Multivariate Regression with Monotone Missing Observation of the Dependent Variables
Raats, V.M.; van der Genugten, Ben; Moors, J.J.A.

Publication date:
2002

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.
MULTIVARIATE REGRESSION WITH MONOTONE MISSING OBSERVATION OF THE DEPENDENT VARIABLES

By V.M. Raats, B.B. van der Genugten, J.J.A. Moors

July 2002
Multivariate regression with monotone missing observations of the dependent variables

V.M. Raats*, B.B. van der Genugten and J.J.A. Moors

Abstract

Multivariate regression is discussed, where the observations of the dependent variables are (monotone) missing completely at random; the explanatory variables are assumed to be completely observed.

We discuss OLS-, GLS- and a certain form of E(stimated) GLS-estimation. It turns out that (E)GLS-estimation uses the preceding dependent variables in a well-structured way. In case of normality, ML-estimation coincides with (E)GLS-estimation. We include (sets of) MANOVA-tables enabling us to perform exact tests on the coefficients based on a (new) generalized Wilks’ distribution.

Only the very special case of the constant as sole explanatory variable has been treated in the literature so far: our model incorporates this missing data problem.

Keywords: generalized Wilks distribution, hypothesis testing, least squares, maximum likelihood, monotone sample data

Jel codes: C12, C13, C19

*V.M.Raats@kub.nl, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
1 Introduction

In multivariate linear regression, dependent variables are written as a linear combination of explanatory variables. An extensive literature exists on multivariate linear regression and much is known about for example the maximum likelihood estimators in case of normality. We only mention Seber (1984) for future reference. However, almost all these results are derived for the situation with either no missing observations or no specific missing data pattern for the dependent variables. One exception is Robins and Rotnitzky (1995), who discuss the semiparametric efficiency in multivariate regression models with missing data and in particular with monotone missing observations for the dependent variables. In this paper we also consider the specific case where the dependent variables follow a monotone missing data pattern: the dependent variables can be ordered in such a way that if an observation of a dependent variable for an item is missing, the observations of all subsequent dependent variables for the same item are also missing. See Schafer (1997) e.g. for a more extensive discussion about monotone data patterns. The explanatory variables are assumed to have been completely observed, i.e. no missing observations occur. Throughout the paper, the data are assumed to be missing completely at random (MCAR). In that case, the missing data mechanism can be ignored for inferences (see Rubin (1976)).

Practical examples of such data patterns are experimental designs where new dependent variables are added during the experiment, panel surveys with drop outs or new members, or repeated audit controls. In a repeated audit control, the assumption of an infallible auditor is dropped and to take possible mistakes of the auditor into account, a subsample of the already checked elements is checked once more (see Moors et al. (2000)).

In Section 2 our model for monotone regression is presented in detail and illustrated with a numerical example. In Section 3 four estimation procedures are discussed: OLS-, GLS-, a certain form of E(stimated) GLS- and ML-estimation. OLS-estimation is solely based on the explanatory variables, while (E)GLS uses the explanatory variables and the preceding dependent variables in a well-structured way. The estimators appear to have a clear geometrical interpretation. Finally, it is shown that under the normality assumption the (E)GLS-estimators for the coefficients coincide with the MLE.

Two special cases of monotone regression are presented next. In Section 4 the situation with no missing observations is discussed; it is shown that our general formulae then reduce to well-known estimation results, such as described by Seber (1984). In Section
the model with the constant term as an explanatory variable, and in particular the model with the constant term as the sole explanatory variable, is described. Under the assumption of normality and observations missing at random (see Rubin (1976)), Anderson (1957), Bhargava (1962), Afifi and Elashoff (1966), Anderson and Olkin (1985), Jinadasa and Tracy (1992) and Fujisawa (1995) all presented expressions for the maximum likelihood estimators (MLE) for the latter model. Our formulae generalize the results of these previous authors considerably.

Section 6 describes estimation under linear restrictions and gives MANOVA-tables to perform exact tests on the coefficients. A general approach for complete data can be found in Van Der Genugten (1997) e.g.. Section 7 reviews the Wishart and Wilks’ distribution and introduces a generalized Wilks’ distribution that describes the distribution of our test statistics in Section 8. In Section 9 the presented estimation and testing techniques are applied to the numerical illustration. The final Section 10 contains the main conclusions and ideas for future research.

2 The model

Consider the multivariate linear regression model with $M$ dependent variables, $k$ (deterministic) explanatory variables and $N$ items. Let $X_{tj} \in \mathbb{R}$ be the observed value of the $j^{th}$ explanatory variable ($j = 1, \ldots, k$) for the $t^{th}$ item; complete data are available for the explanatory variables, so $t = 1, \ldots, N$ for all $j$.

The observations of the dependent variables are incomplete; the dependent variables are divided into $r$ ordered groups according to the pattern of increasing missing data. Group $i$ contains $m_i$ variables for which exactly the first $N_i$ observations are available ($N = N_1 \geq N_2 \geq \ldots \geq N_r$); $M_i = \sum_{j=1}^{i} m_j$ ($i = 1, \ldots, r$; $M_r = M$). The vector $Y_{ti} \in \mathbb{R}^{m_i}$ contains the values of these $m_i$ dependent variables for item $t$. So $Y_{ti}$ is observable for $t = 1, \ldots, N_i$ and missing for $t = N_i + 1, \ldots, N$.

The $r$ (multivariate) regression equations can be written as

$$Y_{ti} = \mu_{ti} + \varepsilon_{ti}, \quad \mu_{ti} = \sum_{j=1}^{k} X_{tj} \beta_{ji}, \quad i = 1, \ldots, r, \quad (2.1)$$

where $\beta_{ji} \in \mathbb{R}^{m_i}$ denotes a vector of completely unknown regression coefficients. For the errors the usual assumptions are

$$E(\varepsilon_{ti}) = 0, \quad Cov(\varepsilon_{ti}, \varepsilon_{sj}) = \delta_{ts} \sigma_{ij}, \quad (2.2)$$
with (completely unknown) non-singular $\Sigma = (\sigma_{ij}) \in \mathbb{R}^{M \times M}$ not depending on the $\beta_{ji}$. We write $\Sigma > 0$. Normality of the errors will be assumed only when used explicitly.

The union of the groups 1 up to $i$ will be denoted by $(i)$, hence $Y_{t(i)} = (Y_{t1} \ldots Y_{ti})' \in \mathbb{R}^{M_i}$, $i = 1, \ldots, r$ and similarly for $\mu_{t(i)}$ and $\varepsilon_{t(i)}$. For OLS-estimation of the coefficients $\beta_{ji}$ we minimize ($N_r+1 := 0$):

$$\sum_{i=1}^{r} \sum_{t=N_{i+1}+1}^{N_i} \varepsilon_{t(i)}' \varepsilon_{t(i)}.$$  \hfill (2.3)

This minimization problem will be solved in Section 3.2.

The error covariance matrix $\Sigma_{(i)(i)}$ of $\varepsilon_{t(i)}$ can be partitioned as follows

$$\Sigma_{(i)(i)} := Cov(\varepsilon_{t(i)}) = Cov \left( \begin{array}{c} \varepsilon_{t(i-1)} \\ \varepsilon_{ti} \end{array} \right) = \begin{bmatrix} \Sigma_{(i-1)(i-1)} & \Sigma_{(i-1)i} \\ \Sigma_{i(i-1)} & \Sigma_{ii} \end{bmatrix}.$$  \hfill (2.4)

Of course, $\Sigma_{(i)(i)} \in \mathbb{R}^{M_i \times M_i}$, $\Sigma_{(i-1)(i-1)} \in \mathbb{R}^{M_i-1 \times M_i-1}$ and $\Sigma_{(i-1)i} \in \mathbb{R}^{M_i-1 \times m_i}$. So $\Sigma_{(r)(r)} = \Sigma$ and $\Sigma_{(1)(1)} = \Sigma_{11}$.

For GLS-estimation we minimize

$$\sum_{i=1}^{r} \sum_{t=N_{i+1}+1}^{N_i} \varepsilon_{t(i)}' \Sigma_{(i)(i)}^{-1} \varepsilon_{t(i)}.$$  \hfill (2.5)

This problem will be treated in Section 3.3. Of course this method can only be applied for known $\Sigma$. Therefore in Section 3.4 we consider EGLS-estimation, where $\Sigma$ is replaced by a specific estimate $\hat{\Sigma}$. In Section 3.5 we consider maximum likelihood estimation under normality; it will be shown that EGLS-estimation coincides with ML-estimation.

The notations are illustrated by means of the following fictitious data.

**Numerical illustration**

As usual, columns of $X$ (and $Y$) refer to variables and rows to items. Not observed values in $Y$ are denoted by parentheses. We nevertheless give these values to compare the results obtained from the incomplete data with the results for the complete data.
\[
X = \begin{bmatrix}
1 & 5 & 5 & 7 \\
1 & 1 & 3 & 1 \\
1 & 3 & 3 & 1 \\
1 & 3 & 1 & 3 \\
1 & 5 & 5 & 7 \\
1 & 1 & 3 & 1 \\
1 & 3 & 3 & 1 \\
1 & 3 & 1 & 3 \\
1 & 4 & 4 & 5 \\
1 & 2 & 3 & 2 \\
1 & 3 & 3 & 2 \\
1 & 3 & 2 & 3
\end{bmatrix}, \\
Y = \begin{bmatrix}
7 & 5 & 6 & 1 \\
5 & 9 & 2 & 4 \\
1 & 1 & 2 & 5 \\
4 & 2 & 0 & 4 \\
5 & 9 & 8 & 4 \\
7 & 8 & 4 & 6 \\
4 & 1 & 8 & 2 \\
3 & 2 & 4 & 1 \\
5 & 7 & 5 & 4 \\
6 & 8 & 6 & (5) \\
6 & (3) & (5) & (6)
\end{bmatrix}
\]

So for example
\[
X_{1,4} = 7, \ Y_{1,1} = 7, \ Y_{1,2} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \ Y_{1(2)} = \begin{bmatrix} 7 \\ 5 \\ 6 \end{bmatrix}, \ Y_{1,3} = 1,
\]
and (2.1) reads for \( i = 2 \):
\[
Y_{t,2} = \beta_{1,2} + X_{t,2} \beta_{2,2} + X_{t,3} \beta_{3,2} + X_{t,4} \beta_{4,2} + \varepsilon_{t,2}, \ t = 1, \ldots, 11.
\]

Note that suffices are separated by a comma whenever confusion threatens.

# 3 Estimation

## 3.1 Notation

We introduce some column-notation for the observed variables and regression coefficients. The index \( i \) refers to group \( i \) and \( (i) \) again to the union of the groups \( 1, 2, \ldots, i \).

\[
X = \begin{bmatrix}
X_{1,1} & X_{1,2} & \cdots & X_{1,k} \\
\vdots & \vdots & & \vdots \\
X_{N_i,1} & X_{N_i,2} & \cdots & X_{N_i,k} \\
\vdots & \vdots & & \vdots \\
X_{N,1} & X_{N,2} & \cdots & X_{N,k} \\
\end{bmatrix} \quad \beta = \begin{bmatrix}
\beta_{1,1} & \cdots & \beta_{1,i-1} & \beta'_{1,i} & \cdots & \beta'_{1,r} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\beta_{j,1} & \cdots & \beta_{j,i-1} & \beta'_{j,i} & \cdots & \beta'_{j,r} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\beta_{k,1} & \cdots & \beta_{k,i-1} & \beta'_{k,i} & \cdots & \beta'_{k,r} \\
\end{bmatrix}
\]

\[
\uparrow \quad \uparrow \quad \uparrow \quad \uparrow
\]

\[
\beta_{(i-1)} \quad \beta_{i} \quad \cdots \quad \beta_{r}
\]
So \( X_i \in \mathbb{R}^{N_i \times k} \) is the matrix with the first \( N_i \) observations of all exogenous variables. The submatrices \( \beta_{(i-1)} \in \mathbb{R}^{k \times M_{i-1}} \) and \( \beta_i \in \mathbb{R}^{k \times m_i} \) of \( \beta \in \mathbb{R}^{k \times M} \) contain the regression coefficients corresponding to groups \((i - 1)\) and \(i\) of dependent variables, respectively. The \( Y_i \) can be grouped in the following way:

\[
Y = \begin{bmatrix}
Y'_{1,1} & \cdots & Y'_{1,i-1} & Y'_{1,i} & \cdots & Y'_{1,r} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
Y'_{N_i,1} & \cdots & Y'_{N_i,i-1} & Y'_{N_i,i} \\
\vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
Y'_{N_i-1,i-1} & \cdots & \vdots & \ddots & \ddots & \vdots \\
Y'_{N_i}
\end{bmatrix}
\]

The matrix \( Y_i \in \mathbb{R}^{N_i \times m_i} \) contains all observations of group \(i\). But the matrix \( Y_{(i-1)} \in \mathbb{R}^{N_i \times M_{i-1}} \) contains only the first \( N_i \) observations of the foregoing groups \((i - 1)\) (with \( Y_{(0)} = 0 \)). We use similar definitions for the \( \mu_{ti} \) and \( \varepsilon_{ti} \).

### 3.2 OLS-estimation

From (2.1) we get \((i = 1, \ldots, r)\)

\[
\begin{align*}
Y_i &= \mu_i + \varepsilon_i, \\
\mu_i &= X_i \beta_i \\
Y_i &= \mu_{(i-1)} + \varepsilon_{(i-1)}, \quad \mu_{(i-1)} = X_i \beta_{(i-1)}.
\end{align*}
\]

Then the OLS-criterion (2.3) can be written as

\[
\sum_{i=1}^{r} \text{tr}(\varepsilon'_i \varepsilon_i).
\]

So the OLS-estimates can be found by columnwise orthogonal projection. We define the following relevant spaces and accompanying characteristics:

\[
\begin{align*}
L_i &= \mathcal{R}(X_i) : \text{the space spanned by the columns of } X_i \\
H_i &\in \mathbb{R}^{N_i \times N_i} : \text{the orthogonal projection matrix of } L_i \\
U_i &= I_{N_i} - H_i : \text{the orthogonal projection matrix of } L_i^\perp \\
l_i &= \text{dim}(L_i) = r(X_i), \quad r_i = \text{dim}(L_i^\perp) = N_i - l_i.
\end{align*}
\]

Clearly each column of \( \mu_i \) is element of \( L_i \). To indicate this we write \((\mu_i)_h \in L_i \ (\forall h)\).

**Theorem 1** The OLS-estimators for \( \mu_i \) and \( \varepsilon_i \) \((i = 1, \ldots, r)\) are

\[
\begin{align*}
Z_i &:= H_i Y_i, \\
E_i &:= Y_i - Z_i = U_i Y_i.
\end{align*}
\]
Proof. The OLS-criterion (3.2) is the sum of $r$ inner products of the error terms. Since the mean $\mu_i$ only appears in the $i^{th}$ inner product, (3.2) is minimized by separate minimization of the $r$ inner products. The minimization of term $i$ is based on the following orthogonal decomposition of the error term:

$$\varepsilon_i = Y_i - \mu_i = H_i(Y_i - \mu_i) + U_i(Y_i - \mu_i) = (Z_i - \mu_i) + U_iY_i.$$  

The first part of the orthogonal decomposition is zero for the OLS-estimators (3.4).  

The OLS-estimators $b_i$ for the regression coefficients $\beta_i$ are given by

$$b_i = G_iX'_iY_i \quad \text{with} \quad G_i = (X'_iX_i)^{-}. \quad (3.5)$$

As estimator $S$ for the covariance matrix $\Sigma$, we take the sample covariance matrix corrected for degrees of freedom. Using a similar partition for $S$ as for $\Sigma$ in (2.4), this leads to the covariance estimator

$$S_{ii} = E'_iE_i/r_i, \quad (S_{(i-1)i})_g = (E'_iE_{(i-1)})_g/\sqrt{r_g r_i} \quad \text{for} \quad g = 1, ..., i-1. \quad (3.6)$$

Here $(A)_g$ denotes the columns $M_g-1$ through $M_g$ of $A$, i.e. all columns corresponding to group $g$.

The correction for the degrees of freedom takes place to reduce the bias of $S$ and to secure $S \geq 0$. (The latter is the case since $S$ is the inner product $(FD)'(FD)$ where $F$ denotes the matrix with the elements of the residuals $E_i$ as the corresponding elements for the available observations, and zeros for the missing observations, while $D$ is a diagonal matrix with diagonal elements $aa$ ($1 \leq a \leq M$) defined by $1/\sqrt{r_i}$ for $M_{i-1} < a \leq M_i$.)

3.3 GLS-estimation

GLS-estimation is usually applied in the (not very likely) situation where the covariance matrix $\Sigma$ is known but not the regression coefficients. In this case the GLS-estimators are BLUE and outperform the OLS-estimators. Let

$$\begin{cases}
\alpha_i := \Sigma^{-1}_{(i-1)(i-1)} \Sigma_{(i-1)i} \in \mathbb{R}^{M_{i-1} \times m_i} \\
\zeta_{ti} := \alpha'_i \varepsilon_{ti(i-1)} \in \mathbb{R}^{m_i \times 1} \\
\eta_{ti} := \varepsilon_{ti} - \zeta_{ti} \in \mathbb{R}^{m_i \times 1} \\
\nu_{ti} := \mu_{ti} + \zeta_{ti} \in \mathbb{R}^{m_i \times 1}.
\end{cases} \quad (3.7)$$
Note that \( Y_{t(0)} = \varepsilon_{t(0)} = 0 \), so \( \zeta_{t1} = 0 \), \( \eta_{t1} = \varepsilon_{t1} \) and \( \nu_{t1} = \mu_{t1} \). Then \( \eta_{t1}, \ldots, \eta_{tr} \) are uncorrelated, \( \eta_{ti} \) and \( \nu_{ti} \) are uncorrelated and
\[
\begin{align*}
E(\zeta_{ti}) &= E(\eta_{ti}) = 0, \\
\Delta_{ii} &:= \text{Cov}(\zeta_{ti}) = \alpha_i \Sigma_{(i-1)(i-1)}^{-1} \alpha_i, \\
\Gamma_{ii} &:= \text{Cov}(\eta_{ti}) = \Sigma_{ii} - \Delta_{ii}.
\end{align*}
\tag{3.8}
\]
From the definitions it is clear that in case of normality
\[
\begin{align*}
\nu_{ti} &= E(Y_{ti}|Y_{t(i-1)}) = \mu_{ti(i-1)}, \\
\Gamma_{ii} &= \text{Cov}(Y_{ti}|Y_{t(i-1)}) = \Sigma_{ii(i-1)}.
\end{align*}
\tag{3.9}
\]
To avoid confusion with the block notation, the (more common) latter symbols will not be used. Clearly
\[
\Sigma_{(i)(i)}^{-1} = \begin{bmatrix}
\Sigma_{(i-1)(i-1)}^{-1} + \alpha_i \Gamma_{ii}^{-1} \alpha_i & -\alpha_i \Gamma_{ii}^{-1} \\
-\alpha_i \Gamma_{ii}^{-1} & \Gamma_{ii}^{-1}
\end{bmatrix}
\]
and so
\[
\varepsilon_{t(i)}' \Sigma_{(i)(i)}^{-1} \varepsilon_{t(i)} = \varepsilon_{t(i-1)}' \Sigma_{(i-1)(i-1)}^{-1} \varepsilon_{t(i-1)} + \eta_{ti} \Gamma_{ii}^{-1} \eta_{ti}.
\]
Therefore, the GLS-criterion (2.5) can be written as
\[
\sum_{i=1}^{r} \sum_{t=1}^{N_i} \eta_{ti} \Gamma_{ii}^{-1} \eta_{ti}.
\tag{3.10}
\]
For the \( \zeta_{ti}, \eta_{ti} \) and \( \nu_{ti} \) we use the same block notation as for the \( Y_{ti} \) (and \( \mu_{ti}, \varepsilon_{ti} \)). So we get from (3.7) \( i = 1, \ldots, r \)
\[
\begin{align*}
Y_i &= \nu_i + \eta_i, \\
\nu_i &= \mu_i + \zeta_i, \\
\zeta_i &= \varepsilon_{(i-1)} \alpha_i, \\
\varepsilon_i &= \zeta_i + \eta_i.
\end{align*}
\tag{3.11}
\]
Now, the GLS-criterion (3.10) can be written as
\[
\begin{align*}
\sum_{i=1}^{r} \text{tr}(\Gamma_{ii}^{-1} \eta_{i} \eta_{i}') &= \sum_{i=1}^{r} \text{tr}(\Gamma_{ii}^{-1} (Y_i - \nu_i)'(Y_i - \nu_i)) \\
&= \sum_{i=1}^{r} \text{tr}(\Gamma_{ii}^{-1} (Y_i - Y_{(i-1)} \alpha_i + \mu_{(i-1)} \alpha_i - \mu_i)'(Y_i - Y_{(i-1)} \alpha_i + \mu_{(i-1)} \alpha_i - \mu_i)).
\end{align*}
\tag{3.12}
\]
Using the notation of Theorem 1, this leads to Theorem 2.
Theorem 2 The GLS-estimators for $\mu_i$ and $\varepsilon_i$, ($i = 1, ..., r$) are

\[
\begin{align*}
\tilde{\mu}_i &:= H_i(Y_i - Y_{(i-1)} \alpha_i + \tilde{\mu}_{(i-1)} \alpha_i) \\
\tilde{\varepsilon}_i &:= Y_i - \tilde{\mu}_i
\end{align*}
\]

and can be written as

\[
\begin{align*}
\tilde{\mu}_i &= Z_i - H_i \tilde{\zeta}_i, \quad \tilde{\zeta}_i := \tilde{\varepsilon}_{(i-1)} \alpha_i; \\
\tilde{\varepsilon}_i &= E_i + H_i \tilde{\zeta}_i, \quad \tilde{\eta}_i := \tilde{\varepsilon}_i - \tilde{\zeta}_i.
\end{align*}
\]

Proof. The GLS-criterion (3.12) is a summation over all groups. Clearly the mean $\mu_i$ not only appears in the $i^{th}$ term but also in all subsequent terms. So minimization of (3.12) has to take place in a sequential way, starting with group $r$. The following orthogonal decomposition of $\eta_i$ is essential:

\[
Y_i - Y_{(i-1)} \alpha_i + \mu_{(i-1)} \alpha_i - \mu_i = \\
H_i(Y_i - Y_{(i-1)} \alpha_i + \mu_{(i-1)} \alpha_i - \mu_i) + U_i(Y_i - Y_{(i-1)} \alpha_i + \mu_{(i-1)} \alpha_i - \mu_i) = \\
H_i(Y_i - Y_{(i-1)} \alpha_i + \mu_{(i-1)} \alpha_i - \mu_i) + U_i(Y_i - Y_{(i-1)} \alpha_i).
\]

Regardless of the value of $\Gamma_{rr}$ the first term of the orthogonal decomposition is zero for $\tilde{\mu}_r$ given by (3.13) and therefore $\tilde{\mu}_r$ is the GLS-estimator for $\mu_r$. As a consequence of the substitution of the GLS-estimator in (3.12), $\mu_{r-1}$ only appears in the $(r-1)^{th}$ term, etcetera. Since $Y_{(i-1)} = \mu_{(i-1)} = 0$ for $i = 1$, repeated application of the preceding argumentation results in the closed form GLS-estimators (3.13). Relation (3.14) follows directly from (3.13) and the definitions of the OLS-estimators (3.4).

From expression (3.13), it is clear that the GLS-estimates have to be determined sequentially, i.e. only after the GLS-estimates for group $i - 1$ are determined, it is possible to determine the estimates for group $i$. So the ML-estimators in the proof of Theorem 2 are derived sequentially starting with the last group, while the actual estimates are determined sequentially starting with the first group.

The GLS-estimators $\tilde{\beta}_i$ for $\beta_i$ are given by

\[
\tilde{\beta}_i = G_i X_i'(Y_i - \tilde{\zeta}_i).
\]

The GLS-estimators $\tilde{\mu}_i$ are BLUE. So the $\tilde{\beta}_i$ are BLUE for estimable $\beta_i$.

The achieved minimum of the GLS-criterion (2.5), (3.10) or (3.12) is

\[
\sum_{i=1}^{r} \text{tr}(\Gamma_{ii}^{-1} \tilde{\eta}_i' \tilde{\eta}_i).
\]
3.4 EGLS-estimation

In the more common situation in which both the regression coefficients and the covariance matrix are unknown, EGLS is often applied. For EGLS we have to estimate $\Sigma$, or equivalently, the $(\alpha_i, \Gamma_{ii})$ $i = 1, \ldots, r$. From the expressions (3.13) for the GLS-estimators $\tilde{\mu}_i$ and $\tilde{\epsilon}_i$ it is clear that they depend on the $\alpha_i$ but not on the $\Gamma_{ii}$. Therefore only the EGLS-estimators $\hat{\alpha}_i$ for the $\alpha_i$ are relevant for the EGLS-estimators $\hat{\mu}_i$ and $\hat{\epsilon}_i$ for $\mu_i$ and $\epsilon_i$, respectively. We choose the $\hat{\alpha}_i$ that minimize (3.16), which is equivalent to minimizing (3.12) simultaneously to $\alpha_i$ and $\beta_i$. This minimization takes place by orthogonal projection onto extended spaces $L(i) \supset L_i$. We define

$$
\begin{align*}
L(i) &= \mathcal{R}(X_i Y(i-1)) = L_i \oplus \mathcal{R}(Y(i-1)), \ (\text{with } Y(0) := 0) \\
H(i) &\in \mathbb{R}^{N_i \times N_i} : \text{orthogonal projection matrix of } L(i) \\
U(i) &= I_{N_i} - H(i) : \text{orthogonal projection matrix of } L(i) \\
l(i) &= \dim(L(i)), \ r(i) = \dim(L(i) \cap \ker(q)) = N_i - l(i).
\end{align*}
$$

(3.17)

\textbf{Theorem 3} The EGLS-estimators for $\nu_i$ and $\eta_i$ ($i = 1, \ldots, r$) are

$$
\hat{\nu}_i := H(i) Y_i, \quad \hat{\eta}_i := Y_i - \hat{\nu}_i = U(i) Y_i.
$$

(3.18)

\textbf{Proof.} The EGLS-estimators for $\nu_i$ and $\eta_i$ follow straightforwardly from the orthogonal decomposition (compare the proof of Theorem 2):

$$
Y_i - \nu_i = \ H(i)(Y_i - \nu_i) + U(i)(Y_i - \nu_i) = (H(i)Y_i - \nu_i) + U(i)Y_i.
$$

So the EGLS-estimators are given by (3.18) regardless of $\Gamma_{ii}$. \hfill \Box

Similar to GLS, the EGLS-estimation procedure takes place sequentially, starting with group 1.

Since $Y(i-1) = \mu(i-1) + \tilde{\epsilon}(i-1)$ and $\mathcal{R}(X_i Y(i-1)) = \mathcal{R}(X_i \tilde{\epsilon}(i-1))$, the EGLS-estimators $\hat{\beta}_i$ and $\hat{\alpha}_i$ for $\beta_i$ and $\alpha_i$, respectively, are given by

$$
\begin{bmatrix}
\hat{\beta}_i \\
\hat{\alpha}_i
\end{bmatrix} = \begin{bmatrix}
X_i' \\
\tilde{\epsilon}_i(i-1)
\end{bmatrix} \begin{bmatrix}
Y_i \\
X_i \tilde{\epsilon}_i(i-1)
\end{bmatrix}, \ \text{with } G(i) = \begin{bmatrix}
X_i' X_i & X_i \tilde{\epsilon}_i(i-1) \\
\tilde{\epsilon}_i(i-1)' X_i & \tilde{\epsilon}_i(i-1)' \tilde{\epsilon}_i(i-1)
\end{bmatrix}^{-1}.
$$

(3.19)

Since $\tilde{\epsilon}_i(0) = 0$, we can always take $\hat{\beta}_1 = b_1$ given by (3.5).

\textbf{Theorem 4} The EGLS-estimators for $\mu_i$, $\xi_i$ and $\epsilon_i$ are

$$
\begin{align*}
\hat{\mu}_i &= Z_i - H_i \hat{\xi}_i, \quad \hat{\xi}_i = \tilde{\epsilon}(i-1) \hat{\alpha}_i, \quad \hat{\epsilon}_i = E_i + H_i \hat{\xi}_i.
\end{align*}
$$

(3.20)
Proof. EGLS corresponds to partial regression of $Y_i$ on $X_i$ and $\tilde{\varepsilon}_{(i-1)}$. The following orthogonal decomposition corresponds to partial regression:

$$Y_i - Y_{(i-1)}\alpha_i + \mu_{(i-1)}\alpha_i - \mu_i = H_i(Y_i - Y_{(i-1)}\alpha_i + \mu_{(i-1)}\alpha_i) + (U_i - U_{(i)}) (Y_i - Y_{(i-1)}\alpha_i + \mu_{(i-1)}\alpha_i) + U_{(i)} Y_i.$$ 

Following a similar reasoning as in the proof of Theorem 2, relation (3.20) follows from this orthogonal decomposition. \[\] 

The geometrical interpretations and the underlying relations of the OLS and EGLS-estimators are shown in Figure 1.

**Figure 1:** Geometrical interpretation

The fit $Z_i$ and the residuals $E_i$ of OLS are the (columnwise) orthogonal projections of $Y_i$ on $\mathcal{R}(X_i)$ and $\mathcal{R}(X_i)^\perp$, respectively. In EGLS, the fit $\hat{\mu}_i$ is the orthogonal projection of $Y_i$ on $\mathcal{R}(X_i, \tilde{\varepsilon}_{(i-1)})$ with residuals $(\hat{\eta}_i)_{h} \perp \mathcal{R}(X_i, \tilde{\varepsilon}_{(i-1)})$. Figure 1 illustrates that $Z_i$ and $\hat{\mu}_i$ (and therefore $E_i$ and $\tilde{\varepsilon}_i$) coincide when $\mathcal{R}(\tilde{\varepsilon}_{(i-1)}) \subset \mathcal{R}(X_i)^\perp$. Note that the equality $\tilde{\varepsilon}_i = E_i$ only holds if $X_i$ and $\tilde{\varepsilon}_{(i-1)}$ are orthogonal; this is in general not the case.

Similar to OLS, we build the EGLS-estimator $\hat{S}$ for the covariance matrix as the sample variance corrected for degrees of freedom, i.e.

$$
\begin{cases}
\hat{S}_{ii} = \frac{\varepsilon_i^2 \tilde{\varepsilon}_i}{r_i} \\
(\hat{S}_{(i-1)})_g = (\varepsilon_i^2 \tilde{\varepsilon}_{(i-1)})_g / \sqrt{r_g r_i} & \text{for } g = 1, \ldots, i-1.
\end{cases}
$$

(3.21)
Again, we have $\hat{S} \geq 0$. To ensure $\hat{S} > 0$, we may impose the regularity condition $N_r \geq M_r + l_r$. As a consequence $r(\epsilon_{(i-1)}) = M_{i-1}$ a.s., $l(i) = l_i + M_{i-1}$ a.s. and the estimates $\hat{\alpha}_i$ for the regression coefficients $\alpha_i$ are unique a.s.

### 3.5 Maximum likelihood

For maximum likelihood estimation we make the additional assumption that the error terms $\epsilon_{ti}$ have (simultaneously for all $t$ and $i$) a normal distribution. It follows that,

$$L(Y_{t(i-1)} Y_{ti}) = N_M \left( \left( \begin{array}{c} \mu_{t(i-1)} \\ \mu_{ti} \end{array} \right), \left( \begin{array}{cc} \Sigma_{(i-1)(i-1)} & \Sigma_{(i-1)i} \\ \Sigma_{i(i-1)} & \Sigma_{ii} \end{array} \right) \right).$$

The distribution of the observations is characterized by the unknown parameter $\theta = (\beta, \Sigma) \in \Theta$. We write $|A| = \det(A)$.

**Theorem 5** The likelihood of the observations $Y$ is given by

$$L(\theta; Y) = \exp\left\{ -\frac{1}{2} \sum_{i=1}^{r} tr(\Gamma^{-1}_{ii}(\hat{\nu}_i - \nu_i)'(\hat{\nu}_i - \nu_i)) \right\} \prod_{i=1}^{r} \left( \frac{(2\pi)^m |\Gamma_{ii}|^{-\frac{1}{2}}}{\sqrt{\pi}} \exp\left\{ -\frac{1}{2} tr(\Gamma^{-1}_{ii}\hat{\eta}_i\hat{\eta}_i) \right\} \right).$$

(3.22)

**Proof**

$$L(\theta; Y) \overset{1}{=} \prod_{i=1}^{r} \prod_{t=1}^{N_i} p(Y_{ti}|Y_{t(i-1)})$$

$$\overset{2}{=} \prod_{i=1}^{r} \left( \frac{(2\pi)^m |\Gamma_{ii}|^{-\frac{1}{2}}}{\sqrt{\pi}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{N_i} (Y_{ti} - \nu_{ti})\Gamma^{-1}_{ii}(Y_{ti} - \nu_{ti}) \right\} \right)$$

$$\overset{3}{=} \prod_{i=1}^{r} \left( \frac{(2\pi)^m |\Gamma_{ii}|^{-\frac{1}{2}}}{\sqrt{\pi}} \exp\left\{ -\frac{1}{2} tr(\Gamma^{-1}_{ii}(Y_i - \nu_i)'(Y_i - \nu_i)) \right\} \right)$$

$$\overset{4}{=} \prod_{i=1}^{r} \left( \frac{(2\pi)^m |\Gamma_{ii}|^{-\frac{1}{2}}}{\sqrt{\pi}} \exp\left\{ -\frac{1}{2} tr(\Gamma^{-1}_{ii}(\hat{\nu}_i - \nu_i)'(\hat{\nu}_i - \nu_i)) - \frac{1}{2} tr(\Gamma^{-1}_{ii}\hat{\eta}_i\hat{\eta}_i) \right\} \right).$$

Equality 1 holds by conditioning; note that $Y_{t(0)} = 0$. Given $Y_{t(i-1)}$, $\nu_{ti}$ is fixed and (3.9) implies $L(Y_{ti}|Y_{t(i-1)}) = N_{m_i}(\nu_{ti}, \Gamma_{ii})$; because of the row independence, the conditional densities can be substituted into the likelihood which results in equality 2. Equality 3 is obtained by writing the likelihood in terms of matrices $Y_i$ instead of the columns $Y_{ti}$. The fourth equality is based on the orthogonal decomposition of $Y_i$ in $\hat{\nu}_i$ and $\hat{\eta}_i$ (according to (3.18)). Since $\hat{\eta}_i$ is the orthogonal projection of $Y_i$ onto $L_{(i)}^\perp$, $\hat{\eta}_i$ is orthogonal to both $\hat{\nu}_i$ and $\nu_i$.

From the likelihood (3.22) we see that $\hat{\nu}_1, ..., \hat{\nu}_r, \hat{\eta}_1, ..., \hat{\eta}_r, \hat{\eta}_r$ form a sufficient statistic. So the MLE can be expressed in terms of these statistics.
In case of known $\Sigma$, it is clear from the proof of Theorem 5 that maximization of the likelihood (3.22) coincides with minimization of the GLS-criterion (3.12) and that the MLE will coincide with the GLS-estimators. So in case of normality, the GLS-estimators are MVUE.

We will now discuss the more interesting situation of ML-estimation in case of unknown covariance matrix.

**Theorem 6** The MLE for $\nu_i$ and $\Gamma_{ii}$ are $\hat{\nu}_i$ in (3.18) and

$$\hat{\Gamma}_{ii} = \frac{\hat{n}_i^2}{N_i}$$

respectively. The maximized likelihood is given by

$$\sup_{\theta \in \Theta} L(\theta; Y) = (2\pi)^{-\frac{1}{2} \sum_{i=1}^{r} N_i m_i} \prod_{i=1}^{r} \left[ \frac{\hat{n}_i^2}{N_i} \right]^{-\frac{N_i}{2}} \exp \left\{ \frac{1}{2} (\Gamma_{ii}^{-1} \hat{n}_i) \right\}.$$  \( (3.24) \)

**Proof.** The MLE are obtained by maximization of the likelihood (3.22) w.r.t. all $\nu_i$ and $\Gamma_{ii}$, respectively. Now (3.22) is maximized by $\nu_i = \hat{\nu}_i$, regardless the value of $\Gamma_{ii}$. Therefore $\hat{\nu}_i$ is the MLE for $\nu_i$, even in case of unknown $\Gamma_{ii}$. Substitution of $\hat{\nu}_i$ in (3.22) gives

$$\sup_{\nu_i} L(\theta; Y) = \prod_{i=1}^{r} \left\{ \{(2\pi)^{m_i} |\Gamma_{ii}|\}^{-\frac{N_i}{2}} \exp \left\{ \frac{1}{2} (\Gamma_{ii}^{-1} \hat{n}_i^2) \right\} \right\}.$$  \( (3.24) \)

This has to be maximized w.r.t the $\Gamma_{ii}$. Matrix differentiation of the separate $r$ factors, renders $\hat{\Gamma}_{ii} = \frac{\hat{n}_i^2}{N_i}$ as MLE. Substitution of the $\hat{\nu}_i$ and $\hat{n}_i$ into (3.22) results in (3.24).

Since the MLE for $\nu_i$ coincides with the EGLS-estimator $\hat{\nu}_i$, the MLE for $\mu_i$, $\zeta_i$ and $\varepsilon_i$ coincide with the EGLS-estimators $\hat{\mu}_i$, $\hat{\zeta}_i$ and $\hat{\varepsilon}_i$. So in case of identifiable $\alpha_i$ and $\beta_i$, the EGLS-estimators $\hat{\beta}_i$ and $\hat{\alpha}_i$ equal the MLE. Though the regression coefficients $\alpha_i$ are always identifiable, this is not true for $\beta_i$. In case of non-unique $\hat{\beta}_i$ we choose the MLE equal to the EGLS-estimators for $\beta_i$.

The MLE $\hat{\Sigma}$ for the covariance matrix follows sequentially from the relations (3.7) and (3.8), and from the MLE (3.23):

$$\left\{ \begin{array}{l} \hat{\Sigma}_{11} = \hat{\Gamma}_{11} \\
\hat{\Sigma}_{(i-1)i} = \hat{\Sigma}_{(i-1)(i-1)} \hat{\alpha}_i, \quad \hat{\Delta}_{ii} = \alpha_i^2 \hat{\Sigma}_{(i-1)(i-1)} \hat{\alpha}_i, \quad \hat{\Sigma}_{ii} = \hat{\Gamma}_{ii} + \hat{\Delta}_{ii} \end{array} \right. \text{ for } i = 2, \ldots, r. \quad (3.25)$$

Note that the difference between the estimators $\hat{\Sigma}$ and $\hat{S}$ in (3.21) is not just caused by the introduction of the number of degrees of freedom. E.g. from the expressions

$$\hat{\Sigma}_{22} = \hat{n}_2^2 / N_2 + \alpha_2^2 \hat{\varepsilon}_1 \hat{\alpha}_2 / N_1, \quad \hat{S}_{22} = \hat{n}_2^2 / r_2 + \alpha_2^2 \hat{\varepsilon}_1^2 \hat{\alpha}_2 / r_2$$
we see that the difference is caused by taking other residuals as well.

It is not straightforward which one of the covariance estimators $S$, $\hat{S}$ or $\hat{\Sigma}$ has the smallest bias. The bias of $\hat{\Sigma}$ will probably be decreased by correcting for the degrees of freedom. Replacing $N_i$ by $r_i$ in (3.23) gives an unbiased estimator for $\Gamma_{ii}$; $\Sigma$ can still be estimated according to relation (3.25). A major drawback of this correction is that the estimator for $\Sigma$ depends on the particular division of the data into groups, even in case of no missing observations. This problem is solved by substituting $r_i$ for $N_i$ in (3.23) and still estimating $\Sigma$ by relation (3.25). Though this does not result in an unbiased estimator for $\Gamma_{ii}$, the estimator for $\Sigma$ is unique in case of no missing observations and the bias of this estimator is probably smaller than the bias of the MLE $\hat{\Sigma}$. The construction of an unbiased estimator for $\Sigma$ and the analysis of the bias of the current covariance estimators $S$, $\hat{S}$ and $\hat{\Sigma}$ are left for future research. In this paper, we restrict ourselves to (3.6), (3.21) and (3.25).

4 No missing observations

In the model formulation of Section 2 the restrictions $N_{i-1} \geq N_i$ are imposed instead of $N_{i-1} > N_i$. In case of the last restrictions the division of the data into different group is always unique, while this is not true for the first restrictions: if there are several variables with the same number of observations, all the variables together can be defined as one group, but it is also possible to define multiple groups. In case of different groups with the same number of observations, the $\hat{\varepsilon}_j$ of the previous dependent variables with the same number of observations as the dependent variables of group $i$ are orthogonal to $X_i$. Since the regression of $Y_i$ onto the $X_i$ and $\hat{\varepsilon}_{(i-1)}$ coincides with partial regression (see e.g. Greene (1993)), the estimators $\hat{\mu}_i$ and $\hat{\varepsilon}_i$ will not depend on the group composition.

The situation with no missing observations ($N = N_1 = ... = N_r$) is a special case of the presented model. By constructing just one group, it is straightforward that the OLS- and (E)GLS-estimators for $\mu_i$ are identical: $Z_i = \hat{\mu}_i$. As a consequence the covariance estimators (3.6) and (3.21) are identical and unique.

The uniqueness and equality of the OLS- and EGLS-estimators can also be shown sequentially by the estimation procedure. From both Figure 1 and formula (3.19) for the regression coefficients, we can see that the OLS- and EGLS-estimators are identical when $\mathcal{R}(\hat{\varepsilon}_{(i-1)}) \subset \mathcal{R}(X_i)^{\perp}$. That this is true for the situation with no missing observations can be directly deduced from the estimation procedure. In case of complete data, we have $X = X_1 = X_2 = ... = X_r$ and $\mathcal{R}(X_i) = \mathcal{R}(X)$ for $i = 1, ..., r$. The iterations in the
EGLS-estimation procedure show

Step 1: \( \hat{\mu}_1 \in \mathcal{R}(X) \), \( \hat{\varepsilon}_1 \in \mathcal{R}(X)^\perp \)

Step \( i \) (\( i = 2, \ldots, r \)): \( \hat{\varepsilon}_{(i-1)} = [\hat{\varepsilon}_1 \hat{\varepsilon}_2 \ldots \hat{\varepsilon}_{i-1}] \in \mathcal{R}(X)^\perp \)
\( \Rightarrow \hat{\varepsilon}_i = \hat{\varepsilon}_{(i-1)} \hat{\alpha}_i \in \mathcal{R}(\hat{\varepsilon}_{(i-1)}) \subset \mathcal{R}(X)^\perp \) and \( \hat{\eta}_i \in \mathcal{R}(X)^\perp \)
\( \Rightarrow \hat{\varepsilon}_i = \hat{\varepsilon}_i + \hat{\eta}_i \in \mathcal{R}(X)^\perp \)
\( \Rightarrow \mathcal{R}(\hat{\varepsilon}_{(i)}) = \mathcal{R}(\hat{\varepsilon}_1 \hat{\varepsilon}_2 \ldots \hat{\varepsilon}_i) \subset \mathcal{R}(X)^\perp \).

So \( \hat{\varepsilon}_{(i-1)} \in \mathcal{R}(X)^\perp \), \( Z_i = \hat{\mu}_i \) and as a consequence \( S = \hat{S} \).

For the case of complete data, the MLE in Theorem 6 must be identical to the standard result known from literature, as well as the maximized likelihood (3.24). To show the latter we make use of the following two properties:

\[
\begin{cases}
(a) \hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_r \text{ are orthogonal} \\
(b) E_{(r)} = [E_1 E_2 \ldots E_r] = [\hat{\varepsilon}_1 \hat{\varepsilon}_2 \ldots \hat{\varepsilon}_r] = [\hat{\eta}_1 \hat{\eta}_2 \ldots \hat{\eta}_r]A,
\end{cases}
\]

with \( A \) an upper triangular invertible matrix with unit diagonal elements. The existence of such a matrix follows form (a slightly modified) QR-factorization of \( E_{(r)} \). Hence,

\[
\sup_{\theta \in \Theta} L(\theta; Y) = (2\pi e)^{-\frac{1}{2}NM} \prod_{i=1}^r |\hat{\eta}_i^2/N|^{-\frac{N}{2}}
= (2\pi e)^{-\frac{1}{2}NM} |[\hat{\eta}_1 \hat{\eta}_2 \ldots \hat{\eta}_r] [\hat{\eta}_1 \hat{\eta}_2 \ldots \hat{\eta}_r]/N|^{-\frac{N}{2}}
= (2\pi e)^{-\frac{1}{2}NM} |[\hat{\eta}_1 \hat{\eta}_2 \ldots \hat{\eta}_r]A'([\hat{\eta}_1 \hat{\eta}_2 \ldots \hat{\eta}_r]A/N)^{-\frac{N}{2}}
= (2\pi e)^{-\frac{1}{2}NM} |E'_{(r)}E_{(r)}/N|^{-\frac{N}{2}}.
\]

The first equality follows from \( N = N_1 = \ldots = N_r \) and \( X = X_1 = \ldots = X_r \). The second and third equality are based on property (a) and \(|A| = 1\). The last equality follows from (b). The final expression can be found in Seber (1984), p. 407.

### 5 The constant term

Often the first explanatory variable is the constant term. We denote the corresponding regression coefficients by \( \beta_c \in \mathbb{R}^{1 \times M} \) (\( c = \text{constant} \)); the regression coefficients of the other explanatory variables are denoted by \( \beta_v \in \mathbb{R}^{(k-1) \times M} \) (\( v = \text{variable} \)). Expanding this notation we can write

\[
\beta = \begin{bmatrix} \beta_c \\ \beta_v \end{bmatrix}, \quad b = \begin{bmatrix} b_c \\ b_v \end{bmatrix}, \quad \tilde{\beta} = \begin{bmatrix} \tilde{\beta}_c \\ \tilde{\beta}_v \end{bmatrix}, \quad X = [1_N X_v],
\]

with \( X_v \in \mathbb{R}^{N_1 \times (k-1)} \). The subindices \( i \) and \( (i-1) \) have a similar meaning as in the preceding sections, so for example, \( X_{vi} \) contains the first \( N_i \) rows of \( X_v \).
LS-estimation with the constant term corresponds to orthogonal projections on $\mathcal{R}(1_{N_i})$ and the centered spaces $\tilde{L}_i$ and $\tilde{L}_{(i)}$ defined as

$$
\begin{align*}
\tilde{L}_i + \mathcal{R}(1_{N_i}) &= L_i \\
\tilde{L}_{(i)} + \mathcal{R}(1_{N_i}) &= L_{(i)}
\end{align*}
$$

and $\tilde{L}_i \perp \mathcal{R}(1_{N_i})$, $\tilde{L}_{(i)} \perp \mathcal{R}(1_{N_i})$, $\tilde{l}_i = \dim(\tilde{L}_i) = l_i - 1$, $\tilde{l}_{(i)} = \dim(\tilde{L}_{(i)}) = l_{(i)} - 1$.

The mean and centered observations coincide with orthogonal projections of the observations on $\mathcal{R}(1_{N_i})$ and the centered spaces:

$$
\begin{align*}
\tilde{X}_{vi} &= \frac{1}{N_i} 1_{N_i}' X_{vi} \in \mathbb{R}^{1 \times k} \\
\tilde{Y}_i &= \frac{1}{N_i} 1_{N_i}' Y_i \in \mathbb{R}^{1 \times m_i} \\
\tilde{\Sigma}_{(i-1)} &= \frac{1}{N_i} 1_{N_i}' \Sigma_{(i-1)} \in \mathbb{R}^{1 \times m_{i-1}} \\
\tilde{\epsilon}_{(i-1)} &= \frac{1}{N_i} 1_{N_i}' \epsilon_{(i-1)} \in \mathbb{R}^{1 \times m_{i-1}}
\end{align*}
$$

Note that $\tilde{Y}_{(i-1)} \neq \tilde{Y}_1 \tilde{Y}_2 \ldots \tilde{Y}_{(i-1)}$ and $\tilde{\epsilon}_{(i-1)} \neq 0$.

The LS-estimators can be expressed in terms of the means and the centered observations, e.g. the EGLS-estimators (or equivalently the MLE in case of unknown $\Sigma$) read

$$
\begin{align*}
\begin{bmatrix}
\hat{\beta}_{vi} \\
\hat{\alpha}_i
\end{bmatrix}
= \tilde{G}_{(i)}
\begin{bmatrix}
\tilde{X}_{vi}' \\
\tilde{\epsilon}_{(i-1)}'
\end{bmatrix}
\tilde{Y}_i \\
\text{with}
\tilde{G}_{(i)}
\begin{bmatrix}
\tilde{X}_{vi}' \tilde{X}_{vi} \\
\tilde{\epsilon}_{(i-1)}' \tilde{X}_{vi} \\
\tilde{\epsilon}_{(i-1)}' \tilde{\epsilon}_{(i-1)}
\end{bmatrix}
\end{align*}
$$

(5.1)

The model with only the constant term $X_i = 1_{N_i}$ has received a lot of attention in literature, especially maximum likelihood estimation under the normality assumption. Anderson (1957) derived the MLE for $r = 2$ and $m_1 = m_2 = 1$ and suggested an approach to determine the MLE for general $r$. Bhargava (1962) derived the MLE for general $r$. Following the approach suggested by Anderson (1957), Afifi and Elashoff (1967) confirmed the findings of Bhargava (1962) for the regression coefficients, but presented a different, incorrect MLE for the covariance matrix. Jinadasa and Tracy (1992) derived the correct MLE for general $r$ by matrix differentiation which resulted in rather complicated expressions. Fujisawa (1995) presented the MLE for general $r$ in recursive form, which coincide with the MLE given by Bhargava (1962) and Jinadasa and Tracy (1995). For the model with only the constant term $\tilde{\epsilon}_{(i-1)}$ and $\tilde{\epsilon}_i$ coincide with $\tilde{Y}_{(i-1)}$ and $\tilde{Y}_i$ respectively, and the MLE (5.1) for the regression coefficients reduce to the same expressions as found by Fujisawa (1995):

$$
\begin{align*}
\hat{\beta}_{ci} &= \hat{\mu}_i = Y_1 \\
\hat{\alpha}_i &= (\tilde{Y}_{(i-1)}' \tilde{Y}_{(i-1)})^{-1} (\tilde{Y}_{(i-1)} \tilde{Y}_i) \\
\hat{\beta}_{ci} &= \hat{\mu}_i = Y_i - (\tilde{Y}_{(i-1)} - \hat{\mu}_{(i-1)}) \hat{\alpha}_i, \text{ for } i = 2, \ldots, r.
\end{align*}
$$
The MLE $\hat{\Gamma}_{ii}$ are determined by substituting the MLE for the regression coefficients into (3.23), leading to the same covariance estimators as found by Fujisawa (1995):

$$
\left\{ \begin{array}{l}
\hat{\Gamma}_{11} = \frac{\tilde{Y}_1\tilde{Y}_1}{N_i} \\
\hat{\Gamma}_{ii} = (\tilde{Y}_i - \tilde{Y}_{(i-1)}\hat{\alpha}_i)'(\tilde{Y}_i - \tilde{Y}_{(i-1)}\hat{\alpha}_i)/N_i \quad \text{for } i = 2, \ldots, r.
\end{array} \right.
$$

As an additional check we applied the estimation procedure to the data set considered by Little and Rubin (1987, p.118) and later by Jinadasa and Tracy (1995, p.43). Our estimates coincide exactly with their outcomes.

## 6 Restricted models

So far we just have considered (unrestricted) models in which $(\mu_i)_h \in L_i$ and $(\nu_i)_h \in L(i)$. In a restricted model, $p_i$ linear constraints are imposed on the parameters $\beta_i$ and $\alpha_i$.

To save space, for two matrices $P \in \mathbb{R}^{p \times c}$ and $Q \in \mathbb{R}^{q \times c}$ we will write $\begin{bmatrix} P \\ Q \end{bmatrix}$ as $[P; Q]$.

Hence the linear constraints can be written as $C_i[\beta_i; \alpha_i] = 0$ with $C_i \in \mathbb{R}^{p_i \times (k+M_i-1)}$ for $i = 1, \ldots, r$. So for $i = 1, \ldots, r$ the unknown $\beta_i$ and $\alpha_i$ are restricted to $\mathcal{N}(C_i)$, the null space of $C_i$. Then $\nu_i = [X_i Y_{(i-1)}][\beta_i; \alpha_i]$ is restricted to $\mathcal{R}(\langle X_i, Y_{(i-1)} \rangle(\mathcal{N}(C_i)))$, the image of $\mathcal{N}(C_i)$ under the linear transformation $(X_i, Y_{(i-1)})$. The linear space $L(i) = \mathcal{R}(X_i Y_{(i-1)})$ can be split into two orthogonal subspaces: $L_0(i)$ and $L_1(i)$, which (with some additional characteristics) are defined as

$$
\begin{align*}
L_0(i) & = \mathcal{R}(\langle X_i, Y_{(i-1)} \rangle(\mathcal{N}(C_i))), & L_1(i) & = L(i), & L_1(i) & \perp L_0(i) \\
H_{0(i)}: \text{projection matrix of } L_{0(i)}, & l_{0(i)} = \dim(L_{0(i)}) \\
H_{1(i)}: \text{projection matrix of } L_{1(i)}, & l_{1(i)} = \dim(L_{1(i)}) = l_{(i)} - l_{0(i)} \\
U_{0(i)}: \text{projection matrix of } L_{0(i)}^+, & L_{0(i)}^+ \perp L_{1(i)}^+ \\
& r_{0(i)} = \dim(L_{0(i)}^+) = N_i - l_{0(i)}.
\end{align*}
$$

So $L_{0(i)}^+ = L_{1(i)} \perp L_{0(i)}^+$. Quantities relating to $L_{0(i)}$ and $L_{1(i)}$ are denoted by a primary subindex 0 and 1, respectively. The following (identifiable) testing problem will be considered:

$$
\begin{align*}
H_0: \{ \forall i : C_i[\beta_i; \alpha_i] = 0 \} & \text{ against } H_1: \{ \exists i : C_i[\beta_i; \alpha_i] \neq 0 \}, \\
\text{or equivalently,} & \\
H_0: \{ \forall i, h : (\nu_i)_h \in L_{0(i)} \} & \text{ against } \quad H_1: \{ \exists i, h : (\nu_i)_h \in L_{1(i)}; \forall i, h : (\nu_i)_h \in L_{0(i)} \}.
\end{align*}
$$

The relevant test statistics for (6.1) can be based on orthogonal projections onto the $L_{1(i)}$ and $L_{0(i)}^+$. Note that the testing problem (6.1) includes not only restrictions on the
but also on the correlation of variables of different groups. The (usual) test with homogeneous constraints on the \( \beta_i \) is given by \( C_i = [C \ 0] \) with \( C \in \mathbb{R}^{p \times k} \).

Similar to the unrestricted model, we can distinguish between OLS- and (E)GLS-estimation. The orthogonal decompositions for OLS and (E)GLS are similar to the ones in the unrestricted model. Since the covariance matrix is in general unknown, GLS can not be applied. We will only discuss EGLS, since the GLS-estimators outperform the OLS-estimators and the EGLS-estimators are asymptotically equivalent to the GLS-estimators (though they of course do not share the same small sample properties). Moreover, the EGLS-estimators coincide with the MLE under the normality assumption.

The whole procedure for EGLS-estimation for the restricted model is similar to the one described in Section 3.4 for the unrestricted model: only the subspaces \( L_{(i)} \) have to be replaced by \( L_{0(i)} \). Formulae (3.18) and (3.20) through (3.25) still hold for the restricted model if we add a subindex 0. The estimators \( \hat{\beta}_{0i} \) and \( \hat{\alpha}_{0i} \) for \( \beta_i \) and \( \alpha_i \) respectively, are given (similar to (3.19)) by

\[
\begin{align*}
\hat{\beta}_{0i} &= G_{0i}X_i'Y_i, \text{ with } G_{0i} \in \mathbb{R}^{k \times k} \\
\text{defined by} & \begin{bmatrix} X_i'X_i & C_i' \\ C_i & 0 \end{bmatrix}^{-1} \\
\end{align*}

\text{for } i = 2, ..., r : \begin{align*}
\begin{bmatrix} \hat{\beta}_{0i} \\ \hat{\alpha}_{0i} \end{bmatrix} &= G_{0(i)} \begin{bmatrix} X_i' \\ \tilde{\varepsilon}'_{0(i-1)} \end{bmatrix} Y_i \text{ with } G_{0(i)} \in \mathbb{R}^{(k+M_{i-1}) \times (k+M_{i-1})} \\
\text{defined by} & \begin{bmatrix} X_i'X_i & X_i'\tilde{\varepsilon}_{0(i-1)} \\ \tilde{\varepsilon}'_{0(i-1)}X_i & \tilde{\varepsilon}'_{0(i-1)}\tilde{\varepsilon}_{0(i-1)} + C_i' \end{bmatrix}^{-1}. \\
\end{align*}

\[
\Lambda_{0i} = \frac{\hat{\eta}_i'\hat{\eta}_i}{\hat{\eta}_i'\hat{\eta}_i + \hat{\nu}_i'\hat{\nu}_i}
\]

The required statistics for the LR-test (based on EGLS) can be summarized into a collection of non-centered MANOVA-tables for \( i = 1, ..., r \). In the tables the abbreviations SS, DF and R stand for Sum of Squares, Degrees of Freedom and Restricted, respectively.

**Table 1:** Collection of non-centered MANOVA-tables \( (i = 1, ... r) \)

<table>
<thead>
<tr>
<th>Model</th>
<th>Space</th>
<th>SS</th>
<th>DF</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. model</td>
<td>( L_{0(i)} )</td>
<td>( \hat{\nu}<em>{0i}'\hat{\nu}</em>{0i} )</td>
<td>( l_{0(i)} )</td>
<td>( \text{Total} )</td>
</tr>
<tr>
<td>Difference</td>
<td>( L_{1(i)} )</td>
<td>( \hat{\nu}<em>{1i}'\hat{\nu}</em>{1i} )</td>
<td>( l_{1(i)} )</td>
<td>( \Lambda_{0i} = \frac{\hat{\eta}_i'\hat{\eta}_i}{\hat{\eta}_i'\hat{\eta}_i + \hat{\nu}_i'\hat{\nu}_i} )</td>
</tr>
<tr>
<td>Model</td>
<td>( L_{(i)} )</td>
<td>( \hat{\nu}<em>{(i)}'\hat{\nu}</em>{(i)} )</td>
<td>( l_{(i)} )</td>
<td>( \text{Total} )</td>
</tr>
<tr>
<td>Error</td>
<td>( \hat{\nu}<em>{(i)}'\hat{\nu}</em>{(i)} )</td>
<td>( \hat{\eta}<em>{(i)}'\hat{\eta}</em>{(i)} )</td>
<td>( r_{(i)} )</td>
<td>( \text{Total} )</td>
</tr>
<tr>
<td>Total</td>
<td>( \mathbb{R}^{N_i} )</td>
<td>( Y_i'Y_i )</td>
<td>( N_i )</td>
<td>( \text{Total} )</td>
</tr>
</tbody>
</table>
The column testing is mentioned here for future use; note that \( \hat{\eta}_i = U_{0(i)} Y_i = \hat{\nu}_{1i} + \hat{\eta}_i \).

If the constant term is included as an explanatory variable, often the centered MANOVA-tables are presented, provided that no restrictions are imposed on the constant term. The abbreviation C stands for Corrected (or Centered):

**Table 2**: Collection of centered MANOVA-tables \( (i = 1, \ldots, r) \)

<table>
<thead>
<tr>
<th>Model</th>
<th>Space</th>
<th>SS</th>
<th>DF</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.R. model</td>
<td>( L_{0(i)} )</td>
<td>( \nu_0' \hat{\nu}_0 )</td>
<td>( l_{0(i)} )</td>
<td>( \Lambda_{0i} = \frac{</td>
</tr>
<tr>
<td>Difference</td>
<td>( L_{1(i)} )</td>
<td>( \nu_{1i}' \hat{\nu}_{1i} )</td>
<td>( l_{1(i)} )</td>
<td></td>
</tr>
<tr>
<td>C. model</td>
<td>( L_i )</td>
<td>( \hat{\nu}'_i \hat{\eta}_i )</td>
<td>( l_i )</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>( L_i^\perp )</td>
<td>( \hat{\eta}'_i \hat{\eta}_i )</td>
<td>( r_i )</td>
<td></td>
</tr>
<tr>
<td>C. total</td>
<td>( \mathcal{R}(1_{N_i})^\perp )</td>
<td>( Y_i' Y_i )</td>
<td>( N_i - 1 )</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>( \mathcal{R}(1_{N_i}) )</td>
<td>( N_i Y_i' Y_i )</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( \mathcal{R}^{N_i} )</td>
<td>( Y_i' Y_i )</td>
<td>( N_i )</td>
<td></td>
</tr>
</tbody>
</table>

The inner products in the non-centered MANOVA-tables are acquired by adding the inner products of the corresponding means to the centered inner products, e.g. \( \hat{\nu}'_i \hat{\nu}_i = \hat{\nu}'_i \hat{\nu}_i + N_i Y_i' Y_i \). Since the terms \( \hat{\nu}_{1i} \) and the errors \( \hat{\eta}_i \) in the non-centered MANOVA-tables are centered if a constant is included in the model, they are identical to the corresponding inner products in the centered MANOVA-tables.

Linear restrictions are often imposed to solve identification problems or to test or estimate submodels. So suppose (not necessary identifiable) linear restrictions \( C_i[\beta; \alpha] = 0 \) have already been imposed and \( q_i \) additional linear constraints are considered of the form \( D_i[\beta; \alpha] = 0 \) with \( D_i \in \mathbb{R}^{q_i \times (k+M_i-1)} \). Then the unknown \( \beta_i \) and \( \alpha_i \) are restricted to \( \mathcal{N}([C_i; D_i]) \), the null space of \( [C_i; D_i] \). Similar to the (single) restricted model, the linear space \( L_{0(i)} \) can be split into the subspaces \( L_{00(i)} = \mathcal{R}((X_i Y_{i-1}),(\mathcal{N}([C_i; D_i]))) \) and \( L_{01(i)} \), the orthogonal complement of \( L_{00(i)} \) w.r.t. \( L_{0(i)} \). The test statistics for (identifiable)

\[
\begin{array}{c}
H_{00} : \{ \forall i : C_i[\beta_i; \alpha_i] = D_i[\beta_i; \alpha_i] = 0 \} \text{ against } \\
H_{01} : \{ \exists i : D_i[\beta_i; \alpha_i] \neq 0 \wedge \forall i : C_i[\beta_i; \alpha_i] = 0 \}
\end{array}
\]

or equivalently,

\[
\begin{array}{c}
H_{00} : \{ \forall i, h : (\nu_i)_h \in L_{00(i)} \} \text{ against } \\
H_{01} : \{ \exists i, h : (\nu_i)_h \in L_{00(i)} - L_{00(i)}; \forall i, h : (\nu_i)_h \in L_{00(i)} \}
\end{array}
\]  

(6.3)

can be based on orthogonal projections onto the \( L_{01(i)} \) and \( L_{00(i)}^\perp \). The estimation procedure of the preceding sections can again be applied to the double restricted model.
similar as to the restricted model. For estimation under the (not necessarily identifiable) double restrictions \([C_i; D_i][\beta_i; \alpha_i] = 0 \forall i\), we can use again (6.2) with \(C_i\) replaced by \([C_i; D_i]\).

All information of the unrestricted, restricted and double restricted models required for the described tests can be summarized in combined centered MANOVA-tables for \(i = 1, \ldots, r\), assuming that the model contains the constant as an explanatory variable and that no restrictions are imposed on this constant. This combined centered MANOVA-table can be obtained by adding Table 3 to the top of the centered MANOVA-table in Table 2. Here D stands for double:

### Table 3: Double restricted centered inner products \((i = 1, \ldots, r)\)

<table>
<thead>
<tr>
<th>Model</th>
<th>Space</th>
<th>SS</th>
<th>DF</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. D. Restricted model</td>
<td>(L_{00(i)})</td>
<td>(\tilde{\nu}<em>{00i}\tilde{\nu}</em>{00i})</td>
<td>(l_{00(i)})</td>
<td>(\Lambda_{00i} = \frac{</td>
</tr>
<tr>
<td>Difference</td>
<td>(L_{01(i)})</td>
<td>(\tilde{\nu}<em>{01i}\tilde{\nu}</em>{01i})</td>
<td>(l_{01(i)})</td>
<td></td>
</tr>
</tbody>
</table>

From Tables 1, 2 and 3 relations between the unrestricted, restricted and double restricted statistics can be deduced such as \(\tilde{\eta}_{00i}\tilde{\eta}_{00i} = \tilde{\eta}_{0i}\tilde{\eta}_{0i} + \tilde{\nu}_{01i}\tilde{\nu}_{01i}\). This relation will be used in Section 8.

### 7 Some distributions and orthogonal projections

We define the Wishart distribution \(W_d\) as follows: let \(Y = [Y_1 \ldots Y_n]'\) and \(\mu = [\mu_1 \ldots \mu_n]'\) with independent \(Y_i \sim N_d(\mu_i, \Sigma), \Sigma \geq 0\). Then

\[
W = Y'Y = \sum_{i=1}^{n} Y_iY_i' \sim W_d(n, \Sigma; \Delta) \quad \text{(with } \Delta = \mu'\mu)\]

where \(W_d(n, \Sigma; \Delta)\) denotes the noncentral Wishart distribution with dimension \(d\), degrees of freedom \(n\), dispersion matrix \(\Sigma\) and non-centrality matrix \(\Delta\). The central Wishart distribution is \(W_d(n, \Sigma) = W_d(n, \Sigma; 0)\). The standard Wishart distribution is \(W_d(n) = W_d(n, I_d)\). Our notation is the same as the one of Gupta and Nagar (2000), except for the non-centrality matrix which they define as \(\Theta = \Sigma^{-1}\Delta\) for nonsingular \(\Sigma\).

The properties of the projections follow from the following projection theorem (compare Gupta & Nagar (2000), Theorems 7.8.3 and 7.8.5).
Theorem 7 Let $L_0$ and $L_1$ be linear subspaces of $\mathbb{R}^n$ with $L_0 \perp L_1$. Denote the orthogonal projection matrices of $L_0$ and $L_1$ by $P_0$ respectively $P_1$ and let $l_0 = \text{dim}(L_0)$. Then, for $Y' = [Y_1 \ldots Y_n] \in \mathbb{R}^{d \times n}$, with uncorrelated $Y_i$, $\text{Cov}(Y_i) = \Sigma$ and $E(Y) = \mu$, 

$$
\begin{cases}
P_0 Y \text{ and } P_1 Y \text{ are uncorrelated}, \\
E(P_0 Y) = P_0 \mu, \\
\text{Cov}(\text{vec}(P_0 Y)) = \Sigma \otimes P_0.
\end{cases}
$$

If in addition the $Y_i$ are normally distributed, then 

$$
\begin{cases}
P_0 Y \text{ and } P_1 Y \text{ are independent}, \\
Y' P_0 Y \sim W_d(l_0, \Sigma; \mu' P_0 \mu).
\end{cases}
$$

In the next section a generalization of the Wilks’ distribution is used. For the (usual) Wilks’ distribution we follow the same notation as e.g. Rencher (1998): let $B \sim W_d(s)$, $C \sim W_d(t)$, $B$ and $C$ independent. Then 

$$
\Lambda = \frac{|B|}{|B + C|} \sim \Lambda_{d,t,s},
$$

where $\Lambda_{d,t,s}$ denotes the Wilks’ distribution with parameters $d$, $t$ and $s$. We define the generalized Wilks’ distribution $\Lambda_{A,D,T,S}$ with parameters $A$, $D$, $T$ and $S \in \mathbb{R}^{1 \times r}$ as follows: let $\Lambda_i \sim \Lambda_{d_i,t_i,s_i}$ be independent and $a_i \in [0, 1]$ with $a_1 = 1$. Then

$$
\prod_{i=1}^r \Lambda_i^{a_i} \sim \Lambda_{A,D,T,S}.
$$

The vector $A$ contains the exponents $a_i$ of the separate factors as elements, $D$ the $d_i$, $T$ the $t_i$ and $S$ the $s_i$ ($i = 1, \ldots, r$).

8 Testing

We assume normally distributed errors now. From the projection Theorem 7 (applied to $L_{(i)}$ and $L_{(i)}^\perp$) we get the following conditional properties given $Y_{(i-1)}$:

$$
\begin{cases}
\hat{\nu}_i \text{ and } \hat{\eta}_i \text{ are independent, normally distributed conditional under } Y_{(i-1)}, \\
E(\hat{\nu}_i|Y_{(i-1)}) = H_{(i)} \nu_i = \nu_i, \quad E(\hat{\eta}_i|Y_{(i-1)}) = U_{(i)} \nu_i = 0, \\
\text{Cov}(\text{vec}(\hat{\nu}_i)|Y_{(i-1)}) = \Gamma_{ii} \otimes H_{(i)}, \quad \text{Cov}(\text{vec}(\hat{\eta}_i)|Y_{(i-1)}) = \Gamma_{ii} \otimes U_{(i)}, \\
\mathcal{L}(\hat{\nu}_i|Y_{(i-1)}) = W_{m_i} (l_{(i)}; \nu'_i \nu_i), \quad \mathcal{L}(\hat{\eta}_i|Y_{(i-1)}) = W_{m_i} (r_{(i)}; \Gamma_{ii}).
\end{cases}
$$
Since \( \nu \), for the model with only the constant term as explanatory variable, \( \text{LR}_t = \text{W}_{mi}(r_{(i)}, \Gamma_{ii}) \) leads to the conclusion that \( \hat{\eta}_j \) (\( j < i \)) is a function of \( Y_{(j-1)} \) and \( Y_j \) and therefore of \( Y_{(i-1)} \) and the individual observations \( Y_{t(j)} \), \( t = N_i + 1, \ldots, N_j \). The latter are independent of \( \hat{\eta}_i \) because of the row independence of the observations (see (2.2)).

Now consider the likelihood ratio test for the (identifiable) hypothesis (6.1). Denote the restricted parameter space of \( \theta = (\beta, \Sigma) \) by \( \Theta_0 \). From (3.24) the likelihood ratio \( LR_0 \) for (6.1) is given by

\[
LR_0 = \frac{\sup_{\theta \in \Theta_0} L(\theta; Y)}{\sup_{\theta \in \Theta} L(\theta; Y)} = \prod_{i=1}^{r} \left( \frac{\left| \hat{\eta}_i \hat{\eta}_i^\prime \right|}{\left| \hat{\eta}_0 \hat{\eta}_0^\prime \right|} \right)^{\frac{N_i}{2}} = \prod_{i=1}^{r} \left( \frac{\left| \hat{\eta}_i \hat{\eta}_i^\prime \right|}{\left| \hat{\eta}_0 \hat{\eta}_0^\prime + \hat{\nu}_i \hat{\nu}_i^\prime \right|} \right)^{\frac{N_i}{2}} = \prod_{i=1}^{r} L_{00}^{N_i}. \tag{8.3}
\]

For the model with only the constant term as explanatory variable, \( LR_0 \) reduces to the test statistic which Bhargava (1962) derived. Hao and Krishnamoorthy (2001) discussed that test statistic in more detail; in both papers its distribution was approximated.

Since \( \nu_i H_{1(i)} \nu_i = 0 \) under \( H_0 \) of (6.1), applying Theorem 7 to \( L_{0(i)} \), \( L_{1(i)} \) and \( L_{1(i)}^\bot \) leads to the conclusion that \( \hat{\nu}_{1,1} \hat{\nu}_{1,1}, \hat{\nu}_{1,2} \hat{\nu}_{1,2}, \ldots, \hat{\nu}_{1,r} \hat{\nu}_{1,r}, \hat{\nu}_{1,1} \hat{\eta}_1, \hat{\nu}_{1,2} \hat{\eta}_2, \ldots, \hat{\nu}_{1,r} \hat{\eta}_r \) are independent under \( H_0 \) (compare (8.2)). Now Theorem 8 follows directly.

**Theorem 8** Under \( H_0 : \{ \forall i : C_i[\beta_i; \alpha_i] = 0 \} \):

\[
(LR_0)^{\frac{4}{2}} \sim \Lambda_{A,D,T,S}, \text{with} \quad a_i = N_i/N_1, \quad d_i = m_i, \quad t_i = l_{(i)}, \quad s_i = r_{(i)}, \quad \text{for} \ i = 1, \ldots, r. \tag{8.4}
\]

Denote the double restricted parameterspace of \( \theta = (\beta, \Sigma) \) by \( \Theta_{00} \). The likelihood ratio \( LR_{00} \) for (6.3) becomes

\[
LR_{00} = \frac{\sup_{\theta \in \Theta_{00}} L(\theta; Y)}{\sup_{\theta \in \Theta_0} L(\theta; Y)} = \prod_{i=1}^{r} \left( \frac{\left| \hat{\nu}_{0i} \hat{\nu}_{0i}^\prime \right|}{\left| \hat{\nu}_{0i} \hat{\nu}_{0i}^\prime + \hat{\nu}_{01} \hat{\nu}_{01}^\prime \right|} \right)^{\frac{N_i}{2}} = \prod_{i=1}^{r} L_{00}^{N_i}. \tag{8.5}
\]

Since \( \nu_i H_{0(i)} \nu_i = 0 \) under \( H_{00} \), applying Theorem 7 to \( L_{00(i)} \), \( L_{01(i)} \) and \( L_{01(i)}^\bot \) leads to the conclusion that \( \hat{\nu}_{0,11} \hat{\nu}_{0,11}, \hat{\nu}_{0,12} \hat{\nu}_{0,12}, \ldots, \hat{\nu}_{0,r} \hat{\nu}_{0,r}, \hat{\nu}_{0,1} \hat{\nu}_{01}, \hat{\nu}_{0,2} \hat{\nu}_{02}, \ldots, \hat{\nu}_{0r} \hat{\nu}_{0r} \) are independent under \( H_{00} \) (compare (8.2)). Corollary 9 follows directly.
Corollary 9 Under $H_{00}$: $\{ \forall i : C_i[\beta_i; \alpha_i] = D_i[\beta_0; \alpha_0] = 0 \}$:

$$(LR_{00})^2 \sim \Lambda_{A,D,T,S}, \text{ with } a_i = N_i/N_1 \quad d_i = m_i,$$

$$t_i = l_{01(i)}, \quad s_i = r_{0(i)}, \text{ for } i = 1, \ldots, r. \quad (8.6)$$

Note that in both (8.4) and (8.6) $T$ contains the degrees of freedom of the null hypothesis, while $S$ contains the degrees of freedom of the error terms under the alternative hypothesis.

9 A numerical illustration

In this section the estimation and testing procedures are applied to the numerical example described in Section 2. All the tests are performed on a 5% significance level.

The OLS-estimation is straightforward by columnwise regression of the dependent variables on only the explanatory variables. Since the covariance matrix is not known, GLS cannot be applied. To obtain the EGLS-estimates, the orthogonal projections described in Section 3.4 have to be sequentially performed for groups $i = 1, 2, 3$. For $i = 1$ this gives $\hat{\mu}_1 = Z_1$, $\hat{\varepsilon}_1 = E_1$ while $\hat{\beta}_1$ coincides with the OLS-estimate (3.5). For $i = 2, 3$, $\hat{\nu}_i$ and $\hat{\eta}_i$ follow from (3.18), the EGLS-estimates $\hat{\beta}_i$ and $\hat{\alpha}_i$ are sequentially determined according to (3.19) and $\hat{\mu}_i$ and $\hat{\varepsilon}_i$ follow from (3.20). The EGLS-estimate $\hat{S}$ follows from (3.21) and the ML-estimate $\hat{\Sigma}$ is determined according to (3.23) and (3.25).

We will discuss six tests, of which one in more detail. Assume that we are particularly interested in the testing problem (6.1) with $C_i = [C \ 0]$ and $C = [0 \ 0 \ 0 \ 1]$, and in (6.3) with $D_i = [D \ 0]$ and $D = [0 \ 0 \ 1 \ 0]$. The estimates for the corresponding restricted and double restricted model are given in appendix A.2 and A.3. The results for the complete data are presented in appendix B for comparison. Neither the estimation technique nor the missing observations result in large differences in the estimates. The latter phenomenon seems logical in view of the relative small number of missing observations.

Appendix A.4 contains the combined centered MANOVA-tables with the required statistics to perform the two LR tests discussed above. For testing the significance of the fourth explanatory variable, the LR statistic is determined according to (8.3); we found $LR^2_{00} = 0.3070$. From the MANOVA-tables and the structure of the dataset, it follows that

$$(LR_{00})^2 \sim \Lambda_{[1 \ 11/12 \ 10/12],[1 \ 2 \ 1],[1 \ 1 \ 1],[8 \ 6 \ 3]}.$$
Since we do not have an analytical expression available yet for the quantiles of the generalized Wilks’ distribution, the critical values are determined by means of a simulation run of size 1,000,000. In each drawing the inner products \( \tilde{\eta}_i \tilde{\eta}_i \) and \( \tilde{\nu}_i \tilde{\nu}_i \) are simulated from a standard normal distribution (for \( i = 1, \ldots, r \)). Subsequently the ratio of the determinants is calculated according to (8.3). Finally, the critical value is determined as the 5%-quantile from the resulting empirical distribution. The computer program (in MATLAB) is available on request.

The main results for this test and several others tests are contained in Table 4. This table contains the null and alternative hypotheses, the values of the corresponding test statistics and the critical values for the performed tests on a 5% significance level. The tests are performed for both the dataset with missing observations and the complete data. In tests 1 through 5, \( LR^{2}_0 \) is the test statistic; in the last test \( LR^{2}_{00} \).

For the complete data, these test statistics coincide with the usual test statistic Wilks’ lambda except when restrictions are imposed on the \( \alpha_i \). In the latter situation, which applies to tests 1 and 2, the division in groups is essential for the precise null hypothesis. Hence to obtain comparable tests for the complete and incomplete data cases, the division in groups has to be the same.

So our test statistics for the complete data coincide with Wilks’ lambda for tests 3 through 6; the corresponding critical values are given by e.g. Kres (1983), p. 32. In Table 4 the abbreviations TS and CV stand for Test Statistic and Critical Value.

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>Alternative hypothesis</th>
<th>Incomplete data</th>
<th>Complete data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TS</td>
<td>CV</td>
</tr>
<tr>
<td>1. ( \forall i : \alpha_i = \beta_i = 0 )</td>
<td>( \exists i : \beta_i \neq 0 ) or ( \alpha_i \neq 0 )</td>
<td>5.234 \times 10^{-5}</td>
<td>0.0078</td>
</tr>
<tr>
<td>2. ( \forall i : \alpha_i = 0 )</td>
<td>( \exists i : \alpha_i \neq 0 )</td>
<td>0.2174</td>
<td>0.0670</td>
</tr>
<tr>
<td>3. ( \forall i : \beta_i = 0 )</td>
<td>( \exists i : \beta_i \neq 0 )</td>
<td>0.0019</td>
<td>0.0148</td>
</tr>
<tr>
<td>4. ( \forall i : \beta_{4i} = 0 )</td>
<td>( \exists i : \beta_{4i} \neq 0 )</td>
<td>0.0240</td>
<td>0.0262</td>
</tr>
<tr>
<td>5. ( \forall i : \beta_{3i} = 0 )</td>
<td>( \exists i : \beta_{3i} \neq 0 )</td>
<td>0.3070</td>
<td>0.1348</td>
</tr>
<tr>
<td>6. ( \forall i : \beta_{4i} = \beta_{3i} = 0 )</td>
<td>( \exists i : \beta_{3i} \neq 0 ) or ( \forall i : \beta_{4i} = 0 )</td>
<td>0.4474</td>
<td>0.2053</td>
</tr>
</tbody>
</table>

From the results in Table 4 it can be concluded that, for example, the null hypothesis 5 of an insignificant fourth explanatory variable is not rejected. The conclusions for all the tests are identical for the complete and incomplete data. This seems (again) logical in view of the relative small number of missing observations.
10 Conclusions and further research

This paper discussed estimation and testing for a linear regression model with complete observations for the explanatory variables and a monotone missing data structure for the dependent variables. A sequential OLS- and (E)GLS-estimation procedure in terms of linear regression by means of orthogonal projections was presented. This lead to LS-estimators (and MLE since the MLE were shown to coincide with the EGLS-estimators) with a clear geometrical interpretation. Exact tests for restricted and double restricted models were presented. So there is no need for an EM-algorithm or other approximation procedures.

Our results generalize existing literature in two directions: firstly, the standard multivariate regression model is extended to monotone missing dependent variables, and secondly, the monotone missing data problem considered by several authors is extended to the case of not just the constant as explanatory variable.

The small sample properties of our estimators have not been analyzed in detail yet. Especially the small sample properties of the EGLS-estimators are interesting for further research. To derive these properties a similar approach as the one of Krishnamoorthy and Pannala (1999) or Kanda and Fujikoshi (1998) for the model with only the constant term could be followed. It would also be interesting to look at alternative estimators for the covariance matrix such as for example presented by Krishnamoorthy (1991) for the model with only the constant term.

The LR-test for linear restrictions on the explanatory variables or the intergroup-correlations have been extensively discussed. Alternative tests, such as for example constructed by Krishnamoorthy and Pannala (1998) for the model with only the constant term, are left for further research.

References


Anderson, T.W. (1957), Maximum likelihood estimates for a multivariate normal distribution when some observations are missing, Journal of the American Statistical Association 52, 200-203

Bhargava, R.P. (1962), Multivariate tests of hypotheses with incomplete data, Ph.D dissertation, Stanford University, Stanford, CA


Kanda, T. and Y. Fujikoshi (1998), Some basic properties of the MLE’s for a multivariate normal distribution with monotone missing data, American journal of mathematical and management sciences 18, 161-190


Appendix A: Missing data

A.1 The unrestricted model

**OLS-estimates**

\[ b = \begin{bmatrix} 2.0000 & 5.0000 & 5.0000 & 3.4107 \\ 1.0000 & -1.0000 & 1.0000 & 0.9821 \\ 1.0000 & 2.0000 & 0.0000 & 0.1964 \\ -1.0000 & -1.0000 & -1.0000 & -1.0536 \end{bmatrix} \]

\[ S = \begin{bmatrix} 2.2500 & 1.2027 & 2.4054 & -0.6959 \\ 1.2027 & 2.5714 & 0.0000 & -0.0496 \\ 2.4045 & 0.0000 & 10.2857 & -2.7775 \\ -0.6959 & -0.0496 & -2.7775 & 2.1964 \end{bmatrix} \]

**EGLS-estimates**

\[ \hat{\beta} = \begin{bmatrix} 2.0000 & 5.4091 & 5.8182 & 3.1919 \\ 1.0000 & -1.0000 & 1.0000 & 0.9815 \\ 1.0000 & 1.8636 & -0.2727 & 0.2694 \\ -1.0000 & -0.9545 & -0.9091 & -1.0774 \end{bmatrix} \]

\[ \hat{S} = \begin{bmatrix} 2.2500 & 1.2756 & 2.5511 & -0.7382 \\ 1.2756 & 2.6246 & 0.1063 & -0.0951 \\ 2.5511 & 0.1063 & 10.4982 & -2.8377 \\ -0.7382 & -0.0951 & -2.8377 & 2.2139 \end{bmatrix} \]

**ML-estimate**

\[ \hat{\Sigma} = \begin{bmatrix} 1.5000 & 1.0227 & 2.0455 & -0.5480 \\ 1.0227 & 1.7758 & 0.2789 & -0.1050 \\ 2.0455 & 0.2789 & 7.1033 & -1.7858 \\ -0.5480 & -0.1050 & -1.7858 & 1.3169 \end{bmatrix} \]
### A2 The restricted model

**OLS-estimates**

\[
b_0 = \begin{bmatrix}
4.0000 & 7.0769 & 7.0769 & 5.4839 \\
-0.3333 & -2.3187 & -0.3187 & -0.4113 \\
0.6667 & 1.6374 & -0.3626 & -0.1774 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
S_0 = \begin{bmatrix}
3.3333 & 2.5526 & 3.6132 & 0.9083 \\
2.5526 & 3.7335 & 1.4835 & 1.1833 \\
3.6132 & 1.4835 & 10.4835 & -0.8794 \\
0.9083 & 1.1833 & -0.8794 & 3.6014
\end{bmatrix}
\]

**EGLS-estimates**

\[
\hat{b}_0 = \begin{bmatrix}
4.0000 & 7.3889 & 7.5185 & 5.6474 \\
-0.3333 & -2.2593 & -0.2346 & -0.3881 \\
0.6667 & 1.5185 & -0.5309 & -0.2238 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
\hat{S}_0 = \begin{bmatrix}
3.3333 & 2.6022 & 3.6835 & 0.9496 \\
2.6022 & 3.7762 & 1.5439 & 1.2442 \\
3.6835 & 1.5439 & 10.5690 & -0.8337 \\
0.9496 & 1.2442 & -0.8337 & 3.6166
\end{bmatrix}
\]

**ML-estimate**

\[
\Sigma_0 = \begin{bmatrix}
2.5000 & 2.0278 & 2.8704 & 0.7295 \\
2.0278 & 2.7629 & 1.1464 & 0.9570 \\
2.8704 & 1.1464 & 7.1999 & -0.5300 \\
0.7295 & 0.9570 & -0.5300 & 2.4869
\end{bmatrix}
\]

### A3 The double restricted model

**OLS-estimates**

\[
b_{00} = \begin{bmatrix}
5.0000 & 9.6818 & 6.5000 & 5.2000 \\
0.0000 & -1.5000 & -0.5000 & -0.5000 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
S_{00} = \begin{bmatrix}
3.6000 & 4.0247 & 2.8460 & 0.6037 \\
4.0247 & 7.0152 & 0.5000 & 0.6128 \\
2.8460 & 0.5000 & 9.5000 & -0.6835 \\
0.6037 & 0.6128 & -0.6835 & 3.2000
\end{bmatrix}
\]

**EGLS-estimates**

\[
\hat{b}_{00} = \begin{bmatrix}
5.0000 & 9.7813 & 6.5703 & 5.1376 \\
0.0000 & -1.5000 & -0.5000 & -0.5000 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
\hat{S}_{00} = \begin{bmatrix}
3.6000 & 4.0352 & 2.8535 & 0.5898 \\
4.0352 & 7.0272 & 0.5085 & 0.5848 \\
2.8535 & 0.5085 & 9.5060 & -0.6961 \\
0.5898 & 0.5848 & -0.6961 & 3.2049
\end{bmatrix}
\]

**ML-estimate**

\[
\Sigma_{00} = \begin{bmatrix}
3.0000 & 3.2812 & 2.3203 & 0.3876 \\
3.2812 & 5.5320 & 0.2623 & 0.2392 \\
2.3203 & 0.2623 & 7.6689 & -0.6188 \\
0.3876 & 0.2392 & -0.6188 & 2.5704
\end{bmatrix}
\]
A4 The collection of centered MANOVA-tables

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space</strong></td>
<td><strong>SS</strong></td>
<td><strong>DF</strong></td>
</tr>
<tr>
<td>$\tilde{L}_{00(1)}$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$L_{01(1)}$</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>$\tilde{L}_{0(1)}$</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>$L_{1(1)}$</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>$\tilde{L}_{1(1)}$</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>$\tilde{R}_{(12)}$</td>
<td>36</td>
<td>11</td>
</tr>
<tr>
<td>$\tilde{R}_{(12)}$</td>
<td>300</td>
<td>1</td>
</tr>
<tr>
<td>$\tilde{R}^{12}$</td>
<td>336</td>
<td>12</td>
</tr>
<tr>
<td>$\tilde{R}^{11}$</td>
<td>399</td>
<td>303</td>
</tr>
<tr>
<td>$\tilde{R}^{10}$</td>
<td>167</td>
<td>10</td>
</tr>
</tbody>
</table>

$L_{R0} = 0.3070, \quad LR_{00} = 0.4474.$
Appendix B: complete data

B1 The unrestricted model

LS-estimates

\[
\begin{bmatrix}
2.0000 & 5.0000 & 5.0000 & 4.0000 \\
1.0000 & -1.0000 & 1.0000 & 1.0000 \\
1.0000 & 2.0000 & 0.0000 & 0.0000 \\
-1.0000 & -1.0000 & -1.0000 & -1.0000
\end{bmatrix}
\]

ML-estimate

\[
\begin{bmatrix}
1.5000 & 0.7500 & 1.5000 & 0.0000 \\
0.7500 & 1.5000 & 0.0000 & 0.0000 \\
1.5000 & 0.0000 & 6.0000 & -1.5000 \\
0.0000 & 0.0000 & -1.5000 & 1.5000
\end{bmatrix}
\]

B2 Restricted model

LS-estimates

\[
\begin{bmatrix}
4.0000 & 7.0000 & 7.0000 & 6.0000 \\
-0.3333 & -2.3333 & -0.3333 & -0.3333 \\
0.6667 & 1.6667 & -0.3333 & -0.3333 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

ML-estimate

\[
\begin{bmatrix}
2.5000 & 1.7500 & 2.5000 & 1.0000 \\
1.7500 & 2.5000 & 1.0000 & 1.0000 \\
2.5000 & 1.0000 & 7.0000 & -0.5000 \\
1.0000 & 1.0000 & -0.5000 & 2.5000
\end{bmatrix}
\]
B3 The double restricted model

**LS-estimates**

\[ b_{00} = \begin{bmatrix}
5.0000 & 9.5000 & 6.5000 & 5.5000 \\
0.0000 & -1.5000 & -0.5000 & -0.5000 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix} \]

\[ S_{00} = \begin{bmatrix}
3.6000 & 3.6000 & 2.7000 & 0.9000 \\
3.6000 & 6.7500 & 0.4500 & 0.4500 \\
2.7000 & 0.4500 & 8.5500 & -0.4500 \\
0.9000 & 0.4500 & -0.4500 & 3.1500
\end{bmatrix} \]

**ML-estimate**

\[ \hat{\Sigma}_{00} = \begin{bmatrix}
3.0000 & 3.0000 & 2.2500 & 0.7500 \\
3.0000 & 5.6250 & 0.3750 & 0.3750 \\
2.2500 & 0.3750 & 7.1250 & -0.3750 \\
0.7500 & 0.3750 & -0.3750 & 2.6250
\end{bmatrix} \]