have been obtained.

Thus, the likelihood can be maximized with respect to eight parameters, $(\mu_1,\ \sigma_{11},\ \beta_{10},\ \beta_1,\ \sigma_{2|1},\ \beta_{20},\ \beta_2,\ \sigma_{3|1,2})$, and the estimators of the initial parameters can be retrieved from there. Again, the estimators of μ_1 and σ_{11} are given by (13), the estimators of μ_2, σ_{22} and σ_{12} are given by (19), and the estimators of μ_3, σ_{33} and σ_{23} equal

$$\begin{aligned}
\hat{\mu}_3 &= \frac{1}{n} \sum_{i=1}^{n} x_{3i}, \\
\hat{\sigma}_{33} &= \hat{\sigma}_{3|1,2} + \hat{\beta}_2^2 \, \hat{\sigma}_{2|1} = \frac{1}{n} \sum_{i=1}^{n} (x_{3i} - \hat{\mu}_3)^2, \\
\hat{\sigma}_{23} &= \hat{\beta}_2 \, \hat{\sigma}_{2|1} = \frac{1}{n} \sum_{i=1}^{n} (x_{3i} - \hat{\mu}_3)\big(x_{2i} - \hat{\mu}_2 - \hat{\beta}_1 (x_{1i} - \hat{\mu}_1)\big),
\end{aligned} \tag{27}$$

where $\hat{\beta}_1$ and $\hat{\sigma}_{2|1}$ are derived in Section 3.2, and

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\sum_{i=1}^{n} (x_{3i} - \hat{\mu}_3) \left(x_{2i} - \hat{\mu}_2 - \hat{\beta}_1 \, (x_{1i} - \hat{\mu}_1)\right)}{\sum_{i=1}^{n} \left(x_{2i} - \hat{\mu}_2 - \hat{\beta}_1 \, (x_{1i} - \hat{\mu}_1)\right)^2}, \\
\hat{\sigma}_{3|1,2} &= \frac{1}{n} \sum_{i=1}^{n} (x_{3i} - \hat{\mu}_{3|1,2i})^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \Big((x_{3i} - \hat{\mu}_3)^2 - \hat{\beta}_2^2 \, \big((x_{2i} - \hat{\mu}_2)^2 - \hat{\beta}_1^2 \, (x_{1i} - \hat{\mu}_1)^2 \big) \Big), \\
\hat{\mu}_{3|1,2i} &= \hat{\beta}_{20} + \hat{\beta}_2 (x_{2i} - \hat{\beta}_1 x_{1i}), \quad \hat{\beta}_{20} = \hat{\mu}_3 - \hat{\beta}_2 (\hat{\mu}_2 - \hat{\beta}_1 \hat{\mu}_1).
\end{aligned} \tag{28}$$

It is seen that the estimators of the mean coincide with the standard MLE, and are therefore unbiased and consistent. The estimators of all elements of the covariance matrix, besides $\hat{\sigma}_{23}$, coincide with the standard MLE and are therefore consistent but not unbiased. To study the properties of $\hat{\sigma}_{23}$, we first look on the conditional expectation and conditional variance of $\hat{\beta}_2$. Taking into account (22) and (25) one has

$$E\big(\hat{\beta}_2 | x_{1i}, x_{2i}\big) = E \left(\frac{\sum_{i=1}^{n} (x_{3i} - \hat{\mu}_3) \left(x_{2i} - \hat{\mu}_2 - \hat{\beta}_1 \, (x_{1i} - \hat{\mu}_1)\right)}{\sum_{i=1}^{n} \left(x_{2i} - \hat{\mu}_2 - \hat{\beta}_1 \, (x_{1i} - \hat{\mu}_1)\right)^2} \mid x_{1i}, x_{2i} \right)$$

$$= \frac{\sum_{i=1}^{n} \beta_2 \big(x_{2i} - \hat{\mu}_2 - \hat{\beta}_1 \, (x_{1i} - \hat{\mu}_1)\big)^2}{\sum_{i=1}^{n} \left(x_{2i} - \hat{\mu}_2 - \hat{\beta}_1 \, (x_{1i} - \hat{\mu}_1)\right)^2} = \beta_2,$$

$$
\begin{aligned}
Var\big(\hat\beta_2|x_{1i},x_{2i}\big) &= Var\left(\frac{\sum_{i=1}^n (x_{3i}-\hat\mu_3)\left(x_{2i}-\hat\mu_2-\hat\beta_1\,(x_{1i}-\hat\mu_1)\right)}{\sum_{i=1}^n\,\left(x_{2i}-\hat\mu_2-\hat\beta_1\,(x_{1i}-\hat\mu_1)\right)^2}\mid x_{1i},x_{2i}\right)\\
&= \frac{\sum_{i=1}^n \left(x_{2i}-\hat\mu_2-\hat\beta_1\,(x_{1i}-\hat\mu_1)\right)^2 Var\big((x_{3i}-\hat\mu_3)|x_{1i},x_{2i}\big)}{\left(\sum_{i=1}^n\,\left(x_{2i}-\hat\mu_2-\hat\beta_1\,(x_{1i}-\hat\mu_1)\right)^2\right)^2}\\
&= \frac{\sigma_{3|1,2}}{\sum_{i=1}^n\,\left(x_{2i}-\hat\mu_2-\hat\beta_1\,(x_{1i}-\hat\mu_1)\right)^2}. \qquad (29)
\end{aligned}
$$

Finally,

$$
E\big(\hat\beta_2\big)=E\Big(E\big(\hat\beta_2\mid x_{1i},x_{2i}\big)\Big)=\beta_2,
$$

$$
\begin{aligned}
Var\big(\hat\beta_2\big) &= E\Big(Var\big(\hat\beta_2\mid x_{1i},x_{2i}\big)\Big)+Var\Big(E\big(\hat\beta_2\mid x_{1i},x_{2i}\big)\Big)\\
&= \sigma_{3|1,2}\,E\left(\frac{1}{\sum_{i=1}^n\,\left(x_{2i}-\hat\mu_2-\hat\beta_1\,(x_{1i}-\hat\mu_1)\right)^2}\right)\\
&= \sigma_{3|1,2}\,E\left(E\left(\frac{1}{\sum_{i=1}^n\,\left(x_{2i}-\hat\mu_2-\hat\beta_1\,(x_{1i}-\hat\mu_1)\right)^2}\mid x_{1i}\right)\right)\\
&= \frac{\sigma_{3|1,2}}{\sigma_{2|1}}\frac{1}{n-4}. \qquad (30)
\end{aligned}
$$

The last equality here follows from the fact that for

$$
Y\equiv\frac{\sum_{i=1}^n\,\left(x_{2i}-\hat\mu_2-\hat\beta_1\,(x_{1i}-\hat\mu_1)\right)^2}{\sigma_{2|1}}\mid x_{1i}\sim\chi^2_{n-2},\;\; E(Y^{-1})=\frac{1}{n-4}. \qquad (31)
$$

According to (30) $\hat\beta_2$ is an unbiased and consistent estimator of β_2.

3.4 General p-variate case

In this section we extend the results of the previous section to a general p.

It is observed that p-variate normality implies

$$
\begin{aligned}
x_{1i} &\sim N_1(\mu_1,\sigma_{11}),\\
x_{2|1i} &\sim N_1(\mu_{2|1i},\sigma_{2|1}),\\
&\vdots\\
x_{p|1,2,\ldots,p-1;i} &\sim N_1(\mu_{p|1,2,\ldots,p-1;i},\sigma_{p|1,2,\ldots,p-1}), \qquad (32)
\end{aligned}
$$

and the likelihood equals

$$L_p = \Big(2\pi\sigma_{11}\Big)^{-n/2}\exp\Big(-\frac{1}{2\sigma_{11}}\sum_{i=1}^{n}(x_{1i}-\mu_1)^2\Big)$$

$$\times\prod_{k=2}^{p}\Big(2\pi\sigma_{k|1,\dots k-1}\Big)^{-n/2}\exp\Big(-\frac{1}{2\sigma_{k|1,\dots k-1}}\sum_{i=1}^{n}(x_{ki}-\mu_{k|1,\dots k-1})^2\Big), \quad (33)$$

where

$$\mu_{p|1,2,\dots,p-1;i} = \mu_p + (0\ \ \dots\ \ 0\ \ \sigma_{p,p-1})\ \mathbf{\Sigma}_{p-1}^{-1}\begin{pmatrix} x_{1i}-\mu_1 \\ \vdots \\ x_{p-2;i}-\mu_{p-2} \\ x_{p-1;i}-\mu_{p-1}\end{pmatrix},$$

$$\sigma_{p|1,2,\dots,p-1} = \sigma_{p,p} - (0\ \ \dots\ \ 0\ \ \sigma_{p,p-1})\ \mathbf{\Sigma}_{p-1}^{-1}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sigma_{p-1,p}\end{pmatrix}. \quad (34)$$

Again, it is easy to see that

$$\begin{aligned}\mu_{p|1,2,\dots,p-1;i} &= \mu_p + \beta_{p-1}(x_{p-1;i}-\mu_{p-1|1,\dots,p-2;i}) \\ &= \dots = \beta_{p-1,0} + \beta_{p-1}(x_{p-1;i}-\beta_{p-2}(x_{p-2;i}-\beta_{p-3}(\dots))),\end{aligned}$$

$$\sigma_{p|1,2,\dots,p-1} = \sigma_{p,p} - \beta_{p-1}^2\ \sigma_{p-1|1,2,\dots,p-2} = \frac{|\mathbf{\Sigma}_p|}{|\mathbf{\Sigma}_{p-1}|}, \quad (35)$$

where

$$\beta_{p-1,0} = \mu_p - \beta_{p-1}(\mu_{p-1}-\beta_{p-2}(\mu_{p-2}-\beta_{p-3}(\dots))),\ \ \beta_{p-1} = \sigma_{p,p-1}\frac{|\mathbf{\Sigma}_{p-2}|}{|\mathbf{\Sigma}_{p-1}|}. \quad (36)$$

The likelihood can now be maximized with respect to $3p-1$ parameters, $(\mu_1,\ \sigma_{11};\ \beta_{10},\ \beta_1,\ \sigma_{2|1},\ \dots,\ \beta_{p-1,0},\ \beta_{p-1},\ \sigma_{p|1,\dots,p-1})$, and the estimators of initial parameters, can be carried out from there. In general, the estimators of parameters for $p \geq 2$ can be written as

$$
\begin{aligned}
\hat{\mu}_p &= \frac{1}{n}\sum_{i=1}^{n} x_{pi}, \\
\hat{\sigma}_{p,p} &= \hat{\sigma}_{p|1,\dots,p-1} + \hat{\beta}_{p-1}^2\ \hat{\sigma}_{p-1|1,\dots,p-2} = \frac{1}{n}\sum_{i=1}^{n}(x_{pi}-\hat{\mu}_p)^2, \\
\hat{\sigma}_{p-1,p} &= \hat{\beta}_{p-1}\ \hat{\sigma}_{p-1|1,\dots,p-2} \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_{p,i}-\hat{\mu}_p)\big(x_{p-1,i}-\hat{\mu}_{p-1}-\hat{\beta}_{p-2}\big(x_{p-2,i}-\hat{\mu}_{p-2}-\hat{\beta}_{p-3}(\dots)\big)\big),
\end{aligned}
\tag{37}
$$

with the estimators of μ_1 and σ_{11} being given by (13). Here

$$
\begin{aligned}
\hat{\beta}_{p-1} &= \frac{\sum_{i=1}^{n}(x_{p,i}-\hat{\mu}_p)\Big(x_{p-1,i}-\hat{\mu}_{p-1}-\hat{\beta}_{p-2}\left(x_{p-2,i}-\hat{\mu}_{p-2}-\hat{\beta}_{p-3}(\dots)\right)\Big)}{\sum_{i=1}^{n}\Big(x_{p-1,i}-\hat{\mu}_{p-1}-\hat{\beta}_{p-2}\left(x_{p-2,i}-\hat{\mu}_{p-2}-\hat{\beta}_{p-3}(\dots)\right)\Big)^2}, \\
\hat{\sigma}_{p|1,\dots,p-1} &= \frac{1}{n}\sum_{i=1}^{n}(x_{pi}-\hat{\mu}_{p|1,\dots,p-1;i})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big((x_{p,i}-\hat{\mu}_p)^2-\hat{\beta}_{p-1}^2\big((x_{p-1,i}-\hat{\mu}_{p-1})^2-\hat{\beta}_{p-2}^2(\dots)^2\big)\Big) = \frac{|\hat{\boldsymbol{\Sigma}}_{p-1}|}{|\hat{\boldsymbol{\Sigma}}_p|}, \\
\hat{\mu}_{p|1,\dots,p-1;i} &= \hat{\beta}_{p-1,0}+\hat{\beta}_{p-1}\big(x_{p-1;i}-\hat{\beta}_{p-2}(x_{p-2;i}-\hat{\beta}_{p-3}(\dots))\big), \\
\hat{\beta}_{p-1,0} &= \hat{\mu}_p-\hat{\beta}_{p-1}\big(\hat{\mu}_{p-1}-\hat{\beta}_{p-2}(\hat{\mu}_{p-2}-\hat{\beta}_{p-3}(\dots))\big).
\end{aligned}
\tag{38}
$$

The unbiasedness and the consistency of these estimators will be analyzed in details in a forthcoming publication. It should be noted however that the estimators of the mean coincide with the standard MLEs, and are therefore unbiased and consistent. The estimators of the diagonal elements of the banded covariance matrix coincide with the standard MLEs for the unstructured covariance matrix and are therefore consistent but not unbiased.

4 The likelihood ratio test

Making inference about hypothesis often relies on the theory of the likelihood ratio statistic. A likelihood-ratio test, denoted as Λ, is a statistical test in which a ratio is computed between the maximum likelihood of a result under two different hypotheses. The numerator corresponds to the maximum likelihood of an observed result under the null hypothesis, H_0, the denominator corresponds to the maximum likelihood of an observed result under the alternative hypothesis, H_1:

$$\Lambda = \frac{\sup_{H_0} L(H_0)}{\sup_{H_1} L(H_1)}. \tag{39}$$

The likelihood ratio is between 0 and 1. Lower values of the likelihood ratio mean that the observed result was less likely to occur under the null hypothesis. Higher values mean that the observed result was more likely to occur under the null hypothesis.

In most cases the exact distribution of the likelihood ratio corresponding to specific hypotheses is very difficult to determine. A convenient result tells us that under certain regularity conditions the distribution of $-2\ln\Lambda$ will tend to be a χ^2 distribution for large sized samples. The likelihood-ratio test rejects the null hypothesis if the value of this statistic is too small.

In many problems, it is desired to test the hypothesis $H_0 : \theta \in \Omega_0$ against $H_1 : \theta \in \Omega$, where Ω is the k-dimensional parameter space and Ω_0 is an r-dimensional ($r < k$) subset of Ω. Wilks (1938) proved that in such a case, when the null hypothesis is nested within the alternative hypothesis the distribution of the statistic $-2\ln\Lambda$ is asymptotically χ^2 with $k - r$ degrees of freedom.

4.1 $H_0 : \mathbf{\Sigma}_{kj} = 0$ for all $k, j = 1, ..., q;\ k \neq j$; unstructured covariance matrix

Let $\mathbf{X} : p \times n$ and $\mathbf{M} : p \times n$ be given by (1)-(2) and $\mathbf{\Sigma}$ be a positive definite unstructured covariance matrix.

Suppose that $\mathbf{X}$, $\mathbf{M}$, $\mathbf{\Sigma}$ can be partitioned as (see, for example, Muirhead (2005))

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_q \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_q \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_q \end{pmatrix},$$

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \dots & \mathbf{\Sigma}_{1q} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \dots & \mathbf{\Sigma}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{q1} & \mathbf{\Sigma}_{q2} & \dots & \mathbf{\Sigma}_{qq} \end{pmatrix}, \tag{40}$$

where

$$\mathbf{X}_k = (\mathbf{X}_{k1} : \mathbf{X}_{k2} : \ldots : \mathbf{X}_{kn}) = (\mathbf{X}_{ki}) : p_k \times n, \quad \mathbf{M}_k = \boldsymbol{\mu}_k \mathbf{1}_n' : p_k \times n,$$
$$\boldsymbol{\Sigma}_{kj} : p_k \times p_j \text{ for } k,j = 1, ..., q, \quad \sum_{k=1}^{q} p_k = \sum_{j=1}^{q} p_j = p. \tag{41}$$

We wish to test the null hypothesis that the submatrices (subsamples) $\mathbf{X}_k, ..., \mathbf{X}_j$ $(k, j = 1, ..., q;\ k \neq j)$ are independent, i.e.,

$$H_0: \ \boldsymbol{\Sigma}_{kj} = 0 \text{ for all } k,j = 1, ..., q;\ k \neq j, \tag{42}$$

against the alternative H_1 that H_0 is not true.

Let $\boldsymbol{\Sigma}^*$ be the covariance matrix when H_0 is true,

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & 0 & \ldots & 0 \\ 0 & \boldsymbol{\Sigma}_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{\Sigma}_{qq} \end{pmatrix}. \tag{43}$$

Then the likelihood function becomes

$$L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*) = \prod_{k=1}^{q} L_{p_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{kk}), \tag{44}$$

where

$$L_{p_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{kk}) = \Big(2\pi|\boldsymbol{\Sigma}_{kk}|\Big)^{-n/2} \exp\Big(-\frac{1}{2}\boldsymbol{\Sigma}_{kk}^{-1}\sum_{i=1}^{n}(\mathbf{X}_{ki} - \boldsymbol{\mu}_k)^2\Big), \tag{45}$$

and $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_{kk}$ and $\mathbf{X}_{ki}$ are defined in (40)-(41). It follows that

$$\sup_{\boldsymbol{\mu}, \boldsymbol{\Sigma}^*} L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*) = \prod_{k=1}^{q} \sup_{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{kk}} L_{p_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{kk}) = \prod_{k=1}^{q} L_{p_k}(\widehat{\boldsymbol{\mu}}_k, \widehat{\boldsymbol{\Sigma}}_{kk}), \tag{46}$$

where the hats over $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_{kk}$ mean that we deal with standard MLEs, which are known to be the mean and the sample covariance matrix of the corresponding subsample. Finally,

$$\sup_{\boldsymbol{\mu}, \boldsymbol{\Sigma}^*} L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*) = c \prod_{k=1}^{q} |\widehat{\boldsymbol{\Sigma}}_{kk}|^{-n/2}, \tag{47}$$

where $|\mathbf{\Sigma}_{kk}|$ is the determinant of the corresponding submatrix and

$$c = (2\pi)^{-pn/2} \exp\Big(-\frac{pn}{2} \Big). \tag{48}$$

The likelihood function under H_1 is given by (6) and the maximum likelihood estimators in the case of an unstructured covariance matrix are known to be the sample mean and the sample covariance matrix, therefore

$$\sup_{\boldsymbol{\mu},\mathbf{\Sigma}} L_p(\boldsymbol{\mu}, \mathbf{\Sigma}) = L_p(\widehat{\boldsymbol{\mu}}, \widehat{\mathbf{\Sigma}}) = (2\pi)^{-pn/2} \exp\Big(-\frac{pn}{2} \Big)\, |\widehat{\mathbf{\Sigma}}|^{-n/2} = c\, |\widehat{\mathbf{\Sigma}}|^{-n/2}. \tag{49}$$

The likelihood ratio test is then given by (see, for example, Muirhead (2005))

$$\Lambda = \frac{|\widehat{\mathbf{\Sigma}}_p|^{n/2}}{\prod_{k=1}^{q} |\widehat{\mathbf{\Sigma}}_{kk}|^{n/2}} = \left(\frac{|\widehat{\mathbf{\Sigma}}|}{\prod_{k=1}^{q} |\widehat{\mathbf{\Sigma}}_{kk}|} \right)^{n/2}. \tag{50}$$

For the unstructured covariance matrix of size $p \times p$ the number of parameters to be estimated under H_1 is $p(p+1)/2$, whereas the number of parameters to be estimated under H_0 is $\sum_{k=1}^{q} p_k(p_k+1)/2 = (\sum_{k=1}^{q} p_k^2 + p)/2$. Thus the difference in the number of parameters is $f = (p^2 - \sum_{k=1}^{q} p_k^2)/2$ and the distribution of statistics $-2\ln\Lambda$ is asymptotically χ^2 with f degrees of freedom.

4.2 $H_0 : \sigma_{kj} = 0$ for all $k, j = 1, ..., p; k \neq j$; banded covariance matrix

Let $\mathbf{X}$ be given by (1) and $\mathbf{\Sigma}$ by (9). We wish to test the null hypothesis that the subvectors $\mathbf{x}_k, ..., \mathbf{x}_j$ $(k, j = 1, ..., p;\ k \neq j)$ are independent, i.e.,

$$H_0 :\ \sigma_{kj} = 0 \ \text{ for all } \ k, j = 1, ..., p;\ k \neq j, \tag{51}$$

against the alternative H_1 that H_0 is not true.

Let $\mathbf{\Sigma}^*$ be the covariance matrix when H_0 is true,

$$\mathbf{\Sigma}^* = \begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{pmatrix}. \tag{52}$$

Then the likelihood function becomes

$$L_p(\boldsymbol{\mu}, \mathbf{\Sigma}^*) = \prod_{k=1}^{p} L_1(\mu_k, \sigma_{kk}), \tag{53}$$

where

$$L_1(\mu_k, \sigma_{kk}) = \Big(2\pi\sigma_{kk}\Big)^{-n/2} \exp\Big(- \frac{1}{2\sigma_{kk}} \sum_{i=1}^{n} (x_{ki} - \mu_k)^2 \Big). \tag{54}$$

Thus,

$$\sup_{\boldsymbol{\mu},\boldsymbol{\Sigma}^*} L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*) = \prod_{k=1}^{p} \sup_{\mu_k,\sigma_{kk}} L_1(\mu_k, \sigma_{kk}) = \prod_{k=1}^{p} L_1(\widehat{\mu}_k, \widehat{\sigma}_{kk}) = c \prod_{k=1}^{p} (\widehat{\sigma}_{kk})^{-n/2}, \tag{55}$$

where $\widehat{\mu}_k$ and $\widehat{\sigma}_{kk}$ are standard MLEs, and c is given by (48).

The likelihood function under H_1 is given by (33). Let $\widetilde{\sup}_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the likelihood value when our estimators have been inserted in the likelihood. Using the approach presented in Section 3, one has

$$\begin{aligned}
\widetilde{\sup_{\boldsymbol{\mu},\boldsymbol{\Sigma}}} L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Big(2\pi\hat{\sigma}_{11}\Big)^{-n/2} \exp\Big(- \frac{1}{2\hat{\sigma}_{11}} \sum_{i=1}^{n} (x_{1i} - \hat{\mu}_1)^2 \Big) \\
&\quad \times \prod_{k=2}^{p} \Big(2\pi\hat{\sigma}_{k|1,\dots k-1}\Big)^{-n/2} \exp\Big(- \frac{1}{2\hat{\sigma}_{k|1,\dots k-1}} \sum_{i=1}^{n} (x_{ki} - \hat{\mu}_{k|1,\dots k-1})^2 \Big) \\
&= c\ (\hat{\sigma}_{11})^{-n/2} \prod_{k=2}^{p} \big(\hat{\sigma}_{k|1,\dots k-1}\big)^{-n/2},
\end{aligned} \tag{56}$$

with the estimators of μ_1 and σ_{11} being given by (13) and those of $\mu_{k|1,\dots k-1}$ and $\sigma_{k|1,\dots k-1}$ by (38). Finally, it reduces to

$$\widetilde{\sup_{\boldsymbol{\mu},\boldsymbol{\Sigma}}} L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = c \left(\hat{\sigma}_{11} \prod_{k=2}^{p} \frac{|\hat{\boldsymbol{\Sigma}}_k|}{|\hat{\boldsymbol{\Sigma}}_{k-1}|} \right)^{-n/2} = c\ |\hat{\boldsymbol{\Sigma}}_p|^{-n/2}. \tag{57}$$

We have indicated in Section 3 that our estimators for the banded covariance matrix are consistent and therefore are asymptotically equivalent to the maximum likelihood estimators. Thus, we can use the estimators to construct a test similar to the traditional likelihood ratio test. This likelihood-based test is given by

$$\Lambda_r = \frac{\sup_{\boldsymbol{\mu},\boldsymbol{\Sigma}^*} L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*)}{\widetilde{\sup}_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \frac{|\hat{\boldsymbol{\Sigma}}_p|^{n/2}}{\prod_{k=1}^{p} (\widehat{\sigma}_{kk})^{n/2}} = \left(\frac{|\hat{\boldsymbol{\Sigma}}_p|}{\prod_{k=1}^{p} (\widehat{\sigma}_{kk})} \right)^{n/2}, \tag{58}$$

and the distribution of $-2\ln\Lambda_r$ is asymptotically that of χ^2. As far as we have shown in Section 3 that the estimators of the diagonal elements of the covariance matrix coincide with the standard MLEs, therefore we replaced $\widehat{\sigma}_{kk}$ by our explicit estimators $\hat{\sigma}_{kk}$ in (58).

Formally, (57)-(58) coincide with the expressions for the unstructured covariance matrix in the previous subsection, see (49)-(50). However, here $\mathbf{\Sigma}_p$ stands for the banded covariance matrix. There is also a significant difference in the number of degrees of freedom of the χ^2 distribution. For a banded covariance matrix of size $p \times p$ the number of parameters to be estimated under H_1 is $2p-1$, whereas the number of parameters to be estimated under H_0 is p. Thus the difference in the number of parameters is $f = p-1$ and the distribution of $-2\ln\Lambda_r$ is asymptotically χ^2 with $f = p-1$ degrees of freedom.

For example, if $p = 3$, the likelihood ratio test is given by

$$\Lambda_r = \left(\frac{|\hat{\mathbf{\Sigma}}_3|}{\hat{\sigma}_{11}\hat{\sigma}_{22}\hat{\sigma}_{33}} \right)^{n/2} \tag{59}$$

and the statistics follows asymptotically that of χ^2 with 2 degrees of freedom.

4.3 $H_0:\ \sigma_{kj} = 0$ for some $k, j = 1, ..., p;\ k \neq j$; banded covariance matrix

The model (40)-(42) can be used when one wants to test the null hypothesis

$$H_0:\ \sigma_{kj} = 0 \ \text{ for some } \ k, j = 1, ..., p;\ k \neq j, \tag{60}$$

against the alternative H_1 that H_0 is not true.

For example, if $p = 3$ and one wants to test the null hypothesis

$$H_0:\ \sigma_{12} = 0, \tag{61}$$

against the alternative H_1 that H_0 is not true. It is easy to see that one may use (40)-(42), with

$$\begin{aligned} &\mathbf{\Sigma}_3 = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}, \\ &\mathbf{\Sigma}_{11} = \begin{pmatrix} \sigma_{11} \end{pmatrix}, \ \mathbf{\Sigma}_{12} = \begin{pmatrix} \sigma_{12} & 0 \end{pmatrix}, \ \mathbf{\Sigma}_{22} = \begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{pmatrix}. \end{aligned} \tag{62}$$

In such a case,

$$\sup_{\boldsymbol{\mu}, \mathbf{\Sigma}^*} L_p(\boldsymbol{\mu}, \mathbf{\Sigma}^*) = c\left(|\hat{\mathbf{\Sigma}}_{11}||\hat{\mathbf{\Sigma}}_{22}|\right)^{-n/2}, \quad \widetilde{\sup_{\boldsymbol{\mu}, \mathbf{\Sigma}}} L_p(\boldsymbol{\mu}, \mathbf{\Sigma}) = c\, |\hat{\mathbf{\Sigma}}_3|^{-n/2}. \tag{63}$$

The test is then based on

$$\Lambda_r = \left(\frac{|\hat{\mathbf{\Sigma}}_3|}{|\hat{\mathbf{\Sigma}}_{11}||\hat{\mathbf{\Sigma}}_{22}|} \right)^{n/2} = \left| \frac{\hat{\sigma}_{11}(\hat{\sigma}_{22}\hat{\sigma}_{33} - \hat{\sigma}_{23}^2) - \hat{\sigma}_{33}\hat{\sigma}_{12}^2}{\hat{\sigma}_{11}(\hat{\sigma}_{22}\hat{\sigma}_{33} - \hat{\sigma}_{23}^2)} \right|^{n/2}, \tag{64}$$

where the asymptotical distribution of $-2\ln\Lambda_r$ is χ^2 with 1 degrees of freedom.

The null hypothesis

$$H_0: \ \sigma_{23} = 0, \tag{65}$$

implies in (40)-(42)that

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix},$$

$$\mathbf{\Sigma}_{11} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \ \mathbf{\Sigma}_{22} = \begin{pmatrix} \sigma_{33} \end{pmatrix}, \ \mathbf{\Sigma}_{12} = \begin{pmatrix} 0 \\ \sigma_{23} \end{pmatrix}. \tag{66}$$

In such a case,

$$\sup_{\boldsymbol{\mu},\mathbf{\Sigma}^*} L_p(\boldsymbol{\mu}, \mathbf{\Sigma}^*) = c\left(|\hat{\mathbf{\Sigma}}_{11}||\hat{\mathbf{\Sigma}}_{22}|\right)^{-n/2}, \quad \widetilde{\sup_{\boldsymbol{\mu},\mathbf{\Sigma}}} L_p(\boldsymbol{\mu}, \mathbf{\Sigma}) = c\,|\hat{\mathbf{\Sigma}}_3|^{-n/2}. \tag{67}$$

The test is then given by

$$\Lambda_r = \left(\frac{|\hat{\mathbf{\Sigma}}_3|}{|\hat{\mathbf{\Sigma}}_{11}||\hat{\mathbf{\Sigma}}_{22}|} \right)^{n/2} = \left| \frac{\hat{\sigma}_{33}(\hat{\sigma}_{11}\hat{\sigma}_{22} - \hat{\sigma}_{12}^2) - \hat{\sigma}_{11}\hat{\sigma}_{23}^2}{\hat{\sigma}_{33}(\hat{\sigma}_{11}\hat{\sigma}_{22} - \hat{\sigma}_{12}^2)} \right|^{n/2}, \tag{68}$$

and the asymptotical distribution of $-2\ln\Lambda_r$ is χ^2 with 1 degrees of freedom.

Another example could be the case with $p = 6$, where one wants to test the null hypothesis

$$H_0: \ \sigma_{34} = \sigma_{56} = 0, \tag{69}$$

against the alternative H_1 that H_0 is not true. Then (40)-(42) equals

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \mathbf{\Sigma}_{23} \\ \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} & \mathbf{\Sigma}_{33} \end{pmatrix},$$

$$\mathbf{\Sigma}_{11} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ 0 & \sigma_{32} & \sigma_{33} \end{pmatrix}, \ \mathbf{\Sigma}_{22} = \begin{pmatrix} \sigma_{44} & \sigma_{45} \\ \sigma_{54} & \sigma_{55} \end{pmatrix}, \ \mathbf{\Sigma}_{33} = \begin{pmatrix} \sigma_{66} \end{pmatrix},$$

$$\mathbf{\Sigma}_{12} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \sigma_{34} & 0 \end{pmatrix}, \ \mathbf{\Sigma}_{13} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \ \mathbf{\Sigma}_{23} = \begin{pmatrix} 0 \\ \sigma_{56} \end{pmatrix}. \tag{70}$$

Now,

$$\sup_{\boldsymbol{\mu},\boldsymbol{\Sigma}^*} L_p(\boldsymbol{\mu},\boldsymbol{\Sigma}^*) = c\left(|\hat{\boldsymbol{\Sigma}}_{11}||\hat{\boldsymbol{\Sigma}}_{22}||\hat{\boldsymbol{\Sigma}}_{33}|\right)^{-n/2}, \quad \widetilde{\sup_{\boldsymbol{\mu},\boldsymbol{\Sigma}}} L_p(\boldsymbol{\mu},\boldsymbol{\Sigma}) = c\,|\hat{\boldsymbol{\Sigma}}_6|^{-n/2}. \tag{71}$$

The test is then given by

$$\Lambda_r = \left(\frac{|\hat{\boldsymbol{\Sigma}}_6|}{|\hat{\boldsymbol{\Sigma}}_{11}||\hat{\boldsymbol{\Sigma}}_{22}||\hat{\boldsymbol{\Sigma}}_{33}|}\right)^{n/2}, \tag{72}$$

with

$$\begin{aligned} |\hat{\boldsymbol{\Sigma}}_{11}| &= \hat{\sigma}_{11}\hat{\sigma}_{22}\hat{\sigma}_{33} - \hat{\sigma}_{11}\hat{\sigma}_{23}^2 - \hat{\sigma}_{33}\hat{\sigma}_{12}^2, \quad |\hat{\boldsymbol{\Sigma}}_{22}| = \hat{\sigma}_{44}\hat{\sigma}_{55} - \hat{\sigma}_{45}^2, \quad |\hat{\boldsymbol{\Sigma}}_{33}| = \hat{\sigma}_{66}, \\ |\hat{\boldsymbol{\Sigma}}_6| &= |\hat{\boldsymbol{\Sigma}}_{11}||\hat{\boldsymbol{\Sigma}}_{22}||\hat{\boldsymbol{\Sigma}}_{33}| - |\hat{\boldsymbol{\Sigma}_{11}}|\hat{\sigma}_{44}\hat{\sigma}_{56}^2 - \hat{\sigma}_{34}^2(\hat{\sigma}_{11}\hat{\sigma}_{22} - \hat{\sigma}_{12}^2)(\hat{\sigma}_{55}\hat{\sigma}_{66} - \hat{\sigma}_{56}^2), \end{aligned} \tag{73}$$

and the asymptotical distribution of $-2\ln\Lambda_r$ is χ^2 with 2 degrees of freedom.

5 Simulation

The examples presented here illustrate the results obtained in Section 3. We will compare the explicit estimators derived in our study for the mean and covariance matrix with the true values.

In each simulation a sample of 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).

A small sample with the number of observations equal to $n = 10$, a moderate sample with $n = 100$, and a large sample with $n = 1000$ were considered. Simulations were repeated $N = 1000$ times and estimators were averaged.

Banded covariance structures with m=1 and $p = (3, 4, 5, 6, 8, 10)$ were analyzed and the results of the simulation study are presented in the tables given below.

5.1 True values vs explicit estimators for the mean and covariance matrix

Here p stands for the dimension of observations in the sample, i.e. the size of the covariance matrix, and n for a number of observations in a sample.

Table 1. $p = 3$

	True values	**Estimators**		
		$n = 10$	$n = 100$	$n = 1000$
μ_1	1.0000	1.0062	1.0005	0.9995
μ_2	2.0000	1.9966	1.9997	2.0015
μ_3	3.0000	2.9939	3.0017	3.0030

	True values	**Estimators**		
		$n = 10$	$n = 100$	$n = 1000$
σ_{11}	2.0000	1.8494	1.9698	1.9999
σ_{22}	3.0000	2.7277	2.9662	2.9936
σ_{33}	4.0000	3.6174	3.9574	4.0036
σ_{12}	1.0000	0.9414	0.9819	1.0006
σ_{23}	2.0000	1.6344	1.9677	1.9957

Table 2. p=4

	True values	**Estimators**		
		$n = 10$	$n = 100$	$n = 1000$
μ_1	1.0000	1.0169	0.9982	1.0002
μ_2	2.0000	2.0082	1.9871	1.9989
μ_3	3.0000	3.0165	2.9986	2.9965
μ_4	4.0000	4.0232	3.9961	4.0035

	True values	**Estimators**		
		$n = 10$	$n = 100$	$n = 1000$
σ_{11}	2.0000	1.8204	1.9794	1.9995
σ_{22}	3.0000	2.7000	2.9881	2.9954
σ_{33}	4.0000	3.6449	3.9790	3.9951
σ_{44}	5.0000	4.6039	4.9367	5.0033
σ_{12}	1.0000	0.9002	0.9944	1.0003
σ_{23}	2.0000	1.6003	1.9759	1.9939
σ_{34}	1.0000	0.8332	0.9633	0.9956

Table 3. p=5

	True values	Estimators		
		$n = 10$	$n = 100$	$n = 1000$
μ_1	1.0000	1.0120	1.0016	0.9964
μ_2	2.0000	1.9979	2.0025	1.9991
μ_3	3.0000	2.9849	3.0018	3.0042
μ_4	4.0000	4.0054	3.9912	4.0004
μ_5	5.0000	5.0107	4.9936	4.9989

	True values	Estimators		
		$n = 10$	$n = 100$	$n = 1000$
σ_{11}	2.0000	1.8340	1.9763	1.9987
σ_{22}	3.0000	2.7016	2.9597	3.0057
σ_{33}	4.0000	3.5490	3.9554	4.0023
σ_{44}	5.0000	4.4250	4.9717	5.0031
σ_{55}	6.0000	5.4363	5.9244	5.9836
σ_{12}	1.0000	0.9579	0.9914	0.9985
σ_{23}	2.0000	1.5541	1.9515	2.0056
σ_{34}	1.0000	0.7870	1.0003	0.9939
σ_{45}	2.0000	1.6075	1.9502	1.9987

Table 4. p=6

	True values	Estimators		
		$n = 10$	$n = 100$	$n = 1000$
μ_1	1.0000	1.0014	1.0064	0.9993
μ_2	2.0000	1.9926	1.9922	2.0031
μ_3	3.0000	2.9884	2.9843	3.0030
μ_4	4.0000	4.0248	3.9946	3.9995
μ_5	5.0000	5.0094	4.9967	5.0008
μ_6	6.0000	6.0098	5.9919	5.9984

	True values	Estimators		
		$n = 10$	$n = 100$	$n = 1000$
σ_{11}	2.0000	1.8152	1.9780	2.0014
σ_{22}	3.0000	2.7969	2.9692	3.0017
σ_{33}	4.0000	3.6454	3.9568	4.0008
σ_{44}	5.0000	4.3763	4.9740	4.9905
σ_{55}	6.0000	5.4598	5.9381	5.9864
σ_{66}	7.0000	6.5996	6.9157	6.9704
σ_{12}	1.0000	0.9523	0.9805	1.0034
σ_{23}	2.0000	1.6375	1.9676	1.9991
σ_{34}	1.0000	0.8158	0.9734	1.0019
σ_{45}	2.0000	1.5502	1.9681	1.9929
σ_{56}	3.0000	2.5652	2.9344	2.9868

Table 5. p=8

	True values	Estimators		
		$n = 10$	$n = 100$	$n = 1000$
μ_1	1.0000	1.0162	0.9987	1.0015
μ_2	2.0000	2.0125	2.0053	2.0012
μ_3	3.0000	3.0067	3.0024	3.0011
μ_4	4.0000	4.0238	4.0005	4.0030
μ_5	5.0000	5.0007	4.9993	5.0028
μ_6	6.0000	6.0223	5.9991	6.0020
μ_7	5.0000	5.0185	5.0079	4.9989
μ_8	4.0000	4.0158	4.0036	4.0017

	True values	Estimators		
		$n = 10$	$n = 100$	$n = 1000$
σ_{11}	2.0000	1.7645	1.9812	2.0012
σ_{22}	3.0000	2.7091	2.9638	2.9984
σ_{33}	4.0000	3.6563	3.9383	3.9959
σ_{44}	5.0000	4.5560	4.9521	5.0014
σ_{55}	6.0000	5.3711	5.9683	5.9912
σ_{66}	7.0000	6.2825	6.9551	6.9852
σ_{77}	6.0000	5.2672	5.9502	6.0034
σ_{88}	5.0000	4.5454	4.9850	5.0048
σ_{12}	1.0000	0.8562	0.9913	1.0001
σ_{23}	2.0000	1.6602	1.9483	1.9959
σ_{34}	1.0000	0.7859	0.9804	1.0031
σ_{45}	2.0000	1.5737	1.9655	1.9912
σ_{56}	3.0000	2.4004	2.9528	2.9956
σ_{67}	2.0000	1.5227	1.9878	1.9956
σ_{78}	1.0000	0.8537	0.9613	0.9998

Table 6. p=10

	True values	Estimators		
		$n = 10$	$n = 100$	$n = 1000$
μ_1	1.0000	0.9926	1.0005	0.9997
μ_2	2.0000	2.0005	1.9997	1.9998
μ_3	3.0000	3.0048	3.0034	2.9979
μ_4	4.0000	3.9489	3.9930	3.9969
μ_5	5.0000	4.9874	4.9986	5.0030
μ_6	6.0000	6.0208	6.0037	6.0019
μ_7	5.0000	4.9872	5.0196	4.9980
μ_8	4.0000	4.0081	3.9998	3.9977
μ_9	3.0000	3.0128	2.9988	2.9982
μ_{10}	2.0000	1.9976	2.0024	2.0001

	True values	Estimators		
		$n = 10$	$n = 100$	$n = 1000$
σ_{11}	2.0000	1.8204	1.9754	1.9932
σ_{22}	3.0000	2.6713	2.9648	2.9971
σ_{33}	4.0000	3.6196	3.9598	3.9959
σ_{44}	5.0000	4.4362	4.9438	4.9880
σ_{55}	6.0000	5.4208	5.9384	5.9958
σ_{66}	7.0000	6.2828	6.9378	6.9943
σ_{77}	6.0000	5.4992	5.9232	6.0029
σ_{88}	5.0000	4.4700	4.9506	4.9904
σ_{99}	4.0000	3.6308	3.9199	3.9939
$\sigma_{10,10}$	3.0000	2.7154	2.9984	2.9920
σ_{12}	1.0000	0.8789	0.9784	0.9996
σ_{23}	2.0000	1.5962	1.9587	1.9932
σ_{34}	1.0000	0.8173	0.9838	0.9968
σ_{45}	2.0000	1.6463	1.9597	1.9960
σ_{56}	3.0000	2.4161	2.9247	2.9936
σ_{67}	2.0000	1.6141	1.9659	1.9993
σ_{78}	1.0000	0.7389	1.0032	1.0042
σ_{89}	2.0000	1.5883	1.9340	1.9936
$\sigma_{9,10}$	1.0000	0.8441	0.9798	0.9921

5.2 Discussion

The numerical examples presented above show that our explicit estimators for the mean and covariance matrix resemble the true values. However, in the case of small sample study, with a number of observations in a sample $n = 10$, the estimators and true values are not always very close to each other, especially for larger p. However, already for moderate sample study, with $n = 100$, the averages provide reasonable agrement. In general, results of large sample study, with $n = 1000$, are much better, especially for smaller p.

6 Conclusions

In this paper, we have presented a simple algorithm for estimating the mean and covariance matrix for a multivariate normal distribution with banded covariance structure of order $m = 1$. It is shown that the estimator of the mean coincides with the MLE and is unbiased and consistent. The estimator of the covariance matrix is found to be consistent. Simulations confirm that the estimators are accurate.

Acknowledgement

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

References

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

Anderson, T. W. (1984). *Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, New York, USA, 2nd edition.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer, New York, 2nd edition.

Brown, H. and Prescott, R. (1999). *Applied Mixed Models in Medicine.* Statistics in Practice. Wiley, New York.

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*, 46:784–787.

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

Chiu, T. Y. M., Leonard, T., and Tsui, K. W. (1996). The matrix-logarithm covariance model. *J. Amer. Statist. Assoc.*, 91:198–210.

Dempster, A. M. (1972). Covariance selection. *Biometrics*, 28:157–175.

Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2003). *Analysis of Longitudinal Data.* Oxford University Press, Oxford.

Efron, B. and Morris, C. (1976). Multivariate empirical bayes and estimation of covariance matrices. *Ann. Statist.*, 4:22–32.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Royal Soc. A*, 222:309–368.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. Wiley, New York, 2nd edition.

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

Haff, L. R. (1980). Empirical bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics*, 8:586–597.

Hotelling, H. (1931). The generalization of Student's ratio. *The Annals of Mathematical Statistics*, 2:360–378.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 361–379. University of California Press, Berkeley.

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

Leonard, T. and Hsu, J. S. J. (1992). Bayesian inference for a covariance matrix. *Ann. Statist.*, 20:1669–1696.

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. (2005). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York, USA.

Neyman, J. and Pearson, E. (1938). Contribution to the theory of testing statistical hypothesis. *Stat. Res. Mem.*, 2:25–57. (Also in Neyman, J., and Pearson, E. (1967): Joint Statistical Papers. Cambridge: The University Press).

Pal, N. (1993). Estimating the normal dispersion matrix and the precision matrix from a decision theoretic point of view: A review. *Statistical Papers*, 34:1–26.

Pourahmadi, M. (1999). Joint mean-covariance models with application to longitudinal data: unconstrained parameterisation. *Biometrika*, 86:677–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, 87:425–435.

Sinha, B. K. and Ghosh, M. (1987). Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss. *Statistics and Decisions*, 5:201–227.

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

Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*. Elsevier North Holland, Inc., New York, USA. p.47.

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

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. John Wiley and Sons, New York, USA.

Wilks, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrika*, 3:309–326.

Wilks, S. S. (1938). The large sample distribution of the likelihood ratios for testing composite hypothesis. *The Annals of Mathematical Statistics*, 9:60–62.

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.

Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20:32–52.

Wong, F., Carter, C. K., and Kohn, R. (2003). Efficient estimation of covariance selection models. *Biometrika*, 90:809–830.

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

Yang, R. and Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22:1195–1211.