# Explicit Estimators of Parameters in the Growth Curve Model with Linearly Structured Covariance Matrices

**Martin Ohlson and Dietrich von Rosen**

# Explicit Estimators of Parameters in the Growth Curve Model with Linearly Structured Covariance Matrices

Martin Ohlson[1]

*Department of Mathematics*
*Linköping University, SE-581 83 Linköping, Sweden*

Dietrich von Rosen

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

## Abstract

Estimation of parameters in the classical Growth Curve model when the covariance matrix has some specific linear structure is considered. In our examples maximum likelihood estimators can not be obtained explicitly and must rely on optimization algorithms. Therefore explicit estimators are obtained as alternatives to the maximum likelihood estimators. From a discussion about residuals, a simple non-iterative estimation procedure is suggested which gives explicit and consistent estimators of both the mean and the linear structured covariance matrix.

**Keywords**: Growth curve model, linearly structured covariance matrix, explicit estimators, residuals.

---

[1]E-mail address to the correspondence author: martin.ohlson@mai.liu.se

# 1 Introduction

The Growth Curve model introduced by Potthoff and Roy (1964) has been extensively studied over many years. It is a generalized multivariate analysis of variance model (GMANOVA) which belongs to the curved exponential family. The mean structure for the Growth Curve model is bilinear in contrary to the ordinary MANOVA model where it is linear. For more details about the Growth Curve model see e.g., Kollo and von Rosen (2005), Kshirsagar and Smith (1995), Srivastava Khatri (1979), Srivastava and von Rosen (1999).

In the MANOVA model, when dealing with measurements on $k$ equivalent psychological tests, one of the first to consider patterned covariance matrices was Wilks (1946). A covariance matrix with equal diagonal elements and equal off-diagonal elements, i.e., a so called uniform structure was studied. The model was extended by Votaw (1948) to a set of blocks where each block had a uniform structure.

Olkin and Press (1969) considered a circular stationary model, where variables are thought of as being equally spaced around a circle, and the covariance between two variables depends only on the distance between the variables. Olkin (1973) studied a multivariate version in which each element was a matrix, and the blocks were patterned.

More generally, group symmetry covariance models may be of interest since they generalize the above models, see for example, Andersson (1975), Jensen (1988), and Perlman (1987). In Nahtman (2006) marginal permutation invariant covariance matrices were considered and it was proven that permutation invariance implies a specific structure for the covariance matrices. In particular shift permutation invariance generates invariant matrices with a Toeplitz structure, e.g., see Christensen (2007) and Marin and Dhorne (2002).

Furthermore, Anderson (1973) studied when the covariance matrix can be written as a linear combination of known symmetric matrices but the coefficients of the linear combinations are unknown parameters to be estimated. Chaudhuri et al. (2007) considered graphical models and derived an algorithm for estimating covariance matrices under the constraint that certain covariances are zero. As a special case of the model discussed by Chaudhuri et al. (2007), Ohlson et al. (2009) studied banded covariance matrices, i.e., covariance matrices with so called $m$-dependence structure.

For the Growth Curve model, when no assumption about the covariance matrix was made, Potthoff and Roy (1964) originally derived a class of weighted estimators for the mean parameter matrix. Khatri (1966) extended this result and showed that the maximum likelihood estimator also is a weighted estimator. Under a certain covariance structure, Rao (1967) and

Reinsel (1982) have shown that the unweighted estimator also is the maximum likelihood estimator. Furthermore, Chinchilli and Walter (1984) has derived the likelihood ratio test for this type of covariance matrix.

Several other types of structured covariance matrices, utilized by the Growth Curve model, do also exist. For example, Khatri (1973) derived the likelihood ratio test for the intraclass covariance structure and Arnold (1981) and Lee (1988) considered the uniform covariance structure. The autoregressive covariance structure which is natural for time series and repeated measurements have been discussed by Fujikoshi et al. (1990), Hudson (1983), and Lee (1988).

Closely connected to the intraclass covariance structure is the random effects covariance structure studied by Rao (1965, 1975), Reinsel (1982, 1984), and Ware (1985). More recently, the random-effect covariance structure have been considered for the mixed MANOVA-GMANOVA models and the Extended Growth Curve models, e.g., see Yokoyama (1995, 1996, 1997).

Inference on the mean parameters strongly depends on the estimated covariance matrix. The covariance matrix for the estimator of the mean is always a function of the covariance matrix. Hence, when testing the mean parameters the estimator of the covariance matrix is very important. Originally, many estimators of the covariance matrix were obtained from non-iterative least squares methods. When computer sources became stronger and covariance matrices with structures were considered iterative methods were introduced such as the maximum likelihood method and the restricted maximum likelihood method, among others. Nowadays, when data sets are very large, non-iterative methods have again become of interest.

In this paper we will study patterned covariance matrices which are linearly structured, i.e., see Kollo and von Rosen (2005, Definition 1.3.7). The goal is not just to obtain reasonable explicit estimators, but also to explore some new inferential ideas which later can be applied to more general models.

The fact that the mean structure is bilinear will result in decompositions of tensor spaces instead of linear spaces as in MANOVA. The estimation procedure which is proposed in this paper will rely on this decomposition. Calculations do not depend on the distribution of the observations, i.e., the normal distribution. However, when studying properties of the estimators the normal distribution is considered.

The organization of this paper is as follows. In Section 2 the main idea is introduced and the decomposition generated by the design matrices is given. In order to support the decomposition presented in Section 2 maximum likelihood estimators for the non-patterned case are presented in Section 3. Fur-

thermore, in Section 4 explicit estimators for patterned covariance matrices in the Growth Curve model are derived. The section will start with a treatment of patterned covariance matrices in the MANOVA model and then it is shown how these estimators may be used when finding overall estimators with the attractive property of being explicit. Finally, some properties of the proposed estimators will be presented in Section 5 and in Section 6 several numerical examples are given.

## 2  Main idea

Throughout this paper matrices will be denoted by capital letters, vectors by bold lower case letters, and scalars and elements of matrices by ordinary letters.

Some general ideas of how to estimate parameters in the Growth Curve model will be presented in this section. The model is defined as follows.

**Definition 2.1** *Let* $\mathbf{X} : p \times n$ *and* $\mathbf{B} : q \times k$ *be the observation and parameter matrices, respectively, and let* $\mathbf{A} : p \times q$ *and* $\mathbf{C} : k \times n$ *be the within and between individual design matrices, respectively. Suppose that* $q \leq p$ *and* $r + p \leq n$, *where* $r = \mathrm{rank}(\mathbf{C})$. *The Growth Curve model is given by*

$$\mathbf{X} = \mathbf{ABC} + \mathbf{E}, \tag{1}$$

*where the columns of* $\mathbf{E}$ *are assumed to be independently* $p-$*variate normally distributed with mean zero and an unknown positive definite covariance matrix* $\boldsymbol{\Sigma}$, *i.e.,* $\mathbf{E} \sim N_{p,n}\left(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{I}_n\right)$.

The estimators of parameters in the model will be derived via a fairly heuristic approach but, among others, the advantage is that it presents a clear way, as illustrated in Section 4, to find explicit estimators of covariance matrices with complicated structures. Our starting point for estimating parameters in the Growth Curve model are the two jointly sufficient statistics, the "mean" $\mathbf{X}\mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^{-}\mathbf{C}$ and the sum of squares matrix

$$\mathbf{S} = \mathbf{X}\left(\mathbf{I} - \mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^{-}\mathbf{C}\right)\mathbf{X}. \tag{2}$$

The distribution of the "mean" and the sum of squares matrix are given by

$$\mathbf{X}\mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^{-}\mathbf{C} \sim N_{p,n}\left(\mathbf{ABC}, \boldsymbol{\Sigma}, \mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^{-}\mathbf{C}\right) \tag{3}$$

and

$$\mathbf{S} = \mathbf{X} \left( \mathbf{I} - \mathbf{C}' \left( \mathbf{CC}' \right)^{-} \mathbf{C} \right) \mathbf{X} \sim W_p \left( \mathbf{\Sigma}, n - r \right), \tag{4}$$

where $^-$ denotes an arbitrary g-inverse, $r = \text{rank}(\mathbf{C})$, $N_{p,n}(\bullet, \bullet, \bullet)$ stands for the matrix normal distribution and $W_p(\bullet, \bullet)$ for the Wishart distribution. Observe that $\mathbf{S}$ and its distribution is independent of the parameter $\mathbf{B}$. If $\mathbf{\Sigma}$ is known we have from least squares theory the estimator (i.e., the BLUE)

$$\widetilde{\mathbf{ABC}} = \mathbf{A} \left( \mathbf{A}' \mathbf{\Sigma}^{-1} \mathbf{A} \right)^{-} \mathbf{A}' \mathbf{\Sigma}^{-1} \mathbf{XC}' \left( \mathbf{CC}' \right)^{-} \mathbf{C}. \tag{5}$$

In this expression there are two projectors involved, $\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{\Sigma}^{-1}$ and $\mathbf{C}'(\mathbf{CC}')^{-}\mathbf{C}$. Here $\mathbf{\Sigma}$ is included in one of the projectors and indeed we are working with the space given by the tensor product $\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A}) \otimes \mathcal{C}(\mathbf{C}')$, where $\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})$ stands for the linear space generated by the columns of $\mathbf{A}$ with an inner product defined via $\mathbf{\Sigma}$ as $< \boldsymbol{x}, \boldsymbol{y} > = \boldsymbol{x}'\mathbf{\Sigma}^{-1}\boldsymbol{y}$. If there is no subscript, as in $\mathcal{C}(\mathbf{C}')$, it means that the standard inner product is assumed.

As a basis for the inference in our models we perform a decomposition of the whole tensor space into three parts:

$$\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A}) \otimes \mathcal{C}(\mathbf{C}') \boxplus (\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A}) \otimes \mathcal{C}(\mathbf{C}'))^{\perp}$$
$$= (\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A}) \otimes \mathcal{C}(\mathbf{C}')) \boxplus \mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})^{\perp} \otimes \mathcal{C}(\mathbf{C}') \boxplus \mathcal{V} \otimes \mathcal{C}(\mathbf{C}')^{\perp}, \tag{6}$$

where $\mathcal{V}$ represents the whole space and $\boxplus$ denotes the orthogonal direct sum of subspaces. The space $\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A}) \otimes \mathcal{C}(\mathbf{C}')$ is used to estimate $\mathbf{ABC}$ and the other two are used to create residuals. If $\mathbf{\Sigma}$ is unknown it should be estimated and a general idea is to use the variation in the residual which for the Growth Curve model is build up by three subresiduals (see Seid Hamid and von Rosen (2006), von Rosen (1995)). However, for our purposes two of the three residuals are merged so that they agree with the decomposition in (6):

$$\left( \mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A}) \otimes \mathcal{C}(\mathbf{C}') \right)^{\perp} = \mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})^{\perp} \otimes \mathcal{C}(\mathbf{C}') \boxplus \mathcal{V} \otimes \mathcal{C}(\mathbf{C}')^{\perp}. \tag{7}$$

For an illustration of the spaces considered above see Figure 1.

The main problem is that $\mathbf{\Sigma}$ is involved in $\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})$ and therefore the two residuals can not immediately be used to estimate $\mathbf{\Sigma}$. However, we make the following important observation: The role of $\mathbf{\Sigma}$ is twofold; it is used as a weight matrix in $\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{\Sigma}^{-1}$ in order to obtain an estimator of $\mathbf{B}$ with small variance, and it describes the variation in data.
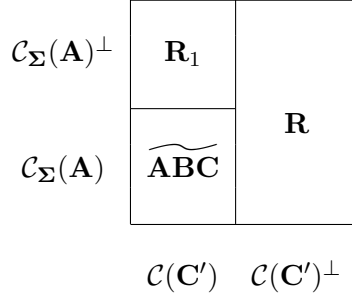
$$\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})^{\perp} \quad \mathbf{R}_1$$
$$\mathbf{R}$$
$$\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A}) \quad \widetilde{\mathbf{ABC}}$$

$$\mathcal{C}(\mathbf{C}') \quad \mathcal{C}(\mathbf{C}')^{\perp}$$

Figure 1: Decomposition of the space generated by the design matrices $\mathbf{A}$ and $\mathbf{C}$. The matrices $\mathbf{R}_1$ and $\mathbf{R}$ given in (8) and (9), respectively, are theoretical residuals used later in the paper.

The theoretical residuals used in this paper which correspond to the subspace decomposition are given by (see also Figure 1)

$$\mathbf{R}_1 = \mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C} - \widetilde{\mathbf{ABC}}$$
$$= \left(\mathbf{I} - \mathbf{A}\left(\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A}\right)^{-}\mathbf{A}'\mathbf{\Sigma}^{-1}\right)\mathbf{X}\mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^{-}\mathbf{C} \quad (8)$$
$$\mathbf{R} = \mathbf{X}\left(\mathbf{I} - \mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^{-}\mathbf{C}\right). \quad (9)$$

Here $\mathbf{R}_1$ is obtained from $\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})^{\perp} \otimes \mathcal{C}(\mathbf{C}')$ and $\mathbf{R}$ from $\mathcal{V} \otimes \mathcal{C}(\mathbf{C}')^{\perp}$. However, since $\mathbf{\Sigma}$ is unknown it has to be estimated in order to make it possible to find an expressions for (5) as well as (8).

It is focused on explicit estimators and we start studying $\mathbf{\Sigma}$ in $\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})$. The matrix of the sum of squares equals $\mathbf{S} = \mathbf{R}\mathbf{R}'$ and since $\mathbf{S}$ is independent of $\mathbf{B}$, $n^{-1}\mathbf{S} \xrightarrow{p} \mathbf{\Sigma}$ ($\xrightarrow{p}$ denotes convergence in probability) and $E[\mathbf{S}] = (n-r)\mathbf{\Sigma}$ we may use as estimator of $\mathbf{\Sigma}$ in $\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{\Sigma}^{-1}$ a function of $\mathbf{S}$, e.g., $n^{-1}\mathbf{S}$. Hence, instead of $\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{\Sigma}^{-1}$ we obtain $\mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1}$ which means that the decomposition in (7) should be replaced by

$$(\mathcal{C}_{\mathbf{S}}(\mathbf{A}) \otimes \mathcal{C}(\mathbf{C}'))^{\perp} = \mathcal{C}_{\mathbf{S}}(\mathbf{A})^{\perp} \otimes \mathcal{C}(\mathbf{C}') \boxplus \mathcal{V} \otimes \mathcal{C}(\mathbf{C}')^{\perp},$$

i.e., $\mathbf{S}$ is used instead of $\mathbf{\Sigma}$ when defining the inner product.

Thus, since the total variation is described by the sum of the squared residuals a natural estimator is

$$n\widehat{\mathbf{\Sigma}} = \mathbf{S} + \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1', \quad (10)$$

where $\widehat{\mathbf{R}}_1 = \left(\mathbf{I} - \mathbf{A}\left(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A}\right)^{-}\mathbf{A}'\mathbf{S}^{-1}\right)\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}$.

# 3 Maximum likelihood estimators

We will present the well known maximum likelihood estimators for the parameters in an ordinary Growth Curve model with a non-patterned covariance matrix $\mathbf{\Sigma}$. The estimators show that the heuristic method presented in the previous section is relevant and that the maximum likelihood approach perfectly fits into it.

The maximum likelihood estimator for the mean parameter $\mathbf{B}$ in the Growth Curve model is given by many authors, e.g., see Khatri (1966), Kollo and von Rosen (2005), and Srivastava and Khatri (1979), and equals

$$\widehat{\mathbf{B}}_{ML} = \left(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A}\right)^{-} \mathbf{A}'\mathbf{S}^{-1}\mathbf{X}\mathbf{C}' \left(\mathbf{C}\mathbf{C}'\right)^{-} + (\mathbf{A}')^o\mathbf{Z}_1 + \mathbf{A}'\mathbf{Z}_2\mathbf{C}^{o\prime}, \qquad (11)$$

where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are arbitrary matrices and $\mathbf{S}$ is given in (2). We have used the notation $\mathbf{A}^o$ for any matrix of full rank which is spanning the orthogonal complement to $\mathcal{C}(\mathbf{A})$, i.e., $\mathcal{C}(\mathbf{A}^o) = \mathcal{C}(\mathbf{A})^{\perp}$.

If $\mathbf{A}$ and $\mathbf{C}$ are of full rank, i.e., $\text{rank}(\mathbf{A}) = q$ and $\text{rank}(\mathbf{C}) = k$, the estimator in (11) reduces to one unique estimator:

$$\widehat{\mathbf{B}}_{ML} = \left(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A}\right)^{-1} \mathbf{A}'\mathbf{S}^{-1}\mathbf{X}\mathbf{C}' \left(\mathbf{C}\mathbf{C}'\right)^{-1}. \qquad (12)$$

Furthermore, the maximum likelihood estimator of $\mathbf{\Sigma}$ is given by

$$n\widehat{\mathbf{\Sigma}}_{ML} = \left(\mathbf{X} - \mathbf{A}\widehat{\mathbf{B}}_{MLE}\mathbf{C}\right) \left(\mathbf{X} - \mathbf{A}\widehat{\mathbf{B}}_{MLE}\mathbf{C}\right)' = \mathbf{S} + \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1', \qquad (13)$$

where the residual $\widehat{\mathbf{R}}_1$ as before equals

$$\widehat{\mathbf{R}}_1 = \mathbf{X}\mathbf{C}' \left(\mathbf{C}\mathbf{C}'\right)^{-1} \mathbf{C} - \mathbf{A}\widehat{\mathbf{B}}_{ML}\mathbf{C}. \qquad (14)$$

Note that $\mathbf{S}$ does not depend on the parameter $\mathbf{B}$ and we know that

$$\frac{1}{n-r}\mathbf{S} \xrightarrow{p} \mathbf{\Sigma}. \qquad (15)$$

Furthermore, from (11) it follows that

$$\mathbf{A}\widehat{\mathbf{B}}_{ML}\mathbf{C} = \mathbf{A} \left(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A}\right)^{-} \mathbf{A}'\mathbf{S}^{-1}\mathbf{X}\mathbf{C}' \left(\mathbf{C}\mathbf{C}'\right)^{-} \mathbf{C} \qquad (16)$$

is always unique, i.e., the expression does not depend on the choice of g-inverses, and therefore $\widehat{\mathbf{\Sigma}}_{ML}$ is also always uniquely estimated.

# 4 Explicit estimators in the growth curve model with a linearly structured covariance matrix

In this section we will derive explicit estimators for the parameters in the Growth Curve model with a covariance matrix which belongs to special class of patterned matrices, i.e., the class of linearly structured matrices, which is presented in the next definition.

**Definition 4.1** *A matrix $\mathbf{\Sigma} = (\sigma_{ij})$ is linearly structured if the only linear structure between the elements is given by $|\sigma_{ij}| = |\sigma_{kl}|$ and there exists at least one $(i,j) \neq (k,l)$ so that $|\sigma_{ij}| = |\sigma_{kl}|$.*

Hence, assume that we have the Growth Curve model

$$\mathbf{X} = \mathbf{ABC} + \mathbf{E},$$

defined in Definition 2.1, but with

$$\mathbf{E} \sim N_{p,n}\left(\mathbf{0}, \mathbf{\Sigma}^{(p)}, \mathbf{I}_n\right),$$

where $\mathbf{\Sigma}^{(p)}$ is a linearly structured covariance matrix.

The estimation procedure which is proposed in this paper will rely on the decomposition of the whole space generated by the design matrices, see Figure 1. When estimating $\mathbf{\Sigma}^{(p)}$ the idea is to use the residual variation as when we obtained the estimator for $\mathbf{\Sigma}$ in the unstructured case. Thus we will consider $\mathbf{S}$ and $\widehat{\mathbf{R}}_1 \widehat{\mathbf{R}}_1'$ and the total residual variation is the sum of these two terms. The problem is how to combine the information from the residuals since the covariance matrix $\mathbf{\Sigma}^{(p)}$ is patterned.

A fundamental idea, which was presented in Section 2, was to decompose the space $\mathcal{V} \otimes \mathcal{C}(\mathbf{C}')^\perp$ in order to estimate the inner product in $\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})$.

Different structures on the covariance matrix may lead to different estimation procedures. Which procedure is the best depends on which linear structure the covariance matrix $\mathbf{\Sigma}^{(p)}$ has.

In this paper we will apply a universal least squares approach and minimize

$$\text{tr}\left\{\left(\mathbf{S} - (n-r)\,\mathbf{\Sigma}^{(p)}\right)\left(\mathbf{S} - (n-r)\,\mathbf{\Sigma}^{(p)}\right)\right\} \tag{17}$$

with respect to $\mathbf{\Sigma}^{(p)}$. For notational convenience $\mathbf{\Sigma}$ will be used instead of $\mathbf{\Sigma}^{(p)}$.

Let $\text{vec}\boldsymbol{\Sigma}(K)$ be the columnwise vectorized form of $\boldsymbol{\Sigma}^{(p)}$ where all 0 and repeated elements (by absolute value) have been disregarded. For example,

$$\boldsymbol{\Sigma}^{(p)} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ 0 & \sigma_{23} & \sigma_{33} \end{pmatrix}$$

gives

$$\text{vec}\boldsymbol{\Sigma}(K) = (\sigma_{11}, \sigma_{12}, \sigma_{22}, \sigma_{23}, \sigma_{33})'.$$

Expression (17) will be differentiated with respect to $\text{vec}\boldsymbol{\Sigma}(K)$ and the collection of partial derivatives, i.e., the matrix derivative to be used, is defined as

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{d\,\text{vec}'\mathbf{Y}}{d\,\text{vec}\mathbf{X}}.$$

For details of how to use matrix derivatives, in particular for linearly structured matrices, see Kollo and von Rosen (2005, Section 1.4). Now,

$$\frac{d\,\text{tr}\left\{(\mathbf{S} - (n-r)\boldsymbol{\Sigma})(\mathbf{S} - (n-r)\boldsymbol{\Sigma})\right\}}{d\,\boldsymbol{\Sigma}(K)}$$

$$= -2(n-r)\frac{d\,\boldsymbol{\Sigma}}{d\,\boldsymbol{\Sigma}(K)}\text{vec}(\mathbf{S} - (n-r)\boldsymbol{\Sigma}) = \mathbf{0}. \qquad (18)$$

Moreover,

$$\frac{d\,\boldsymbol{\Sigma}}{d\,\boldsymbol{\Sigma}(K)} = \left(\mathbf{T}^{+}\right)', \qquad (19)$$

where $\mathbf{T}^{+}$ is the Moore-Penrose inverse of $\mathbf{T}$ defined in Kollo and von Rosen (2005, Theorem 1.3.11.), i.e., $\mathbf{T}$ is a matrix such that

$$\text{vec}\boldsymbol{\Sigma}(K) = \mathbf{T}\text{vec}\boldsymbol{\Sigma}. \qquad (20)$$

The explicit structure and theory around $\mathbf{T}$ and $\mathbf{T}^{+}$ is not of interest to this paper. From (18) and the relation

$$\text{vec}\boldsymbol{\Sigma} = \mathbf{T}^{+}\text{vec}\boldsymbol{\Sigma}(K) \qquad (21)$$

we obtain the linear equation system

$$\left(\mathbf{T}^{+}\right)'\text{vec}\mathbf{S} = (n-r)\left(\mathbf{T}^{+}\right)'\mathbf{T}^{+}\text{vec}\boldsymbol{\Sigma}(K). \qquad (22)$$

8

which gives

$$(n - r)\,\text{vec}\mathbf{\Sigma}(K) = \left((\mathbf{T}^+)'\mathbf{T}^+\right)^-(\mathbf{T}^+)'\text{vec}\mathbf{S} + \left((\mathbf{T}^+)'\mathbf{T}^+\right)^o \mathbf{z},$$

where $\mathbf{z}$ is an arbitrary vector. Hence, the unique estimator is given by

$$\text{vec}\mathbf{\Sigma}^{(p)} = \mathbf{T}^+\text{vec}\mathbf{\Sigma}(K) = \frac{1}{n-r}\mathbf{T}^+\left((\mathbf{T}^+)'\mathbf{T}^+\right)^-(\mathbf{T}^+)'\text{vec}\mathbf{S},$$

i.e., we have a first estimator for $\mathbf{\Sigma}^{(p)}$ given by

$$\text{vec}\widehat{\mathbf{\Sigma}}_1^{(p)} = \frac{1}{n-r}\mathbf{T}^+\left((\mathbf{T}^+)'\mathbf{T}^+\right)^-(\mathbf{T}^+)'\text{vec}\mathbf{S}. \qquad (23)$$

Now, because of $\mathcal{C}(\mathbf{T}^+) = \mathcal{C}(\mathbf{T}')$ and the uniqueness property of projectors, the estimator (23) can be written

$$\text{vec}\widehat{\mathbf{\Sigma}}_1^{(p)} = \frac{1}{n-r}\mathbf{T}'\left(\mathbf{T}\mathbf{T}'\right)^-\mathbf{T}\text{vec}\mathbf{S}. \qquad (24)$$

Following the ideas of Section 2, we may consider $\mathcal{C}_{\widehat{\mathbf{\Sigma}}_1}(\mathbf{A})$ instead of $\mathcal{C}_{\mathbf{\Sigma}}(\mathbf{A})$. From Figure 1 it follows that the estimator of $\mathbf{ABC}$ is obtained by projection on $\mathcal{C}_{\widehat{\mathbf{\Sigma}}_1}(\mathbf{A}) \otimes \mathcal{C}(\mathbf{C}')$, i.e., a natural estimator is given by

$$\mathbf{A}\widehat{\mathbf{B}}\mathbf{C} = \mathbf{A}\left(\mathbf{A}'\widehat{\mathbf{\Sigma}}_1^{-1}\mathbf{A}\right)^-\mathbf{A}'\widehat{\mathbf{\Sigma}}_1^{-1}\mathbf{X}\mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^-\mathbf{C}. \qquad (25)$$

When deriving the final estimator for $\mathbf{\Sigma}^{(p)}$ the idea is to use the residual variation as when we obtained the estimator for $\mathbf{\Sigma}$ in the unstructured case. Thus we will consider $\mathbf{S}$ and $\widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1'$ and the total residual variation is the sum of these two terms. The problem is how to combine the information from the residuals since $\mathbf{\Sigma}^{(p)}$ is a patterned matrix. The distribution of $\mathbf{S}$ is Wishart. Moreover, given the inner product, i.e., conditioning on $\mathbf{S}$, we have

$$\widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1'|\mathbf{S} \sim W_p\left(\widehat{\mathbf{P}}\mathbf{\Sigma}^{(p)}\widehat{\mathbf{P}}', r\right),$$

where the projector $\widehat{\mathbf{P}}$ is given by

$$\widehat{\mathbf{P}} = \mathbf{I} - \mathbf{A}\left(\mathbf{A}'\widehat{\mathbf{\Sigma}}_1^{-1}\mathbf{A}\right)^-\mathbf{A}'\widehat{\mathbf{\Sigma}}_1^{-1}. \qquad (26)$$

Furthermore, since $\widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' = \widehat{\mathbf{P}}\mathbf{S}_0\widehat{\mathbf{P}}'$, where $\mathbf{S}_0 = \mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^-\mathbf{C}\mathbf{X}'$ and $\mathbf{S}$ is independent of $\mathbf{S}_0$ it is very natural to condition $\widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1'$ with respect to $\mathbf{S}$. The variation caused by estimating the inner product is not of any direct

interest and is indeed misleading if using it in the estimation of $\mathbf{\Sigma}^{(p)}$. Again for notational convenience $\mathbf{\Sigma}$ will be used instead of $\mathbf{\Sigma}^{(p)}$. Moreover, the notation $(\mathbf{Q})()'$ is used instead of $(\mathbf{Q})(\mathbf{Q})'$. Once again we will perform a least squares approach and minimize

$$
\mathrm{tr}\left\{ \left( \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' + \mathbf{S} - \left( r\widehat{\mathbf{P}}\mathbf{\Sigma}\widehat{\mathbf{P}}' + (n-r)\mathbf{\Sigma} \right) \right)()' \right\}
$$
$$
= \left( \mathrm{vec}\left( \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' + \mathbf{S} \right) - \widehat{\mathbf{\Psi}}\mathrm{vec}\mathbf{\Sigma} \right)'(), \qquad (27)
$$

where

$$
\widehat{\mathbf{\Psi}} = r\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}} + (n-r)\mathbf{I}, \qquad (28)
$$

with respect to $\mathbf{\Sigma}^{(p)}$. Expression (27) will now be differentiated with respect to $\mathrm{vec}\mathbf{\Sigma}(K)$ and the collection of partial derivatives is given by

$$
\frac{d\left( \mathrm{vec}\left( \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' + \mathbf{S} \right) - \widehat{\mathbf{\Psi}}\mathrm{vec}\mathbf{\Sigma} \right)'()}{d\mathbf{\Sigma}(K)}
$$
$$
= -2\frac{d\mathbf{\Sigma}}{d\mathbf{\Sigma}(K)}\mathbf{\Psi}'\left( \mathrm{vec}\left( \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' + \mathbf{S} \right) - \widehat{\mathbf{\Psi}}\mathrm{vec}\mathbf{\Sigma} \right) = \mathbf{0}. \qquad (29)
$$

Thus, from (19) and (29) we obtain

$$
(\mathbf{T}^+)'\widehat{\mathbf{\Psi}}'\left( \mathrm{vec}\left( \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' + \mathbf{S} \right) - \widehat{\mathbf{\Psi}}\mathrm{vec}\mathbf{\Sigma} \right) = \mathbf{0}
$$

which gives

$$
(\mathbf{T}^+)'\widehat{\mathbf{\Psi}}'\mathrm{vec}\left( \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' + \mathbf{S} \right) = (\mathbf{T}^+)'\widehat{\mathbf{\Psi}}'\widehat{\mathbf{\Psi}}\mathbf{T}^+\mathrm{vec}\mathbf{\Sigma}(K). \qquad (30)
$$

Since

$$
\mathcal{C}\left( (\mathbf{T}^+)'\widehat{\mathbf{\Psi}}'\mathrm{vec}\left( \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' + \mathbf{S} \right) \right) \subseteq \mathcal{C}\left( (\mathbf{T}^+)'\widehat{\mathbf{\Psi}}' \right) = \mathcal{C}\left( (\mathbf{T}^+)'\widehat{\mathbf{\Psi}}'\widehat{\mathbf{\Psi}}\mathbf{T}^+ \right)
$$

equation (30) is consistent and a general solution is given by

$$
\mathrm{vec}\mathbf{\Sigma}(K) = \left( (\mathbf{T}^+)'\widehat{\mathbf{\Psi}}'\widehat{\mathbf{\Psi}}\mathbf{T}^+ \right)^{-}(\mathbf{T}^+)'\widehat{\mathbf{\Psi}}'\mathrm{vec}\left( \widehat{\mathbf{R}}_1\widehat{\mathbf{R}}_1' + \mathbf{S} \right)
$$
$$
+ ((\mathbf{T}^+)'\widehat{\mathbf{\Psi}}'\widehat{\mathbf{\Psi}}\mathbf{T}^+)^o\mathbf{z}, \qquad (31)
$$

where $\mathbf{z}$ is an arbitrary vector. Furthermore, using (21) we have the unique estimator of $\mathbf{\Sigma}^{(p)}$ The result is formulated in the next theorem.

**Theorem 4.1** *The least squares estimator which solves* (29) *is given by*

$$\text{vec}\widehat{\boldsymbol{\Sigma}}^{(p)} = \mathbf{T}^+ \left( (\mathbf{T}^+)' \, \widehat{\boldsymbol{\Psi}}' \widehat{\boldsymbol{\Psi}} \mathbf{T}^+ \right)^- (\mathbf{T}^+)' \, \widehat{\boldsymbol{\Psi}}' \text{vec} \left( \widehat{\mathbf{R}}_1 \widehat{\mathbf{R}}_1' + \mathbf{S} \right), \qquad (32)$$

*where*

$$\begin{aligned}
\widehat{\mathbf{R}}_1 &= \left( \mathbf{I} - \mathbf{A} \left( \mathbf{A}' \widehat{\boldsymbol{\Sigma}}_1^{-1} \mathbf{A} \right)^- \mathbf{A}' \widehat{\boldsymbol{\Sigma}}_1^{-1} \right) \mathbf{X} \mathbf{C}' \left( \mathbf{C} \mathbf{C}' \right)^- \mathbf{C}, \\
\widehat{\boldsymbol{\Psi}} &= r \left( \mathbf{I} - \mathbf{A} \left( \mathbf{A}' \widehat{\boldsymbol{\Sigma}}_1^{-1} \mathbf{A} \right)^- \mathbf{A}' \widehat{\boldsymbol{\Sigma}}_1^{-1} \right) \otimes \left( \mathbf{I} - \mathbf{A} \left( \mathbf{A}' \widehat{\boldsymbol{\Sigma}}_1^{-1} \mathbf{A} \right)^- \mathbf{A}' \widehat{\boldsymbol{\Sigma}}_1^{-1} \right) \\
&\quad + (n - r) \mathbf{I}
\end{aligned}$$

*and* $\widehat{\boldsymbol{\Sigma}}_1$ *is given in* (24)*. Moreover,* $\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}$ *is presented in* (25)*.*

## 5 Properties of the proposed estimators

The proposed estimators (23) (see also (24)), (25) and (32) are ad hoc based least square estimators. Hence, it is important to prove some properties, e.g., unbiasedness and consistency. We will start with the following lemma.

**Lemma 5.1** *The estimator* $\widehat{\boldsymbol{\Sigma}}_1^{(p)}$ *given in* (23) *is a consistent estimator of* $\boldsymbol{\Sigma}^{(p)}$*, i.e.,* $\widehat{\boldsymbol{\Sigma}}_1^{(p)} \xrightarrow{p} \boldsymbol{\Sigma}^{(p)}$*.*

PROOF. We have from (15), that $\frac{1}{n-r}\text{vec}\mathbf{S} \xrightarrow{p} \text{vec}\boldsymbol{\Sigma}^{(p)}$. Hence, from (20), (21) and (23) we have

$$\begin{aligned}
\text{vec}\widehat{\boldsymbol{\Sigma}}_1^{(p)} &= \frac{1}{n-r} \mathbf{T}^+ \left( (\mathbf{T}^+)' \mathbf{T}^+ \right)^- (\mathbf{T}^+)' \text{vec}\mathbf{S} \\
&\xrightarrow{p} \mathbf{T}^+ \left( (\mathbf{T}^+)' \mathbf{T}^+ \right)^- (\mathbf{T}^+)' \text{vec}\boldsymbol{\Sigma}^{(p)} \\
&= \mathbf{T}^+ \left( (\mathbf{T}^+)' \mathbf{T}^+ \right)^- (\mathbf{T}^+)' \mathbf{T}^+ \text{vec}\boldsymbol{\Sigma}(K) \\
&= \mathbf{T}^+ \text{vec}\boldsymbol{\Sigma}(K) = \text{vec}\boldsymbol{\Sigma}^{(p)}
\end{aligned}$$

which completes the proof. $\qquad\square$

Thus, consistency for the estimator $\widehat{\boldsymbol{\Sigma}}_1^{(p)}$ (23) is established and now we can also prove some properties for the estimators (25) and (32). Since the estimator for the mean $\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}$ has dimension $p \times n$ it is pointless to discuss the asymptotic behavior when $n$ tends to infinity. Hence, we will prove the asymptotic properties for the first $m$ columns of $\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}$, i.e., let $\mathbf{C}_m$ be the first $m$ columns in $\mathbf{C}$.

**Theorem 5.2** *Let the estimator* $\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}$ *be given in* (25). *Then*

(i) $\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}$ *is an unbiased estimator of* $\mathbf{ABC}$, *i.e.,* $\mathrm{E}\left(\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}\right) = \mathbf{ABC}$,

(ii) $\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}_m$ *is asymptotic equivalent to*

$$\widetilde{\mathbf{ABC}}_m \sim N_{p,n}\left(\mathbf{ABC}_m, \mathbf{A}\left(\mathbf{A}'\boldsymbol{\Sigma}^{(p)-1}\mathbf{A}\right)^-\mathbf{A}', \mathbf{C}'_m\left(\mathbf{CC}'\right)^-\mathbf{C}_m\right),$$

*i.e.,* $\left\|\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}_m - \widetilde{\mathbf{ABC}}_m\right\| = \mathrm{tr}\left\{\left(\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}_m - \widetilde{\mathbf{ABC}}_m\right)()'\right\} \xrightarrow{p} 0.$

PROOF. (i) Since $\mathbf{S}$ given in (2) and $\mathbf{XC}'$ are independent, $\widehat{\boldsymbol{\Sigma}}_1^{(p)}$ given in (23) and $\mathbf{XC}'$ are also independent. Hence, the expectation of $\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}$ is given by

$$\mathrm{E}\left(\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}\right) = \mathrm{E}\left(\mathbf{A}\left(\mathbf{A}'\widehat{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A}\right)^-\mathbf{A}'\widehat{\boldsymbol{\Sigma}}_1^{-1}\right)\mathrm{E}\left(\mathbf{XC}'\left(\mathbf{CC}'\right)^-\mathbf{C}\right)$$

$$= \mathrm{E}\left(\mathbf{A}\left(\mathbf{A}'\widehat{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A}\right)^-\mathbf{A}'\widehat{\boldsymbol{\Sigma}}_1^{-1}\right)\mathbf{ABCC}'\left(\mathbf{CC}'\right)^-\mathbf{C} = \mathbf{ABC},$$

where the second equality follows from $\mathrm{E}\left(\mathbf{XC}'\left(\mathbf{CC}'\right)^-\mathbf{C}\right) = \mathbf{ABC}$ and the last equality from $\mathbf{A}\left(\mathbf{A}'\widehat{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A}\right)^-\mathbf{A}'\widehat{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A} = \mathbf{A}$.

(ii) Let $\epsilon > 0$ be arbitrary and $\mathbf{M}$ an arbitrary constant matrix. Then

$$\mathrm{P}\left(\left\|\mathbf{A}\widehat{\mathbf{B}}\mathbf{C}_m - \widetilde{\mathbf{ABC}}_m\right\| > \epsilon\right) = \mathrm{P}\left(\left\|\left(\mathbf{Q}_{\widehat{\boldsymbol{\Sigma}}_1} - \mathbf{Q}_{\boldsymbol{\Sigma}}\right)\mathbf{XC}'\left(\mathbf{CC}'\right)^-\mathbf{C}_m\right\| > \epsilon\right)$$

$$= \mathrm{P}\left(\left\|\left(\mathbf{Q}_{\widehat{\boldsymbol{\Sigma}}_1} - \mathbf{Q}_{\boldsymbol{\Sigma}}\right)\mathbf{XC}'\left(\mathbf{CC}'\right)^-\mathbf{C}_m\right\| > \epsilon,\right.$$
$$\left.\mathbf{MM}' - \mathbf{XC}'\left(\mathbf{CC}'\right)^-\mathbf{C}_m\mathbf{C}'_m\left(\mathbf{CC}'\right)^-\mathbf{CX}' > \mathbf{0}\right)$$

$$+ \mathrm{P}\left(\left\|\left(\mathbf{Q}_{\widehat{\boldsymbol{\Sigma}}_1} - \mathbf{Q}_{\boldsymbol{\Sigma}}\right)\mathbf{XC}'\left(\mathbf{CC}'\right)^-\mathbf{C}_m\right\| > \epsilon,\right.$$
$$\left.\mathbf{MM}' - \mathbf{XC}'\left(\mathbf{CC}'\right)^-\mathbf{C}_m\mathbf{C}'_m\left(\mathbf{CC}'\right)^-\mathbf{CX}' \leq \mathbf{0}\right)$$

$$< \mathrm{P}\left(\left\|\left(\mathbf{Q}_{\widehat{\boldsymbol{\Sigma}}_1} - \mathbf{Q}_{\boldsymbol{\Sigma}}\right)\mathbf{M}\right\| > \epsilon\right)$$

$$+ \mathrm{P}\left(\mathbf{MM}' - \mathbf{XC}'\left(\mathbf{CC}'\right)^-\mathbf{C}_m\mathbf{C}'_m\left(\mathbf{CC}'\right)^-\mathbf{CX}' \leq \mathbf{0}\right),$$

where

$$\mathbf{Q}_{\boldsymbol{\Sigma}} = \mathbf{A}\left(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}\right)^-\mathbf{A}'\boldsymbol{\Sigma}^{-1},$$

12

and if $\mathbf{Y}$ is a square matrix, $\mathbf{Y} > \mathbf{0}$ means that $\mathbf{Y}$ is positive definite and $\mathbf{Y} \leq \mathbf{0}$ means that $\mathbf{Y}$ is not positive definite, respectively. From Lemma 5.1 we have $\mathbf{Q}_{\widehat{\mathbf{\Sigma}}_1} \xrightarrow{p} \mathbf{Q}_{\mathbf{\Sigma}}$ and hence, $\mathrm{P}\left( \left\| \left( \mathbf{Q}_{\widehat{\mathbf{\Sigma}}_1} - \mathbf{Q}_{\mathbf{\Sigma}} \right) \mathbf{M} \right\| > \epsilon \right) \to 0$. Furthermore, for some vector $\boldsymbol{\alpha} : p \times 1$ we have

$$\mathrm{P}\left( \mathbf{MM}' - \mathbf{XC}' \left( \mathbf{CC}' \right)^- \mathbf{C}_m \mathbf{C}_m' \left( \mathbf{CC}' \right)^- \mathbf{CX}' \leq \mathbf{0} \right) \tag{33}$$

$$= \mathrm{P}\left( \boldsymbol{\alpha}' \mathbf{XC}' \left( \mathbf{CC}' \right)^- \mathbf{C}_m \mathbf{C}_m' \left( \mathbf{CC}' \right)^- \mathbf{CX}' \boldsymbol{\alpha} \geq \boldsymbol{\alpha}' \mathbf{MM}' \boldsymbol{\alpha} \right)$$

$$\leq \frac{\mathrm{tr}\left\{ \mathbf{C}_m' \left( \mathbf{CC}' \right)^- \mathbf{C}_m \right\} \boldsymbol{\alpha}' \mathbf{\Sigma} \boldsymbol{\alpha} + \boldsymbol{\alpha}' \mathbf{ABC}_m \mathbf{C}_m' \mathbf{B}' \mathbf{A}' \boldsymbol{\alpha}}{\boldsymbol{\alpha}' \mathbf{MM}' \boldsymbol{\alpha}},$$

where we have used the Markov inequality. Since $\mathrm{tr}\left\{ \mathbf{C}_m' \left( \mathbf{CC}' \right)^- \mathbf{C}_m \right\} \leq \mathrm{tr}\left\{ \mathbf{C}_m' \left( \mathbf{C}_m \mathbf{C}_m' \right)^- \mathbf{C}_m \right\} = \mathrm{rank}\left( \mathbf{C}_m \right)$, we can choose the arbitrary matrix $\mathbf{M}$ such that the probability (33) is sufficiently small. The proof is complete. $\square$

**Theorem 5.3** *The estimator $\widehat{\mathbf{\Sigma}}^{(p)}$ given in (32) is a consistent estimator of $\mathbf{\Sigma}^{(p)}$, i.e., $\widehat{\mathbf{\Sigma}}^{(p)} \xrightarrow{p} \mathbf{\Sigma}^{(p)}$.*

PROOF. Using Lemma 5.1 and Cramér-Slutsky's theorem (Cramér, 1946) we have

$$\widehat{\mathbf{P}} \xrightarrow{p} \mathbf{P} = \mathbf{I} - \mathbf{A} \left( \mathbf{A}' \mathbf{\Sigma}^{-1} \mathbf{A} \right)^- \mathbf{A}' \mathbf{\Sigma}^{-1}$$

and

$$\widehat{\mathbf{\Psi}} \xrightarrow{p} \mathbf{\Psi} = r\mathbf{P} \otimes \mathbf{P} + (n - r)\mathbf{I},$$

where $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{\Psi}}$ are given in (26) and (28), respectively. Then

$$\mathrm{vec}\widehat{\mathbf{\Sigma}}^{(p)} = \mathbf{T}^+ \left( (\mathbf{T}^+)' \widehat{\mathbf{\Psi}}' \widehat{\mathbf{\Psi}} \mathbf{T}^+ \right)^- (\mathbf{T}^+)' \widehat{\mathbf{\Psi}}' \mathrm{vec}\left( \widehat{\mathbf{R}}_1 \widehat{\mathbf{R}}_1' + \mathbf{S} \right)$$

$$\xrightarrow{p} \mathbf{T}^+ \left( (\mathbf{T}^+)' \mathbf{\Psi}' \mathbf{\Psi} \mathbf{T}^+ \right)^- (\mathbf{T}^+)' \mathbf{\Psi}' \mathrm{vec}\left( r\mathbf{P} \mathbf{\Sigma} \mathbf{P}' + (n - r)\mathbf{\Sigma} \right)$$

$$= \mathbf{T}^+ \left( (\mathbf{T}^+)' \mathbf{\Psi}' \mathbf{\Psi} \mathbf{T}^+ \right)^- (\mathbf{T}^+)' \mathbf{\Psi}' \mathbf{\Psi} \mathrm{vec}\mathbf{\Sigma}$$

$$= \mathbf{T}^+ \left( (\mathbf{T}^+)' \mathbf{\Psi}' \mathbf{\Psi} \mathbf{T}^+ \right)^- (\mathbf{T}^+)' \mathbf{\Psi}' \mathbf{\Psi} \mathbf{T}^+ \mathrm{vec}\mathbf{\Sigma}(K)$$

$$= \mathbf{T}^+ \mathrm{vec}\mathbf{\Sigma}(K) = \mathrm{vec}\mathbf{\Sigma}^{(p)},$$

since $\mathbf{\Psi}$ has full rank and thus the proof is complete. $\square$

# 6    Examples

**Example 1 (Dental Data, Potthoff & Roy 1964)** Dental measurements on eleven girls and sixteen boys at four different ages $(8, 10, 12, 14)$ were taken. Each measurement is the distance, in millimeters, from the center of pituitary to pteryo-maxillary fissure. Suppose linear growth curves describe the mean growth for both the girls and the boys. Then we may use the Growth Curve model where the observation, parameter and design matrices are given as follows (notice the non-traditional way of presenting the $4 \times 27$ observation matrix)

$$\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{27})$$

$$= \begin{pmatrix}
21 & 21 & 20.5 & 23.5 & 21.5 & 20 & 21.5 & 23 & 20 & \ldots \\
16.5 & 24.5 & 26 & 21.5 & 23 & 20 & 25.5 & 24.5 & 22 & \ldots \\
\ldots & 24 & 23 & 27.5 & 23 & 21.5 & 17 & 22.5 & 23 & 22 \\
20 & 21.5 & 24 & 24.5 & 23 & 21 & 22.5 & 23 & 21 & \ldots \\
19 & 25 & 25 & 22.5 & 22.5 & 23.5 & 27.5 & 25.5 & 22 & \ldots \\
\ldots & 21.5 & 20.5 & 28 & 23 & 23.5 & 24.5 & 25.5 & 24.5 & 21.5 \\
21.5 & 24 & 24.5 & 25 & 22.5 & 21 & 23 & 23.5 & 22 & \ldots \\
19 & 28 & 29 & 23 & 24 & 22.5 & 26.5 & 27 & 24.5 & \ldots \\
\ldots & 24.5 & 31 & 31 & 23.5 & 24 & 26 & 25.5 & 26 & 23.5 \\
23 & 25.5 & 26 & 26.5 & 23.5 & 22.5 & 25 & 24 & 21.5 & \ldots \\
19.5 & 28 & 31 & 26.5 & 27.5 & 26 & 27 & 28.5 & 26.5 & \ldots \\
\ldots & 25.5 & 26 & 31.5 & 25 & 28 & 29.5 & 26 & 30 & 25
\end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} b_{01} & b_{02} \\ b_{11} & b_{12} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \mathbf{1}'_{11} & \mathbf{0}'_{16} \\ \mathbf{0}'_{11} & \mathbf{1}'_{16} \end{pmatrix}.$$

The maximum likelihood estimators for the parameter matrix and the non-patterned covariance matrix are given by

$$\widehat{\mathbf{B}}_{ML} = \begin{pmatrix} 17.4254 & 15.8423 \\ 0.4764 & 0.8268 \end{pmatrix}$$

and

$$\widehat{\boldsymbol{\Sigma}}_{ML} = \begin{pmatrix}
5.1192 & 2.4409 & 3.6105 & 2.5222 \\
2.4409 & 3.9279 & 2.7175 & 3.0623 \\
3.6105 & 2.7175 & 5.9798 & 3.8235 \\
2.5222 & 3.0623 & 3.8235 & 4.6180
\end{pmatrix}.$$

14

Assume that the covariance matrix has Toeplitz structure but with different variances, i.e.,

$$\mathbf{\Sigma}^{(p)} = \begin{pmatrix} \sigma_1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & \sigma_2 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & \sigma_3 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & \sigma_4 \end{pmatrix}.$$

The $\mathbf{T}$ matrix in (20) equals

$$\mathbf{T} = \frac{1}{12} \begin{pmatrix} 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 3 & 3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 12 \end{pmatrix}.$$

The estimates for the parameter matrix and the covariance matrix, (25) and (32) respectively, are given by (for comparisons, the maximum likelihood estimates calculated in Proc Mixed in SAS are also presented)

$$\widehat{\mathbf{B}} = \begin{pmatrix} 17.4647 & 15.6624 \\ 0.4722 & 0.8437 \end{pmatrix}, \qquad \widehat{\mathbf{B}}_{ML} = \begin{pmatrix} 17.4116 & 16.0252 \\ 0.4758 & 0.8216 \end{pmatrix}$$

and

$$\widehat{\mathbf{\Sigma}} = \begin{pmatrix} 5.4809 & 3.2756 & 3.5978 & 2.7136 \\ 3.2756 & 4.2452 & 3.2756 & 3.5978 \\ 3.5978 & 3.2756 & 6.2373 & 3.2756 \\ 2.7136 & 3.5978 & 3.2756 & 4.9514 \end{pmatrix},$$

$$\widehat{\mathbf{\Sigma}}_{ML} = \begin{pmatrix} 5.3929 & 3.2767 & 3.5284 & 2.5024 \\ 3.2767 & 5.1759 & 3.2767 & 3.5284 \\ 3.5284 & 3.2767 & 5.4134 & 3.2767 \\ 2.5024 & 3.5284 & 3.2767 & 4.3192 \end{pmatrix}.$$

In the next we assume that the covariance matrix is Toeplitz and obtain

$$\mathbf{T} = \frac{1}{12} \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 3 \\ 0 & 2 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 2 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 3 & 3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 & 0 & 0 \end{pmatrix},$$

$$\widehat{\mathbf{B}} = \begin{pmatrix} 17.4051 & 16.2589 \\ 0.4764 & 0.7955 \end{pmatrix}, \qquad \widehat{\mathbf{B}}_{ML} = \begin{pmatrix} 17.4092 & 16.2603 \\ 0.4759 & 0.7972 \end{pmatrix}$$

and

$$\widehat{\mathbf{\Sigma}} = \begin{pmatrix} 5.2217 & 3.2946 & 3.5934 & 2.7191 \\ 3.2946 & 5.2217 & 3.2946 & 3.5934 \\ 3.5934 & 3.2946 & 5.2217 & 3.2946 \\ 2.7191 & 3.5934 & 3.2946 & 5.2217 \end{pmatrix},$$

$$\widehat{\mathbf{\Sigma}}_{ML} = \begin{pmatrix} 4.9438 & 3.0506 & 3.4053 & 2.3421 \\ 3.0506 & 4.9438 & 3.0506 & 3.4053 \\ 3.4053 & 3.0506 & 4.9438 & 3.0506 \\ 2.3421 & 3.4053 & 3.0506 & 4.9438 \end{pmatrix}.$$

Another well known covariance structure is the compound symmetry structure given by

$$\mathbf{\Sigma}^{(p)} = \begin{pmatrix} \sigma & \rho & \rho & \rho \\ \rho & \sigma & \rho & \rho \\ \rho & \rho & \sigma & \rho \\ \rho & \rho & \rho & \sigma \end{pmatrix}.$$

If this structure holds we obtain

$$\mathbf{T} = \frac{1}{12} \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 3 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix},$$

$$\widehat{\mathbf{B}} = \begin{pmatrix} 17.3727 & 16.3406 \\ 0.4795 & 0.7844 \end{pmatrix}, \qquad \widehat{\mathbf{B}}_{ML} = \begin{pmatrix} 17.3727 & 16.3406 \\ 0.4796 & 0.7844 \end{pmatrix}$$

and

$$\widehat{\mathbf{\Sigma}} = \begin{pmatrix} 5.2127 & 3.3013 & 3.3013 & 3.3013 \\ 3.3013 & 5.2127 & 3.3013 & 3.3013 \\ 3.3013 & 3.3013 & 5.2127 & 3.3013 \\ 3.3013 & 3.3013 & 3.3013 & 5.2127 \end{pmatrix},$$

$$\widehat{\mathbf{\Sigma}}_{ML} = \begin{pmatrix} 4.9052 & 3.0306 & 3.0306 & 3.0306 \\ 3.0306 & 4.9052 & 3.0306 & 3.0306 \\ 3.0306 & 3.0306 & 4.9052 & 3.0306 \\ 3.0306 & 3.0306 & 3.0306 & 4.9052 \end{pmatrix}.$$

16

We conclude from the above examples that even if we have only 27 observations the proposed estimates are very close to the maximum likelihood estimates.

Finally, the asymptotic behavior of the estimators is illustrated. (25) and (32) when $\mathbf{\Sigma}^{(p)}$ is a banded matrix defined in Ohlson et al. (2009) as

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

where

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

**Example 2 (Simulation study)** In each simulation a sample of size $n = 500$ observations was randomly generated from a $p$-variate Growth Curve model using MATLAB Version 7.4.0 (The Mathworks Inc., Natick, MA, USA). Next, the explicit estimates were calculated in each simulation. Simulations were repeated 500 times and the average values of the obtained estimates were calculated.

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

**(A)** *Simulations for $p = 5$, $m = 1$*

Data was generated with parameters

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \mathbf{1}'_{n/2} & \mathbf{0}'_{n/2} \\ \mathbf{0}'_{n/2} & \mathbf{1}'_{n/2} \end{pmatrix},$$

where $\mathbf{1}_{n/2}$ and $\mathbf{0}_{n/2}$ are vectors of ones and zeroes, respectively, and

$$\mathbf{\Sigma}^{(p)} = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 3 & -2 & 0 & 0 \\ 0 & -2 & 4 & -1 & 0 \\ 0 & 0 & -1 & 5 & 2 \\ 0 & 0 & 0 & 2 & 6 \end{pmatrix}.$$

Based on 500 simulations the average estimates are given by

$$\widehat{\mathbf{B}} = \begin{pmatrix} 0.9999 & 1.0125 \\ 1.0015 & 0.9968 \end{pmatrix},$$

and

$$\widehat{\mathbf{\Sigma}}^{(p)} = \begin{pmatrix} 2.0019 & 0.9903 & 0 & 0 & 0 \\ 0.9903 & 2.9796 & -1.9933 & 0 & 0 \\ 0 & -1.9933 & 4.0189 & -0.9963 & 0 \\ 0 & 0 & -0.9963 & 4.9963 & 1.9887 \\ 0 & 0 & 0 & 1.9887 & 6.0042 \end{pmatrix}.$$

**(B)** *Simulations for $p = 4$, $m = 2$*

Corresponding to the previous case the model is defined through

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \mathbf{1}'_{n/2} & \mathbf{0}'_{n/2} \\ \mathbf{0}'_{n/2} & \mathbf{1}'_{n/2} \end{pmatrix}$$

and

$$\mathbf{\Sigma}^{(p)} = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 3 & 2 & 1 \\ 1 & 2 & 4 & 1 \\ 0 & 1 & 1 & 5 \end{pmatrix}.$$

After 500 simulations average explicit estimates equal

$$\widehat{\mathbf{B}} = \begin{pmatrix} 1.0027 & 0.9797 \\ 1.0065 & 1.0054 \end{pmatrix},$$

and

$$\widehat{\mathbf{\Sigma}}^{(p)} = \begin{pmatrix} 1.9933 & 0.9947 & 0.9924 & 0 \\ 0.9947 & 2.9820 & 1.9950 & 1.0190 \\ 0.9924 & 1.9950 & 4.0091 & 1.0479 \\ 0 & 1.0190 & 1.0479 & 4.9935 \end{pmatrix}.$$

From the above simulations one conclusion is that the explicit estimates derived in this paper perform very well and are close to the true values.

# References

[1] Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Stat.*, 1(1):135–141.

[2] Andersson, S. (1975). Invariant normal models. *Ann. Stat.*, (1975) 3(1):132–154.

[3] Arnold, S. F. (1981). *Theory of Linear Models and Multivariate Analysis.* Wiley, New York, pp. 209–238.

[4] Chaudhuri, S., Drton, M., and Richardson, T. S. (2007). Estimation of a covariance matrix with zeros. *Biom.*, 94(1):199–216.

[5] Chinchilli, V. M. and Walter Jr, H. C. (1984). A likelihood ratio test for a patterned covariance matrix in a multivariate growth-curve model. *Biometrics*, 40:151–156.

[6] Christensen, L.P.B. (2007). An EM-algorithm for band-Toeplitz covariance matrix estimation. *In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*

[7] Cramér, H. (1946). *Mathematical Methods Of Statistics.* Princeton University Press, Princeton, pp. 254-255.

[8] Fujikoshi, Y., Kanda, T., and Tanimura, N. (1990). The Growth curve model with an autoregressive covariance structure. *Ann. Inst. Stat. Math.*, 42:533–542.

[9] Hudson, I. L. (1983). Asymptotic tests for Growth curve models with autoregressive errors. *Austral. J. Stat.*, 25:413–424.

[10] Jensen, S. T. (1988). Covariance hypotheses which are linear in both the covariance and the inverse covariance. *The Annals of Statistics*, 16(1):302–322.

[11] Khatri, C. G. (1966). A note on a MANOVA model applied to problems in Growth Curve. *Ann. Inst. Stat. Math.*, 18:75–86.

[12] Khatri, C. G. (1973). Testing some covariance structures under a Growth Curve model. *Journal of Multivariate Statistics*, 3:102–116.

[13] Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices.* Springer, New York.

[14] Kshirsagar, A. M. and Smith, W. B.. (1995). *Growth Curves.* Marcel Dekker, New York.

[15] Lee, J. C. (1988). Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association*, 83:432–440.

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

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

[18] Ohlson, M., Andrushchenko, Z., and von Rosen, D. (2009). Explicit estimators under m-dependence for a multivariate normal distribution. *Accepted to Annals of the Institute of Statistical Mathematics.*

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

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

[21] Perlman, M. D. (1987). Group symmetry covariance models. *Stat. Sci.*, 2(4):421–425.

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

[23] Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, 52:447–458.

[24] Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol I (L. M. LeCam and J. Neyman, eds.)*, pp. 355–372.

[25] Rao, C. R. (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics*, 31(2):545–554.

[26] Reinsel, G. (1982). Multivariate repeated-measurement or Growth Curve models with multivariate random-effects covariance structure. *Journal of the American Statistical Association*, 77:190–195.

[27] Reinsel, G. (1984). Estimation and prediction in a multivariate random effects generalized linear model. *Journal of the American Statistical Association*, 79:406–414.

[28] Seid Hamid, J. and von Rosen, D. (2006). Residuals in the extended Growth Curve model. *Scandinavian Journal of Statistics*, 33(1):121–138.

[29] Srivastava, M. and Khatri, C. (1979). *An Introduction to Multivariate Statistics.* Elsevier North Holland, New York, USA.

[30] Srivastava, M. S. and von Rosen, D. (1999). Growth Curve model. *In: Multivariate Analysis, Design of Experiments, and Survey Sampling (S. Ghosh, ed.)*, Statistics: Textbooks and Monographs 159, Dekker, New York, pp. 547–578.

[31] von Rosen, D. (1995). Residuals in the Growth Curve model. *Ann. Inst. Stat. Math.*, 47(1):129–136.

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

[33] Ware, J. H. (1985). Linear models for the analysis of longitudinal studies linear models for the analysis of longitudinal studies. *The American Statistician*, 39(2):95–101.

[34] Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 17:257–28.

[35] Yokoyama, T. (1995). Statistical inference on some mixed MANOVA-GMANOVA models with random effects. *Hiroshima Math. J.*, 25(3):441–474.

[36] Yokoyama, T. (1996). Extended Growth Curve models with random-effects covariance structures. *Communications in statistics. Theory and methods*, 25(3):571–584.

[37] Yokoyama, T. (1997). Tests for a family of random-effects covariance structures in a multivariate Growth Curve model. *Journal of Statistical Planning and Inference*, 65(2):281–292.