M-METHODS IN MULTIVARIATE LINEAR MODELS

by

Julio da Motta Singer
Universidade de Sao Paulo, Brazil

and

Pranab Kumar Sen
Department of Biostatistics
University of North Carolina, Chapel Hill, NC U.S.A.

Institute of Statistics Mimeo Series No. 1424

January 1983

M-METHODS IN MULTIVARIATE LINEAR MODELS

Julio da Motta Singer*
Universidade de Sao Paulo, Brazil

and

Pranab Kumar Sen†
University of North Carolina, Chapel Hill

Two types of M-estimators of the parameters of the multivariate linear model are considered. The first type corresponds to coordinatewise M-estimators; the second is an extension of Maronna's (1976) proposal for the multivariate location case. Under some general conditions on the error distribution, on the design matrix and on the score functions, the asymptotic distribution of the coordinatewise M-estimator is derived by using an asymptotic linearity result due to Jurečková (1977). It is also indicated how the same approach can be employed to obtain the

---

AMS Subject classifications: 62E20, 62H12, 62H15.

Key words and phrases: Asymptotic distributions, asymptotic linearity, robust procedures, influence functions.

*This work was done while this author 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.

†This work was partially supported by the National Heart, Lung, and Blood Institute, NIH, under Contract NIH-NHLBI-71-2243-L.

asymptotic distribution of the Maronna-type estimator under the additional assumption of elliptical symmetry for the error distribution. Asymptotic tests of hypotheses about the parameters are proposed. Finally, robustness properties of both types of estimators are discussed.

1. INTRODUCTION

Consider the multivariate linear model:

$$\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{1.1}$$

where $\underset{\sim}{Y}$ (nxp) is a matrix of observable random variables (r.v.'s), $\underset{\sim}{X}$ (nxr) is a matrix of known constants, assumed of full rank r<n, $\underset{\sim}{\beta}$ (rxp) is a matrix of unknown parameters, and $\underset{\sim}{\varepsilon}$ (nxp) is a matrix of random errors, the rows of which are assumed independently distributed with common distribution F, mean vector $\underset{\sim}{0}$, and positive definite (p.d.) covariance matrix $\underset{\sim}{\Sigma}$.

Estimation and hypothesis testing under this model have been extensively considered in the literature when F is assumed multivariate Normal. In the univariate case (p=1), a variety of methods have been proposed, which allow for more general assumptions on the error distribution; among these are the M, L, and R approaches, which are discussed in Huber (1981, ch. 7). In particular, much attention has been devoted to the class of M-methods, which includes maximum likelihood (ML) and least squares (LS) methods as special cases and also allow for the construction of estimators and tests that are robust against departures from normality. Very little attention, however, has been directed to the dual multivariate problem, though we expect the normality assumption to be more easily violated by multivariate data, not only because of the larger number of coordinates where departures can occur, but also because of the dependence among them.

Maronna (1976) proposed a simultaneous M-estimation procedure for the multivariate location and dispersion parameters; under the assumption of elliptical symmetry of the underlying distribution, and some conditions on the score function, he established consistency and asymptotic normality of the estimators. Carroll (1978) examined some further asymptotic properties of these estimators. Pendergast and Broffitt (1981) extended Maronna's procedure to the growth curve model and showed consistency and asymptotic normality in the special case when the n observations can be divided into k distinct groups, each assumed to have its own growth curve and common dispersion.

In this paper we consider the multivariate linear model (1.1), treating $\underset{\sim}{\beta}$ as the parameter of interest and $\underset{\sim}{\Sigma}$ as a nuisance parameter. We are concerned with both the estimation and hypothesis testing problems.

In section 2 we define both types of estimators and state the assumptions required for the subsequent sections.

In section 3 we derive the asymptotic distribution of the coordinatewise M-estimator by applying an asymptotic linearity result similar to that of Jurečková (1977). The proof of the basic result is presented in the appendix, and though it follows the general guidelines of Jurečková's paper, it requires neither her concordance-discordance assumption on the elements of the design matrix nor the dispersion parameter to be known. Klein and Yohai (1981) also consider a similar result; however, the proof presented here is simpler and requires less stringent assumptions. We also consider two equi-efficient asymptotic test procedures.

In section 4 we indicate how a similar argument to that of the previous section can be employed to derive the asymptotic distribution of the Maronna-type M-estimator under the additional assumption of elliptical symmetry for the underlying error distribution.

Finally, in section 5 we comment on the construction of robust M-estimators of both types by examining their influence functions and breakdown points.

The following notation will be used throughout:

i) $\underset{\sim}{y}_k(p\times 1)$, $\underset{\sim}{\varepsilon}_k(p\times 1)$, and $\underset{\sim}{x}_k(r\times 1)$, $(k=1,\dots,n)$ will denote the transposes of the k^{th} row of $\underset{\sim}{Y}$, $\underset{\sim}{\varepsilon}$, and $\underset{\sim}{X}$, respectively;

ii) $\underset{\sim}{\beta}_j(r\times 1)$, $(j=1,\dots,p)$ will denote the j^{th} column of $\underset{\sim}{\beta}$;

iii) $\underset{\sim}{A}^*$(bax1) denotes the "rolled out by rows" version of any (axb) matrix $\underset{\sim}{A}$.

2. DEFINITIONS AND ASSUMPTIONS

Consider the multivariate linear model (1.1) and assume that $\underset{\sim}{\Sigma}$ is known. Following the lines of Maronna (1976), we can define an M-estimator of $\underset{\sim}{\beta}$ as a solution $\hat{\underset{\sim}{\beta}}$ to:

$$M_{nij}(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}}) = \sum_{k=1}^{n} u(d_k)x_{ki}\,(y_{kj} - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}_j) = 0, \qquad (2.1)$$

$(i=1,\ldots,r;\ j=1,\ldots,p)$.

where:

$$d_k^2 = (\underset{\sim}{y}_k - \hat{\underset{\sim}{\beta}}^T\underset{\sim}{x}_k)^T\underset{\sim}{\Sigma}^{-1}\,(\underset{\sim}{y}_k - \hat{\underset{\sim}{\beta}}^T\underset{\sim}{x}_k) \qquad (2.2)$$

and $u(d)$ is a score function defined for $d \geq 0$. In matrix notation (2.1) can be written as $\underset{\sim}{M}_n(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}}) = \underset{\sim}{0}$, where $\underset{\sim}{M}_n(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}}) = \{M_{nij}(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}})\}$. If $\underset{\sim}{\Sigma}$ is unknown, the M-estimator is defined by replacing it by an estimate $\hat{\underset{\sim}{\Sigma}}$ in (2.1).

The class of M-estimators defined by (2.1) can be viewed as a generalization of the class of ML estimators under the assumption of elliptical symmetry for the error distribution. In fact, if the corresponding density is of the form:

$$f(\underset{\sim}{\varepsilon}_k) = |\underset{\sim}{\Sigma}|^{-\frac{1}{2}}\, h\left\{(\underset{\sim}{\varepsilon}_k^T\,\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\varepsilon}_k)^{\frac{1}{2}}\right\}$$

where h is a density in R^p, a ML estimator of $\underset{\sim}{\beta}$ can be obtained from (2.1) by taking $u(d) = -d^{-1}\partial \log h(d)/\partial d$. In the multivariate Normal case, $u(d)=1$. As we shall see below, robust Maronna-type M-estimators can be obtained by a convenient choice of the score function u.

Here we note that for elliptically symmetric distributions, $\underset{\sim}{\Sigma}$ does not necessarily correspond to the covariance matrix, but to a scalar multiple of it. Nevertheless, we shall still refer to it as the covariance matrix.

An alternative approach within the context of M-estimation corresponds to the following generalization of the LS method. First let:

$$\underset{\sim}{\gamma}_k = \left\{\mathrm{diag}(\sigma_{11}^{-\frac{1}{2}},\ldots,\sigma_{pp}^{-\frac{1}{2}})\right\}(\underset{\sim}{y}_k - \underset{\sim}{\beta}^T \underset{\sim}{x}_k), \quad (k=1,\ldots,n),$$

$\underset{\sim}{h}: R^p \to R^p$ be a vector valued function, and

$$\underset{\sim}{\Gamma} = E\left\{\underset{\sim}{h}(\underset{\sim}{\gamma}_k)\underset{\sim}{h}^T(\underset{\sim}{\gamma}_k)\right\}, \text{ assumed p.d.}$$

Then define an M-estimator of $\underset{\sim}{\beta}$ as the value $\hat{\underset{\sim}{\beta}}$ minimizing:

$$Q(\underset{\sim}{\beta}) = \sum_{k=1}^{n} \underset{\sim}{h}^T(\underset{\sim}{\gamma}_k)\underset{\sim}{\Gamma}^{-1}\underset{\sim}{h}(\underset{\sim}{\gamma}_k).$$

or equivalently as a solution to:

$$\frac{\partial}{\partial \underset{\sim}{\beta}} Q(\underset{\sim}{\beta}) = \underset{\sim}{0}. \qquad (2.3)$$

If for each coordinate h_j $(j=1,\ldots,p)$ of $\underset{\sim}{h}$ we define

$\rho_j(x) = h_j^2(x)$ and $\psi_j(x) = \partial\rho_j(x)/\partial x$, it can be easily shown that a solution to (2.3) is equivalent to a solution to:

$$M_{nij}(\underset{\sim}{Y},\underset{\sim}{\sigma},\hat{\underset{\sim}{\beta}}) = \sum_{k=1}^{n} x_{ki}\ \psi\left\{\sigma_{jj}^{-\frac{1}{2}}(y_{kj} - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}_j)\right\} = 0, \qquad (2.4)$$

$(i=1,\ldots,r;\ j=1,\ldots,p)$.

where $\underset{\sim}{\sigma} = (\sigma_{11},\ldots,\sigma_{pp})^T$. A matrix notation similar to that of the previous case can also be considered here. If $\underset{\sim}{\sigma}$ is unknown, we can define the estimator by replacing it by an estimate $\hat{\underset{\sim}{\sigma}}$ in (2.4).

Note that solving (2.4) is equivalent to obtaining M-estimators of each column of $\underset{\sim}{\beta}$ individually. Thus we shall refer to the corresponding estimator as the coordinatewise M-estimator.

If we take $h_j(x)=x$, $(j=1,\ldots,p)$ we obtain the ordinary multivariate LS estimator. Alternatively a robust estimator could be obtained by defining:

$$h_j(x) = \begin{cases} 2^{-\frac{1}{2}}x & \text{if } |x| \le k \\ [k(|x| - k/2)]^{\frac{1}{2}}\ \mathrm{sign}(x) & \text{if } |x| > k \end{cases}$$

where k is a positive constant; this corresponds to Huber's proposal for $\psi_j(=\psi)$. In this case the generalized LS method will de-emphasize outliers coordinatewise, as opposed to the corresponding Maronna-type method with $u(d) = \psi(d)/d$, which de-emphasizes outliers by considering all coordinates simultaneously.

One advantage of the coordinatewise method over the one suggested by Maronna is that it offers the possibility of choosing different score functions for different coordinates; this could be useful in the case where some coordinates are known to produce more outliers than the others. In this paper, for reasons of notational ease, we shall consider the same score function ψ for each coordinate, with no loss of generality. Also, in order to compute the estimate of $\underset{\sim}{\beta}$ we do not need to estimate the entire covariance matrix $\underset{\sim}{\Sigma}$ as in Maronna's case, but only its diagonal $\underset{\sim}{\sigma}$.

Furthermore, if we select score functions to de-emphasize outliers as in Huber's suggestion, outliers in one coordinate should have less influence in the estimates of parameters related to other coordinates than with Maronna's proposal. Since each column of the coordinatewise M-estimator $\hat{\underset{\sim}{\beta}}$ is computed independently, outliers in one coordinate will only influence elements of $\hat{\underset{\sim}{\beta}}$ corresponding to other coordinates through the covariance matrix.

Finally, we point that the assumptions required to derive the asymptotic distributions are less stringent for the coordinatewise method than for Maronna's. We summarize below the assumptions to be used in the subsequent sections.

A1. The distribution function F of the error r.v.'s is absolutely continuous with density function f such that $f_i'(\underset{\sim}{\varepsilon}) = \partial f(\underset{\sim}{\varepsilon})/\partial \varepsilon_i$ exists, i=1,...,p.

A2. F has a finite and p.d. Fisher information matrix with respect to location, $\underset{\sim}{I}(f) = \{I_{ij}(f)\}$, where:

$$I_{ij}(f) = \int [f_i'(\underset{\sim}{\varepsilon}) f_j'(\underset{\sim}{\varepsilon}) / \{f(\underset{\sim}{\varepsilon})\}^2] \, f(\underset{\sim}{\varepsilon}) d\underset{\sim}{\varepsilon}, \quad (i,j=1,\ldots,p)$$

A3. F has a finite and p.d. Fisher information matrix with respect to scale, $\underset{\sim}{I}_1(f) = \{I_{1ij}(f)\}$, where:

$$I_{1ij}(f) = -1 + \int [\varepsilon_i \varepsilon_j f_i'(\underset{\sim}{\varepsilon}) f_j'(\underset{\sim}{\varepsilon}) / \{f(\underset{\sim}{\varepsilon})\}^2] \, f(\underset{\sim}{\varepsilon}) d\underset{\sim}{\varepsilon}, \quad (i,j=1,\ldots,p)$$

A4. F is elliptically symmetric, i.e., the density function f is given by:

$$f(\underset{\sim}{\varepsilon}) = |\underset{\sim}{\Sigma}|^{-\frac{1}{2}} \, h\{(\underset{\sim}{\varepsilon}^T \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\varepsilon})^{\frac{1}{2}}\}$$

where h is a density in R^P, such that $h'(x) = \partial h(x)/\partial x$ exists.

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

i) Noether's condition: $\lim_{n\to\infty} \left\{ \max_{1 \le k \le n} x_{ki}^2 \Big/ \sum_{k=1}^{n} x_{ki}^2 \right\} = 0$, $(i=1,\ldots,r)$.

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

C1. The score function ψ is a nonconstant function expressible in the form $\psi(x) = \sum_{j=1}^{s} \psi_j(x)$, where each ψ_j is monotone and is either an absolutely continuous function on any bounded interval in R, with derivative ψ_j' almost everywhere, or is a step function. It also satisfies:

i) $\int \psi(x) f(x) dx = 0.$

ii) $\rho^2 = \int \psi^2(x) f(x) dx < \infty.$

iii) $\omega = - \sigma^{-1} \int \psi(x) f'(x) dx = \sigma^{-1} \int \psi'(x) f(x) dx < \infty.$

(i) is satisfied, for example, when ψ is skew-symmetric and F is symmetric, which is the case usually considered in the literature.

C2. The score function u(d) is nonnegative, nonincreasing, and continuous for $d \geq 0$. To facilitate notation and comparison with the univariate case we define $u(d) = \psi(d)/d$, where ψ is a bounded function.

3. ASYMPTOTIC PROPERTIES OF THE COORDINATEWISE M-ESTIMATOR

In order to obtain the asymptotic distribution of the coordinatewise M-estimator defined by (2.4), we first consider the extension of two results due to Jurečková (1977).

The first result relates to the linearization of $\underset{\sim}{M}_n(\underset{\sim}{Y}, \hat{\underset{\sim}{\sigma}}, \hat{\underset{\sim}{\beta}})$ in a neighborhood of the true parameter $\underset{\sim}{\beta}$ and is given by the following theorem, the proof of which is presented in appendix A.

THEOREM 3.1. Under assumptions A1-A3, B, and C1 it follows that:

$$P\left\{\sup n^{-\frac{1}{2}} \| \underset{\sim}{M}_n^*(\underset{\sim}{Y},\hat{\underset{\sim}{\sigma}},\hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n^*(\underset{\sim}{Y},\underset{\sim}{\sigma},\underset{\sim}{\beta}) + n(\underset{\sim}{V}\otimes \underset{\sim}{W})(\hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*)\| \geq \varepsilon :\right.$$

$$\left. n^{\frac{1}{2}} \| \hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*\| \leq K,\ n^{\frac{1}{2}} \| \hat{\underset{\sim}{\sigma}}-\underset{\sim}{\sigma}\| \leq L\right\} \to 0 \text{ as } n\to\infty$$

holds for all $K>0$, $L>0$ and $\varepsilon>0$. (3.1)

where:

$$\underset{\sim}{W} = \text{diag}\{w_1,\ldots,w_p\},\ w_j = \sigma_{jj}^{-\frac{1}{2}} \int \psi'(\gamma_{kj}) f(\underset{\sim}{\gamma}_k) d\underset{\sim}{\gamma}_k, \quad (3.2)$$

$(j=1,\ldots,p)$.

The second result relates to the boundedness in probability of $n^{\frac{1}{2}} \| \hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*\|$, and corresponds to a slight modification of Lemma 5.2 of Jurečková (1977). In the special case where $p=1$ (univariate linear model) and the variance is known, she showed that given $\eta>0$, $\min_{\hat{\beta}} n^{-\frac{1}{2}} \| \underset{\sim}{M}_n^*(\underset{\sim}{Y},\underset{\sim}{\sigma},\hat{\underset{\sim}{\beta}})\| \geq \eta$ with probability approaching 1 outside the compact set $n^{\frac{1}{2}} \| \hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*\| \leq K$, which is equivalent to showing that $n^{\frac{1}{2}} \| \hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*\|$ is bounded in probability. Recalling that her $\underset{\sim}{X}_N$, c_{ji}, $\underset{\sim}{\Delta}^o$ and $\underset{\sim}{\Delta}$ correspond to $\underset{\sim}{Y}$, $n^{-\frac{1}{2}} x_{ij}$, $n^{\frac{1}{2}}\underset{\sim}{\beta}$, and $n^{\frac{1}{2}}\hat{\underset{\sim}{\beta}}$ in our notation, her proof also goes through in the more general case considered here by defining:

$$X_i^* = (X_i - \underset{\sim}{\Delta}^o \underset{\sim}{c}^{(i)})/\hat{\sigma} \text{ and } M(\zeta) = \sum_{i=1}^{n} c_i^* \psi(X_i^* + \zeta c_i^*/\hat{\sigma})$$

where $\hat{\sigma}$ is a $n^{\frac{1}{2}}$ consistent estimate of σ, and applying the result to each coordinate separately.

We may now consider the following theorem:

THEOREM 3.2. Under assumptions A1-A3, B, and C1 it follows that:

$$n^{\frac{1}{2}}(\hat{\beta}^* - \beta^*) \approx N_{rp}(0, V^{-1} \otimes W^{-1} \Phi W^{-1})$$

where:

$$\Phi = \{\phi_{ij}\}, \; \phi_{ij} = \int \psi(\gamma_{ki})\psi(\gamma_{kj}) f(\gamma_k) d\gamma_k, \; (i,j=1,\ldots,p) \quad (3.3)$$

Proof: From Theorem 3.1 and the fact that $n^{\frac{1}{2}} \| \hat{\beta}^* - \beta^* \|$ is bounded in probability, it follows that $n^{\frac{1}{2}}(V \otimes W)(\hat{\beta}^* - \beta^*)$ has the same asymptotic distribution as $n^{-\frac{1}{2}} M_n(Y, \sigma, \beta)$. Now, by applying Cramér-Wold's Theorem and Theorem V.1.2 of Hájek-Šidák (1967) it can be shown that $n^{-\frac{1}{2}} M_n(Y, \sigma, \beta) \approx N(0, V \otimes \Phi)$ and the result follows.

The asymptotic distribution derived above is useful to construct tests of hypotheses of the form:

$$H: C\beta U = K \quad (3.4)$$

where $C(c \times r)$ and $U(p \times u)$ are known matrices of full row and column ranks $c(\leq r)$ and $u(\leq p)$ respectively, and $K(c \times u)$ is any known matrix. By making a linear transformation and a convenient reparametrization, testing (3.4) under the model (1.1) is equivalent to testing

$$H_0: \eta = 0 \quad (3.5)$$

under the following model:

$$Z = [A \; Q] \begin{bmatrix} \xi \\ \eta \end{bmatrix} + \zeta = D\alpha + \zeta \quad (3.6)$$

where $\underset{\sim}{Z} = \underset{\sim}{Y}\underset{\sim}{U}$, $\underset{\sim}{\zeta} = \underset{\sim}{\varepsilon}\underset{\sim}{U}$, $\underset{\sim}{A}$(nxr-c) and $\underset{\sim}{Q}$(nxc) are known matrices, $\underset{\sim}{\xi}$(r-cxu) and $\underset{\sim}{\eta}$(cxu) are the unknown parameters, $\underset{\sim}{D} = [\underset{\sim}{A}\ \underset{\sim}{Q}]$ and $\underset{\sim}{\alpha} = [\underset{\sim}{\xi}^T\ \underset{\sim}{\eta}^T]^T$.

Let H and h respectively denote the distribution function and the density function of the transformed (u-dimensional) r.v.'s and let $\underset{\sim}{W}_H$ and $\underset{\sim}{\Sigma}_H$ respectively correspond to (3.2) and (3.3) defined in terms of H. Note that the linear transformation and the reparametrization preserve assumptions A1-A3, B, and C1.

We shall outline below two different approaches which produce asymptotically equi-efficient tests for local Pitman-type alternatives of the form:

$$H_n:\ \underset{\sim}{\eta}_n = \underset{\sim}{\eta} = n^{-\frac{1}{2}}\underset{\sim}{\Delta} \qquad (3.7)$$

where $\underset{\sim}{\Delta}(\neq \underset{\sim}{0})$ is a fixed (cxu) matrix.

From assumptions A1-A3, B, and C1 it follows as in Hájek-Šidák (1967, ch. 6) that the sequence of probability measures under H_n is contiguous to that under H_o. This contiguity implies that the result of theorem 3.1 also holds under H_n. Then, letting:

$$\underset{\sim}{R} = \lim_{n\to\infty} n^{-1}\ \underset{\sim}{D}^T\underset{\sim}{D} = \lim_{n\to\infty} n^{-1} \begin{bmatrix} \underset{\sim}{A}^T\underset{\sim}{A} & \underset{\sim}{A}^T\underset{\sim}{Q} \\ \underset{\sim}{Q}^T\underset{\sim}{A} & \underset{\sim}{Q}^T\underset{\sim}{Q} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{R}_{11} & \underset{\sim}{R}_{12} \\ \underset{\sim}{R}_{21} & \underset{\sim}{R}_{22} \end{bmatrix}$$

we can proceed as in theorem 3.2 to show that:

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\eta}}^* - \underset{\sim}{\eta}^*) \approx N_{cu}(\underset{\sim}{\Lambda}_1^*, \underset{\sim}{P}^{-1} \otimes \underset{\sim}{W}_H^{-1}\underset{\sim}{\Sigma}_H\underset{\sim}{W}_H^{-1}), \qquad \text{where}$$

$$\underset{\sim}{P} = \underset{\sim}{R}_{22} - \underset{\sim}{R}_{21}\, \underset{\sim}{R}_{11}^{-1}\, \underset{\sim}{R}_{12}$$

and $\underset{\sim}{\Lambda}_1^* = \underset{\sim}{0}^*$ if H_o holds or $\underset{\sim}{\Lambda}_1^* = \underset{\sim}{\Delta}^*$ if H_n holds.

From well known results related to the Wishart distribution we get:

$$\underset{\sim}{H} = \hat{\underset{\sim}{\eta}}^T[\underset{\sim}{Q}^T\underset{\sim}{Q}-\underset{\sim}{Q}^T\underset{\sim}{A}(\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\underset{\sim}{A}^T\underset{\sim}{Q}]\hat{\underset{\sim}{\eta}} \approx W_u(c,\underset{\sim}{W}_H^{-1}\underset{\sim}{\Sigma}_H\underset{\sim}{W}_H^{-1},\underset{\sim}{\Omega}_1) \qquad (3.8)$$

where:

$$\underset{\sim}{\Omega}_1 = \underset{\sim}{\Lambda}_1^T\, \underset{\sim}{P}\, \underset{\sim}{\Lambda}_1$$

Similarly to the Normal theory case, we propose to base the asymptotic tests on functions of the characteristic roots of $\underset{\sim}{H}\underset{\sim}{E}^{-1}$, where:

$$\underset{\sim}{E} = \hat{\underset{\sim}{W}}_H^{-1}\, \hat{\underset{\sim}{\Sigma}}_H\, \hat{\underset{\sim}{W}}_H^{-1} \qquad (3.9)$$

and $\hat{\underset{\sim}{W}}_H$ and $\hat{\underset{\sim}{\Sigma}}$ are $n^{\frac{1}{2}}$-consistent estimators of $\underset{\sim}{W}_H$ and $\underset{\sim}{\Sigma}_H$ respectively.

From (3.8) and (3.9) it follows by a standard argument that both tr $\underset{\sim}{H}\underset{\sim}{E}^{-1}$ (Lawley-Hotelling's trace) and $n \log\{|n\underset{\sim}{E}+\underset{\sim}{H}|/|n\underset{\sim}{E}|\}$ (Wilks' likelihood ratio) are asymptotically distributed as χ^2_{cu} (tr $\underset{\sim}{\Omega}_E$), where

$$\underset{\sim}{\Omega}_E = \underset{\sim}{\Sigma}_H^{-\frac{1}{2}}\, \underset{\sim}{W}_H\underset{\sim}{\Lambda}_1^T\, \underset{\sim}{P}\underset{\sim}{\Lambda}_1\, \underset{\sim}{W}_H\underset{\sim}{\Sigma}_H^{-\frac{1}{2}}$$

Furthermore, the asymptotic null distribution of $ch_1(\underset{\sim}{H}\underset{\sim}{E}^{-1})$ (Roy's largest root) can be shown to be that of the largest characteristic root of a $W_u(c,\underset{\sim}{I},\cdot)$ distribution.

Following the suggestions of Pendergast and Broffitt (1981) a reasonable small sample approximation to the asymptotic distributions of the test statistics would correspond to adopting the Normal theory degrees of freedom for $\underset{\sim}{E}$; this, however, needs further investigation.

The second approach for hypothesis testing consists of an extension of the method proposed by Sen (1982) for univariate linear models. This approach leads to test statistics which do not require the estimation of $\underset{\sim}{W}_H$ and are equi-efficient to the ones derived through the previous approach.

First let $\underset{\sim}{M}_{n1}(\hat{\underset{\sim}{\xi}},\hat{\underset{\sim}{\eta}})$ and $\underset{\sim}{M}_{n2}(\hat{\underset{\sim}{\xi}},\hat{\underset{\sim}{\eta}})$ respectively correspond to the first (r-c) and the last c rows of $\underset{\sim}{M}_n(\underset{\sim}{Z},\hat{\underset{\sim}{\sigma}},\hat{\underset{\sim}{\beta}})$. Then let $\tilde{\underset{\sim}{\xi}}$ be a coordinatewise M-estimator of $\underset{\sim}{\xi}$ under the null hypothesis, that is a solution to $\underset{\sim}{M}_{n1}(\tilde{\underset{\sim}{\xi}},\underset{\sim}{0}) = \underset{\sim}{0}$, and define $\tilde{\underset{\sim}{M}}_{n2} = \underset{\sim}{M}_{n2}(\tilde{\underset{\sim}{\xi}},\underset{\sim}{0})$. Now observe that:

$$
\begin{aligned}
n^{-\frac{1}{2}}\, \tilde{\underset{\sim}{M}}{}^{*}_{n2} = n^{-\frac{1}{2}} \Big\{ & \underset{\sim}{M}{}^{*}_{n2}[\underset{\sim}{\xi}+(\tilde{\underset{\sim}{\xi}}-\underset{\sim}{\xi}),\underset{\sim}{0}] - \underset{\sim}{M}{}^{*}_{n2}(\underset{\sim}{\xi},\underset{\sim}{0}) \\
& + n(\underset{\sim}{R}_{21} \otimes \underset{\sim}{W}_H)(\tilde{\underset{\sim}{\xi}}{}^{*}-\underset{\sim}{\xi}^{*})\Big\} + n^{-\frac{1}{2}} \Big\{\underset{\sim}{M}{}^{*}_{n2}(\underset{\sim}{\xi},\underset{\sim}{0}) \\
& - n(\underset{\sim}{R}_{21} \otimes \underset{\sim}{W}_H)(\tilde{\underset{\sim}{\xi}}{}^{*}-\underset{\sim}{\xi}^{*})\Big\}. \qquad (3.10)
\end{aligned}
$$

From the contiguity mentioned above it follows by Theorem 3.1 that the first term in (3.10) converges in probability to zero under either H_o or H_n; furthermore, the same theorem implies that $n^{\frac{1}{2}}(\tilde{\underset{\sim}{\xi}}{}^{*}-\underset{\sim}{\xi}^{*})$ has the same asymptotic distribution as $(\underset{\sim}{R}_{11}^{-1} \otimes \underset{\sim}{W}_H^{-1})$ x $n^{-\frac{1}{2}}\, \underset{\sim}{M}{}^{*}_{n1}(\underset{\sim}{\xi},\underset{\sim}{0})$. Therefore we obtain:

$$n^{-\frac{1}{2}} \tilde{M}^*_{n2} = n^{-\frac{1}{2}} \left\{ M^*_{n2}(\xi,0) - (R_{21}R_{11}^{-1} \otimes I) M^*_{n1}(\xi,0) \right\} + o_p(1).$$

Proceeding as in Theorem 3.2 we may show that:

$$n^{-\frac{1}{2}} \tilde{M}^*_{n2} \approx N_{cu}(\Lambda^*_2, P \otimes \Sigma_H),$$

where $\Lambda^*_2 = 0$ if H_o holds or $\Lambda^*_2 = (P \otimes W_H)\Delta^*$ if H_n holds. Consequently, we obtain:

$$H = \tilde{M}^T_{n2} \, [Q^TQ - Q^TA(A^TA)^{-1}A^TQ]^{-1}\tilde{M}_{n2} \approx W_u(c, \Sigma_H, \Omega_2),$$

where:

$$\Omega_2 = \Lambda^T_2 \, P^{-1} \, \Lambda_2 .$$

Taking $E = \hat{\Sigma}_H$ we conclude that the asymptotic distributions of the usual test statistics are the same as those derived previously.

4. ASYMPTOTIC PROPERTIES OF THE MARONNA-TYPE M-ESTIMATOR

Basically, the same approach of section 3 can be employed to obtain the asymptotic distribution of the Maronna-type M-estimator defined by (2.1).

First we present the analog of the basic asymptotic linearity theorem for this case; the proof is given in appendix B.

THEOREM 4.1. Under assumptions A1-A4, B, and C2 it follows that:

$$P\left\{ \sup n^{-\frac{1}{2}} \| M^*_n(Y,\hat{\Sigma},\hat{\beta}) - M^*_n(Y,\Sigma,\beta) + n(V \otimes Ib_o)(\hat{\beta}^* - \beta^*) \| \geq \varepsilon : n^{\frac{1}{2}} \| \hat{\beta}^* - \beta^* \| \leq K, \; n^{\frac{1}{2}} \| \hat{\Sigma}^{\frac{1}{2}} - \Sigma^{\frac{1}{2}} \| \leq L \right\} \to 0 \text{ as } n \to \infty$$

holds for all $K>0$, $L>0$, $\varepsilon>0$. (4.1) where:

$$b_o = E_F \left\{p^{-1}\psi'(d_k) + (1-p^{-1})^{-1}\psi(d_k)/d_k\right\}.$$

Next we point that boundedness in probability of $n^{\frac{1}{2}}\|\hat{\beta}^*-\beta^*\|$ follows from a straightforward extension of Lemma 5.2 of Jurečková (1977). Besides the differences in notation discussed in section 3, the main change relates to the definition of her M function, which in our case should be given by:

$$M(\zeta) = n^{-\frac{1}{2}} \sum_{k=1}^{n} u \left[\left\{(\varepsilon_k+\zeta w_k)^T \hat{\Sigma}^{-1}(\varepsilon_k+\zeta w_k)\right\}^{\frac{1}{2}}\right] w_k^T(\varepsilon_k+\zeta w_k),$$

where:

$$w_k = (\beta-\beta^{(1)})^T x_k,$$

and $\beta^{(1)}$ is any point such that $n^{\frac{1}{2}} \|\beta^{(1)*}- \beta^*\| = K$.

Then the asymptotic distribution of the Maronna-type M-estimator is given by the following theorem:

THEOREM 4.2. Under assumptions A1-A4, B, and C2 it follows that:

$$(\hat{\beta}^*-\beta^*) \approx N_{rp}(0,V^{-1} \otimes a_o b_o^{-2} \Sigma),$$

where:

$$a_o = E_F \left\{p^{-1} \psi^2(d_k)\right\}.$$

Proof: Similar to that of Theorem 3.2, with the use of Theorem 4.1 in lieu of Theorem 3.1.

Asymptotic tests as those of section 3 can be obtained similarly. The H matrices are computed in the same way and the E

matrices are given by: $\underset{\sim}{E} = \hat{a}_o \hat{b}_o^{-2} \underset{\sim}{U}^T \hat{\underset{\approx}{\Sigma}} \underset{\sim}{U}$ under the first approach and by $\underset{\sim}{E} = \hat{a}_o \underset{\sim}{U}^T \hat{\underset{\approx}{\Sigma}} \underset{\sim}{U}$ under Sen's approach, where $\hat{a}_o$, $\hat{b}_o$ and $\hat{\underset{\approx}{\Sigma}}$ are $n^{\frac{1}{2}}$-consistent estimators of a_o, b_o and $\underset{\approx}{\Sigma}$ respectively.

5. ROBUSTNESS PROPERTIES OF THE PROPOSED ESTIMATORS

We shall briefly discuss two heuristic indicators of robustness: the influence function and the breakdown point.

First assume that $\underset{\sim}{x}_k$ are independent identically distributed r.v.'s which are independent of $\underset{\sim}{\varepsilon}_k$, $(k=1,\ldots,n)$ and let G be their common distribution function. Consider the gross error model:

$$H_t = (1-t)H + t\delta(\underset{\sim}{\varepsilon}_o, \underset{\sim}{x}_o),$$

where $H=F \cdot G$ and $\delta(\underset{\sim}{\varepsilon}_o, \underset{\sim}{x}_o)$ is a point mass. Define $\underset{\sim}{\beta}(H_t)$ as a solution to:

$$\int \underset{\sim}{x}\, \psi\left[\sigma_{ii}^{-\frac{1}{2}} \left\{y_i - \underset{\sim}{x}^T \underset{\sim}{\beta}_i(H_t)\right\}\right] dH_t = \underset{\sim}{0}, \quad (i=1,\ldots,p).$$

Letting G correspond to a distribution which associates probability n^{-1} to each point $\underset{\sim}{x}_k$, $(k=1,\ldots,n)$ and letting $n \to \infty$, a standard argument yields the asymptotic influence function of the coordinatewise M-estimator:

$$\underset{\sim}{i}(\underset{\sim}{\varepsilon}_o, \underset{\sim}{x}_o; \underset{\sim}{\beta}, F) = \underset{\sim}{V}^{-1} \underset{\sim}{x}_o \otimes \underset{\sim}{W}^{-1} \underset{\sim}{\psi}(\underset{\sim}{\gamma}_o)$$

where $\underset{\sim}{\psi}(\underset{\sim}{\gamma}_o) = \left\{\psi(\gamma_{01}),\ldots,\psi(\gamma_{0p})\right\}^T$ and $\underset{\sim}{\gamma}_o = \left\{\text{diag}(\sigma_{11}^{-\frac{1}{2}},\ldots,\sigma_{pp}^{-\frac{1}{2}})\right\} \underset{\sim}{\varepsilon}_o$.

Similarly one can obtain the asymptotic influence function of the Maronna-type M-estimator:

$$\underset{\sim}{i}(\underset{\sim}{\varepsilon}_0, \underset{\sim}{x}_0; \underset{\sim}{\beta}, F) = \underset{\sim}{V}^{-1} \underset{\sim}{x}_0 \otimes (b_0 d_0)^{-1} \psi(d_0) \underset{\sim}{\varepsilon}_0 .$$

In both cases the estimators will only be locally robust if both the ψ function is bounded and $\underset{\sim}{x}_0$ is only allowed to take values on a bounded set. This is generally the case when $\underset{\sim}{X}$ corresponds to a designed experiment and we take ψ as Huber's function.

Note that similarly to the univariate location case as in Huber (1981, ch. 1), the asymptotic variances given by theorems 3.2 and 4.2 are reproduced by taking $\int \underset{\sim}{i}\,\underset{\sim}{i}^T dH$.

Following Maronna, Bustos and Yohai (1979), define the gross error breakdown point by:

$$t^* = \sup\{t : \exists\ K, \text{ constant such that } \|\underset{\sim}{\beta}(H_t)\| \leq K,\ \forall\ \delta(\underset{\sim}{\varepsilon}_0, \underset{\sim}{x}_0)\}.$$

In Theorem 5.1, the proof of which is presented in appendix C, we show that for either case, global robustness as measured by a positive breakdown point can only be achieved for very simple designs (equivalent to cell means models) and only if we allow contamination at the points defining the design space. Furthermore, it deteriorates very rapidly as the number of parameters increase.

THEOREM 5.1. Suppose that:

i) the independent variables can only assume s (finite) values;

ii) $\lim_{n\to\infty} n_\ell/n = p_\ell$, ($\ell$=1,...s), where n_ℓ is the number of observations for which the value of the independent variable is $\underset{\sim}{x}_{\ell}$;

iii) the function ψ is bounded.

Then, if s=r (the design is equivalent to a cell means model), and contamination is only allowed at the points $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_r$, we have $t^* = \min_{\ell=1,\ldots,r} (1+p_{\ell}^{-1})^{-1}$. Otherwise, $t^*=0$.

Maronna, Bustos and Yohai (1979) considered a class of estimators with positive gross error breakdown points in more general situations, but these required a weighting procedure on both the dependent and independent variables. We are not concerned about this class of estimators in this paper.

The proof of Theorem 5.1 assumes known variance. In the case where it is not known, the breakdown point of the estimator $\hat{\underset{\sim}{\beta}}$ depends on the breakdown point of the scale estimator, and it is still not clear what preliminary scale estimator to use if we want to maintain global robustness.

APPENDIX A

Observing that the multivariate case (p>1) will follow by applying the univariate result (p=1) to each coordinate separately, we can restrict ourselves without loss of generality to the latter case. We will consider the usual univariate notation $\underset{\sim}{y}(n\times 1)$, $\underset{\sim}{\beta}(r\times 1)$, $\underset{\sim}{\varepsilon}(n\times 1)$, and σ^2 throughout and drop the subscript corresponding to the different variates for notational ease.

We first prove a lemma which is equivalent to Theorem 3.1 when σ^2 is known. The proof will be outlined in four steps, only the first of which we detail, since it is the only one significantly different from the corresponding step in Jurečková (1977, Theorem 4.1).

LEMMA A1. Under assumptions A1, A2, B, and C1 it follows that (3.1) holds with $\hat{\sigma}^2$ replaced by σ^2.

<u>Proof</u>: We provide the proof in several steps.

<u>Step 1</u>: Write $M_{ni}(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}})$ as a sum of a finite number of components which are monotone functions of each element of $\hat{\underset{\sim}{\beta}}$ when the others are held fixed.

In this direction, for fixed i and h, (i,h=1,...,r), let:

$$S_{ih}(0) = \{k: \mathrm{sign}(x_{ki}) = \mathrm{sign}(x_{kh})\}$$

$$S_{ih}(1) = \{k: \mathrm{sign}(x_{ki}) \neq \mathrm{sign}(x_{kh})\}$$

and observe that:

$$\sum_{k\varepsilon S_{ih}(0)} x_{ki}\ \psi\left\{\sigma^{-1}(y_k - \sum_{j\neq h} x_{kj}\hat{\beta}_j - x_{kh}\hat{\beta}_h)\right\}$$

is non-increasing in $\hat{\beta}_h$ for non-decreasing ψ and non-decreasing in $\hat{\beta}_h$ for non-increasing ψ; the relations are inverted if the summation is over $k\varepsilon S_{ih}(1)$.

For each i (i=1,...,r) the set of all observations can be partitioned into 2^r disjoint sets S_{ig}, (g=1,...,2^r) formed by intersections of the type:

$$S_{ig} = \bigcap_{h=1}^{p} S_{ih}(\delta_g),$$

where δ_g is either 0 or 1.

Then we may write:

$$M_{ni}(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}}) = \sum_{g=1}^{2^r} M_{nig}(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}}),$$

where:

$$M_{nig}(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}}) = \sum_{k\varepsilon S_{ig}} x_{ki}\ \psi\left\{\sigma^{-1}(y_k - \underset{\sim}{x}_k^T\ \hat{\underset{\sim}{\beta}})\right\}$$

is a monotone function of each element of $\hat{\underset{\sim}{\beta}}$ when the others are kept fixed.

Therefore, all we have to show is that:

$$P\left\{n^{-\frac{1}{2}}\left|M_{nig}(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}}) - M_{nig}(\underset{\sim}{y},\sigma^2,\underset{\sim}{\beta}) + n\omega\ \underset{\sim}{v}_{ig}^T(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})\right| \geq \varepsilon:\ n^{\frac{1}{2}}\ \|\ \hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}\ \| \leq K\right\} \to 0 \text{ as } n\to\infty$$

holds for all K>0, ε>0, (i=1,...,r; g=1,...,2^r), where $\underset{\sim}{v}_{ig} = \left\{v_{i1g},\ldots,v_{irg}\right\}^T$ and $v_{ihg} = \lim_{n\to\infty} n^{-1} \sum_{k\varepsilon S_{ig}} x_{ki}x_{kh}$, $(h=1,\ldots,r)$.

If for any of the sets S_{ig} $(i=1,\ldots,r;\ g=1,\ldots,2^r)$, the number of elements remains finite as $n\to\infty$, the result follows trivially; we only have to be concerned with those sets S_{ig} with cardinality $\to \infty$ as $n\to\infty$.

Step 2: Note that without loss of generality we may take $\sigma=1$. For $i=1,\ldots,r$ and $g=1,\ldots,2^r$, define:

$$T_{nig} = - \sum_{k\varepsilon S_{ig}} \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}}\, f'(\varepsilon_k)/f(\varepsilon_k)$$

$$\log L_{nig} = \sum_{k\varepsilon S_{ig}} \log\left\{f(y_k - \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}})/f(\varepsilon_k)\right\}$$

$$p_{nig} = \begin{cases} 1 & \text{if } S_{ig} = \emptyset, \\ \prod_{k\varepsilon S_{ig}} f(\varepsilon_k) & \text{otherwise.} \end{cases}$$

$$q_{nig} = \begin{cases} 1 & \text{if } S_{ig} = \emptyset, \\ \prod_{k\varepsilon S_{ig}} f(y_k - \underset{\sim}{x}_k^T\underset{\sim}{\beta}) & \text{otherwise.} \end{cases}$$

Then, using the contiguity results of Hájek-Sidák (1967, ch. 6) for location alternatives, obtain the asymptotic distribution of $n^{-\frac{1}{2}} M_{nig}(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}})$ under the sequence of probability measures defined by either p_{nig} or q_{nig} when $\hat{\underset{\sim}{\beta}}$ is in the compact set $\Theta = \{\hat{\underset{\sim}{\beta}}\varepsilon R^r:\ n^{\frac{1}{2}} \|\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}\| \le K\}$.

Step 3: Truncate the score function conveniently, and then use Lebesgue's Dominated Convergence Theorem, Chebyshev's inequality, and the asymptotic distributions of step 2 to show that the asymptotic linearity result holds for a fixed $\hat{\underset{\sim}{\beta}} \in \Theta$.

Step 4: Use the monotonicity property of $M_{nig}(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}})$ to replace the operation of taking the supremum over Θ by that of taking the maximum over a finite number of grid points, thus showing that the result holds uniformly in $\hat{\underset{\sim}{\beta}} \in \Theta$.

Proof of theorem 3.1. Consider the decomposition:

$$\underset{\sim}{M}_n(\underset{\sim}{y},\hat{\sigma}^2,\hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n(\underset{\sim}{y},\sigma^2,\underset{\sim}{\beta}) + n\omega \underset{\sim}{V}(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}) =$$

$$\left\{\underset{\sim}{M}_n(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n(\underset{\sim}{y},\sigma^2,\underset{\sim}{\beta}) + n\omega \underset{\sim}{V}(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})\right\} +$$

$$\left\{\underset{\sim}{M}_n(\underset{\sim}{y},\hat{\sigma}^2,\hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}})\right\}$$

By applying Lemma A1 to the first term, all we have to show is that:

$$P\left\{\sup n^{-\frac{1}{2}} \| \underset{\sim}{M}_n(\underset{\sim}{y},\hat{\sigma}^2,\hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}} \| \geq \varepsilon : \; n^{\frac{1}{2}} |\hat{\sigma}-\sigma| \leq L\right\} \to 0$$

as $n\to\infty$ holds for all $L>0$, $\varepsilon>0$.

This can be done in lines very similar to the proof of Lemma A1. We outline the major modifications.

Step 1: Partition each of the sets $S_{ih}(\delta)$, $(i,h=1,\ldots,r;$ $\delta=0,1)$ into two disjoint sets, one of which contains those k's such that $\text{sign}(x_{ki}) = \text{sign}(y_k - \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}})$ and the other to its complement. Then observe that:

$$\partial\left[x_{ki}\ \psi\left\{\hat{\sigma}^{-1}(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})\right\}\right]\Big/\partial\hat{\sigma} = -\ \hat{\sigma}^{-2}x_{ki}(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})\ \psi'\left\{\hat{\sigma}^{-1}(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})\right\}$$

has the same sign within each of these sets. Defining 2^{r+1} disjoint sets S_{ig} $(g=1,\ldots,2^{r+1})$ in a similar way as in step 1 of Lemma A1, we may write $M_{ni}(\underset{\sim}{y},\hat{\sigma}^2,\hat{\underset{\sim}{\beta}})$ as the sum of a finite number of functions which are monotone in each element of $(\hat{\underset{\sim}{\beta}}^T,\sigma)$ when the others are kept fixed.

<u>Step 2</u>: Following the ideas of Jurečková and Sen (1982), consider the reparametrization $(\underset{\sim}{\beta}^T,\sigma) \to (\underset{\sim}{\beta}^T,\zeta)$ where $\zeta = \log\sigma$. Without loss of generality let $\zeta = 0$ and let $d = n^{-\frac{1}{2}}\delta$ where δ is a positive constant. Define for $i=1,\ldots,r;\ g=1,\ldots,2^{r+1}$:

$$T_{nig} = -d \sum_{k\epsilon S_{ig}} \left\{1 + (y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})\ f'(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})\Big/f(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})\right\}$$

$$\log L_{nig} = \sum_{k\epsilon S_{ig}} \log e^{-d}\left[f\left\{(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})\big/e^d\right\}\Big/f(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})\right]$$

$$p_{nig} = \begin{cases} 1 & \text{if } S_{ig} = \emptyset, \\ \prod_{k\epsilon S_{ig}} f(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}) & \text{otherwise.} \end{cases}$$

$$q_{nig} = \begin{cases} 1 & \text{if } S_{ig} = \emptyset, \\ \prod_{k\epsilon S_{ig}} e^{-d} f\{(y_k - \underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}})/e^d\} & \text{otherwise.} \end{cases}$$

Proceed as in step 2 of Lemma A1 using the contiguity results of Hájek-Šidák (1967, ch. 6) for scale alternatives.

<u>Steps 3 and 4</u>: Essentially the same as in Lemma A1.

APPENDIX B

First we present three lemmas which are needed in the proof of Theorem 4.1. The first is an analog of Lemma A1; the other two correspond to the scale alternatives counterpart of Lemmas 4.1 and 4.2 of Patel (1973).

LEMMA B1. Under assumptions A1, A2, A4, B, and C2 it follows that (4.1) holds when $\hat{\Sigma}$ is replaced by Σ.

<u>Proof</u>: Essentially the same as that of Lemma A1 with the obvious replacement of the univariate r.v.'s by multivariate elliptically symmetric ones. We shall only detail step 1.

Let $\Sigma = I$ without loss of generality and define the sets S_{ig} $(i=1,\ldots,r;\ g=1,\ldots,2^r)$ as in Lemma A1. Then write:

$$M_{nij}(Y,\Sigma,\hat{\beta}) = \sum_{g=1}^{2^r} M_{nijg}(Y,\Sigma,\hat{\beta}), \quad (i=1,\ldots,r;\ j=1,\ldots,p) \quad (B.1)$$

where:

$$M_{nijg}(Y,\Sigma,\hat{\beta}) = \sum_{k\varepsilon S_{ig}} \psi(d_k) d_k^{-1}\, x_{ki}(y_{kj} - x_k^T \hat{\beta}_j),$$
$$(i=1,\ldots,r;\ j=1,\ldots,p). \quad (B.2)$$

Now let g be such that $\mathrm{sign}(x_{ki}) = \mathrm{sign}(x_{kh})$ for all $k\varepsilon S_{ig}$. Then observe that:

$$\partial d_k / \partial \hat{\beta}_{hj} = -d_k^{-1}\, x_{kh}(y_{kj} - x_k^T \hat{\beta}_j)$$

which implies:

$$\partial\left\{x_{ki}(y_{kj}-\underset{\sim}{x}_k^T\underset{\sim}{\hat\beta}_j)\right\}/\partial\underset{\sim}{\hat\beta}_{hj} = -d_k^{-1}\ x_{ki}x_{kh}\left[1 - \left\{d_k^{-1}(y_{kj}-\underset{\sim}{x}_k^T\underset{\sim}{\hat\beta}_j)\right\}^2\right] < 0$$

with probability 1. Therefore it follows that:

$$-d_k^{-1}\ x_{ki}(y_{kj}-\underset{\sim}{x}_k^T\underset{\sim}{\hat\beta}_j) \text{ is } \begin{cases} \geq 0 \text{ and decreasing for } \hat\beta_{hj} < y_{kj}/x_{kh} \\ < 0 \text{ and decreasing otherwise} \end{cases} \tag{B.3}$$

Recalling from assumption C2 that $\psi(d_k)$ is nonnegative and non-decreasing for $d_k \geq 0$, it follows that regarded as a function of $\hat\beta_{hj}$,

$$\psi(d_k) \text{ is } \begin{cases} \geq 0 \text{ and nonincreasing for } \hat\beta_{hj} < y_{kj}/x_{kh} \\ \geq 0 \text{ and nondecreasing otherwise} \end{cases} \tag{B.4}$$

From (B.3) and (B.4) using the rule for the derivative of a product of two functions, it follows that (B.2) is a nonincreasing function of $\hat\beta_{hj}$.

In a similar way we can show that (B.2) is a nondecreasing function of $\hat\beta_{hj}$ when g is such that $\text{sign}(x_{ki}) \neq \text{sign}(x_{kh})$ for all $k\varepsilon S_{ig}$.

Now without loss of generality let $\underset{\sim}{S}_t = \underset{\sim}{\hat\Sigma}_t^{\frac{1}{2}} = \underset{\sim}{\Sigma}^{\frac{1}{2}} + t\underset{\sim}{\Delta}$ where $\underset{\sim}{\Delta}$ is a matrix of constants and $t = n^{-\frac{1}{2}}$. Then define:

$$f_t(\underset{\sim}{\varepsilon}) = |\underset{\sim}{S}_t^2|^{-\frac{1}{2}}\ h(d_t), \quad \text{where } d_t^2 = \underset{\sim}{\varepsilon}^T S_t^{-2}\underset{\sim}{\varepsilon}$$

$$s_t(\underset{\sim}{\varepsilon}) = \{f_t(\underset{\sim}{\varepsilon})\}^{\frac{1}{2}}$$

and let $f_t'(\underset{\sim}{\varepsilon}) = \partial f_t(\underset{\sim}{\varepsilon})/\partial t$ and $s_t'(\underset{\sim}{\varepsilon}) = \partial s_t(\underset{\sim}{\varepsilon})/\partial t$.

LEMMA B2. Under assumptions A1-A4 we have:

$$C = \int \left[\frac{s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon})}{t} - s_o'(\underset{\sim}{\varepsilon}) \right]^2 d\underset{\sim}{\varepsilon} \to 0 \text{ as } t \to 0.$$

Proof: Without loss of generality let $\underset{\sim}{\Sigma} = \underset{\sim}{I}$. Then using matrix derivatives as in MacRae (1974) we get:

$$f_t'(\underset{\sim}{\varepsilon}) = -h(d_t)\ \mathrm{tr}\left\{ |\underset{\sim}{S}_t^2| \underset{\sim}{S}_t^{-2} \underset{\sim}{\Delta} \right\} - d_t^{-1} h'(d_t)\ \mathrm{tr}\left\{ \underset{\sim}{S}_t^{-2} \underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}^T \underset{\sim}{S}_t^{-2} \underset{\sim}{\Delta} \right\}$$

$$f'(\underset{\sim}{\varepsilon}) = -h(d_o)\ \mathrm{tr}\ \underset{\sim}{\Delta} - d_o^{-1}\ h'(d_o) \underset{\sim}{\varepsilon}^T \underset{\sim}{\Delta}\underset{\sim}{\varepsilon}.$$

which implies:

$$s_o'(\underset{\sim}{\varepsilon}) = -2^{-1}\left[h(d_o)\ \mathrm{tr}\ \underset{\sim}{\Delta}/\{h(d_o)\}^{\frac{1}{2}} - d_o^{-1} h'(d_o)\ \underset{\sim}{\varepsilon}^T \underset{\sim}{\Delta}\underset{\sim}{\varepsilon}/\{h(d_o)\}^{\frac{1}{2}} \right].$$

By assumption A4 it follows that:

$$\int [s_o'(\underset{\sim}{\varepsilon})]^2 d\underset{\sim}{\varepsilon} < \infty. \qquad \text{(B.5)}$$

By the Cauchy-Schwarz inequality:

$$\left[s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon}) \right]^2 = \left[\int_o^t s_w'(\underset{\sim}{\varepsilon})\, dw \right]^2$$

$$\leq \int_o^t \left[s_w'(\underset{\sim}{\varepsilon}) \right]^2 dw \int_o^t dw = t \int_o^t \left[s_w'(\underset{\sim}{\varepsilon}) \right]^2 dw$$

Then we may write:

$$\int \left[\frac{s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon})}{t} \right]^2 d\underset{\sim}{\varepsilon} \leq \int \frac{1}{t} \int_o^t [s_w'(\underset{\sim}{\varepsilon})]^2\, dw d\underset{\sim}{\varepsilon}$$

$$= \int [s_w'(\underset{\sim}{\varepsilon})]^2\, d\underset{\sim}{\varepsilon} \qquad \text{(B.6)}$$

Now, since $[\{s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon})\}/t]^2 \to [s_o'(\underset{\sim}{\varepsilon})]^2$ as $t \to 0$, it follows from (B.5) and Fatou's Lemma that:

$$\lim_{t\to 0} \int\left[\frac{s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon})}{t}\right]^2 d\underset{\sim}{\varepsilon} \geq \int\left[s_o'(\underset{\sim}{\varepsilon})\right]^2 d\underset{\sim}{\varepsilon} \qquad \text{(B.7)}$$

Using (B.6) it follows that equality must hold in (B.7) and according to criterion (4.5) in Hájek (1962), the functions $[\{s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon})\}/t]^2$ are uniformly integrable, which implies the result.

Next define:

$$W_n = 2 \sum_{k=1}^{n} \left[\left\{f_t(\underset{\sim}{\varepsilon}_k)/f(\underset{\sim}{\varepsilon}_k)\right\}^{\frac{1}{2}} - 1\right]$$

$$p_n = \prod_{k=1}^{n} f(\underset{\sim}{\varepsilon}_k)$$

$$q_n = \prod_{k=1}^{n} f_t(\underset{\sim}{\varepsilon}_k)$$

LEMMA B3. Under assumptions A1-A4 we have:

$$W_n \approx N(-b^2/4,\ b^2),$$

where:

$$b^2 = \int\left[\text{tr}\ \underset{\sim}{\Delta} + \frac{h'(d_o)}{h(d_o)} \frac{1}{d_o} \underset{\sim}{\varepsilon}^T \underset{\sim}{\Delta}\underset{\sim}{\varepsilon}\right]^2 h(d)\, d\underset{\sim}{\varepsilon}.$$

<u>Proof</u>: Under p_n we have:

$$EW_n = 2 \sum_{k=1}^{n} \int\left[\frac{s_t(\underset{\sim}{\varepsilon}_k)}{s_o(\underset{\sim}{\varepsilon}_k)} - 1\right] s_o^2(\underset{\sim}{\varepsilon}_k)\, d\underset{\sim}{\varepsilon}_k = -n \int\left[s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon})\right]^2 d\underset{\sim}{\varepsilon}.$$

Then observe that by Lemma B2:

$$\left|\left\{\int\left[\frac{s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon})}{t}\right]^2 d\underset{\sim}{\varepsilon}\right\}^{\frac{1}{2}} - \left\{\int\left[s_o'(\underset{\sim}{\varepsilon})\right]^2 d\underset{\sim}{\varepsilon}\right\}^{\frac{1}{2}}\right| \leq C^{\frac{1}{2}} \to 0$$

as $t \to 0$ (B.8)

Therefore we can approximate the first integral in (B.8) by the second and then:

$$EW_n \approx -nt^2 \int [s_0'(\underset{\sim}{\varepsilon})]^2 d\underset{\sim}{\varepsilon} \to -b^2/4 \qquad (B.9)$$

Now introduce the statistics:

$$T_n = -t \sum_{k=1}^{n} \left[tr\, \underset{\sim}{\Delta} + h'(d_o)\underset{\sim}{\varepsilon}_k^T \underset{\sim}{\Delta} \underset{\sim}{\varepsilon}_k / \left\{ d_o h(d_o) \right\} \right] = 2t \sum_{k=1}^{n} s_o'(\underset{\sim}{\varepsilon}_k)$$

and observe that:

$$ET_n = -t \sum_{k=1}^{n} \int \frac{\partial}{\partial t} f_t(\underset{\sim}{\varepsilon}_k) \Big|_{t=0} d\underset{\sim}{\varepsilon}_k = -t \sum_{k=1}^{n} \frac{\partial}{\partial t} \int f_t(\underset{\sim}{\varepsilon}_k) d\underset{\sim}{\varepsilon}_k \Big|_{t=0} = 0$$

$$\text{Var } T_n = 4t^2 \sum_{k=1}^{n} \int [s_o'(\underset{\sim}{\varepsilon}_k)]^2 d\underset{\sim}{\varepsilon}_k \to b^2$$

Furthermore, by the Central Limit Theorem it follows that:

$$T_n \approx N(0, b^2) \qquad (B.10)$$

Finally, note that by Lemma B2 we have:

$$\text{Var}(W_n - T_n) \le \sum_{k=1}^{n} \int \left[2 \frac{s_t(\underset{\sim}{\varepsilon}) - s_o(\underset{\sim}{\varepsilon})}{s_o(\underset{\sim}{\varepsilon})} - 2t \frac{s_o'(\underset{\sim}{\varepsilon})}{s_o(\underset{\sim}{\varepsilon})} \right]^2 s_o^2(\underset{\sim}{\varepsilon}) d\underset{\sim}{\varepsilon}$$

$$= 4t\,C \to 0 \text{ as } t \to 0 \qquad (B.11)$$

From (B.9)-(B.11) the result follows.

<u>Proof of theorem 4.1</u>. Considering a similar decomposition as in the proof of Theorem 3.1 and in view of Lemma B1, all we have to show is that:

$$P\left\{\sup n^{-\frac{1}{2}} \| \underset{\sim}{M}_n^*(\underset{\sim}{Y},\hat{\underset{\approx}{\Sigma}},\hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n^*(\underset{\sim}{Y},\underset{\approx}{\Sigma},\hat{\underset{\sim}{\beta}}) \| \geq \varepsilon :\ n^{\frac{1}{2}} \| \hat{\underset{\approx}{\Sigma}}^{\frac{1}{2}} - \underset{\approx}{\Sigma}^{\frac{1}{2}} \| \leq L\right\} \to 0$$

as $n \to \infty$ holds for all $L>0$, $\varepsilon>0$.

First define sets $S_{ih}(\delta)$, $(i,h=1,\ldots,r;\ \delta=0,1)$ as in Lemma A1. Then partition each of these sets into two disjoint sets, one containing those k's for which $\mathrm{sign}(y_{k\ell} - \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}}_\ell) = \mathrm{sign}(y_{km} - \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}}_m)$, $\ell=1$, $m=2$ and the other corresponding to its complement. Repeat the procedure with each new set by letting $\ell,m=1,\ldots,p$, $\ell<m$. Then define $2^{r+p(p-1)/2}$ disjoint sets S_{ig}, $(g=1,\ldots,2^{r+p(p-1)/2})$ in a similar way as was done in Lemma A1. Note that within each of these sets $M_{nij}(\underset{\sim}{Y},\hat{\underset{\approx}{\Sigma}},\hat{\underset{\sim}{\beta}})$ is a monotone function in each element of $(\hat{\underset{\approx}{\Sigma}},\hat{\underset{\sim}{\beta}})$ when the others are held fixed. The rest of the proof follows in the lines of that of Lemma A1. Lemmas B2 and B3 are used in step 3.

APPENDIX C

Proof of theorem 5.1. We first consider the univariate case (p=1). Note that we can restrict ourselves to the case $\underset{\sim}{X} = \underset{\sim}{I}$ and observe that the only component of $\underset{\sim}{\beta}(H_t)$ affected by the contamination is the one corresponding to the nonnull element of $\underset{\sim}{x}_o$. Furthermore we assume σ^2 known, and take $\sigma = 1$ without loss of generality. Now, for every sequence $\underset{\sim}{\beta}(H_{tn})$, where: $H_{tn} = (1-t)H + t\delta(\varepsilon_n, \underset{\sim}{x}_o)$, we must have:

$$-t\ \psi\left\{y_n - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn})\right\}\underset{\sim}{x}_o = (1-t)\ \frac{n_o}{n} \int \psi\left\{y - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn})\right\} dF(\varepsilon)\underset{\sim}{x}_o \qquad \text{(C.1)}$$

where n_o can take the values n_ℓ ($\ell=1,\ldots,r$) according to the value taken by $\underset{\sim}{x}_o$. Equating the coefficients of $\underset{\sim}{x}_o$ and taking absolute value on both sides we obtain:

$$t\left|\psi\left\{y_n - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn})\right\}\right| = (1-t)\ \frac{n_o}{n}\left| \int \psi\left\{y - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn})\right\} dF(\varepsilon)\right| \qquad \text{(C.2)}$$

Suppose that for all $\delta(\varepsilon_o, \underset{\sim}{x}_o)$ we have $\| \underset{\sim}{\beta}(H_{tn}) \| \le K$. Then we can choose a sequence (y_n, x_o) such that $|\psi\{y_n - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn})\}| \to M$ as $n \to \infty$ where $M = \sup|\psi(\varepsilon)|$. Letting $n \to \infty$ in (C.2) we get:

$$t \lim_{n\to\infty} \left|\psi\left\{y_n - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn})\right\}\right| \le (1-t) \lim_{n\to\infty} \frac{n_o}{n} \int \left|\psi\left\{y - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn})\right\}\right| dF(\varepsilon)$$

which implies $t \le (1+p_o^{-1})^{-1}$. Since this relation must hold for $\underset{\sim}{x}_o = \underset{\sim}{x}_\ell$ ($\ell=1,\ldots,r$) we have:

$$t \le \min_{\ell=1,\ldots,r} \left(1 + p_\ell^{-1}\right)^{-1}. \qquad \text{(C.3)}$$

Now suppose that there is a sequence $\delta(\varepsilon_n, \underset{\sim}{x}_o)$ such that $\| \underset{\sim}{\beta}(H_{tn}) \| \to \infty$ as $n \to \infty$. Then from (C.2) letting $n \to \infty$, we get:

$$t \lim_{n\to\infty} \left| \psi\left\{ y_n - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn}) \right\} \right| = (1-t) \lim_{n\to\infty} \left| \int \psi\left\{ y - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn}) \right\} dF(\varepsilon) \right|$$

which implies:

$$t\, M \geq (1-t) p_o \left| \int \lim_{n\to\infty} \psi\left\{ y - \underset{\sim}{x}_o^T \underset{\sim}{\beta}(H_{tn}) \right\} dF(\varepsilon) \right| = (1-t) p_o M$$

and then, $t \geq (1+p_o^{-1})^{-1}$. Therefore,

$$t \geq \max_{\ell=1,\ldots,r} \left(1 + p_\ell^{-1}\right)^{-1} \tag{C.4}$$

From (C.3), (C.4) and the definition of the gross error breakdown point we get $t^* = \min_{\ell=1,\ldots,r} \left(1 + p_\ell^{-1}\right)^{-1}$. Note that if $n_\ell = n^*$, $(\ell=1,\ldots,r)$ then $t^* = (1+r)^{-1}$.

The last part of the theorem follows from the fact that if any $\underset{\sim}{x}_\ell$ $(\ell=1,\ldots,r)$ has a nonnull component in the direction orthogonal to $\underset{\sim}{x}_o$, the only solution to (C.1) is $t=0$.

The extension of the theorem to the multivariate case is routine and therefore omitted.

REFERENCES

[1] Carroll, R.J. (1978). On the asymptotic distribution of multivariate M-estimates. J. Multivariate Analysis 8, 361-371.

[2] Hájek, J. (1962). Asymptotically most powerful rank-order tests. Ann. Math. Statist. 33, 1124-1147.

[3] Hájek, J. and Šidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.

[4] Huber, P.J. (1981). Robust Statistics. Wiley, New York.

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

[6] Jurecková, J. and Sen, P.K. (1982). Simultaneous M-estimator of the common location and the scale-ratio in the two-sample problem. Math. Operationsforsch. Statist., Ser. Statistics 13, 163-169.

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

[8] MacRae, E.C. (1974). Matrix derivatives with an application to an adaptive linear decision problem. Ann. Statist. 2, 337-346.

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

[10] Maronna, R., Bustos, O., and Yohai, V. (1979). Bias and efficiency robustness of general M-estimators for regression with random carriers, in Proceedings of the 1979 workshop on smoothing techniques for curve estimation. Lecture notes in Mathematics 757, Springer-Verlag, Heidelberg.

[11] Patel, K.M. (1973). Hájek-Šidák approach to the asymptotic distribution of multivariate rank order statistics. J. Multivariate Analysis 3, 57-70.

[12] Pendergast, J.F. and Broffitt, J.D. (1981). Robust estimation in growth curve models, personal communication (paper presented at the 141st Annual Meeting of the American Statistical Association, Detroit).

[13] Sen, P.K. (1982). On M-tests in linear models. Biometrika 69, 245-248.