Department of Mathematics

Bilinear regression with rank restrictions on the mean and the dispersion matrix

Tatjana von Rosen and Dietrich von Rosen

LiTH-MAT-R--2018/10--SE
Bilinear regression with rank restrictions on the mean and the dispersion matrix

Tatjana von Rosen† and Dietrich von Rosen‡,*

†Stockholm University, Sweden
‡Swedish University of Agricultural Sciences, Sweden
*Linköping University, Sweden

August 10, 2018

Abstract

A bilinear regression model with rank restrictions imposed on the mean-parameter matrix and on the dispersion matrix is studied. Maximum likelihood inspired estimates are derived. The approach generalizes classical reduced rank regression analysis and principal component analysis. It is shown via a simulation study and a real example that even for small dimensions the method works as well as reduced rank regression analysis whereas the approach in this article also can be used when the dimension is large.

Keywords: Growth curve model; likelihood based estimates; rank restrictions; singular dispersion matrix.

1 Introduction

The rapid development of technology enables the use of various types of sensors together with advanced computer facilities in modern empirical studies which often results in data with quite complex structures. Well established models do not exist in this case since one only has a vague idea what should be included in a model adequately describing the data. For example, a huge number of chemical soil characteristics can be observed which, however, are all governed by a few latent processes, or brain activity can be measured via EEG-signals from a large number of electrodes which are placed on the scalp though there are a few latent processes which direct the signals. At the end of this section another example is presented in some detail.

Our philosophy is that latent processes should be looked upon as non-observable processes on which we do not want to put a distribution. The concept of latent variables appears in the literature in different contexts. Sometimes latent variables are motivating mixed linear models. Other examples where latent
variable are mentioned are finite mixtures of regression models and the terms latent class models and latent factor models are also often used. In this article, we relate latent variables to rank restrictions on parameter matrices which according to data analysis seems to be a rewarding strategy.

Statistical modelling under a normality assumption often means joint modelling of the mean and dispersion. The available literature comprises results on using latent variables to model the mean (reduced rank regression, cointegration; e.g. see Anderson, 1951; Johansen, 1991; Reinsel & Velu, 1998) or to model the dispersion (principal component analysis, among other methods; e.g. see Jolliffe, 2002). To our knowledge there is, however, no results concerning jointly modelling multiprocesses, i.e. the case when simultaneously some latent processes are affecting the mean whereas other latent processes steer the dispersion.

In this article the focus is on the bilinear regression model (BRM) which is often called the growth curve model or generalized multivariate analysis model (GMANOVA). The model was introduced by Potthoff & Roy (1964) although other authors had earlier considered similar models. For general references about the model, e.g. see Woolson & Leeper (1980), von Rosen (1991), Kshirsagar & Smith (1995) or von Rosen (2018).

As has been mentioned above the inference in this article is based on the fact that latent processes can be described through rank restrictions on parameters. For references to results and applications of rank restrictions on the mean, including the mean structure of the BRM, we refer the readers to the book by Reinsel & Velu (1998), Albert & Kshirsagar (1993), Reinsel & Velu (2003) or to a recent work by von Rosen & von Rosen (2017). Today high-dimensional statistical analysis is rapidly developing. A pure high-dimensional perspective on rank restrictions on the mean is presented in an interesting article by Kar&gin (2015) and Chen & Huang (2012) combined reduced rank regression with a variable selection method. Further, there exists an extensive literature on linear models when the dispersion matrix is proportional to a known positive semi-definite matrix (the Gauss-Markov model), e.g. see Rao (1973a, 1979), Nordström (1984), Baksalary et al. (1992). However, when the dispersion matrix is unknown only a few references exist (see Wong & Cheng 2001, Srivastava & von Rosen 2002).

The aim of this work is to find estimates of the mean and dispersion parameters under simultaneous rank restrictions on the matrix of regression coefficients and the dispersion matrix.

Grizzle & Allen (1969) published a data set which later was analyzed by Albert & Kshirsagar (1993) who used a reduced rank restriction on the mean parameter in order to discriminate groups of independent observations which is linked to canonical correlation analysis (see Tso, 1981; Johansen, 1988). The data set consists of 36 dogs randomly assigned to four groups. For each dog measurements of potassium were obtained at seven equally spaced time points and then a reduced rank growth curve model was applied. As Grizzle & Allen (1969), the authors used a cubic regression model and tested what rank on the mean parameters would be naturally to assume. Albert & Kshirsagar (1993) suggested the rank to be one. Moreover, Albert & Kshirsagar (1993) aimed
to reduce the dimensionality in data via rank restrictions on the mean (similar to canonical correlation analysis) whereas in this article we also employ ideas from principal component analysis and additionally reduce the data via rank restrictions on the dispersion matrix.

The following notation will be used. Bold upper cases denote matrices: $C(A)$ is the column vector space generated by the columns of $A$ and $C(A)^+$ denotes its orthogonal complement; $A^o$ denotes any matrix of full rank which generates $C(A)^+$. The orthogonal projector on $C(A)$ is denoted $P_A$ and equals $P_A = A(A'A)^{-1}A'$, where "$ - " denotes an arbitrary generalized inverse (g-inverse). Moreover, we will often write $(Q)(Q)'$ instead of $(Q)(Q)'$, where $Q$ represents any matrix expression. Additionally necessary notation will be introduced in the following sections.

2 The model

Consider the $BRM$ with normally distributed error which without loss of generality can be written as

$$X = ABC + E,$$  \hspace{1cm} (1)

where $X$: $p \times n$ consists of independent response vectors, $A$: $p \times q$ is a known within-individuals design matrix, $C$: $k \times n$ is a known between-individuals design matrix, $E \sim N_{p,n}(0, \Sigma, I_n)$, i.e. $E$ is matrix normally distributed (see Ohlson et al., 2013), the mean matrix $B$: $q \times k$ and the dispersion matrix $\Sigma$: $p \times p$, are unknown parameter matrices. A distinguishing characteristic of the $BRM$ studied here are the following reduced-rank restrictions on the model parameters

$$r(B) = f < \min(q, k),$$  \hspace{1cm} (2)

$$r(\Sigma) = r < p, r \geq f.$$  \hspace{1cm} (3)

Thus, $\Sigma$ is positive semi-definite instead of being positive definite, which is a usual assumption. As mentioned in the introduction, the main aim of this article is to estimate the parameters $B$ and $\Sigma$ in (1) when the rank restrictions (2) and (3) hold so that the model can be used for prediction. Note that in the majority of published articles $B$ is interpretable or one wants to find interpretable scores such as in Albert & Kshirsagar (1993) which can be difficult in practice. However, it is possible that $B$ constitutes an object with rank restrictions corresponding to latent processes where we really do not know the implications of these processes. The latent process can have a clear interpretation of the “between individuals” structure of the model but can also be interpretable in terms of the “within individuals” structure. In both cases (precise or vague interpretation of $B$) the mathematics will be the same when deriving estimators, i.e. the only thing which matters is the interpretation of the estimators. In this article we focus on statistical inference rather than interpreting the results which can depend on the data and purpose of the empirical study.
3 Preliminaries

We will present three lemmas which are used to derive the estimates in model (1). The first lemma is essentially a restatement of Theorem 1.5.3 in Srivastava & Khatri (1979).

**Lemma 3.1.** For any $\Theta: q \times k$, with rank restrictions, say $r(\Theta) = f$, there exist $\Theta_1: q \times f$, $\Theta_2: f \times k$, $r(\Theta_i) = f$, $i = 1, 2$, such that $\Theta = \Theta_1 \Theta_2$.

Note that $\Theta_1$ and $\Theta_2$ together include more unknown elements than $\Theta$. Thus, the estimation strategy is to consider a larger parameter space than the original one, i.e. some type of imbedding takes place. On the other hand, if $r(\Theta) = f$ there exist $k - f$ linear combinations of $\Theta$ which equal 0, say $L \Theta = 0$, where $L$ is of size $(q - f) \times q$, or $L \Theta' = 0$, where $L$ is of size $(k - f) \times k$. Both conditions imply, since $L$ is unknown, that $\Theta = \Theta_1 \Theta_2$, and thus the factorization of $\Theta$ is a natural approach to apply. However, what type of restrictions exist can be important for the interpretation of the analysis. The next lemma is also well known.

**Lemma 3.2.** (spectral decomposition) Let $\Sigma: p \times p$ be positive semi-definite of rank $r(\Sigma) = r$. Then there exists a semi-orthogonal matrix $\Gamma$, i.e. $\Gamma' \Gamma = I$, and a diagonal matrix $\Delta$: $r \times r$ with positive diagonal elements, such that $\Sigma = \Gamma \Delta \Gamma'$.

It is well known that $C(\Sigma) \subseteq C(\Sigma)$, when $S \sim W_p(\Sigma, n)$. Furthermore, if $n \geq r(\Sigma)$, then $C(S) = C(\Sigma)$ holds with probability 1 (e.g. see Srivastava & von Rosen, 2002, Lemma 3.1). In this article, $S = X(I - P_{C'})X' \sim W_p(\Sigma, n - r(C))$. Let $S = HDH'$ where $H$ consists of eigenvectors of $S$ and $D$ is a diagonal matrix with eigenvalues of $S$ on its diagonal. If we have observations $X_o$, i.e. realizations of $X$, we can calculate $H_o$ which is the observed $H$, where however eigenvectors corresponding to the $p - r$ smallest eigenvalues of $X_o(I - P_{C'})X_o'$ have been removed, and $D_o$ is the diagonal matrix consisting of the observed eigenvalues of $X_o(I - P_{C'})X_o'$, where the $p - r$ smallest eigenvalues have been removed. Then

$$S_o = H_o D_o H_o'$$

(4)

does not represent the observed $S$ but a quantity corresponding to $S$, where the $p - r$ smallest eigenvalues have been put to 0. The way of defining $S_o$ is fundamental for the results of this article.

Furthermore, throughout the paper we make the following two crucial assumptions.

**Assumption 3.1.** For the model in (1) it is assumed that $C(X_o) \subseteq C(A) + C(\Sigma)$.

**Assumption 3.2.** For the model in (1), if $n \geq r + r(C)$, it is assumed that $C(S_o) = C(\Sigma)$.

Removing the $p - r$ smallest eigenvalues in (4) is not obvious. However, Assumption 3.2, i.e. $C(S_o) = C(\Sigma)$, is motivated by $C(S) = C(\Sigma)$ which as noted
before holds with probability 1. Therefore, under Assumption 3.2,

$$C(H_o) = C(S_o) = C(\Sigma) = C(\Gamma).$$  \hspace{1cm} (5)

It is worth noting the equality $C(H_o) = C(S_o)$. Instead of Assumption 3.1 we can use $C(X_o) \subseteq C(A) + C(S)$, and using Assumption 3.2, $C(S)$ can be replaced by $C(S_o)$. In reality it means that before analysing data, $X_o$ has to be preprocessed, i.e. $X_o$ should be projected on $C(A) + C(S_o)$, and then Assumption 3.1 holds.

**Lemma 3.3.** Let $\Gamma$ and $H_o$, defined in Lemma 3.2 and (4), respectively, be the matrices of eigenvectors corresponding to $\Sigma$ and $S_o$, respectively, and suppose $n \geq r + r(C)$. Then, for some orthogonal matrix $Q$, under Assumption 3.2,

$$\Gamma = H_oQ.$$

**Proof.** The proof follows from (5) and a few calculations: $I_r = \Gamma'\Gamma = Q' H_o' H_o Q = Q'Q$.

**Corollary 3.4.** Let $\Gamma$ and $S_o$ be defined as in Lemma 3.3, $X_o$ is the realization of $X$ in (1) where also $C$ is given. Then, under Assumption 3.2,

$$\Gamma' X_o (I - P_{C'}) X_o' \Gamma - \Gamma' S_o \Gamma = 0.$$

**Proof.** According to Lemma (3.3) $\Gamma = H_oQ$ and thus

$$Q' H_o' (X_o (I - P_{C'}) X_o' - S_o) H_o Q = 0.$$

**Corollary 3.5.** Let $\Gamma$ and $S_o$ be defined as in Lemma 3.3. Then

$$\Gamma (\Gamma' S_o \Gamma)^{-1} \Gamma' = S_o^+,$$

where “+” denotes the Moore-Penrose generalized inverse (e.g. see Rao, 1973b; p. 26).

**Proof.** Since $\Gamma = H_oQ$ for some orthogonal $Q$, the Moore-Penrose generalized inverse is established by verification of its four defining conditions.

The next lemma can be found for example in Rao (1973b; pp. 64-65), and it follows from the Poincaré Separation theorem concerning eigenvalues.

**Lemma 3.6.** Let $F: p \times q$ satisfy $F'F = I_q$, and let $U: p \times p$ be positive definite. Then

$$|F'UF| \geq \prod_{i=1}^{q} \lambda_i,$$

where $\lambda_q \geq \lambda_{q-1} \geq \cdots \geq \lambda_1$ are the $q$ smallest eigenvalues of $U$.

Equality holds in Lemma 3.6 if $F$ comprises the eigenvectors corresponding to the $q$ smallest eigenvalues of $U$ which follows, for example, from Lemma 3.2.
4 Estimation

In this section the goal is to estimate $B$ and $\Sigma$ in (1) under the constraints in (2) and (3). The following decomposition of the whole space,

$$
\mathcal{C}(\Sigma) \oplus \mathcal{C}(\Sigma)^\perp \cap \mathcal{C}(A : \Sigma) \oplus \mathcal{C}(A : \Sigma)^\perp,
$$

where $\oplus$ denotes the "orthogonal sum" of linear spaces, yields a one-one transformation of the original model presented in (1):

$$
\begin{pmatrix}
\Gamma' \\
A'(I - \Gamma') \\
(A : \Sigma)^o'
\end{pmatrix}
X,
$$

where it has been utilized that $\mathcal{C}((I - \Gamma')A) = \mathcal{C}(\Sigma)^\perp \cap \mathcal{C}(A : \Sigma)$ (e.g. see Kollo & von Rosen, 2005; Theorem 1.2.16). Thus we obtain three relations which are fundamental for obtaining the estimates of the parameters:

\begin{align}
\Gamma'X &= \Gamma'ABC + \Delta^{1/2}\tilde{E}, \quad \tilde{E} \sim N_r,n(0,I_r,I_n), \quad (6) \\
A'(I - \Gamma')X &= A'(I - \Gamma')ABC, \quad (7) \\
(A : \Sigma)^o'X &= 0. \quad (8)
\end{align}

Note that " = " in (6) stands for equality in distribution whereas " = " in (7) and (8) means equality with probability 1. To perform likelihood inference we should replace $X$ by $X_o$ in these equations. In (6) the density is a function of $X_o$. Because of Assumption 3.1, (8) is trivially satisfied when $X$ is replaced by $X_o$, and will not be considered anymore. We call (6) to be the random part of the original model and (7) to be the deterministic part. Next, (7) will be exploited. From Lemma 3.3 it follows that we should study

$$
A'(I - H_oH_o')X_o = A'(I - H_oH_o')ABC
$$

which is identical to

$$
A'H_o^oH_o'^{o'}X_o = A'H_o^oH_o'^{o'}ABC. \quad (9)
$$

Thus, data put restrictions on $B$ which is unusual and not obvious how to handle. One strategy is to ignore them but here we will think of that utilizing the restrictions will improve our estimators. Assumption 3.1 implies that

$$
\mathcal{C}(A'H_o^oH_o'^{o'}X_o) \subseteq \mathcal{C}(A'H_o^oH_o'^{o'}A'). \quad (10)
$$

Moreover, it will be further assumed that

$$
\mathcal{C}(X_o'H_o^o) \subseteq \mathcal{C}(C') \quad (11)
$$

and therefore instead of $X_o'H_o^o$ its orthogonal projection on $\mathcal{C}(C')$,

$$
P_C.X_o'H_o^o,
$$

where $P_C$ denotes the orthogonal projector onto $\mathcal{C}(C')$. 


will be used when estimating the parameters. The projection could also have taken place in a preprocessing step of $X_o$. Note that $\mathcal{C}(X' H_o^o) \subseteq \mathcal{C}(C')$ holds with probability 1. Hence, we replace (9) by

$$A' H_o^o H_o^o X_o P_{C'} = A' H_o^o H_o^o A B C,$$  \hspace{1cm} (12)

and have now a consistent system of linear equations. A general solution to (12) is given by (see Kollo & von Rosen, 2005; Theorem 1.3.4)

$$B = (A' H_o^o A)^{-1} A' H_o^o H_o^o X_o C'(CC')^{-} + (A' H_o^o)^{\circ} \Theta_1 + A' H_o^o H_o^o A \Theta_2 C'$$

where $\Theta_1: (q - r(A' H_o^o)) \times k$, and $\Theta_2: q \times (k - r(C))$, are new parameters.

However, $H_o^o A$ has to be considered in some detail. The first observation is that if $H_o^o A = 0$ or $H_o^o A = 0$, which in principle never will hold mathematically, we cannot use the model in (1), with rank restrictions given by (2) and (3), and thus it will be assumed that $H_o^o A = 0$ or $H_o^o A = 0$ will not occur, which was implicitly assumed in the above discussion. In the following subsections two cases will be studied:

(a) $\mathcal{C}(A) \cap \mathcal{C}(\Sigma) = \{0\}$, $q \leq p - r$, which implies $r(H_o^o A) = q$;

(b) $\mathcal{C}(A) \cap \mathcal{C}(\Sigma) \neq \{0\}$.

In (b) there exists a special case, $r(H_o^o A) = p - r < q$, which, however, will not be considered because the treatment of the model under this condition follows the case discussed in the forthcoming Subsection 4.2. The only difference is that if $r(H_o^o A) = p - r$ holds, Assumption 3.1 is not needed, since in this case $\mathcal{C}(A) + \mathcal{C}(\Sigma)$ spans the whole space and $X_o$ belongs always to this space.

### 4.1 Estimation under the condition $\mathcal{C}(A) \cap \mathcal{C}(\Sigma) = \{0\}$

If $q \leq p - r$, $\mathcal{C}(A) \cap \mathcal{C}(\Sigma) = \{0\}$ yields $r(H_o^o A) = q$ and we get from (13) the relation

$$B = (A' H_o^o A)^{-1} A' H_o^o H_o^o X_o C'(CC')^{-}$$

because now $(A' H_o^o)^{\circ} = 0$. However, the estimation of $\Theta_2$ which only can be performed via (6), is impossible, because the term involving $B$, given by (14), into (6). Thus we can only estimate $\Sigma$ but not $B$.

**Theorem 4.1.** Let the model be defined in (1) with rank restrictions given by (2) and (3). Assume that $\mathcal{C}(A) \cap \mathcal{C}(\Sigma) = \{0\}$, $H_o^o A \neq 0$, $H_o^o A = 0$, $\mathcal{C}(X_o) \subseteq \mathcal{C}(A) + \mathcal{C}(\Sigma)$ and $\mathcal{C}(X_o H_o) \subseteq \mathcal{C}(C')$. Put

$$Y_o = (I - A(A' H_o^o H_o^o A)^{+} A' H_o^o H_o^o) X_o.$$
Then

\[ n\tilde{\Sigma} = \tilde{\Gamma}_o \tilde{\Delta}_o \tilde{\Gamma}_o', \]

where the estimate \( n\tilde{\Delta}_o \) equals the diagonal matrix with the \( r \) largest eigenvalues of \( Y_o Y_o' \) on the diagonal, and \( \tilde{\Gamma}_o \) comprises the eigenvectors of \( Y_o Y_o' \) corresponding to these eigenvalues.

Proof. The proof follows from the proof of Theorem 4.5. \( \square \)

4.2 Estimation under the condition \( C(A) \cap C(\Sigma) \neq \{0\} \)

According to (2) there exists the rank restriction \( r(B) = f \) and it will be assumed that \( f - r(H_o' X_o) < \min((q - r(H_o' X_o)), k) \). To estimate parameters the idea is to insert an appropriate chosen \( B \) in (6), for example based on (13). It seems reasonable to consider, instead of (13), the following solution

\[ B = (A'H_o' H_o A)^+ A'H_o' H_o X_o C'(CC')^- + (A'H_o')^o \Theta_1, \]

(15)

because when inserting this \( B \) into (6) \( \Theta_2 \) disappears, and when choosing \( B \) according to (15) the effect of the rank restrictions will be "maximal". Thus, the challenge is to put appropriate rank restrictions on \( \Theta_1 \) so that \( r(B) = f \) holds. It is noted that

\[ r((A'H_o' H_o A)^+ A'H_o' H_o X_o C'(CC')^-) : (A'H_o')^o \Theta_1) \]

\[ = r((A'H_o' H_o A)^+ A'H_o' H_o X_o C'(CC')^-) + r((A'H_o')^o \Theta_1), \]

since \((A'H_o')^o (A'H_o' H_o A)^+ = 0\) because \( C((A'H_o' H_o A)^+) = C(A'H_o' H_o A)\) (see Harville, 1997; Theorem 20.5.1). It can be remarked that it is essential that the Moore-Penrose generalized inverse has been chosen. Hence, from (15)

\[ f = \quad r((A'H_o' H_o A)^+ A'H_o' H_o X_o C'(CC')^-) + r((A'H_o')^o \Theta_1) \]

\[ = \quad r(A'H_o' H_o X_o C'(CC')^-) + r((A'H_o')^o \Theta_1) \]

\[ = \quad r(A'H_o' H_o X_o C') + r(\Theta_1), \]

where it has been utilized that \( r((A'H_o')^o \Theta_1) = r(\Theta_1 : A'H_o') - r(A'H_o') = r(\Theta_1), \) with an underlying assumption that the spaces generated by \( A'H_o' \) and \( \Theta_1 \), respectively, are disjoint. The relations in (10) and (11) show that

\[ f = \quad r(H_o' X_o) + r(\Theta_1) \]

and therefore \( \Theta_1 \) can be factored as \( \Theta_1 = \Psi_1 \Psi_2 \), with

\[ \Psi_1 : (q - r(H_o' A)) \times (f - r(H_o' X_o)), \]

\[ \Psi_2 : (f - r(H_o' X_o)) \times k. \]
The likelihood for

The model in (16) can be rewritten as

\[ \Gamma'X = \Gamma'A(A'H_o^\circ H_o^\circ A)^+ A'H_o^\circ H_o^\circ X_o + \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C + \Delta^{1/2} \tilde{E}, \]

(16)

where \( \tilde{E} \sim N_{r,n}(0,I_r,I_n) \). Via this model the parameters \( B, \Psi_1, \Psi_2, \Delta, \Gamma \) and \( \Sigma \), or sometimes linear combinations of them, e.g. \( L'B \) or \( L'\Psi_1 \), for some \( L \), will be discussed.

Let

\[ Y_o = (I - A(A'H_o^\circ H_o^\circ A)^+ A'H_o^\circ H_o^\circ)X_o, \quad (17) \]

\[ Y = X - A(A'H_o^\circ H_o^\circ A)^+ A'H_o^\circ H_o^\circ X_o. \quad (18) \]

The model in (16) can be rewritten as

\[ \Gamma'Y = \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C + \Delta^{1/2} \tilde{E}. \]

The likelihood for \( \Gamma'Y \) equals

\[ L(\Psi_1, \Psi_2, \Gamma, \Delta) = c|\Delta|^{-n/2} \exp \left\{ -\frac{1}{2} \text{tr} \left\{ \Delta^{-1}(\Gamma'Y_o - \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C)(')^d \right\} \right\} \]

\[ \leq c|\frac{1}{n}((\Gamma'Y_o - \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C)(')^d - \frac{1}{2} \text{tr}(\Gamma'Y_o - \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C)(')^d | \frac{1}{2} \exp \left\{ -\frac{1}{2} \text{tr}(\Gamma'Y_o - \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C)(')^d \right\}, \quad (19) \]

where \( c = (2\pi)^{-n/2} \) and \( \{W\}_d \) is the diagonal matrix with non-zero elements equal to elements on the main diagonal of \( W \). The proof of the inequality can be obtained from the proof of Theorem 1.10.4 in Srivastava & Khatri (1978). Equality in (19) holds if and only if

\[ n\Delta = \{(\Gamma'Y_o - \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C)(')^d \}. \quad (20) \]

Moreover, from (19) it follows that we shall study the following determinant

\[ |\{(\Gamma'Y_o - \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C)(')^d \}| \]

\[ \geq |\{(\Gamma'Y_o - \Gamma'A(A'H_o^\circ)^o \Psi_1 \Psi_2 C)(')^d \}| \]

(21)

and equality holds if \( \Gamma \) comprises the eigenvectors corresponding to the \( r \) smallest eigenvalues of \( (Y_o - A(A'H_o^\circ)^o \Psi_1 \Psi_2 C)(') \). Thus, \( \Gamma \) can be estimated if \( \Psi_1 \) and \( \Psi_2 \) can be estimated.

Now the goal is to minimize the determinant on the right hand side of (21). Let \( S_o \) be defined as in (4), where \( H_o \) consists of the \( r \) eigenvectors corresponding to the \( r \) largest eigenvalues of \( Y_o(I - P_C')Y_o' \) which are placed in the
diagonal matrix \( D_o \). Hence, the right hand side of (21) equals (see Corollary 3.4)

\[
\begin{align*}
& |(\Gamma'Y_o(I - PC')Y_o' - So)\Gamma + \Gamma' So\Gamma \\
& + (\Gamma'Y_o PC' - \Gamma' A(A'H_o')\Psi_1\Psi_2 C)(')|
\end{align*}
\]

\[
= |\Gamma' So\Gamma + (\Gamma'Y_o PC' - \Gamma' A(A'H_o')\Psi_1\Psi_2 C)(')|
\]

\[
= |\Gamma' So\Gamma||I + (\Gamma'Y_o PC' - \Gamma' A(A'H_o')\Psi_1\Psi_2 C)'(\Gamma' So\Gamma)^{-1}
\]

\[
\times (\Gamma'Y_o PC' - \Gamma' A(A'H_o')\Psi_1\Psi_2 C)|. 
\] (22)

Let

\[
T = A(A'H_o')\Psi_1, 
\] (23)

which is of size \( p \times (f - r(H_o'X_o)) \). In order to estimate \( \Psi_2 \) as a function of \( \Psi_1 \), we are going to apply the same technique used to find maximum likelihood estimates in the growth curve model (e.g. see Kollo & von Rosen, 2005; pp. 358–361). We can note that Tso (1981) used the the same approach as in this article whereas Johansen (1988) estimated first \( \Psi_1 \) as a function of \( \Psi_2 \). Below some calculations are presented showing the estimation approach. An important result is (for a proof, e.g. see Kollo & von Rosen, 2005; Theorem 1.2.25)

\[
(\Gamma' So\Gamma)^{-1} = (\Gamma' So\Gamma)^{-1}\Gamma'T(\Gamma' (\Gamma' So\Gamma)^{-1}\Gamma'T) - \Gamma'T(\Gamma' So\Gamma)^{-1} + (\Gamma'T)^{\circ}(\Gamma'T)^{\circ} (\Gamma' So\Gamma)(\Gamma'T)^{\circ} - (\Gamma'T)^{\circ} Y_o PC'. 
\] (24)

Applying this result yields that the right hand side of (22) is larger or equal to

\[
|\Gamma' So\Gamma||I + PC' Y_o' \Gamma(\Gamma'T)^{\circ}(\Gamma'T)^{\circ} \Gamma' So\Gamma(\Gamma'T)^{\circ} - (\Gamma'T)^{\circ} Y_o PC'|
\] (25)

and equality holds if and only if

\[
T'T(\Gamma' So\Gamma)^{-1}\Gamma'(Y_o PC' - T\Psi_2 C) = 0,
\]

which is identical to (see Corollary 3.5)

\[
T'S_o^+ (Y_o PC' - T\Psi_2 C) = 0.
\]

This is a consistent system of linear equations in \( \Psi_2 \) and all solutions satisfy

\[
T\Psi_2 C = T(T'S_o^+ T)^{-1}T'S_o^+ Y_o PC'. 
\] (26)

Note that the solutions are independent of the choice of \( \Gamma \) and that the inverse in (26) exists. Moreover, (25) is identical to

\[
|D_o^{-1}|I + PC' Y_o' H_o (H_o' T)^{\circ} ((H_o' T)^{\circ} D_o (H_o' T)^{\circ} - (H_o' T)^{\circ} H_o' Y_o PC' |. 
\] (27)
which also is independent of $\mathbf{\Upsilon}$. Now an appropriate linear combination of $\mathbf{\Psi}_1$ which is included in $\mathbf{T}$ will be estimated. Using (24) again, the relation in (27) will be manipulated, i.e. (27) is identical to

$$
|D_o^{-1}|\{I + P_C'Y_o'\hat{S}_o^\dagger Y_oP_C|
$$

$$
-P_C'Y_o'H_oD_o^{-1}H_o'T(T'H_oD_o^{-1}H_o'T)^{-1}T'H_oD_o^{-1}H_o'Y_oP_C|.
$$

(28)

Let

$$
F = D_o^{-1/2}H_o'T(T'H_oD_o^{-1}H_o'T)^{-1/2},
$$

(29)

where the square root is supposed to be symmetric. Then (28) can be written

$$
|D_o^{-1}|\{I + P_C'Y_o'\hat{S}_o^\dagger Y_oP_C|
$$

$$
\times|I - F'D_o^{-1/2}H_o'Y_oP_C|(I + P_C'Y_o'\hat{S}_o^\dagger Y_oP_C)^{-1}P_C'Y_o'H_oD_o^{-1/2}F|
$$

$$
= |D_o^{-1}|\{I + P_C'Y_o'\hat{S}_o^\dagger Y_oP_C|
$$

$$
\times|F'(I - D_o^{-1/2}H_o'Y_oP_C|(I + P_C'Y_o'\hat{S}_o^\dagger Y_oP_C)^{-1}P_C'Y_o'H_oD_o^{-1/2})F|.
$$

(30)

Since (e.g. see Kollo & von Rosen, 2005; Proposition 1.3.5) the following matrix

$$
I - D_o^{-1/2}H_o'Y_oP_C(I + P_C'Y_o'\hat{S}_o^\dagger Y_oP_C)^{-1}P_C'Y_o'H_oD_o^{-1/2}
$$

$$
= (I + D_o^{-1/2}H_o'Y_oP_CY_o'H_oD_o^{-1/2})^{-1}
$$

is positive definite and $F$ is of full rank, Lemma 3.6 can be applied. Thus, a lower bound of (30) is given by

$$
|D_o^{-1}|\{I + P_C'Y_o'\hat{S}_o^\dagger Y_oP_C|\prod_{i=1}^g \delta_i.
$$

(31)

where $\delta_1, \ldots, \delta_g (g = f - r(H_o'X_o))$ are the $g$ smallest eigenvalues of

$$
(I_r + D_o^{-1/2}H_o'Y_oP_CY_o'H_oD_o^{-1/2})^{-1}
$$

(32)

which do not depend on any unknown parameters. Therefore we have to find an estimator of $\mathbf{\Psi}_1$ so that the lower bound is attained. Let

$$
\hat{F} = (v_1, v_2, \ldots, v_g),
$$

where $\{v_i\}$ are the eigenvectors which correspond to the $g$ eigenvalues $\{\delta_i\}$. If it is possible to solve the following non-linear equation in $\mathbf{\Psi}_1$, see equation (29),

$$
\hat{F} = D_o^{-1/2}H_o'T(T'H_oD_o^{-1}H_o'T)^{-1/2}
$$

$$
= D_o^{-1/2}H_o'A(\hat{A}'H_o^o)^{\dagger}\mathbf{\Psi}_1\{\mathbf{\Psi}_1(\hat{A}'H_o^o)^{\dagger}A'H_oD_o^{-1}H_o'A(\hat{A}'H_o^o)^{\dagger}\mathbf{\Psi}_1\}^{-1/2},
$$

(33)
there is chance to find an estimator of $\Psi_1$ but it is known from the usual BRM that without any rank condition on $A$ it is not possible to estimate $\Psi_1$. The linear combinations

$$H_o^\prime A(A'H_o^o)\Psi_1 = H_o'T$$

are however estimable and

$$H_o^\prime A(A'H_o^o)\Psi_1 = D_o^{1/2} \hat{F}$$

which is verified by inserting this expression in (33) and using that the eigenvectors $\hat{F}$ are of unit length, i.e. $\hat{F}^\prime \hat{F} = I$. Note that $A(A'H_o^o)$ is connected to the fact that we only consider $B$ in (13) to be a function of $\Theta_1$, and $H_o$ is used because we are now working with the random part of the model in (6). If $A = I$ the parameter $\Psi_1$ is always estimable whereas if $\Sigma$ is positive definite $A\Psi_1$ is always estimable and if additionally $r(A) = q$, $\Psi_1$ is estimable. Moreover, if $r'(H_o'A_o) = p - r$ then $C(H_o) \subseteq C(A)$ and $\Psi_1$ is estimable.

**Proposition 4.2.** Let the model be defined in (1) with rank restrictions given by (2) and (3). If $C(A) \cap C(\Sigma) \neq \{0\}$, $H_o^o'A \neq 0$, $H_o'A \neq 0$, $C(X_o) \subseteq C(A) + C(\Sigma)$, $C(X_o'H_o) \subseteq C(C)$ and

$$\tilde{\Psi}_2 C = \hat{F} D_o^{-1/2} H_o'Y_o P_C,$$

$$H_o^\prime A(A'H_o^o)\Psi_1 = D_o^{1/2} \hat{F},$$

where $D_o$, $Y_o$ and $\hat{F}$ are presented in (4), (17) and (33), respectively. Then

$$H_o^\prime ABC$$

$$= H_o^\prime A(A'H_o^o)H_o'A^\prime + H_o'X_oP_{C^o} + D_o^{1/2} \hat{F} D_o^{-1/2} H_o'Y_o P_{C^o}.$$

**Motivation:** First we show the relation for $\tilde{\Psi}_2 C$. From (26) it follows that

$$\tilde{\Psi}_2 C = (\tilde{T}' S_o^T \tilde{T})^{-1} \tilde{T}' S_o Y_o P_{C^o} = (\tilde{T}' H_o D_o^{-1} H_o' \tilde{T})^{-1} \tilde{T}' H_o D_o^{-1} H_o' Y_o P_{C^o},$$

where according to (23) and (34) $H_o' \tilde{T} = H_o^\prime A(A'H_o^o)\Psi_1 = D_o^{1/2} \hat{F}$. A few calculations show that $\tilde{T}' S_o^T \tilde{T} = \hat{F} \hat{F} = I_g, g = f - r(H_o X_o)$ and $\tilde{T}' S_o Y_o = \hat{F} D_o^{-1/2} H_o'Y_o$ which establish $\tilde{\Psi}_2 C$.

From (13) one can see that

$$H_o^\prime ABC = H_o^\prime A(A'H_o^o)A^\prime + A'H_o^o H_o'X_o P_{C^o} + H_o^\prime A(A'H_o^o)\Psi_1 \tilde{\Psi}_2 C,$$

and using (34) together with $\tilde{\Psi}_2 C$, presented above, implies the expression for $H_o^\prime ABC$.

In Proposition 4.2 we have estimated $ABC$ within the random part of the model. However, in most realistic situations we can make a stronger statement.
Proposition 4.3. Assume that $\mathcal{C}(A) \cap \mathcal{C}(\Sigma)^\perp = \{0\}$, $r(A) = q$, $r(C) = k$, and let the matrices be as in Proposition 4.2. Then $B$ can be estimated using
\[
\hat{B} = (A'H_o H_o A)' + A'H_o H_o X_o C'(CC')^{-1} \\
+ (A'H_o H_o A)^{-1} A'H_o D^{1/2} \tilde{F} \tilde{F}' D^{-1/2} H_o Y_o C'(CC')^{-1}.
\]

Motivation: By assumption it follows that $(A'H_o H_o A)^{-1}$ exists and therefore premultiplying $H_o A \hat{B} C$ by $(A'H_o H_o A)^{-1} A'H_o$ establishes the statement.

The assumption $\mathcal{C}(A) \cap \mathcal{C}(\Sigma)^\perp = \{0\}$ in Proposition 4.3 is natural since inference about the mean parameters is connected to the random part of the model in (1).

Proposition 4.4. Let the model be defined in (1) with rank restrictions given by (2) and (3). Assume that $\mathcal{C}(A) \cap \mathcal{C}(\Sigma) \neq \{0\}$, $H_o A \neq 0$, $H_o A \neq 0$, $\mathcal{C}(X_o) \subseteq \mathcal{C}(A) + \mathcal{C}(\Sigma)$, $\mathcal{C}(X_o A) \subseteq \mathcal{C}(C)$ and $Y_o$ is given in (17). The estimate $n\hat{\Delta}_o$ equals the diagonal matrix with the $r$ smallest eigenvalues of $(Y_o - H_o A \hat{B} C)$ on its main diagonal, and $\hat{\Gamma}_o$ is given by the eigenvectors of $(Y_o - H_o A \hat{B} C)$ corresponding to the $r$ eigenvalues, where $H_o A \hat{B} C$ is presented in Proposition 4.2. Then
\[
n\hat{\Sigma} = \hat{\Gamma}_o \hat{\Delta}_o \hat{\Gamma}_o'.
\]

Motivation: The result follows from (19), (20), (21) and the fact that $\Gamma' = \Gamma'(H_o A H_o + H_o A H_o') = \Gamma'H_o A H_o'$.

Theorem 4.5. If choosing $H_o A \hat{B} C$ as in Proposition 4.2 and $\hat{\Sigma}$ as in Proposition 4.4 the likelihood which corresponds to (6), with restrictions on $B$ given by (7), is maximized.

We end this section by studying $H_o A \hat{B} C$ when $n \to \infty$. Firstly it is observed that the "largest" eigenvalues of $\frac{1}{n}S$ converge in probability to the non-zero eigenvalues of $\Sigma$ which for simplicity are supposed to be of multiplicity 1. Moreover, the corresponding eigenvectors of $\frac{1}{n}S$ which correspond to $H_o$ converge to $\Gamma$. We will not be able to say anything about $H_o$ and therefore in our expressions $H_o$ is kept and not replaced by any random variable. To show that the eigenvectors converge it can be noted that $S = ZZ'$, for $Z \sim N_{p,n-r(C)}(0, \Sigma, I)$, $r(Z) = r$, and then its density which is defined on a subspace (see Srivastava & Khatri, p. 43) can be used. Based on this density $D$ and $H$ ($D$ and $H$ correspond to $D_o$ and $H_o$) are maximum likelihood estimators and therefore they are also consistent. Hence, in probability,
\[
H \to \Gamma, \quad \frac{1}{n}D \to \Delta.
\]

Now we study what happens with the random version of $H_o A \hat{B} C$ (which will be called $H_o A \hat{B} C$):
\[
H_o A \hat{B} C = H_o A (A'H_o H_o'A)^+ A'H_o H_o' X_o P C', \\
+ D^{1/2} \tilde{F} \tilde{F}' D^{-1/2} H'o Y_o P C',
\]
(35)
where $Y$ is given by (18) and now $\widehat{F}$ is a random variable. The crucial point is to observe that the random version of (32) converges to $I_r$. This holds since $\sqrt{n}D^{-1/2} \to \Delta^{-1/2}$ and $n^{-1}YP_C Y' \to 0$ in probability, under some mild assumptions on $C$, which are supposed to hold. Therefore the eigenvalues of (32) converge to 1. Since $\widehat{F}F'$ is an eigenprojector we will construct our result so that its limit also is an eigenprojector. Thus $\widehat{F} \to (I_g : 0)'$, in probability, implying the result

$$\widehat{FP}' \to \begin{pmatrix} I_g & 0 \\ 0 & 0 \end{pmatrix}, \quad \widehat{F}'F = I_g,$$

where obviously $\begin{pmatrix} I_g & 0 \\ 0 & 0 \end{pmatrix}$ is a projector.

**Theorem 4.6.** The expression $H'AB'C$ in (35) is asymptotically equivalent to

$$\Gamma' A (A'H_o H_o A) + A' H_o H_o X_P C' + \begin{pmatrix} I_g & 0 \\ 0 & 0 \end{pmatrix} \Gamma' Y P_C',$$

which is normally distributed.

A final remark is that if assuming $H_o' = 0$, i.e. $\Sigma$ is positive definite, then the estimates in Proposition 4.2 and Proposition 4.4 for $B$ and $\Sigma$ under rank restrictions on $B$ are identical to known estimates of the parameters of the growth curve model with rank restrictions on $B$.

### 5 Illustrations

**Example 5.1.** Firstly we present a small simulation study showing that our estimators make sense. Since $B$ with rank restrictions is difficult to interpret, the focus will be on prediction, i.e. $H_o' \widehat{X}C'(CC')^{-1} = H_o'AB$ will be considered, and is obtained from Proposition 4.2. Note that for small $n$ there is rather much variation in $H_o$. It is not obvious how to evaluate $H_o' \widehat{X}$ but we have decided to compare $H_o' \widehat{X}C'(CC')^{-1}$ with $H_o' P_A X C'(CC')^{-1}$ which for a fixed $H_o$ is an unbiased estimator of $H_o' AB$, irrespectively if there exist rank restrictions on $B$ or there are no rank restrictions.

Let

$$A = \begin{pmatrix} 1 & 8 & 64 \\ 1 & 10 & 100 \\ 1 & 12 & 144 \\ 1 & 14 & 196 \end{pmatrix}, \quad B = \begin{pmatrix} 17.1 & 37.1 & 34.2 \\ 0.54 & 0.23 & 1.08 \\ -0.003 & 0.0020 & -0.006 \end{pmatrix},$$

$$C = \left( \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes 1_{10} : \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes 1_g : \begin{pmatrix} 0 \\ 0 \end{pmatrix} \otimes 1_s \right), \quad \Sigma_o = \begin{pmatrix} 5.1 & 2.4 & 3.6 & 2.5 \\ 3.9 & 2.7 & 3.1 & 4.6 \\ 6.0 & 3.8 & 4.6 \end{pmatrix}.$$
This means that we have a growth curve model with polynomial growth of order two and three groups, respectively consisting of 10, 9 and 8 independent observations. Data were generated from the model $X = ABC + E$, where $E \sim N_p(0, \Sigma)$ with $\Sigma = HDH'$, where $D$ is a diagonal matrix which consists of the three largest eigenvalues of $\Sigma_0$ and $H$ is the matrix of corresponding eigenvectors. Thus $r(\Sigma) = 3$. Note that the third column of $B$ equals two times its first column, meaning that the rank of $B$ equals 2. Then 1000 simulations took place and $H'\hat{X}C'(CC')^{-1}$ was compared to $H'PAXC'(CC')^{-1}$, where $H'\hat{X}C'(CC')^{-1}$ follows from Proposition 4.2. Since the size of these matrices are $3 \times 3$ there are nine elements which are to be compared. The results of the simulation study are presented in Table 1. One conclusion is that there is almost no difference between the two estimators besides that $H'\hat{X}C'(CC')^{-1}$ is of rank equal to two and the other estimator is of rank three. Thus, we can conclude that our approach makes sense from a point of view that even in small samples the estimators seem to be appropriate.

Table 1: Let $p = \text{vec}(H'\hat{X}C'(CC')^{-1})$ and $p_i$ be the $i$th element of $p$. Correspondingly, let $q = \text{vec}(H'PAXC'(CC')^{-1})$ and let $q_i$ be the $i$th element of $q$.

<table>
<thead>
<tr>
<th></th>
<th>$p_1$</th>
<th>$q_1$</th>
<th>$p_2$</th>
<th>$q_2$</th>
<th>$p_3$</th>
<th>$q_3$</th>
<th>$p_4$</th>
<th>$q_4$</th>
<th>$p_5$</th>
<th>$q_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>44.6</td>
<td>44.6</td>
<td>78.5</td>
<td>78.4</td>
<td>89.1</td>
<td>89.2</td>
<td>4.54</td>
<td>4.53</td>
<td>8.08</td>
<td>8.09</td>
</tr>
<tr>
<td>std</td>
<td>1.4</td>
<td>1.4</td>
<td>1.8</td>
<td>1.8</td>
<td>1.9</td>
<td>1.9</td>
<td>3.6</td>
<td>3.7</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>min</td>
<td>39.7</td>
<td>39.8</td>
<td>70.1</td>
<td>69.9</td>
<td>80.3</td>
<td>80.3</td>
<td>-1.9</td>
<td>-2.4</td>
<td>-1.6</td>
<td>-1.7</td>
</tr>
<tr>
<td>max</td>
<td>39.7</td>
<td>48.2</td>
<td>83.9</td>
<td>84.3</td>
<td>94.7</td>
<td>94.8</td>
<td>19.4</td>
<td>19.0</td>
<td>34.6</td>
<td>34.6</td>
</tr>
<tr>
<td>$p_6$</td>
<td>9.08</td>
<td>9.08</td>
<td>4.61</td>
<td>4.56</td>
<td>8.68</td>
<td>8.79</td>
<td>9.22</td>
<td>9.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_7$</td>
<td>7.3</td>
<td>7.4</td>
<td>3.6</td>
<td>3.6</td>
<td>6.2</td>
<td>6.2</td>
<td>7.2</td>
<td>7.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_8$</td>
<td>-3.7</td>
<td>-4.7</td>
<td>-1.9</td>
<td>-2.5</td>
<td>-1.6</td>
<td>-1.5</td>
<td>-3.8</td>
<td>-3.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>38.7</td>
<td>38.9</td>
<td>20.3</td>
<td>20.1</td>
<td>35.7</td>
<td>35.8</td>
<td>40.7</td>
<td>40.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example 5.2.** Now we turn to the real data example utilized by Albert & Kshirsagar (1993) which was mentioned in the introduction. Group averages for potassium values are plotted separately for each group of dogs in Figure 5.1.
Figure 1: The dogs’ data used in Grizzle & Allen (1969) and Albert & Kshirsagar (1993) are presented. Each line corresponds to one group of dogs where the average values are plotted over time.

The $A$ and $C$ matrices equal

$$A = \begin{pmatrix}
0.378 & -0.567 & 0.546 & -0.408 \\
0.378 & -0.378 & 0 & 0.408 \\
0.378 & -0.189 & -0.327 & 0.40 \\
0.378 & 0 & -0.436 & 0 \\
0.378 & 0.189 & -0.327 & -0.408 \\
0.378 & 0.378 & 0 & -0.408 \\
0.378 & 0.567 & 0.546 & 0.408 \\
\end{pmatrix}$$

$$C = \left( \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \otimes 1_9 : \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \otimes 1_{10} : \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \otimes 1_8 : \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \otimes 1_9 \right).$$

We will assume that $r(\Sigma)$ equals 7, 6 or 5. When $r(\Sigma) = 7$, i.e. the dispersion matrix is of full rank, Albert & Kshirsagar (1993) developed a testing strategy to decide about the rank of $B$. We will compare four models by assuming

$$r(\Sigma) = 7, r(B) = 4; r(\Sigma) = 7, r(B) = 1; r(\Sigma) = 6, r(B) = 2$$

or $r(\Sigma) = 5, r(B) = 3$.

The first case is a model without any rank restrictions. In the second alternative we only have rank restrictions on the mean parameters, which was the case
Table 2: Let $p_{ij} = (H_o' \tilde{X}C'(CC')^{-1})_{ij}$, $i = 1, l$, $l$ stands for the last row, $j = 1, 2, 3, 4$, where $H_o' \tilde{X}C'(CC')^{-1}$ follows from Proposition 4.2.

<table>
<thead>
<tr>
<th></th>
<th>$p_{11}$</th>
<th>$p_{12}$</th>
<th>$p_{13}$</th>
<th>$p_{14}$</th>
<th>$p_{l1}$</th>
<th>$p_{l2}$</th>
<th>$p_{l3}$</th>
<th>$p_{l4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r(\Sigma) = 7, r(B) = 4$</td>
<td>12.3</td>
<td>9.2</td>
<td>10.7</td>
<td>10.1</td>
<td>-0.003</td>
<td>0.06</td>
<td>0.07</td>
<td>0.1</td>
</tr>
<tr>
<td>$r(\Sigma) = 7, r(B) = 1$</td>
<td>11.7</td>
<td>10.0</td>
<td>10.3</td>
<td>10.3</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>$r(\Sigma) = 6, r(B) = 2$</td>
<td>12.4</td>
<td>9.3</td>
<td>10.6</td>
<td>10.0</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>$r(\Sigma) = 5, r(B) = 3$</td>
<td>11.9</td>
<td>10.2</td>
<td>10.2</td>
<td>10.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

treated by Albert & Kshirsagar (1993), and for the other two cases we have rank restrictions on both the mean parameters and the dispersion matrix. As in the simulation study (Example 5.1) we will compare with $H_o' \tilde{X}C'(CC')^{-1}$ where the choice of $H_o$ depends on $r(\Sigma)$, in particular the length of the column in $H_o$ depends on the rank assumption. Moreover, it is impossible to present all elements in $H_o' \tilde{X}C'(CC')^{-1}$ and therefore the focus will be on the first and last rows of $H_o' \tilde{X}C'(CC')^{-1}$ of which each consists of four elements. Table 2 shows similar prediction values for the different models. For the last row, i.e. $p_{l1}$-$p_{l4}$, some differences appear but it should not be forgotten that different $H_o$ matrices are used in the different models so these minor differences are not unexpected.

Thus, as with the simulation study, we find an agreement between existing methods and our new approach where rank restrictions on the mean are combined with rank restrictions on the dispersion. Moreover, if $p$ is large those existing methods mentioned above cannot be used because the dispersion matrix cannot be estimated and $S$ cannot be inverted which in our approach can take place. Therefore, we believe that it is worth to continue to study the new approach presented in this article.

Acknowledgement

This research has been supported by The Swedish Foundation for Humanities and Social Sciences (P14-0641:1) and the Swedish Research Council (2017-03003). The comments made by an anonymous referee have significantly improved the paper.

References


