RT-MAE-8510

M-METHODS IN GROWTH CURVE ANALYSIS

by

Júlio da Motta Singer

and

Pranab K. Sen

Palavras Chaves: Asymptotic normality, computational algorithm, elliptically symmetric distribution, growth curve model, M-procedures, robustness
(Key words)

Classificação AMS: 62G35, 62H15
(AMS Classification)

- Junho de 1985 -

M-METHODS IN GROWTH CURVE ANALYSIS

Julio M. Singer
Departamento de Estatística, Universidade de São Paulo
Brazil

Pranab K. Sen
Department of Biostatistics, University of North Carolina at Chapel Hill
USA

ABSTRACT: Two types of multivariate M-estimators of the parameter matrix in the standard growth curve model are obtained via the Potthoff-Roy transformation. Their asymptotic distributions are derived through an extension of the methods considered in Singer and Sen (1985) and computational algorithms are suggested. The problem of optimality in the choice of the transformation matrix is discussed and a numerical example is used to compare Huber-type M-estimates to the Normal maximum likelihood estimates.

AMS Subject Classifications: 62G35 , 62H15

Key words: Asymptotic normality, computational algorithm, elliptically symmetric distribution, growth curve model, M-procedures, robustness.

1. Introduction

In the statistical literature, the label "longitudinal data" is generally associated with situations in which the response of each observational unit is measured along a certain period of time or over several doses of a certain drug. Many techniques are available to analyze this type of data; among these, Growth Curve Analysis refers to multivariate methods of fitting curves to the average response of the populations under study. In this setting, we are particularly interested in the following growth curve model suggested by Potthoff and Roy (1964):

$$\underset{\sim}{Z} = \underset{\sim}{X}\underset{\sim}{\beta}\underset{\sim}{G} + \underset{\sim}{\tau} \tag{1.1}$$

where $\underset{\sim}{Z}$ is a (nxq) matrix of observable random variables, $\underset{\sim}{X}$ is a (nxr) between--subject specification matrix of known constants assumed of full rank r(<n), $\underset{\sim}{\beta}$ is a (rxp) matrix of unknown parameters, $\underset{\sim}{G}$ is a (pxq) within-subject specification matrix of known constants, assumed of full rank $p(\leq q)$ and $\underset{\sim}{\tau}$ is a (nxq) matrix of random errors. The rows of $\underset{\sim}{\tau}$ are assumed to follow independent distributions with distribution function $F^{(1)}$, location vector $\underset{\sim}{0}$ and positive definite (p.d.) scatter matrix $\underset{\sim}{\Sigma}^{(1)} = ((\sigma_{ij}^{(1)}))$ in the sense of Maronna (1976). This model has been extensively studied in the literature in the case where $F^{(1)}$ is assumed to be multi--normal. The reader is referred to Woolson and Leeper (1980), Timm (1980) and Geisser (1980) for recent reviews on the subject. In this paper we are concerned with the study of model (1.1) under some more general assumptions on $F^{(1)}$. In particular we consider two types of M-estimators of the parameter matrix $\underset{\sim}{\beta}$ obtained (as in the Normal theory case) after transforming the growth curve model (1.1) to the standard multivariate linear model:

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{1.2}$$

where $\underset{\sim}{X}$ and $\underset{\sim}{\beta}$ are as in (1.1), $\underset{\sim}{Y}$ is a (nxp) matrix of (transformed) random variables and $\underset{\sim}{\varepsilon}$ is a (nxp) matrix of (transformed) random errors, the rows of which are independently distributed with distribution function F, location vector $\underset{\sim}{0}$ and scatter matrix $\underset{\sim}{\Sigma} = ((\sigma_{ij}))$, assumed p.d.

In Section 2 we define the two types of M-estimators of $\underset{\sim}{\beta}$ under model (1.2) and summarize the assumptions required to derive their asymptotic distributions. Furthermore, we indicate that if these assumptions are satisfied by the components of the original growth curve model (1.1) they will also be satisfied by those of the reduced model (1.2) obtained via the Potthoff-Roy transformation. We then present the corresponding asymptotic distributions as a direct consequence of the results of Singer and Sen (1985) and propose computational algorithms. In Section 3 we discuss optimality in the choice of the Potthoff-Roy transformation. Finally, in Section 4, we compare the Normal maximum likelihood estimates with Huber-type robust M-estimates in a numerical example with data from the statistical literature.

2. The proposed M-estimators: computational algorithms and asymptotic distributions

Consider the model (1.2) with the following notation:

(i) $\underset{\sim}{y}_k(p\times 1)$, $\underset{\sim}{\varepsilon}_k(p\times 1)$ and $\underset{\sim}{x}_k(r\times 1)$, $k=1,\dots,n$ denote the transposes of the row vectors of $\underset{\sim}{Y}$, $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{X}$, respectively; (ii) $\underset{\sim}{\beta}_j(r\times 1)$ denotes the j^{th} column of $\underset{\sim}{\beta}$, $j=1,\dots,p$. Also let $\underset{\sim}{A}^*(ba\times 1)$ denote the row expansion of an $(a\times b)$ matrix $\underset{\sim}{A}$. Following the lines of Maronna (1976) we may define an M-estimator of $\underset{\sim}{\beta}$ as a solution $\hat{\underset{\sim}{\beta}}$ to:

$$M_{nij}(\underset{\sim}{Y},\hat{\underset{\sim}{\Sigma}},\hat{\underset{\sim}{\beta}}) = \sum_{k=1}^{n} u(d_k)x_{ki}(y_{kj} - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}_j) = 0 \text{ , } i=1,\dots,r \text{ ; } j=1,\dots,p \quad (2.1)$$

where $d_k^2 = (\underset{\sim}{y}_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})^T \hat{\underset{\sim}{\Sigma}}^{-1}(\underset{\sim}{y}_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})$, $u(d)$ is a score function defined for $d \geq 0$ and $\hat{\underset{\sim}{\Sigma}}$ is a $n^{1/2}$-consistent estimator of $\underset{\sim}{\Sigma}$. $\hat{\underset{\sim}{\beta}}$ can be viewed as a generalized maximum likelihood estimator in the case where F is elliptically symmetric; note that $u(d) = 1$ defines the maximum likelihood estimator for F Normal.

An alternative (coordinatewise) M-estimator which corresponds to a generalized least squares estimator was suggested by Gnanadesikan (1977, p.136) and is defined as a solution $\hat{\underset{\sim}{\beta}}$ to:

$$M_{nij}(\underset{\sim}{Y},\hat{\underset{\sim}{\sigma}},\hat{\underset{\sim}{\beta}}) = \sum_{k=1}^{n} x_{ki}\psi_j\{(y_{kj} - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}_j)/\hat{\sigma}_j\} = 0 \text{ , } i=1,\dots,r \text{ ; } j=1,\dots,p \quad (2.2)$$

where $\hat{\underset{\sim}{\sigma}} = (\hat{\sigma}_1,\dots,\hat{\sigma}_p)^T$ is an $n^{1/2}$-consistent estimator of $\underset{\sim}{\sigma} = (\sigma_1,\dots,\sigma_p)^T$, $\sigma_j^2 = \sigma_{jj}$, $j=1,\dots,p$ and the ψ_j are suitable score functions (for simplicity we take $\psi_1 = \dots = \psi_p = \psi$); note that $\psi(x) = x$ produces the least squares estimator .

In order to derive the asymptotic distributions of the proposed M-estimators, the following assumptions (which are standard in the literature on M-methods) are required:

A1. F is absolutely continuous with a symmetric density function f such that $f_j'(\underset{\sim}{\varepsilon}) = (\partial/\partial\varepsilon_j)f(\underset{\sim}{\varepsilon})$, j=1,...,p exist almost everywhere (a.e.).

A2. F has finite and positive definite (p.d.) Fisher information matrices with respect to both location and scale.

A3. F is elliptically symmetric, i.e. its density function is given by $f(\underset{\sim}{\varepsilon}) = |\underset{\sim}{\Sigma}|^{-1/2}h\{(\underset{\sim}{\varepsilon}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\varepsilon})^{1/2}\}$ where h is a scalar multiple of a density in $\mathbb{R}$.

B. The elements of the specification matrix $\underset{\sim}{X}$ satisfy:

(i) Noether's condition: $\max_{1\le k\le n} \{\underset{\sim}{x}_k^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{x}_k\} \to 0$ as $n \to \infty$,

(ii) $\lim_{n\to\infty} \{n^{-1}(\underset{\sim}{X}^T\underset{\sim}{X})\} = \underset{\sim}{V} = (\underset{\sim}{v}_1,\ldots,\underset{\sim}{v}_r)$ is a p.d. matrix

C1. The score function ψ can be expressed as a sum of two nondecreasing skew--symmetric functions $\psi_{(1)}$ and $\psi_{(2)}$ where $\psi_{(1)}$ is absolutely continuous on any bounded interval in $\mathbb{R}$ with derivative $\psi_{(1)}'$ a.e. and $\psi_{(2)}$ is a step function defined as follows. For some positive integer s, assume that there exist open intervals $E_\ell = (a_\ell, a_{\ell+1})$, $\ell=0,\ldots,s$, with $a_0 = -\infty < a_1 < \ldots < a_s < a_{s+1} = \infty$ such that $\psi_{(2)}(x) = \theta_\ell$ for $x \in E_\ell$, $0 \le \ell \le s$ where θ_ℓ are real numbers (not all equal). Conventionally we let $\psi_{(2)}(a_\ell) = (\theta_{\ell-1} + \theta_\ell)/2$ for $\ell=1,\ldots,s$. The function ψ also satisfies (i) $\rho_j^2 = \int\psi^2(\varepsilon/\sigma_j)f_j(\varepsilon)d\varepsilon < \infty$ and (ii) $w_j = -\int\psi(\varepsilon/\sigma_j)f_j'(\varepsilon)d\varepsilon < \infty$, j=1,...,p, where f_j is the marginal density corresponding to the j^{th} coordinate.

C2. The score function u(·) is nonnegative, nonincreasing and continuous. To facilitate comparison with the score function defined above we let $u(d)=\psi(d)/d$

where $\psi = \psi_{(1)}$ and is assumed bounded.

Using an asymptotic linearity result due to Jurečková (1977), Singer and Sen (1985) obtained the asymptotic distribution of the coordinatewise M-estimator (2.2) under A1, A2, B and C1 and also that of the Maronna-type M-estimator (2.1) under A1-A3, B and C2. They also indicated how one can construct tests of linear hypotheses of the form H: $\mathbf{C}\boldsymbol{\beta}\mathbf{U} = \mathbf{K}$ where $\mathbf{C}(c \times r)$ and $\mathbf{U}(p \times u)$ are known matrices of full row and column ranks $c(\leq r)$ and $u(\leq p)$ respectively and $\mathbf{K}(c \times u)$ is any known matrix.

Now consider the growth curve model (1.1); then, following the suggestion of Pothoff and Roy (1964) let $\mathbf{H}_1$ be any $(g \times p)$ matrix such that $\mathbf{G}\mathbf{H}_1 = \mathbf{I}$ and observe that by making the transformation $\mathbf{Y} = \mathbf{Z}\mathbf{H}_1$, $\boldsymbol{\varepsilon} = \boldsymbol{\tau}\mathbf{H}_1$ the original model (1.1) reduces to (1.2). Therefore we may use (2.1) and (2.2) to obtain estimates of the parameter matrix $\boldsymbol{\beta}$ and since assumptions A1-C2 are invariant with respect to linear transformations their asymptotic distributions may be obtained from the results of Singer and Sen (1985). In the coordinatewise case we have:

$$n^{1/2}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*) \approx N_{rp}(\mathbf{0}, \mathbf{V}^{-1} \otimes \mathbf{W}^{-1}\boldsymbol{\Phi}\mathbf{W}^{-1}) \tag{2.3}$$

where $\mathbf{W} = \mathrm{diag}(w_1, \ldots, w_p)$, $w_j = -\int \psi(\varepsilon/\sigma_j) f_j'(\varepsilon) d\varepsilon$, $\boldsymbol{\Phi} = ((\phi_{ij}))$,

$\phi_{ij} = \int \psi(\varepsilon_i/\sigma_i)\psi(\varepsilon_j/\sigma_j) f(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon}$, $i,j = 1, \ldots, p$. For the Maronna-type M-estimator we have:

$$n^{1/2}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*) \approx N_{rp}(\mathbf{0}, \mathbf{V}^{-1} \otimes ab^{-2}\mathbf{H}_1^T \boldsymbol{\Sigma} \mathbf{H}_1) \tag{2.4}$$

where $a = E_F\{\psi^2(d_k)/p\}$, $b = E_F\{p^{-1}\psi'(d_k) + (1-p)^{-1}\psi(d_k)/d_k\}$

and $d_k^2 = (\mathbf{y}_k - \boldsymbol{\beta}^T \mathbf{x}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_k - \boldsymbol{\beta}^T \mathbf{x}_k)$.

In general, the M-estimators defined by either (2.1) or (2.2) must be obtained by iterative methods; we propose here two computational procedures which are generalizations of the Type 2 algorithms discussed in Klein and Yohai (1981) and essentially correspond to approximations of Newton-Raphson's method. For the coordinatewise case, the algorithm is given by:

$$\begin{cases} \hat{\beta}^*_{(0)} = \tilde{\beta}^* \\ \hat{\beta}^*_{(m+1)} = \hat{\beta}^*_{(m)} + \{(X^TX) \otimes \hat{W}(\hat{\beta}_{(m)})\}^{-1} M^*_n(\hat{\beta}_{(m)}) \quad , m \geq 0 \end{cases} \tag{2.5}$$

where $\hat{W}(T) = \text{diag}\{\hat{w}_1(T),\ldots,\hat{w}_p(T)\}$, $\hat{w}_j(T) = (n\hat{\sigma}_j)^{-1}[\sum_{k=1}^{n} \psi'_{(1)}\{(y_{kj}-x_k^T T_j)/\hat{\sigma}_j\} + $ $+ n \sum_{\ell=1}^{s} (\theta_\ell-\theta_{\ell-1})\hat{f}_j(a_\ell)]$, $\hat{f}_j(a_\ell) = n^{1/2}\{\hat{F}_j(a_\ell+n^{-1/2})-\hat{F}_j(a_\ell-n^{-1/2})\}/2$, $\hat{F}_j$ is the empirical distribution function of the standardized residuals corresponding to the j^{th} coordinate, $\tilde{\beta}$ and $\hat{\sigma}_j$ are $n^{1/2}$-consistent estimators of β and σ_j, respectively and $M_n(T)$ is given by $M_{nij}(T) = \sum_{k=1}^{n} x_{ki}\psi\{(y_{kj}-x_k^T T_j)/\hat{\sigma}_j\}$, $i=1,\ldots,r$, $j=1,\ldots,p$.

A similar algorithm for the Maronna-type M-estimator is given by:

$$\begin{cases} \hat{\beta}^*_{(0)} = \tilde{\beta}^* \\ \hat{\beta}^*_{(m+1)} = \hat{\beta}^*_{(m)} + \{(X^TX) \otimes \hat{b}_0(\hat{\beta}_{(m)})I\}^{-1} M^*_n(\hat{\beta}_{(m)}) \quad , m \geq 0 \end{cases} \tag{2.6}$$

where $\hat{b}_0(T) = (pn)^{-1} \sum_{k=1}^{n} \psi'\{\hat{d}_k(T)\} + \{n(1-p^{-1})\}^{-1} \sum_{k=1}^{n} \psi\{\hat{d}_k(T)\}/\hat{d}_k(T)$,

$\hat{d}_k^2(T) = (y_k - T^T x_k)^T \hat{\Sigma}^{-1}(y_k - T^T x_k)$, $M_n(T)$ is given by

$M_{nij}(T) = \sum_{k=1}^{n} u\{\hat{d}_k(T)\}x_{ki}(y_{kj} - x_k^T T_j)$, $i=1,\ldots,r$, $j=1,\ldots,p$ and $\tilde{\beta}$ and $\hat{\Sigma}$ are $n^{1/2}$-consistent estimators of β and Σ , respectively. Convergence (in probability) of (2.5) and (2.6) follows from the asymptotic linearity results (3.1) and (4.1) of

Singer and Sen (1985) respectively.

Least squares estimators can be taken as starting points for the iterations. However, if the score functions are defined to produce robust M-estimators, it is preferable to use some more robust initial estimates, such as the least absolute residual estimator for $\underset{\sim}{\beta}$ and the median absolute deviation estimator for $\underset{\sim}{\sigma}$ although they are computationally more elaborate. In the case of Maronna's method, robust estimates of $\underset{\sim}{\Sigma}$ are difficult to obtain and a few suggestions are presented in Devlin et al. (1981).

3. Optimality in the choice of the Potthoff-Roy transformation

In order to obtain the reduction from (1.1) to (1.2), Potthoff and Roy (1964) suggested that we take

$$\underset{\sim}{H}_1 = \underset{\sim}{A}^{-1}\underset{\sim}{G}^T(\underset{\sim}{G}\underset{\sim}{A}^{-1}\underset{\sim}{G}^T)^{-1} \tag{3.1}$$

as the transformation matrix, where $\underset{\sim}{A}(q\text{x}q)$ is an arbitrary, but fixed p.d. matrix. They showed that if the underlying distribution $F^{(1)}$ is multi-normal, the choice $\underset{\sim}{A} = \underset{\sim}{\Sigma}^{(1)}$ in (3.1) and $u(d) = -d^{-1}(\partial/\partial d)\ \log h(d)$ in (2.1) or $\psi_j(x) = -(\partial/\partial x)\log f_j(x)$ in (2.2) leads to the (fully efficient) Normal maximum likelihood estimator. In our case the optimum choice of $\underset{\sim}{A}$ is not so clear cut, since we are not specifying the functional form of the underlying distribution nor the score functions defining the M-estimators. A possible approach, however, is to maximize the asymptotic efficiency of the M-estimator in its general form at some specified model, say the Normal model. This follows the spirit of Hampel's extremal problem (see Huber (1981, ch. 11) for example) and has a special appeal for robust M-estimators; the optimum choice of $\underset{\sim}{A}$ in such a case would lead to estimators which perform as well as possible at the Normal model while retaining robustness against departures from it.

First observe that we might take $\Sigma^{(1)} = I$ and G such that $GG^T = I$ with no loss of generality. In this case, $\mathrm{Var}(\varepsilon_k) = H_1^T H_1 = H = ((h_{ij}))$, $i,j=1,\dots,p$; recall that H is a function of A. Now suppose that the underlying distribution is unknown but that assumptions A1, A2, B and C1 hold. Expressing the asymptotic covariance matrix of the coordinatewise M-estimator as a function of A we may write $\Gamma(A) = W^{-1}\ddagger W^{-1} = ((\gamma_{ij}(A)))$, where:

$$\gamma_{ij}(A) = h_i h_j [E_F\psi(\varepsilon_i/h_i)\psi(\varepsilon_j/h_j)/\{\int f_i^*(\nu_i)d\psi(\nu_i)\}\{\int f_j^*(\nu_j)d\psi(\nu_j)\}] \quad (3.2),$$

$\int f_i^*(\nu_i)d\psi(\nu_i) = E_F\psi'_{(1)}(\varepsilon_i/h_i) + \sum_{\ell=1}^{s}(\theta_\ell-\theta_{\ell-1})f_i^*(a_\ell)$, $h_i = h_{ii}^{1/2}$ and f_i^*, $i=1,\dots,p$ is the density function of the standardized residuals corresponding to the i^{th} coordinate. Considering the three most usual criteria of multivariate optimality (A,D and E optimality) we can maximize the asymptotic efficiency of the estimator by minimizing either $\mathrm{tr}\{\Gamma(A)\}$, $|\Gamma(A)|$ or $ch_1\{\Gamma(A)\}$, where $\mathrm{tr}(M)$, $|M|$ and $ch_1(M)$ respectively denote the trace, the determinant and the largest characteristic root of a matrix M.

Letting $\lambda_j = \lambda_j(A)$, $j=1,\dots,p$ denote the characteristic roots of $\Gamma(A)$, observe that: (i) $(\partial/\partial\lambda_j)\prod_{i=1}^{p}\lambda_i = \prod_{i\neq j}\lambda_i > 0$, $j=1,\dots,p$ and (ii) $(\partial/\partial\lambda_j)\sum_{i=1}^{p}\lambda_i = 1$, $j=1,\dots,p$, which imply that both $|\Gamma(A)|$ and $\mathrm{tr}\{\Gamma(A)\}$ are increasing functions in each λ_i when all the others are held fixed. Also note that since $\Gamma(A)$ is symmetric and p.d. there exists an orthogonal matrix $P = P(A)$ such that $P\Gamma(A)P^T = \mathrm{diag}\{\lambda_1,\dots,\lambda_p\} = D$ and therefore: $ch_1\{\Gamma(A)\} = \sup_x\{x^T\Gamma(A)x/x^Tx\} = \sup_x\{x^TP^TDPx/x^TP^TPx\} = \sup_y \sum_{i=1}^{p}\lambda_i y_i^2/y^Ty$ where $y = Px$. Now, since $(\partial/\partial\lambda_j)\sum_{i=1}^{p}\lambda_i y_i^2/y^Ty = y_j^2/y^Ty > 0$, $j=1,\dots,p$, we conclude that $ch_1\{\Gamma(A)\}$ is also an increasing function in each λ_i when all the others are held fixed. Thus, to maximize the asymptotic efficiency of the estimator according to any of the above criteria it suffices to minimize each λ_j separately. For symmetry reasons we must have $\lambda_j = \lambda = \lambda(\underset{\sim}{A})$, $j=1,\ldots,p$ at the minimum, which implies $\underset{\sim}{P}\underset{\sim}{\Gamma}(\underset{\sim}{A})\underset{\sim}{P}^T=\lambda\underset{\sim}{I}$ and consequently $\underset{\sim}{\Gamma}(\underset{\sim}{A}) = \lambda\underset{\sim}{P}^T\underset{\sim}{P} = \lambda\underset{\sim}{I}$. Then, from (3.2) we conclude that at the minimum:

$$\underset{\sim}{\Gamma}(\underset{\sim}{A}) = h_{ii}(\underset{\sim}{A})E_F\psi^2(\varepsilon_i/h_i)/\{\int f_i^*(\nu_i)d\psi(\nu_i)\}^2\underset{\sim}{I}$$

Now observe that because of scale-invariance only $h_{ii}(\underset{\sim}{A})$ depends on the choice of $\underset{\sim}{A}$ in the expression above. Therefore the problem reduces to that of the minimization of $h_{ii}(\underset{\sim}{A})$ and from the classical theory it follows that the minimum at the Normal model corresponds to the choice $\underset{\sim}{A} = \underset{\sim}{I}$ (or $\underset{\sim}{A} = \underset{\sim}{\Sigma}$ in the general case).

To obtain a similar result for Maronna-type M-estimators we need the additional assumption that the underlying distribution is elliptically symmetric. In such a case the distribution of the (transformed) random errors $\underset{\sim}{\varepsilon}_k$ is also elliptically symmetric and has covariance matrix $\underset{\sim}{H}_1^T\underset{\sim}{\Sigma}\underset{\sim}{H}_1$. Now observe that the transformation $\underset{\sim}{\nu}_k = (\underset{\sim}{H}_1^T\underset{\sim}{\Sigma}\underset{\sim}{H}_1)^{-1/2}\underset{\sim}{\varepsilon}_k$ preserves elliptical symmetry and implies that $Var(\underset{\sim}{\nu}_k) = \underset{\sim}{I}$. Consequently, $d_k^2 = \underset{\sim}{\varepsilon}_k^T(\underset{\sim}{H}_1^T\underset{\sim}{\Sigma}\underset{\sim}{H}_1)^{-1}\underset{\sim}{\varepsilon}_k = \underset{\sim}{\nu}_k^T\underset{\sim}{\nu}_k$ which implies that both a and b in (2.4) are invariant with respect to the choice of $\underset{\sim}{H}_1$ (or equivalently of $\underset{\sim}{A}$). Then, using an argument similar to that of the coordinatewise case it follows from (2.4) and (3.1) that the optimal choice at the Normal model corresponds to $\underset{\sim}{A} = \underset{\sim}{\Sigma}$.

4. Numerical Illustration

The practical appeal of the methoods described in Section 2 is mainly related to the possibility of inference based on estimates which are robust with respect to the presence of outliers or extreme values in the data. In this section we illustrate such a capability through the actual computation of Huber--type M-estimates for a set of data previously considered in the literature by Potthoff and Roy (1964). It is well known that the robustness of such M-estimates in linear models is impaired by the presence of outliers in the explanatory variables. However, as indicated in Singer and Sen (1985), the proposed M-estimates are reasonably robust when the specification matrix $\underset{\sim}{X}$ corresponds to a cell means model, as in the case under consideration. In the presence of high leverage points - that is, of points where the diagonal of the hat matrix $\underset{\sim}{H} = \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T$ exceeds a certain bound -- Huber-type M-estimates should be used with caution. In particular, some other type of robust method such as the efficient bounded-influence regression procedures suggested by Krasker and Welsch (1982) should be considered.

The dataset consists of measurements of the distance in millimeters from the centre of the pituitary to the pterio-maxillar fissure on each of 11 girls and 16 boys at the ages of 8, 10, 12 and 14 years. The matrices $\underset{\sim}{Z}, \underset{\sim}{X}, \underset{\sim}{\beta}$ and $\underset{\sim}{G}$ in (1.1) respectively correspond to $\underset{\sim}{X}_0$, $\underset{\sim}{A}, \underset{\sim}{\xi}$ and $\underset{\sim}{P}$ in Potthoff and Roy (1964). As in that case,we assumed that a linear growth was appropriate to model the mean distance as a function of time; furthermore, for the sake of simplicity, we carried out the computations with $\rho=0$ in their expression (13). The intention here is only to illustrate the computational results as opposed to present an exhaustive analysis of the data.

The algorithms (2.5) and (2.6) were programmed using the MATRIX procedure of the Statistical Analysis System (SAS) package. For both the coordinatewise and the Maronna-type M-estimates, the score function was of the form $\psi(x) = \min(|x|,k)\ \mathrm{sign}(x)$ where k is an appropriate tuning constant. For teh Maronna-type M-estimate the tuning constant $k=k_M$ was computed as a solution to $P\{\underset{\sim}{\varepsilon}^T \underset{\sim}{\varepsilon} \leq k_M^2\} = .8664$ where $\underset{\sim}{\varepsilon} \sim N_p(\underset{\sim}{0},\underset{\sim}{I})$; the corresponding tuning constant $k=k_C$ for the coordinatewise M-estimate was chosen in such a way that the volume of the "truncation" region was the same as that of the Maronna-type M-estimate. In our case p=2, k_M=2.01 and k_C 1.78 which correspond to k=1.5 in the univariate situation. The starting value for the iterative computation of both types of estimates was the least squares estimate; in the coordinatewise case we used the median absolute deviation multiplied by 1.48 (to make it approximately unbiased at the Normal model) as an estimate of scale; the usual $\underset{\sim}{S}$ matrix was used to estimate the scatter matrix for the Maronna-type M-estimate algorithm. The convergence criteria consisted of stopping iterations when the Euclidean norm of the difference betweeen the estimates from two consecutive steps was < .001. The algorithms converged in 4 iterations and the results are presented in Table 4.1 along with the least squares (or Normal maximum likelihood) estimates for comparison; estimates of the corresponding asymptotic standard errors are indicated in parentheses.

Table 4.1: Least squares (LS), coordinatewise (CO) and Maronna-type (MA) M-estimates of the parameter matrix for the Potthoff and Roy (1964) data

		Intercept	Slope
Girls	LS	22.648 (0.586)	0.480 (0.104)
	MA	22.673 (0.537)	0.480 (0.095)
	CO	22.694 (0.591)	0.480 (0.086)
Boys	LS	24.969 (0.486)	0.784 (0.086)
	MA	24.946 (0.445)	0.752 (0.079)
	CO	24.890 (0.490)	0.745 (0.071)

There is quite close agreement among all the three estimates which is an indication that outliers do not seem to be a problem in this dataset. In fact we computed the weight function $\psi(x)/x$ and in the coordinatewise case only 5 of the 54 values had weights < 1, all but one being > .72; in the Maronna-type case only 3 of the 27 values had weights < 1, all being > .56.

For illustrative purposes we followed an idea of Pendergast and Broffitt (1981) and introduced one "outlier" by supposing that row 4 of the observation matrix read: (35.0 45.0 50.0 65.0) instead of (23.5 24.5 25.0 26.5); this could have happened had the data been punched on cards and the above row been misaligned by one column. We recalculated the values of the three estimators using the same parameters as in the previous case. The results are indicated in Table 4.2.

Table 4.2: Least squares (LS), coordinatewise (CO) and Maronna-type (MA) M-estimates of the parameter matrix for the Potthoff and Roy (1964) data with an hypothetical outlier

		Intercept	Slope
Girls	LS	24.818 (1.617)	0.868 (0.267)
	MA	23.477 (1.191)	0.651 (0.196)
	CO	22.861 (0.692)	0.548 (0.103)
Boys	LS	24.969 (1.341)	0.784 (0.221)
	MA	24.984 (0.989)	0.759 (0.163)
	CO	24.947 (0.574)	0.752 (0.086)

As expected, the least squares procedure was the one most affected by the change: not only were the point estimates for the parameters related to the group of girls considerably different from the values obtained with the original data, but also the estimated standard errors were inflated in comparison to the ones obtained through the two robust M-methods. Although both M-methods succeeded in downweighting the outlier, the performance of the coordinatewise M-estimator was somewhat better than that of the Maronna-type one. This is probably related to the fact that a robust estimate of scatter was used in the computation of the former, while a non-robust one was considered for the latter. We conjecture that improved results for both types of estimates would be obtained if better estimates of scale were available, at least in situations with no high leverage points such as the one being analyzed.

With a rather tentative spirit we also indicated the possibility of testing hypotheses via M-emthods; we considered tests of the hypothesis of no difference between the boys'and the girls'curves using both the Wald-type tests and the score-type tests suggested in Singer and Sen (1985). In all cases the test statistics follow asymptotic chi-squared distributions with 2 degrees of freedom. The results are indicated in Table 4.3.

Table 4.3: Asymptotic M-tests of the hypothesis of equal curves for girls and boys - Potthoff and Roy (1964) data (p-values in parentheses).

		Wald's tests		Score tests	
		Coordinatewise	Maronna	Coordinatewise	Maronna
Original Data	LHT, RLR,	11.6564 (.0029)	14.1500 (.0009)	9.1012 (.0106)	10.3958 (.0055)
	WLR	9.1851 (.0101)	10.7646 (.0046)	7.5034 (.0235)	8.4089 (.0149)
"Outlier" Introduced	KHT, RLR,	6.0395 (.0488)	1.7456 (.4178)	4.8173 (.0899)	1.6513 (.4380)
	WLR	5.1932 (.0753)	1.6199 (.4449)	4.2495 (.1195)	1.5394 (.4632)

LHT: Lawley-Hotelling's trace statistic

RLR: Roy's largest root statistic

WLR: Wilk's likelihood ratio statistic

When the original data were considered, all tests lead to the rejection of the null hypothesis; although the associated P-values were of different magnitudes, they seemed to be consistent with the corresponding results from the Normal theory case where tests based on scores tend to be more conservative than Wald-type tests. The tests did not seem to be as robust to the introduction of an outlier as expected and we conjecture that this is due both to the moderate sample size (n=27) and to the use of non-robust estimates of scatter, especially in the Maronna-type case. Clearly some further research is needed to clarify these issues.

ACKNOWLEDGEMENTS

The work of the first author was done while he was on leave at the University of North Carolina, Chapel Hill; it was partially supported by a grant from CAPES, an agency of the Brazilian Ministry of Education and Culture. The work of the second author was partially supported by the Contract NIH-NHLBI--71-2243-L.

REFERENCES

[1] Devlin, S.J., Gnanadesikan, R. and Kettenring, J.R. (1981). Robust estimation of dispersion matrices and principal components. J.Amer.Statist.Assoc. 76, 354-362.

[2] Geisser, S. (1980). Growth Curve Analysis. In Handbook of Statistics, Vol. 1 (ed. P.R. Krishnaiah), North Holland, Amsterdam, pp. 89-115.

[3] Gnanadesikan, R. (1977). Methods for statistical Data Analysis of Multivariate Observations. Wiley, New York.

[4] Grizzle, J.E. and Allen, D.M. (1969). Analysis of growth and dose response curves. Biometrics 25, 357-381

[5] Huber, P.J. (19[illegible]). Robust Statistics. John Wiley, New York.

[6] Jurečkovā, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. Ann. Statist. 5, 464-472.

[7] Klein, R. and Yohai, V.J. (1981). Asymptotic behaviour of iterative M-estimators for the linear model. Commun. Statist. A10, 2373-2388

[8] Krasker, W.S. and Welsch, R.E. (1982). Efficient bounded-influence regression estimation. J. Amer. Statist. Assoc. 77, 595-604

[9] Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. Ann. Statist. 4, 51-67.

[10] Pendergast, J.F. and Broffit, J.D.(1981). Robust estimation in growth curve models. Personal Communication (paper presented at the 141st Annual Meeting of the American Statistical Association, Detroit)

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

[12] SAS User's Guide, 1979 Edition, Hewlig, J.T. and Council, K.A. (eds.), SAS Institute, Cary, NC.

[13] Singer, J.M. and Sen, P.K. (1985). M-methods in multivariate linear models. J. Multivar. Anal. (in press)

[14] Timm, N.H. (1980). Multivariate analysis of variance of repeated measurements. In Handbook of Statistics, Vol.1 (ed. P.R. Krishnaiah), North Holland, Amsterdam, pp. 41-87

[15] Woolson, R.F. and Leeper, J.D. (1980). Growth curve analysis of complete and incomplete longitudinal data. Commun. Statist. A9, 1491-1513

oOo

RELATÓRIO TÉCNICO

DO

DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS

7901 - BORGES, W. de S. On the limiting distributios of the failure time of composite material. São Paulo, IME-USP,1979,22p.

7902 - GALVES, A.; LEITE, J.G.; ROUSSIGNOL, M. The invariance principle for the one-dimensional symmetric simple exclusion process. São Paulo, IME-USP, 1979, 9p.

8001 - MENTZ, R.P. et al. Exploratory fitting of autoregressive and moving average models to well-behaved time series data. São Paulo, IME-USP, 1980, 16p.

8002 - MORETTIN, P.A. Walsh spectral analysis. São Paulo, IME-USP, 1980, 27p.

8003 - RODRIGUES, J. Robust estimation and finite population. São Paulo, IME-USP, 1980, 13p.

8004 - BORGES, W. de S. & RODRIGUES, F.W. On the axiomatic theory of multistate coherent structures. São Paulo, IME-USP, 1980, 10p.

8005 - MORETTIN, P.A. A central limit theorem for stationary processes. São Paulo, IME-USP, 1980, 5p.

8101 - DANTAS, C.A.B. & COLUCCI, E. A Simulation program for emergency services-II, São Paulo, IME-USP, 1981, 14p.

8102 - ANDJEL,E.D. Invariant measures for the zero range process. São Paulo, IME-USP, 1981, 55p.

8103 - ANDJEL, E.D. The asymmetric simple exclusion process on Z^d. São Paulo, IME-USP, 1981, 13p.

8104 - MORETTIN, P.A. & TOLOI, C.M.C., Accuracy of forecasting with special reference to the Box-Jenkins and Bayesian Methodologies. São Paulo, IME-USP, 1981, 41p.

8105 - PINO, F.A. & MORETTIN, P.A., Intervention analysis applied to Brazilian coffee and milk time series. São Paulo .IME-USP. 1981, 36p.

8106 - BORGES, W.S. & RODRIGUES, J., Testing for new better than used in expectation. São Paulo, IME-USP, 1981, 7p.

8107 - FAHMY, S.; PEREIRA, C.A.B.; PROSCHAN, F., The influence of the sample on the posterior distribution. São Paulo, IME-

8108 - PERES, C.A., Asymptotic efficiency of the likelihood ratio conditional test for multinomial distributions. São Paulo IME-USP, 1981, 29p.

8109 - PERES, C.A., Testing the effect of blocking in a randomized complete block design (RCBD). São Paulo, IME-USP, 1981, 14p.

8110 - BASU, D. & PEREIRA, C.A.B., On the Bayesian analysis fo categorical data: the problem of nonresponse. São Paulo, IME-USP, 1981, 13p.

8201 - BASU, D. & PEREIRA, C.A.B., Conditional independence in statistics, São Paulo, IME-USP, 1982, 37p.

8202 - BASU, D. & PEREIRA, C.A.B., A note on Blackwell sufficiency and a Skibinsky characterization of distributions. São Paulo, IME-USP, 1982, 12p.

8203 - PERES, C.A., On the interpretation of the parameters of the quadratic model for cell survival ofter irradiation. São Paulo, IME-USP, 1982, 22p.

8204 - GALVES, A., et al. Rescaling the stirring process. São Paulo IME-USP, 1982, 23p.

8205 - RODRIGUES, J., On the asymptotic theory for the fixed size confidence ellipsoids. São Paulo, IME-USP, 1982, 14p.

8206 - PEREIRA, C.A.B. & RODRIGUES, J., Robust linear predicition in finite populations. São Paulo, IME-USP, 1982, 14p.

8207 - MORETTIN, P.A., Walsh-Fourier transforms. São Paulo,IME-USP 1982, 15p.

8208 - PERES, C.A. & MORETTIN, P.A., Building bridges between the academic and real wordls-some observations from South America. São Paulo, IME-USP, 1982 16p.

8209 - PEREIRA, C.A.B. & ROGATKO, A., The Hardy-Weinberg equilibrium under a Bayesian perspective. São Paulo, IME-USP , 1982, 16p.

8210 - MORETTIN, P.A., The Levinson algorithm and its applications in time series analysis. São Paulo, IME-USP, 1982, 16p.

8211 - RODRIGUES, J., A Note on Maximized and Conditional Likelihood Functions. São Paulo, IME-USP, 1982, 9p.

8301 - PEREIRA,C.A.B., Stopping rules and conditional inference in 2x2 contigence tables. São Paulo, IME-USP, 1983, 7p.

8302 - BOLFARINE, H., PEREIRA, C.A.B. & RODRIGUES, J., Robust Linear Prediction in Finite Populations: A Bayesian Perspective. São Paulo, IME-USP, 1983, 21p.

8303 - MORETTIN, P.A. et al., Rainfall at Fortaleza, Ceará, Brazil Revisited. São Paulo, IME-USP, 1983, 33p.

8304 - MORETTIN, P.A. & TOLOI, C.M.C., Evaluation of Forecasting Procedures: A Case Study. São Paulo, IME-USP, 1983,30p.

8305 - PERES, C.A., et al., Educating and training undergraduate applied statisticians. São Paulo, IME-USP, 1983, 13p.

8306 - PEREIRA, C.A.B., & LINDLEY, D.V., Examples Questioning the Use of Partial Likelihood. São Paulo, IME-USP, 1983, 10p.

8307 - MORETTIN, P.A. et al., Statistics in South America. São Paulo, IME-USP, 1983, 10p.

8308 - LINDLEY, D.V., Royal Statistical Society 150th Anniversary. São Paulo, IME-USP, 1983, 19p.

8309 - ANDJEL, E.D., Invariant Measures and Long Time Behaviour of the Smoothing Process. São Paulo, IME-USP, 1983, 25p.

8310 - BOLFARINE, H. et al., A General Theory of Prediction in Finite Populations. São Paulo, IME-USP, 1983, 42p.

8401 - BOLFARINE, H. & RODRIGUES, J., Characterization of Alternative Models for Robust Linear Prediction in Finite Populations. São Paulo, IME-USP, 1984, 12p.

8402 - PEREIRA, C.A.B. et al., Inversão de Condicionamento. São Paulo. IME-USP, 1984, 30p.

8403 - BOLFARINE, H. & RODRIGUES, J., On Bayesian Prediction of the Population Variance in Finite Populations. São Paulo, IME-USP, 1984, 21p.

8404 - ZACKS, S., Bayes Sequential Estimation of the Size of a Finite Population. São Paulo, IME-USP, 1984, 23p.

8405 - ROGATKO, A. et al., Bayesian Method for the Estimation of Penetrance: Application to Mandibulofacial and Fronto-Nasal Dysostoses. São Paulo, IME-USP, 1984, 67p.

8406 - SHIBATA, R., Identification and Selection of ARMA models. São Paulo, IME-USP, 1984, 17p.

8407 - MORETTIN, P.A. & MESQUITA, A.R., A Phase Angle Test for Periodic Components in Time Series. São Paulo, IME-USP, 1984, 27p.

8408 - SHIBATA, R., Selection of Regression Variables. São Paulo; IME-USP, 1984, 11p.

8409 - ESTON, V.R. et al., Chthamalus Bissinuatus (Cirripedia) and Brachidontes Solisianus (Bivalvia) Spatial Interactions: A Stochastic Model. São Paulo, IME-USP, 1984, 32p.

8410 - PINO, F.A. & MORETTIN, P.A., Forecasting Linear Combinations of Time Series. São Paulo, IME-USP, 1984, 30p.

8411 - SCHONMANN, R.H., <u>Metastability for the Contact Process</u>, São Paulo, IME-USP, 1984, 29p.

8412 - SCHONMANN, R.H., <u>Central Limit Theorem for the Contact Process</u>. São Paulo, IME-USP, 1984, 10p.

8413 - ANDRADE, D.F. & BOLFARINE, H., <u>Estimation in Covariance Components Models with Unequal Intraclass Variances</u>. São Paulo, IME-USP, 1984, 10p.

8501 - RODRIGUES, J. et al., <u>The EM-Algorithm for Finding the ML-Predictor for Finite Populations in Two-Stage Sampling</u>. São Paulo, IME-USP, 1985, 15p.

8502 - BOLFARINE, H. & RODRIGUES, J., <u>A Missing Value Approach to the Prediction Problem in Finite Populations</u>. São Paulo, IME-USP, 1985, 16p.

8503 - SCHONMANN, R., <u>A New Proof of the Complete Convergence Theorem for Contact Processes in Several Dimensions with Large Infection Parameter</u>. São Paulo, IME-USP, 1985, 11p.

8504 - ZACKS, S. & RODRIGUES, J., <u>A Note on the EM-Algorithm for Maximum Likelihood Estimation in Sampling from a Finite Population with a Multinormal Superpopulation Model</u>. São Paulo, IME-USP, 1985, 5p.

8505 - ANDJEL, E.D., <u>Convergence to a non Extremal Equilibrium Measure in the Exclusion Process</u>. São Paulo, IME-USP, 1985, 15p.

8506 - IRONY, T.Z. & PEREIRA, C.A.B., <u>Exact Tests for Equality of Two Proportions: Fisher x Bayes</u>. São Paulo, IME-USP, 1985, 30p.

8507 - SCHONMANN, R.H. & VARES, M.E., <u>The Survival of the Large Dimensional Basic Contact Process</u>. São Paulo, IME-USP, 1985, 14p.

8508 - ACHCAR, J.A., <u>Modelos de Regressão com Dados Censurados</u>. São Paulo, IME-USP, 1985, 18p.

8509 - ACHCAR, J.A. & BOLFARINE, H., Use of Accurate Approximations for Posterior Densities in Regression Models with Censored data. São Paulo, IME-USP, 1985, 21p.