Estimation in high-dimensional analysis and multivariate linear models

Tõnu Kollo, Tatjana von Rosen, and Dietrich von Rosen

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2008:12
ISSN 1651-8543

Estimation in high-dimensional analysis and multivariate linear models

Tõnu Kollo[1]
Institute of Mathematical Statistics, University of Tartu, Estonia

Tatjana von Rosen
Department of Statistics, Stockholm University, Sweden

Dietrich von Rosen
Centre of Biostochastics
Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract

When the number of variables compared with the number of observations is large this paper presents a new approach of estimating the parameters describing the mean structure in the Growth Curve model. An explicit estimator is obtained which is unbiased, consistent and asymptotically normally distributed. Its variance is also derived.

Keywords: Growth Curve model; high-dimensional analysis, Kolmogorov asymptotics, parameter estimation; $\frac{p}{n}$-asymptotics.

AMS subject classification: 62H12, 62H99.

[1]E-mail address to the correspondence author: Tonu.Kollo@ut.ee

1 Introduction

In this paper we derive estimators of the mean parameters in general multivariate linear models in a high-dimensional setting. The focus will be on the Growth Curve model introduced by Potthoff & Roy (1964) but also more general models, such as the Sum of Profiles model (Verbyla & Venables, 1988) or generalized Growth Curve models (see Kollo & von Rosen, 2005, Chapter 4), could have been considered.

If more variables (p) than observations (n) should be studied, most high-dimensional approaches construct some kind of summary statistic or study some function of the parameters. For example, test statistics, discriminant functions, spectral densities or different score functions are studied under different types of asymptotics. Sometimes the general case $p \to \infty, n \to \infty$ (see Srivastava & Du, 2008) is studied but more commonly the condition $\frac{p}{n} \to c > 0$, for some constant c, is added to the assumptions. The asymptotics under this condition will be referred to as $\frac{p}{n}$-asymptotics. In the literature this asymptotics is also called Kolmogorov asymptotics. In Takemura & Sheena (2005) and Sheena & Takemura (2008) an alternative approach in high-dimensional statistical analysis is considered.

Under the $\frac{p}{n}$-asymptotics the spectral density is the most widely studied statistic (e.g. see the books by Girko, 1990, and Bai & Silverstein, 2006). Concerning some inference problems on eigenvalues in a high-dimensional setting we refer to Schott (2006). Moreover, the literature concerning testing procedures for the mean is growing: e.g. see Läuter et al. (1996, 1998), Läuter et al. (2005), Srivastava & Fujikoshi, (2006), Schott (2007) and Srivastava & Du (2008), where also many other references are given. Tests about the covariance structure such as the sphericity test have, among others, been considered by Ledoit & Wolf (2002, 2004), and Srivastava (2005, 2006a). Several authors, when treating high-dimensional problems, have proposed to use shrinkage (regularized) estimators in order to obtain meaningful results (see Sancetta, 2008). It is interesting to observe that, in particular, high-dimensional discrimination analysis (supervised classification), has frequently been considered over the years (see Girko, 1990; Pavlenko & von Rosen, 2001, 2004; Srivastava, 2006b).

In general, a classical likelihood analysis is usually inappropriate to use, in high-dimensional analysis. Moreover, estimation is a harder problem than testing. For testing, one should construct a one-dimensional criterion which should reflect the null hypothesis and which is independent of nuisance parameters. For estimation we often have many parameters which should be estimated and sometimes the number of nuisance parameters increases to infinity which makes inference much more complicated.

The Growth Curve model is defined as follows:

$$\boldsymbol{X} = \boldsymbol{ABC} + \boldsymbol{E}, \tag{1.1}$$

where $\boldsymbol{X}$: $p \times n$ is the data matrix, $\boldsymbol{B}$: $q \times k$ is a matrix of unknown parameters, $\boldsymbol{A}$: $p \times q$ and $\boldsymbol{C}$: $k \times n$ are known design matrices, $\boldsymbol{E} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$ is a normally distributed error matrix, where $\boldsymbol{e}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ are i.i.d. and $\boldsymbol{\Sigma}$ is an unknown positive definite parameter matrix. The model in (1.1) belongs to the curved exponential family and has a relatively long history. For reviews of the model see Woolson & Leeper (1980), von Rosen (1991) or Srivastava & von Rosen (1999). Kshirsagar & Smith (1995) have written a book on this model and for a recent contribution see Kollo & von Rosen (2005, Chapter 4), where the model and some extensions are presented. If $\boldsymbol{A} = \boldsymbol{I}_p$, we have the standard MANOVA model. However, in the context of the present paper the MANOVA model is not appropriate to study since the number of mean parameters turns to infinity as $p \to \infty$. For the Growth Curve model the mean parameter space is independent of p and n whereas the covariance matrix $\boldsymbol{\Sigma}$ increases in size with p.

Sufficient statistics for the Growth Curve model are

$$\boldsymbol{S} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{CC}')^{-}\boldsymbol{C})\boldsymbol{X}' \tag{1.2}$$

and

$$\boldsymbol{XC}'(\boldsymbol{CC}')^{-}\boldsymbol{C}.$$

Due to the normality assumption, i.e. since the distribution is symmetric around the mean, in order to estimate the mean parameters it is natural to consider

$$\begin{aligned}
&\tfrac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{ABC})(\boldsymbol{X} - \boldsymbol{ABC})'\} \\
&\quad = \tfrac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{XC}'(\boldsymbol{CC}')^{-}\boldsymbol{C} - \boldsymbol{ABC})(\boldsymbol{XC}'(\boldsymbol{CC}')^{-}\boldsymbol{C} - \boldsymbol{ABC})'\} \\
&\qquad + \tfrac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\}.
\end{aligned}$$

The factor $1/p$ is used to handle the increase in size of $tr(\bullet)$ when $p \to \infty$, i.e. the trace functions have been averaged.

The results in the paper are presented in three sections. Section 2 presents the asymptotic distribution for

$$\boldsymbol{T}_1 = \tfrac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\} \tag{1.3}$$

as well as

$$\boldsymbol{T}_2 = \frac{1}{p} tr\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{C})(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{C})'\}, \quad (1.4)$$

and these results are utilized in Section 3 where estimators of $\boldsymbol{B}$ are obtained. Finally, in Section 4 the mean and dispersion matrix of the estimator of $\boldsymbol{B}$ as well as its asymptotic distribution will be obtained which turns out to be normal.

2 Asymptotics of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$

In high-dimensional analysis, one often considers $\frac{1}{p}tr(\boldsymbol{S})$ or $\frac{1}{p}tr(\boldsymbol{S}^2)$ (e.g. see Ledoit & Wolf, 2002 or Srivastava, 2005) but then the asymptotics depends on $\boldsymbol{\Sigma}$. However, because of normality $\boldsymbol{S}$ is Wishart distributed, ($\boldsymbol{S} \sim W_p(\boldsymbol{\Sigma}, n')$, $n' = n - r(\boldsymbol{C})$), and from Kollo & von Rosen (2005, Corollary 2.4.2.2, p. 238) it follows that $\boldsymbol{T}_1$, given in (1.3), is chi-square distributed with n' degrees of freedom. Hence, the characteristic function $\varphi_{\boldsymbol{T}_1}(t)$ equals

$$\varphi_{\boldsymbol{T}_1}(t) = (1 - i\, t\tfrac{2}{p})^{-pn'/2},$$

where i is the imaginary unit. If taking the logarithm of the characteristic function and expanding it as a power series in p and n, it follows that

$$\begin{aligned}
\ln \varphi_{\boldsymbol{T}_1}(t) &= -pn'/2 \ln(1 - i\, t\tfrac{2}{p}) = \frac{pn'}{2} \sum_{j=1}^{\infty} \left(\tfrac{2}{p}\right)^j \tfrac{1}{j} i^k\, t^j \\
&= i\, tn' - \frac{n'p}{2}\frac{2^2}{p^2}\frac{1}{2}t^2 + \frac{n'p}{2}\frac{2^3}{p^3} i^3 \frac{1}{3} t^3 + \cdots \\
&\approx i\, tn' - \frac{n'p}{2}\frac{2^2}{p^2}\frac{1}{2}t^2.
\end{aligned}$$

This implies that under $\frac{p}{n}$-asymptotics

$$\frac{\frac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\} - n'}{\sqrt{\frac{n'}{p}}} \overset{\text{a}}{\sim} N(\mathbf{0}, 2), \quad (2.1)$$

where $\overset{\text{a}}{\sim}$ means "asymptotically distributed as".

Representing $\boldsymbol{T}_2$, given in (1.4), as $\boldsymbol{T}_2 = \frac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{V}'\}$, where

$$\boldsymbol{V} = \boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{C} \quad (2.2)$$

with $\boldsymbol{VV}' \sim W_p(\boldsymbol{\Sigma}, r)$, $r = r(\boldsymbol{C})$, it is observed that since in this case the number of degrees of freedom of the distribution is fixed, the speed of convergence is slower than the speed of convergence of $\boldsymbol{T}_1$.

The logarithm of the characteristic function of $\boldsymbol{T}_2$ equals

$$\ln \varphi_{\boldsymbol{T}_2}(t) = -\tfrac{rp}{2}\ln(1 - i\,t\tfrac{2}{\sqrt{p}}).$$

Thus,

$$\begin{aligned}\ln \varphi_{\boldsymbol{T}_2}(t) &= -\tfrac{rp}{2}\ln(1 - i\,t\tfrac{2}{\sqrt{p}}) = \tfrac{rp}{2}\sum_{j=1}^{\infty} p^{-\frac{j}{2}} 2^j \tfrac{1}{j}\, i^j\, t^j \\ &= i\,tr\sqrt{p} - rt^2 + i^3\, t^3 rp^{-\frac{1}{2}}\tfrac{1}{3} + \cdots\end{aligned}$$

and

$$\frac{\frac{1}{\sqrt{p}}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{VV}'\} - r\sqrt{p}}{\sqrt{r}} \overset{\mathrm{a}}{\sim} N(\boldsymbol{0}, 2). \tag{2.3}$$

Hence, the following theorem has been verified:

Theorem 2.1. *Under $\frac{p}{n}$-asymptotics (2.1) holds, and for any n and $p \to \infty$, (2.3) holds.*

Since $\boldsymbol{S}$ and $\boldsymbol{XC}'(\boldsymbol{CC}')^{-}\boldsymbol{C}$ are sufficient statistics we may note that (2.1) and (2.3) include the relevant information for estimating the mean parameters of the Growth Curve model. Thus, based on (2.1) and (2.3) an asymptotic likelihood approach may be presented.

3 Estimation of $\boldsymbol{B}$ in the Growth Curve model

From the previous section it follows that an asymptotic likelihood function for (1.1) based on $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ is proportional to

$$exp\{-\tfrac{1}{4}(pn'(\tfrac{1}{pn'}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\} - 1)^2)\}exp\{-\tfrac{1}{4}(pr(\tfrac{1}{pr}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{VV}'\} - 1)^2)\}. \tag{3.1}$$

Following the likelihood principle we maximize this function. However, since $\boldsymbol{\Sigma}$ is assumed to be of full rank and unstructured, and $\boldsymbol{S}$ may be singular if $\frac{p}{n} \to c > 1$, from (3.1) it is impossible to get appropriate estimators for all elements of $\boldsymbol{\Sigma}$ and $\boldsymbol{B}$. However, we are only interested in the estimation of $\boldsymbol{B}$ and its variance. Therefore we will study the two terms in (3.1) separately, and suggest an approach similar to the restricted maximum likelihood approach. Let us start with the first term, i.e.

$$(\frac{1}{pn'}\,tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\}-1)^2. \tag{3.2}$$

By choosing

$$\widehat{\boldsymbol{\Sigma}}^{-1}=n'\boldsymbol{S}^{-} \tag{3.3}$$

the expression in (3.2) equals 0, where $\boldsymbol{S}^{-}$ denotes an arbitrary g-inverse of $\boldsymbol{S}$. The main drawback with this estimator is that it is not unique. However, it is natural to suppose that $r(\boldsymbol{S}^{-})=r(\boldsymbol{S})$ which implies that $\boldsymbol{S}^{-}$ is a reflexive g-inverse, i.e. $\boldsymbol{S}^{-}\boldsymbol{S}\boldsymbol{S}^{-}=\boldsymbol{S}^{-}$ holds besides the defining condition $\boldsymbol{S}\boldsymbol{S}^{-}\boldsymbol{S}=\boldsymbol{S}$. If $r(\boldsymbol{S})<r(\boldsymbol{S}^{-})$ ($r(\boldsymbol{S})>r(\boldsymbol{S}^{-})$ never holds) it means that we can estimate more elements in $\boldsymbol{\Sigma}^{-1}$ than in $\boldsymbol{\Sigma}$ which does not make sense. It follows from Khatri & Mitra (1976) that there is only one g-inverse which is positive semi definite (p.s.d.) and reflexive, i.e. we may choose the g-inverse to be the Moore-Penrose g-inverse which will be denoted $\boldsymbol{S}^{+}$. In the next we replace $\boldsymbol{\Sigma}^{-1}$ by $n'\boldsymbol{S}^{+}$ in the second exponent in (3.1) and thus have to minimize

$$(\frac{n'}{pr}\,tr\{\boldsymbol{S}^{+}\boldsymbol{V}\boldsymbol{V}'\}-1)^2.$$

Differentiating this expression and taking into account (2.2) we get with respect to $\boldsymbol{B}$ the equation

$$(\frac{n'}{pr}tr\boldsymbol{S}^{+}\boldsymbol{V}\boldsymbol{V}'-1)\boldsymbol{A}'\boldsymbol{S}^{+}(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}-\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})\boldsymbol{C}'=\boldsymbol{0}.$$

With probability 1, the expression $(\frac{n'}{pr}tr\boldsymbol{S}^{+}\boldsymbol{V}\boldsymbol{V}'-1)$ differs from 0, and thus the following linear equation in $\boldsymbol{B}$ emerge:

$$\boldsymbol{A}'\boldsymbol{S}^{+}(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}-\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})\boldsymbol{C}'=\boldsymbol{0}. \tag{3.4}$$

This equation is consistent if the column space relation $C(\boldsymbol{A}'\boldsymbol{S}^{+}) = C(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})$ holds which is true since $\boldsymbol{S}^{+}$ is p.s.d. Hence, for $\boldsymbol{B}$ in (3.4) the general solution can be written (see Kollo & von Rosen, 2005; Theorem 1.3.4)

$$\widehat{\boldsymbol{B}}=(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}+(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{o}\boldsymbol{Z}_1+\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A}\boldsymbol{Z}_2\boldsymbol{C}^{o'}, \tag{3.5}$$

where $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are arbitrary matrices, and $(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{o}$ and $\boldsymbol{C}^{o}$ are any arbitrary matrices spanning the orthogonal complement to $C(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})$ and $C(\boldsymbol{C})$, respectively. From here we obtain the following result:

Theorem 3.1. *The estimator $\widehat{\boldsymbol{B}}$, given in (3.5), is unique and with probability 1 equals*

$$\widehat{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}, \tag{3.6}$$

if and only if $r(\boldsymbol{A}) = q < \min(p, n')$, $r(\boldsymbol{C}) = k$ *and* $C(\boldsymbol{A}) \cap C(\boldsymbol{S})^{\perp} = \{\boldsymbol{0}\}$, *where* $\boldsymbol{S}$ *is given in (1.2).*

Proof. We only show that $C(\boldsymbol{A})\cap C(\boldsymbol{S})^{\perp}=\{\boldsymbol{0}\}$ is equivalent to $(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{o}=\boldsymbol{0}$:

$$r(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A}) = r(\boldsymbol{A}'\boldsymbol{S}^{+}) = r(\boldsymbol{A} : (\boldsymbol{S}^{+})^{o}) - r((\boldsymbol{S}^{+})^{o}) = r(\boldsymbol{A}) \tag{3.7}$$

with probability 1, since $C(\boldsymbol{A}) \cap C(\boldsymbol{S}^{+})^{\perp} = C(\boldsymbol{A}) \cap C(\boldsymbol{S})^{\perp} = \{\boldsymbol{0}\}$ with probability 1. The relation $C(\boldsymbol{S}^{+}) = C(\boldsymbol{S})$ holds because $\boldsymbol{S}^{+}\boldsymbol{S} = \boldsymbol{S}\boldsymbol{S}^{+}$. $\square$

If $\boldsymbol{S}$ is of full rank ($p \leq n'$), $\widehat{\boldsymbol{B}}$ in (3.6) is identical to the maximum likelihood estimator (see Kollo & von Rosen, 2005; p. 360).

4 Properties of $\widehat{\boldsymbol{B}}$

In this section it is assumed that $\widehat{\boldsymbol{B}}$ is unique. Since $\boldsymbol{X}\boldsymbol{C}$ and $\boldsymbol{S}$ are independently distributed

$$\begin{aligned} E[\widehat{\boldsymbol{B}}] &= E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}]E[\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}] \\ &= E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}]\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}. \end{aligned}$$

The dispersion matrix

$$D[\widehat{\boldsymbol{B}}] = E[vec(\widehat{\boldsymbol{B}} - \boldsymbol{B})vec'(\widehat{\boldsymbol{B}} - \boldsymbol{B})],$$

where $vec(\cdot)$ is the usual vec-operator, is much more complicated to obtain. Since $D[\boldsymbol{X}] = \boldsymbol{I} \otimes \boldsymbol{\Sigma}$,

$$D[\widehat{\boldsymbol{B}}] = (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}].$$

In the next the expectation will be written in a canonical form. There exist always an orthogonal matrix $\boldsymbol{\Gamma}$ and a non-singular matrix $\boldsymbol{T}$ such that

$$\boldsymbol{A}' = \boldsymbol{T}(\boldsymbol{I}_q : \boldsymbol{0})\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{\frac{1}{2}}. \tag{4.1}$$

Let

$$\boldsymbol{U} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}\boldsymbol{\Sigma}^{-\frac{1}{2}} \sim W_p(\boldsymbol{I}_p, n'),$$

and thus

$$
\begin{aligned}
&E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}] \\
&\quad = (\boldsymbol{T}')^{-1}E[\left((\boldsymbol{I}_q:\boldsymbol{0})\boldsymbol{U}^{+}\left(\begin{smallmatrix}\boldsymbol{I}_q\\ \boldsymbol{0}\end{smallmatrix}\right)\right)^{-1}(\boldsymbol{I}_q:\boldsymbol{0})\boldsymbol{U}^{+}\boldsymbol{U}^{+}\left(\begin{smallmatrix}\boldsymbol{I}_q\\ \boldsymbol{0}\end{smallmatrix}\right) \\
&\qquad \times \left((\boldsymbol{I}_q:\boldsymbol{0})\boldsymbol{U}^{+}\left(\begin{smallmatrix}\boldsymbol{I}_q\\ \boldsymbol{0}\end{smallmatrix}\right)\right)^{-1}]\boldsymbol{T}^{-1}.
\end{aligned} \tag{4.2}
$$

Since $\widehat{\boldsymbol{B}}$ does not depend on the order of observations, we may suppose that the observations are ordered so that

$$
\boldsymbol{U}^{+} = \begin{pmatrix}\boldsymbol{W}^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}, \qquad \boldsymbol{W} \sim W_s(\boldsymbol{I}_s, n'), \quad s = \min(p, n').
$$

By assumption $q \leq s$, and we may partition $\boldsymbol{W}$ as

$$
\boldsymbol{W} = \begin{pmatrix}\boldsymbol{W}_{11} & \boldsymbol{W}_{12}\\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22}\end{pmatrix}, \qquad \begin{pmatrix} q\times q & q\times(s-q)\\ (s-q)\times q & (s-q)\times(s-q)\end{pmatrix}
$$

and correspondingly $\boldsymbol{W}^{-1}$ as

$$
\boldsymbol{W}^{-1} = \begin{pmatrix}\boldsymbol{W}^{11} & \boldsymbol{W}^{12}\\ \boldsymbol{W}^{21} & \boldsymbol{W}^{22}\end{pmatrix}, \qquad \begin{pmatrix} q\times q & q\times(s-q)\\ (s-q)\times q & (s-q)\times(s-q)\end{pmatrix}.
$$

Thus (4.2) equals

$$
\begin{aligned}
&(\boldsymbol{T}')^{-1}\boldsymbol{T}^{-1} + (\boldsymbol{T}')^{-1}E[\boldsymbol{W}^{11})^{-1}\boldsymbol{W}^{12}\boldsymbol{W}^{12}\boldsymbol{W}^{11})^{-1}](\boldsymbol{T}')^{-1} \\
&\quad = (\boldsymbol{T}')^{-1}\boldsymbol{T}^{-1} + (\boldsymbol{T}')^{-1}E[\boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{21}](\boldsymbol{T}')^{-1}.
\end{aligned}
$$

From calculations in Kollo & von Rosen (2005, p. 413) it follows that

$$
D[\widehat{\boldsymbol{B}}] = \frac{n'-1}{n'+q-s-1}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1},
$$

since $(\boldsymbol{T}')^{-1}\boldsymbol{T}^{-1} = (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}$. Note that when $p/n > 1$

$$
D[\widehat{\boldsymbol{B}}] = \frac{n'-1}{q-1}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}. \tag{4.3}
$$

Moreover, $(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1} \sim W_q((\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}, n'+q-s)$ and therefore an unbiased estimator of $(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}$ is given by

$$
\frac{1}{n'+q-s}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}.
$$

Thus, an unbiased estimator of the $D[\widehat{\boldsymbol{B}}]$ is

$$\widehat{D[\widehat{\boldsymbol{B}}]} = \frac{n'-1}{(q-1)(n'+q-s)}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}. \tag{4.4}$$

Finally we will find the asymptotic distribution of $\widehat{\boldsymbol{B}}$. Using (4.1)

$$\widehat{\boldsymbol{B}} = (\boldsymbol{T}')^{-1}\left((\boldsymbol{I}_q:\boldsymbol{0})\boldsymbol{U}^{+}\left(\begin{smallmatrix}\boldsymbol{I}_q\\ \boldsymbol{0}\end{smallmatrix}\right)\right)^{-1}(\boldsymbol{I}_q:\boldsymbol{0})\boldsymbol{U}^{+}\boldsymbol{Y}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1},$$

where $\boldsymbol{Y} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p, \boldsymbol{I}_n)$. Since $\frac{1}{n'}\boldsymbol{W} \xrightarrow{\mathrm{p}} \boldsymbol{I}_s$ ($\xrightarrow{\mathrm{p}}$ denotes convergence in probability)

$$n'\boldsymbol{U}^{+} \xrightarrow{\mathrm{p}} \begin{pmatrix}\boldsymbol{I}_s & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}.\end{pmatrix}$$

Thus,

$$\widehat{\boldsymbol{B}} \quad - \quad (\boldsymbol{T}')^{-1}\left((\boldsymbol{I}_q:\boldsymbol{0})\boldsymbol{I}_s\left(\begin{smallmatrix}\boldsymbol{I}_q\\ \boldsymbol{0}\end{smallmatrix}\right)\right)^{-1}(\boldsymbol{I}_q:\boldsymbol{0})\boldsymbol{I}_s\boldsymbol{Y}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} \xrightarrow{\mathrm{p}} \boldsymbol{0}$$

and using $(\boldsymbol{T}')^{-1}(\boldsymbol{I}_q:\boldsymbol{0})\boldsymbol{Y} = (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$ it follows that

$$\widehat{\boldsymbol{B}} - (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} \xrightarrow{\mathrm{p}} \boldsymbol{0}.$$

This means that $\widehat{\boldsymbol{B}}$ is asymptotically normally distributed as well as it shows that the proposed estimation procedure leads to a natural estimator because when $\boldsymbol{\Sigma}$ is known, we would have obtained the best (in the sense of smallest variance) estimator $\widehat{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}$. Moreover, under mild conditions on $\boldsymbol{C}$, i.e. the columns of $\boldsymbol{C}$ should not differ too much, the matrix $\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}$ converges in probability to its mean $\boldsymbol{A}\boldsymbol{B}$. This implies that $\widehat{\boldsymbol{B}}$ is a consistent estimator of $\boldsymbol{B}$ even when $p > n$. The results of this section are summarized in the next theorem

Theorem 4.1. *Let $\widehat{\boldsymbol{B}}$ be given by (3.6). Then,*

(i) *$\widehat{\boldsymbol{B}}$ is an unbiased and consistent (under some conditions on $\boldsymbol{C}$) estimator of $\boldsymbol{B}$;*

(ii) *the dispersion matrix of $\widehat{\boldsymbol{B}}$, $D[\widehat{\boldsymbol{B}}]$, is given by (4.3);*

(iii) *the estimated dispersion matrix $\widehat{D[\widehat{\boldsymbol{B}}]}$ is given by (4.4);*

(iv) *$\widehat{\boldsymbol{B}}$ is asymptotically equivalent to a random variable with a matrix normal distribution.*

Remark: The results show that if $n \to \infty$ or the $\frac{p}{n}$-asymptotics holds the estimator of the mean parameter, proposed by the approach of this paper, behaves in the same way, i.e the large number of dispersion parameters does not seriously influence the estimator of $\boldsymbol{B}$.

Acknowledgement

Tõnu Kollo is grateful to the Estonian Research Foundation for support through the grant GMTMS7435. Tatjana von Rosen's research is supported by the Estonian Research Foundation, grant GMTMS6702.

References

Bai, Z. and Silverstein, J.W. (2006). *Spectral analysis of large dimensional random matrices*, Mathematics Monograph Series, **2**, Beijing, Science Press.

Fang, K.-T., Wang, S.G., and von Rosen, D. (2006). Restricted Expected Multivariate Least Squares. *J. Multivar. Anal.*, **97**, 619–632.

Girko, V.L. (1990). *Theory of random determinants.* Translated from the Russian. Mathematics and its Applications (Soviet Series), **45**, Dordrecht, Kluwer Academic Publishers Group.

Khatri, C.G. and Mitra, S.K (1976). Hermitian and nonnegative definite solutions of linear matrix equations. *SIAM J. Appl. Math.* **31**, 579–585.

Kollo, T. and von Rosen, D. (2005). *Advanced multivariate statistics with matrices. Mathematics and Its Applications*, (New York), **579**, Dordrecht, Springer.

Kshirsagar, A.M. and Smith, W.B. (1995). *Growth curves.* Statistics: Textbooks and Monographs, **145**, New York, Marcel Dekker, Inc.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30**, 1081–1102.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* **88**, 365–411.

Läuter, J., Glimm, E., and Eszlinger, M. (2005). Search for relevant sets of variables in a high-dimensional setup keeping the familywise error rate, *Statist. Neerlandica* **59**, 298–312.

Läuter, J., Glimm, E., and Kropf, S. (1996). New multivariate tests for data with an inherent structure, *Biometrical J.* **38**, 5–23. Erratum: *Biometrical J.* **40**, (1996): 1015.

Läuter, J., Glimm, E., and Kropf, S. (1998). Multivariate tests based on left-spherically distributed linear scores, *Ann. of Statist.* **26**, 1972–1988. Erratum: *Ann. of Statist.* **27**, 1441.

Pavlenko, T. and von Rosen, D. (2001). Estimating the discriminant function when the number of variables is large, *Statistics* **36**, 191–213.

Pavlenko, T. and von Rosen, D. (2004). On the optimal weighting of high-dimensional Bayesian networks, *Adv. Appl. Statist.* **4**, 357–377.

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

von Rosen, D. (1991). The Growth Curve model: A review. *Commun. Statist.-Theor. Meth.* **20**, 2791–2822.

Sancetta, A. (2008). Sample covariance shrinkage for high dimensional dependent data. *J. Multivariate Anal.* **99**, 949–967.

Schott, J.R. (2006). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *J. Multivariate Anal.* **97**, 827–843.

Schott, J.R. (2007). Some high-dimensional tests for a one-way MANOVA. *J. Multivariate Anal.* **98**, 1825–1839.

Sheena, Y. and Takemura, A. (2008). Asymptotic distribution of Wishart matrix for block-wise dispersion of population eigenvalues. *J. Multivariate Anal.* **99**, 751–775.

Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high dimensional data, *J. Japan Statist. Soc.* **35**, 251–272.

Srivastava, M.S. (2006a). Some tests criteria for the covariance matrix with fewer observations than the dimension. *Acta Comment. Univ. Tartu. Math.* **10**, 77–93.

Srivastava, M.S. (2006b). Minimum distance classification rules for high dimensional data, *J. Multivariate Anal.* **97**, 2057–2070.

Srivastava, M.S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99**, 386–402.

Srivastava, M.S. and Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *J. Multivariate Anal.* **97**, 1927–1940.

Srivastava, M.S. and von Rosen, D. (1999). Growth curve models. *Multivariate analysis, design of experiments, and survey sampling*, 547–578. Textbooks Monogr., **159**, New York, Marcel Dekker, Inc.

Takemura, A. and Sheena, Y. (2005). Distribution of eigenvalues and eigenvectors of Wishart matrix when the eigenvalues are infinitely dispersed and its application to minimax estimation of covariance matrix. *J. Multivariate Anal.* **94**, 271–299.

Verbyla, A.P. and Venables, W.N. (1988). An extension of the growth curve model. *Biometrika* **75**, 129–138.

Woolson, R.F. and Leeper, J.D. (1980). Growth curve analysis of complete and incomplete longitudinal data. *Comm. Statist. A—Theory Methods* **9**, 1491–1513.