Regression models with unknown singular covariance matrix

Muni S. Srivastava a, Dietrich von Rosen b,∗

aDepartment of Statistics, University of Toronto, 100 St. George Street, Toronto, Canada, Ont. M5S 1A1
bDepartment of Statistics, Swedish University of Agricultural Sciences, P.O. Box 7013, SE-750 07 Uppsala, Sweden

Received 12 February 2001; accepted 24 February 2002
Submitted by G.P.H. Styan

Abstract

In the analysis of the classical multivariate linear regression model, it is assumed that the covariance matrix is nonsingular. This assumption of nonsingularity limits the number of characteristics that may be included in the model. In this paper, we relax the condition of nonsingularity and consider the case when the covariance matrix may be singular. Maximum likelihood estimators and likelihood ratio tests for the general linear hypothesis are derived for the singular covariance matrix case. These results are extended to the growth curve model with a singular covariance matrix. We also indicate how to analyze data where several new aspects appear.

© 2002 Published by Elsevier Science Inc.

Keywords: Estimators; Growth curve model; GMANOVA; Multivariate regression; Rank restriction; Singular covariance matrix; Tests

1. Introduction

Consider the model

\[ Y = BX + \Sigma^{1/2}E, \]

where \( Y : p \times n, B : p \times q, \xi : q \times k, X : k \times n, \) the rank of \( \Sigma = \Sigma^{1/2} \Sigma^{1/2} \) denoted by \( r(\Sigma) = r \leq p \) and known, and the elements of \( E \) are i.i.d. \( N(0, 1) \). It is assumed

*a Corresponding author.
E-mail addresses: srivasta@utstat.toronto.edu (M.S. Srivastava), dietrich@math.uu.se (D. von Rosen).
that \( \xi \) and \( \Sigma \) are unknown, \( \Sigma^{1/2} \) is a positive semi-definite factorization of the covariance matrix \( \Sigma \) and \( B \) and \( X \) are matrices of known constants. When the covariance matrix \( \Sigma \) is known, some special cases of this model have been considered by Mitra and Rao [6] and Rao and Mitra [11, p. 203–206]. However, when \( \Sigma \) is unknown, only least squares estimators in the regression model and some ad hoc testing procedures have been considered by Khatri [4].

When \( B = I_p \) and \( \xi \colon p \times k \), then, (1.1) becomes a model for the multivariate regression. The general model (1.1) is known as the growth curve model in the literature, introduced and developed by Rao [8,9] and Potthoff and Roy [7]. The maximum likelihood estimators of the parameters in the general case were given by Khatri [3]. For a review of the model see [5,13,19]. In this paper we consider the growth curve model (1.1) as well as the multivariate linear regression model when the covariance matrix \( \Sigma \) is singular. The case of a known covariance matrix \( \Sigma \) has been considered by Khatri [4]. In this paper we obtain maximum likelihood estimators of \( \xi \) and \( \Sigma \) and derive the likelihood ratio test (LRT) for the general linear hypothesis. The distribution of the LRT is given. We show how to analyze a data set in which we first determine the rank of the sample covariance matrix, something similar to principal component analysis. Having determined the rank \( r \), we present methods for estimating the parameters and obtain the likelihood ratio tests. The organization of the paper is as follows. In Section 2 we introduce the singular multivariate normal distribution and give a version of its pdf. Results for one-sample and two-sample inference problem are summarized in this section, the proofs of which can easily be obtained for the general regression model discussed in Section 3. In this section we obtain the MLE, derive the likelihood ratio test (LRT) and establish its exact distribution. Then results are generalized to the growth curve model in Section 4. An example is discussed in Section 5. Throughout the report = now and then stands for equality with probability 1. It will be clear from the content when equality holds with probability 1. Nevertheless, sometimes we remind the reader that equality holds with probability 1.

### 2. Singular multivariate normal distributions

Consider a \( p \)-dimensional random vector \( y \) which is normally distributed with mean vector \( \mu \) and covariance matrix, \( \Sigma \), denoted by \( N_p(\mu, \Sigma) \). When \( \Sigma \) is positive definite (\( \Sigma > 0 \)), the probability density function (pdf) of \( y \) is uniquely defined except on sets of probability zero. However, when the covariance matrix \( \Sigma \) is singular and of rank \( r \) the density is restricted to an \( r \)-dimensional subspace, see [2, p. 290] and [10, p. 527–528], and [18, p. 4], hereafter referred to as S & K. This pdf is not uniquely defined as shown in S & K [18]. A version of such a pdf was first given by Khatri [4], using a generalized inverse \( \Sigma^- \) of \( \Sigma \), where \( \Sigma^- \) is defined by \( \Sigma \Sigma^- \Sigma = \Sigma \). Since \( \Sigma \) is of rank \( r \), it has only \( r \) non-zero eigenvalues, say \( \lambda_1 \geq \cdots \geq \lambda_r > 0 \). The corresponding \( p \times r \) matrix of eigenvectors will be denoted by \( \Gamma \), and a generalized
inverse is given by $\Sigma^{-} = \Gamma A^{-1} \Gamma'$, where $A = \text{diag}(\lambda_1, \ldots, \lambda_r)$ is an $r \times r$ diagonal matrix with diagonal elements as $\lambda_1, \ldots, \lambda_r$. Indeed the given $\Sigma^{-}$ is a Moore–Penrose inverse and will be denoted $\Sigma^{+}$. Then from Khatri [4], a version of the pdf of $y$ is given by

$$(2\pi)^{-(1/2)r} |A|^{-1/2} \exp \left(-\frac{1}{2} (y - \mu)' \Sigma^{-}(y - \mu) \right),$$

(2.1)

where $|C|$ stands for the determinant of a square matrix $C$ and with probability 1

$$\Gamma \mu' = \Gamma' y,$$

(2.2)

for a $p \times (p - r)$ semi-orthogonal matrix $\Gamma'$, orthogonal to $\Gamma$, i.e.

$$\Gamma \Gamma' = I_{p-r}, \quad \Gamma' \Gamma = 0.$$  

(2.3)

The likelihood function for a random sample of size $n$ with observation matrix $Y = (y_1 : \ldots : y_n)$ is given by

$$L(\mu, A, \Gamma) = (2\pi)^{-(1/2)r n} |A|^{-(1/2)n} \times \text{etr} \left\{ -\frac{1}{2} A^{-1} (\Gamma' Y - \Gamma' \mu 1') (\Gamma' Y - \Gamma' \mu 1')' \right\},$$

(2.4)

where $\text{etr} A$ stands for the exponential of the trace of the matrix $A$ and

$$\Gamma \mu' = \Gamma' Y$$

(2.5)

with probability 1; here $I$ is an $n \times 1$ row vector of ones, i.e. $1' = (1, \ldots, 1)$. Since $1'(\frac{1}{2} I) 1' = 1', \frac{1}{2} I$ is a g-inverse of $1'$. Let

$$P_i = I - \frac{1}{n} 11' \quad \text{and} \quad P_i = I - \frac{1}{n} 11',$$

(2.6)

where $I_i$ is an $n_i \times 1$ column vector of ones. Then, from (2.5) we get $G' Y P = 0$ giving

$$G' = U (I - (YP)(YP)^{-})$$

(2.7)

for any $(p - r) \times p$ matrix $U$ such that $G' G = I_{p-r}$ and since $YP$ is a rank $r$, $(YP)(YP)^{-}$ is a idempotent matrix of rank $r$ putting some additional restrictions on $U$ but does not determine $G$ uniquely. From (2.3) it is clear that the uniqueness of $G$ is not required. In fact, our analysis will not depend on the choice of $G$. It is only assumed that $G' \Gamma = 0$, i.e. the space generated by the columns of $G$ which are orthogonal to $\Gamma$ and it appears that we have complete knowledge of that space. The likelihood function given in (2.4) will be used to obtain the maximum likelihood estimates of $\Gamma$, $A$ and the “non-fixed” part of $\mu$. These are given in the following.

**Theorem 2.1.** Let $Y \sim N_{p,n}(\mu 1', \Sigma, I_p)$, where $\Sigma$ is of rank $r(\Sigma) = r$, and $\Sigma = \Gamma A \Gamma'$. For $P$ defined in (2.6), let $S = Y P Y'$, and $H$ be the $p \times r$ matrix of eigenvectors corresponding to the $r$ largest eigenvalues of $S$, denoted by $L = \text{diag}(l_1, \ldots, l_r)$. Then the MLE of $\mu$, $A$, $\Gamma$ and $\Sigma$ are respectively given by
The proof of this theorem will follow the general result on the regression model given in Section 3. Similarly, the proofs of the following two theorems can also be obtained from Section 3.

**Theorem 2.2.** Let \( Y \sim N_{p,n}(\mu', \Sigma, I_n) \), where \( r(\Sigma) = r \). The LRT for testing the hypothesis \( H : \Gamma' \mu = 0 \) against the alternative \( A : \Gamma' \mu \neq 0 \) is based on the statistic

\[
T^2_r = n \bar{y}' \left( \hat{\Gamma} S \hat{\Gamma}' \right)^{-1} \hat{\Gamma}' \bar{y},
\]

where

\[
f - r + 1 \frac{T^2_r}{r} \sim F_{r,f-r+1}, \quad f = n - 1,
\]

\( \hat{\Gamma} \) and \( S \) are defined in Theorem 2.1.

**Theorem 2.3.** Let \( Y_1 \sim N_{p,n_1}(\mu_1', \Sigma, I_{n_1}) \) and \( Y_2 \sim N_{p,n_2}(\mu_2', \Sigma, I_{n_2}) \), where \( r(\Sigma) = r \), be independently distributed. For \( P_t \) defined in (2.6) let \( S = Y_1 P_1 Y_1' + Y_2 P_2 Y_2 = HLH' \). The likelihood ratio test for testing the hypothesis that \( H : \Gamma' \mu_1 = \Gamma' \mu_2 \) against the alternative \( A : \Gamma' \mu_1 \neq \Gamma' \mu_2 \) is based on the statistic

\[
T^2_r = \frac{n_1 n_2}{n} (\bar{y}_1 - \bar{y}_2)' \left( \hat{\Gamma} S \hat{\Gamma}' \right)^{-1} (\bar{y}_1 - \bar{y}_2),
\]

where

\[
f - r + 1 \frac{T^2_r}{r} \sim F_{r,f-r+1}, \quad f = n_1 + n_2 - 2.
\]

We shall now examine the test statistic of Theorem 2.2. For \( \Sigma \) of rank \( r \), we have \( \hat{\Gamma}' \hat{S} \hat{\Gamma} = L = \text{diag}(l_1, \ldots, l_r) \). Thus

\[
T^2_r = n \bar{y}' L^{-1} \hat{\Gamma}' \bar{y} = n \sum_{i=1}^{r} \frac{z_i^2}{l_i},
\]

where \( z = (z_1, \ldots, z_r)' = \hat{\Gamma}' \bar{y} \), the mean vector of the first \( r \) sample principal components. For large samples, we know that \( (f - r + 1)T^2_r \) is asymptotically chi-square distributed with \( r \) degree of freedom. Theorem 2.2 gives the exact distribution.

The results of Theorems 2.1–2.3 are great improvements over the ones that might have been used although none have been mentioned or used in the statistical literature except in one case by Rao and Mitra [11, pp. 204–206] for the problem of classification but that too when the covariance matrix \( \Sigma \) is known. Principal components,
however, among others have been used for testing or assessing multivariate normality of the data, see [14,17].

For practical applications of these results, there remains however, a problem of how to determine the rank \( r \) of the covariance matrix \( \Sigma \). In real data analysis involving high-dimensional data, the sample covariance matrix \( f^{-1}S \), \( f = n - 1 \), may never be exactly singular, and thus we may never know \( r \). It is tempting to look for a test for the rank of the covariance matrix \( \Sigma \), something similar to finding the number of factors in factor analysis. But here there is no comparison group available. For example, the simplest test that comes to mind from principal components analysis is to use the test statistic based on

\[
\frac{l_1 + \cdots + l_r}{l_1 + \cdots + l_p},
\]

where the \( l_i \)s are the eigenvalues of \( S \). But under the hypothesis, this statistic takes the value 1 with probability 1. Thus, in practice we will have to follow the pragmatic approach of principal components analysis to decide on the rank. It should also be mentioned that the problem of determining the rank of \( \Sigma \) is quite different than the problem of determining the rank of the regression matrix as considered by Anderson [1].

### 3. Estimation in multivariate regression model

The multivariate regression model is given by

\[
Y = \xi X + \Sigma^{1/2} E, \tag{3.1}
\]

where \( Y : p \times n \), \( X : k \times n \), \( r(X) \leq k \), \( \xi : p \times k \), \( \Sigma^{1/2} \) is a symmetric square root, \( E = (e_{ij}) \) with \( e_{ij} \) i.i.d. \( N(0, 1) \), \( \Sigma : p \times p \) and \( r(\Sigma) = r \leq p \). It is assumed that \( X \) is a known matrix and \( \xi \) is the matrix of unknown mean parameters. We consider the problem of estimating \( \xi \) when the unknown covariance matrix \( \Sigma \) is of rank \( r(\Sigma) = r \leq p \).

Consider an orthogonal matrix \((\Gamma, \Gamma^o)\), where \( \Gamma : p \times r \) and \( \Gamma^o : p \times (p - r) \) are such that

\[
\Sigma = \Gamma \Lambda \Gamma^o, \quad \Gamma^o \Sigma = 0, \quad \Gamma^o \Gamma = 0, \quad \Gamma \Gamma^o = I_r, \quad \Gamma^o \Gamma^o = I_{p-r}
\]

and \( A = \text{diag}(\lambda_1, \ldots, \lambda_r) \) is the diagonal matrix of non-zero eigenvalues of \( \Sigma \). Then (3.1) can be written as

\[
\begin{pmatrix}
\Gamma^o \\
\Gamma^o
\end{pmatrix}
Y = \begin{pmatrix}
\Gamma^o \\
\Gamma^o
\end{pmatrix}
\xi X + \begin{pmatrix}
\Gamma^o \Sigma^{1/2} E \\
0
\end{pmatrix}. \tag{3.2}
\]

It follows from (3.2) that with probability 1

\[
\Gamma^o Y = \Gamma^o \xi X. \tag{3.3}
\]
Thus, (3.3) may be regarded as a linear system of equations in $\xi$. A general solution of (3.3) is given by (see [11, p. 24])

$$
\xi = \Gamma^o\Gamma^o'YX - + \Theta - \Gamma^o\Gamma^o'\Theta XX^{-1}
$$

(3.4)

with probability 1, where $\Theta : q \times k$ is a matrix of parameters. In the above we have used the fact that $(\Gamma^o)^{-1} = \Gamma^o$, and $\Gamma^o'YX^{-1} = \Gamma^o Y$ with probability 1. It may be noted that if the $k \times n$ matrix $X$ is of full rank $k$, then the $k \times k$ idempotent matrix $XX^{-1}$ is of rank $k$ and hence must equal $I_k$. Thus with probability 1 (3.4) becomes

$$
\xi = \Gamma^o\Gamma^o'YX - + \Gamma^o\Theta.
$$

(3.5)

Thus, in either case whether we insert the solution (3.4) or (3.5) in the random part of the model (3.2) we get

$$
\Gamma'Y = \Gamma'\xi + \Gamma'\Sigma^{1/2}E
$$

$$
= \Gamma'\Theta X + \Gamma'\Sigma^{1/2}E,
$$

(3.6)

since from (3.4) $\Gamma'\xi = \Gamma'\Theta$. Thus, the number of unknown mean parameters $(pk)$ remains the same except that now we should be working with $\Theta$ instead of $\xi$.

### 3.1. Maximum likelihood estimators

For simplicity of presentation we shall assume that the rank of the $k \times n$ matrix $X$ is $k$. The likelihood function of $\Theta, \Gamma$, and $A$ is given by

$$
L(\Theta, \Lambda, \Gamma) = c|A|^{-1/(2m)}\exp\left\{-\frac{1}{2}\Lambda^{-1}(\Gamma'Y - \Gamma'\Theta X)(\Gamma'Y - \Gamma'\Theta X)'\right\},
$$

(3.7)

where $c = (2\pi)^{-n/2}$. Let

$$
R = X'(XX')^{-1}X, \quad P = I - R,
$$

$$
S = YPY',
$$

(3.8)

$$
\Gamma'\hat{\Theta} = \Gamma'YX'(XX')^{-1}.
$$

Then,

$$
\text{tr}\Lambda^{-1}(\Gamma'Y - \Gamma'\Theta X)(\Gamma'Y - \Gamma'\Theta X)' = \text{tr}\Lambda^{-1}(\Gamma'Y - \Gamma'\hat{\Theta} X)(\Gamma'Y - \Gamma'\hat{\Theta} X)' + \text{tr}\Lambda^{-1}(\Gamma'\hat{\Theta} X - \Gamma'\Theta X)(\Gamma'\hat{\Theta} X - \Gamma'\Theta X)'
$$

$$
\geq \text{tr}\Lambda^{-1}\Gamma'(Y - \hat{\Theta} X)'(Y - \hat{\Theta} X)',
$$

where equality holds at

$$
\Gamma'\hat{\Theta} = \Gamma'\Theta
$$

since $r(X) = k$. Hence, from (3.4) and since $\Gamma^o'YX^{-1} = \Gamma^o'YX'(XX')^{-1}$ holds with probability 1 it follows that
Lemma 3.1. Let $S \sim W_p \left( \Sigma, n \right)$. Then, with probability 1, $\mathcal{C}(S) \subseteq \mathcal{C}(\Sigma)$ and if $n \geq r = r(\Sigma)$, $\mathcal{C}(S) = \mathcal{C}(\Sigma)$.

Proof. Since $S$ is Wishart distributed it follows that $S = ZZ'$, where $Z \sim N_{p,n}(0, \Sigma, I)$. Furthermore, $Z = \Psi U$ where $\Sigma = \Psi \Psi'$, $\Psi : p \times r$ and the elements in $U : r \times n$ are i.i.d. $N(0, 1)$. Thus $\mathcal{C}(S) = \mathcal{C}(Z) \subseteq \mathcal{C}(\Psi) = \mathcal{C}(\Sigma)$ and if $n \geq r$, $r(U) = r$ with probability 1 and then equality holds.

In order to obtain a positive definite estimate of $\Gamma' \Sigma \Gamma$, we shall assume that $n - r(X) \geq r(\Sigma) = r$ and let

$$S = HLH',$$  

(3.9)

where $H : p \times r$ is semiorthogonal, i.e. $H'H = I_r$, and $L = (l_i)$ is a diagonal matrix of positive eigenvalues of $S$. However, note that (3.9) holds with probability 1 and in explicit calculations we have to consider a diagonal matrix $L$ which consists of the $r$ largest eigenvalues of $S$. As previously $\Sigma = \Gamma \Lambda \Gamma'$. Then, from Lemma 3.1 it follows with probability 1 that

$$\mathcal{C}(\Gamma) = \mathcal{C}(\Sigma) = \mathcal{C}(S) = \mathcal{C}(H).$$

Thus,

$$\Gamma = H \Psi'$$

for some, $\Psi : r \times r$ of full rank. Furthermore, since $\Gamma' \Gamma = I$ and $H'H = I$ we have that

$$I = \Gamma' \Gamma = \Psi' H' H \Psi' = \Psi' \Psi.'$$  

(3.10)

Since $\Psi$ is square and of full rank it follows from (3.10) that $\Psi$ is orthogonal.

The likelihood function (3.7) after maximizing with respect to $\Theta$ becomes

$$L(\hat{\Theta}, A, \Gamma) = c |A|^{-1/2} \exp \left\{ - \frac{1}{2} A^{-1} \Gamma' ST \right\}$$

$$= c |A|^{-1/2} \exp \left\{ - \frac{1}{2} A^{-1} \Gamma' HLH' \Gamma \right\}$$

$$= c |A|^{-1/2} \exp \left\{ - \frac{1}{2} A^{-1} \Psi L \Psi \right\}$$

$$\leq c \prod_{i=1}^{r} \lambda_i^{-1/2} \exp \left\{ - \frac{1}{2} l_i / \lambda_i \right\}$$  

(3.11)
from [18, Theorem 1.10.2(ii)], where \( l_1 > \cdots > l_r \) and \( \lambda_1 > \cdots > \lambda_r \) are now the ordered values of the eigenvalues; this distinction will not be maintained in what follows for the sake of simplicity of presentation.

The equality in (3.11) is obtained for \( \Psi = I \). From the proof of Theorem 1.10.4 in [18] we obtain
\[
\lambda_i^{-(1/2)n} \exp \left\{ -\frac{1}{2} l_i / \lambda_i \right\} \leq (l_i/n)^{-(1/2)n} e^{-(1/2)n},
\]
the equality is achieved at \( \hat{\lambda}_i = l_i/n \). Thus, we obtain the following.

**Theorem 3.1.** Assume \( r(X) = k \) and \( n - k \geq r(X) \). The MLEs of the parameters in (3.1) are given by
\[
\hat{\Lambda} = \frac{1}{n} L, \\
\hat{\Gamma} = H, \\
\hat{\Sigma} = -\frac{1}{n} HLH', \\
\hat{\xi} = Y'X(XX)^{-1},
\]
where \( H \) is the matrix of eigenvectors which correspond to the \( r \) largest eigenvalues, \( l_i \), of \( S \) and \( L = \text{diag}(l_1, \ldots, l_r) \).

**Remark.** If \( X \) is not of full rank \( k \), i.e. \( r(X) < k \), then
\[
\hat{\xi} = Y'X(XX')^{-1} + U(I_k - XX')^{-1},
\]
for an arbitrary \( p \times k \) matrix \( U \); for proof, see [20].

### 3.2. Testing of hypothesis

In this section, we shall consider the problem of testing the hypothesis
\[
H : \Theta C = 0, \quad A : \Theta C \neq 0
\]
for a given \( k \times m \) matrix \( C \) of rank \( m \). That is,
\[
\Theta = \Theta_1 C',
\]
where \( \theta_1 : p \times (k - m) \) is a matrix of new parameters and \( C' : (k - m) \times k \) of rank \( k - m \). Thus, we have a model where the random part is given by
\[
\Gamma'Y = \Gamma'\Theta_1 C'X + \Gamma'\Sigma^{1/2}E
\]
and the maximum likelihood estimators of \( \Theta_1, \Gamma, \) and \( \Lambda \) can be obtained from the previous theorem. The result can be stated as follows.
Theorem 3.2. Assume \( r(X) = k \) and \( n - k + m \geq r \). Let
\[
G_1 = YX'(XX')^{-1}C(C'XX')^{-1}C'(XX')^{-1}XY',
\]
(3.13)\[
S_H = S + G_1,
\]
(3.14)
and \( H_H \) be a matrix of eigenvectors corresponding to the \( r = r(\Sigma) \) largest eigenvalues of \( \lambda_{H1} \) of \( S_H \) and \( H_HH_H = I \). Let \( L_H = \text{diag}(\lambda_{H1}, \ldots, \lambda_{Hr}) \) be the diagonal matrix of the \( r \) largest eigenvalues of \( S_H \). Then, the maximum likelihood estimators of the parameters under \( H \), given by (3.12), equal
\[
\hat{\lambda} = \frac{1}{n}L_H,
\]
\[
\hat{\Gamma} = H_H,
\]
\[
\hat{\Sigma} = \frac{1}{n}H_HL_HH_H,
\]
\[
\hat{\Gamma}'\hat{\xi} = \hat{\Gamma}'\hat{\Theta}_1 = \hat{\Gamma}'YX'C(C'XX'C)^{-1},
\]
\[
\hat{\xi} = YX'(XX)^{-1} - \hat{\Gamma}'YX'(XX)^{-1}C(C'XX'C)^{-1}C'(XX)^{-1}.
\]

It follows from Theorems 3.1–3.2 and (3.7) that the likelihood ratio test for testing \( H \) versus \( A \) is based on the statistic
\[
\lambda^{-1} = \prod_{i=1}^{r} l_{H_{i}}/l_{i},
\]
(3.15)
where \( l_{i} > \cdots > l_{r} \) and \( l_{H_{1}} > \cdots > l_{H_{r}} \) are the \( r \) largest ordered roots of \( S \) and \( S_{H} \), respectively. Moreover,
\[
\lambda^{-1} = \frac{\hat{\Gamma}'H_H\hat{\xi}_H}{|\hat{\Gamma}'\hat{\xi}|} = \frac{|\hat{\Gamma}'H_H(S + G_1)\hat{\xi}_H|}{|\hat{\Gamma}'\hat{\xi}_H|} = \frac{\hat{\Gamma}'H_H\hat{\xi}_H}{|\hat{\Gamma}'\hat{\xi}_H|}(I + \hat{\Gamma}'H_H\hat{\xi}_H)^{-1}\hat{\Gamma}'H_HG_1).
\]
Since, from Lemma 3.1
\[
\mathcal{C}(\hat{\Gamma}) = \mathcal{C}(\Sigma) = \mathcal{C}(\Gamma) = \mathcal{C}(\hat{\Gamma}_H)
\]
it follows that
\[
\hat{\Gamma}_H(\hat{\Gamma}_H'\hat{\xi}_H)^{-1}\hat{\Gamma}_H' = \Gamma(\Gamma'S\Gamma)^{-1}\Gamma'
\]
and there exists an orthogonal matrix \( \Psi \) such that
\[
\hat{\Gamma} = \hat{\Gamma}_H\Psi.
\]
Hence,
\[
\frac{|\hat{\Gamma}_H\hat{\xi}_H|}{|\hat{\Gamma}'\hat{\xi}|} = 1.
\]
with probability 1 and the LRT, with $G_1$ defined in (3.13), is given by

$$
\lambda^{-1} = |I + \Gamma'(SS')^{-1}\Gamma G_1|
= |I + (\Gamma'S\Gamma)^{-1}\Gamma'YY'(XX')^{-1}C(C'XX')^{-1}C'(XX')^{-1}XY'\Gamma'|
$$

(3.16)

Furthermore,

$$
Y'(XX')^{-1}G_1| = |Y'(XX')^{-1}XY'|
$$

and

$$
\begin{align*}
\lambda = & \frac{|SSSE|}{|SSSE + SS(\text{TR})|} \\
& \sim \text{Wr}(\Gamma', I, n - k),
\end{align*}
$$

Thus, the test statistics under $H_0$ is independent of $S$ and $S'$. Hence, we can use standard results from multivariate regression models. In the subsequent

$$
U_{p,q,n} = \frac{|SSSE|}{|SSSE + SS(\text{TR})|},
$$

denotes the standard test statistic in multivariate analysis [16, p. 97] where SSE = sum of squares due to error and SS(\text{TR}) = sum of squares due to hypothesis, $p$ denotes that SSE and SS(\text{TR}) are of size $p \times p$, $q$ is the number of degrees of freedom due to hypothesis and $n$ is the number of degrees of freedom due to error.

**Theorem 3.3.** Let $\lambda$ be given by (3.15). The LRT rejects $H$ given by (3.11) if

$$
\lambda = U_{r,m,n-k} < c,
$$

where $c$ is chosen according to the significance level.

In order to calculate a percentage point, i.e. to determine $c$, in the distribution of $\lambda = U_{r,m,n-k}$, good approximating formulas are available [16, Theorem 4.2.1].

4. Growth curve model

The growth curve model is given in (1.1). Then, as in the multivariate regression case,

$$
\begin{pmatrix}
\Gamma' \\
\Gamma'
\end{pmatrix}
Y = \begin{pmatrix}
\Gamma' \\
\Gamma'
\end{pmatrix}
B\xi X + \begin{pmatrix}
\Gamma'\Sigma^{1/2}E \\
0
\end{pmatrix},
$$

(4.1)

where

$$
\begin{align*}
\Gamma\Gamma' &= 0, \\
\Gamma'\Sigma &= 0, \\
\Gamma' &= I_r, \\
\Gamma : p \times r, \\
\Sigma &= \Gamma A\Gamma',
\end{align*}
$$

where $A = (\lambda_{ij})$ is a diagonal matrix of positive eigenvalues of $\Sigma$ and $\Gamma': p \times (p - r)$ such that $\Gamma'\Gamma = I_{p-r}$. We shall assume that the $p \times q$ matrix $B$ is of rank $q$ and
that the \( k \times n \) matrix \( X \) is of rank \( k \). The results for the general case without these restrictions are available in the technical report of Srivastava and von Rosen [20].

It follows from (4.1) that with probability 1

\[
\Gamma'Y = \Gamma' B\xi X. \tag{4.2}
\]

Thus, (4.2) may be regarded as a linear system of equations in \( \xi \). Equation (4.2) has a solution if and only if (see [11, p. 24])

\[
\Gamma' B\xi X = \Gamma' B(\Gamma' B)\Gamma' YX^- X.
\]

Since, with probability 1,

\[
\Gamma' YX^- X = \Gamma' B\xi XX^- X = \Gamma' B\xi X = \Gamma' Y
\]

and

\[
(\Gamma' B)(\Gamma' B)'\Gamma' Y = (\Gamma' B)(\Gamma' B)'\Gamma' B\xi X = \Gamma' B\xi X = \Gamma' Y,
\]

the above condition is satisfied. Since \( X \) is of rank \( k \) the idempotent matrix \( XX^- \) is \( I_k \) and hence the general solution of (4.2) is given by

\[
\xi = (\Gamma' B)^{-1}(\Gamma' YX^- + \Theta_0 - (\Gamma' B)^{-1}(\Gamma' B)^{-1}XX^-) = (\Gamma' B)^{-1}(\Gamma' YX^- + (I - (\Gamma' B)^{-1}(\Gamma' B))\Theta_0), \tag{4.4}
\]

where \( \Theta_0 \) is an arbitrary \( q \times k \) parameter matrix. Thus, the random part of the growth curve model (4.1) is given by

\[
\Gamma' Y = \Gamma' B\xi X + \Gamma' \Sigma^{1/2} E
\]

where

\[
= \Gamma' B[(\Gamma' B)^{-1}XX^- + \Theta_0 - (\Gamma' B)^{-1}XX^-] + \Gamma' B[I_q - (\Gamma' B)^{-1}XX^-]X + \Gamma' \Sigma^{1/2} E
\]

\[
= \Gamma' B[(\Gamma' B)^{-1}YY^- + \Gamma' B[I_q - (\Gamma' B)^{-1}YY^-]X + \Gamma' \Sigma^{1/2} E, \tag{4.5}
\]

since with probability 1 \( \Gamma' YX^- X = \Gamma' Y \) as shown above. We note that the rank of the matrix \( \Gamma' B \), satisfies

\[
\rho(\Gamma' B) = l \leq \min(p - r, q).
\]

Then, we need to consider the following three cases:

(a) \( l = \rho(\Gamma' B) = q \leq p - r \),
(b) \( l = \rho(\Gamma' B) = p - r < q \),
(c) \( l = \rho(\Gamma' B) < \min(q, p - r) \).

Remark. The condition in (a) is equivalent to \( \rho(B) \cap \rho(\Gamma) = \{0\} \). The condition in (b) states that \( \rho(B) \cap \rho(\Gamma) \neq \{0\} \) but that \( \rho(B) + \rho(\Gamma) \) spans the whole space whereas (c) means that \( \rho(B) \cap \rho(\Gamma) \neq \{0\} \) and that \( \rho(B) + \rho(\Gamma) \) does not span the whole space. Furthermore, observe that \( \rho(\Gamma) = \rho(\Sigma) \).
Let us consider the case (a) first. Since \((\Gamma\theta^\prime B)^{-1}\Gamma\theta^\prime B\) is a \(q \times q\) idempotent matrix of rank \(q\), it implies that
\[
(\Gamma\theta^\prime B)^{-1}\Gamma\theta^\prime B = I_q.
\]
Thus, in this case
\[
\Gamma^\prime Y = \Gamma^\prime B(\Gamma\theta^\prime B)^{-1}\Gamma\theta^\prime Y + \Gamma^\prime \Sigma^{1/2}E
\]
\[
= \Gamma^\prime BB^{-1} + YY^\prime \Sigma^{1/2}E.
\]
since \((\Gamma\theta^\prime B)^{-1}\Gamma\theta^\prime = B^{-1}\) as \(r(\Gamma\theta^\prime B) = q = r(B)\), see [18, p. 13]. Thus in this case there are no unknown parameters \(\xi\).

We shall consider the case (b) in which \(r(\Gamma\theta^\prime B) = p - r < q\). Since \(r(\Gamma\theta^\prime) = r(B)\) it follows from [18, p. 13] that
\[
B(\Gamma\theta^\prime B)^{-1} = (\Gamma\theta^\prime)^{-1} = \Gamma\theta^\prime.
\]
Hence, from (4.5)
\[
\Gamma^\prime Y = \Gamma^\prime \Gamma\theta^\prime Y + \Gamma^\prime (I - \Gamma\theta^\prime B)\Theta_0X + \Gamma^\prime \Sigma^{1/2}E
\]
\[
= \Gamma^\prime BB\Theta_0X + \Gamma^\prime \Sigma^{1/2}E.
\]
(4.6)
where \(\Theta_0\) is a \(q \times k\) matrix of parameters. For the case (c), we have the general result given in (4.5). That is, by defining
\[
Z = (I_p - B(\Gamma\theta^\prime B)^{-1}\Gamma\theta^\prime)Y
\]
we can write (4.5) as
\[
\Gamma^\prime Z = \Gamma^\prime B(I_p - (\Gamma\theta^\prime B)^{-1}\Gamma\theta^\prime B)\Theta_0X + \Sigma^\prime \Sigma^{1/2}E.
\]
Since \(\Gamma\theta^\prime B\) is a \((p - r) \times q\) matrix of rank \(l \leq p - r \leq q\), we can write (see [18, p. 11]),
\[
I - (\Gamma\theta^\prime B)^{-1}\Gamma\theta^\prime B = VV^\prime,
\]
where \(C\) is a \(q \times (q - l)\) matrix of rank \(q - l\). Thus,
\[
B(I - (\Gamma\theta^\prime B)^{-1}\Gamma\theta^\prime B)\Theta_0 = BV^\prime \Theta_0 \equiv A\Theta_1,
\]
and
\[
\Gamma^\prime Z = \Gamma^\prime A\Theta_1X + \Gamma^\prime \Sigma^{1/2}E.
\]
(4.7)
where \(A = BV: p \times (q - l), \ r(A) = q - l, \ \Theta_1 = V^\prime \Theta_0: (q - l) \times k\). Thus we have to estimate \(\Theta_1\) in this model.

Since there are no unknown mean parameters in case (a), we do not need to consider this case by assuming that \(p - r \leq q\). We next focus on case (b). Case (c) is a straightforward extension but the calculations become more involved since \(Z\) is a function of \(\Gamma\theta^\prime\) and will not be treated in this paper. For details we refer to our technical report [20].
4.1. Maximum likelihood estimators when \( r(\Gamma'\mathbf{B}) = p - r \leq q \)

In this subsection, we consider the case when the rank of the matrix \( \Gamma'\mathbf{B} \) is \( p - r \) is less than \( q \). We shall further assume that \( q \leq r \) since otherwise it will be a MANOVA model. From (4.7) the likelihood function for the random part is given by

\[
L(\Theta_0, \Gamma, \Lambda) = (2\pi)^{-n_\eta/2} |\Lambda|^{-n/2} \times \text{etr} \left\{ -\frac{1}{2} \Lambda^{-1} (\Gamma'\mathbf{Y} - \Gamma'\mathbf{B}\Theta_0\mathbf{X})(\Gamma'\mathbf{Y} - \Gamma'\mathbf{B}\Theta_0\mathbf{X}') \right\}.
\]

Clearly for given \( \Gamma \) and \( \Theta_0 \) the MLE of \( \Lambda \) is given by \( \text{diag}((\Gamma'\mathbf{Y} - \Gamma'\mathbf{B}\Theta_0\mathbf{X})(\Gamma'\mathbf{Y} - \Gamma'\mathbf{B}\Theta_0\mathbf{X})') \). Since \( |\text{diag}(\mathbf{A})| \geq |\mathbf{A}| \) for any matrix \( \mathbf{A} \), to obtain the MLE of \( \Theta_0 \) for a given \( \Gamma \), we need to minimize the determinant of \( (\Gamma'\mathbf{Y} - \Gamma'\mathbf{B}\Theta_0\mathbf{X})(\Gamma'\mathbf{Y} - \Gamma'\mathbf{B}\Theta_0\mathbf{X})' \). From [18, Theorem 1.10.3, p. 24] it follows that the determinant is minimized at

\[
\Gamma'\mathbf{B}\hat{\Theta}_0 = \Gamma'\mathbf{B}(\Gamma'(\Gamma'\mathbf{S})^{-1}\Gamma'\mathbf{B})^{-1}\Gamma'(\Gamma'\mathbf{S})^{-1}\Gamma'\mathbf{Y}(\mathbf{X}')^{-1}, \quad (4.8)
\]

where \( \mathbf{S} \) is given in (3.8). Since, \( r(\Gamma'(\Gamma'\mathbf{S})^{-1}) = r(\mathbf{S}) \), it follows from [18, p. 13] that

\[
\Gamma(\Gamma'(\Gamma'\mathbf{S})^{-1})^{-1} = \mathbf{S}^{-1}
\]

which is independent of \( \Gamma \). From (3.9) and (3.8) it follows that \( \Gamma = \mathbf{H}\Psi \) for some orthogonal \( \Psi \) which implies that the given \( \mathbf{S}^{-1} \) is independent of \( \Gamma \). Indeed \( \mathbf{S}^{-1} \) is the Moore–Penrose inverse which will be denoted \( \mathbf{S}^{+} \). Hence, the MLE of \( \Theta_0 \) satisfies

\[
\Gamma'\mathbf{B}\hat{\Theta}_0 = \Gamma'\mathbf{B}(\mathbf{B}'\mathbf{S}^{+} \mathbf{B})^{-1}\mathbf{B}'\mathbf{S}^{+}\mathbf{Y}(\mathbf{X}')^{-1}.
\]

Note that

\[
(\Gamma'\mathbf{Y} - \Gamma'\mathbf{B}\hat{\Theta}_0\mathbf{X})(\Gamma'\mathbf{Y} - \Gamma'\mathbf{B}\hat{\Theta}_0\mathbf{X})' \\
= \Gamma'(\mathbf{Y} - \mathbf{B}\hat{\Theta}_0\mathbf{X})(\mathbf{Y} - \mathbf{B}\hat{\Theta}_0\mathbf{X})' \Gamma' \\
= \Gamma'(\mathbf{Y} - \mathbf{B}(\mathbf{B}'\mathbf{S}^{+} \mathbf{B})^{-1}\mathbf{B}'\mathbf{S}^{+}\mathbf{Y})(\mathbf{Y} - \mathbf{B}(\mathbf{B}'\mathbf{S}^{+} \mathbf{B})^{-1}\mathbf{B}'\mathbf{S}^{+}\mathbf{Y})' \Gamma' \\
= \Gamma'\mathbf{T}\Gamma' \\
\]

To obtain the MLE of \( \Gamma \), we need to minimize the determinant

\[
|\Gamma'\mathbf{T}\Gamma'|
\]

with respect to \( \Gamma \). Let \( \mathbf{M} \) be the \( p \times r \) matrix of eigenvectors corresponding to the non-zero \( r \) eigenvectors of the matrix \( \mathbf{T} \). From Lemma 3.1, we know that

\[
\Gamma = \mathbf{M}\Psi
\]

for an \( r \times r \) orthogonal matrix \( \Psi \). Hence

\[
|\Gamma'\mathbf{T}\Gamma'| = |\mathbf{M}'\mathbf{T}\mathbf{M}|.
\]

Note that \( \mathbf{T} \) can be simplified as

\[
\mathbf{T} = \mathbf{S} + (\mathbf{I} - \mathbf{G}\mathbf{S}^{+})\mathbf{Y}\mathbf{R}'(\mathbf{I} - \mathbf{S}^{+}\mathbf{G}),
\]
where
\[ G = B(B'S^+B)^{-1}B', \quad R = X'(XX')^{-1}X. \]

**Theorem 4.1.** Let \( r = r(\Sigma) \) and \( M \) be a \( p \times r \) matrix of eigenvectors corresponding to the \( r \) largest eigenvalues \( d_i \) of the \( p \times p \) matrix
\[
T = S + (I - GS^+)YRY'(I - S^+G) = (Y - B\hat{\Theta}_0X)(Y - B\hat{\Theta}_0X)'.
\]

If \( n - k \geq r \), and \( r(\Gamma'\Theta) = p - r \leq q \), then the MLE of \( A, \Gamma, \) and \( \zeta \) are given by
\[
\hat{\Lambda} = \frac{1}{n}(\text{diag}(d_1, \ldots, d_r))^{-1} = \frac{1}{n}D, \\
\hat{\Gamma} = M, \\
\hat{\Sigma} = \frac{1}{n}MDM', \\
\hat{\Gamma}'\hat{B}\zeta = \hat{\Gamma}'B\hat{\Theta}_0 = \hat{\Gamma}'GS^+YX'(XX')^{-1}.
\]

From (4.3) and the fact that \( B(\Gamma'\Theta) = (\Gamma'\Theta)^{-1} = \Gamma' \), we get
\[
\hat{B} = B(\Gamma'\Theta) = B\Gamma'(XX')^{-1} + B(I - (\Gamma'\Theta)^{-1}\Gamma'\Theta)\hat{\Theta}_0 = \Gamma\Gamma'\Gamma'(XX')^{-1} + \hat{\Gamma}'\hat{B}\hat{\Theta}_0 = \Gamma\Gamma'\Gamma'(XX')^{-1} + \hat{\Gamma}'GS^+YX'(XX')^{-1} = (I - \hat{\Gamma}'\Gamma'(XX')^{-1} + \hat{\Gamma}'GS^+YX'(XX')^{-1}.
\]

Next we consider the problem of estimating the parameters of the model (4.6) with an additional restriction on the parameter matrix \( \Theta_0 \), namely
\[ F\Theta_0C = 0, \quad (4.9) \]

where \( F : q_1 \times q \), \( r(F) = q_1 \) and \( C : k \times k_1 \), \( r(C) = k_1 \). Since \( F\Theta_0C = 0 \) is equivalent to
\[ (FF')^{-1/2}F\Theta_0C(C'C)^{-1/2} = 0, \]

we may assume without any loss of generality that in (4.9) \( FF' = I_q \) and \( C'C = I_k \).

Let \( K = (F'^\prime, F') : q \times q \) and \( N = (C'^\prime, C) \) be orthogonal matrices of order \( q \times q \) and \( k \times k \), respectively. Then
\[
B\Theta_0X = BK'K\Theta_0NN'X = \tilde{B}\left(\begin{array}{ll}
F'^\prime\Theta_0C & F'^\prime\Theta_0C \\
C'^\prime\Theta_0C & C'^\prime\Theta_0C
\end{array}\right)\tilde{X} = \tilde{B}\left(\begin{array}{l}
\eta_1 \\
\eta_2 \\
\eta_3 \\
\eta_4
\end{array}\right)\tilde{X}.
\]
where $\tilde{B} = BK'$, $\tilde{X} = N'X$, $\eta_1$, $\eta_2$, and $\eta_3$ are matrices of parameters. Since $\eta_4 = F^tC$, we need to find the MLEs of $(\eta_1', \eta_2')', \eta_3$, and $\Lambda$ in the model (4.6), with restrictions in (4.9), rewritten as

$$
\Gamma'Y = \Gamma'B \begin{pmatrix} \eta_1 \\ \eta_3 \\ 0 \end{pmatrix} \tilde{X} + \Gamma'\Sigma^{1/2}E
= \Gamma'\tilde{B}\delta \tilde{X} + \Gamma'\tilde{B}_2\eta_2\tilde{X}_2 + \Gamma'\Sigma^{1/2}E,
$$

where

$$
\delta = \begin{pmatrix} \eta_1 \\ \eta_3 \end{pmatrix}, \quad \tilde{X} = (\tilde{X}_1', \tilde{X}_2'), \quad \tilde{B} = (\tilde{B}_1, \tilde{B}_2').
$$

Since the columns of $\tilde{B}_2$ form a subset of the columns of $\tilde{B}$, it is a nested model introduced by S & K [18, p. 197] and studied by von Rosen [12] and Srivastava [15], among others. Let

$$
\tilde{R}_1 = \tilde{X}_1'(X_1\tilde{X}_1')^{-1}\tilde{X}_1, \quad \tilde{P}_1 = I - \tilde{R}_1
$$

$$
\tilde{S} = \tilde{Y}\tilde{P}_1(I - \tilde{X}_1'(X_2\tilde{P}_1\tilde{X}_2')^{-1}\tilde{X}_2)\tilde{P}_1\tilde{Y}'
$$

$$
\tilde{S}_2 = (Y - \tilde{B}_2\tilde{X}_2)\tilde{X}_2(X_2\tilde{P}_1\tilde{X}_2')'Y - (Y - \tilde{B}_2\tilde{X}_2)\tilde{X}_2(X_2\tilde{P}_1\tilde{X}_2')'.
$$

Then, from [15] the MLE of $\eta_2$ and $\delta$ for fixed $\Gamma$ are given by

$$
\Gamma'\tilde{B}_2\hat{\eta}_2 = \Gamma'\tilde{B}_2(I'\tilde{S}_2^{-1}\tilde{B}_2)'^{-1}\tilde{B}_2'\Gamma(I'\tilde{S}_2^{-1}\tilde{B}_2)'^{-1}\Gamma'Y\tilde{P}_1\tilde{X}_2(X_2\tilde{P}_1\tilde{X}_2')^{-1}
= \Gamma'\tilde{B}(\tilde{B}_2'\tilde{S} + \tilde{B}_2')^{-1}\tilde{S} + \Gamma'Y\tilde{P}_1\tilde{X}_2(X_2\tilde{P}_1\tilde{X}_2')^{-1}
$$

and

$$
\Gamma'\tilde{B}\hat{\delta} = \Gamma'\tilde{B}(\tilde{B}_2'\tilde{S}_2^{-1}\tilde{B}_2')^{-1}\tilde{B}_2'\tilde{S}_2^{-1}(Y - \tilde{B}_2\tilde{X}_2)\tilde{X}_2(X_2\tilde{X}_2)'.
$$

Let $\tilde{M}$ be the $p \times r$ matrix of eigenvectors corresponding to the $r$ largest eigenvalues $\tilde{l}_i$ of the $p \times p$ matrix

$$
\tilde{T}_1 = (Y - \tilde{B}\tilde{\delta}\tilde{X}_1 - \tilde{B}_2\tilde{\eta}_2\tilde{X}_2)(Y - \tilde{B}\tilde{\delta}\tilde{X}_1 - \tilde{B}_2\tilde{\eta}_2\tilde{X}_2)'.
$$

Then

$$
\tilde{\Gamma} = \tilde{M}
$$

and

$$
\tilde{\Lambda}_H = \text{diag}(\tilde{l}_1, \ldots, \tilde{l}_r),
$$

where $\tilde{\Lambda}_H$ is the MLE of $\Lambda$ under the restriction (4.9). Thus the LRT for testing the hypothesis $F^tC = 0$ is based on the statistic

$$
\lambda = \frac{|M'\tilde{T}_1\tilde{M}|}{|M'\tilde{T}_1\tilde{M}|} = \frac{|\Gamma'\tilde{T}|}{|\Gamma'\tilde{T}|}.$$
with probability 1. The distribution of $\lambda$ is thus $U_{q,k,n-k-r+q}$, where $U$ is the statistic defined earlier.

5. Applications

The presented results cannot be applied when working with real data. The model states that $Y \in \mathcal{C}(B) + \mathcal{C}(\Sigma)$ and that $\dim \mathcal{C}(\Sigma) = r \leq p$. In Lemma 3.1 it was shown that if $n - r(X) \geq r$, $\mathcal{C}(\Sigma) = \mathcal{C}(S)$ with probability 1. However, small deviations from the model as well as events which occur with probability 0 will often give a non-singular $S$ which contradicts the model assumption $r(\Sigma) = r$.

How should we proceed when applying the results of the paper? One idea is to construct the matrix $S$ from data such that $r(S) = r$. This is easily carried out by keeping the $r$ largest eigenvalues, $l_i$, of $S$ and then with the help of the corresponding $r$ eigenvectors, which are collected in $H$, construct a new $S$ by putting

$$ S = HLH', $$

where $L = (l_i) : r \times r$.

Moreover, we have to guarantee

$$ Y \in \mathcal{C}(B) + \mathcal{C}(\Sigma) = \mathcal{C}(B) + \mathcal{C}(S) $$

which can be achieved by projecting $Y$ on $\mathcal{C}(B) + \mathcal{C}(S)$ and then we can work with the projected observations. However, if case (b) in Section 3 holds $Y$ is always in $\mathcal{C}(B) + \mathcal{C}(\Sigma)$ and thus no projection has to be performed.

In the next we are going to illustrate the results of Section 4 with the help of the well known dental data set [7]. Data is presented in Table 1. We emphasize that the purpose is just to illustrate our approach and to show the effect of various assumptions concerning the rank of $\Sigma$. A principal component analysis based on $S$ suggests that $r = 2$.

Data is collected in $Y : 4 \times 27$, where the $i$th column of $Y$ consists of the four measurements from the $i$th individual. Furthermore,

$$ B' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} I_{11} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ I_{16} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix}, $$

where $I_i$ is a vector of $i$th ones. Under model (1.1), i.e.

$$ Y = B\xi X + \Sigma^{1/2}E, $$

when it is supposed that $\Sigma$ is of full rank, the maximum likelihood estimator for $\xi$ equals

$$ \hat{\xi} = (B'S^{-1}B)^{-1}B'S^{-1}YX'(XX')^{-1} \quad (5.1) $$
Table 1
Data from 11 girls and 16 boys at ages 8, 10, 12 and 14

<table>
<thead>
<tr>
<th>Individual</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Girls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>20</td>
<td>21.5</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>21.5</td>
<td>24</td>
<td>25.5</td>
</tr>
<tr>
<td>3</td>
<td>20.5</td>
<td>24</td>
<td>24.5</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>23.5</td>
<td>24.5</td>
<td>25</td>
<td>26.5</td>
</tr>
<tr>
<td>5</td>
<td>21.5</td>
<td>23</td>
<td>22.5</td>
<td>23.5</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>21</td>
<td>21</td>
<td>22.5</td>
</tr>
<tr>
<td>7</td>
<td>21.5</td>
<td>22.5</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>23</td>
<td>23.5</td>
<td>24</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>21.5</td>
</tr>
<tr>
<td>10</td>
<td>16.5</td>
<td>19</td>
<td>19</td>
<td>19.5</td>
</tr>
<tr>
<td>11</td>
<td>24.5</td>
<td>25</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td><strong>Boys</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>26</td>
<td>25</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>13</td>
<td>21.5</td>
<td>22.5</td>
<td>23</td>
<td>26.5</td>
</tr>
<tr>
<td>14</td>
<td>23</td>
<td>22.5</td>
<td>24</td>
<td>27.5</td>
</tr>
<tr>
<td>15</td>
<td>25.5</td>
<td>27.5</td>
<td>26.5</td>
<td>27</td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>23.5</td>
<td>22.5</td>
<td>26</td>
</tr>
<tr>
<td>17</td>
<td>24.5</td>
<td>25.5</td>
<td>27</td>
<td>28.5</td>
</tr>
<tr>
<td>18</td>
<td>22</td>
<td>22</td>
<td>24.5</td>
<td>26.5</td>
</tr>
<tr>
<td>19</td>
<td>24</td>
<td>21.5</td>
<td>24.5</td>
<td>25.5</td>
</tr>
<tr>
<td>20</td>
<td>23</td>
<td>20.5</td>
<td>31</td>
<td>26</td>
</tr>
<tr>
<td>21</td>
<td>27.5</td>
<td>28</td>
<td>31</td>
<td>31.5</td>
</tr>
<tr>
<td>22</td>
<td>23</td>
<td>23</td>
<td>23.5</td>
<td>25</td>
</tr>
<tr>
<td>23</td>
<td>21.5</td>
<td>23.5</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>24</td>
<td>17</td>
<td>24.5</td>
<td>26</td>
<td>29.5</td>
</tr>
<tr>
<td>25</td>
<td>22.5</td>
<td>25.5</td>
<td>25.5</td>
<td>26</td>
</tr>
<tr>
<td>26</td>
<td>23</td>
<td>24.5</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>27</td>
<td>22</td>
<td>21.5</td>
<td>23.5</td>
<td>25</td>
</tr>
</tbody>
</table>

and using the data in Table 1 gives

\[ \hat{\xi} = \begin{pmatrix} 17.42 & 15.84 \\ 0.476 & 0.827 \end{pmatrix}. \]

Note that the first column in \( \hat{\xi} \) represents the girls and the second column the boys.
If instead of the maximum likelihood estimator an unweighted estimator is used, i.e. an estimator which is independent of the covariance estimator we get

\[ \hat{\xi} = (B'B)^{-1}B'YX'(XX')^{-1} \] (5.2)

and

\[ \hat{\xi} = \begin{pmatrix} 17.37 & 16.34 \\ 0.480 & 0.784 \end{pmatrix}. \]
When \( r(\Sigma) = 4 \) the maximum likelihood estimator of \( \Sigma \) is given by
\[
\hat{\Sigma} = \begin{pmatrix}
5.12 & 2.44 & 3.61 & 2.52 \\
2.44 & 3.93 & 2.72 & 3.06 \\
3.61 & 2.72 & 5.98 & 3.82 \\
2.52 & 3.06 & 3.82 & 4.62
\end{pmatrix}.
\]

Now we will apply the results of Section 4 and assume that \( r(\Sigma) = 3 \). Hence we first modify \( S \). The eigenvalues of \( S \) equal 382, 67, 54, 23. The corresponding eigenvectors are given by
\[
H = \begin{pmatrix}
0.48 & -0.74 & 0.38 & -0.28 \\
0.42 & 0.38 & 0.61 & 0.55 \\
0.59 & -0.13 & -0.69 & 0.41 \\
0.50 & 0.54 & -0.08 & -0.67
\end{pmatrix}.
\]

Hence, the modified \( S \) of rank 3, which is based on the eigenvectors which correspond to the three largest eigenvalues, equals
\[
S = HLH' = \begin{pmatrix}
134 & 71 & 100 & 63 \\
71 & 98 & 68 & 91 \\
100 & 68 & 158 & 110 \\
63 & 91 & 110 & 114
\end{pmatrix}.
\]

This \( S \) can be compared to the original \( S \)
\[
S = \begin{pmatrix}
135 & 68 & 98 & 68 \\
68 & 105 & 73 & 83 \\
98 & 73 & 161 & 103 \\
68 & 83 & 103 & 125
\end{pmatrix}.
\]

In order to apply the results of Section 4 we have to decide if \( \mathcal{C}(B) \cap \mathcal{C}(\Sigma) \neq \{0\} \) or \( \mathcal{C}(B) \cap \mathcal{C}(\Sigma) = \{0\} \), i.e. if case (a) in Section 4 applies. However, from (5.3) the eigenvector which corresponds to the largest eigenvalue seems almost to be proportional to \((1, 1, 1, 1)'\). Thus since \( \mathcal{C}(S) = \mathcal{C}(\Sigma) \) with probability 1 it follows that it is reasonable to assume that \( \mathcal{C}(B) \cap \mathcal{C}(\Sigma) \neq \{0\} \). According to our approach in the next step we should project the observations \( Y \) on \( \mathcal{C}(B) + \mathcal{C}(S) \) which in this case will not affect \( Y \) since \( \mathcal{C}(B) + \mathcal{C}(S) \) spans the whole space, even if it is supposed that \( \mathcal{C}(B) \cap \mathcal{C}(\Sigma) \neq \{0\} \), i.e. we are in case (b) of Section 4. Application of Theorem 4.1 gives, when \( r(\Sigma) = 3 \),
\[
\hat{x} = \begin{pmatrix}
17.56 & 13.20 \\
0.462 & 1.08
\end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix}
5.59 & 2.53 & 3.45 & 1.83 \\
2.53 & 3.65 & 2.57 & 3.47 \\
3.45 & 2.57 & 5.95 & 4.28 \\
1.83 & 3.47 & 4.28 & 4.67
\end{pmatrix}.
\]

The reader is reminded that \( \hat{x} \) is not the MLE but an analogue of the full rank case.
Acknowledgements

We are grateful to two referees who provided us with detailed comments on an earlier version of the paper. Support from the Natural Sciences and Engineering Research Council (INSE ARC) of Canada and the Swedish Natural Research Council (NFR) are also gratefully acknowledged.

References