Multivariate Linear Regression Model with Elliptically Contoured Distributed Errors and Monotone Missing Dependent Variables

A. Batsidis and K. Zografos[1]

University of Ioannina, Department of Mathematics, 451 10 Ioannina, GREECE

Abstract

In this paper, the multivariate linear regression model is studied under the assumptions that the error term of this model is described by the elliptically contoured distribution and the observations on the response variables are of a monotone missing pattern. It is primarily concerned with estimation of the model parameters', as well as, with the development of the likelihood ratio test in order to examine the existence of linear constraints on the regression coefficients. In this context, the multivariate linear regression model with the constant term as a sole explanatory variable is also studied and leads to estimators of the location and scale of elliptically contoured distributions with monotone missing data. A numerical example is presented for the explanation of the results.

MSC: 62J12, 62H12, 62H15.

Keywords: Monotone missing data; Elliptically contoured distributions; Multivariate linear regression analysis; Estimation; Consistency of Estimators; Hypothesis Testing; Generalized Wilk's distribution;

1 Introduction

Multivariate linear regression analysis is a well known statistical technique which helps to predict values of responses, dependent variables, from a set of explanatory, independent, variables. It is a popular statistical tool used in almost every branch of science and engineering. The classic linear multivariate regression model is analyzed assuming the error matrix has a multivariate normal distribution with zero mean matrix and a positive definite dispersion matrix. The role of the multivariate normal distribution is seminal in probability theory and statistics. However, many statistical papers and empirical studies show that the normal distribution is not capable of exhibiting important properties encountered in finance and economics, among other research areas. A well known insufficiency of the normal distributions are their light tails which fail to formulate for instance, observations of rates of return on common stock, according to Fama (1965) and Blattberg and Gonedes (1974). In this respect, there has been intense research in the use of

[1] Corresponding author.
E-mail address: kzograf@cc.uoi.gr

nonnormal distributions in financial area. The papers by Zellner (1976) and Sutradhar and Ali (1986), include an extensive overview of relevant literature where the error term of the multivariate regression model can have nonnormal distributions and in particular t-distribution in practice. To tackle insufficiencies of normal distributions researchers focus on the broader class of elliptic distributions the last three decades. They provide useful alternatives to the multivariate normal distribution and many of the nice properties of the multivariate normal model holds for elliptic distributions. This generalized family of multivariate distributions, includes as representatives, the multivariate normal, multivariate t-distribution, Pearson type II and VII, multivariate symmetric Kotz type distribution. For a comprehensive monograph on elliptically contoured distributions see for example Fang and Zhang (1990), Fang *et al.* (1990) and Gupta and Varga (1993). Elliptic distributions and in particular the multivariate t-distribution have been considered by several authors to formulate the errors in the multivariate regression model. We refer, among others, to Zellner (1976), Sutradhar and Ali (1986), Galea *et al.* (1997), Liu (2002), Diaz-Garcia *et al.* (2003) and references therein.

In these and other treatments the multivariate linear regression model with nonnormal errors is studied under the assumption that complete data are available for the response and the explanatory variables. The investigation of this model in the case of incomplete data is particularly appealing from a theoretical as well as a practical viewpoint and it has occupied the literature of the subject. In this direction, Little (1992) and Rao and Toutenburg (1999) review the literature of regression analysis with missing values in the independent variables, while Robins and Rotnitzky (1995) discuss the semiparametric efficiency in multivariate regression models with missing data and in particular with monotone missing data for the response variable. Liu (1996) considers Bayesian estimation of multivariate linear regression models using fully observed explanatory variables and possible missing values from response variables. Tang *et al.* (2003) consider the same model with missing data in the response variables, when the nonresponse mechanism depends on the underlined values of the responses and hence is nonignorable. In a recent paper Raats *et al.* (2002) consider the problem of multivariate linear regression analysis, in the context of normally distributed error terms, for the specific case where the observations of the dependent variables appear a monotone missing pattern. Monotone missing data is a particular type of missing data which is common in practice (cf. Hao and Krishnamoorthy (2001)) and on the other hand, a non-monotone data set can be made monotone or nearly so by reordering the variables according to their missingness rates (cf. Schafer (1997), p. 218). There is an increasing interest in the development of statistical methods for handling monotone missing data from normal or elliptical populations (cf., for instance, Kanda and Fujikoshi (1998), Krishnamoorthy and Pannala (1998), Hao and Krishnamoorthy (2001), Chung and Han (2000), Batsidis and Zografos (2005) and references therein).

In this paper we extend the classic multivariate linear regression model in two aspects: on the one hand adopting elliptically contoured distributed errors and on the other hand considering monotone missing data for the response variables. More specifically, we consider a $p-$dimensional vector of response variables on a $q-$dimensional vector of explanatory variables when the explanatory are completely observed while the responses have missing values of a monotone pattern. These data are assumed to be missing completely at random (MCAR), that is the missing data mechanism can be ignored for inference (cf. Rubin (1976)). Practical examples of such data patterns, according to Raats *et al.* (2002), are experimental designs where new dependent variables are added during the experiment, panel surveys with drop outs or new members.

In the frame described previously, in the next section some preliminary concepts will be presented related to the elliptic family of distributions, monotone missing data and multivariate linear regression model. The necessary notation is also stated. In Section 3, the explicit form of the maximum likelihood estimators (MLE) will be derived for the parameters of the model. In Section 4, we will obtain the likelihood ratio test statistic in order to examine the existence of linear constraints on the regression parameters. In the final Section 5, we illustrate the results of this paper to a numerical example. In the Appendix, we deal with the special case of the constant term as sole explanatory variable. This case has been treated previously in the literature in the context of multivariate normal distribution by many authors (cf., for instance, Anderson (1957), Jinadasa and Tracy (1992), Fujisawa (1995)) and in the framework of elliptically contoured distribution by Batsidis and Zografos (2005). It will be shown that in this particular case of a constant term as sole explanatory variable, the main results of this paper lead to the similar ones mentioned above.

2 The model and Preliminaries

Let us suppose that the $N \times p$ random matrix $Y = (y_1, y_2, ..., y_N)^t$ has an elliptical distribution with an $N \times p$ location matrix μ and an $Np \times Np$ scale matrix $\Sigma \otimes I_N$, with $y_i \in R^p$, $i = 1, ..., N$, Σ a positive definite matrix of order p, I_N the identity matrix of order N and $\otimes$ denotes the Kronecker product of the respective matrices. Then, its density function is given by

$$|\Sigma|^{-N/2} f\left[tr\left\{\Sigma^{-1}(Y-\mu)^t(Y-\mu)\right\}\right], \tag{1}$$

where f is a one-dimensional real valued function such that (cf. Gupta and Varga (1993, p. 31))

$$0 \leq \int_0^\infty u^{(Np/2)-1} f(u)\, du < \infty.$$

We use in this case the notation $Y \sim EC_{N \times p}(\mu, \Sigma \otimes I_N)$, and we call $f(\cdot)$ the probability density function generator (p.d.f. generator).

Hence, we note (cf. Anderson *et al.* (1986)) that this distribution is written as a univariate elliptic distribution and then all the properties of univariate elliptic models, are still valid. Moreover, $y_1, y_2, ..., y_N$ can be considered as N uncorrelated realizations from a p-dimensional elliptic population (cf. Diaz-Garcia *et al.* (2003), Gupta and Varga (1993)) and relation (1) is in fact the likelihood function of them. We will present in the sequel some properties related to the univariate elliptic models.

Consider a $N \times p$ random matrix Y from an elliptically contoured distribution, as mentioned above, with unknown location matrix μ and unknown scale matrix $\Sigma \otimes I_N$, with Σ a positive definite matrix of order p. Let Y be partitioned as $(Y_1, Y_2, ..., Y_k)$, where Y_i is a $N \times p_i$ random matrix, $i = 1, ..., k$, and $p_1 + p_2 + ... + p_k = p$.

Following the notation of Kanda and Fujikoshi (1998), let the partitions of the location parameter μ and the scale matrix Σ, according to the ones of Y, be

$$\mu_{[i]} = (\mu_1, \mu_2, ..., \mu_i) \quad \text{and} \quad \Sigma_{(1,...,i)(1,...,i)} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdot & \cdot & \Sigma_{1i} \\ \Sigma_{21} & \Sigma_{22} & \cdot & \cdot & \Sigma_{2i} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \Sigma_{i1} & \Sigma_{i2} & \cdot & \cdot & \Sigma_{ii} \end{pmatrix}, \quad i = 1, ..., k,$$

where μ_j, is $N \times p_j$–dimensional and $\Sigma_{j\ell}$, are $p_j \times p_\ell$ matrices, with Σ_{jj} positive definite for $j, \ell = 1, ..., i$. Then, it is well known (cf. Fang and Zhang (1990), Gupta and Varga (1993)), that each Y_i, is distributed as an elliptically contoured $EC_{N \times p_i}(\mu_i, \Sigma_{ii} \otimes I_N)$, $i = 1, ..., k$.

Let now the transformation of the initial parameters μ and Σ to η and $\Delta = (\Delta_{ij})$, defined respectively by

$$\eta_1 = \mu_1, \;\; \eta_i = \mu_i - \mu_{[i-1]}\Delta_{(1...i-1)i}, \quad i = 2, ..., k,$$

$$\Delta_{11} = \Sigma_{11}, \;\; \Delta_{12} = \Delta_{21}^t = \Sigma_{11}^{-1}\Sigma_{12}, \;\; \Delta_{jj} = \Sigma_{jj} - \Sigma_{j(1...j-1)}\Sigma^{-1}_{(1...j-1)(1...j-1)}\Sigma_{(1...j-1)j},$$

$$\Delta_{(1...j-1)j} = \begin{pmatrix} \Delta_{1j} \\ \cdot \\ \cdot \\ \Delta_{j-1j} \end{pmatrix} = \Delta^t_{j(1...j-1)} = \Sigma^{-1}_{(1...j-1)(1...j-1)}\Sigma_{(1...j-1)j}, \tag{2}$$

for $j = 2, ..., k$. Under this notation we can easily seen that the conditional distribution of $Y_i|Y_{[i-1]}$, $Y_{[i-1]} = (Y_1, ..., Y_{i-1})$, is an elliptically contoured with a p.d.f. generator g_i, location parameter $\eta_i + y_{[i-1]}\Delta_{(1...i-1)i}$, $i = 2, ..., k$, and conditional covariance matrix $Cov\left(Y_i|Y_{[i-1]}\right)$, $i = 2, ..., k$, respectively

$$Cov\left(Y_2|Y_1\right) = \Delta_{22}h_2\left[tr\left\{\Delta_{11}^{-1}(Y_1 - \eta_1)^t(Y_1 - \eta_1)\right\}\right] \otimes I_N, \quad \text{and for } i = 3,...,k,$$

$$\begin{aligned} Cov\left(Y_i|Y_{[i-1]}\right) = (\Delta_{ii} \otimes I_N) \times h_i \Big\{tr\Big[&\left(Y_1 - \eta_1 ... Y_{i-1} - \eta_{i-1} - Y_{[i-2]}\Delta_{(1...i-2)i-1}\right) \times \\ diag\left(\Delta_{11}^{-1}, ..., \Delta_{i-1i-1}^{-1}\right) &\left(\, Y_1 - \eta_1 \;\; ... \;\; Y_{i-1} - \eta_{i-1} - Y_{[i-2]}\Delta_{(1...i-2)i-1}\,\right)^t\Big\}, \end{aligned} \tag{3}$$

where h_i, $i = 2, ..., k$, is a scalar function. This is a well known result appeared, for instance, in Fang *et al.* (1990, p. 45, 67) and Gupta and Varga (1993, p. 63). The expression for h_2 of specific elliptic models, like Pearson type VII etc., has been derived in Batsidis and Zografos (2005), for the case of univariate elliptic models.

Consider now the multivariate linear regression model with p dependent variables, q explanatory variables and N_1 items. This means that we consider the model

$$Y = XB + E, \tag{4}$$

where Y is a $N_1 \times p$ observation matrix of responses, X is a known $N_1 \times q$ model matrix of full column rank q, B is a $q \times p$ matrix of regression parameters with unknown values and E is a $N_1 \times p$ random matrix with $E = (\epsilon_1, \epsilon_2, ..., \epsilon_{N_1})^t$, where $\epsilon_i \in R^p$, $i = 1, ..., N_1$. The matrix E is known as error matrix. Further, we assume that the error matrix E has an elliptical distribution $EC_{N_1 \times p}(0, \Sigma \otimes I_{N_1})$ and hence the density of Y is given by (1) with $\mu = XB$. This is typically called the elliptical multivariate linear regression model (cf. among others Diaz-Garcia *et al.* (2003)) and extends the respective classic linear model by using elliptic distribution for the error matrix E instead of multivariate normal distribution.

Motivated by the work of Raats *et al.* (2002), let us modify the classical elliptical multivariate linear regression model mentioned above. We assume that the observations of the dependent variables are incomplete and can be divided into k, $k \geq 2$, ordered groups according to the pattern of increasing missing rate. Group r contains p_r variables for which exactly the first N_r observations are available, $r = 1, ..., k$, with $N_1 > N_2 > ... > N_k$ and $p_1 + p_2 + ... + p_k = p$. Thus Y has the following form

$$Y = \begin{pmatrix} y_{11}^t & y_{21}^t & \cdot & \cdot & y_{k1}^t \\ y_{12}^t & y_{22}^t & \cdot & \cdot & y_{k2}^t \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ y_{1N_k}^t & y_{2N_k}^t & & & y_{kN_k}^t \\ \cdot & \cdot & & & \\ y_{1N_2}^t & y_{2N_2}^t & & & \\ \cdot & & & & \\ y_{1N_1}^t & & & & \end{pmatrix}, \tag{5}$$
$$\quad \uparrow Y_1 \quad \uparrow Y_2 \qquad\qquad \uparrow Y_k$$

where each y_{rj} is p_r-dimensional vector, $j = 1, ..., N_r$, $r = 1, ..., k$, with $p_1+p_2+...+p_k = p$ and we denote by $Y_r = (y_{r1}, y_{r2}, ..., y_{rN_r})^t$ the $N_r \times$ p_r matrix which contains all the available observations of group r, with $r = 1, ..., k$. Such a pattern is called k-step monotone missing pattern (cf. for example, Kanda and Fujikoshi (1998)).

We will present in the sequel the notation of this paper. The error matrix E is given by

$$E = \begin{pmatrix} \epsilon_{11}^t & \epsilon_{21}^t & \cdot & \epsilon_{k1}^t \\ \epsilon_{12}^t & \epsilon_{22}^t & \cdot & \epsilon_{k2}^t \\ \cdot & \cdot & \cdot & \cdot \\ \epsilon_{1N_1}^t & \epsilon_{2N_1}^t & & \epsilon_{kN_1}^t \end{pmatrix}, \tag{6}$$

where each ϵ_{rj} is p_r-dimensional vector, $r = 1, ..., k$, $j = 1, ..., N_1$, with $p_1+p_2+...+p_k = p$. Similar to Y_r, we denote by E_r the $N_r \times$ p_r matrix given by $E_r = (\epsilon_{r1}, \epsilon_{r2}, ..., \epsilon_{rN_r})^t$,

$r = 1, ..., k$. We will denote by $Y_{(r-1)}$ and $E_{(r-1)}$ the $N_r \times \, p_{(r-1)}$ matrices, where $p_{(r-1)} = p_1 + ... + p_{r-1}$. These matrices contain the first N_r observations of the foregoings groups for $r = 2, ..., k$, while $Y_{(0)} = E_{(0)} = 0$. This means that

$$Y_{(r-1)} = \begin{pmatrix} y_{11}^t & \cdots & y_{r-1,1}^t \\ y_{12}^t & \cdots & y_{r-1,2}^t \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ y_{1N_r}^t & \cdots & y_{r-1,N_r}^t \end{pmatrix}, E_{(r-1)} = \begin{pmatrix} \epsilon_{11}^t & \cdots & \epsilon_{r-1,1}^t \\ \epsilon_{12}^t & \cdots & \epsilon_{r-1,2}^t \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \epsilon_{1N_r}^t & \cdots & \epsilon_{r-1,N_r}^t \end{pmatrix}, \; r = 2, ..., k. \quad (7)$$

Further let X_{lj} denotes the observed value of the l explanatory variable, $l = 1, ..., q$, for the j item, $j = 1, ..., N_1$. We assume that complete data are available for the explanatory variables and X is of the following form

$$X = \begin{pmatrix} X_{11} & X_{21} & \cdot & \cdot & X_{q1} \\ X_{12} & X_{22} & \cdot & \cdot & X_{q2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{1N_1} & X_{2N_1} & \cdot & \cdot & X_{qN_1} \end{pmatrix}. \quad (8)$$

Following the notation of Raats *et al.* (2002) let us denote by X_r the $N_r \times q$ matrix which contains the first N_r observations of all explanatory variables, $r = 1, ..., k$. According to this we have that

$$X_r = \begin{pmatrix} X_{11} & X_{21} & \cdot & \cdot & X_{q1} \\ X_{12} & X_{22} & \cdot & \cdot & X_{q2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{1N_r} & X_{2N_r} & \cdot & \cdot & X_{qN_r} \end{pmatrix}, \; r = 1, ..., k. \quad (9)$$

Moreover, we will assume that the matrix B of the regression parameters has the form

$$B = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdot & \cdot & \beta_{1k} \\ \beta_{21} & \beta_{22} & \cdot & \cdot & \beta_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \beta_{q1} & \beta_{q2} & \cdot & \cdot & \beta_{qk} \end{pmatrix} = (B_1, B_2, ..., B_k), \quad (10)$$

where each β_{lr}, is $1 \times p_r$, $l = 1, ..., q$, $r = 1, ..., k$, with $p_1 + p_2 + ... + p_k = p$ and B_r denotes the $q \times p_r$ submatrices of B. Denote also by $B_{(r-1)}$, the $q \times \, p_{(r-1)}$, $p_{(r-1)} = p_1 + ... + p_{r-1}$, submatrices of B, defined as follows

$$B_{(r-1)} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1,r-1} \\ \beta_{21} & \cdots & \beta_{2,r-1} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \beta_{q1} & \cdots & \beta_{q,r-1} \end{pmatrix}, \; r = 2, ..., k. \quad (11)$$

Using the notation introduced by relations (5)-(11) the multivariate linear regression model (4) can be expressed in the following equivalent form

$$y_{rj}^t = \sum_{l=1}^{q} X_{lj}\beta_{lr} + \epsilon_{rj}^t, r = 1, ..., k, j = 1, ..., N_r,$$

$$Y_r = X_r B_r + E_r = \mu_r + E_r, \quad r = 1, ..., k, \tag{12}$$

and

$$Y_{(r-1)} = X_r B_{(r-1)} + E_{(r-1)} = \mu_{(r-1)} + E_{(r-1)}, \quad r = 2, ..., k. \tag{13}$$

3 Estimation of the model

In this section we will present the maximum likelihood approach for the estimation of the regression coefficients as well as the parameters of the elliptically contoured distribution in the presence of monotone missing data in the response variables. In particular, in the next subsection the maximum likelihood estimators (MLE) of B and Σ are derived. Consistent estimators of the parameters B, as well as, of the covariance matrix of the elliptically distributed error matrix E will be derived in Subsection 3.2 below.

3.1 Maximum likelihood estimators

Following the maximum likelihood approach we seek values of the unknown B and Σ that maximize the likelihood function. The likelihood function is the joint density of Y and it is denoted by $L_Y(B, \Sigma)$.

Theorem 1 *Consider the multivariate linear regression model given by (4), under the assumption that the error matrix E is distributed according to an elliptic distribution $EC_{N_1 \times p}(0, \Sigma \otimes I_{N_1})$. On the basis of monotone missing pattern observations for the response variables of the form (5), the MLE of B and Σ are respectively*

$$\widehat{B} = \left(\widehat{B}_1, \widehat{B}_2, ..., \widehat{B}_k\right) \quad and \quad \widehat{\Sigma} = \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} & \cdot & \cdot & \widehat{\Sigma}_{1k} \\ \widehat{\Sigma}_{21} & \Sigma_{22} & \cdot & \cdot & \widehat{\Sigma}_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \widehat{\Sigma}_{k1} & \widehat{\Sigma}_{k2} & \cdot & \cdot & \widehat{\Sigma}_{kk} \end{pmatrix},$$

where

$$\begin{aligned} \widehat{B}_1 &= (X_1^t X_1)^- X_1^t Y_1, \\ \begin{pmatrix} \widehat{B}_r \\ \widehat{\Delta}_{(1...r-1)r} \end{pmatrix} &= \begin{pmatrix} X_r^t X_r & X_r^t e_{(r-1)} \\ e_{(r-1)}^t X_r & e_{(r-1)}^t e_{(r-1)} \end{pmatrix}^- \begin{pmatrix} X_r^t \\ e_{(r-1)}^t \end{pmatrix} Y_r, \ \mathit{for}\ r = 2, ..., k, \ k \geq 2, \end{aligned}$$

with $e_{(r-1)} = Y_{(r-1)} - \widehat{\mu}_{(r-1)}$, *while*

$$\begin{aligned}
&\widehat{\Sigma}_{11} = \widehat{\Delta}_{11} = \lambda_{\max}(g_1)Q(\widehat{B}_1),\\
&\widehat{\Sigma}_{r(1...r-1)} = \widehat{\Delta}_{r(1...r-1)}\widehat{\Sigma}_{(1...r-1)(1...r-1)},\\
&\widehat{\Delta}_{rr} = \frac{\xi_{k,\max}(g_r)}{h_r}Q(\widehat{B}_r, \widehat{\Delta}_{(1...r-1)r})\\
&\widehat{\Sigma}_{rr} = \widehat{\Delta}_{rr} + \widehat{\Delta}_{r(1...r-1)}\widehat{\Sigma}_{(1...r-1)(1...r-1)}\widehat{\Delta}_{(1...r-1)r},
\end{aligned}$$

for $r = 2, ..., k,\ k \geq 2$, *where* $Q(\widehat{B}_1) = (Y_1 - X_1\widehat{B}_1)^t(Y_1 - X_1\widehat{B}_1)$, $Q(\widehat{B}_r, \widehat{\Delta}_{(1...r-1)r})$ *denotes the quantity*

$$\left(Y_r - \begin{pmatrix} X_r & e_{(r-1)} \end{pmatrix}\begin{pmatrix} \widehat{B}_r \\ \widehat{\Delta}_{(1...r-1)r} \end{pmatrix}\right)^t \left(Y_r - \begin{pmatrix} X_r & e_{(r-1)} \end{pmatrix}\begin{pmatrix} \widehat{B}_r \\ \widehat{\Delta}_{(1...r-1)r} \end{pmatrix}\right)$$

and h_r, *the function related to the conditional covariance matrix* $Cov\left(Y_r|Y_{(r-1)}\right)$, $r = 2, ..., k,\ k \geq 2$, *given by (3). Moreover,* g_1 *and* g_r, $r = 2, ..., k$, *are the nonincreasing, by assumption, p.d.f. generators respectively, of the marginal density of* Y_1 *and the conditional density* $Y_r|Y_{(r-1)}$, *while* $\lambda_{\max}(g_1)$ *denotes the point at which the function* $\lambda^{-N_1p_1/2}g_1(p_1/\lambda)$ *arrives at its maximum and* $\xi_{r,\max}(g_r)$ *the point at which the function* $\xi^{-N_rp_r/2}g_r(p_r/\xi)$ *arrives at its maximum,* $r = 2, ..., k,\ k \geq 2$.

Proof. The proof follows the conditional likelihood approach introduced by Anderson (1957). Writing the joint density as the product of the marginal and conditional densities functions and taking into account relation (1) and the reparametrization (2), we can express the likelihood function as follows

$$L_Y(\mu, \Sigma) = L_{Y_1}(\eta_1, \Delta_{11})L_{Y_2|Y_{(1)}}(\eta_2, \Delta_{12}, \Delta_{22})...L_{Y_k|Y_{(k-1)}}(\eta_k, \Delta_{(1...k-1)k}, \Delta_{kk}). \tag{14}$$

In view of (2), there is a one-to-one correspondence (cf. Anderson (1957), Little and Rubin (2002, p. 135)) between the initial (μ, Σ) and the natural parameters (η, Δ) in the conditional approach. Therefore, it is enough to derive the MLE of (η, Δ). We will obtain, at the beginning, the MLE of η_1 and Δ_{11} based on L_{Y_1}, and then, replacing in the expression of $L_{Y_2|Y_{(1)}}$, η_1 and Δ_{11} by their MLE, we will derive the MLE of η_2, Δ_{12} and Δ_{22} based on the conditional likelihood. We repeat this procedure until the last part of the product given by the right-side of relation (14). Therefore, the MLE of η_1 and Δ_{11} will be obtained, at the beginning, by the maximization, with respect to η_1 and Δ_{11}, of the first part of the likelihood which, using relation (1), is given by

$$L_{Y_1}(\eta_1, \Delta_{11}) = |\Delta_{11}|^{-N_1/2}g_1\left(tr\left\{\Delta_{11}^{-1}(Y_1 - \eta_1)^t(Y_1 - \eta_1)\right\}\right), \tag{15}$$

where g_1, is the nonincreasing, by assumption, p.d.f. generator function of the marginal density of Y_1. By monotonicity of g_1, for a given $\Delta_{11} > 0$, we have to minimize according to η_1, the quantity $(Y_1 - \eta_1)^t(Y_1 - \eta_1)$. Thus, based on the relations (2) and (12), we have to minimize with respect to B_1, the quantity $(Y_1 - X_1B_1)^t(Y_1 - X_1B_1)$. This quantity is minimized by $\widehat{B}_1 = (X_1^tX_1)^-X_1^tY_1$.

Hence the concentrated likelihood is

$$L_{Y_1}(\widehat{B}_1, \Delta_{11}) = |\Delta_{11}|^{-N_1/2} g_1 \left[tr(\Delta_{11}^{-1} Q(\widehat{B}_1)) \right],$$

with

$$Q(\widehat{B}_1) = (Y_1 - X_1\widehat{B}_1)^t (Y_1 - X_1\widehat{B}_1). \tag{16}$$

Following now the steps of the proof of Theorem 4.1.1, of Fang and Zhang (1990), the MLE of Δ_{11} is given by

$$\widehat{\Delta}_{11} = \lambda_{\max}(g_1) Q(\widehat{B}_1). \tag{17}$$

We will now concentrate on the MLE of the parameters η_2, Δ_{21} and Δ_{22}. The conditional likelihood $L_{Y_2|Y_{(1)}} = L_{Y_2|Y_{(1)}}(\eta_2, \Delta_{12}, \Delta_{22})$, in view of relations (1) and (3), is defined by

$$\begin{aligned} L_{Y_2|Y_{(1)}} = \; & |\Delta_{22} h_2|^{-N_2/2} \\ & \times g_2 \left\{ tr \left[(h_2 \Delta_{22})^{-1} (Y_2 - \eta_2 - Y_{(1)}\Delta_{12})^t (Y_2 - \eta_2 - Y_{(1)}\Delta_{12}) \right] \right\} \end{aligned} \tag{18}$$

with h_2 related to the conditional covariance matrix $Cov\,(Y_2|Y_1)$, defined by (3). If we replace in the expression of $L_{Y_2|Y_{(1)}}$, η_1 and Δ_{11}, by their MLE, then using monotonicity of g_2, the maximum of $L_{Y_2|Y_{(1)}}(\eta_2, \Delta_{12}, \Delta_{22})$ with respect to η_2 and Δ_{12} arrives at the values of η_2 and Δ_{12} which minimize the quantity

$$(Y_2 - \eta_2 - Y_{(1)}\Delta_{12})^t (Y_2 - \eta_2 - Y_{(1)}\Delta_{12}).$$

From relation (2), we have that $\eta_2 = \mu_2 - \mu_{(1)}\Delta_{12}$. Hence, if we replace $\mu_{(1)}$ by its MLE we have to minimize with respect to μ_2 and Δ_{12} the quantity

$$(Y_2 - \mu_2 - (Y_{(1)} - \widehat{\mu}_{(1)})\Delta_{12})^t (Y_2 - \mu_2 - (Y_{(1)} - \widehat{\mu}_{(1)})\Delta_{12}).$$

Using the fact that $\mu_2 = X_2 B_2$, we have to minimize with respect to B_2 and Δ_{12} the quantity

$$(Y_2 - X_2 B_2 - (Y_{(1)} - \widehat{\mu}_{(1)})\Delta_{12})^t (Y_2 - X_2 B_2 - (Y_{(1)} - \widehat{\mu}_{(1)})\Delta_{12}),$$

or equivalently the quantity

$$\left(Y_2 - \begin{pmatrix} X_2 & e_{(1)} \end{pmatrix} \begin{pmatrix} B_2 \\ \Delta_{12} \end{pmatrix} \right)^t \left(Y_2 - \begin{pmatrix} X_2 & e_{(1)} \end{pmatrix} \begin{pmatrix} B_2 \\ \Delta_{12} \end{pmatrix} \right),$$

where $e_{(1)} = Y_{(1)} - \widehat{\mu}_{(1)}$. After some algebra we obtain that

$$\begin{pmatrix} \widehat{B}_2 \\ \widehat{\Delta}_{12} \end{pmatrix} = \begin{pmatrix} X_2^t X_2 & X_2^t e_{(1)} \\ e_{(1)}^t X_2 & e_{(1)}^t e_{(1)} \end{pmatrix}^{-} \begin{pmatrix} X_2^t \\ e_{(1)}^t \end{pmatrix} Y_2. \tag{19}$$

In order to derive the MLE of the parameter Δ_{22}, we have to maximize the quantity

$$L_{Y_2|Y_{(1)}}(\widehat{\eta}_2, \widehat{\Delta}_{21}, \Delta_{22}) = \left| h_2(\widehat{\eta}_1, \widehat{\Delta}_{11}) \times \Delta_{22} \right|^{-N_2/2} g_2 \left\{ tr \left[\left(h_2(\widehat{\eta}_1, \widehat{\Delta}_{11}) \times \Delta_{22} \right)^{-1} Q(\widehat{B}_2, \widehat{\Delta}_{12}) \right] \right\},$$

where

$$Q(\widehat{B}_2, \widehat{\Delta}_{12}) = \left(Y_2 - \begin{pmatrix} X_2 & e_{(1)} \end{pmatrix} \begin{pmatrix} \widehat{B}_2 \\ \widehat{\Delta}_{12} \end{pmatrix}\right)^t \left(Y_2 - \begin{pmatrix} X_2 & e_{(1)} \end{pmatrix} \begin{pmatrix} \widehat{B}_2 \\ \widehat{\Delta}_{12} \end{pmatrix}\right). \quad (20)$$

Following again the steps of the proof of Theorem 4.1.1, of Fang and Zhang (1990), we obtain the MLE of $\Delta_{22} = \Sigma_{2\cdot 1}$, that is

$$\widehat{\Delta}_{22} = \widehat{\Sigma}_{2\cdot 1} = \frac{\xi_{2,\max}(g_2)}{h_2(\widehat{\eta}_1, \widehat{\Delta}_{11})} Q(\widehat{B}_2, \widehat{\Delta}_{12}). \quad (21)$$

If we repeat the same procedure, the last term $L_{Y_k|Y_{(k-1)}} = L_{Y_k|Y_{(k-1)}}(\eta_k, \Delta_{(1\ldots k-1)k}, \Delta_{kk})$, of the likelihood function (14) becomes

$$\begin{aligned} &L_{Y_k|Y_{(k-1)}} = |\Delta_{kk} h_k|^{-N_k/2} \times \\ &g_k\left[tr\left\{(\Delta_{kk} h_k)^{-1}(Y_k - \eta_k - Y_{(k-1)}\Delta_{(1\ldots k-1)k})^t (Y_k - \eta_k - Y_{(k-1)}\Delta_{(1\ldots k-1)k})\right\}\right], \end{aligned} \quad (22)$$

where the scalar functions g_k and h_k are respectively the p.d.f. generator of $Y_k|Y_{(k-1)}$, and the function related to the conditional covariance matrix of $Y_k|Y_{(k-1)}$, $k \geq 2$, defined in (3).

In a similar manner, as in the case of the maximization of $L_{Y_2|Y_{(1)}}$, we obtain the MLE estimator of $\begin{pmatrix} B_k \\ \Delta_{(1\ldots k-1)k} \end{pmatrix}$ to be

$$\begin{pmatrix} X_k^t X_k & X_k^t e_{(k-1)} \\ e_{(k-1)}^t X_k & e_{(k-1)}^t e_{(k-1)} \end{pmatrix}^{-} \begin{pmatrix} X_k^t \\ e_{(k-1)}^t \end{pmatrix} Y_k, \quad (23)$$

where $e_{(k-1)} = Y_{(k-1)} - \widehat{\mu}_{(k-1)}$. Moreover, we have that

$$\widehat{\Delta}_{kk} = \frac{\xi_{k,\max}}{h_k} Q(\widehat{B}_k, \widehat{\Delta}_{(1\ldots k-1)k}), \quad (24)$$

for $k \geq 3$, with $\xi_{k,\max}$ being the point at which $\xi^{-N_k p_k/2} g_k(p_k/\xi)$ arrives at its maximum and $Q(\widehat{B}_k, \widehat{\Delta}_{(1\ldots k-1)k})$ is

$$\left(Y_k - \begin{pmatrix} X_k & e_{(k-1)} \end{pmatrix} \begin{pmatrix} \widehat{B}_k \\ \widehat{\Delta}_{(1\ldots k-1)k} \end{pmatrix}\right)^t \left(Y_k - \begin{pmatrix} X_k & e_{(k-1)} \end{pmatrix} \begin{pmatrix} \widehat{B}_k \\ \widehat{\Delta}_{(1\ldots k-1)k} \end{pmatrix}\right). \quad (25)$$

Based on the previous discussion, the MLE $\widehat{\mu}$ and $\widehat{\Sigma}$ of the initial parameters μ and Σ, can be obtained by using relation (2). Hence, using the relations

$$\begin{aligned} &\widehat{\Sigma}_{11} = \widehat{\Delta}_{11}, \\ &\widehat{\Sigma}_{21} = \widehat{\Delta}_{21}\widehat{\Sigma}_{11}, \\ &\widehat{\Sigma}_{22} = \widehat{\Sigma}_{2\cdot 1} + \widehat{\Sigma}_{21}\widehat{\Sigma}_{11}^{-1}\widehat{\Sigma}_{12} = \widehat{\Delta}_{22} + \widehat{\Delta}_{21}\widehat{\Delta}_{11}\widehat{\Delta}_{12}, \\ &\widehat{\Sigma}_{k(1\ldots k-1)} = \widehat{\Delta}_{k(1\ldots k-1)}\widehat{\Sigma}_{(1\ldots k-1)(1\ldots k-1)}, \quad k \geq 3, \\ &\widehat{\Sigma}_{kk} = \widehat{\Delta}_{kk} + \widehat{\Delta}_{k(1\ldots k-1)}\widehat{\Sigma}_{(1\ldots k-1)(1\ldots k-1)}\widehat{\Delta}_{(1\ldots k-1)k}, \quad k \geq 3, \end{aligned}$$

we obtain the desired results. ■

Remark 1. a) An application of Theorem 1 for $N_1 = N_2 = ... = N_k$, leads to the MLE of B and Σ of the elliptic multivariate linear regression model, in the complete data case (cf. Diaz-Garcia *et al.* (2003)).
b) Taking into account Theorem 1, we can obtain after some algebra, the following equivalent expressions for the MLE of $\Delta_{(1...r-1)r}$ and B_r,

$$\widehat{\Delta}_{(1...r-1)r} = \left(e^t_{(r-1)} U_r e_{(r-1)}\right)^- e^t_{(r-1)} U_r Y_r \tag{26}$$

and

$$\widehat{B}_r = \left(X^t_r X_r\right)^- X^t_r \left[Y_r - e_{(r-1)} \widehat{\Delta}_{(1...r-1)r}\right] \tag{27}$$

for $r = 2, ..., k$, $k \geq 2$, with $U_r = I - X_r \left(X^t_r X_r\right)^- X^t_r$. The above expressions will be used later in the study of the consistency property of the MLE of Theorem 1.

A particular multivariate linear regression model is the constant term model obtained from (4) when $X = 1_{N_1}$, with 1_{N_1} the $N_1 \times 1$ unity vector and $B = (B_1, B_2, ..., B_k)$, where each B_i is $1 \times p_i$ dimensional with $p_1 + p_2 + ... + p_k = p$. The $N_1 \times p$ error random matrix E is supposed again elliptically distributed $EC_{N_1 \times p}(0, \Sigma \otimes I_{N_1})$. In other words, we assume the elliptical multivariate linear regression model with the constant term as explanatory variable, under the extra assumption that there exist monotone missing data for the response variables.

This model has received a lot of attention in literature under the assumption of multivariate distributed error terms. We refer among others to Anderson (1957), Jinadasa and Tracy (1992), Fujisawa (1995) and references therein. These results have been also obtained by Raats *et al.* (2002). Our aim is to prove that the MLE for the regression coefficients, as well as, for Σ obtained previously reduce to the same expressions determined recently by Batsidis and Zografos (2005) in the case of elliptically contoured distributions.

In the sequel we will denote

$$\overline{Y}_r = \tfrac{1}{N_r} \sum_{\nu=1}^{N_k} y^t_{r\nu} = \tfrac{1}{N_r} 1^t_{N_r} Y_r \text{ and } S_{rr,r} = (Y_r - 1_{N_r}\overline{Y}_r)^t (Y_r - 1_{N_r}\overline{Y}_r),\ r \geq 1$$
$$\overline{Y}_{(r-1)} = \tfrac{1}{N_r} 1^t_{N_r} Y_{(r-1)} \text{ and } S_{(1...r-1)r,r} = \left(Y_{(r-1)} - 1_{N_r}\overline{Y}_{(r-1)}\right)^t \left(Y_r - 1_{N_r}\overline{Y}_r\right),\ r \geq 2$$
$$S_{(1...r-1)(1...r-1),r} = \left(Y_{(r-1)} - 1_{N_r}\overline{Y}_{(r-1)}\right)^t \left(Y_{(r-1)} - 1_{N_r}\overline{Y}_{(r-1)}\right),\ r \geq 2$$
$$S_{r\cdot(1...r-1),r} = S_{rr,r} - S_{r(1...r-1),r} S^{-1}_{(1...r-1)(1...r-1),r} S_{(1...r-1)r,r},\ r \geq 2.$$

In the next proposition we obtain the estimators of the constant term model. The proof of the proposition is outlined in the Appendix.

Proposition 2 *Consider the multivariate linear regression model given by (4), with the constant term as a sole explanatory variable, under the assumption that the error matrix E is distributed according to an elliptical distribution $EC_{N_1 \times p}(0, \Sigma \otimes I_{N_1})$. On the basis of a monotone missing pattern observations for the response variables, the MLE of B and Σ are* *respectively* $\widehat{B} = \left(\widehat{B}_1, \widehat{B}_2, ..., \widehat{B}_k\right)$ *and* $\widehat{\Sigma} = \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} & \cdot & \cdot & \widehat{\Sigma}_{1k} \\ \widehat{\Sigma}_{21} & \widehat{\Sigma}_{22} & \cdot & \cdot & \widehat{\Sigma}_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \widehat{\Sigma}_{k1} & \widehat{\Sigma}_{k2} & \cdot & \cdot & \widehat{\Sigma}_{kk} \end{pmatrix}$, *where* $\widehat{B}_1 = \overline{Y}_1$ *and* $\widehat{B}_r = \overline{Y}_r - \left(\overline{Y}_{(r-1)} - \widehat{B}_{(r-1)}\right) \widehat{\Delta}_{(1...r-1)r}$, *with* $\widehat{\Delta}_{(1...r-1)r} = S^{-1}_{(1...r-1)(1...r-1),r} S_{(1...r-1)r,r}$, *for* $r = 2, ..., k$, *while* $\widehat{\Sigma}_{11} = \widehat{\Delta}_{11} = \lambda_{\max}(g_1) S_{11,1}$ *and* $\widehat{\Sigma}_{rr}$ *is given by the relation* $\widehat{\Sigma}_{rr} = \widehat{\Delta}_{rr} + \widehat{\Delta}_{r(1...r-1)} \widehat{\Sigma}_{(1...r-1)(1...r-1)} \widehat{\Delta}_{(1...r-1)r}$, *for* $r = 2, ..., k$, *where* $\widehat{\Delta}_{rr} = \frac{\xi_{r,\max}}{h_r} S_{r\cdot(1...r-1),r}$, *with* $g_1, h_r, g_r, \lambda_{\max}, \xi_{r,\max}$ *as given in Theorem 1, for* $r = 2, ..., k$.

3.2 The consistency property

Consistency refers to a limiting property of an estimator and it is usually considered a basic requirement of an inference procedure. Our object in this subsection is to derive the consistent estimators of the parameters B of the model (4), as well as, of the covariance matrix of the elliptically distributed error matrix E of this model. Similar work has been made previously by Sutradhar and Ali (1986) for multivariate $t-$distributed errors in the model (4) and complete data for the responses. Recently, Raats *et al.* (2004) derived the respective consistent estimators under the assumption of normally distributed error matrix and monotone missing observations for the responses Y of the model (4). In this respect the results of this subsection generalize in both aspects the results of the above mentioned papers.

It will be shown, in the sequel, that the MLE $\widehat{B}$ derived in Theorem 1 is a consistent estimator of the parameters B of the model (4). A similar conclusion it is not true in general for the MLE $\widehat{\Sigma}$ of the scale matrix Σ of the elliptically distributed error matrix E. This point will be clarified later on in a remark at the end of this subsection. We will use the symbol $\underset{N_i \to \infty}{p\lim}$ to denote converge in probability, as $N_i \to \infty$, $i = 1, ..., k$. Convergence in probability of random matrices is considered in the element-wise sense.

Consider the multivariate linear regression model (4), that is, $Y = XB + E$ and suppose, as above, that $E \sim EC_{N_1 x p}(0, \Sigma \otimes I_{N_1})$. The covariance matrix of E is $Cov(E) = \Omega \otimes I_{N_1}$, where Ω is a positive definite $p \times p$ matrix related with the scale matrix by the equality $\Omega = c\Sigma$, where $c = -2\Psi'(0)$ is a constant which depends on the characteristic function Ψ of E (cf., for instance, Gupta and Varga (1993, p. 33)). Let the following partition for the covariance matrix Ω:

$$\Omega_{(1,...,i)(1,...,i)} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \cdot & \cdot & \Omega_{1i} \\ \Omega_{21} & \Omega_{22} & \cdot & \cdot & \Omega_{2i} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \Omega_{i1} & \Omega_{i2} & \cdot & \cdot & \Omega_{ii} \end{pmatrix} = c\Sigma_{(1,...,i)(1,...,i)}, \quad i = 1, ..., k,$$

similar to the partition defined previously for the matrix Σ. Further, denote by

$$\Omega_{(1\ldots r-1)r} = \Omega^t_{r(1\ldots r-1)} = \begin{pmatrix} \Omega_{1r} \\ \Omega_{2r} \\ \cdot \\ \cdot \\ \Omega_{r-1,r} \end{pmatrix} = c\Sigma_{(1\ldots r-1)r},\ r = 2, \ldots, k.$$

Under these circumstances the consistent estimators of B and Ω are derived in the next theorem. The keynote in the proof of the theorem is the convergence in probability of submatrices of the error matrix E to suitable submatrices of Ω. Although it is immediate in the case of a multivariate normal distributed error matrix E, in view of the weak law of large numbers, it is not so obvious in the case of an elliptic error matrix E considered here. The next lemma establishes the convergence in probability of submatrices of E. The results of the lemma are proved under the additional assumption that the elliptic distribution $EC_{N_1 \times p}(0, \Sigma \otimes I_{N_1})$ of E possesses a consistency property as it is defined and studied by Kano (1994). Consistency property of the elliptical density function of E is equivalent, according to Theorem 1 of Kano (1994), to the fact that the characteristic function of E and hence the constant c, which appears in the $Cov(E) = (c\Sigma) \otimes I_{N_1}$, does not depend on the dimension of the distribution. This assumption permits to prove the following lemma.

Lemma 3 *Consider the multivariate linear regression model (4) and suppose that the density function of the error matrix $E \sim EC_{N_1 \times p}(0, \Sigma \otimes I_{N_1})$, obeys the consistency property of Kano (1994). Then, under the assumption that $\lim_{N_i \to \infty} \left(\frac{1}{N_i} X_i^t X_i\right)^-$ exists, $i = 1, \ldots, k$,*

i) $p\lim_{N_r \to \infty} \left(\frac{1}{N_r} E^t_{(r-1)} X_r\right) = 0,$

ii) $p\lim_{N_r \to \infty} \left(\frac{1}{N_r} E^t_{(r-1)} E_{(r-1)}\right) = c\Sigma_{(1\ldots r-1)(1\ldots r-1)},$

iii) $p\lim_{N_r \to \infty} \left(\frac{1}{N_r} E^t_{(r-1)} E_r\right) = c\Sigma_{(1\ldots r-1)r}.$

Proof. We will prove part ii). The proof of parts i) and iii) are similar and they are omitted. If $\left(E^t_{(r-1)} E_{(r-1)}\right)_{ij}$ denotes the (i, j) submatrix of $E^t_{(r-1)} E_{(r-1)}$, then

$$\frac{1}{N_r} \left(E^t_{(r-1)} E_{(r-1)}\right)_{ij} = \frac{1}{N_r} \sum_{l=1}^{N_r} \epsilon_{il} \epsilon^t_{jl},\ r = 2, \ldots, k.$$

We observe that $\epsilon_{il} \epsilon^t_{jl}$, $l = 1, \ldots, N_r$, are uncorrelated $p_i \times p_j$ random matrices with $E\left(\epsilon_{il} \epsilon^t_{jl}\right) = Cov\left(\epsilon_{il} \epsilon^t_{jl}\right) = \Omega_{ij} = c\Sigma_{ij}$, where c does not depend on the sample size N_r, taking into account the consistency, in Kano's (1994) sense, of the density of the error matrix E. If we apply the weak law of large numbers (cf. Rao (1973, p. 112)) to the sequence of random matrices $\epsilon_{i1} \epsilon^t_{j1}, \epsilon_{i2} \epsilon^t_{j2}, \ldots, \epsilon_{iN_r} \epsilon^t_{jN_r}$, we have that

$$p\lim_{N_r \to \infty} \left\{\frac{1}{N_r} \left(E^t_{(r-1)} E_{(r-1)}\right)_{ij}\right\} = p\lim_{N_r \to \infty} \left\{\frac{1}{N_r} \sum_{l=1}^{N_r} \epsilon_{il} \epsilon^t_{jl}\right\} = \Omega_{ij} = c\Sigma_{ij},$$

hence

$$\underset{N_r\to\infty}{p\lim}\left\{\frac{1}{N_r}\left(E^t_{(r-1)}E_{(r-1)}\right)\right\} = \Omega_{(1\dots r-1)(1\dots r-1)} = c\Sigma_{(1\dots r-1)(1\dots r-1)},$$

which proves the desired result. ■

Lemma 4 *Under the assumptions of Lemma 3,*

$$\underset{N_r\to\infty}{p\lim}\,\widehat{\Delta}_{(1\dots r-1)r} = \underset{N_r\to\infty}{p\lim}\left\{\left(\tfrac{1}{N_r}E^t_{(r-1)}E_{(r-1)}\right)^-\left(\tfrac{1}{N_r}E^t_{(r-1)}E_r\right)\right\}.$$

Proof. Using relation (26), we obtain after some algebra that

$$\widehat{\Delta}_{(1\dots r-1)r} = \left(E^t_{(r-1)}U_rE_{(r-1)}\right)^- E^t_{(r-1)}U_rE_r, \text{ for } r = 2,\dots,k,$$

where $U_r = I - X_r(X^t_rX_r)^-X^t_r = I - H_r$. Hence,

$$\underset{N_r\to\infty}{p\lim}\,\widehat{\Delta}_{(1\dots r-1)r} = \underset{N_r\to\infty}{p\lim}\left\{\left(\frac{1}{N_r}E^t_{(r-1)}U_rE_{(r-1)}\right)^-\right\}\underset{N_r\to\infty}{p\lim}\left(\frac{1}{N_r}E^t_{(r-1)}U_rE_r\right). \quad (28)$$

Using the fact that $U_r = I - H_r$, we obtain that

$$\begin{aligned}\underset{N_r\to\infty}{p\lim}\left(\frac{1}{N_r}E^t_{(r-1)}U_rE_r\right) &= \underset{N_r\to\infty}{p\lim}\left(\frac{1}{N_r}E^t_{(r-1)}E_r\right) - \underset{N_r\to\infty}{p\lim}\left(\frac{1}{N_r}E^t_{(r-1)}H_rE_r\right)\\ &= \underset{N_r\to\infty}{p\lim}\left(\frac{1}{N_r}E^t_{(r-1)}E_r\right), \quad (29)\end{aligned}$$

because of

$$\begin{aligned}\underset{N_r\to\infty}{p\lim}\left(\tfrac{1}{N_r}E^t_{(r-1)}H_rE_r\right) &= \underset{N_r\to\infty}{p\lim}\left\{\tfrac{1}{N_r}E^t_{(r-1)}X_r\right\}\lim_{N_r\to\infty}\left\{\left(\tfrac{1}{N_r}X^t_rX_r\right)^-\right\}\\ &\times\underset{N_r\to\infty}{p\lim}\left\{\tfrac{1}{N_r}X^t_rE_r\right\} = 0,\end{aligned}$$

in view of Lemma 3 i). Moreover, taking into account again Lemma 3 i)

$$\underset{N_r\to\infty}{p\lim}\left(\frac{1}{N_r}E^t_{(r-1)}H_rE_{(r-1)}\right) = 0,$$

and hence

$$\underset{N_r\to\infty}{p\lim}\left(\frac{1}{N_r}E^t_{(r-1)}U_rE_{(r-1)}\right) = \underset{N_r\to\infty}{p\lim}\left\{\frac{1}{N_r}E^t_{(r-1)}E_{(r-1)}\right\}. \quad (30)$$

Therefore from (28), taking into consideration the intermediate results (29) and (30), we obtain that

$$\underset{N_r\to\infty}{p\lim}\,\widehat{\Delta}_{(1\dots r-1)r} = \underset{N_r\to\infty}{p\lim}\left\{\left(\frac{1}{N_r}E^t_{(r-1)}E_{(r-1)}\right)^-\left(\frac{1}{N_r}E^t_{(r-1)}E_r\right)\right\},$$

which is the desired result. ■

We are now ready to derive the consistent estimators of B and Ω in the next theorem.

Theorem 5 *Under the assumption that* $\lim_{N_i\to\infty}\left\{\left(\frac{1}{N_i}X_i^tX_i\right)^-\right\}$ *exists we have that:*

a) $p\lim_{N_r\to\infty}\widehat{\Delta}_{(1\ldots r-1)r}=\Delta_{(1\ldots r-1)r}$*, for* $r=2,\ldots,k$

b) $p\lim_{N_i\to\infty}\widehat{B}_i=B_i$*, for* $i=1,\ldots,k$ *and*

c) $p\lim_{N_1\to\infty}\widetilde{\Delta}_{11}=\Omega_{11}$*, while* $p\lim_{N_r\to\infty}\widetilde{\Delta}_{rr}=\Omega_{rr}-\Omega_{r(1\ldots r-1)}\Omega^{-1}_{(1\ldots r-1)(1\ldots r-1)}\Omega_{(1\ldots r-1)r}$*,*

with $\widetilde{\Delta}_{11}=\frac{1}{N_1}Q(\widehat{B}_1)$ *and* $\widetilde{\Delta}_{rr}=\frac{1}{N_r}Q(\widehat{B}_r,\widehat{\Delta}_{(1\ldots r-1)r})$*, for* $r=2,\ldots,k$*. Moreover,* $\widehat{B}=\left(\widehat{B}_1,\widehat{B}_2,\ldots,\widehat{B}_k\right)$ *and* $\widehat{\Delta}_{(1\ldots r-1)r}$*, as well as,* $Q(\widehat{B}_1)$ *and* $Q(\widehat{B}_r,\widehat{\Delta}_{(1\ldots r-1)r})$*,* $r=2,\ldots,k$*, are defined in Theorem 1.*

Proof. a) Based on Lemma 3 ii) and iii) it can be easily seen that

$$\begin{aligned}p\lim_{N_r\to\infty}\left\{\left(\frac{1}{N_r}E^t_{(r-1)}E_{(r-1)}\right)^-\left(\frac{1}{N_r}E^t_{(r-1)}E_r\right)\right\} &= \Sigma^{-1}_{(1\ldots r-1)(1\ldots r-1)}\Sigma_{(1\ldots r-1)r}\\ &= \Delta_{(1\ldots r-1)r}.\end{aligned}\tag{31}$$

Relation (31) and Lemma 4 complete the proof of part a).
b) Motivated by Raats *et al.* (2004), we will prove that $p\lim_{N_i\to\infty}\widehat{B}_i=B_i$, for $i=1,\ldots,k$, by using an induction argument. For $i=1$, we have that

$$\widehat{B}_1=(X_1^tX_1)^-X_1^tY_1$$

and taking into consideration (12) we obtain

$$\widehat{B}_1=(X_1^tX_1)^-X_1^t(X_1B_1+E_1)=B_1+(X_1^tX_1)^-X_1^tE_1.$$

Hence

$$\begin{aligned}p\lim_{N_1\to\infty}\left(\widehat{B}_1-B_1\right) &= p\lim_{N_1\to\infty}\left\{\left(\frac{1}{N_1}X_1^tX_1\right)^-\frac{1}{N_1}X_1^tE_1\right\}\\ &= \lim_{N_1\to\infty}\left(\frac{1}{N_1}X_1^tX_1\right)^- p\lim_{N_1\to\infty}\left(\frac{1}{N_1}X_1^tE_1\right)=0,\end{aligned}$$

in view of Lemma 3 i).

Afterwards, using the induction assumption that $p\lim_{N_i\to\infty}\widehat{B}_i=B_i$, for $i=1,\ldots,k-1$, which implies that $p\lim_{N_k\to\infty}\widehat{B}_{(k-1)}=B_{(k-1)}$, we are going to prove that $p\lim_{N_k\to\infty}\widehat{B}_k=B_k$. Based on relation (27) and taking into account (12), we have that

$$\widehat{B}_k=B_k+\left(X_k^tX_k\right)^-X_k^t\left[E_k-e_{(k-1)}\widehat{\Delta}_{(1\ldots k-1)k}\right],$$

or equivalently

$$\widehat{B}_k - B_k = \left(X_k^t X_k\right)^- X_k^t \left[E_k - e_{(k-1)}\widehat{\Delta}_{(1\ldots k-1)k}\right].$$

Hence,

$$\underset{N_k\to\infty}{p\lim}\left(\widehat{B}_k - B_k\right) = \underset{N_k\to\infty}{p\lim}\left\{\left(X_k^t X_k\right)^- X_k^t \left[E_k - e_{(k-1)}\widehat{\Delta}_{(1\ldots k-1)k}\right]\right\}.$$

Using the fact that $\underset{N_k\to\infty}{p\lim}\,\widehat{B}_{(k-1)} = B_{(k-1)}$, $\underset{N_k\to\infty}{p\lim}\,\widehat{\Delta}_{(1\ldots k-1)k} = \Delta_{(1\ldots k-1)k}$, we have that

$$\begin{aligned}\underset{N_k\to\infty}{p\lim}\left(\widehat{B}_k - B_k\right) &= \lim_{N_k\to\infty}\left(\frac{1}{N_k}X_k^t X_k\right)^- \underset{N_k\to\infty}{p\lim}\left(\frac{1}{N_k}X_k^t E_k\right)\\ &\quad - \lim_{N_k\to\infty}\left(\frac{1}{N_k}X_k^t X_k\right)^- \underset{N_k\to\infty}{p\lim}\left(\frac{1}{N_k}X_k^t E_{(k-1)}\right)\Delta_{(1\ldots k-1)k} = 0,\end{aligned}$$

in view of Lemma 3 i). Therefore, the proof of part b) of the theorem is now completed.
c) In order to prove this part of the theorem we note that

$$Q(\widehat{B}_1) = (Y_1 - X_1\widehat{B}_1)^t(Y_1 - X_1\widehat{B}_1) = E_1^t E_1,$$

and so in view of Lemma 3 ii) we have that

$$\underset{N_1\to\infty}{p\lim}\,\widetilde{\Delta}_{11} = \underset{N_1\to\infty}{p\lim}\left\{\frac{1}{N_1}Q(\widehat{B}_1)\right\} = \underset{N_1\to\infty}{p\lim}\left(\frac{1}{N_1}E_1^t E_1\right) = \Omega_{11}.$$

Afterwards, based on relation (27), we obtain that

$$\begin{aligned}Y_r - X_r\widehat{B}_r - e_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r} &= U_r\left(Y_r - e_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}\right)\\ &= U_r\left(E_r - e_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}\right),\end{aligned} \tag{32}$$

with $U_r = I - X_r\left(X_r^t X_r\right)^- X_r^t$. Taking into account that U_r is symmetric and idempotent, we have that

$$Q(\widehat{B}_r, \widehat{\Delta}_{(1\ldots r-1)r}) = \left(E_r - e_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}\right)^t U_r\left(E_r - e_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}\right), \quad r = 2, \ldots, k.$$

From this relation after some algebra we have

$$\begin{aligned}Q(\widehat{B}_r, \widehat{\Delta}_{(1\ldots r-1)r}) &= E_r^t U_r E_r - E_r^t U_r e_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}\\ &\quad -\widehat{\Delta}_{(1\ldots r-1)r}^t e_{(r-1)}^t U_r E_r + \widehat{\Delta}_{(1\ldots r-1)r}^t e_{(r-1)}^t U_r e_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}\end{aligned}$$

and using that $U_r e_{(r-1)} = U_r Y_{(r-1)} = U_r E_{(r-1)}$, we reach the relation

$$\begin{aligned}Q(\widehat{B}_r, \widehat{\Delta}_{(1\ldots r-1)r}) &= E_r^t U_r E_r - E_r^t U_r E_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}\\ &\quad -\widehat{\Delta}_{(1\ldots r-1)r}^t E_{(r-1)}^t U_r E_r + \widehat{\Delta}_{(1\ldots r-1)r}^t E_{(r-1)}^t U_r E_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}.\end{aligned}$$

Based on $\widehat{\Delta}_{(1\ldots r-1)r} = \left(E^t_{(r-1)}U_rE_{(r-1)}\right)^- E^t_{(r-1)}U_rE_r$, we obtain that

$$Q(\widehat{B}_r, \widehat{\Delta}_{(1\ldots r-1)r}) = E^t_rU_rE_r - E^t_rU_rE_{(r-1)}\left(E^t_{(r-1)}U_rE_{(r-1)}\right)^- E^t_{(r-1)}U_rE_r.$$

Therefore

$$Q(\widehat{B}_r, \widehat{\Delta}_{(1\ldots r-1)r}) = E^t_rU_rE_r - \widehat{\Delta}^t_{(1\ldots r-1)r}E^t_{(r-1)}U_rE_{(r-1)}\widehat{\Delta}_{(1\ldots r-1)r}$$

Afterwards, using that $\underset{N_r\to\infty}{p\lim}\widehat{\Delta}_{(1\ldots r-1)r} = \Delta_{(1\ldots r-1)r}$, for $r = 2, \ldots, k$, and applying Lemma 3 ii) and iii) to the matrices $\frac{1}{N_r}E^t_{(r-1)}E_r$, $\frac{1}{N_r}E^t_rE_r$ and $\frac{1}{N_r}E^t_{(r-1)}E_{(r-1)}$, we complete the proof of the theorem. ■

Remark 2. a) We derived in Theorem 5 above the consistent estimators of the parameters B and Ω of the model $Y = XB + E$, $E \sim EC_{N_1\times p}(0, \Sigma\otimes I_{N_1})$, with $Cov\,(E) = \Omega\otimes I_{N_1} = (c\Sigma)\otimes I_{N_1}$. Theorem 5 is only valid for members of $EC_{N_1\times p}(0, \Sigma\otimes I_{N_1})$ which obey the consistency property of Kano (1994), that is for scale mixtures of normal distributions, which include the multivariate normal and multivariate t distributions, as particular cases. Even for these special members of the elliptic family of distributions, the MLE $\widehat{B}$ of B is also a consistent estimator of B. The same is not always true for the MLE $\widehat{\Sigma}$ in view of Theorem 5 c).

It can be easily seen that the MLE $\widehat{\Sigma}$ of Σ is a consistent estimator of the covariance matrix Ω, which means that $\widehat{\Delta}_{11}$ and $\widehat{\Delta}_{rr}$ are consistent estimators of Ω_{11} and $\Omega_{rr} - \Omega_{r(1\ldots r-1)}\Omega^{-1}_{(1\ldots r-1)(1\ldots r-1)}\Omega_{(1\ldots r-1)r}$, respectively, for $r = 2, \ldots, k$, if the following conditions are fulfilled

$$\lim_{N_1\to\infty}(\lambda_{\max}N_1) = 1 \text{ and } \underset{N_r\to\infty}{p\lim}\left(\frac{\xi_{r,\max}}{h_r}N_r\right) = 1. \tag{33}$$

In the Appendix 1 in Batsidis and Zografos (2005) explicit expressions have been obtained for the quantities $\lambda_{\max}$ and $\xi_{r,\max}$ of specific elliptic distributions. Taking into account this appendix it can be easily seen that (33) are only satisfied for the multivariate normal and t distributions. Hence, in summary, the MLE $\widehat{\Sigma}$ is consistent estimator of the covariance matrix Ω only for the multivariate normal and t distributions. The same has been proved by Sutradhar and Ali (1986) for the case of complete data in the responses. b) If we apply Theorem 5 in the special case of the design matrix $X = 1_{N_1}$, with 1_{N_1} the $N_1\times 1$ unity vector, then we obtain the consistent estimators, in the light of the above remark, for the location parameter, as well as, for the covariance matrix of elliptical distributions studied in Batsidis and Zografos (2005).

4 Test of hypotheses

In the previous section, we obtained the MLEs of the regression parameters B, as well as, of the scale matrix Σ of the model (4), under the assumption that monotone missing data, of the form (5), are available in the response variables. In this section, we will obtain in the context mentioned above, the likelihood ratio test statistic for testing the hypothesis

$$H_0 : C_i B_i = 0_{M_i \times p_i}, \forall\, i = 1, ..., k,$$

with C_i a $M_i \times q$ known coefficient matrix, of rank $M_i \leq q$ and $0_{M_i \times p_i}$, the $M_i \times p_i$ matrix with zero elements. This null hypothesis expresses the existence of M_i linear constraints on the parameters B_i.

In order to derive the likelihood ratio test statistic, we note that the null hypothesis

$$H_0 : C_i B_i = 0, \forall\, i = 1, ..., k,$$

can be equivalently stated in the form

$$H_0 : C_1 B_1 = 0_{M_1 \times p_1} \text{ and } \begin{pmatrix} C_i & 0_{M_i \times p_{(i-1)}} \end{pmatrix} \begin{pmatrix} B_i \\ \Delta_{(1...i-1)i} \end{pmatrix} = 0_{M_i \times p_i}, \forall\, i = 2, ..., k, \quad (34)$$

where $p_{(i-1)} = p_1 + ... + p_{i-1}$ and $p_1 + ... + p_k = p$, for $i \geq 2$.

4.1 Likelihood Ratio Test Statistic

The likelihood function is given by

$$L_Y(B, \Sigma) = L_{Y_1}(\eta_1, \Delta_{11}) L_{Y_2|Y_{(1)}}(\eta_2, \Delta_{12}, \Delta_{22}) ... L_{Y_k|Y_{(k-1)}}(\eta_k, \Delta_{(1...k-1)k}, \Delta_{kk}),$$

where $L_{Y_1}(\eta_1, \Delta_{11})$, $L_{Y_2|Y_{(1)}}(\eta_2, \Delta_{12}, \Delta_{22})$ and $L_{Y_k|Y_{(k-1)}}(\eta_k, \Delta_{(1...k-1)k}, \Delta_{kk})$ are defined by (15), (18) and (22) respectively. Using the MLE, obtained in Theorem 1, it can be easily seen that

$$\begin{aligned} \sup_{B,\Sigma} L(B, \Sigma) &= \lambda_{\max}^{-\frac{N_1 p_1}{2}} g_1\left(\frac{p_1}{\lambda_{\max}}\right) \left|Q(\widehat{B}_1)\right|^{-\frac{N_1}{2}} \xi_{2,\max}^{-\frac{N_2 p_2}{2}} g_2\left(\frac{p_2}{\xi_{2,\max}}\right) \left|Q(\widehat{B}_2, \widehat{\Delta}_{12})\right|^{-\frac{N_2}{2}} \\ &\quad \times ... \times \xi_{k,\max}^{-\frac{N_k p_k}{2}} g_k\left(\frac{p_k}{\xi_{k,\max}}\right) \left|Q(\widehat{B}_k, \widehat{\Delta}_{(1...k-1)k})\right|^{-N_k/2}, \end{aligned} \quad (35)$$

where $Q(\widehat{B}_1)$, $Q(\widehat{B}_2, \widehat{\Delta}_{12})$ and $Q(\widehat{B}_k, \widehat{\Delta}_{(1...k-1)k})$ are defined by (16), (20) and (25) respectively, while $\lambda_{\max}$ and $\xi_{k,\max}$, $k \geq 2$, are defined in Theorem 1.

In order to derive the $\sup_{B,\Sigma} L(B, \Sigma)$, under the null hypothesis $H_0 : C_i B_i = 0$, $\forall\, i = 1, ..., k$, or the equivalent null hypothesis given by (34), following the procedure of Theorem 1, which is based on the conditional likelihood approach, we have, at the beginning, to maximize the quantity $L_{Y_1}(\eta_1, \Delta_{11})$ given by (15), subject to the constraint $C_1 B_1 = 0$. After some algebra, the constraint MLE $\widetilde{B}_1$ of B_1 is given by the following relation

$$\widetilde{B}_1 = \widehat{B}_1 - \left(X_1^t X_1\right)^- C_1^t \left[C_1 \left(X_1^t X_1\right)^- C_1^t\right]^- C_1 \widehat{B}_1. \quad (36)$$

Following again the procedure of Theorem 1 we easily verify that

$$\widetilde{\Delta}_{11} = \lambda_{\max}(g_1) Q(\widetilde{B}_1). \tag{37}$$

After that, we concentrate our interest on the maximization of $L_{Y_2|Y_{(1)}}(\eta_2, \Delta_{12}, \Delta_{22})$, which is given by (18), subject to $\begin{pmatrix} C_2 & 0_{M_2 \times p_1} \end{pmatrix} \begin{pmatrix} B_2 \\ \Delta_{12} \end{pmatrix} = 0_{M_2 \times p_2}$. After some algebra we obtain that the MLE of $\begin{pmatrix} B_2 \\ \Delta_{12} \end{pmatrix}$ is

$$\begin{aligned}
\begin{pmatrix} \widetilde{B}_2 \\ \widetilde{\Delta}_{12} \end{pmatrix} &= \begin{pmatrix} X_2^t X_2 & X_2^t \widetilde{e}_{(1)} \\ \widetilde{e}_{(1)}^t X_2 & \widetilde{e}_{(1)}^t \widetilde{e}_{(1)} \end{pmatrix}^{-} \begin{pmatrix} X_2^t \\ \widetilde{e}_{(1)}^t \end{pmatrix} Y_2 - \begin{pmatrix} X_2^t X_2 & X_2^t \widetilde{e}_{(1)} \\ \widetilde{e}_{(1)}^t X_2 & \widetilde{e}_{(1)}^t \widetilde{e}_{(1)} \end{pmatrix}^{-} \begin{pmatrix} C_2^t \\ 0^t \end{pmatrix} \\
&\quad \times \left[\begin{pmatrix} C_2 & 0 \end{pmatrix} \begin{pmatrix} X_2^t X_2 & X_2^t \widetilde{e}_{(1)} \\ \widetilde{e}_{(1)}^t X_2 & \widetilde{e}_{(1)}^t \widetilde{e}_{(1)} \end{pmatrix}^{-} \begin{pmatrix} C_2^t \\ 0^t \end{pmatrix} \right]^{-} \begin{pmatrix} C_2 & 0 \end{pmatrix} \\
&\quad \times \begin{pmatrix} X_2^t X_2 & X_2^t \widetilde{e}_{(1)} \\ \widetilde{e}_{(1)}^t X_2 & \widetilde{e}_{(1)}^t \widetilde{e}_{(1)} \end{pmatrix}^{-} \begin{pmatrix} X_2^t \\ \widetilde{e}_{(1)}^t \end{pmatrix} Y_2,
\end{aligned}$$

with $\widetilde{e}_{(1)} = Y_{(1)} - \widetilde{\mu}_{(1)}$ and 0 the $M_2 \times p_1$ zero matrix.

Following again Theorem 1, we obtain that

$$\widetilde{\Delta}_{22} = \widetilde{\Sigma}_{2\cdot 1} = \frac{\xi_{2,\max}(g_2)}{h_2(\widetilde{\eta}_1, \widetilde{\Delta}_{11})} Q(\widetilde{B}_2, \widetilde{\Delta}_{12}),$$

where

$$Q(\widetilde{B}_2, \widetilde{\Delta}_{12}) = \left(Y_2 - \begin{pmatrix} X_2 & \widetilde{e}_{(1)} \end{pmatrix} \begin{pmatrix} \widetilde{B}_2 \\ \widetilde{\Delta}_{12} \end{pmatrix} \right)^t \left(Y_2 - \begin{pmatrix} X_2 & \widetilde{e}_{(1)} \end{pmatrix} \begin{pmatrix} \widetilde{B}_2 \\ \widetilde{\Delta}_{12} \end{pmatrix} \right).$$

In a similar manner, the maximization of the quantity $L_{Y_k|Y_{(k-1)}}(\eta_k, \Delta_{(1\ldots k-1)k}, \Delta_{kk})$, defined by (22), subject to the constraints $\begin{pmatrix} C_k & 0_{M_k \times p_{(k-1)}} \end{pmatrix} \begin{pmatrix} B_k \\ \Delta_{(1\ldots k-1)i} \end{pmatrix} = 0_{M_k \times p_k}$, implies that

$$\begin{aligned}
\begin{pmatrix} \widetilde{B}_k \\ \widetilde{\Delta}_{(1\ldots k-1)k} \end{pmatrix} &= \begin{pmatrix} X_k^t X_k & X_k^t \widetilde{e}_{(k-1)} \\ \widetilde{e}_{(k-1)}^t X_k & \widetilde{e}_{(k-1)}^t \widetilde{e}_{(k-1)} \end{pmatrix}^{-} \begin{pmatrix} X_k^t \\ \widetilde{e}_{(k-1)}^t \end{pmatrix} Y_k - \\
&\quad \begin{pmatrix} X_k^t X_k & X_k^t \widetilde{e}_{(k-1)} \\ \widetilde{e}_{(k-1)}^t X_k & \widetilde{e}_{(k-1)}^t \widetilde{e}_{(k-1)} \end{pmatrix}^{-} \begin{pmatrix} C_k^t \\ 0^t \end{pmatrix} \\
&\quad \times \left[\begin{pmatrix} C_k & 0 \end{pmatrix} \begin{pmatrix} X_k^t X_k & X_k^t \widetilde{e}_{(k-1)} \\ \widetilde{e}_{(k-1)}^t X_k & \widetilde{e}_{(k-1)}^t \widetilde{e}_{(k-1)} \end{pmatrix}^{-} \begin{pmatrix} C_k^t \\ 0^t \end{pmatrix} \right]^{-} \\
&\quad \times \begin{pmatrix} C_k & 0 \end{pmatrix} \begin{pmatrix} X_k^t X_k & X_k^t \widetilde{e}_{(k-1)} \\ \widetilde{e}_{(k-1)}^t X_k & \widetilde{e}_{(k-1)}^t \widetilde{e}_{(k-1)} \end{pmatrix}^{-} \begin{pmatrix} X_k^t \\ \widetilde{e}_{(k-1)}^t \end{pmatrix} Y_k,
\end{aligned}$$

with 0 the $M_k \times p_{(k-1)}$ zero matrix and

$$\widetilde{\Delta}_{kk} = \frac{\xi_{k,\max}(g_k)}{h_k} Q(\widetilde{B}_k, \widetilde{\Delta}_{(1\dots k-1)k}),$$

for $k \geq 3$. Hence, we obtain that

$$\begin{aligned} \sup_{B,\Sigma,H_0:C_iB_i=0} L(B,\Sigma) &= \lambda_{\max}^{-\frac{N_1 p_1}{2}} g_1 \left(\frac{p_1}{\lambda_{\max}}\right) \left|Q(\widetilde{B}_1)\right|^{-\frac{N_1}{2}} \xi_{2,\max}^{-\frac{N_2 p_2}{2}} g_2 \left(\frac{p_2}{\xi_{2,\max}}\right) \left|Q(\widetilde{B}_2, \widetilde{\Delta}_{12})\right|^{-\frac{N_2}{2}} \\ &\times \dots \times \xi_{k,\max}^{-\frac{N_k p_k}{2}} g_k \left(\frac{p_k}{\xi_{k,\max}}\right) \left|Q(\widetilde{B}_k, \widetilde{\Delta}_{(1\dots k-1)k})\right|^{-\frac{N_k}{2}}. \end{aligned} \tag{38}$$

The likelihood ratio test statistic for testing the hypothesis $H_0 : C_iB_i = 0, \forall\, i = 1, \dots, k$, or its equivalent form (34) is

$$\Lambda = \frac{\sup\limits_{B,\Sigma,H_0:C_iB_i=0} L(B,\Sigma)}{\sup\limits_{B,\Sigma} L(B,\Sigma)},$$

and taking into account (35) and (38), Λ becomes

$$\Lambda = \frac{\sup\limits_{B,\Sigma,H_0:C_iB_i=0} L(B,\Sigma)}{\sup\limits_{B,\Sigma} L(B,\Sigma)} = \frac{\left|Q(\widetilde{B}_1)\right|^{-\frac{N_1}{2}} \left|Q(\widetilde{B}_2, \widetilde{\Delta}_{12})\right|^{-\frac{N_2}{2}} \times \dots \times \left|Q(\widetilde{B}_k, \widetilde{\Delta}_{(1\dots k-1)k})\right|^{-\frac{N_k}{2}}}{\left|Q(\widehat{B}_1)\right|^{-\frac{N_1}{2}} \left|Q(\widehat{B}_2, \widehat{\Delta}_{12})\right|^{-\frac{N_2}{2}} \times \dots \times \left|Q(\widehat{B}_k, \widehat{\Delta}_{(1\dots k-1)k})\right|^{-\frac{N_k}{2}}}. \tag{39}$$

The investigation of the distribution of the test statistic Λ is now in order and it is the subject of the next subsection.

4.2 Distribution of Λ

We observe that the likelihood ratio test statistic, obtained above, coincides with the similar one for testing the same hypothesis $H_0 : C_iB_i = 0, \forall\, i = 1, \dots, k$, under the assumption of multivariate normal error (cf. Raats *et al.* (2002)). This point will help us to derive the null distribution of the test statistic Λ. Indeed, by Theorems 8.1.2 and 8.1.3 of Gupta and Varga (1993), we can easily verify that the null distribution of statistic Λ, given in (39), is invariant in the class of elliptical distributions. Hence the null distribution of Λ is the same as the null distribution of Λ in the case of multivariate normal distributed errors in the model (4). This last distribution of Λ has been studied by Raats *et al.* (2002) and Raats (2004). Before we will proceed with the derivation of the distribution of Λ and for the sake of completeness, we give the definition of the said generalized Wilk's distribution.

Definition 1 *Let* $\Lambda_i = \frac{|A_i|}{|A_i + C_i|}$, *with* $A_i \sim W_{d_i}(s_i)$ *and* $C_i \sim W_{d_i}(t_i)$ *independent of* A_i, *and* $W_{d_i}(s_i)$, $i = 1, \dots, k$, *denotes the Wishart distribution. Suppose that* Λ_i *are independent and follow Wilk's* $\Lambda-$*distribution* $\Lambda(d_i, t_i, s_i)$, $i = 1, \dots, k$. *Then the product* $\prod\limits_{i=1}^{k} \Lambda_i^{a_i}$, *with* $a_1 = 1$, $a_i \in (0,1)$, *for* $i = 2, ..., k$, *follows the generalized Wilk's distribution* $\Lambda_{A,D,T,S}$ *with parameters* $A = [a_1, ..., a_k], D = [d_1, ..., d_k]$, $T = [t_1, ..., t_k]$ *and* $S = [s_1, ..., s_k]$.

In order to produce the null distribution of the test statistic Λ, we can prove, taking into account relation (39), that

$$\Lambda^{2/N_1} = \prod_{i=1}^{k} \Lambda_i^{a_i}, \tag{40}$$

with

$$\Lambda_1 = \frac{\left|Q(\widehat{B}_1)\right|}{\left|Q(\widetilde{B}_1)\right|} \text{ and } \Lambda_i = \frac{\left|Q(\widehat{B}_i, \widehat{\Delta}_{(1\ldots i-1)i})\right|}{\left|Q(\widetilde{B}_i, \widetilde{\Delta}_{(1\ldots i-1)i})\right|}, \text{ for } i = 2, ..., k, \tag{41}$$

where $a_i = \frac{N_i}{N_1}$, for $i = 1, ..., k$.

After some algebra, we can obtain that

$$\begin{aligned} Q(\widetilde{B}_1) &= Q(\widehat{B}_1) + Y_1^t X_1 \left(X_1^t X_1\right)^- C_1^t \left\{C_1 \left(X_1^t X_1\right)^- C_1^t\right\}^- C_1 \left(X_1^t X_1\right)^- X_1^t Y_1 \\ &= Q(\widehat{B}_1) + Y_1^t P_{11} Y_1, \end{aligned} \tag{42}$$

with

$$P_{11} = X_1 \left(X_1^t X_1\right)^- C_1^t \left\{C_1 \left(X_1^t X_1\right)^- C_1^t\right\}^- C_1 \left(X_1^t X_1\right)^- X_1^t, \tag{43}$$

while for $i = 2, ..., k$, we have that

$$Q(\widetilde{B}_i, \widetilde{\Delta}_{(1\ldots i-1)i}) = Q(\widehat{B}_i, \widehat{\Delta}_{(1\ldots i-1)i}) + Y_i^t P_{1i} Y_i, \tag{44}$$

where P_{1i}, $i = 2, ..., k$, is the $N_i \times N_i$ idempotent matrix given by the following relation

$$\begin{aligned} P_{1i} = & \begin{pmatrix} X_i^t \\ \widetilde{e}_{(i-1)}^t \end{pmatrix}^t \begin{pmatrix} X_k^t X_k & X_k^t \widetilde{e}_{(k-1)} \\ \widetilde{e}_{(k-1)}^t X_k & \widetilde{e}_{(k-1)}^t \widetilde{e}_{(k-1)} \end{pmatrix}^- \begin{pmatrix} C_i^t \\ 0_{M_i \times p_{(i-1)}}^t \end{pmatrix} \\ & \times \left[\begin{pmatrix} C_i^t \\ 0_{M_i \times p_{(i-1)}}^t \end{pmatrix}^t \begin{pmatrix} X_k^t X_k & X_k^t \widetilde{e}_{(k-1)} \\ \widetilde{e}_{(k-1)}^t X_k & \widetilde{e}_{(k-1)}^t \widetilde{e}_{(k-1)} \end{pmatrix}^- \begin{pmatrix} C_k^t \\ 0_{M_k \times p_{(k-1)}}^t \end{pmatrix} \right]^- \\ & \times \begin{pmatrix} C_i^t \\ 0_{M_i \times p_{(i-1)}}^t \end{pmatrix}^t \begin{pmatrix} X_k^t X_k & X_k^t \widetilde{e}_{(k-1)} \\ \widetilde{e}_{(k-1)}^t X_k & \widetilde{e}_{(k-1)}^t \widetilde{e}_{(k-1)} \end{pmatrix}^- \begin{pmatrix} X_i^t \\ \widetilde{e}_{(i-1)}^t \end{pmatrix}. \end{aligned} \tag{45}$$

Hence from (40) using relations (41), (42) and (44) we have that

$$\Lambda^{2/N_1} = \frac{\left|Q(\widehat{B}_1)\right|}{\left|Q(\widehat{B}_1) + Y_1^t P_{11} Y_1\right|} \prod_{i=2}^{k} \left(\frac{\left|Q(\widehat{B}_i, \widehat{\Delta}_{(1\ldots i-1)i})\right|}{\left|Q(\widehat{B}_i, \widehat{\Delta}_{(1\ldots i-1)i}) + Y_i^t P_{1i} Y_i\right|} \right)^{a_i}. \tag{46}$$

Moreover, from (16), we have that

$$Q(\widehat{B}_1) = Y_1^t P_{01} Y_1, \tag{47}$$

with

$$P_{01} = I_{N_1} - X_1 \left(X_1^t X_1 \right)^- X_1^t. \quad (48)$$

In this context, taking into account that the quantity $Q(\widehat{B}_i, \widehat{\Delta}_{(1...i-1)i})$ is equal to

$$\left(Y_i - \begin{pmatrix} X_i & e_{(i-1)} \end{pmatrix} \begin{pmatrix} \widehat{B}_i \\ \widehat{\Delta}_{(1...i-1)i} \end{pmatrix} \right)^t \left(Y_i - \begin{pmatrix} X_i & e_{(i-1)} \end{pmatrix} \begin{pmatrix} \widehat{B}_i \\ \widehat{\Delta}_{(1...i-1)i} \end{pmatrix} \right),$$

for $i = 2, ..., k$, we can obtain after some algebra that

$$Q(\widehat{B}_i, \widehat{\Delta}_{(1...i-1)i}) = Y_i^t P_{0i} Y_i, \text{ for } i = 2, ..., k, \quad (49)$$

where P_{0i}, $i = 2, ..., k$, is the $N_i \times N_i$ idempotent matrix given by the following relation

$$P_{0i} = I_{N_i} - \begin{pmatrix} X_i & e_{(i-1)} \end{pmatrix} \begin{pmatrix} X_i^t X_i & X_i^t e_{(i-1)} \\ e_{(i-1)}^t X_i & e_{(i-1)}^t e_{(i-1)} \end{pmatrix}^- \begin{pmatrix} X_i^t \\ e_{(i-1)}^t \end{pmatrix}, \text{ for } i = 2, ..., k. \quad (50)$$

Using these results, it is immediate to see, applying Theorem 7.8.3 of Gupta and Nagar (2000), that under H_0 the random quantity $Q(\widehat{B}_i, \widehat{\Delta}_{(1...i-1)i})$ is described by a Wishart distribution, $W_{p_i}(rank(P_{0i}), \Delta_{ii})$, for $i = 1, ..., k$, while the quantity $Y_i^t P_{1i} Y_i$ can be easily proved that follows a Wishart distribution $W_{p_i}(rank(P_{1i}), \Delta_{ii})$. Moreover, applying Theorem 7.8.5 of Gupta and Nagar (2000), we can easily seen that the $Q(\widehat{B}_i, \widehat{\Delta}_{(1...i-1)i})$ is independent of $Y_i^t P_{1i} Y_i$, because of $P_{0i} P_{1i} = 0_{N_i \times N_i}$, for $i = 1, ..., k$. Hence, the distribution of Λ_i, given by (41), is Wilk's $\Lambda(d_i = p_i, t_i = rank\left(P_{1i}\right), s_i = rank(P_{0i}))$, for $i = 1, ..., k$, with $rank(A)$ being the rank of the matrix A. Thus from Definition 1, we have that under the null hypothesis, the likelihood ratio test statistic Λ of relation (39), follows the generalized Wilk's distribution, $\Lambda_{A,D,T,S}$ with parameters $A = [a_1, ..., a_k]$, $D = [d_1, ..., d_k]$, $S = [s_1, ..., s_k]$ and $T = [t_1, ..., t_k]$, which are given by the following relations

$$a_i = \frac{N_i}{N_1}, d_i = p_i, s_i = rank(P_{0i}) \text{ and } t_i = rank\left(P_{1i}\right). \quad (51)$$

Since we do not have available an analytical expression for the quantiles of the generalized Wilk's distribution, the critical values for testing the hypothesis under examination, are determined by simulation. In order to avoid this procedure, Raats (2004) proved that the generalized Wilk's distribution can be approximated by χ^2-distributions. In particular, motivated from Theorem 3.1 of Raats (2004), we have that a second order approximation of the distribution of

$$V = -2 \log \left(\prod_{i=1}^{k} \Lambda_i^{a_i} \right),$$

is

$$P(V \leq v) = (1 - w_2) P\left(\chi_f^2 \leq vq\right) + w_2 P\left(\chi_{f+4}^2 \leq vq\right) + O(N^{-3}), \quad (52)$$

where

$$
\begin{aligned}
f &= \sum_{i=1}^{k}\sum_{j=1}^{d_i} t_i, \\
q &= \frac{1}{4f}\sum_{i=1}^{k}\sum_{j=1}^{d_i} \frac{t_i}{a_i}\left(2s_i - 2j + t_i\right), \\
w_2 &= -\frac{f}{4} + \frac{1}{48q^2}\sum_{i=1}^{k}\sum_{j=1}^{d_i} \frac{t_i}{a_i^2}\left\{3\left(s_i + j + 1\right)\left(s_i + j + t_i + 1\right) + \left(t_i - 2\right)\left(t_i - 1\right)\right\}.
\end{aligned}
\tag{53}
$$

Also from Raats (2004) we have that a first order approximation is given by the relation

$$
P(V \le v) = P\left(\chi_f^2 \le vq\right). \tag{54}
$$

The results of Subsections 4.1 and 4.2 are summarized in the next theorem.

Theorem 6 *Under the assumptions of Theorem 1, the likelihood ratio criterion for testing the hypothesis* $H_0 : C_i B_i = 0_{M_i \times p_i}, \forall\, i = 1, ..., k$, *with* C_i *a* $M_i \times q$ *known coefficient matrix, of rank* $M_i \le q$, *is*

$$
\Lambda^{2/N_1} = \frac{\left|Q(\widehat{B}_1)\right|}{\left|Q(\widehat{B}_1) + Y_1^t P_{11} Y_1\right|} \prod_{i=2}^{k} \left(\frac{\left|Q(\widehat{B}_i, \widehat{\Delta}_{(1...i-1)i})\right|}{\left|Q(\widehat{B}_i, \widehat{\Delta}_{(1...i-1)i}) + Y_i^t P_{1i} Y_i\right|} \right)^{a_i},
$$

with $Q(\widehat{B}_1)$ *and* $Q(\widehat{B}_i, \widehat{\Delta}_{(1...i-1)i})$, *for* $i \ge 2$, *are given by (47) and (49), respectively, while* P_{11} *and* P_{1i}, *for* $i \ge 2$, *are given by (42) and (47), respectively. The test statistic follows under the null hypothesis a generalized Wilks distribution* $\Lambda_{A,D,T,S}$ *with parameters* $A = [1, \frac{N_2}{N_1}, ..., \frac{N_k}{N_1}], D = [p_1, ..., p_k]$, $S = [rank(P_{01}), ..., rank(P_{0k})]$ *and* $T = [rank(P_{11}), ..., rank(P_{1k})]$. *Moreover, a second order approximation of the distribution of* $-2\log\Lambda^{2/N_1}$ *is given by relations (52) and (53), while a first order approximation is given by (54).*

Remark 3. If we apply the results of Theorem 6 in the special case of the constant term model obtained from (4), when $X = 1_{N_1}$, with 1_{N_1} the $N_1 \times 1$ unity vector and $B = (B_1, B_2, ..., B_k)$, where each B_i is $1 \times p_i$ dimensional with $p_1 + p_2 + ... + p_k = p$, under the further assumption that $C_i = 1$, we reach the results obtained previously in Krishnamoorty and Pannala (1998).

5 Numerical Example

Sutradhar and Ali (1986) dealt with a multivariate linear regression model under the assumption that the errors have a multivariate t-distribution. This model, which is a direct multidimensional generalization of Zellner's (1976) regression model is used in the area of the stock market analysis. Sutradhar and Ali (1986), in order to illustrate their results, considered a stock market problem relating to four selected firms, having the regression model

$$y_{ij} = a_i + \beta_i m_j + \epsilon_{ij}, i = 1, ..., 4, \; j = 1, ..., 20,$$

where y_{ij} denotes the monthly return on \$100 of capital invested on the i stock during the j month, while m_j denotes the weighted average of these returns during the j month for the aggregate of all stocks trading on the New York Stock Exchange and the error variable was assumed to have a t-distribution. The available data are given in Table 1 of Sutradhar and Ali (1986).

Therefore, if we use the notation of Section 2 we state the regression model

$$Y = XB + E,$$

where Y is a 20×4 observation matrix of responses, X is a known 20×2 model matrix of full column rank, with the elements of the first column equal to 1, $B = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{pmatrix}$ is a 2×4 matrix of regression parameters with unknown values, and the error matrix E is a 20×4 random matrix, which has a matrix variate t distribution $Mt_{20\times4}(0, \Sigma \otimes I_{20})$.

In the sequel and in order to illustrate the main results of this paper, we discard from Table 1 of Sutradhar and Ali (1986) the last six observations on the fourth firm-response variable. That means that we obtain a 2−step monotone missing sample for the response variables with

$$p_1 = 3, \; p_2 = 1, \; N_1 = 20 \text{ and } N_2 = 14.$$

Using the complete data set of Sutradhar and Ali (1986), the estimators $\widehat{B}_c$ of the parameters of the model and the estimator $\widehat{\Sigma}_c$ of Σ are

$$\widehat{B}_c = \begin{pmatrix} -0.2749 & -0.8904 & 0.2159 & -2.0095 \\ 1.1815 & 1.0132 & 0.9513 & 1.1196 \end{pmatrix}$$

and

$$\widehat{\Sigma}_c = \begin{pmatrix} 7.8015 & 6.0204 & -4.0966 & -0.8160 \\ 6.0204 & 12.3543 & -1.1775 & -3.6845 \\ -4.0966 & -1.1775 & 13.6160 & 9.5150 \\ -0.8160 & -3.6845 & 9.5150 & 22.6846 \end{pmatrix}.$$

Using the estimation procedure introduced in Section 3, the respective MLE's are

$$\widehat{B}_M = \begin{pmatrix} -0.2749 & -0.8904 & 0.2159 & -1.0708 \\ 1.1815 & 1.0132 & 0.9513 & 1.1299 \end{pmatrix}$$

and

$$\widehat{\Sigma}_M = \begin{pmatrix} 7.8015 & 6.0204 & -4.0966 & -2.9186 \\ 6.0204 & 12.3543 & -1.1775 & -7.6747 \\ -4.0966 & -1.1775 & 13.6160 & 6.1773 \\ -2.9186 & -7.6747 & 6.1773 & 16.7292 \end{pmatrix}.$$

The MLE's based on partially complete data (the complete data obtained by discarding the additional data on the first $p_1 = 3$ components) are:

$$\widehat{B}_{PC} = \begin{pmatrix} -0.3057 & -0.8884 & 0.2673 & -1.0616 \\ 1.0550 & 0.7955 & 0.7361 & 1.1234 \end{pmatrix}$$

and

$$\widehat{\Sigma}_{PC} = \begin{pmatrix} 9.4591 & 6.2461 & -5.4823 & -2.9323 \\ 6.2461 & 13.8200 & -2.3845 & -9.4512 \\ -5.4823 & -2.3845 & 8.7979 & 3.7541 \\ -2.9323 & -9.4512 & 3.7541 & 17.0437 \end{pmatrix},$$

respectively.

Using measures such as the scale ratio and the likelihood displacement, which was used to measure the influence of dropping observations by Diaz-Garcia *et al.* (2003), we can see that the estimators based on the whole monotone data are closer to the similar one based on complete data than to the estimators based on partially complete data. Thus, in this sense, using the whole monotone data we are able to recover the information lost due to the deletion of the six observations.

In order to illustrate Theorem 6, let us consider in the sequel the following hypothesis testing problem, related to the numerical example of this section. Suppose that we want to test, for instance, according to the notation of Section 4, the null hypothesis

$$H_0 : C_i B_i = 0, \forall i = 1, 2, \text{ against } H_\alpha : \exists i \in \{1, 2\} \text{ such that } C_i B_i \neq 0,$$

where

$$B_1 = \begin{pmatrix} a_1 & a_2 & a_3 \\ \beta_1 & \beta_2 & \beta_3 \end{pmatrix}, B_2 = \begin{pmatrix} a_4 \\ \beta_4 \end{pmatrix}, \text{with } C_i = \begin{pmatrix} -1 & 1 \end{pmatrix}, i = 1, 2.$$

Following the results of Section 4, we obtain that the test statistic of Theorem 6 follows under the null hypothesis a generalized Wilk's distribution $\Lambda_{A,D,T,S}$ with parameters $A = [1, 0.7], D = [3, 1]$, $T = [1, 1]$ and $S = [18, 9]$. Moreover $\Lambda = 0.0016$, $V = -2 \log \Lambda^{2/N_1} = 1.2863$, while from relation (53) we have that $f = 4$, $q = 7.7054$ and $w_2 = 0.0158$, with $p-$value $= 0.0456$.

According to the standard procedure for the same hypothesis (cf. for instance Diaz-Garcia *et al.* (2003), Siotani *et al.* (1985, pp. 298-299), Muirhead (1982, pp. 458-460) and references therein) based on partially complete data with $N = 14$, the value of the likelihood ratio test statistic is 0.0032 with $p-$value $= 0.0883$. We observe that the test statistic based on the monotone data, provides more evidence against the null hypothesis than the statistic based on partially complete data. If, in addition, we test the hypothesis by using the whole sample, the value of the test statistic, based on the standard procedure, is 0.0025 with $p-$value$= 0.0487$. This value is closer to the respective value of the test proposed in this paper, than the similar one obtained by using partially complete data method.

Acknowledgements

This research was partially funded by the program "Heraklitos" of the Operational Program for Education and Initial Vocational Training of the Hellenic Ministry of Education under the 3rd Community Support Framework and the European Social Fund.

Appendix.

Proof of Proposition 2. Using a mathematical induction argument, we will prove, at the beginning, that the desired results are true for a $k = 2$ step monotone pattern. Then, under the assumption that the conclusions of Proposition 2 holds for $k-1$ step pattern, we will prove that it is also true for k step monotone pattern.

In the sequel, we will use the following relations, which can be easily obtained,

$$U_r Y_r = \left(I - \frac{1}{N_r} 1_{N_r} 1_{N_r}^t\right) Y_r = Y_r - 1_{N_r} \overline{Y}_r, \tag{55}$$

and

$$Y_r^t U_r Y_r = (U_r Y_r)^t U_r Y_r = \left(Y_r - 1_{N_r} \overline{Y}_r\right)^t \left(Y_r - 1_{N_r} \overline{Y}_r\right) = S_{rr,r} \tag{56}$$

with $U_r = I - \frac{1}{N_r} 1_{N_r} 1_{N_r}^t$, for $r = 1, ..., k$. Moreover,

$$\begin{aligned} e_{(r-1)}^t U_r e_{(r-1)} &= \left(Y_{(r-1)} - 1_{N_r} \overline{Y}_{(r-1)}\right)^t \left(Y_{(r-1)} - 1_{N_r} \overline{Y}_{(r-1)}\right) \\ &= S_{(1...r-1)(1...r-1),r}, \end{aligned} \tag{57}$$

and for $r = 2, ..., k$,

$$\begin{aligned} e_{(r-1)}^t U_r Y_r &= \left(Y_{(r-1)} - 1_{N_r} \widehat{B}_{(r-1)}\right)^t \left(Y_r - 1_{N_r} \overline{Y}_r\right) \\ &= \left(Y_{(r-1)} - 1_{N_r} \overline{Y}_{(r-1)}\right)^t \left(Y_r - 1_{N_r} \overline{Y}_r\right) = S_{(1...r-1)r,r}. \end{aligned} \tag{58}$$

Relation (57) can be easily proved, combining that

$$\begin{aligned} e_{(r-1)}^t e_{(r-1)} &= \left(Y_{(r-1)} - 1_{N_r} \widehat{B}_{(r-1)}\right)^t \left(Y_{(r-1)} - 1_{N_r} \widehat{B}_{(r-1)}\right) \\ &= \left(Y_{(r-1)} - 1_{N_r} \overline{Y}_{(r-1)}\right)^t \left(Y_{(r-1)} - 1_{N_r} \overline{Y}_{(r-1)}\right) \\ &\quad + \left(\overline{Y}_{(r-1)} - \widehat{B}_{(r-1)}\right)^t 1_{N_r}^t 1_{N_r} \left(\overline{Y}_{(r-1)} - \widehat{B}_{(r-1)}\right), \end{aligned}$$

and

$$\begin{aligned} e_{(r-1)}^t 1_{N_r} 1_{N_r}^t e_{(r-1)} &= \left(1_{N_r}^t e_{(r-1)}\right)^t \left(1_{N_r}^t e_{(r-1)}\right) \\ &= N_r^2 \left(\overline{Y}_{(r-1)} - \widehat{B}_{(r-1)}\right)^t \left(\overline{Y}_{(r-1)} - \widehat{B}_{(r-1)}\right). \end{aligned}$$

Now we are ready to present the proof of Proposition 2. For a $k = 2$ step monotone pattern, using Theorem 1 and in view of $X_1 = 1_{N_1}$, we have that

$$\widehat{B}_1 = (X_1^t X_1)^- X_1 Y_1 = (1_{N_1}^t 1_{N_1})^{-1} 1_{N_1}^t Y_1 = \frac{1}{N_1} \sum_{\nu=1}^{N_1} y_{1\nu}^t = \overline{y}_{1,1}^t = \overline{Y}_1.$$

Moreover from Theorem 1 we have that $\widehat{\Delta}_{11} = \lambda_{\max}(g_1)Q(\widehat{B}_1)$, where $Q(\widehat{B}_1) = (Y_1 - X_1\widehat{B}_1)^t(Y_1 - X_1\widehat{B}_1)$. Because of $X_1 = 1_{N_1}$ and $\widehat{B}_1 = \overline{Y}_1$, we obtain that

$$\widehat{\Delta}_{11} = \lambda_{\max}(g_1)(Y_1 - 1_{N_1}\overline{Y}_1)^t(Y_1 - 1_{N_1}\overline{Y}_1) = \lambda_{\max}(g_1)S_{11,1}.$$

Following Remark 1b) we have that the MLE estimator of Δ_{12}, for this special case, is

$$\widehat{\Delta}_{12} = \left[e_{(1)}^t U_2 e_{(1)}\right]^- e_{(1)}^t U_2 Y_2,$$

with $e_{(1)} = Y_{(1)} - \widehat{\mu}_{(1)} = Y_{(1)} - 1_{N_2}\widehat{B}_{(1)} = Y_{(1)} - 1_{N_2}\overline{Y}_1$ and $U_2 = I - \frac{1}{N_2}1_{N_2}1_{N_2}^t$. Taking into account (57) and (58), we have that

$$\widehat{\Delta}_{12} = S_{11,2}^{-1}S_{12,2}. \tag{59}$$

Moreover, using Remark 1 b) we have that

$$\begin{aligned}
\widehat{B}_2 &= \left(1_{N_2}^t 1_{N_2}\right)^- 1_{N_2}^t \left(Y_2 - e_{(1)}\widehat{\Delta}_{12}\right) \\
&= \frac{1}{N_2}1_{N_2}^t Y_2 - \frac{1}{N_2}1_{N_2}^t e_{(1)}\widehat{\Delta}_{12} \\
&= \overline{Y}_2 - (\overline{Y}_{(1)} - \overline{Y}_1)\widehat{\Delta}_{12}
\end{aligned}$$

because it holds that $\frac{1}{N_2}1_{N_2}^t e_{(1)} = \frac{1}{N_2}1_{N_2}^t\left(Y_{(1)} - 1_{N_2}\overline{Y}_1\right) = \overline{Y}_{(1)} - \overline{Y}_1$.

In order to obtain $\widehat{\Delta}_{22}$, we have to compute the quantity $Q(\widehat{B}_2, \widehat{\Delta}_{12})$, which is given, using Theorem 1 and relation (32), by the following relation

$$\begin{aligned}
Q(\widehat{B}_2, \widehat{\Delta}_{12}) &= \left(Y_2 - e_{(1)}\widehat{\Delta}_{12}\right)^t U_2 \left(Y_2 - e_{(1)}\widehat{\Delta}_{12}\right) \\
&= Y_2^t U_2 Y_2 - \widehat{\Delta}_{21} e_{(1)}^t U_2 Y_2 - Y_2^t U_2 e_{(1)}\widehat{\Delta}_{12} + \widehat{\Delta}_{21} e_{(1)}^t U_2 e_{(1)}\widehat{\Delta}_{12}.
\end{aligned}$$

Using (56)-(59) we obtain that

$$\begin{aligned}
Q(\widehat{B}_2, \widehat{\Delta}_{12}) &= S_{22,2} - \widehat{\Delta}_{21}S_{12,2} - S_{21,2}\widehat{\Delta}_{12} + \widehat{\Delta}_{21}S_{11,2}\widehat{\Delta}_{12} \\
&= S_{22,2} - S_{21,2}S_{11,2}^{-1}S_{12,2} = S_{2\cdot 1,2}.
\end{aligned}$$

Assuming now that the desired results hold for a $k-1$ step monotone missing pattern, we are going to prove that Proposition 2 remains valid for a k step pattern. The MLE estimator of $\Delta_{(1\ldots k-1)k}$, for $k \geq 3$ is

$$\widehat{\Delta}_{(1\ldots k-1)k} = \left[e_{(k-1)}^t U_k e_{(k-1)}\right]^- e_{(k-1)}^t U_k Y_k,$$

with $e_{(k-1)} = Y_{(k-1)} - \widehat{\mu}_{(k-1)} = Y_{(k-1)} - 1_{N_k}\widehat{B}_{(k-1)}$. In view of relations (57) and (58), we obtain

$$\widehat{\Delta}_{(1\ldots k-1)k} = S_{(1\ldots k-1)(1\ldots k-1),k}^{-1} S_{(1\ldots k-1)k,k}. \tag{60}$$

Moreover, using Remark 1 b)

$$\begin{aligned}\widehat{B}_k &= \left(1_{N_k}^t 1_{N_k}\right)^{-} 1_{N_k}^t \left(Y_k - e_{(k-1)}\widehat{\Delta}_{(1\ldots k-1)k}\right)\\ &= \frac{1}{N_k} 1_{N_k}^t Y_k - \frac{1}{N_k} 1_{N_k}^t e_{(k-1)}\widehat{\Delta}_{(1\ldots k-1)k}\\ &= \overline{Y}_k - (\overline{Y}_{(k-1)} - \widehat{B}_{(k-1)})\widehat{\Delta}_{(1\ldots k-1)k}, \text{ for } k \geq 3,\end{aligned}$$

because it holds that $1_{N_k}^t e_{(k-1)} = N_k \left(\overline{Y}_{(k-1)} - \widehat{B}_{(k-1)}\right)$.

In order to compute the estimator of $\widehat{\Delta}_{kk}$, we have to compute the quantity $Q(\widehat{B}_k, \widehat{\Delta}_{(1\ldots k-1)k})$, which is given in view of Theorem 1 and (32), by the following relation

$$\begin{aligned}Q(\widehat{B}_k, \widehat{\Delta}_{(1\ldots k-1)k}) &= \left(Y_k - e_{(k-1)}\widehat{\Delta}_{(1\ldots k-1)k}\right)^t U_k \left(Y_k - e_{(k-1)}\widehat{\Delta}_{(1\ldots k-1)k}\right)\\ &= Y_k^t U_k Y_k + \widehat{\Delta}_{k(1\ldots k-1)} e_{(k-1)}^t U_k e_{(k-1)} \widehat{\Delta}_{(1\ldots k-1)k}\\ &\quad -\widehat{\Delta}_{k(1\ldots k-1)} e_{(k-1)}^t U_k Y_k - Y_k^t U_k e_{(k-1)} \widehat{\Delta}_{(1\ldots k-1)k}.\end{aligned}$$

Using relations (56)- (58) and relation (60), we reach the following relation

$$\begin{aligned}Q(\widehat{B}_k, \widehat{\Delta}_{(1\ldots k-1)k}) &= S_{kk,k} - S_{k(1\ldots k-1),k} S^{-1}_{(1\ldots k-1)(1\ldots k-1),k} S_{(1\ldots k-1)k,k}\\ &= S_{k\cdot(1\ldots k-1),k}.\end{aligned}$$

References

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

[2] Anderson, T. W., Fang, K. T. and Hsu, H. (1986). Maximum-likelihood estimators and likelihood-ratio criteria for multivariate elliptically contoured distributions. *The Canadian Journal of Statistics,* **14**, 55-59.

[3] Batsidis, A. and Zografos, K. (2005). Statistical inference for location and scale of elliptically contoured models with monotone missing data. *Journal of Statistical Planning and Inference* (In Press).

[4] Blattberg, R. C. and Gonedes, N. J. (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business,***47**, 244-280.

[5] Chung, H. and Han, C. (2000). Discriminant analysis when a block of observations is missing. *Ann. Ins. Statist. Math.,* **52,** 544-556.

[6] Diaz-Garcia, J. A., Rojas, Manuel Galea and Leiva-Sanchez, V. (2003). Influence diagnostics for elliptical multivariate linear regression models. *Communications in Statistics-Theory and Methods*, **32**, 625-642.

[7] Fama, E. F. (1965). The behaviour of stock market prices. *Journal of Business,* **38**, 34-105.

[8] Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London, New York.

[9] Fang, K. T and Zhang, Y. T. (1990). *Generalized Multivariate Analysis.* Science Press Beijing and Springer-Verlag, Berlin.

[10] Fujisawa, H. (1995). A note on the maximum likelihood estimators for multivariate normal distribution with monotone data. *Communications in Statistics-Theory and Methods*, **24**, 1377-1382.

[11] Galea, M., Paula, G. A. and Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, **46**, 71-79.

[12] Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions.* Chapman and Hall.

[13] Gupta, A. K. and Varga, T. (1993). *Elliptically Contoured Models in Statistics.* Kluwer Academic Publishers.

[14] Hao, J. and Krishnamoorthy, K. (2001). Inferences on a normal covariance matrix and generalized variance with monotone missing data. *J. Multiv. Analysis*, **78**, 62-82.

[15] Jinadasa, K. G. and Tracy, D. S. (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample. *Communications in Statistics-Theory and Methods* , **21**, 41-50.

[16] Kanda, T. and Fujikoshi, Y. (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.

[17] Kano, Y. (1994). consistency property of elliptical probability density functions. *Journal of Multivariate Analysis*, **51**, 139-147.

[18] Krishnamoorthy, K. and Pannala, M. K. (1998). Some simple test procedures for normal mean vector with incomplete data. *Ann. Ins. Statist. Math.,* **50**, 531-542.

[19] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* Wiley, New York.

[20] Little, R. J. A. (1992). Regression with Missing X's: A review. *Journal of American Statistical Association*, **87**, 1227-1237.

[21] Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of American Statistical Association*, **91**, 1219-1227.

[22] Liu, S. (2002). Local Influence in multivariate elliptical linear regression models. *Linear Algebra and Its Applications*, **354**, 159-174.

[23] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley and Sons, New York.

[24] Raats, V. M., van der Genugten, B. B. and Moors, J. J. A. (2002). Multivariate regression with monotone missing observations of the dependent variables. CentER Discussion Paper No. 2002-63(corrected version). Tilburg University. http://center.uvt.nl/phd_stud/raats/paper3c.pdf

[25] Raats, V. M., van der Genugten, B. B. and Moors, J. J. A. (2004). Asymptotics of multivariate regression with consecutively added variables. CentER Discussion Paper No.2004-77, Tilburg University. http://greywww.kub.nl:2080/greyfiles/center/2004/doc/77.pdf

[26] Raats, V. M. (2004). Approximations of the generalized Wilks' distribution. CentER Discussion Paper No.2004-85, Tilburg University. http://greywww.kub.nl:2080/greyfiles/center/2004/doc/85.pdf

[27] Rao, C. R. (1973). *Linear statistical inference and its applications.* John Wiley and Sons, New York.

[28] Rao, C. R. and Toutenburg, H. (1999). *Linear models: least squares and alternatives.* Springer- Verlag.

[29] Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of American Statistical Association*, **90**, 122-129.

[30] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581-592.

[31] Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data.* Chapman and Hall.

[32] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook.* American Sciences Press.

[33] Sutradhar, B. C. and Ali, M. M. (1986). Estimation of the parameters of a regression model with a multivariate t error variable. *Communications in Statistics-Theory and Methods*, **15**, 429-450.

[34] Tang, G., Little, R. J. A. and Raghunathan, T. E. (2003). Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika*, **90**, 747-764.

[35] Zellner, A. (1976). Bayesian and Non-Bayesian Analysis of the regression model with multivariate student-t error terms. *Journal of American Statistical Association*, **71**, 400-405.