M-METHODS IN MULTIVARIATE LINEAR MODELS

by

Julio da Motta Singer

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1441

June 1983

M-METHODS IN MULTIVARIATE LINEAR MODELS

by

Julio da Motta Singer

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics

Chapel Hill

1983

Approved by

Advisor

Reader

Reader

SINGER, JULIO M. M-Methods in Multivariate Linear Models. (Under the direction of PRANAB K. SEN).

Two types of M-estimators of the parameters of the multivariate linear model are considered: the first one is based on a coordinatewise scoring procedure while the second utilizes a scoring method based on a Mahalanobis-type distance.

Under some general conditions on the error distribution, on the design matrix and on the score function, the asymptotic distribution of the coordinatewise M-estimator is derived by means of an asymptotic linearity result. It is then indicated how a similar approach can be employed to obtain the asymptotic distribution of the second type of M-estimators under the additional assumption of elliptical symmetry of the underlying error distribution. Robustness properties of both types of estimators are studied through their influence functions and breakdown points. Two families of robust M-estimators of the above types are compared through the Pitman asymptotic relative efficiency.

Two types of M-tests for the general linear hypothesis are also suggested: one is a direct analogue of Wald's test and the other corresponds to a χ^2 test in which estimates under the more constrained hypothesis alone are required. Their asymptotic distributions are derived and it is shown that they are equi-efficient under local Pitman-type alternatives.

It is indicated how the above M-methods can be extended to growth curve models and some more general linear models.

Computational algorithms are proposed and the suggested procedures are illustrated by a numerical example with data from the statistical literature.

ACKNOWLEDGEMENTS

I have had the honour of writing this dissertation under the direction of Dr. Pranab K. Sen, and I wish to express my sincere gratitude for his invaluable advice and support during this important period of my career. I am also thankful to the other members of my advisory committee, Dr. Roger C. Grimson, Dr. Ronald W. Helms, Dr. Lawrence L. Kupper and Dr. James C. Lamb.

I am deeply indebted to the Department of Statistics of the University of Sao Paulo and to CAPES, an agency of the Brazilian Ministry of Education and Culture for providing financial support throughout my doctoral program.

Special appreciation goes to Ms. Judy Harrelson for the excellent typing job.

The friends I made in Chapel Hill helped me overcome the burden of spending all these years away from my parents Edith and Jose. Particularly I mention Ed Stanek and his family whose warm friendship will not be forgotten.

My wife Maria Lucia and my daughter Carolina contributed the biggest effort of all. To them I dedicate this work.

M-METHODS IN MULTIVARIATE LINEAR MODELS

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Multivariate linear models have been of widespread interest to both the fields of theoretical and applied Statistics not only because of the variety of situations to which they are applicable, but also because the output of a typical analysis is usually simple and easy to interpret. The recent advances in the development of computer packages to perform the elaborate calculations inherent to the analysis of these models have made them even more attractive to statisticians in general.

The typical setup suitable for analysis through a multivariate linear model concerns a situation in which p measurements are taken on each of n experimental units. These measurements represent the values of p different characteristics of interest or of the same characteristic under p different conditions. The n experimental units may be classified into r groups each corresponding to a different treatment and covariates of interest may also be considered. Kleinbaum (1970) presents a review of several versions of multivariate linear models which have been studied in the literature and indicates the types of experimental situations for which each version is appropriate.

In matrix notation the Standard Multivariate Linear Model (SMLM) can be written as:

$$\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad (1.1.1)$$

where $\underset{\sim}{Y}(n\times p)$ is a matrix of observable random variables (r.v.'s), $\underset{\sim}{X}(n\times r)$ is a matrix of known constants, assumed of full rank $r < n$, $\underset{\sim}{\beta}(r\times p)$ is a matrix of unknown parameters and $\underset{\sim}{\varepsilon}(n\times p)$ is a matrix of random errors, the rows of which are assumed independently distributed with common distribution and density functions F and f respectively, mean vector $\underset{\sim}{0}$ and positive definite (p.d.) scatter matrix $\underset{\sim}{\Sigma}$.[1]

The analysis of such models generally consists of estimating the unknown parameters and testing hypotheses about its elements. These problems have been extensively considered in the statistical literature when F is assumed to be multivariate Normal. A good summary of the most common results in this subject is presented in Kleinbaum (1970). 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 allows 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 the departures can occur, but also because of the dependence among them.

In this research we are concerned with the application of M-methods to multivariate linear models. Basically we consider model (1.1.1) and

[1]We note here that the term scatter is used to depict both variation and orientation of the distribution. In general, we take $\underset{\sim}{\Sigma}$ as a scalar multiple of the covariance matrix when the latter exists.

study the problem of estimating the unknown parameter matrix $\underset{\sim}{\beta}$ and testing hypotheses about its elements while treating the scatter matrix $\underset{\sim}{\Sigma}$ as a nuisance parameter. Generally we will rely on asymptotic theory for the following reasons:

(i) there are not many well studied multivariate distributions other than the Normal;

(ii) even when alternative distributions are considered, it is difficult to derive the small sample properties of estimators and tests;

(iii) large sample theory permits us to obtain simple forms for these properties even under very general assumptions on the underlying error distribution.

In the remaining sections of this chapter we review the literature on the effects of the departure from the assumption of normality on the properties of the usual LS and ML estimators and tests, then we outline some of the general concepts of robustness and M-methods and indicate how they have been applied to linear models. We also introduce some of the notation to be used throughout this work.

In Chapter 2 we define two types of M-estimators for the parameter matrix $\underset{\sim}{\beta}$ and derive their asymptotic distributions under some general conditions on the error distribution F, on the design matrix $\underset{\sim}{X}$ and on the score functions used to define them. We use these asymptotic distributions to construct tests of hypotheses about the elements of $\underset{\sim}{\beta}$. We also study some robustness properties of the proposed M-estimators.

In Chapter 3 we consider an extension of the results of Chapter 2 to other versions of multivariate linear models such as the Growth Curve Model (GCM) or Kleinbaum's More General Linear Model (MGLM).

In Chapter 4 we compute the Pitman asymptotic relative efficiency (ARE) of one type of M-estimator with respect to the other for several choices of the underlying error distribution. We propose algorithms for computation of the M-estimates and present a numerical example using real data.

Finally in Chapter 5 we summarize the conclusions of this study and propose some topics for further research.

1.2 Notation

The following notation will be used throughout:

(i) Matrices and vectors are underscored by ~ ;

(ii) $\underset{\sim}{A} = \langle a_{ij} \rangle$ indicates that a_{ij} is the element in the i^{th} row and j^{th} column of $\underset{\sim}{A}$;

(iii) $\underset{\sim}{A}^T$ denotes the transpose of $\underset{\sim}{A}$;

(iv) $|\underset{\sim}{A}|$ denotes the determinant of $\underset{\sim}{A}$;

(v) $tr\underset{\sim}{A}$ denotes the trace of $\underset{\sim}{A}$;

(vi) $ch_j(\underset{\sim}{A})$ denotes the j^{th} largest characteristic root of $\underset{\sim}{A}$;

(vii) $\underset{\sim}{A}^*(ba\times 1)$ denotes the "rolled out by rows" version of an $(a\times b)$ matrix $\underset{\sim}{A}$;

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

(ix) $\underset{\sim}{\beta}_j(r\times 1)$, $(j=1,\ldots,p)$ denotes the j^{th} column of $\underset{\sim}{\beta}$.

1.3 Effects of the departure from the normality assumption on the properties of the classical parametric estimators and tests in linear models

As we have noted in the introduction, the classical parametric methods of analysis under the SMLM require assumptions of independence of the among subjects error r.v.'s, homoscedasticity and normality. Since in many cases we do not have ways of verifying the validity of these assumptions it is of special interest to examine the behaviour of the classical estimators and tests under conditions different from the assumed ones. Much research has been conducted in this area but a clear picture of the situation is still not available. One of the reasons for this, as Scheffé (1959, ch.10) pointed out, might be the fact that these assumptions can be violated in many more ways than they can be satisfied.

We summarize below some of the most important conclusions in this field; the reader is referred to Scheffé (1959, ch.10) and Arnold (1981, ch.10, 13) for a general discussion on this topic.

Of the three types of problems mentioned above, dependence of the among subjects error r.v.'s is the most difficult to cope with. In general the dependence structure is unknown and the analysis is possible only under some additional and rather restrictive assumptions, like those of autocorrelation models; even so, only approximate procedures are available. In the case where the dependence structure is known, the analysis can be performed by considering the generalized linear model.

Heteroscedasticity is well known to affect the performance of estimators and test statistics, but in many cases it can be at least partially compensated by using variance stabilizing transformations. The use of this technique, however, may affect the underlying distribution, and the transformed observations may lack normality even in the case where the original data followed a Normal distribution.

Many of the classical estimators and tests are known to be rather robust to mild departures from normality but it can be shown that their performance can be seriously affected by underlying distributions with heavy tails, such as t distributions, or by those usually employed in the literature to model the presence of outliers, such as contaminated Normal distributions.

In this study we restrict ourselves to examining the effects of departures from normality. We present some results on the asymptotic distributions of the LS and the Normal maximum likelihood (NML) estimators when the underlying distribution satisfies only some very general conditions and we also discuss the effects of lack of normality on the level of significance and power of the most common tests.

Working the univariate linear model (1.1.1) (with p=1), Eicker (1967) showed consistency of the LS estimator of the parameter vector $\underset{\sim}{\beta}$ and obtained its asymptotic distribution under certain regularity conditions on the error distribution and on the design matrix. His results do not even require the error r.v.'s to be identically distributed, though this assumption is needed for practical applications. An extension of one of his results to the more general case (p>1) which is of practical interest is given in the following theorem, the proof of which is omitted, since it essentially reduces to that of Eicker by first applying the Cramér-Wold theorem.

<u>Theorem 1.1</u>: Let $\hat{\underset{\sim}{\beta}}_{LS}$ be the LS estimator of $\underset{\sim}{\beta}$ in the SMLM given by (1.1.1); let F_j (j=1,...p) denote the marginal distribution function of the j^{th} coordinate and consider the following additional assumptions:

(i) $\int_{|x|>c} x^2 dF_j(x) \to 0$ as $n \to \infty$ for $c \to \infty$, $(j=1,\ldots p)$;

(ii) $\max_{1\le k\le n} \{\underset{\sim}{x}_k^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{x}_k\} \to 0$ as $n \to \infty$;

(iii) $\lim_{n\to\infty} \{n^{-1}(\underset{\sim}{X}^T\underset{\sim}{X})\} = \underset{\sim}{V}$, p.d.

Then $\sqrt{n}\,(\hat{\underset{\sim}{\beta}}{}^*_{LS} - \underset{\sim}{\beta}^*) \approx N_{rp}(\underset{\sim}{0},\underset{\sim}{V}^{-1} \otimes \underset{\sim}{\Sigma})$.

Note that if the underlying error distribution is Normal the corresponding Fisher information matrix is given by $\underset{\sim}{I}(\underset{\sim}{\beta}) = \underset{\sim}{V} \otimes \underset{\sim}{\Sigma}^{-1}$ and the LS estimator is fully efficient. This might not be true otherwise. Furthermore note that nothing has been said about the rate of convergence; although the asymptotic distribution of $\hat{\underset{\sim}{\beta}}_{LS}$ is Normal, the convergence can be very slow for distributions with heavy tails as is indicated in Cramér (1946, ch.17). In these cases an unreasonably large number of observations is needed for the result to be useful.

Consistency of the LS estimator follows directly from the asymptotic normality, though Eicker has shown that it holds under less restrictive assumptions.

Huber (1967) studied the asymptotic properties of ML estimators and showed that under some mild regularity conditions they are consistent and asymptotically normally distributed even when the true underlying error distribution does not belong to the parametric family defining the ML estimator. A similar result concerning the asymptotic distribution of ML estimators in the specific case of the SMLM is presented in the theorem below. It is less general than Huber's result but also requires less stringent assumptions. The proof essentially follows the lines indicated in Cramér (1946, ch.33) and is omitted.

<u>Theorem 1.2</u>: Let $\hat{\underset{\sim}{\beta}}_F$ be the ML estimator of $\underset{\sim}{\beta}$ in the SMLM given by (1.1.1). Let G and g (instead of F and f) respectively denote the distribution and density functions of the true underlying error distribution. Recall that $\underset{\sim}{\varepsilon}_k = \underset{\sim}{y}_k - \underset{\sim}{\beta}^T \underset{\sim}{x}_k$ and consider the following additional assumptions:

(i) $(\partial/\partial\underset{\sim}{\beta})\log f(\underset{\sim}{\varepsilon}_k)$ and $(\partial^2/\partial\underset{\sim}{\beta}\partial\underset{\sim}{\beta}^T)\log f(\underset{\sim}{\varepsilon}_k)$ exist for all $\underset{\sim}{\beta}\in\mathbb{R}^r$ and are dominated by some G-integrable functions;

(ii) $$E_G\{\sup\| (\partial^2/\partial\underset{\sim}{\beta}\partial\underset{\sim}{\beta}^T)\log f(\underset{\sim}{\varepsilon}_k)|_{\underset{\sim}{\beta}} - (\partial^2/\partial\underset{\sim}{\beta}\partial\underset{\sim}{\beta}^T)\log f(\underset{\sim}{\varepsilon}_k)|_{\underset{\sim}{\beta}^+}\| :$$
$$\|\underset{\sim}{\beta}^+-\underset{\sim}{\beta}\| < \varepsilon\} \to 0 \text{ as } \varepsilon \to 0 ;$$

(iii) $E_G\{(\partial/\partial\underset{\sim}{\beta})\log f(\underset{\sim}{\varepsilon}_k)\} = \underset{\sim}{0}$ for all $\underset{\sim}{\beta}\in\mathbb{R}^r$;

(iv) $E_G\{[(\partial/\partial\underset{\sim}{\beta})\log f(\underset{\sim}{\varepsilon}_k)][(\partial/\partial\underset{\sim}{\beta})\log f(\underset{\sim}{\varepsilon}_k)]^T\} = \underset{\sim}{\Lambda}(\underset{\sim}{\beta})$ is non-singular for all $\underset{\sim}{\beta}\in\mathbb{R}^r$ and continuous in $\underset{\sim}{\beta}$;

(v) $\lim_{n\to\infty} \{n^{-1}(\underset{\sim}{X}^T\underset{\sim}{X})\} = \underset{\sim}{V}$, p.d.

Then $\sqrt{n}\,(\hat{\underset{\sim}{\beta}}{}^*_F-\underset{\sim}{\beta}^*) \approx N_{rp}\{\underset{\sim}{0},\ \underset{\sim}{V}^{-1} \otimes \underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\beta})\}$.

Here we note that (ii) is an alternative to the assumption of the existence of the third derivative of $\log f(\underset{\sim}{\varepsilon}_k)$ required in Cramér's proof. We also note that if G = F then $\underset{\sim}{\Lambda}(\underset{\sim}{\beta}) = \underset{\sim}{I}(\underset{\sim}{\beta})$, the Fisher information matrix, and in this case the ML estimator is fully efficient. This might not be true if $G \neq F$, in which case our previous comments on the rate of convergence are still applicable.

As opposed to the estimation problem, no general results are available to assess the effect of departure from the normality assumption in the hypothesis testing setup. Most of the work in this area has been done for univariate models and within this context the more useful results pertain to the one-way fixed effects analysis of variance problem. Some studies for multi-way classification and random effects models have also been conducted. Only very recently has this type of problem been studied in the multivariate setting. The reader is referred to Scheffé

(1959, ch.10) and to Ito (1980) for general reviews in the area. We shall summarize below some of their most important conclusions.

The effects of non-normality on the level of significance of tests have been more closely examined than the corresponding effects on the power of tests possibly because of the difficulty in dealing with the usually complicated non-central distributions involved.

In the univariate case theoretical results on the effects of lack of normality were usually obtained by expressing the level of significanse and power of tests as functions of the first few cumulants of the underlying error distribution and examining their behaviour for different alternative distributions. Monte Carlo methods have also been employed in similar investigations and there seems to be good agreement between the theoretical and empirical results at least for alternatives such as the Exponential and Log-normal distributions. As in the estimation case, the usual tests are robust if the departures from normality are mild; some distributions, such as those mentioned above, can lead to conservative levels of significance and to higher power than in the Normal case. Furthermore, the power based on the non-Normal distributions approaches that based on the Normal distribution as the sample size increases.

In the multivariate case the few available theoretical results in this area were obtained by investigating the behaviour of an assumed large sample distribution family of the test statistic for different underlying error distributions, or by considering the permutation test approach. Most of the Monte Carlo work relates to the special case of two groups. The conclusions tend to indicate that the effect of non-normality is different for the different test statistics and depends on factors such as the dimensionality, and the number of observations. There are some indications that some of the usual multivariate test statistics may be seriously affected by departures from normality and we expect an increase in the effects as the dimensionality increases.

In view of the above results on the effects of non-normality in estimation theory and of the lack of more precise ones in hypothesis testing it seems reasonable to consider alternative statistical methods that could be applied to situations where outliers are expected. We discuss an alternative in the next section.

1.4 Robustness and M-methods

In this section we introduce some general concepts of robustness and M-methods. The reader is referred to Huber (1981, ch.1-3) and Serfling (1980, ch.6-7) for details.

Huber points out that since the assumptions we usually make for statistical analyses are mere mathematically convenient representations of our usually fuzzy knowledge or belief, we should look for procedures that:

(i) have a reasonably good (optimal or near optimal) efficiency at an assumed model;

(ii) are robust in the sense that small deviations from the model assumptions should impair their performance only slightly, that is, their performance should be close to the nominal value calculated at the model (in terms of the asymptotic variance, for example);

(iii) are such that somewhat larger deviations should not cause a catastrophe.

Robustness in the above sense can be more precisely characterized through the theory of "statistical functions"; we indicate below some of the basic ideas.

Many of the most common estimators can be written as functionals $T_n = T(F_n)$ of the empirical distribution function F_n; these functionals are called statistical functions. For example for the variance parameter σ^2 the appropriate functional is $T(F) = \int\{x-\int x dF(x)\}^2 dF(x)$ and T_n is the sample variance. In general, we have $T_n \to T(F)$ and $\mathcal{L}_F[\sqrt{n}\ \{T_n-T(F)\}] \to N\{0,A(F,T)\}$.

Qualitative robustness can be viewed as a form of continuity of the functional T. We say that the sequence $\{T_n,n\geq 1\}$ is robust at $F = F_0$ if given $\varepsilon > 0$, there exists $\delta > 0$ and $n_o > 0$ such that for all F and $n\geq n_o$:

$$d(F_o,F) \leq \delta \Rightarrow d\{\mathcal{L}_{F_o}(T_n),\ \mathcal{L}_F(T_n)\} \leq \varepsilon$$

where d is a convenient distance function.

Quantitatively robustness can be measured by the amount of change in the asymptotic distribution $\mathcal{L}_F(T_n)$ induced by a variation in F. The most common measure of robustness is the asymptotic breakdown point t*, defined by:

$$t^* = \sup\{t: b(t)<b(1)\}$$

where:

$$b(t) = \lim_{n\to\infty} \sup_{F\in P_t} |\mathrm{med}[\mathcal{L}_F\{T_n-T(F_o)\}]|$$

is the maximum asymptotic bias and P_t is some neighbourhood of F_o, i.e. a Lévy neighbourhood:

$$P_t(F_o) = \{F: \forall x,\ F_o(x-t)-t \leq F(x) \leq F_o(x+t)+t\}$$

or a "contamination" neighbourhood:

$$P_t(F_o) = \{F:F = (1-t)F_o+t\ H \text{ and H is a probability measure}\}\ .$$

It corresponds to the limiting fraction of outliers the estimator defined by T can cope with and therefore is a measure of global robustness.

Another important tool in the study of robustness is the influence function, defined by:

$$i(x;F,T) = \lim_{t\to 0} \{[T\{(1-t)F+t\delta_x\} - T(F)]/t\}$$

where δ_x denotes the pointmass 1 at x. Roughly speaking it measures the influence of a single observation on the value of a statistical function $T(F_n)$. Some useful measures of the robustness of $T(F_n)$ can be derived from its influence function. Among these are the "gross error sensitivity" which measures the effects of contamination of the data by gross errors, the "local shift sensitivity" which measures the effects of rounding or grouping observations and the "rejection point" which is the distance from the centre of symmetry of the distribution, above which all observations are rejected. These concepts were introduced and discussed in detail by Hampel (1974).

Many methods have been proposed to obtain robust estimators and tests of hypothesis. Among these, M-methods (Maximum Likelihood-type methods) stand out as very convenient not only because of their easy statistical interpretation but also because of their fairly tractable mathematical properties. We will restrict ourselves to the study of such methods.

Define the functional T(F) as a solution to

$$\int\psi\{x,T(F)\}dF(x) = 0 \qquad (1.4.1)$$

where $\psi(x,t)$ is a real-valued function. For a random sample $X_1,\ldots X_n$ from F, the corresponding M-estimator of T(F) is a solution to:

$$\sum_{k=1}^{n}\psi(X_k,T_n) = 0 \qquad (1.4.2)$$

In typical cases (1.4.1) and (1.4.2) respectively define the parameter of interest and its estimator and correspond to the problem of minimizing $\int\rho\{x,T(F)\}dF(x)$ and $\sum_{k=1}^{n}\rho(X_k,T_n)$ where $\psi(x,t) = (\partial/\partial t)\rho(x,t)$. It can be easily seen that the choice $\rho(x,t) = -\log f(x,t)$ produces the ordinary ML estimator and in the location case the choice $\rho(x-t)=(x-t)^2$ produces the LS estimator.

In general small sample properties of M-estimators are not available and their theoretical study rely on asymptotic theory. M-tests are usually derived from the asymptotic distributions of M-estimators.

An interesting and useful property of M-estimators is that the corresponding influence function is proportional to the score function ψ; this enables us to produce robust M-estimators by conveniently choosing ψ. For example, bounded ψ functions generate M-estimators which are insensitive to the presence of outliers, ψ functions which are proportional to $-f_o'(x)/f_o(x)$ generate M-estimators which are highly efficient at the model with density f_o, continuous ψ functions generate M-estimators which are insensitive to roundoff errors or grouping.

Among the many ψ functions proposed in the literature which produce robust M-estimators, the most commonly used and most statistically appealing are the ones suggested by Huber (1964):

$$\psi(x) = \begin{cases} x & \text{if } |x| \le k \\ k\,\mathrm{sign}(x) & \text{if } |x| > k \end{cases} \tag{1.4.3}$$

where k is a constant, and by Hampel (1974):

$$\psi(x) = \begin{cases} x & , \quad |x| \le a \\ a\,\mathrm{sign}(x) & , \; a \le |x| \le b \\ a\{x-c\,\mathrm{sign}(x)\}/(b-c) & , \; b \le |x| \le c \\ 0 & , \quad |x| \ge c \end{cases} \tag{1.4.4}$$

where a, b and c are positive constants. Estimators obtained by using (1.4.3) have the property of limiting without entirely eliminating the influence of outliers or extreme observations while those obtained by using (1.4.4) have the property of completely rejecting them.

A large portion of the literature on M-methods has been devoted to the study of location problems. Since location M-estimators are not scale-invariant, in practice we have to take scale into account, at least as a nuisance parameter. Simultaneous estimation of location and scale may pose some theoretical difficulties, but a simpler alternative which produces scale-invariant M-estimators of location is available. It is known in the literature by "M-estimation of location with preliminary estimate of scale" and consists of taking any location-invariant estimator of scale S_n and defining the M-estimator of location T_n as a solution to:

$$\sum_{k=1}^{n} \psi\{(X_k - T_n)/S_n\} = 0 \ .$$

In this case, if S_n is an estimator of $S(F)$, a scale parameter of the underlying distribution, T_n will be an estimator of $T(F)$ defined by:

$$\int \psi[\{x - T(F)\}/S(F)]dF(x) = 0$$

and its asymptotic properties depend on $S(F)$ rather than on S_n. The reader is referred to Huber (1981, ch.6) for details.

We will now discuss some of the most important results on the application of M-methods to multivariate problems.

Maronna (1976) studied the problem of simultaneous M-estimation of location $\underset{\sim}{\mu}$ and scatter $\underset{\sim}{\Sigma}$ for elliptically symmetric distributions with absolutely continuous density functions f given by:

$$f(\underset{\sim}{\varepsilon}) = |\underset{\sim}{\Sigma}|^{-\frac{1}{2}} h\{(\underset{\sim}{\varepsilon}^T \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\varepsilon})^{\frac{1}{2}}\} \qquad (1.4.5)$$

where h is a scalar multiple of a density in $\mathbb{R}$ such that $h'(x) = (\partial/\partial x)h(x)$ exists. He considered solutions to the system:

$$\begin{aligned} E_p u_1\{d(\underset{\sim}{x};\underset{\sim}{\mu},\underset{\sim}{\Sigma})\}(\underset{\sim}{x}-\underset{\sim}{\mu}) &= \underset{\sim}{0} \\ E_p u_2\{d^2(\underset{\sim}{x};\underset{\sim}{\mu},\underset{\sim}{\Sigma})\}(\underset{\sim}{x}-\underset{\sim}{\mu})(\underset{\sim}{x}-\underset{\sim}{\mu})^T &= \underset{\sim}{\Sigma} \end{aligned} \tag{1.4.6}$$

where u_1 and u_2 are real-valued functions satisfying certain conditions and:

$$d^2(\underset{\sim}{x};\underset{\sim}{\mu},\underset{\sim}{\Sigma}) = (\underset{\sim}{x}-\underset{\sim}{\mu})^T\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) \tag{1.4.7}$$

If P is the underlying distribution (1.4.6) defines the parameters being estimated; if P is the empirical distribution (1.4.6) defines the corresponding M-estimators, which can be viewed as generalizations of ML estimators. In fact, taking $u_1(d) = -d^{-1}(\partial/\partial d)\log h(d)$ and $u_2(d^2) = u_1(d)$ we obtain the ML estimators. If otherwise we define

$$u_1(d) = \begin{cases} 1 & , \ 0 \le d \le k \\ k/d & , \ d > k \end{cases}$$

and $u_2(d^2) = u_1(d)$ we obtain M-estimators which de-emphasize outlying or extreme observations based on their Mahalanobis distance to the corresponding mean vector. By making $\psi_i(d) = du_i(d)$, $(i=1,2)$ we facilitate the comparison to the univariate case and it is easy to see that this last suggestion corresponds to Huber's proposal.

Maronna showed that solutions to (1.4.6) exist under certain regularity conditions which include ψ_1 to be bounded and ψ_2 to be bounded and non-decreasing. He proved that the solution is unique when P is the underlying distribution and conjectured that this is also true when P is the empirical distribution. Furthermore he showed that the proposed estimators are consistent, asymptotically Normal and asymptotically independent. He also indicated that departures from the assumption of elliptical symmetry has a more serious effect on the robustness of the M-estimators than longtailedness within that class of distributions and that an increase in dimensionality also affects robustness. He proposed an algorithm for computation of the M-estimates but did not provide a proof of convergence.

Carroll (1978) examined some further asymptotic properties of these estimators.

Collins (1982) proposed some M-estimators of multivariate location which allow for redescending score functions such as (1.4.4). He showed that these estimators are consistent and asymptotically Normal if the underlying distribution belongs to a specified neighbourhood of a class of distributions spherically symmetric about $\underset{\sim}{0}$. He also indicated that as in Maronna's case the proposed estimators have low breakdown points.

1.5 M-methods in linear models

The application of robust procedures to linear models is specially appealing since in such cases outliers are usually more difficult to detect. In this section we review some of the literature on the application of M-methods to linear models and discuss their robustness properties.

Huber (1973) defined M-estimators of the parameter vector $\underset{\sim}{\beta}$ in the univariate linear model given by (1.1.1) with p=1 as a solution to:

$$\sum_{k=1}^{n} \underset{\sim}{x}_k \psi\{(y_k - \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}})/\hat{\sigma}\} = \underset{\sim}{0} \tag{1.5.1}$$

where ψ is a real-valued score function and $\hat{\sigma}$ is a location invariant estimator of a scale parameter σ. He showed that the M-estimator is unique, consistent and asymptotically Normal, provided that among other mild regularity conditions the score function is bounded and non-decreasing and $hr^2 \to 0$ as $n \to \infty$, where $h = \max_{1 \le k \le n} \{\underset{\sim}{x}_k^T (\underset{\sim}{X}^T \underset{\sim}{X})^{-1} \underset{\sim}{x}_k\}$ and r is the dimension of $\underset{\sim}{\beta}$. He also showed that some weaker results hold in the case where $hr \to 0$ as $n \to \infty$.

Since $hr \to 0$ implies $r^2/n \to 0$, even in this case a very large number of observations is required already for a moderate dimension of $\underset{\sim}{\beta}$ and the results have little practical application. Based on formal (with no bounds on the remainder terms) asymptotic expansions of (1.5.1) around the true parameter value and on Monte-Carlo studies, Huber (1973, 1981, ch.7) conjectures that this requirement can be relaxed and that perhaps $r/n \to 0$ is already sufficient as long as the error distribution and the function $\rho(x) = \int\psi(x)dx$ are symmetric. Yohai and Maronna (1979) improved Huber's results by eliminating some boundedness conditions on the derivatives of ψ and showing that $hr^{3/2} \to 0$ as $n \to \infty$ is sufficient for consistency and asymptotic normality of the M-estimator. If the assumption on the monotonicity of ψ is relaxed these results still hold in the local sense.

Following Hampel (1974) it is easy to see that the influence function of the M-estimator of $\underset{\sim}{\beta}$ is given by:

$$\underset{\sim}{i}(\varepsilon_o,\underset{\sim}{x}_o;F,\hat{\underset{\sim}{\beta}}) = \{\sigma\psi(\varepsilon_o/\sigma)/E_F\psi'(\varepsilon_o/\sigma)\}\ \underset{\sim}{V}^{-1}\underset{\sim}{x}_o$$

where $E_F\psi'(\varepsilon_o/\sigma) = \int\psi'(\varepsilon/\sigma)dF(\varepsilon)$. Even if ψ is bounded the influence of a point $(\varepsilon_o,\underset{\sim}{x}_o)$ can be arbitrarily large because $\underset{\sim}{x}_o$ multiplies $\psi(\varepsilon_o/\sigma)$; therefore the corresponding M-estimator can be severely affected by outlying observations in the independent variables. If these are under control, however, M-estimators of $\underset{\sim}{\beta}$ can be made robust by an appropriate choice of the score functions. Some authors like Maronna, Bustos and Yohai (1979) or Krasker and Welsch (1982) have considered alternative bounded-influence estimators by weighting both the dependent and independent variables or by using some jackknife procedures but we shall not discuss these alternatives here. Some further results on the robustness of M-estimators in linear models will be presented in section 2.5.

In general the set of equations (1.5.1) is nonlinear and has to be solved iteratively. Klein and Yohai (1981) classify various of the algorithms proposed in the literature into two categories which we describe below. The first corresponds to algorithms of the form:

$$\begin{cases} \hat{\beta}^{(0)} = \tilde{\beta} \\ \hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + \hat{\sigma}\left[\sum_{k=1}^{n} r\{\hat{\varepsilon}_k^{(m)}/\hat{\sigma}\} x_k x_k^T\right]^{-1} \sum_{k=1}^{n} x_k \psi\{\hat{\varepsilon}_k^{(m)}/\hat{\sigma}\} \quad, m \geq 0 \end{cases}$$

where $\hat{\varepsilon}_k^{(\ell)} = y_k - x_k^T \hat{\beta}^{(\ell)}$, $\tilde{\beta}$ is any reasonable location-scale equivariant estimator of β, $\hat{\sigma}$ is an initial $n^{\frac{1}{2}}$-consistent estimate of scale and r is a conveniently chosen function. It includes the Newton-Raphson method for $r(t) = \psi'(t)$, the iteratively reweighted least squares method for $r(t) = \psi(t)/t$ and the iterative winsorization method for $r(t) =$ constant. The second category corresponds to algorithms which are approximations of those in the first one and are of the form:

$$\begin{cases} \hat{\beta}^{(0)} = \tilde{\beta} \\ \hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + \hat{\sigma}\left[n^{-1}\sum_{k=1}^{n} r\{\hat{\varepsilon}_k^{(m)}/\hat{\sigma}\} \sum_{k=1}^{n} x_k x_k^T\right]^{-1} \sum_{k=1}^{n} x_k \psi\{\hat{\varepsilon}_k^{(m)}/\hat{\sigma}\}, \; m \geq 0. \end{cases}$$

Possible choices for the starting point $\hat{\beta}^{(0)}$ include the LS estimator or the more robust (but more difficult to compute) least absolute residual estimator. To estimate scale the most common suggestion is $\hat{\sigma} = \text{med}|\hat{\varepsilon}_k^{(0)}|$.

Holland and Welsch (1977) discuss some advantages and disadvantages of both types of algorithms for different choices of the r function. Huber (1981) showed the convergence of an algorithm for the simultaneous estimation of β and σ, that is, for obtaining simultaneous solutions of (1.5.1) and

$$\sum_{k=1}^{n} \chi\{(y_k - \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}})/\hat{\sigma}\} = 0$$

where $\chi(x) = x\psi(x) - \rho(x)$. His proof, however, only holds when ψ is monotone. If this is not so one can only obtain local convergence since (1.5.1) may have more than one solution and in this case a good starting point is very important. Klein and Yohai (1981) have shown the convergence of the algorithms discussed previously when ψ is non-monotone provided certain regularity conditions are met; these conditions include starting the process with an estimator $\hat{\underset{\sim}{\beta}}^{(0)}$ such that $(\underset{\sim}{X}^T\underset{\sim}{X})^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}^{(0)} - \underset{\sim}{\beta})$ is bounded in probability and a $n^{\frac{1}{2}}$-consistent estimate of scale. Except in Huber's simultaneous estimation proposal no convergence proofs are available when iteration on $\hat{\sigma}$ is performed.

Bickel (1975) studied M-estimators obtained by iterating the Newton-Raphson algorithm only once (one-step estimators) from a consistent starting point and showed that their asymptotic properties are the same as those of the completely iterated estimators.

Holland and Welsch (1977) through a limited Monte-Carlo study examined the effect of the number of iterations and that of estimating scale on several alternative estimators, including M-estimators. Their results indicated that in general the estimators are more sensitive to the estimation of scale than to the number of iterations.

To test linear hypotheses of the form:

$$H: \underset{\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{0} \tag{1.5.2}$$

where $\underset{\sim}{C}$ is a $(c \times p)$ matrix of rank c, three (asymptotically equi-efficient) M-tests have been proposed in the literature. They are direct analogues of the classical parametric inference tests such as Wald's test, likelihood ratio tests and χ^2 tests in which estimates under the more constrained hypothesis alone are required. We shall briefly introduce them below.

The first M-test corresponds to Wald's test and is given by the following quadratic form in the unrestricted M-estimator $\hat{\beta}$:

$$W_M = (C\hat{\beta})^T\{C(X^TX)^{-1}C^T\}^{-1}(C\hat{\beta})$$

It can be shown that under the null hypothesis (1.5.2) W_M is asymptotically distributed as a central χ^2 with c degrees of freedom (d.f.) if certain regularity conditions hold. Similarly under Pitman-type alternatives of the form:

$$A_n: C\beta = n^{-\frac{1}{2}}\Delta \tag{1.5.3}$$

where Δ ($\neq 0$) is a fixed (c×1) vector, W_M is asymptotically distributed as a non-central χ^2 with c d.f. and a certain non-centrality parameter Ω.

Schrader and Hettmansperger (1980) proposed a likelihood ratio-type test which is an analogue of the likelihood ratio test of classical parametric inference. Their test statistic is given by:

$$F_M = 2\hat{\omega}\hat{\phi}^{-2}\sum_{k=1}^{n}[\rho\{(y_k-x_k^T\hat{\beta}_H)/\hat{\sigma}\}-\rho\{(y_k-x_k^T\hat{\beta})/\hat{\sigma}\}]$$

where $\hat{\beta}_H$ is the restricted estimator of β, $\hat{\sigma}$ is a scale estimate,

$$\hat{\omega} = n^{-1}\sum_{k=1}^{n}\psi'\{(y_k-x_k^T\hat{\beta})/\hat{\sigma}\} \quad \text{and} \quad \hat{\phi}^2 = (n-r)^{-1}\sum_{k=1}^{n}\psi^2\{(y_k-x_k^T\hat{\beta})/\hat{\sigma}\} .$$

These authors showed that under either the null hypothesis (1.5.2) or alternative hypotheses of the type (1.5.3) F_M has the same asymptotic distribution as W_M and thus is asymptotically equivalent to it. They also argued that F_M may be easier to compute in some complicated problems, though the situation is reversed in the more common cases.

Sen (1982) proposed an M-test similar to the classical parametric inference χ^2 test in which estimates under the more constrained hypothesis alone are required. It is asymptotically equivalent to the two previous ones and in many cases computationally more convenient. It also incorporates a larger class of score functions ψ than the one proposed by Schrader and Hettmansperger. To simplify notation we consider a canonical reduction of the model (1.1.1) and without loss of generality we may restrict ourselves to hypotheses of the form H: $\underset{\sim}{\beta}_C = \underset{\sim}{0}$ under the model:

$$\underset{\sim}{Y} = [\underset{\sim}{A}\ \underset{\sim}{Q}] \begin{bmatrix} \underset{\sim}{\beta}_H \\ \underset{\sim}{\beta}_C \end{bmatrix} + \underset{\sim}{\varepsilon}$$

where $\underset{\sim}{A}$ (n×r-c) and $\underset{\sim}{Q}$(n×c) are known matrices and $\underset{\sim}{\beta}_H$(c-r×1) and $\underset{\sim}{\beta}_C$(c×1) are the unknown parameters. The proposed test statistic is given by:

$$T_M = \hat{\phi}_R^{-2}\hat{\underset{\sim}{M}}^T\{\underset{\sim}{Q}^T\underset{\sim}{Q}-\underset{\sim}{Q}^T\underset{\sim}{A}(\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\underset{\sim}{A}^T\underset{\sim}{Q}\}^{-1}\hat{\underset{\sim}{M}}$$

where $\hat{\sigma}_R^2 = n^{-1}\sum_{k=1}^{n}\psi^2\{(y_k-\underset{\sim}{a}_k^T\hat{\underset{\sim}{\beta}}_H)/\hat{\sigma}\}$, $\hat{\underset{\sim}{M}} = \sum_{k=1}^{n}\psi\{(y_k-\underset{\sim}{a}_k^T\hat{\underset{\sim}{\beta}}_H)/\hat{\sigma}\}\underset{\sim}{q}_k$,

$\hat{\underset{\sim}{\beta}}_H$ is a solution to $\sum_{k=1}^{n}\psi\{(y_k-\underset{\sim}{a}_k^T\hat{\underset{\sim}{\beta}}_H)/\hat{\sigma}\}\underset{\sim}{a}_k = \underset{\sim}{0}$, $\underset{\sim}{a}_k$ and $\underset{\sim}{q}_k$ are the transposes of the k^{th} row of $\underset{\sim}{A}$ and Q respectively and $\hat{\sigma}$ is an estimate of scale. If certain regularity conditions hold, he showed that T_M is asymptotically equivalent to the two previous tests under both the null hypothesis $\underset{\sim}{\beta}_C = \underset{\sim}{0}$ or Pitman-type alternatives of the form A_n: $\underset{\sim}{\beta}_C = n^{-\frac{1}{2}}\underset{\sim}{\Delta}$ where $\underset{\sim}{\Delta}(\neq\underset{\sim}{0})$ is a fixed (c×1) vector. The non-centrality parameter of the asymptotic χ^2 distribution under the sequence of alternative hypotheses is given by $\Omega = \omega^2\phi^{-2}\underset{\sim}{\Delta}^T\{\underset{\sim}{Q}^T\underset{\sim}{Q}-\underset{\sim}{Q}^T\underset{\sim}{A}(\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\underset{\sim}{A}^T\underset{\sim}{Q}\}\underset{\sim}{\Delta}$ where $\omega = E_F\psi'(\varepsilon/\sigma)$ and $\phi^2 = E_F\psi^2(\varepsilon/\sigma)$. T_M is simpler to calculate than F_M since it does not require the unrestricted M-estimator $\hat{\underset{\sim}{\beta}}$ nor the computation of $\hat{\omega}$.

Very little attention has been directed to the application of M-methods to multivariate linear models. Pendergast and Broffitt (1981) considered an extension of Maronna's (1976) results to the GCM given by:

$$\underset{\sim}{Y} + \underset{\sim}{X}\,\underset{\sim}{\xi}\,\underset{\sim}{G} + \underset{\sim}{\varepsilon} \tag{1.5.4}$$

where $\underset{\sim}{Y}$, $\underset{\sim}{X}$ and $\underset{\sim}{\varepsilon}$ are as in (1.1.1), $\underset{\sim}{\xi}(r\times q)$ is a matrix of unknown parameters and $\underset{\sim}{G}(q\times p)$ is a matrix of known constants assumed of full rank $q\leq p$.

Their M-estimators are defined as solutions to:

$$
\begin{aligned}
&n^{-1}\sum_{k=1}^{n} u_1(d_k)\underset{\sim}{x}_k(\underset{\sim}{y}_k-\underset{\sim}{G}^T\hat{\underset{\sim}{\xi}}^T\underset{\sim}{x}_k)^T\hat{\underset{\sim}{\Sigma}}^{-1}\underset{\sim}{G}^T = \underset{\sim}{0} \\
&n^{-1}\sum_{k=1}^{n} u_2(d_k^2)(\underset{\sim}{y}_k-\underset{\sim}{G}^T\hat{\underset{\sim}{\xi}}^T\underset{\sim}{x}_k)(\underset{\sim}{y}_k-\underset{\sim}{G}^T\hat{\underset{\sim}{\xi}}^T\underset{\sim}{x}_k)^T = \hat{\underset{\sim}{\Sigma}}
\end{aligned}
\tag{1.5.5}
$$

where u_1 and u_2 are real-valued functions and

$$
d_k^2 = (\underset{\sim}{y}_k-\underset{\sim}{G}^T\hat{\underset{\sim}{\xi}}^T\underset{\sim}{x}_k)^T\hat{\underset{\sim}{\Sigma}}^{-1}(\underset{\sim}{y}_k-\underset{\sim}{G}^T\hat{\underset{\sim}{\xi}}^T\underset{\sim}{x}_k) \ . \tag{1.5.6}
$$

Alternatively, if we let $\underset{\sim}{U}_i = \text{diag}\{u_i(d_1^i),\ldots u_i(d_n^i)\}$, $(i=1,2)$ the above M-estimators can be expressed as solutions to:

$$
\begin{aligned}
&\hat{\underset{\sim}{\xi}} = (\underset{\sim}{X}^T\underset{\sim}{U}_1\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{U}_1\underset{\sim}{Y}\hat{\underset{\sim}{\Sigma}}^{-1}\underset{\sim}{G}^T(\underset{\sim}{G}\hat{\underset{\sim}{\Sigma}}^{-1}\underset{\sim}{G}^T)^{-1} \\
&\hat{\underset{\sim}{\Sigma}} = n^{-1}(\underset{\sim}{Y}-\underset{\sim}{X}\hat{\underset{\sim}{\xi}}\underset{\sim}{G})^T\underset{\sim}{U}_2(\underset{\sim}{Y}-\underset{\sim}{X}\hat{\underset{\sim}{\xi}}\underset{\sim}{G})
\end{aligned}
\tag{1.5.7}
$$

which facilitates the comparison with the LS and the NML estimators.

In the case where the experimental design is such that the n experimental units can be classified into m groups each assumed to have its own growth curve and common scatter, they proposed alternative M-estimators of $\underset{\sim}{\xi}$ and $\underset{\sim}{\Sigma}$ by applying Maronna's proposal to each group separately and then pooling the results. Letting n_ℓ, $\tilde{\underset{\sim}{\mu}}_\ell$ and $\tilde{\underset{\sim}{\Sigma}}_\ell$ denote respectively the number of units, the location M-estimate and the scatter M-estimate in the ℓ^{th} group, $(\ell=1,\ldots m)$ we can express these alternative M-estimators as solutions to:

$$
\begin{aligned}
&\tilde{\underset{\sim}{\xi}} = (\underset{\sim}{X}^T\underset{\sim}{U}\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{U}\underset{\sim}{Y}\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}^T(\underset{\sim}{G}\underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}^T)^{-1} \\
&\tilde{\underset{\sim}{\Sigma}} = (n-r)^{-1}\sum_{\ell=1}^{r}(n_\ell-1)\tilde{\underset{\sim}{\Sigma}}_\ell
\end{aligned}
\tag{1.5.8}
$$

where $U = \text{diag}\{u_{11},\ldots u_{m,n_m}\}$, $u_{\ell k} = \{\sum_{k=1}^{n_\ell} u_1(d_{\ell k})\}^{-1} u_1(d_{\ell k})$, $d_{\ell k}^2 = (\underset{\sim}{y}_k^{(\ell)} - \tilde{\underset{\sim}{\mu}}_\ell)^T \tilde{\underset{\sim}{\Sigma}}_\ell^{-1} (\underset{\sim}{y}_k^{(\ell)} - \tilde{\underset{\sim}{\mu}}_\ell)$ and $\underset{\sim}{y}_k^{(\ell)}$ denotes the vector of observations on the k^{th} unit in the ℓ^{th} group, $(k=1,\ldots n_\ell;\ \ell=1,\ldots m)$. The major difference between $\tilde{\underset{\sim}{\xi}}$ and $\hat{\underset{\sim}{\xi}}$ is the way in which the distance argument of the weight function is computed: in $\tilde{\underset{\sim}{\xi}}$ it is based upon the distance between each observation and its group location vector; in $\hat{\underset{\sim}{\xi}}$ it is based upon information from all the n subjects.

These authors showed that if $n_\ell/n \to \lambda_\ell$ as $n \to \infty$, where $0 < \lambda_\ell < 1$, $(\ell=1,\ldots m)$, Maronna's results imply consistency and asymptotic normality of $\tilde{\underset{\sim}{\xi}}$. They conjecture that this also holds for $\hat{\underset{\sim}{\xi}}$ but a proof was not presented. They proposed to use the asymptotic distributions to test hypotheses about the elements of $\underset{\sim}{\xi}$; as in the Normal case, the test statistics correspond to functions of the characteristic roots of $\underset{\sim}{H}\,\underset{\sim}{E}^{-1}$ where $\underset{\sim}{H}$ and $\underset{\sim}{E}$ are matrices of sums of squares and cross products due to the hypothesis and error, respectively. They also suggested that a reasonable small sample approximation to the asymptotic test would be to use the Normal theory degrees of freedom (n-r-p+q) for the error term instead of the infinite degrees of freedom required in that case.

They present a numerical example but do not comment on the convergence of the algorithm used in the computation of the estimates.

CHAPTER II

M-METHODS IN THE STANDARD MULTIVARIATE LINEAR MODEL: ASYMPTOTIC PROPERTIES AND ROBUSTNESS

2.1 Introduction

In this chapter we are mainly concerned with the study of some asymptotic and robustness properties of M-methods under the SMLM.

As indicated in the literature review, most of the available papers on multivariate M-methods relate to the simple location/scatter problem and are based on Maronna's (1976) proposal which considers the assumption of elliptical symmetry for the error distribution. In many practical situations this assumption may be questionable and often difficult to verify. On the other hand, according to Maronna, departures from this class have a more serious effect on the robustness of the M-estimators than longtailedness within it. These facts make elliptical symmetry of the error distribution a rather restrictive assumption and justifies the search for methods that are robust under a more general class of underlying distributions.

In Section 2.2 we consider a Maronna-type M-estimator for the parameter matrix $\underset{\sim}{\beta}$ in the SMLM and suggest an alternative coordinatewise M-estimator which can be made robust even when the underlying distribution is not elliptically symmetric. We also summarize the assumptions required for the subsequent sections.

In Section 2.3 we derive the asymptotic distribution of the coordinatewise M-estimator by applying an asymptotic linearity result similar to that of Jurečková (1977). Although the proof follows the general guidelines of Jurečková's paper it requires neither her concordance-discordance assumption on the elements of the design matrix nor the scale 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 which essentially correspond to extensions of Wald's test and Sen's (1982) proposal to the multivariate case.

In Section 2.4 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 of the error distribution.

Finally in Section 2.5 we comment on the construction of robust M-estimators of both types by examining their influence functions and breakdown points.

2.2 Definition of the M-estimators and assumptions

Consider the SMLM given by (1.1.1) 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:

$$M_{nij}(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}}) = \sum_{k=1}^{n} u(d_k)x_{ki}(y_{kj}-\underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}_j) = 0 \qquad (2.2.1)$$

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

where $u(d)$ is a real-valued function defined for $d\geq 0$ and

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

In matrix notation (2.2.1) 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 Maronna-type M-estimator is defined by (2.2.1) with $\underset{\sim}{\Sigma}$ replaced by an estimate $\hat{\underset{\sim}{\Sigma}}$.

As discussed in section 1.4 we can view this class of M-estimators as a generalization of the class of ML estimators in the case of elliptically symmetric underlying distribution. Note that $u(d) = 1$ defines the NML estimator. Robust Maronna-type M-estimators can be obtained by a convenient choice of the function u.

An alternative approach within the context of M-estimation corresponds to the following generalization of the LS method. First put $\sigma_j = \sigma_{jj}^{1/2}$ and let:

(i) $\underset{\sim}{\gamma}_k = \{\mathrm{diag}(\sigma_1^{-1},\ldots\sigma_p^{-1})\}(\underset{\sim}{y}_k-\underset{\sim}{\beta}^T\underset{\sim}{x}_k)$, (k-1,...n)

(ii) $\underset{\sim}{h}$: $\mathbb{R}^p \to \mathbb{R}^p$ be a vector valued function

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

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

$$Q(\underset{\sim}{\beta}) = \sum_{k=1}^{n} \underset{\sim}{h}^T(\underset{\sim}{\gamma}_k)\underset{\sim}{\Gamma}^{-1}\underset{\sim}{h}(\underset{\sim}{\gamma}_k) \tag{2.2.2}$$

or, equivalently, as a solution to:

$$(\partial/\partial\underset{\sim}{\beta})Q(\underset{\sim}{\beta}) = \underset{\sim}{0} \tag{2.2.3}$$

For each coordinate h_j (j=1,...p) of $\underset{\sim}{h}$ define $\rho_j(x) = h_j^2(x)$ and $\psi_j(x) = (\partial/\partial x)\rho_j(x)$. Then note that since $\underset{\sim}{\Gamma}$ is p.d. there exists a matrix $\underset{\sim}{\Delta} = < \delta_{ij} >$ such that $\underset{\sim}{\Gamma}^{-1} = \underset{\sim}{\Delta}^T\underset{\sim}{\Delta}$. Therefore (2.2.2) can be can be expressed as:

$$Q(\underset{\sim}{\beta}) = \sum_{k=1}^{n} \sum_{\ell=1}^{p} \{ \sum_{m=1}^{p} \delta_{\ell m} h_m(\gamma_{km}) \}^2 \tag{2.2.4}$$

which implies that (2.2.3) can be written as:

$$\sum_{k=1}^{n} x_{ki} h_j(\gamma_{kj}) h_j'(\gamma_{kj}) \sum_{\ell=1}^{m} \delta_{\ell i} = 0, \tag{2.2.5}$$

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

since:

$$(\partial/\partial\beta_{ij})\gamma_{km} = (\partial/\partial\beta_{ij})\{(y_{km} - \sum_{g=1}^{r} x_{kg}\beta_{gm})/\sigma_m\} = \begin{cases} -x_{ij}/\sigma_j & \text{if } g=i,\ m=j \\ 0 & \text{otherwise.} \end{cases}$$

Finally note that from the definition of ψ_j, (2.2.5) is equivalent to:

$$M_{nij}(\underset{\sim}{Y},\underset{\sim}{\sigma},\hat{\underset{\sim}{\beta}}) = \sum_{k=1}^{n} x_{ki}\psi_j\{(y_{kj} - \underset{\sim}{x}_k^T \underset{\sim}{\beta}_j)/\sigma_j\} = 0 \tag{2.2.6}$$

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

where $\underset{\sim}{\sigma} = (\sigma_1^2,\ldots\sigma_p^2)^T$. 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 M-estimator by (2.2.6) with $\underset{\sim}{\sigma}$ replaced by an estimate $\hat{\underset{\sim}{\sigma}}$.

Note that solving (2.2.6) 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 LS (or NML) estimator. Alternatively a robust estimator can be obtained by defining:

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

where k_j is a positive constant; this corresponds to Huber's proposal for $\psi_j(=\psi)$. In such a case the generalized LS method would de-emphasize outliers or extreme observations coordinatewise, as opposed to the corresponding Maronna-type method with $u(d) = \psi(d)/d$, which de-emphasizes outliers or extreme observations 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 feature could be useful in the case where some coordinates are known to produce more outliers than others. In this work for reasons of notational ease, we shall consider the same score function ψ for each coordinate.

Also, in order to compute the estimate of $\underset{\sim}{\beta}$ we do not need to estimate the entire scatter 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 or extreme observations as in Huber's suggestion, their effect on the estimates of parameters related to coordinates where they are not present should be smaller in the coordinatewise case than in Maronna's case. Since each column of the coordinatewise M-estimator $\hat{\underset{\sim}{\beta}}$ is computed independently, outliers or extreme observations in one coordinate will only influence elements of $\hat{\underset{\sim}{\beta}}$ corresponding to other coordinates through the scatter 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 required in the subsequent sections.

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

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

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

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

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

A4. F is elliptically symmetric, i.e. its density function is given by (1.4.5).

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

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

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

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

(i) $\int\psi(\varepsilon/\sigma_j)f_j(\varepsilon) = 0$

(ii) $\rho_j^2 = \int\psi^2(\varepsilon/\sigma_j)f_j(\varepsilon) < \infty$

(iii) $w_j = -\int\psi(\varepsilon/\sigma_j)f_j'(\varepsilon)d\varepsilon = \sigma_j^{-1}\int\psi'(\varepsilon/\sigma_j)f_j(\varepsilon)d\varepsilon < \infty$

where f_j is the marginal density corresponding to the j^{th} coordinate.

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

Note that (i) in C1 is satisfied when ψ is skew-symmetric and F is symmetric which is the case usually considered in the literature.

2.3 Asymptotic properties of the coordinatewise M-estimator

In this section we derive the asymptotic distribution of the M-estimator defined by (2.2.6) and of some tests of hypotheses about the parameter matrix $\underset{\sim}{\beta}$ under both the null hypothesis and local Pitman-type alternatives.

First we consider the extention of two results due to Jurečková (1977). One is related to the asymptotic linearization of $\underset{\sim}{M}_n(\underset{\sim}{Y},\hat{\underset{\sim}{\sigma}},\hat{\underset{\sim}{\beta}})$ in a neighbourhood of the true value of the parameter matrix $\underset{\sim}{\beta}$ and is given Theorem 2.3.1. The other concerns the boundedness in probability of $n^{\frac{1}{2}}\|\hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*\|$ and is presented in Theorem 2.3.2. These two results are then used in the proof of Theorem 2.3.3 to obtain the required asymptotic distribution.

Before the presentation of the main theorems of this section we prove the following lemma which essentially corresponds to Jurečková's (1977) Theorem 4.1, though our assumptions are less stringent than hers.

It basically corresponds to a univariate version of Theorem 2.3.1 in the case where the scale parameter is known. For ease of notation we drop the variate-identifying subscript.

Lemma 2.3.1: Under assumptions A1-A3, B and C1 it follows that for all $K > 0$, $\varepsilon > 0$:

$$P\{\sup n^{-\frac{1}{2}} \| \underset{\sim}{M}_n(\underset{\sim}{y},\sigma^2,\hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n(\underset{\sim}{y},\sigma^2,\underset{\sim}{\beta}) + nw\underset{\sim}{V}(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}) \| \geq \varepsilon :$$

$$n^{\frac{1}{2}} \| \hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta} \| \leq K\} \to 0 \text{ as } n \to \infty \qquad (2.3.1)$$

Proof: The proof will be outlined in four steps, only the first of which is significantly different from the corresponding step in Jurečková (1977, Theorem 4.1).

Step 1: 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, $(i,h=1,\ldots r)$, let:

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

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

and observe that:

$$\sum_{k \in S_{ih}(0)} x_{ki}\psi\{(y_k - \sum_{j \neq h} x_{kj}\hat{\beta}_j - x_{kh}\hat{\beta}_h)/\sigma\}$$

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

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:

$$S_{ig} = \bigcap_{h=1}^{p} S_{ih}(\delta_{hg})$$

where δ_{hg} is either 0 or 1.

Then we may write:

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

where

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

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

Therefore all we have to show is that for all $K > 0$, $\varepsilon > 0$, (i=1,...r; g=1,...2^r):

$$P\{\sup n^{-\frac{1}{2}}|M_{nig}(\underset{\sim}{y},\sigma^2,\hat{\beta}) - M_{nig}(\underset{\sim}{y},\sigma^2,\beta) + nw\, \underset{\sim}{v}_{ig}^T(\hat{\beta}-\beta)| \geq \varepsilon :$$

$$n^{\frac{1}{2}}\|\hat{\beta}-\beta\| \leq K\} \to 0 \text{ as } n \to \infty . \tag{2.3.2}$$

where $v_{ig} = (v_{i1g},\ldots v_{irg})^T$ and $v_{ihg} = \lim_{n\to\infty} \{n^{-1} \sum_{k\epsilon S_{ig}} x_{ki}x_{kh}\}$, (h=1,...r). Note that $v_{ihg} < \infty$ by assumption B.

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

To simplify notation in the following steps we shall fix $h (1 \le h \le r)$ and prove (2.3.2) for all $\hat{\beta} \in \Theta_h$, where $\Theta_h = \{t \in \mathbb{R}^r: n^{1/2} \| t - \beta \| \le K, t_j = 0 \ \forall j \ne h\}$. Then, using the concept of contiguity as in Hájek and Šidák (1967, ch.6) we extend the result to the more general case $\hat{\beta} \in \Theta$ where $\Theta = \{t \in \mathbb{R}^r : n^{1/2} \| t - \beta \| \le K\}$. Furthermore, with no loss of generality we set $\beta = 0$ and $\sigma = 1$.

Step 2: In this step we obtain the asymptotic distribution of $n^{-1/2} M_{nig}(y, 1, \hat{\beta})$ under the sequences of probability measures p_{nig} and q_{nig} defined by:

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

$$q_{nig} = \begin{cases} 1 & , \text{ if } S_{ig} = \emptyset \\ \prod_{k \in S_{ig}} f(y_k - x_{kh}\beta_h), & \text{ otherwise} \end{cases}$$

Consider the statistic:

$$T_{nig} = - \sum_{k \in S_{ig}} x_{kh} \hat{\beta}_h f'(y_k)/f(y_k)$$

From assumptions A1-A3 and B and Theorem V.1.2 of Hájek and Šidák (1967, ch.5) it follows that under p_{nig}, $T_{nig} \approx N(0, b_{ig}^2)$ where:

$$b_{ig}^2 = \lambda_h^2 v_{hhg} I(f) \quad (< \infty)$$

and λ_h is a positive constant. Note that for every $K > 0$ we can always write $\beta_h = n^{-1/2}\lambda_h$ for some $\lambda_h > 0$ since we assume $n^{1/2} \| \hat{\beta} - 0 \| \le K$.

Now for $(i=1,\ldots r;\ g=1,\ldots 2^r)$ let:

$$\log L_{nig} = \sum_{k \in S_{ig}} \log\{f(y_k - x_{kh}\hat{\beta}_h)/f(y_k)\}$$

and observe that by Theorem VI.2.1 of Hájek and Šidák (1967, ch.6) it follows that *under* p_{nig} :

(a) $\log L_{nig} - T_{nig} + b_{ig}^2/2 \xrightarrow{P} 0$

(b) $\log L_{nig} \approx N(-b_{ig}^2/2,\ b_{ig}^2)$ (2.3.3)

(c) q_{nig} and p_{nig} are contiguous.

Applying the Cramér-Wold theorem and Theorem V.1.2 of Hájek and Šidák (1967, ch.5) it is easy to show that *under* p_{nig} we have:

$$\{T_{nig}, n^{-\frac{1}{2}}M_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}})\}^T \approx N_2(\underset{\sim}{0},\underset{\sim}{\Gamma}) \qquad (2.3.4)$$

where $\underset{\sim}{\Gamma} = \begin{bmatrix} b_{ig}^2 & -\lambda_h w v_{ihg} \\ -\lambda_h w v_{ihg} & \rho^2 v_{iig} \end{bmatrix}$

This result together with (2.3.3.a) imply that: $\{\log L_{nig},\ n^{-\frac{1}{2}}M_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}})\}^T \approx N_2\{(-b_{ig}^2/2,0)^T,\ \underset{\sim}{\Gamma}\}$. Finally, from Lemma VI.1.4 of Hájek and Šidák (1967, ch.6) we may conclude that *under* q_{nig}:

$$n^{-\frac{1}{2}}M_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}}) \approx N(-\lambda_h w v_{ihg},\ \rho^2 v_{iig}) \qquad (2.3.5)$$

<u>Step 3</u>: In this step we show that (2.3.2) holds for any *fixed* $\hat{\underset{\sim}{\beta}} \in \Theta_h$. First replace the score function ψ by the bounded function $\psi^{(m)}$, $(m \geq 2)$, defined by:

$$\psi^{(m)}(x) = \xi^{(m)}\{F(x)\}\ ,\ x \in \mathbb{R}$$

where:

$$\xi^{(m)}(s) = \begin{cases} \xi(1/m) & \text{if } 0<s<1/m \\ \xi(s) & \text{if } 1/m \leq s \leq 1-1/m \\ \xi(1-1/m) & \text{if } 1-1/m<s<1 \end{cases}$$

and $\xi(s) = \psi\{F^{-1}(s)\}$, $0 < s < 1$.

Let $M^{(m)}_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}}) = \sum_{k\in S_{ig}} x_{ki}\psi^{(m)}(y_k - x_{kh}\hat{\beta}_h)$ and note that for a fixed m we have:

$$\mathrm{Var}[n^{-\frac{1}{2}}\{M^{(m)}_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}}) - M^{(m)}_{nig}(\underset{\sim}{y},1,\underset{\sim}{0})\}] \leq$$

$$\leq n^{-1}\sum_{k\in S_{ig}} x^2_{ki}\int\{\psi^{(m)}(y_k - x_{kh}\hat{\beta}_h) - \psi^{(m)}(y_k)\}^2 dF(y_k) \to 0$$

as $n \to \infty$

by Lebesgue's Dominated Convergence Theorem. Applying Chebyshev's inequality it follows that given $\varepsilon > 0$:

$$P\{n^{-\frac{1}{2}}|M^{(m)}_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}}) - M^{(m)}_{nig}(\underset{\sim}{y},1,\underset{\sim}{0}) - E\, M^{(m)}_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}})| \geq \varepsilon\} \to 0 \text{ as } m\to\infty \quad (2.3.6)$$

Now observe that:

$$n^{-1}E\{M_{nig}(\underset{\sim}{y},1,\underset{\sim}{0}) - M^{(m)}_{nig}(\underset{\sim}{y},1,\underset{\sim}{0})\}^2 =$$

$$= n^{-1}\sum_{k\in S_{ig}} x^2_{ki}\int_0^1\{\xi(s) - \xi^{(m)}(s)\}^2 ds \to 0 \text{ as } n \to \infty$$

uniformly in n by Lebesgue's Monotone Convergence Theorem. Applying Chebyshev's inequality it follows that given $\eta > 0$:

$$P\{n^{-\frac{1}{2}}|M_{nig}(\underset{\sim}{y},1,\underset{\sim}{0}) - M^{(m)}_{nig}(\underset{\sim}{y},1,\underset{\sim}{0})| \geq \eta\} \to 0 \text{ as } m \to \infty \quad (2.3.7)$$

holds uniformly in n. Now from the contiguity of q_{nig} and p_{nig} and Lemma 3.5 of Jurečková (1969) it follows that given $\eta > 0$:

$$P\{n^{-\frac{1}{2}}|M_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}}) - M^{(m)}_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}})| \geq \eta : \hat{\underset{\sim}{\beta}}\in\Theta_h\} \to 0 \text{ as } m\to\infty,\ n\to\infty \quad (2.3.8)$$

Then from (2.3.5)-(2.3.8) we may conclude that:

$$P\{n^{-\frac{1}{2}}|M_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}}) - M_{nig}(\underset{\sim}{y},1,\underset{\sim}{0})+nw\ v_{ihg}\hat{\beta}_h| \geq \varepsilon:\ \hat{\underset{\sim}{\beta}}\epsilon\Theta_h\} \to 0 \qquad (2.3.9)$$

as $n \to \infty$

holds for $(i=1,\ldots r;\ g=1,\ldots 2^r)$.

Step 4: In this step we show that the linear approximation of (2.3.9) is uniform in $\hat{\underset{\sim}{\beta}}\epsilon\Theta_h$.

First we consider a partition $\{a_0,a_1,\ldots a_m\}$ of $[-K,K]$ such that $-K = a_0 < a_1 < \ldots < a_m = K$ and $|a_\ell - a_{\ell-1}| \leq \varepsilon/(2M|w|)$ $(\ell=1,\ldots m)$ where $M = \max\limits_{\substack{1\leq i\leq r\\ 1\leq g\leq 2^r}} v_{ihg}$. By the monotonicity of $M_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}})$ in each element of $\hat{\underset{\sim}{\beta}}$ when the others are held fixed we can write:

$$\sup_{\hat{\underset{\sim}{\beta}}\,\epsilon\,\Theta_h} n^{-\frac{1}{2}}|M_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}})-M_{nig}(\underset{\sim}{y},1,\underset{\sim}{0})+nw\ v_{ihg}\hat{\beta}_h| \leq$$

$$\leq \max_{\ell=0,\ldots m} n^{-\frac{1}{2}}|M_{nig}(\underset{\sim}{y},1,\hat{\underset{\sim}{\beta}}^{(\ell)})-M_{nig}(\underset{\sim}{y},1,\underset{\sim}{0})+nw\ v_{ihg}\hat{\beta}_h^{(\ell)}|+\varepsilon/2$$

where $\hat{\beta}_h^{(\ell)} = n^{-\frac{1}{2}}a_\ell$ and $\hat{\beta}_j^{(\ell)} = 0$ for all $j \neq h$. This enables us to substitute the operation of taking the sumpremum over Θ_h by that of taking the maximum over a finite number of points in Θ_h and the desired result follows.

Finally we note that (2.3.2) follows from the contiguity of p_{nig} and q_{nig} defined in terms of $\hat{\underset{\sim}{\beta}}\epsilon\Theta$ instead of $\hat{\underset{\sim}{\beta}}\epsilon\Theta_h$.

Theorem 2.3.1: Under assumptions A1-A3, B and C1 it follows that for all $K > 0$, $L > 0$ and $\varepsilon > 0$:

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

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

where $\underset{\sim}{W} = \text{diag}\{w_1,\ldots w_p\}$, $w_j = \sigma_j^{-1}\int\psi'(\varepsilon/\sigma_j)f_j(\varepsilon)d(\varepsilon)$ (2.3.10)

Proof: Since 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. As in the previous theorem we drop the variate-identifying subscript for the ease of notation.

Consider the decomposition:

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

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

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

By applying Lemma 2.3.1 to the first term all we have to show is that for all $L > 0$ and $\varepsilon > 0$:

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

This can be done in lines very similar to the proof of Lemma 2.3.1. We outline the major modifications.

Step 1: Partition each of the sets $S_{ih}(\delta)$, $(i=1,\ldots r;\ \delta=0,1)$ defined in Step 1 of Lemma 2.3.1 into two disjoint sets, one of which contains those k's such that $\text{sign}(x_{ki}) = \text{sign}(y_k-\underset{\sim}{x}_k^T\underset{\sim}{\beta})$ and the other to its complement. Then observe that:

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

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 Lemma 2.3.1, we may write

$M_{ni}(\underset{\sim}{y},\hat{\sigma},\hat{\underset{\sim}{\beta}})$ as the sum of a finite number of functions which are monotone in each element of $(\hat{\underset{\sim}{\beta}}^T,\hat{\sigma})$ when the others are held fixed.

Step 2: 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. Then define for $(i=1,\dots r;\ g=1,\dots 2^{r+1})$:

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

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

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

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

and proceed as in Step 2 of Lemma 2.3.1 using Hájek and Šidák's (1967, ch.6) contiguity results for scale alternatives.

Steps 3 and 4: Essentially the same as in Lemma 2.3.1.

Theorem 2.3.2: Under the assumptions A1-A3, B and C1 it follows that given $\eta > 0$:

$$P\{\min n^{-\frac{1}{2}}\|\underset{\sim}{M}_n^*(\underset{\sim}{Y},\hat{\underset{\sim}{g}},\hat{\underset{\sim}{\beta}})\| > \eta : n^{\frac{1}{2}}\|\hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*\| \le K, n^{\frac{1}{2}}\|\hat{\underset{\sim}{g}}-\underset{\sim}{g}\| \le L\} \to 0$$

holds for all $K > 0$ and $L > 0$.

Proof: First observe that the multivariate result follows by applying the corresponding univariate result to each coordinate separately. Now recall that in Jurečková's (1977) notation, $\underset{\sim}{X}_N$, c_{ji}, $\underset{\sim}{\Delta}^\circ$ 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 case. Then it is easy to verify that the proof of her Lemma 5.2 also goes through in the more general case considered here (where σ is replaced by a $n^{\frac{1}{2}}$-consistent estimate $\hat{\sigma}$) by making:

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

Essentially we are showing that given $\eta > 0$, $\min n^{-\frac{1}{2}} \| \underset{\sim}{M}^*_n(\underset{\sim}{Y},\hat{\underset{\sim}{\sigma}},\hat{\underset{\sim}{\beta}}) \|$ converges in probability to zero in 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.

Theorem 2.3.3: Under assumptions A1-A3, B and C1 it follows that:

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

where

$$\underset{\sim}{\Phi} = < \phi_{ij} >, \ \phi_{ij} = \int \psi(\varepsilon_i/\sigma_i)\psi(\varepsilon_j/\sigma_j) f(\underset{\sim}{\varepsilon}) d\underset{\sim}{\varepsilon} , \quad (i,j=1,\ldots p) \qquad (2.3.11)$$

and $\underset{\sim}{W}$ is given by (2.3.10).

Proof: From Theorem 2.3.2 and 2.3.3 it follows that $n^{\frac{1}{2}}(\underset{\sim}{V} \otimes \underset{\sim}{W})(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)$ has the same asymptotic distribution as $n^{-\frac{1}{2}}\underset{\sim}{M}_n(\underset{\sim}{Y},\underset{\sim}{\sigma},\underset{\sim}{\beta})$. Now, by applying Cramér-Wold's theorem and Theorem V.1.2 of Hájek-Šidák (1967, ch.5) it can be shown that $n^{-\frac{1}{2}}\underset{\sim}{M}_n(\underset{\sim}{Y},\underset{\sim}{\sigma},\underset{\sim}{\beta}) \approx N_{rp}(\underset{\sim}{0}, \underset{\sim}{V} \otimes \underset{\sim}{\Phi})$ and the result follows.

The asymptotic distribution derived above can be used to construct tests of hypotheses of the form:

$$H : \underset{\sim}{C}\underset{\sim}{\beta}\underset{\sim}{U} = \underset{\sim}{K} \qquad (2.3.12)$$

where $\underset{\sim}{C}(c\times r)$ and $\underset{\sim}{U}(p\times u)$ are known matrices of full row and column ranks $c(\le r)$ and $u(\le p)$ respectively and $\underset{\sim}{K}(c\times u)$ is any known matrix. By making a linear transformation and a convenient reparametrization, testing (2.3.12) is equivalent to testing:

$$H_0 : \underset{\sim}{\eta} = \underset{\sim}{0} \tag{2.3.13}$$

under the following model:

$$\underset{\sim}{Z} = [\underset{\sim}{A}\ \underset{\sim}{Q}]\begin{bmatrix}\underset{\sim}{\xi}\\ \underset{\sim}{\eta}\end{bmatrix} + \underset{\sim}{\zeta} = \underset{\sim}{D}\underset{\sim}{\alpha} + \underset{\sim}{\zeta} \tag{2.3.14}$$

where $\underset{\sim}{Z} = \underset{\sim}{Y}\underset{\sim}{U}$, $\underset{\sim}{\zeta} = \underset{\sim}{\varepsilon}\underset{\sim}{U}$, $\underset{\sim}{A}(n\times r\text{-}c)$ and $\underset{\sim}{Q}(n\times c)$ are known matrices, $\underset{\sim}{\xi}(r\text{-}c\times u)$ and $\underset{\sim}{\eta}(c\times u)$ 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 density function of the transformed (u-dimensional) r.v.'s and let $\underset{\sim}{W}_H$ and $\underset{\sim}{L}_H$ respectively correspond to (2.3.10) and (2.3.11) defined in terms of H. Note that the linear transformation and 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:

$$H_n : \underset{\sim}{\eta}_n = \underset{\sim}{\eta} = n^{-\frac{1}{2}}\underset{\sim}{\Delta} \ . \tag{2.3.15}$$

where $\underset{\sim}{\Delta}\ (\ne\underset{\sim}{0})$ is a fixed $(c\times u)$ matrix.

From assumptions of 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_0. This contiguity implies that the result of Theorems 2.3.1 and 2.3.2 also hold under H_n. Then letting:

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

we can proceed as in Theorem 2.3.3 to show that

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

where $\underset{\sim}{P} = \underset{\sim}{R}_{22}-\underset{\sim}{R}_{21}\underset{\sim}{R}_{11}^{-1}\underset{\sim}{R}_{12}$ and $\underset{\sim}{\Delta}_1 = \underset{\sim}{0}$ if H_0 holds or $\underset{\sim}{\Delta}_1 = \underset{\sim}{\Delta}$ if H_n holds.

From well known results related to the Wishart distribution, as in Arnold (1981, ch.17) for example, we get:

$$\underset{\sim}{H} = \hat{\underset{\sim}{\eta}}^T\{\underset{\sim}{Q}^T\underset{\sim}{Q}-\underset{\sim}{Q}^T\underset{\sim}{A}(\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\underset{\sim}{A}^T\underset{\sim}{Q}\}\ \hat{\underset{\sim}{\eta}} \approx W_u(c,\underset{\sim}{W}_H^{-1}\underset{\sim}{\Sigma}_H\underset{\sim}{W}_H^{-1}, \underset{\sim}{\Omega}_1) \tag{2.3.16}$$

where $\underset{\sim}{\Omega}_1 = \underset{\sim}{\Delta}_1^T\underset{\sim}{P}\underset{\sim}{\Delta}_1$.

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:

$$\underset{\sim}{E} = n\hat{\underset{\sim}{W}}_H^{-1}\hat{\underset{\sim}{\Sigma}}_H\hat{\underset{\sim}{W}}_H^{-1} \tag{2.3.17}$$

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

We indicate below the asymptotic distributions of the most common test statistics, i.e. $\mathrm{tr}(\underset{\sim}{H}\underset{\sim}{E}^{-1})$ (Lawley-Hotelling's trace), $|\underset{\sim}{H}+\underset{\sim}{E}|/|\underset{\sim}{E}|$ (Wilks' likelihood ratio) and $\mathrm{ch}_1(\underset{\sim}{H}\underset{\sim}{E}^{-1})$ (Roy's largest root).

First note that from (2.3.16), (2.3.17) and Slutsky's theorem we obtain:

$$n\underset{\sim}{E}^{-\frac{1}{2}}\underset{\sim}{H}\ \underset{\sim}{E}^{-\frac{1}{2}} \approx W_u(c,\underset{\sim}{I},\underset{\sim}{\Omega}_E)$$

where:

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

Then, following Sverdrup (1952) we can show that: $n\ \mathrm{tr}(\underset{\sim}{H}\underset{\sim}{E}^{-1}) = n\ \mathrm{tr}(\underset{\sim}{E}^{-\frac{1}{2}}\underset{\sim}{H}\underset{\sim}{E}^{-\frac{1}{2}}) \xrightarrow{\mathcal{D}} \mathrm{tr}\ \underset{\sim}{W}$ where $\underset{\sim}{W} \sim \underset{\sim}{W}_u(c,\underset{\sim}{I},\underset{\sim}{\Omega}_E)$ and therefore it follows that:

$$n\ \mathrm{tr}(\underset{\sim}{H}\underset{\sim}{E}^{-1}) \approx \chi^2_{cu}(\mathrm{tr}\ \underset{\sim}{\Omega}_E) \qquad (2.3.18)$$

Now let $\underset{\sim}{A}$ and $\underset{\sim}{B}$ be (u×u) p.d. finite matrices and consider the expansion:

$$\log|\underset{\sim}{A}+n^{-1}\underset{\sim}{B}| = \log|\underset{\sim}{A}| + \log\{1 + n^{-1}\sum_{i=1}^{u}\sum_{j=1}^{u} b_{ij}a^{ij} + 0(n^{-2})\}$$

where a^{ij} is the $(i,j)^{th}$ element of $\underset{\sim}{A}^{-1}$. This implies:

$$n\log(|\underset{\sim}{A} + n^{-1}\underset{\sim}{B}|/|\underset{\sim}{A}|) = n\log(|n\underset{\sim}{A} + \underset{\sim}{B}|/|n\underset{\sim}{A}|) = \mathrm{tr}\ (\underset{\sim}{B}\underset{\sim}{A}^{-1}) + o(n^{-1})$$

Then, making $n\underset{\sim}{A} = \underset{\sim}{E}$ and $\underset{\sim}{B} = \underset{\sim}{H}$ it follows that:

$$n\log(|\underset{\sim}{E}+\underset{\sim}{H}|/|\underset{\sim}{E}|) = n\ \mathrm{tr}(\underset{\sim}{H}\underset{\sim}{E}^{-1}) + o(n^{-1})$$

and we conclude from (2.3.18) that:

$$n\log(|\underset{\sim}{E}+\underset{\sim}{H}|/|\underset{\sim}{E}|) \approx \chi^2_{cu}(\mathrm{tr}\ \underset{\sim}{\Omega}_E) \qquad (2.3.19)$$

To obtain the asymptotic null distribution of $ch_1(\underset{\sim}{H}\underset{\sim}{E}^{-1})$ it is sufficient to note that by Sverdrup (1952) we have:

$$n\ ch_1(\underset{\sim}{H}\underset{\sim}{E}^{-1}) \approx ch_1(\underset{\sim}{W}) \qquad (2.3.20)$$

where $\underset{\sim}{W} \sim W_u(c,\underset{\sim}{I},\cdot)$. Tables for the upper percentage points of the distribution of $ch_1(\underset{\sim}{W})$ are available, for example, in Pearson and Hartley (1976, Table 51).

In practice it may be more appropriate to use n-r instead of n as a multiplying factor in (2.3.18) and (2.3.20) and Bartlett's approximation, m=n-r(p-c+1)/2 instead of n as the multiplying factor in (2.3.19).

The tests above are analogues of Wald's tests in the classical setup. We now present an extension of Sen's (1982) approach which produces tests equivalent to the classical χ^2 tests in which estimates under the more constrained hypothesis alone are required. They 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 $\underset{\sim}{M}_n(\underset{\sim}{Z},\hat{\underset{\sim}{\sigma}},\hat{\underset{\sim}{\alpha}})$. Then let $\tilde{\underset{\sim}{\xi}}$ be a coordinatewise M-estimate 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 $\underset{\sim}{M}_{n2} = \underset{\sim}{M}_{n2}(\tilde{\underset{\sim}{\xi}},\underset{\sim}{0})$. Now observe that:

$$n^{-\frac{1}{2}}\tilde{\underset{\sim}{M}}^*_{n2} = n^{-\frac{1}{2}}[\underset{\sim}{M}^*_{n2}\{\underset{\sim}{\xi}+(\tilde{\underset{\sim}{\xi}}-\underset{\sim}{\xi}),\underset{\sim}{0}\}-\underset{\sim}{M}^*_{n2}(\underset{\sim}{\xi},\underset{\sim}{0})+n(\underset{\sim}{R}_{21}\otimes\underset{\sim}{W}_H)(\tilde{\underset{\sim}{\xi}}^*-\underset{\sim}{\xi}^*)] +$$

$$+ n^{-\frac{1}{2}}[\underset{\sim}{M}^*_{n2}(\underset{\sim}{\xi},\underset{\sim}{0})-n(\underset{\sim}{R}_{21}\otimes\underset{\sim}{W}_H)(\tilde{\underset{\sim}{\xi}}^*-\underset{\sim}{\xi}^*)] \qquad (2.3.21)$$

From the contiguity mentioned above it follows by Theorem 2.3.1 that the first term in (2.3.21) converges in probability to zero under H_n as well as H_0; furthermore the same theorem implies that $n^{\frac{1}{2}}(\tilde{\underset{\sim}{\xi}}^*-\underset{\sim}{\xi}^*)$ has the same asymptotic distribution as $(\underset{\sim}{R}_{11}^{-1}\otimes \underset{\sim}{W}_H^{-1})n^{-\frac{1}{2}}\underset{\sim}{M}^*_{n1}(\underset{\sim}{\xi},\underset{\sim}{0})$. Therefore we obtain:

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

and proceeding as in Theorem 2.3.3 we may show that:

$$n^{-\frac{1}{2}}\tilde{\underset{\sim}{M}}^*_{n2} \approx N_{cu}(\underset{\sim}{\Delta}^*_2,\underset{\sim}{P}\otimes\underset{\sim}{\Sigma}_H)$$

where $\underset{\sim}{P} = \underset{\sim}{R}_{22}-\underset{\sim}{R}_{21}\underset{\sim}{R}_{11}^{-1}\underset{\sim}{R}_{12}$ and $\underset{\sim}{\Delta}^*_2 = \underset{\sim}{0}^*$ if H_0 holds or $\underset{\sim}{\Delta}^*_2 = (\underset{\sim}{P}\otimes\underset{\sim}{W}_H)\underset{\sim}{\Delta}^*$ if H_n holds. Consequently we get:

$$\underset{\sim}{H} = \tilde{\underset{\sim}{M}}^T_{n2}\{\underset{\sim}{Q}^T\underset{\sim}{Q}-\underset{\sim}{Q}^T\underset{\sim}{A}(\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\underset{\sim}{A}^T\underset{\sim}{Q}\}^{-1}\tilde{\underset{\sim}{M}}_{n2} \approx W_u(c,\underset{\sim}{\Sigma}_H,\underset{\sim}{\Omega}_2)$$

where

$$\underset{\sim}{\Omega}_2 = \underset{\sim}{\Delta}_2^T\underset{\sim}{P}^{-1}\underset{\sim}{\Delta}_2$$

Taking $\underset{\sim}{E} = n\hat{\underset{\sim}{\Sigma}}_H$ it is easy to show that the asymptotic distributions of the usual test statistics are the same as those derived through the previous approach.

2.4 Asymptotic properties of the Maronna-type M-estimator

In this section we derive the asymptotic distribution of the Maronna-type M-estimator defined by (2.2.1). We also propose some tests of hypotheses about the parameter matrix $\underset{\sim}{\beta}$.

Basically the same approach of section 2.3 can be employed in this case; since the proofs are very similar to those of that section we only indicate the major changes.

First we present three lemmas which are needed in the derivation of the main results of this section: Lemma 2.4.1 is the analogue of Lemma 2.3.1; Lemmas 2.4.2 and 2.4.3 correspond to the scale alternatives counterparts of Lemmas 4.1 and 4.2 of Patel (1973). The basic asymptotic linearity result is presented in Theorem 2.4.1 and the asymptotic distribution of the Maronna-type M-estimator is derived in Theorem 2.4.2.

Lemma 2.4.1: Under assumptions A1-A4, B and C2 it follows that for all $K > 0$, $\varepsilon > 0$:

$$P\{\sup n^{-\frac{1}{2}}||\underset{\sim}{M}_n^*(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}})-\underset{\sim}{M}_n^*(\underset{\sim}{Y},\underset{\sim}{\Sigma},\underset{\sim}{\beta})+n(\underset{\sim}{V}\otimes b_0\underset{\sim}{I})(\hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*)||\geq\varepsilon :$$

$$n^{\frac{1}{2}}||\hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*||\leq K\} \to 0 \text{ as } n\to\infty , \tag{2.4.1}$$

where $$b_0 = E_F\{p^{-1}\psi'(d_k)+(1-p^{-1})\psi(d_k)/d_k\} \tag{2.4.2}$$

Proof: The proof is presented in four steps.

Step 1: Let $\underset{\sim}{\Sigma} = \underset{\sim}{I}$ without loss of generality and define the sets $S_{ig}(i=1,\ldots r;\ g=1,\ldots 2^r)$ as in Lemma 2.3.1. Then write:

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

where:

$$M_{hijg}(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}}) = \sum_{k\epsilon S_{ig}} x_{ki}(y_{kj}-\underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}_j)\psi(d_k)/d_k \qquad (2.4.3)$$

$$(i=1,\ldots r;\ j=1,\ldots p;\ g=1,\ldots 2^r)$$

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

$$(\partial/\partial\hat{\beta}_{hj})d_k = -x_{kh}(y_{kj}-\underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}_j)/d_k$$

which implies:

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

with probability 1. Therefore it follows that:

$$x_{ki}(y_{kj}-\underset{\sim}{x}_k^T\hat{\underset{\sim}{\beta}}_j)/d_k \text{ is } \begin{cases} \geq 0 \text{ and decreasing for } \hat{\beta}_{hj} < y_{kj}/x_{kh} \\ < 0 \text{ and decreasing otherwise .} \end{cases} \qquad (2.4.4)$$

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

$$\psi(d_k) \text{ is } \begin{cases} \geq 0 \text{ and nonincreasing for } \hat{\beta}_{hj} < y_{kj}/x_{kh} \\ \geq 0 \text{ and nondecreasing otherwise .} \end{cases} \qquad (2.4.5)$$

From (2.4.4) and (2.4.5), using the rule for the derivative of a product of two functions it follows that (2.4.3) is a nonincreasing function of $\hat{\beta}_{hj}$.

In a similar way we can show that (2.4.3) is a nondecreasing function of $\hat{\beta}_{hj}$ when g is such that $\text{sign}(x_{ki}) \neq \text{sign}(x_{kh})$ for all $k\epsilon S_{ig}$.

Step 2: Define p_{nig}, q_{nig}, T_{nijg} and $\log L_{nijg}$ in a similar way as in Lemma 2.3.1, replacing the univariate r.v.'s by their multivariate counterparts.

Note that the spherical symmetry assumed for F ($\underset{\sim}{\Sigma}=\underset{\sim}{I}$) implies that d_k and $d_k^{-1}\underset{\sim}{Y}_k$ are independent and that the latter is uniformly distributed on the unit spherical surface.

Using this fact and recalling that the results from Hájek and Šidák (1967, ch.6) also hold in the multivariate case, proceed as in Lemma 2.3.1 to show that *under* p_{nig}:

$$\{\log L_{nijg}, n^{-\frac{1}{2}}M_{nijg}(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}})\}^T \approx N_2\{(-b_{ig}^2/2,0)^T,\underset{\sim}{\Gamma}\}$$

where $\underset{\sim}{\Gamma} = \begin{bmatrix} b_{ig}^2 & -\lambda_{hj}v_{ihg}b_0 \\ -\lambda_{hj}v_{ihg}b_0 & a_0 v_{iig} \end{bmatrix}$

and

$$a_0 = E_F\{\psi^2(d_k)/p\} \tag{2.4.6}$$

Also, we can show that *under* q_{nig}:

$$n^{-\frac{1}{2}}M_{nijg}(\underset{\sim}{Y},\underset{\sim}{\Sigma},\hat{\underset{\sim}{\beta}}) \approx N(-\lambda_{hj}v_{ihg}b_0,\ a_0 v_{iig}) .$$

Steps 3 and 4: Essentially the same as in Lemma 2.3.1. Note that since we are assuming ψ bounded we do not need the truncation procedure as in that case. In the last step we choose the partition of the compact set $[-K,K]$ in such a way that $|a_\ell - a_{\ell-1}| \le \varepsilon/(2M|b_0|)$.

Now, without loss of generality let:

$$\underset{\sim}{S}_t = \hat{\underset{\sim}{\Sigma}}^{\frac{1}{2}} = \underset{\sim}{\Sigma}^{\frac{1}{2}} + t\underset{\sim}{\Delta}$$

where $\underset{\sim}{\Delta}$ is a matrix of constants and $t=n^{-\frac{1}{2}}$. Then define:

$$f_t(\underset{\sim}{\varepsilon}) = |\underset{\sim}{S}_t^2|^{-\frac{1}{2}} h(d_t)$$

where $d_t^2 = \underset{\sim}{\varepsilon}^T \underset{\sim}{S}_t^{-2} \underset{\sim}{\varepsilon}$ and let $f_t'(\underset{\sim}{\varepsilon}) = (\partial/\partial t) f_t(\underset{\sim}{\varepsilon})$, $s_t(\underset{\sim}{\varepsilon}) = \{f_t(\underset{\sim}{\varepsilon})\}^{\frac{1}{2}}$ and $s_t'(\underset{\sim}{\varepsilon}) = (\partial/\partial t) s_t(\underset{\sim}{\varepsilon})$.

Lemma 2.4.2: Under assumptions A1-A4 we have:

$$C = \int \left[\frac{s_t(\underset{\sim}{\varepsilon}) - s_0(\underset{\sim}{\varepsilon})}{t} - s_0'(\underset{\sim}{\varepsilon}) \right]^2 d\underset{\sim}{\varepsilon} \to 0 \text{ as } t \to 0 .$$

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

$$f_t'(\underset{\sim}{\varepsilon}) = -h(d_t) \operatorname{tr}(|\underset{\sim}{S}_t^2| \underset{\sim}{S}_t^{-2} \underset{\sim}{\Delta}) - h'(d_t) \operatorname{tr}(\underset{\sim}{S}_t^{-2} \underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}^T \underset{\sim}{S}_t^{-2} \underset{\sim}{\Delta})/d_t$$

and

$$f'(\underset{\sim}{\varepsilon}) = -h(d_0) \operatorname{tr}\underset{\sim}{\Delta} - h'(d_0) \underset{\sim}{\varepsilon}^T \underset{\sim}{\Delta} \underset{\sim}{\varepsilon}/d_0$$

which implies:

$$s_0'(\underset{\sim}{\varepsilon}) = -[h(d_0)\operatorname{tr}\underset{\sim}{\Delta}/\{h(d_0)\}^{\frac{1}{2}} - d_0^{-1} h'(d_0) \underset{\sim}{\varepsilon}^T \underset{\sim}{\Delta} \underset{\sim}{\varepsilon}/\{h(d_0)\}^{\frac{1}{2}}]/2 .$$

By assumption A4 it follows that:

$$\int \{s_0'(\underset{\sim}{\varepsilon})\}^2 d\underset{\sim}{\varepsilon} < \infty \tag{2.4.7}$$

and by the Cauchy-Schwarz inequality:

$$\{s_t(\underset{\sim}{\varepsilon}) - s_0(\underset{\sim}{\varepsilon})\}^2 = \{\int_0^t s_w'(\underset{\sim}{\varepsilon}) dw\}^2 \le \int_0^t \{s_w'(\underset{\sim}{\varepsilon})\}^2 dw \int_0^t dw = t \int_0^t \{s_w'(\underset{\sim}{\varepsilon})\}^2 dw$$

Then we may write:

$$\int\left\{\frac{s_t(\underset{\sim}{\varepsilon})-s_0(\underset{\sim}{\varepsilon})}{t}\right\}^2 d\underset{\sim}{\varepsilon} \le \int \frac{1}{t}\int_0^t\{s_w'(\underset{\sim}{\varepsilon})\}^2 dw d\underset{\sim}{\varepsilon}$$

and therefore we have:

$$\lim_{t\to 0}\int\left\{\frac{s_t(\underset{\sim}{\varepsilon})-s_0(\underset{\sim}{\varepsilon})}{t}\right\}^2 d\underset{\sim}{\varepsilon} \le \int\{s_0'(\underset{\sim}{\varepsilon})\}^2 d\underset{\sim}{\varepsilon} \qquad (2.4.8)$$

Now, since $[\{s_t(\underset{\sim}{\varepsilon})-s_0(\underset{\sim}{\varepsilon})\}/t]^2 \to [s_0'(\underset{\sim}{\varepsilon})]^2$ as $t \to 0$, it follows from (2.4.7) and Fatou's lemma that:

$$\lim_{t\to 0}\int\left\{\frac{s_t(\underset{\sim}{\varepsilon})-s_0(\underset{\sim}{\varepsilon})}{t}\right\}^2 d\underset{\sim}{\varepsilon} \ge \int\{s_0'(\underset{\sim}{\varepsilon})\}^2 d\underset{\sim}{\varepsilon} \qquad (2.4.9)$$

From (2.4.8) and (2.4.9), we conclude that equality must hold in both and according to criterion 4.5 in Hájek (1962) the functions $[\{s_t(\underset{\sim}{\varepsilon})-s_0(\underset{\sim}{\varepsilon})\}/t]^2$ are uniformly integrable, and the result follows.

Next we define:

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

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

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

and consider the following lemma:

Lemma 2.4.3: Under assumptions A1-A4 it follows that:

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

where $$b^2 = \int \left\{ \mathrm{tr}\underset{\sim}{\Delta} + \frac{h'(d_0)}{h(d_0)} \frac{1}{d_0} \underset{\sim}{\varepsilon}^T \underset{\approx}{\Delta}\underset{\sim}{\varepsilon} \right\}^2 h(d_0) d\underset{\sim}{\varepsilon}$$

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

$$EW_n = 2\sum_{k=1}^{n} \int \left\{ \frac{s_t(\underset{\sim}{\varepsilon}_k)}{s_0(\underset{\sim}{\varepsilon}_k)} - 1 \right\} s_0^2(\underset{\sim}{\varepsilon}_k) d\underset{\sim}{\varepsilon}_k = -n\int \{s_t(\underset{\sim}{\varepsilon}) - s_0(\underset{\sim}{\varepsilon})\}^2 d\underset{\sim}{\varepsilon}$$

Then observe that by Lemma 2.4.2:

$$\left| \left[\int \left\{ \frac{s_t(\underset{\sim}{\varepsilon}) - s_0(\underset{\sim}{\varepsilon})}{t} \right\}^2 d\underset{\sim}{\varepsilon} \right]^{\frac{1}{2}} - \left[\int \{s_0'(\underset{\sim}{\varepsilon})\}^2 d\underset{\sim}{\varepsilon} \right]^{\frac{1}{2}} \right| \le C^{\frac{1}{2}} \to 0$$

$$\text{as } t \to 0 \qquad (2.4.10)$$

Therefore we can approximate the first integral in (2.4.10) by the second and we get:

$$EW_n = -nt^2 \int \{s_0'(\underset{\sim}{\varepsilon})\}^2 d\underset{\sim}{\varepsilon} \to -b^2/4 \quad \text{as } n \to \infty \qquad (2.4.11)$$

Now introduce the statistic:

$$T_n = -t \sum_{k=1}^{n} [\mathrm{tr}\underset{\sim}{\Delta} + h'(d_0)\underset{\sim}{\varepsilon}_k^T \Delta \underset{\sim}{\varepsilon}_k / \{d_0 h(d_0)\}] = 2t \sum_{k=1}^{n} s_0'(\underset{\sim}{\varepsilon}_k)$$

and observe that:

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

and:

$$\mathrm{Var}T_n = 4t^2 \sum_{k=1}^{n} \int \{s_0'(\underset{\sim}{\varepsilon}_k)\}^2 d\underset{\sim}{\varepsilon}_k \to b^2 \text{ as } n \to \infty$$

Furthermore, by the Central Limit Theorem it follows that:

$$T_n \approx N(0, b^2) \; . \qquad (2.4.12)$$

Finally note that by Lemma 2.4.2 we have:

$$\operatorname{Var}(W_n - T_n) \le \sum_{k=1}^{n} \int \left\{ 2\, \frac{s_t(\underset{\sim}{\varepsilon}) - s_0(\underset{\sim}{\varepsilon})}{s_0(\underset{\sim}{\varepsilon})} - 2t\, \frac{s_0'(\underset{\sim}{\varepsilon})}{s_0(\underset{\sim}{\varepsilon})} \right\}^2 s_0^2(\underset{\sim}{\varepsilon})\, d\underset{\sim}{\varepsilon} =$$

$$= 4tC \to 0 \text{ as } t \to 0 \qquad (2.4.13)$$

From (2.4.11)-(2.4.13) the result follows.

Theorem 2.4.1: Under assumptions A1-A4, B and C2 it follows that for all $K > 0$, $L > 0$ and $\varepsilon > 0$:

$$P\{\sup n^{-\frac{1}{2}} \| \underset{\sim}{M}_n^*(\underset{\sim}{Y}, \underset{\sim}{\Sigma}, \hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n^*(\underset{\sim}{Y}, \underset{\sim}{\Sigma}, \underset{\sim}{\beta}) + n(\underset{\sim}{V} \otimes b_0 \underset{\sim}{I})(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*) \| \ge \varepsilon :$$

$$n^{\frac{1}{2}} \| \hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^* \| \le K,\ n^{\frac{1}{2}} \| \hat{\underset{\sim}{\Sigma}}^{\frac{1}{2}} - \underset{\sim}{\Sigma}^{\frac{1}{2}} \| \le L\} \to 0 \text{ as } n \to \infty$$

where b_0 is given by (2.4.2).

Proof: Considering a similar decomposition as in the proof of Theorem 2.3.1 and in view of Lemma 2.4.1, all we have to show is that for all $L>0$ and $\varepsilon>0$:

$$P\{\sup n^{-\frac{1}{2}} \| \underset{\sim}{M}_n^*(\underset{\sim}{Y}, \hat{\underset{\sim}{\Sigma}}, \hat{\underset{\sim}{\beta}}) - \underset{\sim}{M}_n^*(\underset{\sim}{Y}, \underset{\sim}{\Sigma}, \hat{\underset{\sim}{\beta}}) \| \ge \varepsilon :\ n^{\frac{1}{2}} \| \hat{\underset{\sim}{\Sigma}}^{\frac{1}{2}} - \underset{\sim}{\Sigma}^{\frac{1}{2}} \| \le L\} \to 0 \text{ as } n \to \infty$$

With no loss of generality we can substitute $\hat{\underset{\sim}{\Sigma}}^{-1}$ and $\underset{\sim}{\Sigma}^{-1}$ for $\hat{\underset{\sim}{\Sigma}}$ and $\underset{\sim}{\Sigma}$ in the proof since we assume these matrices to be p.d.

First define sets $S_{ih}(\delta)$, $(i,h=1,\ldots r;\delta=0,1)$ as in Lemma 2.3.1. Then partition each of these sets into two disjoint sets, one containing those k's for which sign $(y_{k\ell} - \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}}_\ell) = \operatorname{sign}(y_{km} - \underset{\sim}{x}_k^T \hat{\underset{\sim}{\beta}}_m)$, $\ell=1$, $m=2$ 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 2.3.1.

To simplify notation let $\hat{\underset{\sim}{\varepsilon}}_k = \underset{\sim}{y}_k - \hat{\underset{\sim}{\beta}}^T \underset{\sim}{x}_k$, (k=1,...n) and $\hat{\underset{\sim}{\Sigma}}^{-1} = \langle\hat{\sigma}^{\ell m}\rangle$. Then observe that for ℓ, m=1,...p we have:

$$(\partial/\partial\hat{\sigma}^{\ell m})\,[x_{ki}\hat{\varepsilon}_{kj}u\{(\hat{\underset{\sim}{\varepsilon}}_k^T\hat{\underset{\sim}{\Sigma}}^{-1}\underset{\sim}{\varepsilon}_k)^{\frac{1}{2}}\}] =$$

$$= x_{ki}\hat{\varepsilon}_{kj}u'\{(\hat{\underset{\sim}{\varepsilon}}_k^T\hat{\underset{\sim}{\Sigma}}^{-1}\hat{\underset{\sim}{\varepsilon}}_k)^{\frac{1}{2}}\}\hat{\varepsilon}_{k\ell}\hat{\varepsilon}_{km}/(\underset{\sim}{\varepsilon}_k^T\hat{\underset{\sim}{\Sigma}}^{-1}\underset{\sim}{\varepsilon}_k)^{\frac{1}{2}}$$

Since u is a monotone function this derivative has the same sign within each of the sets S_{ig} defined above, which implies that $M_{nijg}(\underset{\sim}{Y},\hat{\underset{\sim}{\Sigma}},\hat{\underset{\sim}{\beta}})$ is a monotone function in each element of $(\hat{\beta},\hat{\Sigma}^{-1})$ when the others are held fixed.

The rest of the proof follows the lines of that of Lemma 2.3.1 and is omitted. Lemmas 2.4.2 and 2.4.3 are useful in the counterpart of Step 3.

The boundedness in probability of $n^{\frac{1}{2}}||\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*||$ follows from a direct extension of Lemma 5.2 of Jurečková (1977). Besides the differences in notation discussed in Theorem 2.3.2, the main change in her proof relates to the definition of her M-function, which in our case should be given by:

$$M(\zeta) = n^{-\frac{1}{2}}\sum_{k=1}^{n} u[\{(\underset{\sim}{\varepsilon}_k+\zeta\underset{\sim}{w}_k)^T\hat{\underset{\sim}{\Sigma}}^{-1}(\underset{\sim}{\varepsilon}_k+\zeta\underset{\sim}{w}_k)\}^{\frac{1}{2}}]\underset{\sim}{w}_k^T(\underset{\sim}{\varepsilon}_k+\zeta\underset{\sim}{w}_k)$$

where $\underset{\sim}{w}_k = (\underset{\sim}{\beta}-\underset{\sim}{\beta}^{(1)})^T\underset{\sim}{x}_k$ and $\underset{\sim}{\beta}^{(1)}$ is any point such that $n^{\frac{1}{2}}||\underset{\sim}{\beta}^{(1)}*-\underset{\sim}{\beta}^*|| = K$.

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

<u>Theorem 2.4.2</u>: Under assumptions A1-A4, B and C2 it follows that:

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}^*-\underset{\sim}{\beta}^*) \approx N_{rp}(\underset{\sim}{0},\underset{\sim}{V}^{-1}\otimes a_0b_0^{-2}\underset{\sim}{\Sigma})$$

where b_0 and a_0 are defined by (2.4.2) and (2.4.6) respectively.

Proof: Similar to that of Theorem 2.3.3 with the use of Theorem 2.4.1 in lieu of Theorem 2.3.1.

Asymptotic tests as those of section 2.3 can be obtained similarly. The $\underset{\sim}{H}$ matrices are computed in the same way and the $\underset{\sim}{E}$ matrices are given by $\underset{\sim}{E} = n\hat{a}_0\hat{b}_0^{-2}\underset{\sim}{U}^T\hat{\underset{\sim}{\Sigma}}\underset{\sim}{U}$ under the first approach and by $\underset{\sim}{E} = n\hat{a}_0\underset{\sim}{U}^T\hat{\underset{\sim}{\Sigma}}\underset{\sim}{U}$ under Sen's approach, where $\hat{a}_0$, $\hat{b}_0$ and $\hat{\underset{\sim}{\Sigma}}$ are $n^{\frac{1}{2}}$-consistent estimators of a_0, b_0 and $\underset{\sim}{\Sigma}$ respectively. The null asymptotic distributions of the test statistics are exactly the same as those of Section 2.3; under the sequence of Pitman-type alternatives the asymptotic distribution of $\mathrm{ntr}(\underset{\sim}{H}\underset{\sim}{E}^{-1})$ or $\mathrm{nlog}(|\underset{\sim}{E}+\underset{\sim}{H}|/|\underset{\sim}{E}|)$ is $\chi^2_{cu}(\mathrm{tr}\underset{\sim}{\Omega}_E)$ where:

$$\underset{\sim}{\Omega}_E = b_0^2a_0^{-1}(\underset{\sim}{U}^T\underset{\sim}{\Sigma}\underset{\sim}{U})^{-\frac{1}{2}}\underset{\sim}{\Delta}^T\underset{\sim}{P}\underset{\sim}{\Delta}(\underset{\sim}{U}^T\underset{\sim}{\Sigma}\underset{\sim}{U})^{-\frac{1}{2}} \quad .$$

2.5 Robustness considerations

In this section we discuss two heuristic indicators of robustness commonly used in the literature: influence functions and breakdown points. In both cases we are concerned only with the asymptotic aspects.

Initially we obtain the asymptotic influence function of both types of M-estimators with preliminary estimates of scale. Then we define the gross error breakdown point and show that for either type of estimator it is positive only under very simple structures of the design matrix.

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

$$H_t = (1-t)H+t\ \delta(\underset{\sim}{\varepsilon}_0,\underset{\sim}{x}_0)$$

where $H = F \cdot G$ and $\delta(\underset{\sim}{\varepsilon}_0, \underset{\sim}{x}_0)$ is a pointmass 1 at $(\underset{\sim}{\varepsilon}_0, \underset{\sim}{x}_0)$. Define $\underset{\sim}{\beta}(H_t)$ as a solution to:

$$\int \underset{\sim}{x}\psi[\{y_j - \underset{\sim}{x}^T \underset{\sim}{\beta}_j(H_t)\}/\sigma_j(H_t)]dH_t = \underset{\sim}{0}\ , \quad (j=1,\dots p) \tag{2.5.1}$$

where $\underset{\sim}{\beta}_j(H_t)$ denotes the j^{th} column of $\underset{\sim}{\beta}(H_t)$ and $\sigma_j(H_t)$ is regarded as a functional of H_t such that $\sigma_j(H) = \sigma_j$. Note also that $\underset{\sim}{\beta}(H) = \underset{\sim}{\beta}$.

To obtain the influence function of the coordinatewise M-estimator we differentiate (2.5.1) with respect to t and evaluate the result at t=0. Writing $\dot{\underset{\sim}{\beta}}_j = (\partial/\partial t)\underset{\sim}{\beta}_j(H_t)\big|_{t=0}$, $\dot{\sigma}_j = (\partial/\partial t)\sigma_j(H_t)\big|_{t=0}$ and $\dot{H} = (\partial/\partial t)H_t\big|_{t=0} = \delta(\underset{\sim}{\varepsilon}_0, \underset{\sim}{x}_0) - F(\underset{\sim}{\varepsilon})G(\underset{\sim}{x})$ we get for $(j=1,\dots p)$:

$$-\int \psi'(\varepsilon_j/\sigma_j)\underset{\sim}{x}[\{\underset{\sim}{x}^T \dot{\underset{\sim}{\beta}}_j \sigma_j + \varepsilon_j \dot{\sigma}_j\}/\sigma_j^2]dF(\underset{\sim}{\varepsilon})dG(\underset{\sim}{x}) + \psi(\varepsilon_{0j}/\sigma_j)\underset{\sim}{x}_0 = \underset{\sim}{0}$$

In general we take ψ odd and assume F symmetric and in this case $\int \psi'(\varepsilon_j/\sigma_j)\varepsilon_j/\sigma_j dF(\underset{\sim}{\varepsilon}) = 0$. Therefore:

$$\dot{\underset{\sim}{\beta}}_j = \{\int \underset{\sim}{x}\underset{\sim}{x}^T dG(\underset{\sim}{x})\}^{-1}\sigma_j \psi(\varepsilon_{0j}/\sigma_j)/\int \psi'(\varepsilon_j/\sigma_j)dF(\underset{\sim}{\varepsilon})\underset{\sim}{x}_0\ .$$

Now, letting G correspond to a distribution which associates probability n^{-1} to each point $\underset{\sim}{x}_k$, $(k=1,\dots n)$ and making $n \to \infty$ we can write the asymptotic influence function of the coordinatewise M-estimator as:

$$\underset{\sim}{i}(\underset{\sim}{\varepsilon}_0, \underset{\sim}{x}_0; F, \hat{\underset{\sim}{\beta}}) = \underset{\sim}{V}^{-1}\underset{\sim}{x}_0 \otimes \underset{\sim}{W}^{-1}\underset{\sim}{\psi}(\underset{\sim}{\gamma}_0) \tag{2.5.2}$$

where $\underset{\sim}{\psi}(\underset{\sim}{\gamma}_0) = \{\psi(\gamma_{01}),\dots\psi(\gamma_{0p})\}^T$ and $\underset{\sim}{\gamma}_0 = \{\mathrm{diag}(\sigma_1^{-1},\dots\sigma_p^{-1})\}\underset{\sim}{\varepsilon}_0$.

In the case of the Maronna-type M-estimator we define $\underset{\sim}{\beta}(H_t)$ as a solution to:

$$\int u\{d(H_t)\}\underset{\sim}{x}\{y_j - \underset{\sim}{x}^T \underset{\sim}{\beta}_j(H_t)\}dH_t = \underset{\sim}{0}\ , \quad (j=1,\dots p) \tag{2.5.3}$$

where:

$$d^2(H_t)=\{\underset{\sim}{y}-\underset{\sim}{\beta}^T(H_t)\underset{\sim}{x}\}^T\underset{\sim}{\Sigma}^{-1}(H_t)\{\underset{\sim}{y}-\underset{\sim}{\beta}^T(H_t)\underset{\sim}{x}\}$$

and $\underset{\sim}{\Sigma}(H_t)$ is a functional of H_t such that $\underset{\sim}{\Sigma}(H) = \underset{\sim}{\Sigma}$.

Differentiating (2.5.3) with respect to t and evaluating the result at t=0 we obtain for (j=1,...p):

$$\int x\{u'(d)\dot{d}^2\varepsilon_{0j}/2d-u(d)\underset{\sim}{x}^T\dot{\underset{\sim}{\beta}}_j\}dF(\underset{\sim}{\varepsilon})dG(\underset{\sim}{x}) + \int\underset{\sim}{x}\{u(d)\varepsilon_j\}d\dot{H} = \underset{\sim}{0} \qquad (2.5.4)$$

where $\dot{d}^2 = (\partial/\partial t)d^2(H_t)|_{t=0}$. Now observe that:

$$\dot{d}^2=\sum_{\ell=1}^{p}\sum_{m=1}^{p}\{-\underset{\sim}{x}^T\dot{\underset{\sim}{\beta}}_\ell\varepsilon_m-\underset{\sim}{x}^T\dot{\underset{\sim}{\beta}}_m\varepsilon_\ell\}\sigma^{\ell m}+\sum_{\ell=1}^{p}\sum_{m=1}^{p}\varepsilon_\ell\varepsilon_m\dot{\sigma}^{\ell m}$$

where $\dot{\sigma}^{\ell m} = (\partial/\partial t)\sigma^{\ell m}(H_t)|_{t=0}$ and $\underset{\sim}{\Sigma}^{-1}(H_t) = \langle\sigma^{\ell m}(H_t)\rangle$, $(\ell,m=1,\ldots p)$. Taking $\underset{\sim}{\Sigma} = \underset{\sim}{I}$ with no loss of generality we can write:

$$\dot{d}^2 = -2\underset{\sim}{x}^T\sum_{\ell=1}^{p}\dot{\underset{\sim}{\beta}}_\ell\varepsilon_\ell + \sum_{\ell=1}^{p}\sum_{m=1}^{p}\dot{\sigma}^{\ell m}\varepsilon_\ell\varepsilon_m \qquad (2.5.5)$$

and substituting (2.5.5) in (2.5.4) we obtain for (j=1,...p):

$$-\int\underset{\sim}{x}\underset{\sim}{x}^T\{u(d)\sum_{\ell=1}^{p}\dot{\underset{\sim}{\beta}}_\ell\varepsilon_\ell\varepsilon_j/d+u(d)\dot{\underset{\sim}{\beta}}_j\}dF(\underset{\sim}{\varepsilon})dG(\underset{\sim}{x}) +$$

$$+2^{-1}\int\underset{\sim}{x}\underset{\sim}{x}^T\sum_{\ell=1}^{p}\sum_{m=1}^{p}\dot{\sigma}^{\ell m}\{u'(d)\varepsilon_\ell\varepsilon_m\varepsilon_j/d\}dF(\underset{\sim}{\varepsilon})dG(\underset{\sim}{x}) +$$

$$+ \underset{\sim}{x}_0u(d_0)\varepsilon_{0j}=\underset{\sim}{0}$$

where $d_0 = \underset{\sim}{\varepsilon}_0^T\underset{\sim}{\varepsilon}_0$. From the assumption of spherical symmetry ($\underset{\sim}{\Sigma} = \underset{\sim}{I}$) it follows that the integrals involving $\varepsilon_\ell\varepsilon_j$ are zero except for $\ell = j$ and the integrals involving $\varepsilon_\ell\varepsilon_m\varepsilon_j$ are all zero. Therefore we may write:

$$\dot{\underset{\sim}{\beta}}_j=\{\int\underset{\sim}{x}\underset{\sim}{x}^TdG(\underset{\sim}{x})\}^{-1}\psi(d_0)\varepsilon_{0j}/(b_0d_0)\underset{\sim}{x}_0$$

Defining G as we did previously and making $n \to \infty$ we can write the asymptotic influence function of the Maronna-type M-estimator as:

$$\underset{\sim}{i}(\underset{\sim}{\varepsilon}_0,\underset{\sim}{x}_0;F,\hat{\underset{\sim}{\beta}}) = \underset{\sim}{V}^{-1}\underset{\sim}{x}_0 \otimes \psi(d_0)/(b_0 d_0)\underset{\sim}{\varepsilon}_0. \tag{2.5.6}$$

By examining the asymptotic influence functions given by (2.5.2) and (2.5.6) it is easy to see that in both cases the M-estimators will be locally robust only 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 suggestion.

Also note that similarly to the univariate location case as in Huber (1981, ch.1), the asymptotic covariance matrices given in Theorems 2.3.3 and 2.4.2 are reproduced by taking $\int \underset{\sim}{i}\underset{\sim}{i}^T dH$.

Now we consider the breakdown points of the proposed estimators. Following Maronna, Bustos and Yohai (1979), define the gross error breakdown point by:

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

In Theorem 2.5.1 below we show that for both types of M-estimators 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 only at the point defining the design space. Furthermore it deteriorates very rapidly as the number of parameters increases.

<u>Theorem 2.5.1</u>: Assume that:

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

(ii) $\lim_{n\to\infty} n_\ell/n = p_\ell$, $(\ell=1,\ldots s)$, 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 and $\underset{\sim}{\Sigma}$ is known.

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 $t^*=0$.

Proof: To simplify notation we consider only the univariate case (p=1); the extension to the multivariate case is straightforward. Also note that without loss of generality we may take $\sigma=1$ and restrict ourselves to the case where $\underset{\sim}{X} = \underset{\sim}{I}$.

Next we show that the only component of $\underset{\sim}{\beta}(H_t)$ affected by the contamination is the one corresponding to the nonnull element of $\underset{\sim}{x}_0$. Note that $\underset{\sim}{\beta}(H)$ is defined by:

$$\sum_{\ell=1}^{r} \underset{\sim}{x}_\ell \int \psi(y-\beta_\ell)dF(\varepsilon)=\underset{\sim}{0}$$

which is equivalent to:

$$\int \psi(y-\beta_\ell)dF(\varepsilon)=0 \quad , \quad (\ell=1,\ldots r) \tag{2.5.7}$$

On the other hand $\underset{\sim}{\beta}(H_t)$ is defined by:

$$(1-t)\sum_{\ell=1}^{r} \underset{\sim}{x}_\ell \int \psi\{y-\beta_\ell(H_t)\}dF(\varepsilon)+t\ \underset{\sim}{x}_0\psi\{y_0-\beta_0(H_t)\}=\underset{\sim}{0}$$

which corresponds to:

$$\begin{aligned} &(1-t)\int \psi\{y-\beta_\ell(H_t)\}dF(\varepsilon)=0, \quad \ell\neq 0 \\ &(1-t)\int \psi\{y-\beta_0(H_t)\}dF(\varepsilon)=0 \end{aligned} \tag{2.5.8}$$

For $t\neq 1$ the first r-1 equations in (2.5.8) are equivalent to the corresponding ones in (2.5.7) and it follows that $\beta_\ell(H_t)=\beta_\ell$, $\ell\neq 0$.

Now, for every sequence $\underset{\sim}{\beta}(H_{tn})$, where $H_{tn}=(1-t)H+t\delta(\varepsilon_n,\underset{\sim}{x}_0)$ we must have:

$$-t\psi\{y_n-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}\underset{\sim}{x}_0 = (1-t)\frac{n_0}{n}\int\psi\{y-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}dF(\varepsilon)\underset{\sim}{x}_0 \tag{2.5.9}$$

where n_0 can take the values n_ℓ (ℓ=1,...r) according to the values taken by $\underset{\sim}{x}_0$. Equating the coefficients of $\underset{\sim}{x}_0$ and taking absolute values on both sides we obtain:

$$t|\psi\{y_n-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}|=(1-t)\frac{n_0}{n}|\int\psi\{y-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}dF(\varepsilon)| \tag{2.5.10}$$

Suppose that for all $\delta(\varepsilon_0,\underset{\sim}{x}_0)$ we have $\|\underset{\sim}{\beta}(H_{tn})\| \le K$. Then we can choose a sequence $(y_n,\underset{\sim}{x}_0)$ such that $|\psi\{y_n-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}| \to M$ as $n \to \infty$ where $M = \sup|\psi(\varepsilon)|$. Letting $n \to \infty$ in (2.5.10) we get:

$$t\lim_{n\to\infty}|\psi\{y_n-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}|\le(1-t)\lim_{n\to\infty}\frac{n_0}{n}\int|\psi\{y-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}|dF(\varepsilon)$$

which implies $t \le (1+p_0^{-1})^{-1}$. Since this relation must hold for $\underset{\sim}{x}_0=\underset{\sim}{x}_\ell$, (ℓ=1,...r) we have:

$$t\le \min_{\ell=1,\dots r}(1+p_\ell^{-1})^{-1} \tag{2.5.11}$$

Now suppose that there is a sequence $\delta(\varepsilon_n,\underset{\sim}{x}_0)$ such that $\|\underset{\sim}{\beta}(H_{tn})\| \to \infty$ as $n \to \infty$. Then from (2.5.10) letting $n \to \infty$, we get:

$$t\lim_{n\to\infty}|\psi\{y_n-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}|= (1-t)\lim_{n\to\infty}|\int\psi\{y-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}dF(\varepsilon)|$$

which implies:

$$tM \ge(1-t)p_0|\int\lim_{n\to\infty}\psi\{y-\underset{\sim}{x}_0^T\underset{\sim}{\beta}(H_{tn})\}dF(\varepsilon)|= (1-t)p_0M$$

and then $t \ge (1+p_0^{-1})^{-1}$. Therefore:

$$t \ge \max_{\ell=1,\dots r}(1+p_\ell^{-1})^{-1} \tag{2.5.12}$$

From (2.5.11), (2.5.12) and the definition of the gross error breakdown point we get $t^* = \min_{\ell=1,\ldots r} (1+p_\ell^{-1})^{-1}$. 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 $\underset{\sim}{x}_0$ the only solution to (2.5.9) is $t = 0$.

The proof of Theorem 2.5.1 assumes known scale. In the case where it is not known the breakdown point of the M-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.

CHAPTER III

M-METHODS IN OTHER MULTIVARIATE LINEAR MODELS

3.1 Introduction

In this chapter we consider the extension of the M-methods studied in Chapter 2 to some more general multivariate linear models such as growth curve models or those allowing for different design matrices for the different variates and/or missing observations.

In Section 3.2 we define the M-estimators for the parameter matrix in a growth curve model and show how the results of Chapter 2 can be applied by means of a linear transformation suggested by Potthoff and Roy (1964). We also discuss the optimality problem related to the choice of the matrix defining the linear transformation. Finally we show how the covariance adjustment technique considered by Rao (1967) and Grizzle and Allen (1969) for the classical Normal theory methods can be extended to the situation discussed here.

In Section 3.3 we indicate how M-estimators and their asymptotic properties can be obtained for Kleinbaum's (1970) MGLM and GGCM.

3.2 M-methods in growth curve models

Consider the GCM given by (1.5.4) and assume that $\underset{\sim}{\Sigma}$ is known.

Following the suggestion of Potthoff and Roy (1964) let $\underset{\sim}{H}_1(p\times q)$ be any matrix such that $\underset{\sim}{G}\underset{\sim}{H}_1 = \underset{\sim}{I}$ and observe that by making the transformation:

$$\underset{\sim}{Z} = \underset{\sim}{Y}\underset{\sim}{H}_1, \ \underset{\sim}{\tau} = \underset{\sim}{\varepsilon}\underset{\sim}{H}_1$$

the initial GCM can be reduced to the SMLM given by:

$$\underset{\sim}{Z} = \underset{\sim}{X}\underset{\sim}{\xi}+\underset{\sim}{\tau} \tag{3.2.1}$$

Since the (linear) transformation does not affect the assumptions A1-C2 of Section 2.2, we can apply the results of Chapter 2 to obtain M-estimators of the parameter matrix $\underset{\sim}{\xi}$ and their asymptotic distributions.

In the coordinatewise case we have:

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\xi}}^*-\underset{\sim}{\xi}^*) \approx N_{rq}(\underset{\sim}{0},\underset{\sim}{V}^{-1}\otimes\underset{\sim}{W}_1^{-1}\underset{\sim}{\ddagger}_1\underset{\sim}{W}_1^{-1}) \tag{3.2.2}$$

where $\underset{\sim}{W}_1$ and $\underset{\sim}{\ddagger}_1$ correspond to (2.3.10) and (2.3.11) defined in terms of the distribution function of the transformed variables $\underset{\sim}{z}_k$, say H_1.

Similarly for the Maronna-type M-estimator we have:

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\xi}}^*-\underset{\sim}{\xi}^*) \approx N_{rq}(\underset{\sim}{0},\underset{\sim}{V}^{-1}\otimes a_1 b_1^{-2}\underset{\sim}{H}_1^T\underset{\sim}{\Sigma}\underset{\sim}{H}_1) \tag{3.2.3}$$

where a_1 and b_1 respectively correspond to (2.4.2) and (2.4.6) defined in terms of H_1.

Potthoff and Roy (1964) suggested that we take:

$$\underset{\sim}{H}_1=\underset{\sim}{A}^{-1}\underset{\sim}{G}^T(\underset{\sim}{G}\underset{\sim}{A}^{-1}\underset{\sim}{G}^T)^{-1} \tag{3.2.4}$$

where $\underset{\sim}{A}(p\times p)$ is an arbitrary but fixed p.d. matrix. They showed that if the underlying distribution is Normal the choice $\underset{\sim}{A}=\underset{\sim}{\Sigma}$ in (3.2.4) and $u(d) = -d^{-1}(\partial/\partial d)\log h(d)$ in (2.2.1) or $\psi_j(x) = -(\partial/\partial x)\log f_j(x)$ in (2.2.6) leads to the fully efficient NML estimator. In the more general situation considered here we cannot expect to attain the Fréchet-Cramér-Rao lower bound since the underlying distribution is not completely specified. We can, however, attempt to maximize the asymptotic efficiency of the

M-estimators in its general form at some specified model, say the Normal model.

First observe that we might take $\Sigma = I$ and G such that $GG^T = I$ with no loss of generality. In this case $\mathrm{Var}(\tau_k) = H_1^T H_1 = H = \langle h_{ij} \rangle$, $(i,j=1,\ldots q)$; note that H is a function of A.

Now suppose that the underlying distribution is unknown and that assumptions A1-A3, B and C1 hold. Expressing the asymptotic covariance matrix of the coordinatewise M-estimator as a function of A we have:

$$\Gamma(A) = W_1^{-1} \Sigma_1 W_1^{-1} = \langle \gamma_{ij}(A) \rangle$$

where:

$$\gamma_{ij} = h_i h_j [E_{H_1} \psi(\tau_i/h_i)\psi(\tau_j/h_j)/\{E_{H_1}\psi'(\tau_i/h_i)\psi'(\tau_j/h_j)\}] \tag{3.2.5}$$

and $h_i = h_{ii}^{\frac{1}{2}}$. 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 $\mathrm{ch}_1\{\Gamma(A)\}$.

Letting $\lambda_j(A) = \mathrm{ch}_j\{\Gamma(A)\}$, $(j=1,\ldots q)$, note that:

(i) $(\partial/\partial\lambda_j) \prod_{i=1}^{q} \lambda_i(A) = \prod_{i\neq j} \lambda_i(A) > 0$, $(j=1,\ldots q)$

(ii) $(\partial/\partial\lambda_j) \sum_{i=1}^{q} \lambda_i(A) = 1$, $(j=1,\ldots q)$

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 such that:

$$P\Gamma(A)P^T = \mathrm{diag}\{\lambda_1(A),\ldots\lambda_q(A)\} = D$$

and therefore:

$$\lambda_1(\underset{\sim}{A}) = \sup_{\underset{\sim}{x}}\{\underset{\sim}{x}^T\underset{\sim}{\Gamma}(\underset{\sim}{A})\underset{\sim}{x}/\underset{\sim}{x}^T\underset{\sim}{x}\} = \sup_{\underset{\sim}{x}}\{\underset{\sim}{x}^T\underset{\sim}{P}^T\underset{\sim}{D}\underset{\sim}{P}\underset{\sim}{x}/\underset{\sim}{x}^T\underset{\sim}{P}^T\underset{\sim}{P}\underset{\sim}{x}\} =$$

$$= \sup_{\underset{\sim}{y}} \sum_{i=1}^{q} \lambda_i(\underset{\sim}{A})y_i^2/\underset{\sim}{y}^T\underset{\sim}{y}$$

where $\underset{\sim}{y} = \underset{\sim}{P}\underset{\sim}{x}$. Now, since $(\partial/\partial\lambda_j)\sum_{i=1}^{q}\lambda_i(\underset{\sim}{A})y_i^2/\underset{\sim}{y}^T\underset{\sim}{y} = y_j^2/\underset{\sim}{y}^T\underset{\sim}{y} > 0$, $(j=1,\ldots q)$ we conclude that $\lambda_1(\underset{\sim}{A})$ is also an increasing function in each λ_i when all the other are held fixed.

In view of these facts we may conclude that the minimum of $\text{tr}\{\underset{\sim}{\Gamma}(\underset{\sim}{A})\}$, $|\underset{\sim}{\Gamma}(\underset{\sim}{A})|$ or of $\text{ch}_1\{\underset{\sim}{\Gamma}(\underset{\sim}{A})\}$ will correspond to $\lambda_i(\underset{\sim}{A}) = \lambda$, $(i=1,\ldots q)$. Therefore $\underset{\sim}{\Gamma}(\underset{\sim}{A})$ must be of the form $\lambda\underset{\sim}{I}$ at the minimum. To see this, note that $\underset{\sim}{P}\underset{\sim}{\Gamma}(\underset{\sim}{A})\underset{\sim}{P}^T = \lambda\underset{\sim}{I}$ which implies $\underset{\sim}{\Gamma}(\underset{\sim}{A}) = \lambda\underset{\sim}{P}^T\underset{\sim}{P} = \lambda\underset{\sim}{I}$. Furthermore from (3.2.5) we conclude that at the minimum:

$$\underset{\sim}{\Gamma}(\underset{\sim}{A}) = h_{ii}(\underset{\sim}{A})\{E_{H_1}\psi^2(\tau_i/h_i)/E_{H_1}^2\psi'(\tau_i/h_i)\}\underset{\sim}{I}$$

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

Now suppose further that the (unknown) underlying distribution is elliptically symmetric and consider the Maronna-type M-estimator.

Then observe that the distribution of $\underset{\sim}{\tau}_k$ is also elliptically symmetric with covariance matrix $\underset{\sim}{H}_1^T\underset{\sim}{\Sigma}\underset{\sim}{H}_1$ and consider the transformation $\underset{\sim}{\nu}_k = (\underset{\sim}{H}_1^T\underset{\sim}{\Sigma}\underset{\sim}{H}_1)^{-\frac{1}{2}}\underset{\sim}{\tau}_k$ which preserves the elliptical symmetry and implies that $\text{Var}(\underset{\sim}{\nu}_k) = \underset{\sim}{I}$. Furthermore $d^2 = \underset{\sim}{\tau}_k^T(\underset{\sim}{H}_1^T\underset{\sim}{\Sigma}\underset{\sim}{H}_1)^{-1}\underset{\sim}{\tau}_k = \underset{\sim}{\nu}_k^T\underset{\sim}{\nu}_k$ which implies that both a_1

and b_1 are invariant with respect to the choice of $\underset{\sim}{H}_1$ (and consequently of $\underset{\sim}{A}$). Then from (3.2.3), (3.2.4) and the Fréchet-Cramér-Rao lower bound it is easy to see that the optimum choice at the Normal model corresponds to $\underset{\sim}{A}=\underset{\sim}{\Sigma}$.

Though the problem of optimality of the choice of $\underset{\sim}{A}$ under more general situations (than at the Normal model) requires further research the above approach can be justified by the fact that it can produce M-estimators which perform as well as possible at the Normal model while retaining robustness against departures from it.

Now note that if $\hat{\underset{\sim}{\Sigma}}$ is such that:

$$n^{\frac{1}{2}}||\hat{\underset{\sim}{\Sigma}}^{\frac{1}{2}}-\underset{\sim}{\Sigma}^{\frac{1}{2}}||= O_p(1) \tag{3.2.6}$$

we get:

$$n^{\frac{1}{2}}||(\underset{\sim}{H}_1^T\hat{\underset{\sim}{\Sigma}}\underset{\sim}{H}_1)^{\frac{1}{2}}-(\underset{\sim}{H}_1^T\underset{\sim}{\Sigma}\underset{\sim}{H}_1)^{\frac{1}{2}}||= O_p(1).$$

Therefore the asymptotic distributions and the optimality results considered above still hold if we replace $\underset{\sim}{\Sigma}$ by an estimate $\hat{\underset{\sim}{\Sigma}}$ satisfying (3.2.6).

Also note that asymptotic tests of hypotheses about the elements of the parameter matrix $\underset{\sim}{\xi}$ can be obtained in the same ways as in Chapter 2.

Now we indicate how the covariance adjustment technique can be used for M-estimators.

Let $\underset{\sim}{H}_1(p\times q)$ and $\underset{\sim}{H}_2(p\times p\text{-}q)$ be any matrices such that $\underset{\sim}{G}\underset{\sim}{H}_1=\underset{\sim}{I}$ and $\underset{\sim}{G}\underset{\sim}{H}_2=\underset{\sim}{0}$. For example take $\underset{\sim}{H}_1$ as in (3.2.4) and $\underset{\sim}{H}_2$ as a basis of the space spanned by the columns of $\underset{\sim}{I}-\underset{\sim}{A}^{-1}\underset{\sim}{G}^T(\underset{\sim}{G}\underset{\sim}{A}^{-1}\underset{\sim}{G}^T)^{-1}\underset{\sim}{G}$.

Now make the transformation:

$$[\underset{\sim}{Z}\ \underset{\sim}{Q}]=\underset{\sim}{Y}[\underset{\sim}{H}_1\ \underset{\sim}{H}_2],\quad [\underset{\sim}{\tau}\ \underset{\sim}{\nu}]=\underset{\sim}{\varepsilon}[\underset{\sim}{H}_1\ \underset{\sim}{H}_2]$$

and consider the conditional (on $\underset{\sim}{Q}$) multivariate linear model:

$$\underset{\sim}{Z}|\underset{\sim}{Q}=\underset{\sim}{X}\,\underset{\sim}{\xi}+\underset{\sim}{Q}\,\underset{\sim}{\eta}+\underset{\sim}{\delta} = [\underset{\sim}{X}\;\underset{\sim}{Q}]\begin{bmatrix}\underset{\sim}{\xi}\\ \underset{\sim}{\eta}\end{bmatrix}+\underset{\sim}{\delta} \tag{3.2.7}$$

Let $H_{1.2}$ denote the distribution function of the r.v.'s $\underset{\sim}{\delta}_k$, which corresponds to the conditional distribution of $\underset{\sim}{\tau}_k$ given $\underset{\sim}{\nu}_k$.

Note also that assumptions A1-A4 are preserved under the (linear) transformation. Furthermore observe that assumption B(i) is also satisfied by the design matrix $[\underset{\sim}{X}\;\underset{\sim}{Q}]$ with probability 1. Now we show that assumption B(ii) is also preserved; more precisely we show that:

$$n^{-1}\begin{bmatrix}\underset{\sim}{X}^T\underset{\sim}{X} & \underset{\sim}{X}^T\underset{\sim}{Q}\\ \underset{\sim}{Q}^T\underset{\sim}{X} & \underset{\sim}{Q}^T\underset{\sim}{Q}\end{bmatrix} \xrightarrow{P} \begin{bmatrix}\underset{\sim}{V} & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{H}_2^T\underset{\sim}{\Sigma}\underset{\sim}{H}_2\end{bmatrix} \tag{3.2.8}$$

From assumption B(ii) it follows that:

$$\underset{\sim}{n}^{-1}\underset{\sim}{X}^T\underset{\sim}{X} \longrightarrow \underset{\sim}{V} \tag{3.2.9}$$

By Kolmogorov's Strong Law of Large Numbers we have:

$$n^{-1}\underset{\sim}{Q}^T\underset{\sim}{Q}=n^{-1}\sum_{k=1}^{n}\underset{\sim}{q}_k\underset{\sim}{q}_k^T \longrightarrow E\underset{\sim}{Q}^T\underset{\sim}{Q}=\underset{\sim}{H}_2^T\underset{\sim}{\Sigma}\underset{\sim}{H}_2 \tag{3.2.10}$$

Now to show that $n^{-1}\underset{\sim}{X}^T\underset{\sim}{Q} \xrightarrow{P} \underset{\sim}{0}$ we can show that for all $\underset{\sim}{\alpha}\in\mathbb{R}^r$ and $\underset{\sim}{\beta}\in\mathbb{R}^{p-q}$, $\underset{\sim}{\alpha} \neq \underset{\sim}{0}$, $\underset{\sim}{\beta} \neq \underset{\sim}{0}$ we get $n^{-1}\underset{\sim}{\alpha}^T\underset{\sim}{X}^T\underset{\sim}{Q}\underset{\sim}{\beta} \xrightarrow{P} 0$. Let:

$$\psi_n(t)=E[\exp\{it\underset{\sim}{\alpha}^T\underset{\sim}{X}^T\underset{\sim}{Q}^T\underset{\sim}{\beta}/n\}]=E[\exp\{it\sum_{k=1}^{n}\alpha^T\underset{\sim}{x}_k\underset{\sim}{q}_k^T\underset{\sim}{\beta}/n\}]=$$

$$=\prod_{k=1}^{n}E[\exp\{itU_k/n\}]=\prod_{k=1}^{n}\phi_k(t/n)\ .$$

Expanding $\phi_k(t/n)$ in a Taylor series we get:

$$\phi_k(t/n)=1+itEU_k/n+i^2tEU_k^2/n^2+o(t^2/n^2)$$

Then observe that:

$$EU_k = E(\underset{\sim}{\alpha}^T \underset{\sim}{x}_k \underset{\sim}{q}_k^T \underset{\sim}{\beta}) = \underset{\sim}{\alpha}^T \underset{\sim}{x}_k E\underset{\sim}{q}_k^T \underset{\sim}{\beta} = 0$$

$$EU_k^2 = E(\underset{\sim}{\alpha}^T \underset{\sim}{x}_k \underset{\sim}{q}_k^T \underset{\sim}{\beta})^2 = \underset{\sim}{\alpha}^T \underset{\sim}{x}_k \underset{\sim}{x}_k^T \underset{\sim}{\alpha} \underset{\sim}{\beta}^T E\underset{\sim}{q}_k \underset{\sim}{q}_k^T \underset{\sim}{\beta} = 0(1)$$

by assumption B(i). Therefore it follows that $\phi_k(t/n) = 1 + o(n^{-1})$ which implies $\psi_n(t) = \{1 + o(n^{-1})\}^n \to 1$ as $n \to \infty$, and then:

$$n^{-1} \underset{\sim}{X}^T \underset{\sim}{Q} \xrightarrow{P} \underset{\sim}{0} \qquad (3.2.11)$$

From (3.2.9)-(3.2.11) we conclude (3.2.8) and Theorem 2.3.3 and 2.4.2 can be applied to obtain the asymptotic distributions of the M-estimators.

In the coordinatewise case we have:

$$n^{1/2}(\hat{\underset{\sim}{\xi}}^* - \underset{\sim}{\xi}^*) \approx N_{rq}(\underset{\sim}{0}, \underset{\sim}{V}^{-1} \otimes \underset{\sim}{W}_{1.2}^{-1} \underset{\sim}{\ddagger}_{1.2} \underset{\sim}{W}_{1.2}^{-1}) \qquad (3.2.12)$$

where $\underset{\sim}{W}_{1.2}$ and $\underset{\sim}{\ddagger}_{1.2}$ respectively correspond to (2.3.10) and (2.3.11) defined in terms of $H_{1.2}$.

For the Maronna-type M-estimator we have:

$$n^{1/2}(\hat{\underset{\sim}{\xi}}^* - \underset{\sim}{\xi}^*) \approx N_{rq}(\underset{\sim}{0}, \underset{\sim}{V}^{-1} \otimes a_{1.2} b_{1.2}^{-2} \underset{\sim}{\Sigma}_{1.2}) \qquad (3.2.13)$$

where $a_{1.2}$ and $b_{1.2}$ respectively correspond to (2.4.2) and (2.4.6) defined in terms of $H_{1.2}$ and $\underset{\sim}{\Sigma}_{1.2}$ is the scatter matrix associated to $H_{1.2}$.

Since neither (3.2.12) nor (3.2.13) depend on the covariate matrix $\underset{\sim}{Q}$ they correspond to the unconditional asymptotic distributions of the proposed estimators.

We can also obtain the asymptotic distributions of the M-estimators of $\underset{\sim}{\eta}$ in a similar way; they are given by expressions similar to (3.2.12) and (3.2.13) with $\underset{\sim}{V}^{-1}$ replaced by $(\underset{\sim}{H}_2^T \underset{\sim}{\Sigma} \underset{\sim}{H}_2)^{-1}$ and dimension (p-q)q instead of rq.

Asymptotic tests about the parameters $[\underset{\sim}{\xi}^T \underset{\sim}{\eta}^T]^T$ can be obtained by the methods of Chapter 2. Though we are basically concerned with $\underset{\sim}{\xi}$ a test of the hypothesis $H:\underset{\sim}{\eta} = \underset{\sim}{0}$ is of special interest since covariance adjustment only makes sense when $\underset{\sim}{\eta} \neq \underset{\sim}{0}$. The decision to include covariates in the analysis should depend on the result of such a test among other factors. Though both approaches discussed in Chapter 2 can be employed, the one proposed by Sen (1982) is more appealing since it avoids the fitting of the covariance adjusted model when the hypothesis $H:\underset{\sim}{\eta} = \underset{\sim}{0}$ is not rejected.

Finally note that as in the Normal theory case the decision to include the columns of $\underset{\sim}{Q}$ (or a subset of them) as covariates should depend on the covariances between the elements of each row of $\underset{\sim}{Z}$ and those of the corresponding row of $\underset{\sim}{Q}$, that is, on the elements of $\underset{\sim}{H}_2^T \underset{\sim}{\Sigma} \underset{\sim}{H}_2$. In practice $\underset{\sim}{\Sigma}$ is usually unknown and the guidelines suggested by Grizzle and Allen (1969) for the choice of covariates can also be employed, provided a robust and $n^{1/2}$-consistent estimate $\hat{\underset{\sim}{\Sigma}}$ is available.

3.3 M-methods in Kleinbaum's generalized multivariate linear models

Using Kleinbaum's (1970) vector versions of the MGLM and GGCM we can obtain M-estimators, M-tests and their asymptotic distributions by a direct application of the (univariate) results of Chapter 2.

Kleinbaum's vector version of the MGLM can be written as:

$$\underset{\sim}{Z} = \underset{\sim}{D}\,\underset{\sim}{\gamma} + \underset{\sim}{\zeta} \tag{3.3.1}$$

where $\underset{\sim}{Z} = (\underset{\sim}{z}_1^T, \ldots \underset{\sim}{z}_p^T)^T$, $\underset{\sim}{D} = \mathrm{diag}(\underset{\sim}{D}_1, \ldots \underset{\sim}{D}_p)$, $\underset{\sim}{\gamma} = (\underset{\sim}{\gamma}_1^T, \ldots \underset{\sim}{\gamma}_p^T)^T$, $\underset{\sim}{\zeta} = (\underset{\sim}{\zeta}_1^T, \ldots \underset{\sim}{\zeta}_p^T)^T$ and $\underset{\sim}{z}_j$, $\underset{\sim}{D}_j$, $\underset{\sim}{\gamma}_j$ and $\underset{\sim}{\zeta}_j$ respectively correspond to the vector of observations, design matrix, parameter vector and error vector for the j^{th} variate, $(j=1,\ldots p)$. The error vector $\underset{\sim}{\zeta}$ is assumed to have mean vector $\underset{\sim}{0}$ and scatter matrix:

$$\underset{\sim}{\Omega} = \begin{bmatrix} \sigma_{11}\underset{\sim}{I}_{n_1} & \sigma_{12}\underset{\sim}{Q}_{12} & \cdots & \sigma_{1p}\underset{\sim}{Q}_{1p} \\ \sigma_{12}\underset{\sim}{Q}_{12}^T & \sigma_{22}\underset{\sim}{I}_{n_2} & \cdots & \sigma_{2p}\underset{\sim}{Q}_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{1p}\underset{\sim}{Q}_{1p}^T & \sigma_{2p}\underset{\sim}{Q}_{2p}^T & \cdots & \sigma_{pp}\underset{\sim}{I}_{n_p} \end{bmatrix}$$

where $\underset{\sim}{Q}_{rs}$ denotes an incidence matrix of 0's and 1's defined by $\underset{\sim}{Q}_{rs} = \langle q_{ij}(rs) \rangle$, and:

$$q_{ij}(rs) = \begin{cases} 1 & \text{If the } i^{th} \text{ component of } \underset{\sim}{z}_r \text{ and the } j^{th} \text{ component of} \\ & \underset{\sim}{z}_s \text{ are observed on the same experimental unit;} \\ 0 & \text{otherwise.} \end{cases}$$

Assuming $\underset{\sim}{\Omega}$ known we can consider the transformation $\underset{\sim}{y}=\underset{\sim}{\Omega}^{-\frac{1}{2}}\underset{\sim}{Z}$, $\underset{\sim}{X}= \underset{\sim}{\Omega}^{-\frac{1}{2}}\underset{\sim}{D}$ and $\underset{\sim}{\varepsilon} = \underset{\sim}{\Omega}^{-\frac{1}{2}}\underset{\sim}{\zeta}$ so (3.3.1) reduces to (1.1.1) with $\underset{\sim}{\beta} = \underset{\sim}{\gamma}$ and $p = 1$. An M-estimator of $\underset{\sim}{\beta}$ can be obtained as a solution to (2.2.6) and its asymptotic distribution can be derived using Theorem 2.3.3 provided:

$$\lim_{n\to\infty}(n^{-1}\underset{\sim}{X}^T\underset{\sim}{X}) = \lim_{n\to\infty}(n^{-1}\underset{\sim}{D}^T\underset{\sim}{\Omega}^{-1}\underset{\sim}{D})=\underset{\sim}{V}, \text{ a p.d. matrix}$$

If $\underset{\sim}{\Omega}$ is unknown we can still define the M-estimator and obtain its asymptotic distribution given that a $n^{\frac{1}{2}}$-consistent estimate $\hat{\underset{\sim}{\Omega}}$ is available. We must be concerned, however, with the same problem faced by Kleinbaum in the Normal theory case regarding the positive-definiteness of the usual estimator (obtained by forming the pooled estimate of σ_{ij} from only those experimental units on which the i^{th} and j^{th} variates are observed simultaneously). Kleinbaum suggests some reasonable alternatives which include the estimation of $\underset{\sim}{\Omega}$ based on a complete subset of the data or on an independent sample. Though a general solution to this problem does not seem possible, it would be convenient to study conditions on the pattern of missing values or on the design matrices or on the covariance structure which would produce $n^{\frac{1}{2}}$-consistent and p.d. estimates of $\underset{\sim}{\Omega}$.

A similar procedure can be used to obtain M-estimators and their asymptotic distributions under the GGCM.

CHAPTER IV

ASYMPTOTIC COMPARISONS AND NUMERICAL EXAMPLE

4.1 Introduction

In this chapter we consider numerically the behaviour of two types of multivariate M-estimators with Huber's score function.

In Section 4.2 we compute the Pitman ARE of the coordinatewise M-estimator with respect to the Maronna-type M-estimator under different underlying distributions.

In Section 4.3 we propose algorithms for the computation of multivariate M-estimators; these algorithms are suggested by the asymptotic linearity results of Chapter 2 and correspond to approximations of the Newton-Raphson method.

Finally, in Section 4.4 we illustrate some of the results considered in this work by means of an example with real data.

4.2 Some asymptotic values

In this section we propose to analyze numerically the asymptotic behaviour of two families of robust M-estimators of the types considered in Chapter 2. In both cases we define the M-estimators by taking Huber's proposal (1.4.3) for the score function ψ.

Initially we consider the Pitman ARE of the coordinatewise M-estimators with respect to the Maronna-type ones. Furthermore, we compare the two families of M-estimators under the assumption of elliptical symmetry for the error distribution since this is a requirement in the derivation of the asymptotic distribution of the Maronna-type estimators.

Two types of (elliptically symmetric) distributions are proposed for the underlying model: the unit multivariate spherical Student distribution with g degrees of freedom, $ST_p(g)$, (g=1,2,3 and 5) and the unit multivariate spherical contaminated Normal distribution with contaminating proportion α and contaminating variance σ^2, $CN_p(\alpha,\sigma^2)$, (α = .05, .10; σ^2=4,9). Both families have been used in the literature to model the presence of outliers (or extreme observations) as in Maronna (1976) or Devlin, Gnanadesikan and Kettenring (1981) for example. The densities of the $ST_p(g)$ and $CN_p(\alpha,\sigma^2)$ distributions are respectively given by:

$$f(\underset{\sim}{\varepsilon})=\Gamma\{(g+p)/2\}\{(\pi g)^{p/2}\Gamma(g/2)\}^{-1}\{1+\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}/g\}$$

$$f(\underset{\sim}{\varepsilon})=(1-\alpha)(2\pi)^{-p/2}\exp(-\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}/2)+\alpha(2\pi\sigma^2)^{-p/2}\exp\{-\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}/(2\sigma^2)\}$$

The special case $ST_p(1)$ is known as the multivariate Cauchy distribution.

We included dimensions p = 2,3,4,5,6,10 and 20 in our study and to allow for some comparison among them we followed Maronna (1976) and parameterized the family of Maronna-type estimators with the number:

$$q = P\{\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon} \le k_M^2\}$$

where $\underset{\sim}{\varepsilon} \sim N_p(\underset{\sim}{0},\underset{\sim}{I})$ and k_M is the tuning constant in Huber's score function. The values chosen for q are .5160, .6826, .8664, .9544 and .9974 which respectively correspond to k_M = .7, 1.0, 1.5, 2.0 and 3.0 in the univariate case.

The tuning constant k_C for the coordinatewise M-estimator is determined in such a way that the volume of the "truncation" region is the same as for the corresponding Maronna-type M-estimator. Since all the proposed underlying distributions are spherical, this implies:

$$k_C = [\pi^{p/2}\{\Gamma(p/2+1)\}^{-1}]^{1/p}\ k_M/2\ .$$

Furthermore, since all the covariance matrices involved are of the form $c\underset{\sim}{I}$, where c is a scalar we only need to report the constant c, which we call the "variance" of the estimator as in Maronna (1976).

The results are presented in Tables 4.2.1 and 4.2.2. The entries are the ARE's obtained by dividing the asymptotic "variance" of the Maronna-type M-estimator by that of the corresponding coordinatewise M-estimator with the same "truncation volume". In parentheses we present the asymptotic "variance" of the coordinatewise M-estimator.

It can be seen from these results that both types of M-estimators perform equivalently well when the underlying distribution is spherical multivariate Normal. As the tails of the underlying distribution get heavier the ARE of the coordinatewise M-estimator with respect to its Maronna-type counterpart decreases, indicating, as expected, that the latter is preferable under the elliptical symmetry assumption. However, this loss of efficiency of the coordinatewise M-estimator is significant only for the extreme cases such as the $ST_p(1)$ or $ST_p(2)$ or for large dimensionalities.

Recalling that under the $ST_p(g)$ distribution the ML estimator has a "variance" given by $1+2/(g+p)$, it can be checked that the proposed estimators do not perform badly if the underlying distribution is of this form except in the extreme cases mentioned above.

Furthermore, since the asymptotic "variance" of the sample mean (NML estimator) is given by $g/(g-2)$ for $g > 2$ or ∞ for $g=1,2$ when the underlying distribution is the $ST_p(g)$ and by $(1-\alpha)+\alpha\sigma^2$ when the underlying distribution is the $CN_p(\alpha,\sigma^2)$ it is clear that both the proposed M-estimators perform better than it if instead of normality any of the alternatives considered here holds.

The problem of evaluating the ARE of the proposed estimators is computationally much more complicated if the underlying distribution though elliptical symmetric is not of the spherical type. We consider here only the unit bivariate Normal distribution with correlation coefficient ρ .

To compute the value $E\{\psi(X)\psi(Y)\}$ needed to obtain the covariance matrix of the asymptotic distribution of the coordinatewise M-estimators we used the following results, the derivation of which are due to Snappin (1983):

$$\int_{-\infty}^{k_2}\int_{-\infty}^{k_1} xg(x,y)\,dxdy = -\phi(k_1)\Phi[(k_2-\rho k_1)(1-\rho^2)^{-\frac{1}{2}}]-\rho\phi(k_2)\Phi[(k_1-\rho k_2)(1-\rho^2)^{-\frac{1}{2}}]$$

and:

$$\begin{aligned}\int_{-\infty}^{k_2}\int_{-\infty}^{k_1} xyg(x,y)\,dxdy = {} & \rho G(k_1,k_2;\rho)+(1-\rho^2)g(k_1,k_2;\rho)\\ & -\rho k_2\phi(k_2)\Phi[(k_1-\rho k_2)(1-\rho^2)^{-\frac{1}{2}}]\\ & -\rho k_1\phi(k_1)\Phi[(k_2-\rho k_1)(1-\rho^2)^{-\frac{1}{2}}]\end{aligned}$$

where ϕ and Φ respectively correspond to the density and distribution functions of the standard Normal distribution and g and G to the density and distribution functions of the unit bivariate Normal distribution with correlation coefficient ρ. Computer programs to evaluate $G(x,y;\rho)$ and $\Phi(x)$ are available, for example, in the IMSL package.

The values of the tuning constants were computed as previously; the values chosen for ρ were .25, .50 and .75. The Pitman ARE's were obtained by taking the square root of the ratio of the generalized variances of the asymptotic distributions of the proposed estimators. The results are presented in Table 4.2.3 and the values in parentheses correspond to the square root of the generalized variance of the asymptotic distribution of the coordinatewise M-estimator.

The performance of both types of M-estimators is still similar under this situation, though, as expected there is a slight loss of efficiency of the coordinatewise M-estimator as the value of ρ increases.

If the underlying distribution is not elliptically symmetric we do not know the asymptotic behaviour of the Maronna-type M-estimator. However, we can still obtain the Pitman ARE of the coordinatewise M-estimator with respect to the NML estimator by referring to Huber's (1967) result regarding the asymptotic behaviour of ML estimators when the assumed error distribution does not match the true one. We conjecture that a similar result must hold for the Maronna-type M-estimator though it has still to be proved.

As an example of a non-elliptically symmetric distribution to model the presence of outliers we considered a contaminated bivariate Normal distribution, the density function of which is given by:

$$f(\underset{\sim}{\varepsilon})=(1-\alpha)(2\pi)^{-1}\exp(-\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}/2)+\alpha(2\pi|\underset{\sim}{\Sigma}|^{\frac{1}{2}})^{-1}\exp(-\underset{\sim}{\varepsilon}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\varepsilon}/2) \qquad (4.2.1)$$

where $\underset{\sim}{\Sigma} = \sigma^2\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$.

The values of the tuning constant k_C, the contaminating proportion α, the contaminating variance σ^2 and the correlation coefficient ρ were those considered previously. The ARE's were obtained by taking the square root of the ratio of the generalized variance of the coordinatewise M-estimator to that of the NML estimator which is given by $|(1-\alpha)\underset{\sim}{I}+\alpha \underset{\sim}{\Sigma}|$. The results are presented in Table 4.2.4.

The values in Table 4.2.4 indicate that the coordinatewise M-estimator can be made (by an appropriate choice of the tuning constant) asymptotically more efficient than the NML estimator if the underlying distribution is given by (4.2.1). The relative efficiency increases as the contaminating proportion and the contaminating variance increase, indicating that the coordinatewise M-estimator can handle well distributions with somewhat heavier tails.

Though the actual computation of the ARE can become fairly difficult, it is easy to conceive of other situations where the coordinatewise M-estimator would perform much better than the NML-estimator. This would be the case if the marginal distribution of any of the coordinates were a heavy-tailed one.

4.3 Computational algorithms

In general the M-estimators defined by either (2.2.1) or (2.2.6) must be obtained by iterative methods. In this section we propose two computational algorithms which are suggested by the asymptotic linearity results given in Theorems 2.3.1 and 2.4.1.

Consider first the following algorithm for the coordinatewise case:

$$\begin{cases} \hat{\underset{\sim}{\beta}}^*_{(0)} = \tilde{\underset{\sim}{\beta}}^* \\ \hat{\underset{\sim}{\beta}}^*_{(m+1)} = \hat{\underset{\sim}{\beta}}^*_{(m)} + \{(\underset{\sim}{X}^T\underset{\sim}{X}) \otimes \hat{\underset{\sim}{W}}(\hat{\underset{\sim}{\beta}}_{(m)})\}^{-1} \underset{\sim}{M}^*_n(\hat{\underset{\sim}{\beta}}_{(m)}) \;, \; m \geq 0 \end{cases} \tag{4.3.1}$$

where:

$$\hat{W}(T) = \text{diag}\{\hat{w}_1(T),\ldots\hat{w}_p(T)\},\ \hat{w}_j(T) = (n\hat{\sigma}_j)^{-1}\sum_{k=1}^{n}\psi'\{(y_{kj}-x_k^T T_j)/\sigma_j\},$$

$$M_{nij}(T) = \sum_{k=1}^{n} x_{ki}\psi\{(y_{kj}-x_k^T T_j)/\hat{\sigma}_j\},\quad (i=1,\ldots r;\ j=1,\ldots p)$$

and $\tilde{\beta}$ and $\hat{\sigma}_j$ are $n^{\frac{1}{2}}$-consistent estimators of β and σ_j respectively.

Note that (4.3.1) is essentially an approximation to Newton's method. The choice of the initial value and the asymptotic linearity result insure that at each step we remain at the neighbourhood of β defined by $\{T: n^{\frac{1}{2}}||T-\beta||\le K, K > 0\}$ with probability arbitrarily close to 1. Convergence (in probability) of $\{\hat{\beta}_{(m)}, m\ge 0\}$ to $\hat{\beta}$ follows from the convergence of Newton's method; for a proof see Stoer and Bulirsch (1980) for example.

In practice we may stop iterating (4.3.1) when for some specified $\varepsilon > 0$, $||\hat{\beta}_{(m)}-\hat{\beta}_{(m-1)}||\le \varepsilon$.

A similar argument holds in the case of the Maronna-type M-estimator and the corresponding algorithm is given by:

$$\begin{cases} \hat{\beta}^*_{(0)} = \tilde{\beta}^* \\ \hat{\beta}^*_{(m+1)} = \hat{\beta}^*_{(m)} + \{(X^T X)\otimes \hat{b}_0(\hat{\beta}_{(m)})I\}^{-1} M^*_n(\hat{\beta}_{(m)}),\ m \ge 0 \end{cases} \tag{4.3.2}$$

where:

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

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

$$M_{nij}(T) = \sum_{k=1}^{n} u\{\hat{d}_k(T)\}x_{ki}(y_{kj}-x_k^T T_j),\quad (i=1,\ldots r;\ j=1,\ldots p)$$

and $\tilde{\beta}$ and $\hat{\Sigma}$ are $n^{\frac{1}{2}}$-consistent estimators of β and Σ respectively.

LS estimators can be taken as starting points for the iterations despite their known lack of robustness. It is preferable to start with more robust estimators such as the Least Absolute Residual estimator for $\underset{\sim}{\beta}$ and the Median Absolute Deviation for $\underset{\sim}{\sigma}$ though they are computationally more elaborate. In the case of Maronna's method robust estimates for $\underset{\sim}{\Sigma}$ are even more difficult to obtain; however, if the data correspond to the k-sample problem, robust estimates for each group can be obtained and then pooled to produce an overall estimate of $\underset{\sim}{\Sigma}$. A few alternative robust estimators of $\underset{\sim}{\Sigma}$ in the simple multivariate location problem are presented in Devlin, Gnanadesikan and Kettenring (1981).

In the coordinatewise case any of the univariate algorithms proposed in the literature can be used to obtain the estimates since we can work with each coordinate separately; this includes Huber's (1981, ch. 7) algorithm for simultaneous estimation of $\underset{\sim}{\beta}$ and σ.

4.4 Numerical example

In this section we consider the actual computation of M-estimates and M-tests in a practical example. Our intention is not to present any exhaustive analysis of the data but simply to illustrate an application of the M-methods studied in this work.

The algorithms of Section 4.3 were programmed using the MATRIX procedure of the Statistical Analysis System (SAS) and were employed to obtain both the coordinatewise and the Maronna-type M-estimates. In either case Huber's score function (1.4.3) was considered and the tuning constants were chosen in such a way to correspond to $k = 1.5$ in the univariate case, a value commonly used in the literature. The starting value for the iterative computation of both types of estimates was the LS estimate. In the coordinatewise case we used the median absolute deviation (MAD) multiplied by 1.48 (to make it approximately unbiased at the Normal model) as an estimate of scale; an estimate of the asymptotic covariance matrix of the coordinatewise M-estimate was used to estimate the scatter matrix in the Maronna-type case. The convergence criteria consisted of stopping the iterations when the Euclidean norm of the difference between the estimates from two consecutive steps was $< .001$.

The data we considered for our example was analyzed by Smith, Gnanadesikan and Hughes (1962) to illustrate the use of the classical MANOVA methods. We repeated part of their analysis using the robust M-methods described above.

The dataset consists of measurements on thirteen physico-chemical characteristics of complete morning urine samples from individuals classified into four weight groups. A subset of these variates is of special interest since their measurements can be obtained by "easy standard analyses"; we considered only these responses for our purposes. Two of the remaining variates were taken as covariables. Both the covariables and the responses are indicated below:

Covariables: V_1 = volume in mℓ
V_2 = (specific gravity-1) $\times 10^3$

Responses: Y_1 = pH
Y_2 = Calcium mg/mℓ
Y_3 = Creatinine mg/mℓ
Y_4 = Chloride mg/mℓ

Similarly to the above authors we considered the model:

$$Y_j = \mu_j + w_{ij}X_2 + w_{2j}X_3 + w_{3j}X_4 + \alpha_j V_1 + \gamma_j V_2 + \varepsilon_j \ , \quad (j=1,\ldots 4)$$

where μ_j is the overall mean effect, w_{ij} (i=1,2,3) are weight effects, α_j and γ_j are regression coefficients and ε_j is the error term. The matrix of observations corresponds to the 1st, 5th, 7th and 8th columns of their matrix $\underset{\sim}{Y}^T$ and both the design and parameter matrices were taken exactly as their $\underset{\sim}{A}$ and $\underset{\sim}{\xi}$ matrices. The number of observations was 45 and the values of the tuning constants for Huber's score function were k_M = 2.65 in the Maronna-type case and k_C = 1.98 in the coordinatewise setup.

The algorithm for computation of the coordinatewise M-estimate converged in 8 iterations; only 7 iterations were needed for the Maronna-type M-estimate. The results are presented in Table 4.4.1 along with the LS estimate for comparison; estimates of the corresponding asymptotic standard deviations are indicated in parentheses.

There are no significant differences among the three estimates, indicating 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 10 elements of the matrix of observations had weights <1, all being >.51; in the Maronna-type case 8 observations had weights <1 but all values were >.68.

For illustration purposes we introduced some "outliers" in the data by supposing that rows 28, 29 and 30 of the matrix of observations read:

	pH	Calcium	Creatinine	Chloride
row 28	0.5	9.003	9.06	4.26
row 29	0.5	6.014	2.26	0.65
row 30	0.5	6.008	0.14	8.48

instead of the original values. This situation could have occurred had the data been punched on cards and the above rows been misaligned by one column. We then recalculated the values of the three estimators using the same parameters as in the previous case. The results are indicated in Table 4.4.2.

As expected, the LS procedure was the one most affected by the changes: not only were the point estimates far off the values obtained with the original data but also the estimated standard deviations were considerably more inflated than the ones obtained through the two robust M-methods. Even though these robust procedures succeeded in detecting the "outliers" (their weights were of the order of .20) they were not as robust as we expected. We recomputed them using the scale estimates based on the original data and the results (not presented here) were much less sensible to the presence of the "outliers". This is an empirical indication that scale estimation might be an important factor in the computation of the robust M-estimators and should deserve some future research.

We also illustrate the proposed M-tests by considering the hypothesis of no difference between the weight groups which was of interest to the original investigators. We considered both the approaches discussed in Chapter 2 and the results are presented in Table 4.4.3. The (asymptotic) distribution of both Lawley-Hotelling's trace statistic and Wilks' likelihood ratio statistic conveniently normalized is χ^2_{12}; Roy's largest root statistic is asymptotically distributed as the largest root of a $W_4(3,\underset{\sim}{I},\cdot)$ distribution.

The p-values for all tests were <.05 and the inclusion of outliers in the data did not affect any of the results in a significant way.

Even though these empirical results do not indicate any aberrant behaviour we suggest that extreme care should be taken when applying the above multivariate M-tests, since not even in the classical Normal theory case there is a clear picture of their behaviour.

CHAPTER V

SUMMARY AND SUGGESTIONS FOR FUTURE RESEARCH

Robust methods are specially appealing in the analysis of large datasets or problems with complicated designs where detection of outliers by traditional methods (such as residual analysis) can be a difficult and time consuming task. Though many papers on related topics have recently appeared in the statistical literature, the increasing number of studies with the above characteristics can be viewed as an incentive for research in this area. The present study constitutes an attempt to address some robust methods in multivariate linear models and our conclusions are summarized below.

We considered two alternative M-estimators for the parameter matrix in the standard multivariate linear model: the first one is based on a coordinatewise scoring procedure while the second utilizes a scoring method based on the Mahalanobis distance and is an extension of Maronna's (1976) proposal for the multivariate location/scatter problem. The coordinatewise M-estimator is simpler to compute and the derivation of its asymptotic distribution does not require the assumption of elliptical symmetry of the underlying error distribution as in the case of the Maronna-type M-estimator. A comparison of the asymptotic efficiency of robust families (with Huber's score function) of both types of M-estimators under different (elliptically symmetric) error distributions indicated that even within this class the Maronna-type M-estimator outperforms the coordinatewise one only in very extreme cases. Algorithms for computation of the M-estimates were proposed and were employed in a practical example. In the case where no serious outliers are present both types of robust M-estimates behaved very similarly to the LS estimate; the presence of outliers in the data did not affect them as much but their performance seemed to depend on the scale estimates.

We also proposed two types of M-tests for the general linear hypothesis which are analogues of some classical Normal theory tests. Both were shown to be asymptotically equi-efficient for local Pitman-type alternatives and their empirical behaviour was similar to that of the M-estimates on which they are based. The analogue of Wald's test is computationally more convenient if one expects the null hypothesis to be rejected; otherwise we favour the second type of M-test which is an extension of Sen's (1982) proposal for the univariate case since it only requires the fitting of the reduced model.

Finally we indicated how these M-methods could be extended to the growth curve model and to some more general multivariate linear models which allow for missing data and/or different design matrices for different variates.

In general we feel that these results tend to favour the use of robust multivariate M-methods in practical applications. In particular we recommend them for cases where no other outlier detection method is employed (as in the use of most computer packages by non-statisticians) since they can prevent "disasters". They are also of interest as alternatives to outlier rejection rules because of their ability to downweight outlying observations smoothly instead of simply eliminating them. We feel, however, that there are still some important issues to be studied before these methods can be applied in a more widespread way. We enumerate below some of these topics along with others that should be considered for further research.

A. Study the effects of different estimates of scale and scatter on the performance of multivariate M-estimators; consider methods which allow for simultaneous estimation of the scatter parameter as well as of the regression parameters. Jurečková and Sen (1982) obtained results on the simultaneous M-estimation of the common location and scale-ratio in the two sample problem; Maronna (1976) considered the case of multivariate location and scatter.

B. Study the asymptotic behaviour of the Maronna-type M-estimator when the underlying distribution is not necessarily elliptically symmetric. A similar type of problem is that considered by Huber (1967) who studied the asymptotic behaviour of ML estimators when the assumed underlying distribution does not match the true one.

C. Study rates of convergence and the behaviour of multivariate M-estimators in linear models for small samples. This topic could be considered for a Monte-Carlo study though some theoretical results on the line of Field (1982) could also be attempted.

D. Study multivariate bounded-influence estimators, i.e. estimators which de-emphasize outliers in the independent variates as well as in the dependent ones. Some recent results for the univariate linear models are given in Maronna, Bustos and Yohai (1979) and Krasker and Welsch (1982).

E. Study robust estimation with respect to departures from the homoscedasticity and independence assumption in the multivariate case. Robust estimation in heteroscedastic univariate linear models has been considered in Carroll and Ruppert (1982) for example.

F. Obtain the asymptotic distribution of the M-estimator proposed by Pendergast and Broffitt (1981) for the growth curve model. We conjecture that Jurečková's asymptotic linearity result would also apply in this case.

G. Study the effect of different choices for the matrix defining the Potthoff-Roy transformation in the analysis of growth curve models. This topic would probably involve some simulation study.

Table 4.2.1 A.R.E. of the coordinatewise M-estimator with respect to the corresponding Maronna-type M-estimator and asymptotic "variance" of the coordinatewise M-estimator.

p	q	$ST_p(1)$	$ST_p(2)$	$ST_p(3)$	$ST_p(5)$	N_p
2	.5160	.8699(2.5993)	.9207(1.7691)	.9429(1.5246)	.9628(1.3424)	.9955(1.0939)
	.6826	.8699(2.8403)	.9243(1.8452)	.9478(1.5561)	.9682(1.3421)	.9984(1.0529)
	.8664	.8711(3.2799)	.9302(2.0019)	.9551(1.6377)	.9754(1.3714)	.9988(1.0188)
	.9544	.8726(3.7472)	.9356(2.1727)	.9616(1.7317)	.9817(1.4139)	.9991(1.0058)
	.9974	.8753(4.7476)	.9444(2.5203)	.9718(1.9172)	.9905(1.4965)	.9999(1.0003)
3	.5160	.8056(2.7660)	.8787(1.8204)	.9124(1.5447)	.9434(1.3402)	.9956(1.0630)
	.6826	.8056(3.0061)	.8834(1.9029)	.9188(1.5849)	.9505(1.3505)	.9976(1.0359)
	.8664	.8069(3.4164)	.8911(2.0520)	.9289(1.6651)	.9610(1.3834)	.9979(1.0135)
	.9544	.8085(3.8383)	.8983(2.2056)	.9380(1.7499)	.9701(1.4223)	.9986(1.0046)
	.9974	.8117(4.7277)	.9102(2.5137)	.9525(1.9138)	.9834(1.0169)	.9999(1.0003)
4	.5160	.7669(2.8816)	.8522(1.8593)	.8927(1.5629)	.9307(1.3438)	.9954(1.0481)
	.6826	.7669(3.1130)	.8573(1.9412)	.8999(1.6049)	.9388(1.3579)	.9969(1.0280)
	.8664	.7680(3.4963)	.8657(1.1552)	.9111(1.6812)	.9510(1.3907)	.9973(1.0110)
	.9544	.7696(3.8837)	.8734(1.1449)	.9213(1.7588)	.9616(1.4264)	.9982(1.0040)
	.9974	.7726(4.6929)	.8867(1.1278)	.9381(1.9079)	.9774(1.4927)	.9997(1.0004)
5	.5160	.7409(2.9664)	.8339(1.8889)	.8789(1.5777)	.9216(1.3481)	.9951(1.0394)
	.6826	.7408(3.1879)	.8391(1.9684)	.8864(1.6195)	.9302(1.3638)	.9963(1.0234)
	.8664	.7418(3.5483)	.8476(2.1002)	.8981(1.6917)	.9433(1.3955)	.9969(1.0097)
	.9544	.7432(3.9084)	.8554(2.2308)	.9088(1.7637)	.9548(1.4287)	.9980(1.0038)
	.9974	.7460(4.6560)	.8692(2.4899)	.9267(1.9016)	.9722(1.4900)	.9996(1.0004)
6	.5160	.7222(3.0316)	.8204(1.9120)	.8685(1.5896)	.9147(1.3522)	.9948(1.0339)
	.6826	.7220(3.2437)	.8255(1.9887)	.8762(1.6305)	.9237(1.3683)	.9959(1.0205)
	.8664	.7228(3.5845)	.8339(2.1135)	.8881(1.6991)	.9373(1.3989)	.9966(1.0088)
	.9544	.7240(3.9224)	.8418(2.2359)	.8990(1.7664)	.9492(1.4300)	.9978(1.0036)
	.9974	.7267(4.6206)	.8557(2.4781)	.9174(1.8955)	.9677(1.4874)	.9996(1.0005)

Table 4.2.1: Continued

p	q	$ST_p(1)$	$ST_p(2)$	$ST_p(3)$	$ST_p(5)$	N_p
10	.5160	.6802(3.1907)	.7888(1.9694)	.8445(1.6200)	.8985(1.3640)	.9938(1.0233)
	.6826	.6799(3.3729)	.7940(2.0360)	.8518(1.6563)	.9076(1.3795)	.9947(1.0150)
	.8664	.6803(3.6583)	.8016(2.1404)	.8632(1.7139)	.9237(1.4057)	.9959(1.0073)
	.9544	.6810(3.9364)	.8087(2.2409)	.8737(1.7692)	.9337(1.4312)	.9973(1.0035)
	.9974	.6828(4.5038)	.8219(2.4388)	.8923(1.8751)	.9537(1.4787)	.9993(1.0007)
20	.5160	.6431(3.3494)	.7608(2.0274)	.8220(1.6516)	.8831(1.3775)	.9926(1.0159)
	.6826	.6427(3.4904)	.7645(2.0791)	.8281(1.6801)	.8911(1.3902)	.9936(1.0112)
	.8664	.6425(3.7064)	.7703(2.1579)	.8375(1.7236)	.9033(1.4102)	.9951(1.0065)
	.9544	.6427(3.9132)	.7760(2.2325)	.8463(1.7646)	.9143(1.4291)	.9967(1.0037)
	.9974	.6436(4.3291)	.7867(2.3790)	.8625(1.8436)	.9335(1.4649)	.9989(1.0012)

Table 4.2.2: A.R.E. of the coordinatewise M-estimator with respect to the corresponding Maronna-type M-estimator and asymptotic "variance" of the coordinatewise M-estimator.

p	q	$CN_p(.05,4)$	$CN_p(.05,9)$	$CN_p(.10,4)$	$CN_p(.10,9)$	N_p
2	.5160	.9890(1.0700)	.9831(1.2024)	.9830(1.2518)	.9718(1.3228)	.9955(1.0939)
	.6826	.9915(1.1325)	.9840(1.1696)	.9853(1.2177)	.9711(1.2989)	.9984(1.0529)
	.8664	.9920(1.1072)	.9820(1.1554)	.9863(1.1209)	.9680(1.3056)	.9988(1.0188)
	.9544	.9934(1.1053)	.9816(1.1687)	.9889(1.2097)	.9680(1.3459)	.9991(1.0058)
	.9974	.9977(1.1229)	.9856(1.2264)	.9961(1.2485)	.9760(1.4663)	.9999(1.0003)
3	.5160	.9845(1.1413)	.9744(1.1769)	.9744(1.2254)	.9553(1.3032)	.9956(1.0630)
	.6826	.9859(1.1185)	.9732(1.1595)	.9757(1.2067)	.9519(1.2960)	.9976(1.0359)
	.8664	.9866(1.1050)	.9700(1.1573)	.9771(1.2017)	.9470(1.3150)	.9979(1.0135)
	.9544	.9891(1.1064)	.9694(1.1730)	.9815(1.2128)	.9470(1.3558)	.9986(1.0046)
	.9974	.9957(1.1225)	.9745(1.2251)	.9993(1.2478)	.9575(1.4639)	.9999(1.0003)
4	.5160	.9809(1.1283)	.9678(1.1664)	.9679(1.2142)	.9433(1.2974)	.9954(1.0481)
	.6826	.9817(1.1127)	.9654(1.1564)	.9685(1.2030)	.9384(1.2980)	.9969(1.0280)
	.8664	.9827(1.1045)	.9616(1.1593)	.9704(1.2030)	.9325(1.3215)	.9973(1.0110)
	.9544	.9857(1.1070)	.9608(1.1753)	.9757(1.2145)	.9320(1.3609)	.9982(1.0040)
	.9974	.9939(1.1219)	.9657(1.2230)	.9895(1.2465)	.9426(1.4596)	.9997(1.0004)
5	.5160	.9780(1.1212)	.9627(1.1613)	.9628(1.2087)	.9342(1.2960)	.9951(1.0394)
	.6826	.9785(1.1098)	.9597(1.1554)	.9629(1.2015)	.9284(1.3008)	.9963(1.0234)
	.8664	.9796(1.1044)	.9555(1.1609)	.9651(1.2041)	.9220(1.3262)	.9969(1.0097)
	.9544	.9830(1.1073)	.9544(1.1765)	.9710(1.2154)	.9210(1.3638)	.9980(1.0038)
	.9974	.9920(1.1212)	.9586(1.2207)	.9863(1.2451)	.9304(1.4551)	.9996(1.0004)
6	.5160	.9757(1.1169)	.9586(1.1585)	.9587(1.2056)	.9271(1.2962)	.9948(1.0339)
	.6826	.9759(1.1081)	.9552(1.1553)	.9586(1.2010)	.9208(1.3035)	.9959(1.0205)
	.8664	.9771(1.1044)	.9509(1.1621)	.9609(1.2050)	.9141(1.3295)	.9966(1.0088)
	.9544	.9807(1.1075)	.9496(1.1773)	.9669(1.2159)	.9125(1.3654)	.9978(1.0036)
	.9974	.9903(1.0098)	.9529(1.2185)	.9832(1.2437)	.9205(1.4507)	.9996(1.0005)

Table 4.2.2: Continued

p	q	$CN_p(.05,4)$	$CN_p(.05,9)$	$CN_p(.10,4)$	$CN_p(.10,9)$	N_p
10	.5160	.9694(1.1097)	.9483(1.1554)	.9480(1.2015)	.9094(1.3009)	.9938(1.0233)
	.6826	.9693(1.1055)	.9447(1.1565)	.9475(1.2013)	.9029(1.3117)	.9947(1.0150)
	.8664	.9707(1.1047)	.9404(1.1649)	.9497(1.2069)	.8958(1.3367)	.9959(1.0073)
	.9544	.9742(1.1077)	.9384(1.1780)	.9554(1.2164)	.8930(1.3670)	.9973(1.0035)
	.9974	.9839(1.1063)	.9387(1.2113)	.9722(1.2392)	.8956(1.4362)	.9993(1.0007)
20	.5160	.9628(1.1058)	.9378(1.1561)	.9371(1.2011)	.8917(1.3101)	.9926(1.0159)
	.6826	.9627(1.1045)	.9347(1.1591)	.9365(1.2029)	.8862(1.3210)	.9936(1.0112)
	.8664	.9639(1.1050)	.9312(1.1669)	.9379(1.2084)	.8801(1.3416)	.9951(1.0065)
	.9544	.9665(1.1074)	.9291(1.1768)	.9418(1.0618)	.8768(1.3643)	.9967(1.0037)
	.9974	.9738(1.1148)	.9269(1.0788)	.9541(1.2322)	.8744(1.4145)	.9989(1.0012)

Table 4.2.3: ARE of the coordinatewise M-estimator with respect to the corresponding Maronna-type M-estimator and square root of the generalized variance of the coordinatewise M-estimator (underlying distribution: unit bivariate Normal distribution with correlation coefficient ρ).

	ρ		
q	.25	.50	.75
.5160	.9931(1.0647)	.9745(.9678)	.9460(.7614)
.6826	.9953(1.0226)	.9859(.9234)	.9691(.7175)
.8664	.9977(.9876)	.9941(.8865)	.9875(.6816)
.9544	.9987(.9742)	.9975(.8724)	.9952(.6679)
.9974	.9999(.9686)	.9998(.8664)	.9996(.6618)

Table 4.2.4: ARE of the coordinatewise M-estimator with respect to the NML estimator and square root of the generalized variance of the coordinatewise M-estimator. (Underlying distribution: contaminated bivariate Normal with density (4.2.1))

		$\alpha = .05$		$\alpha = .10$	
ρ	q	$\sigma^2 = 4$	$\sigma^2 = 9$	$\sigma^2 = 4$	$\sigma^2 = 9$
.25	.5160	.9821(1.1698)	1.1608(1.2022)	1.0359(1.2513)	1.3507(1.3222)
	.6826	1.0147(1.1323)	1.1933(1.1694)	1.0650(1.2170)	1.3758(1.2981)
	.8664	1.0380(1.1069)	1.2081(1.1551)	1.0802(1.1999)	1.3693(1.3043)
	.9544	1.0398(1.1049)	1.1946(1.1681)	1.0728(1.2082)	1.3290(1.3438)
	.9974	1.0238(1.1222)	1.1391(1.2251)	1.0402(1.2460)	1.2217(1.4617)
.50	.5160	.9797(1.1694)	1.1498(1.2017)	1.0279(1.2496)	1.3200(1.3203)
	.6826	1.0123(1.1318)	1.1823(1.1688)	1.0573(1.2149)	1.3453(1.2955)
	.8664	1.0358(1.1061)	1.1974(1.1540)	1.0734(1.1967)	1.3407(1.3000)
	.9544	1.0379(1.1038)	1.1847(1.1663)	1.0672(1.2037)	1.3036(1.3369)
	.9974	1.0227(1.1202)	1.1316(1.2211)	1.0373(1.2384)	1.2042(1.4473)
.75	.5160	.9757(1.1686)	1.1315(1.2008)	1.0149(1.2464)	1.2675(1.3164)
	.6826	1.0084(1.1308)	1.1638(1.1675)	1.0447(1.2107)	1.2933(1.2903)
	.8664	1.0322(1.1046)	1.1796(1.1518)	1.0623(1.1907)	1.2921(1.2915)
	.9544	1.0350(1.1016)	1.1684(1.1629)	1.0582(1.1954)	1.2606(1.3237)
	.9974	1.0211(1.1166)	1.1195(1.2137)	1.0324(1.2252)	1.1746(1.4206)

Table 4.4.1: Least squares (LS), coordinatewise (CO) and Maronna-type (MA) M-estimates of the parameter matrix for the Smith, Gnanadesikan and Hughes (1962) data (with estimates of the asymptotic standard deviations in parentheses).

		pH	Calcium	Creatinine	Chloride
μ	LS	5.7897(0.3364)	-0.1565(0.0958)	2.7392(0.7058)	-3.1415(1.7161)
	CO	5.8339(0.3280)	-0.1559(0.0861)	2.3524(0.6128)	-3.1583(1.6306)
	MA	5.7699(0.3183)	-0.1564(0.0836)	2.4559(0.5947)	-2.9201(1.5824)
w_1	LS	0.1948(0.0732)	0.0028(0.0209)	-0.3694(0.1537)	0.9146(0.3736)
	CO	0.1626(0.0714)	0.0029(0.0187)	-0.2995(0.1334)	0.8182(0.3550)
	MA	0.1978(0.0693)	0.0002(0.0182)	-0.3243(0.1295)	0.8429(0.3444)
w_2	LS	-0.0738(0.0700)	-0.0179(0.0199)	-0.2473(0.1468)	0.6506(0.3570)
	CO	-0.0578(0.0622)	-0.0252(0.0179)	-0.1972(0.1275)	0.6817(0.3392)
	MA	-0.0672(0.0662)	-0.0220(0.0174)	-0.2070(0.1237)	0.5975(0.3292)
w_3	LS	0.0126(0.0846)	-0.0678(0.0241)	0.0876(0.1774)	-0.3111(0.4314)
	CO	0.0389(0.0824)	-0.0678(0.0216)	0.0860(0.1540)	-0.2779(0.4099)
	MA	0.0278(0.0800)	-0.0692(0.0210)	0.0830(0.1495)	-0.2667(0.3978)
α	LS	0.0001(0.0005)	0.0003(0.0002)	-0.0052(0.0011)	0.0049(0.0027)
	CO	-0.0001(0.0005)	0.0003(0.0001)	-0.0046(0.0010)	0.0049(0.0026)
	MA	0.0002(0.0005)	0.0003(0.0001)	-0.0048(0.0010)	0.0046(0.0025)
γ	LS	-0.0130(0.0090)	0.0098(0.0026)	0.0418(0.0189)	0.2695(0.0460)
	CO	-0.0128(0.0088)	0.0097(0.0023)	0.0487(0.0164)	0.2690(0.0437)
	MA	-0.0131(0.0085)	0.0098(0.0022)	0.0477(0.0160)	0.2633(0.0425)

Table 4.4.2: Least squares (LS), coordinatewise (CO) and Maronna-type (MA) M-estimates of the parameter matrix for the Smith, Gnanadesikan and Hughes (1962) data with outliers (estimates of the asymptotic standard deviations in parentheses).

		pH	Calcium	Creatinine	Chloride
μ	LS	3.4718(1.4274)	3.5356(1.9323)	5.8472(1.5930)	-2.8251(2.1633)
	CO	5.0210(0.5418)	0.7986(0.6034)	3.5322(0.9024)	-2.6922(1.9950)
	MA	5.1897(0.4579)	0.6136(0.5100)	3.3783(0.7626)	-3.5533(1.6859)
w_1	LS	0.5712(0.3107)	-0.4816(0.4206)	-0.4853(0.3468)	1.0288(0.4709)
	CO	0.3088(0.1180)	-0.1413(0.1314)	-0.3772(0.1964)	0.8325(0.4343)
	MA	0.3009(0.0997)	-0.1203(0.1110)	-0.4171(0.1660)	0.9388(0.3670)
w_2	LS	0.4065(0.2969)	-0.7224(0.4019)	-0.6684(0.3314)	0.5270(0.4500)
	CO	0.0792(0.1127)	-0.2125(0.1255)	-0.3375(0.1877)	0.5408(0.4150)
	MA	0.0578(0.0952)	-0.1799(0.1061)	-0.3482(0.1586)	0.6980(0.3507)
w_3	LS	-1.1939(0.3588)	1.5489(0.4857)	0.6620(0.4004)	-0.4770(0.5438)
	CO	-0.3585(0.1362)	0.4021(0.1517)	0.1765(0.2268)	-0.2112(0.5015)
	MA	-0.3282(0.1151)	0.3293(0.1282)	0.2757(0.1917)	-0.4443(0.4238)
α	LS	0.0015(0.0023)	-0.0014(0.0031)	-0.0058(0.0025)	0.0068(0.0034)
	CO	0.0006(0.0009)	-0.0002(0.0010)	-0.0054(0.0014)	0.0059(0.0032)
	MA	0.0005(0.0007)	-0.0001(0.0008)	-0.0055(0.0012)	0.0054(0.0027)
γ	LS	0.0486(0.0383)	-0.0962(0.0518)	-0.0655(0.0427)	0.2376(0.0580)
	CO	0.0081(0.0145)	-0.0166(0.0162)	0.0134(0.0242)	0.2422(0.0535)
	MA	0.0022(0.0123)	-0.0112(0.0137)	0.0217(0.0205)	0.2781(0.0452)

Table 4.4.3: Asymptotic M-tests of the hypothesis of equal weight groups for the Smith, Gnanadesikan and Hughes (1962) data. (p-values in parentheses).

	Wald's tests		Sen's tests	
	Coordinatewise	Maronna	Coordinatewise	Maronna
WLR	38.7399(.0001)	42.8978(.0000)	25.3410(.0133)	26.8829(.0080)
LHT	53.7684(.0000)	61.0967(.0000)	29.7837(.0030)	31.9308(.0014)
RLR	40.8146(<.01)	45.7450(<.01)	17.8355(<.05)	19.3229(<.05)

WLR: Wilks' likelihood ratio

LHT: Lawley-Hotelling's trace

RLR: Roy's largest root.

Table 4.4.4: Asymptotic M-tests of the hypothesis of equal weight groups for the Smith, Gnanadesikan and Hughes (1962) data with outliers (p-values in parentheses).

	Wald's tests		Sen's tests	
	Coordinatewise	Maronna	Coordinatewise	Maronna
WLR	29.5552(.0033)	38.5817(0.0001)	22.7199(0.0302)	27.2291(0.0072)
LHT	38.2775(.0001)	55.3948(0.0000)	26.9506(0.0079)	33.5761(0.0008)
RLR	29.2545(<.01)	45.0519(<.01)	18.8965(<.05)	24.0245(<.01)

WLR: Wilks' likelihood ratio

LHT: Lawley-Hotelling's trace

RLR: Roy's largest root.

REFERENCES

1. Arnold, S.F. (1981). The Theory of Linear Models and Multivariate Analysis, Wiley, New York.

2. Bickel, P.J. (1975). One step Huber estimates in the linear model. J. Amer. Statist. Assoc., 70, 428-434.

3. Carroll, R.J. (1978). On the asymptotic distribution of multivariate M-estimates. J. Multivariate Anal., 8, 361-371.

4. Carroll, R.J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. Ann. Statist., 10, 429-441.

5. Collins, J.R. (1982). Robust M-estimators of location vectors. J. Multivariate Anal., 12, 480-492.

6. Cramér, H. (1946). Mathematical Methods of Statistics, Princeton University Press, New Jersey.

7. 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.

8. Eicker, F. (1967). Limit theorems for regression with unequal and dependent errors, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 59-82. Univ. of California Press, Berkeley.

9. Field, C. (1982). Small sample asymptotic expansions for multivariate M-estimates. Ann. Statist., 10, 672-689.

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

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

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

13. Hampel, F.R. (1974). The influence curve and its role in robust estimation. J. Amer. Statist. Assoc., 62, 1179-1186.

14. Holland, P.W. and Welsch, R.E. (1977). Robust regression using iteratively reweighted least squares. Comm. Statist., A6, 813-827.

15. Huber, P.J. (1967). The behaviour of maximum likelihood estimates under nonstandard conditions, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 221-233, Univ. of California Press, Berkeley.

16. Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte-Carlo. Ann. Statist., 1, 799-821.

17. Huber, P.J. (1981). Robust Statistics, Wiley, New York.

18. IMSL Library, Edition 9 (1982), IMSL, Inc., Houston, chapter M.

19. Ito, P.K. (1980). Robustness of ANOVA and MANOVA test procedures, in Handbook of Statistics, vol. 1, Krishnaiah P.R. editor, North Holland, Amsterdam, 199-236.

20. Jurečková, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. Ann. Math. Statist., 40, 1889-1900.

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

22. Jurečková, 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.

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

24. Kleinbaum, D.G. (1970). Estimation and hypothesis testing for generalized multivariate linear models. Institute of Statistics Mimeo Series No. 696, University of North Carolina, Chapel Hill.

25. Krasker, W.S. and Welsch, R.E. Efficient bounded-influence regression estimation. J. Amer. Statist. Assoc. 77, 595-604.

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

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

28. 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, 91-116, Springer-Verlag, Heidelberg.

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

30. Pearson, E.S. and Hartley, H.O. (1976). Biometrika Tables for Statisticians, vol. 2, Cambridge University Press, Cambridge, England.

31. 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).

32. 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.

33. Rao, C.R. (1967). Least squares theory using an estimated dispersion matrix and its applications to measurement of signals, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 355-372, Univ. of California Press, Berkeley.

34. SAS User's Guide, 1979 Edition, Helwig, J.T. and Council, K.A. editors, SAS Institute Inc., Raleigh, N.C.

35. Scheffé, H. (1959). The Analysis of Variance, Wiley, New York.

36. Schrader, R.M. and Hettmansperger, T.P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. Biometrika, 67, 93-101.

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

38. Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics, Wiley, New York.

39. Smith, H., Gnanadesikan, R. and Hughes, J.B. (1962). Multivariate analysis of variance (MANOVA). Biometrics, 18, 22-41.

40. Snappin, S. (1983). An evaluation of smoothed error rate estimators in discriminant analysis. Ph.D. dissertation. Department of Biostatistics - UNC Chapel Hill.

41. Stoer, J. and Bulirsch, R. (1980). Introduction to Numerical Analysis, Springer-Verlag, Heidelberg.

42. Sverdrup, E. (1952). The limit distribution of a continuous function of random variables. Skandinavisk Aktuarietidskrift, 35, 1-10.

43. Yohai, V.J. and Maronna, R.A. (1979). Asymptotic behavior of M-estimators for the linear model. Ann. Statist., 7, 258-268.