Robust Testing in Linear Models: The Infinitesimal Approach

RONCHETTI, Elvezio

DOI: 10.13097/archive-ouverte/unige:24845
ROBUST TESTING IN LINEAR MODELS:
THE INFINITESIMAL APPROACH

DISSE异TATION
submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZUE异CH
for the degree of
DOCTOR OF MATHEMATICS

presented by
ELVEZIO RONCHETTI
Dipl. Math. ETH
born on November 4, 1955
citizen of Monte TI, Switzerland

Accepted on the recommendation of
Prof. Dr. F. Hampel, referee
Prof. Dr. I. Olkin, co-referee
Prof. Dr. C. Field, co-referee

1982
Ai miei genitori
I
E. RONCHETTI

ROBUST TESTING IN LINEAR MODELS:
THE INFINITESIMAL APPROACH

Abstract

The infmitesimal approach to the theory of robust statistics is extended to robust testing. By means of the influence function and the change-of-power function we investigate the properties of infinitesimal robustness of test procedures and we establish an optimality criterion for infinitesimal robustness of tests. This formalism allows us to construct optimally robust tests that maximize the asymptotic power, within a certain class, under the side condition of a bounded influence function.

Further a new infinitesimal concept, the change-of-variance function, is discussed both for the location and the regression model, and the following optimality problem is solved: we find the estimator that minimizes the trace of the asymptotic covariance matrix, within a certain class, under the side condition of a bounded change-of-variance function.

These infinitesimal methods are used for constructing a robust version of the classical F-test for linear models. To accomplish this, we introduce a new general class of test procedures for linear models and compute their influence function and asymptotic power. Then, solving the optimality problem for infinitesimal robustness of tests, we find the
optimally robust test in this class.

One can achieve a robust version of the $F$-test in two steps. First, one looks for the optimally robust test with respect to the influence of the residual ($\rho_c$-test). Secondly, taking into account the influence of position in factor space, one imposes the condition of a bounded total influence, obtaining optimally bounded influence tests which are the natural counterpart to optimally bounded influence estimators.

In this framework we also propose a robust procedure for the selection of the variables.

An alternative class of procedures (called $C(\alpha)$-type tests for linear models) is introduced. It is shown that they are equivalent to the former ones, having the same influence function and the same asymptotic efficiency. Moreover, an interesting connection between the optimally robust $C(\alpha)$-type test and an asymptotically minimax test is shown.

Finally, we compare the asymptotic behaviour of several test procedures under different distributions and we compute two examples based on real data; both show the excellent performance of our optimally robust tests.
Zusammenfassung


Ein weiterer neuer infinitesimaler Begriff, die "Varianzänderungsfunktion" (change-of-variance function), wird für das Lokations- und das Regressionsmodell diskutiert und je das folgende Optimalitätsproblem gelöst: wir finden den Schätzer, welcher unter der Nebenbedingung, dass die "Varianzänderungsfunktion" beschränkt sei, die Spur der asymptotischen Kovarianzmatrix in einer gewissen Klasse von Schätzern minimiert.

Diese infinitesimalen Methoden werden dazu benützt, eine robuste Version des klassischen F-Tests für lineare Modelle zu konstruieren. Das wird folgendermassen erreicht.
Wir führen eine neue allgemeine Klasse von Testverfahren für
lineare Modelle ein und berechnen ihre Einflussfunktion und ihre asymptotische Macht. Dann lösen wir das Optimalitätsproblem für die infinitesimale Robustheit von Tests und finden so den optimalen robusten Test in dieser Klasse.
Diese Robustifizierung des F-Tests wird in zwei Schritten durchgeführt. Im ersten Schritt sucht man ein optimales robustes Testverfahren bezüglich des Einflusses auf das Residuum ($\rho_c$-Test).
In einer zweiten Phase betrachtet man auch den Einfluss auf die Lage im Faktorraum und verlangt, dass der totale Einfluss beschränkt sei. Auf diese Weise erhält man optimale robuste Tests, welche den optimalen robusten Schätzern entsprechen. Ferner schlagen wir in diesem Zusammenhang ein robustes Verfahren für die Variablenauswahl vor.
Eine alternative Klasse von Verfahren (Tests vom Typus C($\alpha$) für lineare Modelle) wird eingeführt. Es wird gezeigt, dass diese Tests zu den obigen äquivalent sind, da sie die gleiche Einflussfunktion und die gleiche asymptotische Effizienz besitzen. Ferner wird eine interessante Beziehung zwischen dem optimalen robusten Test vom Typus C($\alpha$) und einem asymptotischen minimax Test gezeigt.
Am Ende vergleichen wir das asymptotische Verhalten von verschiedenen Testverfahren unter verschiedenen Verteilungen und berechnen zwei auf realen Daten basierende Beispiele: beide zeigen die vorzügliche Leistung unserer optimalen robusten Tests.
Résumé

L'application de techniques infinitésimales dans la statistique robuste est généralisée à la théorie des tests robustes. Nous étudions les propriétés de robustesse des tests en utilisant la fonction d'influence et la "fonction d'influence sur la puissance" (change-of-power function) et formulons un critère d'optimalité pour la robustesse infinitésimale des tests. Ce formalisme nous permet de construire des tests robustes optimaux, qui maximisent la puissance asymptotique parmi les tests ayant une fonction d'influence bornée.

Une autre notion infinitésimale, la "fonction d'influence sur la variance" (change-of-variance function), est définie pour le modèle d'un paramètre de position et pour le modèle de la régression linéaire. On résout également le problème d'optimalité suivant: on trouve l'estimateur qui minimise la trace de la matrice de covariance asymptotique parmi les estimateurs ayant une "fonction d'influence sur la variance" bornée.

Ces méthodes infinitésimales sont employées pour construire une version robuste du F-test classique pour les modèles linéaires. Cela s'effectue de la manière suivante: on introduit une classe nouvelle et générale de tests pour les modèles linéaires et on calcule la fonction d'influence et la
puissance asymptotique. Puis, on résout le problème d'optimalité pour la robustesse infinitésimale des tests en trouvant le test robuste optimal dans cette classe.
On peut obtenir cette version robuste du F-test en deux étapes. Premièrement on cherche le test robuste optimal par rapport à l'influence du résidu ($\rho$-test). Deuxièmement on considère l'influence de la position dans l'espace des facteurs et on impose la condition d'avoir une influence totale bornée. On obtient alors les tests robustes optimaux qui correspondent aux estimateurs robustes optimaux.
Dans ce contexte on propose un critère robuste pour la sélection des variables.

On introduit de plus une classe alternative de tests (tests du type $C(\alpha)$ pour les modèles linéaires). On montre qu'ils sont équivalent aux tests qu'on a introduits auparavant car ils ont la même fonction d'influence et la même efficacité asymptotique. On montre aussi une relation intéressante parmi le test robuste optimal du type $C(\alpha)$ et un test asymptotiquement minimax.

Finalement on compare le comportement asymptotique des tests sous divers distributions et on calcule deux exemples réels: l'un et l'autre montrent l'excellente performance de nos tests robustes optimaux.
VII
E. RONCHETTI
TEST ROBUSTI PER I MODELLI LINEARI E
TECNICHE INFINITESIMALI

Riassunto

L'impiego di tecniche infinitesimali nel campo della statistica robusta è generalizzato alla teoria dei test robusti. Studiamo le proprietà di robustezza dei test per mezzo della funzione d'influenza e della "funzione d'influenza sulla potenza" (change-of-power function) e formuliamo un criterio di ottimalità per la robustezza infinitesimale dei test. Questo formalismo ci permette di costruire i test robusti ottimali che massimizzano la potenza asintotica in una classe di test con la funzione d'influenza limitata.

Un altro concetto infinitesimale, la "funzione d'influenza sulla varianza" (change-of-variance function), viene discusso per il modello di un parametro di localizzazione e per quello della regressione lineare. Viene inoltre risolto il problema seguente: si costruisce la stima che minimizza la traccia della matrice di covarianza asintotica in una classe di stime con la "funzione d'influenza sulla varianza" limitata.

Questi metodi infinitesimali vengono usati per costruire una versione robusta del test F classico per i modelli lineari. Per ottenere ciò viene introdotta una nuova classe di test per i modelli lineari e viene calcolata la loro funzione d'influenza e la loro potenza asintotica. Poi si risolve il problema di
ottimalità per la robustezza infinitesimale dei test, trovando così il test robusto ottimale in questa classe.
È possibile ottenere questa versione robusta del test F in due stadi. Dapprima si cerca il test robusto ottimale rispetto alla influenza del residuo (test $\rho_C$). In seguito, tenendo conto dell'influenza della posizione nello spazio dei fattori, si impone la condizione di avere un'influenza totale limitata. Si ottengono così test robusti ottimali che corrispondono alle stime robuste ottimali.
In questo contesto si propone pure un criterio robusto per la selezione delle variabili.

Si introduce inoltre una classe alternativa di test (test del tipo C($\alpha$) per i modelli lineari). Si dimostra che essi sono equivalenti a quelli introdotti precedentemente in quanto possiedono la stessa funzione d'influenza e la stessa efficienza asintotica. Si stabilisce pure una relazione interessante fra il test robusto ottimale del tipo C($\alpha$) e un test asintoticamente minimax.

Infine si confronta il comportamento asintotico di alcuni test per diverse distribuzioni e si calcolano due esempi basati su dati reali: ambedue dimostrano il rendimento eccellente dei nostri test robusti ottimali.
ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. F. Hampel for his valuable suggestions and his generous guidance throughout the course of this research. His interest and his encouragement have made this work possible.

I am also grateful to Prof. C. Field (Dalhousie University) and Prof. I. Olkin (Stanford University) for their critical reading of a version of the manuscript: their helpful comments led to an improved presentation.

My thanks also go to my friend and colleague P. Rousseeuw for his valuable suggestions and for many stimulating discussions.

Finally, I would like to thank A. Marazzi, M. Schweingruber and W. Stahel for some useful remarks.
CONTENTS

NOTATION

1. INTRODUCTION
   1.1 Scope of the present work and summary 4
   1.2 Robustness: Huber's and Hampel's approach 8
   1.3 Robust tests 13

2. INFINITESIMAL METHODS IN ROBUST STATISTICS 16
   2.1 The influence function for estimators 16
      2.1a Definition of the influence function for estimators 16
      2.1b Interpretation and properties of the influence function 18
      2.1c Hampel's optimality criterion 20
   2.2 The influence function for tests 23
      2.2a Definition of the influence function for tests 23
      2.2b Properties of the influence function for tests 25
      2.2c An optimality criterion for infinitesimal robustness of tests 30
      2.2d Examples: M-tests 32
   2.3 Connections with other influence functions for testing 34
      2.3a Bahadur and Pitman efficiency 34
      2.3b P-values 36
      2.3c Comparison with Lambert's approach 36
      2.3d Eplett's influence function for two-sample rank tests 39
   2.4 Further infinitesimal concepts 42
      2.4a The change-of-variance function 42
      2.4b The change-of-power function 47
3. LINEAR MODELS: ROBUST ESTIMATION AND TESTING
3.1 The model and the classical least squares estimators 48
3.2 The robust estimation theory in linear models 51
3.2a Huber estimators 51
3.2b General M-estimators 54
3.2c The change-of-variance function for general M-estimators 58
3.3 The test problem 65

4. A NEW GENERAL CLASS OF TESTS FOR LINEAR MODELS 68
4.1 The τ-test 69
4.2 The influence function 73
4.3 Asymptotic distributions 80
4.4 Special cases 89
4.4a τ(x,r) = ρ(r) 89
4.4b q = p-1 91
4.4c The density of x is "radially symmetric" with respect to \( \bar{x} \) 93

5. OPTIMALLY BOUNDED INFLUENCE TESTS 95
5.1 The \( \rho_c \)-test 96
5.2 The optimal test for Mallows' type regression 98
5.3 The optimal test for the general M-regression 101

6. A ROBUST SELECTION PROCEDURE 108
6.1 The Akaike Information Criterion and the \( C_P \) statistic 108
6.1a Generalities 108
6.1b Relation between the generalized Akaike statistic and the likelihood ratio test in linear models 110
6.2 A robust procedure
   6.2a Definition
   6.2b Properties of $AIC_{\alpha}^{(rob)}(p)$ and $C_{p;\alpha}^{(rob)}$

7. C(\alpha)-TYPE TESTS FOR LINEAR MODELS
   7.1 Definition of a C(\alpha)-type test
   7.2 The influence function and the asymptotic power of C(\alpha)-type tests
   7.3 Optimally robust C(\alpha)-type tests
      7.3a The optimal C(\alpha)-type test
      7.3b Connection with an asymptotically minimax test proposed by Wang

8. COMPUTATIONAL ASPECTS AND COMPARISON BETWEEN DIFFERENT TESTS
   8.1 Computational aspects
      8.1a Algorithm for the computation of the matrix $U_s$
      8.1b Computation of the asymptotic distribution of the $\tau$-test statistic
   8.2 Asymptotic behaviour of different tests
   8.3 Numerical examples
      8.3a Example 1
      8.3b Example 2

REFERENCES
INDEX
CURRICULUM VITAE
NOTATION

The following notation is used throughout this work.
Let p and q be fixed positive integers such that q < p.
Let R be a \((p \times s)\) matrix \((1 \leq s \leq p)\) and denote its elements by \(r_{ij}\). We define the following associated matrices:

(i) \(R^T\) is the transpose and \(R^{-1}\) the inverse of \(R\);

(ii) \(R^*\) is a \((p \times s)\) matrix with elements \(r_{ij}\) for \(1 \leq i \leq q, 1 \leq j \leq \min(q,s)\), and 0 otherwise:

\[
R^* = \begin{bmatrix} r_{ij} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ q & p-q \\
q & s-q \\
\end{bmatrix}
\]

(iii) \(R^{**} = R - R^*\);

(iv) \(R^+\) is the pseudoinverse of \(R\), that is \(R^+\) is a uniquely determined \((s \times p)\) matrix such that \(RR^+R = R\), \(R^+RR^+ = R^+\), \((RR^+)^T = RR^+\), \((R^+R)^T = R^+R\).

Let \(M\) be a \((p \times p)\) matrix. Then, \(M_{11}, M_{12}, M_{21}, M_{22}\) are the elements of the following partition of \(M\):

\[
M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \\ \end{bmatrix}
\]

\(M_{22.1} := M_{22} - M_{21}M_{11}^{-1}M_{12}\),

\(\text{tr } M := \sum_{j=1}^{p} m_{jj}\).

I denotes the \((p \times p)\) identity matrix and \(\text{diag}(\lambda_1, \ldots, \lambda_p)\) is a \((p \times p)\) diagonal matrix with diagonal elements \(\lambda_i\), \(i=1, \ldots, p\).
The components of a p-vector are denoted by means of upper indices, for example \(x = (x^{(1)}, \ldots, x^{(p)})^T\); \(\bar{x} := (x^{(1)}, \ldots, x^{(q)})^T\), \(\bar{x} := (x^{(q+1)}, \ldots, x^{(p)})^T\).

\(\|\cdot\|\) is the Euclidean norm and \(\|\cdot\|\) is some norm in \(\mathbb{R}^p\).

\(N(0,1)\) is the standard normal distribution, \(\Phi\) its distribution function and \(\phi\) its density.

\(L_F(T)\) denotes the distribution of \(T\) under \(F\).

\(E_H(E_\theta)\) denotes the expectation under \(H\) (respectively under \(F_\theta\) if a parametric model is specified).

We require the function
\[
\rho_c(r) = \begin{cases} 
  r^2/2 & \text{if } |r| \leq c \\
  c|r| - c^2/2 & \text{otherwise}
\end{cases}
\]

and its derivative
\[
\psi_c(r) = \begin{cases} 
  r & \text{if } |r| \leq c \\
  c \cdot \text{sign}(r) & \text{otherwise}
\end{cases}
\]

These functions are defined by Huber(1964) in obtaining robust estimators of location. \(\rho_c(r)\) is called Huber \(\rho\)-function for simplicity.

Moreover,
\[
[r]_a^b = \begin{cases} 
  a & \text{if } r \leq a \\
  r & \text{if } a < r < b \\
  b & \text{if } b < r
\end{cases}
\]

and
\[ l_A(r) = \begin{cases} 1 & \text{if } r \in A \\ 0 & \text{otherwise.} \end{cases} \]

Intervals are denoted by \([a, b], [a, b], \ldots\), where \(-\infty < a < b < +\infty\).

Finally, " := " denotes a definition and \(\Box\) the end of a proof.
1. INTRODUCTION

1.1 SCOPE OF THE PRESENT WORK AND SUMMARY

Since its introduction in 1968, Hampel's influence function (Hampel, 1968, 1974) has played a dominating role in the search for robust estimators, and has stimulated further research in this branch of statistics. Essentially, the influence function is the "first derivative" of an estimator viewed as functional and it describes the normalized influence on the estimate of an infinitesimal observation.

I quote Huber (1972), p. 1052:

"In my opinion, Hampel's influence function is the most important single heuristic tool for constructing robust estimates with specified properties. One will strive for influence functions which are bounded (to limit the influence of any single "bad" observation), which are reasonably continuous in x (to achieve insensitivity against roundoff and grouping effects) and which are reasonably continuous as a function of F (to stabilize the asymptotic variance of the estimate under small changes of F). At the same time, one will try to have an influence function roughly proportional to \(-\log f_0(x)\)'', to achieve a high efficiency at the model distribution \(F_0\)."

Because we see that an analogous device would prove very
useful in the study of robust testing, we use infinitesimal methods in order to construct optimally robust tests. This formalism allows us to investigate the robust properties of the classical tests, especially the classical F-test for linear models, and, solving an optimality problem based on a new optimality criterion for infinitesimal robustness of tests, to propose an optimally robust version in the infinitesimal sense.

The main scope of this work is to fill a gap in the theory of linear models, where one has a satisfactory theory only for the robust estimation of the parameters. On the other hand, we want to show that infinitesimal methods can be used in testing theory too, leading to powerful optimally robust test procedures. In our opinion such methods can be applied successfully in many testing problems.

Summary of the present work.

In the next section of this first chapter we review the basic approaches to robust statistics: Huber's minimax and Hampel's infinitesimal approach. Furthermore, in Section 1.3 we present some known results on robust testing.

Chapter 2 deals with infinitesimal methods in a general framework. In Section 2.1 we review the basic notion of influence function for estimators with the related Hampel's optimality criterion for infinitesimal robustness of estimators. In Section 2.2 we extend these concepts to tests: we work out the new notion of influence function for testing,
study its properties and state an optimality criterion for infinitesimal robustness of tests which is the natural counterpart to Hampel's criterion for estimators.

Different approaches to the influence function for testing are compared in Section 2.3. Section 2.4 is devoted to the new notion of change-of-variance function for estimators and its natural counterpart for tests, the change-of-power function. An optimality problem with respect to the change-of-variance function is solved and a connection with a curve introduced by Eplett (1980) is discussed.

In Chapter 3 we consider a theory of robust estimation in linear models (Section 3.2): we describe properties of robust estimators from the point of view of the influence function, extend the notion of change-of-variance function to the regression situation and show a new optimality result with respect to the change-of-variance function concerning the Hampel-Krasker estimator.

In Section 3.3 the test problem for linear models is discussed.

Chapter 4 introduces a new general class of tests for linear models that are the natural counterpart to general M-estimators (in the sense of Mallows, Hampel-Krasker, Welsch). In Section 4.2 and 4.3 we investigate the asymptotic properties (influence function and asymptotic distribution) of these tests.

Using the results of Chapter 4, we are able to derive optimally bounded influence tests (Section 5.2 and 5.3) that can be viewed as second-step robustified F-tests, the first-step robust version being given by the \( \rho_c \)-test (Section 5.1).
In Chapter 6 a robust selection procedure which fits Huber's regression (Huber estimators and $\rho_c$-test) is proposed.

In Chapter 7 we introduce an alternative class of tests ($C(\alpha)$-type tests) for the case where we test only one linear hypothesis on the parameters (with the dimension of the parameter space greater than 1) (Section 7.1). As far as the asymptotic properties (influence function and asymptotic efficiency) are concerned, $C(\alpha)$-type tests and $\tau$-tests are equivalent (Section 7.2).

Moreover, we find an interesting connection between our optimally robust test and an asymptotically minimax test proposed by Wang (1981) (Section 7.3).

Finally, in Chapter 8 we outline some computational aspects (Section 8.1) and present some numerical results comparing the asymptotic behaviour of several tests (Section 8.2).

We conclude with two numerical examples based on real data: both show the excellent performance of our optimally robust tests (Section 8.3).
1.2 ROBUSTNESS: HUBER'S AND HAMELP'S APPROACH

The first step in the mathematical formalization of a concept of robustness was introduced by Huber (1964), in the form of a minimax criterion for estimators of location parameters.

Let \( z_1, \ldots, z_n \) be \( n \) independent and identically distributed random variables with common distribution function \( F(z-\theta) \), where \( F \) is known and \( \theta \) is to be estimated.

Define the class of M-estimators for the location parameter \( \theta \) by

\[
T_n \text{ minimizes } \sum_{i=1}^{n} \rho(z_i - T_n),
\]

(1.2.1)

where \( \rho \) is a particular function.

Suppose that \( \rho \) has a derivative \( \rho' = \psi \). Then, \( T_n \) satisfies the equation

\[
\sum_{i=1}^{n} \psi(z_i - T_n) = 0.
\]

(1.2.2)

(1.2.1) and (1.2.2) are not always equivalent. Since (1.2.2) has in general no unique solution, we can define the estimator to be essentially the solution of (1.2.2) obtained by Newton's method with the sample median of \( (z_1, \ldots, z_n) \) as starting value (Collins, 1976, p. 71).

Note that \( \rho(r) = r^2 / 2 \) defines the arithmetic mean.

Huber showed the asymptotic normality of such estimators and solved the following problem.

Consider a "neighbourhood" of the model distribution \( F \)
\[ P_\varepsilon (F) := \{ G : G=(1-\varepsilon)F+\varepsilon H, \ H \text{ symmetric} \}, \quad (1.2.3) \]

where \( \varepsilon \in [0,1] \) is fixed (gross-error model), and find the M-estimator which minimizes the maximum of the asymptotic variance within \( P_\varepsilon (F) \).

When \( F \) is the normal distribution, the minimax-robust estimator is given by the so-called Huber \( \psi \)-function \( \psi_\varepsilon \). It is the maximum likelihood estimator under a distribution \( G_0 \) which minimizes the Fisher information \( J(G) \) within \( P_\varepsilon (\phi) \) (least favorable distribution). \( G_0 \) has the density

\[ g_0(r) = (1-\varepsilon)(2\pi)^{-1/2}\exp(-\rho_\varepsilon(r)), \]

where \( \varepsilon \) and \( c \) are related by the equation

\[ (1-\varepsilon)^{-1} = 2\phi(c)-1+2\phi(c)/c. \]

Later, Huber developed a finite-sample approach to robust testing based on a minimax criterion (see Section 1.3). Furthermore, he proposed a generalization of his estimator (1.2.2) for regression models (see Section 3.2) and extended the minimax approach to the covariance matrices problem (Huber, 1977b).

For a general discussion we refer to the survey paper of Huber(1972) and to his book (Huber, 1981).

In his Ph.D. thesis (Hampel, 1968) and later in a series of papers (Hampel, 1971, 1973a, 1974) Hampel studied the stability aspects of robustness and introduced the basic notions of qualitative robustness, breakdown point and influence function. He recognized and worked out the close analogy
between the stability aspects of robustness and the stability of a mechanical structure. I quote Huber (1972), p. 1045:

"(i) the qualitative aspect: a small perturbation should have small effects;
(ii) the breakdown aspect: how big can the perturbation be before everything breaks down;
(iii) the infinitesimal aspect: the effect of infinitesimal perturbations."

Infinitesimal methods are discussed extensively in Chapter 2: the basic notions are the influence function (both for estimators and tests), the change-of-variance function and the change-of-power function.

We first discuss (i) and (ii).

In order to formalize the concept of qualitative robustness, we require a notion of perturbation. This depends on a metric that defines small changes. To do this define the Prohorov distance between two probability distributions $F$, $G$ as

$$\pi(F,G) := \inf \{ \epsilon : P(A) \leq G(A^c) + \epsilon, \forall A \in \mathcal{A} \},$$

(1.2.4)

where $A^c$ is the set of the points with distance from $A$ less than $\epsilon$ and $\mathcal{A}$ is the underlying $\sigma$-algebra. The Prohorov distance formalizes the most common deviations from parametric models due to the following main reasons (Hampel, 1973a, p. 88):

a) rounding and grouping and other "local inaccuracies";
b) the occurrence of "gross errors";
c) the model may have been conceived only as an approximation in the sense of the weak topology.
Qualitative robustness means that small distances between $F$ and $G$ in the Prohorov sense implies small distances between the distributions of the estimator.

**DEFINITION 1.1** (Hampel, 1968, 1971)

A sequence of estimators $\{T_n : n \in \mathbb{N}\}$ is called qualitatively robust at $F$ iff:

$$\forall \varepsilon > 0, \exists \delta > 0, \forall G, \forall n \quad \pi(F,G) < \delta \Rightarrow \pi(L_F(T_n), L_G(T_n)) < \varepsilon.$$ 

**DEFINITION 1.2** (Hampel, 1968, 1971)

A sequence of estimators $\{T_n : n \in \mathbb{N}\}$ is called continuous at $F$ iff

$$\forall \varepsilon > 0, \exists \delta > 0, \exists n_0, \forall n, m \geq n_0, \forall F_n, F_m \text{ (empirical distr.)} \quad \{\pi(F, F_n) < \delta, \pi(F, F_m) < \delta\} \Rightarrow |T_n(F) - T_m(F)| < \varepsilon.$$ 

If the $T_n$'s are continuous in the observations, continuity at $F$ implies qualitative robustness at $F$ (Hampel, 1971, Theorem 1).

A crude global measure of robustness is given by the breakdown point.

**DEFINITION 1.3** (Hampel, 1968, 1971)

Let $\Omega$ be the parameter space and $\{T_n : n \in \mathbb{N}\}$ a sequence of estimators. The breakdown point $\varepsilon^*$ of $\{T_n : n \in \mathbb{N}\}$ at $F$ is defined by

$$\varepsilon^* := \sup\{\varepsilon \leq 1 : \exists \text{ compact } C(\varepsilon) \subseteq \Omega, \text{ such that} \quad \pi(F,G) < \varepsilon \Rightarrow G(T_n \in C(\varepsilon)) \to 1, \text{ as } n \to \infty \}.$$
\( \varepsilon^* \) tells us up to which Prohorov distance from \( F \) the estimator still gives us some indication of the original distribution. Loosely speaking, the breakdown point is the smallest amount of "free contamination" (of arbitrary gross errors) which can carry the estimate over all bounds (Hampel, 1978a, p. 431).

For other approaches to robust statistics, see Beran (1977a, 1977b) and for a summary of different approaches, see Bickel (1976, 1979).
1.3 ROBUST TESTS

In 1965 Huber extended his minimax criterion to tests of a simple hypothesis against a simple alternative (Huber, 1965). Let two distributions $P_0$ (hypothesis) and $P_1$ (alternative) be given and define neighbourhoods $P_j$ of $P_j$, $j=0,1$. Consider the following problem.

For a given $0<\alpha<1$ find a test $\Psi$ which maximizes

\[ \inf \{ E_{H} \Psi : H \in P_1 \} , \text{ under the side condition } \]

\[ \sup \{ E_{H} \Psi : H \in P_0 \} \leq \alpha . \]

Let $z_1, \ldots, z_n$ be $n$ independent and identically distributed observations. Then, under certain conditions (for example, if the $P_j$'s are either $\varepsilon$-contamination (see 1.2.3), Prohorov or total variation neighbourhoods), the solution of (1.3.1) is given by the test

\[ \tilde{\Psi}(z_1, \ldots, z_n) = 1_{\{ T_n(z_1, \ldots, z_n) > C_\alpha \}} , \]

where

\[ T_n(z_1, \ldots, z_n) = \prod_{i=1}^{n} \left[ \frac{p_1(z_i)}{p_0(z_i)} \right]^{c'} , \]

$p_j$ are the densities of $P_j$, $j=0,1$,

$C_\alpha$ is the critical value and

$c', c''$ are some constants defined by the neighbourhoods' structure.

$\tilde{\Psi}$ is called the censored likelihood ratio test (Huber, 1965).
From this exact finite-sample minimax result Huber(1968) derived a corresponding result for location estimators and he showed that the Huber estimator, for some constant \( c \) and for the normal model, is minimax in a finite-sample sense. Note that \( \tilde{\psi} \) defined by (1.3.2) is the likelihood ratio test between a least favorable pair of distributions \( Q_j \in P_j, j=0,1 \). The existence of such a least favorable pair can be studied in a general framework using 2-alternating capacities (Huber and Strassen, 1979; Bednarski, 1980).

It is difficult to determine the maximum level and the minimum power (within the neighbourhoods) of the censored likelihood ratio test. Asymptotic values were given by Huber-Carol (1970) who obtained nontrivial limiting level and power using the technique of shrinking neighbourhoods.

This method has been further used by Riede (1978). He constructed an asymptotic testing model by defining neighbourhoods in terms of \( \epsilon \)-contamination and total variation. He considered a class of tests defined by a test statistic of the form

\[
\tilde{\psi} = \frac{1}{n} \sum_{i=1}^{n} \psi(z_i),
\]

where \( \psi \) is an arbitrary function, and he found the uniquely determined \( \psi \)-function which defines the asymptotically minimax test within this class.

In subsequent work Rieder derived estimates from tests (Rieder, 1980) and later studied the robustness properties of rank tests (Rieder, 1981).
Rieder's asymptotic model was extended by Wang(1981) to the case where there are nuisance parameters. We shall discuss in detail this work in the case of linear models (Section 7.3). Moreover, we shall show the strong connection between our optimally bounded influence test and the asymptotically minimax test proposed by Wang.

The reader should notice that the results listed above are based on a minimax theory. In Chapter 2 we develop an infinitesimal approach to robust testing which allows us to deal with more general testing problems. In the cases where they are comparable, both approaches lead to equivalent results, as far as the structure of the optimally robust tests is concerned.
2. INFINITESIMAL METHODS IN
ROBUST STATISTICS

2.1 THE INFLUENCE FUNCTION FOR ESTIMATORS

The infinitesimal approach is based on the central notion of an influence function (IF). This concept was first introduced by Hampel (1968, 1974) for estimators and plays a central role in the theory of robust statistics.

2.1a Definition of the influence function for estimators

Consider a parametric model \((Z, A, \{F_\theta : \theta \in \Omega\})\) ;\(Z\) (the sample space) is a complete separable metric space, \(A\) is the \(\sigma\)-algebra generated by the topology, \(\{F_\theta : \theta \in \Omega\}\) is a family of distributions (distribution functions) on the measurable space \((Z, A)\) and \(\Omega\) (the parameter space) is a convex subset of \(\mathbb{R}^p\), for some integer \(p\).

Let \(z_1, \ldots, z_n\) be \(n\) independent and identically distributed observations in \(Z\) and let \(\{T_n : n \in \mathbb{N}\}\) be a sequence of estimators of \(\theta\) such that:

\[
(2.1.1) \quad T_n(z_1, \ldots, z_n) = T_n(F_n),
\]

where \(F_n\) is the empirical distribution, that is the distribution which puts mass \(1/n\) at \(z_i\), \(i = 1, \ldots, n\).
(2.1.T2) There exists a functional $T$ on a certain subset of the space of all probability distributions (domain of $T$) into $\mathbb{R}^p$ such that

$$T_n(z_1, \ldots, z_n) \xrightarrow{P} T(G)$$

in probability, when $n \to \infty$ and the observations are distributed according to $G$.

(2.1.T3) We suppose that the estimator $T$ is Fisher-consistent (Hampel, 1968, 1974):

$$T(F_0) = \theta, \quad \forall \theta \in \Theta,$$

that is, given the model, $T$ estimates the right parameter.

**DEFINITION 2.1** (Hampel, 1968, 1974)

Let $\delta_z$ be the distribution which puts mass 1 at any point $z \in Z$. Then, the influence function of $T$ at $F$ is defined by

$$IF(z; T, F) := \lim_{\epsilon \to 0} \frac{T((1-\epsilon)F + \epsilon \delta_z) - T(F)}{\epsilon} \quad (2.1.1)$$

if this limit exists.

There is a close connection between the concept of influence function and the notion of derivative of a functional $T$. Consider for instance the notion of Gâteaux differentiability.

**DEFINITION 2.2** (see Reeds, 1976)

A functional $T$ is called Gâteaux differentiable at $F$ if there exists a function $a_1$ such that

$$\lim_{\epsilon \to 0} \frac{T((1-\epsilon)F + \epsilon G) - T(F)}{\epsilon} = \int a_1(u)dG(u), \forall G. \quad (2.1.2)$$
The vector-valued function $a_1(z)$ can be easily computed putting $G=\delta_z$ in (2.1.2): we see that $a_1(z)$ is nothing but the influence function of $T$. On the other hand, $a_1(z)$ is the first kernel function of $T$ if regarded as von Mises functional (von Mises, 1947).

There exist stronger concepts of differentiability, for example Fréchet- and compact differentiability: for a complete survey we refer to the work of Reeds (1976). Other basic references are von Mises (1947) and Filippova (1962).

2.1b Interpretation and properties of the influence function

More simply, the influence function $IF(z; T, F)$ is the "first derivative" of an estimator viewed as functional $T$ and it describes the normalized influence on the estimate of an infinitesimal observation at $z$. Besides properties such as continuity, shape, points and heights of jumps, the following three parameters of the influence function are of particular interest (in order of priority) from the point of view of robustness.

**DEFINITION 2.3**

Let $|| \cdot ||$ be a norm in $\mathbb{R}^P$. Then, we define:

$$\gamma^*(T, F) := \sup \{ || IF(z; T, F) || : z \in \Omega \}$$  \hspace{1cm} (2.1.3)  

*gross-error-sensitivity*

$$\lambda^*(T, F) := \sup \{ || IF(z; T, F) - IF(v; T, F) || / || z - v || : z \neq v \}$$  \hspace{1cm} (2.1.4)  

*local-shift-sensitivity*

$$\rho^*(T) := \inf \{ c > 0 : IF(z; T, F) = 0, z \in \Omega, || z || (F) > c \}$$  \hspace{1cm} (2.1.5)  

*rejection point*
REMARK

This definition of $\rho^*$ makes sense only for a spheric symmetric distribution $F$. $|| \cdot ||_{(F)}$ denotes a norm in $Z$ which depends on $F$ (for example via $T$).

The interpretation of these parameters is as follows. $\gamma^*$ measures the worst influence which a small fixed amount of contamination can have on the value of the estimator. $\lambda^*$ is a measure of the worst effect of "wiggling" the observations, while all observations further away than $\rho^*$ are rejected completely. From this interpretation it is clear that we often look for estimators with finite $\gamma^*$, $\lambda^*$, $\rho^*$ (in order of priority).

There is also an important relation between the influence function and the asymptotic covariance matrix of the estimator $T$. Suppose that the von Mises expansion for $T$ holds (von Mises, 1947; Filippova, 1962; Reeds, 1976). Then, for some distribution $G$ in a neighbourhood of $F$

$$T(G) = T(F) + \int IF(z; T, F)d(G-F)(z) + ... \quad (2.1.6)$$

Since

$$\int IF(z; T, F)dF = 0,$$

putting $G = F_n$, the empirical distribution, (2.1.6) becomes

$$T_n = T_n(F_n) = T(F) + n^{-1} \sum_{i=1}^{n} IF(z_i; T, F) + ... \quad (2.1.7)$$

If the remainder term becomes negligible when $n \to \infty$, it follows from the central limit theorem that
\[ L_F(n^{1/2}(T_n - T(F))) + N(0, V(T;F)) \]  
(2.1.8)

where \( V(T;F) := \int IF(z;T,F) \cdot IF^T(z;T,F) dF(z) \).

2.1c Hampel’s optimality criterion

By means of the influence function one can describe both the robustness and the efficiency properties of the estimator and it is natural to try and find a balance between robustness and efficiency of the estimator. In order to describe Hampel’s optimality criterion, consider a class of estimators which is flexible and large enough for our purpose: the M-estimators. Consider the situation of Subsection 2.1a and define a class of functions \( \{ \zeta(z, \theta) : z \in \mathbb{Z}, \theta \in \mathbb{O} \subset \mathbb{R}^p \} \), where \( \zeta : \mathbb{Z} \times \mathbb{O} \rightarrow \mathbb{R}^p \).

Then, an \( M \)-estimator is defined as the (or a) solution \( T_n \) of the system of equations

\[ \sum_{i=1}^{n} \zeta(z_i, T_n) = 0. \]  
(2.1.9)

Suppose that \( T \) is Fisher-consistent, that is

\[ \int \zeta(z, \theta) dF_{\theta}(z) = 0. \]  
(2.1.10)

Then, under certain regularity conditions on the family \( \{ \zeta(z, \theta) \} \) (see Stahel, 1981), we obtain

\[ IF(z; \zeta, F_\theta) = -C^{-1}(\zeta, \theta) \cdot \zeta(z, \theta), \]  
(2.1.11)

where the matrix \( C(\zeta, \theta) \) is defined by

\[ C(\zeta, \theta) := \int (\partial \zeta(z, \theta) / \partial \theta) dF_{\theta}(z) \].
(We denote the estimator $T$ by its defining function $\zeta$.)

Huber (1967) gives the conditions under which these estimators are asymptotically normal with covariance matrix

$$\sqrt{\text{IF}(z;\zeta,F_\theta)} \cdot \text{IF}^T(z;\zeta,F_\theta) dF_\theta(z) = C^{-1} \int \zeta(z,\theta) \cdot \zeta^T(z,\theta) dF_\theta(z) C^{-T}.$$  

Formula (2.1.11) shows the simple relation between the influence function of an M-estimator and its defining function $\zeta$ and allows us to derive new estimators with pre-specified robustness properties. For instance, in order to obtain an estimator with a bounded influence function, we have to use a bounded score-function $\zeta$.

Hampel tries to find a balance between robustness and efficiency.

Hampel's optimality criterion is to put a bound on the gross-error-sensitivity

$$\gamma^*(\zeta,F_\theta) = \sup\{ \| \text{IF}(z;\zeta,F_\theta) \| : z \notin B \leq b_\theta \},$$

(2.1.12')

for a given constant $b_\theta$ and for some norm and, under this condition, to minimize the trace of the asymptotic covariance matrix of the estimator at the model:

$$\min\{ \text{tr} \sqrt{\text{IF}(z;\zeta,F_\theta)} \cdot \text{IF}^T(z;\zeta,F_\theta) dF_\theta(z) : \zeta \}.$$  

(2.1.12'')

(2.1.12') is the robustness condition and (2.1.12'') is the efficiency condition. Note that (2.1.12'') can be also interpreted as minimum of the asymptotic mean square error (see
and assume that

(2.2.S1) $\xi_n(\theta)$ converges to $\xi(\theta)$ for all $\theta$;

(2.2.S2) $\xi(\theta)$ is differentiable.

**DEFINITION 2.4** (Ronchetti, 1979; Rousseeuw and Ronchetti, 1979, 1981)

Let $\delta_z$ be the distribution which puts mass 1 at any point $z \in \mathcal{Z}$. Then, the influence function of the test defined by the test statistic $T$ at $F_\theta$ is defined by means of the expression

$$\text{IF}_{\text{test}}(z; T, F_\theta) := \lim_{\epsilon \to 0} \frac{(T((1-\epsilon)F_\theta + \epsilon \delta_z) - T(F_\theta))/\epsilon)}{\xi'(\theta)},$$

for all $z \in \mathcal{Z}$ where the right hand side exists.

**INTERPRETATION**

Suppose $\xi$ is strictly monotone. Then, define

$$U(G) := \xi^{-1}(T(G)).$$

This functional gives the parameter value which the true underlying distribution $G$ would have if it would belong to the model. This $U$ is clearly Fisher-consistent, since

$$U(F_\theta) = \xi^{-1}(T(F_\theta)) = \xi^{-1}(\xi(\theta)) = \theta.$$

Now, the following holds:

$$\text{IF}(z; U, F_\theta) = \lim_{\epsilon \to 0} \frac{(U((1-\epsilon)F_\theta + \epsilon \delta_z) - U(F_\theta))/\epsilon)}{\xi'(\theta)}$$

$$= \lim_{\epsilon \to 0} \frac{(T((1-\epsilon)F_\theta + \epsilon \delta_z) - T(F_\theta))/\epsilon)}{\xi'(\theta)}$$

$$= \frac{\text{IF}(z; T, F_\theta)}{\xi'(\theta)}$$

$$= \text{IF}_{\text{test}}(z; T, F_\theta).$$

This means that the influence function of the test equals the
influence function of a suitable standardization of the test statistic, namely that given by U.

On the other hand, for Fisher-consistent estimators Definition 2.4 coincides with Hampel's Definition 2.1 since in this case

\[ \xi(\theta) = T(F_\theta) = \theta, \]

so U=T.

REMARKS

1. Definition 2.4 has been extended for the two-sample case, see Ronchetti(1979) and Rousseeuw and Ronchetti(1979,1981).

2. In the location model there is an interesting connection between the influence function for tests and the influence function of shift estimators in the sense of Hodges and Lehmann(1963); for details, see Rousseeuw and Ronchetti (1979,1981).

2.2b Properties of the influence function for tests

PROPOSITION 2.1

If (2.1.T1), (2.1.T2), (2.2.S1), (2.2.S2) hold, then

(i) the influence function is invariant to monotone differentiable transformations of the test statistic, that is

\[ IF_{test}(z;h(T),F_\theta) = IF_{test}(z;T,F_\theta), \]

for a monotone differentiable function h;

(ii) \( \int IF_{test}(z;T,F_\theta) dF_\theta(z) = 0.\)
The proof of (i) follows directly from Definition 2.4 and
(ii) follows from the corresponding property of \(IF(z;T,F_\theta)\).

As for estimators, the influence function for tests can be used to measure the infinitesimal robustness of the test and its efficiency (asymptotic power). Let us begin with aspects of efficiency.

Recall that
\[
\xi_n(\theta) = E_{\theta} T_n
\]
and define
\[
\sigma_n^2(\theta) := \text{var}_{\theta} T_n.
\]
Consider a sequence of alternatives \( \theta_n = \theta_0 + \Delta n^{-1/2} \), for a positive \( \Delta \), and take the power at the alternative \( \theta_n \); when \( n \to \infty \), we obtain the asymptotic power
\[
\text{as power} = 1 - \Phi(\frac{\Delta}{\sqrt{n}}(1-\alpha) - \Delta n^{1/2})
\]
where \( \alpha \) is the level of the test and

\[
E=E(T;F_{\theta_0}) \text{ is the asymptotic Pitman efficacy at } F_{\theta_0},
\]
which is defined as (Noether, 1955)

\[
E(T;F_{\theta_0}) := \lim_{n \to \infty} \frac{(\xi_n'(\theta_0) \phi)^2 / (n \sigma_n^2(\theta_0))}. \tag{2.2.1}
\]

Pitman's theorem (see Noether, 1955, with \( m = 1, \delta = 1/2 \)) now states that the asymptotic relative efficiency \( \text{ARE}_{1,2} \) of two tests equals the ratio of their asymptotic Pitman efficacies:

\[
\text{ARE}_{1,2} = E_1 / E_2. \tag{2.2.2}
\]

The asymptotic Pitman efficacy \( E \) describes the efficiency
properties of a test and can be used for comparing different
tests. The next proposition shows that $E$ can be computed
using the influence function of the test.

**PROPOSITION 2.2**

If (2.1.T1), (2.1.T2), (2.1.S1), (2.1.S2),

(2.2.TAS1) $L((T - x_n(\theta))/\sigma_n(\theta)) \rightarrow N(0,1)$, when $n \rightarrow \infty$,
uniformly in $\theta$, $\theta_0 < \theta < \theta_0 + d$ ($d > 0$),

(2.2.TAS2) $\lim_{n \rightarrow \infty} \xi_n^{'}(\theta)/\xi_n^{'}(\theta_0) = 1$
$\lim_{n \rightarrow \infty} \sigma_n(\theta)/\sigma_n(\theta_0) = 1$,
where $\theta_n = \theta_0 + \Delta n^{-1/2}$,

(2.2.TAS3) $V(T_n; F_{\theta_0}) = f(IF(z; T, F_{\theta_0}))^2 dF_{\theta_0}(z)$,
where $V(T_n; F_{\theta_0})$ is the asymptotic variance of $T_n$,

then

(i) as power $= 1 - \Phi(\Phi^{-1}(1-\alpha) - \Delta n^{-1/2})$ and 
$E(T; F_{\theta_0})^{-1} = f(IF_{\text{test}}(z; T, F_{\theta_0}))^2 dF_{\theta_0}(z)$;

(ii) $f(IF_{\text{test}}(z; T, F_{\theta_0}))^2 dF_{\theta_0}(z) \geq (J(F_{\theta_0}))^{-1}$,
where $J(F_{\theta_0})$ is the Fisher information.

The test defined by $T$ is asymptotically efficient at $F$
(meaning we have equality) iff $IF_{\text{test}}(z; T, F_{\theta_0})$ is
proportional to $[\partial \log f_\theta(z)/\partial \theta]_{\theta = \theta_0}$, where $f_\theta$ is the
density of $F_{\theta}$.

**Proof.**

(ii) follows directly from Huber(1977a, p. 23) because the
functional $U$ (see Subsection 2.2a) is Fisher-consistent.

(i) : We have

$$\begin{align*}
\text{IF}_{\text{test}}(z;T,F_{\theta_0}) &= \lim_{\varepsilon \to 0} \frac{(T((1-\varepsilon)F_{\theta_0} + \varepsilon F_{\theta_0}) - T(F_{\theta_0}))/\varepsilon)}{\xi'(\theta_0)} \\
&= \text{IF}(z;T,F_{\theta_0})/\xi'(\theta_0)
\end{align*}$$

Combining (2.2.TAS3) with $V(T_n;F_{\theta_0}) = \lim_{n \to \infty} n \cdot \sigma^2_{n}(\theta_0)$,

$$\int (\text{IF}_{\text{test}}(z;T,F_{\theta_0}))^2 dF_{\theta_0}(z) \text{ becomes}$$

$$\lim_{n \to \infty} n \cdot \sigma^2_{n}(\theta_0)/(\xi'_n(\theta_0))^2 = (E(T;F_{\theta_0}))^{-1}.$$

Conditions (2.2.TAS1) and (2.2.TAS2) are taken from Noether (1955) and guarantee that the formula for the asymptotic power holds.

Now one can write the asymptotic Pitman efficacy $E$, the asymptotic power and the asymptotic relative efficiency in terms of the influence function of the test.

For the two-sample case Rousseauw and Ronchetti (1979, 1981) obtained similar results.

We now examine the influence of contamination on the level and on the power of the test. To define the asymptotic power, one constructs again a sequence of alternatives $\theta_n = \theta_0 + \Delta n^{-1/2}$, but now there is contamination in these $F_{\theta_n}$. (Note that some point mass contaminates the null hypothesis and the alternative.) When $\theta_n$ tends to $\theta_0$, this contamination must tend to zero equally fast, or its effect will soon dominate everything else and give divergence. Huber-Carol (1970) and Rieder (1978)
consider the same kind of contamination.

**DEFINITION 2.5**

Let \( \delta_z \) be the distribution which puts mass 1 at any point \( z \in \mathbb{Z} \). Define, with \( \varepsilon_n = \varepsilon \cdot n^{-1/2} \)

\[
F_n^{(1)}(z, \varepsilon, n) := (1-\varepsilon_n)F_{\theta_0, n} + \varepsilon_n \delta_z,
\]

\[
F_n^{(2)}(z, \varepsilon, n) := (1-\varepsilon_n)P_{\theta_0, n} + \varepsilon_n \delta_z
\]

and

\[
U_n(z_1, \ldots, z_n) := \xi_n^{-1}(T_n(z_1, \ldots, z_n)),
\]

\[
P_n(z, \varepsilon) := F_n^{(1)} \{ U_n > C_n^{(\alpha)} \}
\]

\[
L_n(z, \varepsilon) := F_n^{(2)} \{ U_n > C_n^{(\alpha)} \}
\]

where \( C_n^{(\alpha)} \) is given by \( P_{\theta_0, n} \{ U_n > C_n^{(\alpha)} \} = \alpha \), the nominal level.

The **level influence function** is defined by

\[
LIF(z; T, F_{\theta_0}) := \lim_{n \to \infty} (\partial / \partial \varepsilon) L_n(z, \varepsilon) \bigg|_{\varepsilon = 0},
\]

(2.2.3)

and the **power influence function** by

\[
PIF(z; T, F_{\theta_0}) := \lim_{n \to \infty} (\partial / \partial \varepsilon) P_n(z, \varepsilon) \bigg|_{\varepsilon = 0},
\]

(2.2.4)

if the right hand sides exist

(see Ronchetti, 1979; Rousseeuw and Ronchetti, 1979).

If

\[
L_n^{1/2}(U_n - U(G))/\{(IF(z; U, G))^2dG(z)\}^{1/2} \to \mathcal{N}(0,1),
\]

when \( n \to \infty \), then, under sufficient regularity conditions that allow us to interchange integral and derivative,
\[ \text{LIF}(z; T, F_{\theta_0}) = E^{1/2} \cdot \phi^{-1}(1-\alpha) \cdot \text{IF}_{\text{test}}(z; T, F_{\theta_0}) \]  
\[ \text{PIF}(z; T, F_{\theta_0}) = E^{1/2} \cdot \phi^{-1}(1-\alpha) - \Delta E^{1/2} \cdot \text{IF}_{\text{test}}(z; T, F_{\theta_0}). \]

For details, see Ronchetti (1979), Rousseeuw and Ronchetti (1979).

These formulae show that \( \text{IF}_{\text{test}} \), which is proportional to the influence function of the test statistic, describes the approximate effect of contamination on the level and on the power of the test. Compare also with the results of Section 2.3.

**2.2c An optimality criterion for infinitesimal robustness of tests**

In this subsection we introduce an analogue of Hampel's criterion for testing. We shall use it to construct optimally robust tests, see Subsection 2.2d (M-tests) and especially Chapter 5 and 7 (tests for linear models).

First we need an efficiency condition. Following Pitman's approach we have to try and maximize the asymptotic power of the test. In view of Proposition 2.2(i) this is equivalent to minimizing \( \int (\text{IF}_{\text{test}}(z; T, F_{\theta_0}))^2 dF_{\theta_0}(z) \).

We solve this extremal problem under a side condition which is the (infinitesimal) robustness condition.

This condition is

\[ \sup\{ |\text{IF}_{\text{test}}(z; T, F_{\theta_0})| : z \in Z \} \leq b_{\theta_0}, \]

for a given constant \( b_{\theta_0} \), in the case we can standardize the
test statistic as proposed in Subsection 2.2a and 2.2b. On the other hand, condition (2.2.7) is equivalent to the following

$$\gamma^*(T, F_{\theta_0}) = \sup\{|IF(z; T, F_{\theta_0})| : z \in \mathbb{Z} \leq b_{\theta_0}\}$$  \hspace{1cm} (2.2.8)$$
since both influence functions are proportional. Condition (2.2.8) makes sense even in the case we cannot perform the standardization of the test statistic (for example for the model $F_{\theta}(z) = \phi((z-\theta)/\theta^{1/2})$). Therefore, a general optimality criterion for infinitesimal robustness of tests can be stated as follows.

Consider a class $C$ of tests depending on the observations only through a test statistic $T_n(z_1, \ldots, z_n)$. Then, find a test which maximizes the asymptotic power within $C$, under the side condition of a bound on the influence function of the test statistic at the null hypothesis.
2.2d Example: M-tests

Consider the model of Subsection 2.2a and the class of tests with critical regions \( \{ \widetilde{T}_n(z_1, \ldots, z_n) > C_n \} \), where

\[
\widetilde{T}_n(z_1, \ldots, z_n) = n^{-1/2} \sum_{i=1}^{n} \psi(z_i)
\]

(2.2.10)

and \( \psi = \psi_{\theta_0} \) is a function that satisfies \( \int \psi^2 dF_{\theta_0} < \infty \).

Without loss of generality suppose that \( \int \psi dF_{\theta_0} = 0 \).

These tests were introduced by Rieder (1978). We call them M-tests.

Note that in the location case we can derive M-estimators from these tests using the same construction we apply to get R-estimators from R-tests, namely finding an estimator \( S_n \) such that the test statistic \( T_n \) equals 0 for the sample \( z_1 - S_n, \ldots, z_n - S_n \) (see Ronchetti, 1979).

In order to compute the influence function of such tests, we must standardize \( \widetilde{T}_n \) into a von Mises functional of the empirical distribution function

\[
T_n = n^{-1/2} \widetilde{T}_n = \int \psi dF_n,
\]

and with the notation of 2.2a, we have

\[
E_n(\theta) = E_0 T_n = \int \psi dF_0 = \xi(\theta).
\]

If \( \left[ (\partial \psi / \partial \theta) / \psi dF_0 \right]_{\theta = \theta_0} < \infty \), then

\[
I_{\text{test}}(z; \psi, F_{\theta_0}) = \psi(z) / \left[ (\partial \psi / \partial \theta) / \psi dF_\theta \right]_{\theta = \theta_0},
\]

(2.2.11)

\[
E(\psi; F_{\theta_0}) = \left( \left[ (\partial \psi / \partial \theta) / \psi dF_\theta \right]_{\theta = \theta_0} \right)^2 / \int \psi^2 dF_{\theta_0}.
\]

(2.2.12)

Now, we can look for the test which maximizes the asymptotic
power (or equivalently, the asymptotic Pitman efficacy E) within the class of the M-tests, under the side condition of a bound on its influence function.

In view of (2.2.11) and (2.2.12) we have to solve the following extremal problem.

For a given \( k_{\theta_0} \) find \( \psi \) which minimizes

\[
\psi^2 dF_{\theta_0} / \left( \int (3 \nabla \theta) \psi dF_{\theta_0} \right)_{\theta = \theta_0}^2,
\]

under the side condition

\[
\sup \{ |\psi(z)| / \left[ (3 \nabla \theta) \psi dF_{\theta_0} \right]_{\theta = \theta_0} : z \in Z \} \leq k_{\theta_0}.
\]

The function \( \tilde{\psi} \) which solves (2.2.13) can be found using Lemma 5 of Hampel(1968,1974). It is of the form

\[
\tilde{\psi}(z) = \tilde{\psi}_{\theta_0}(z) = \left[ \left( (3 \nabla \theta) \log f_\theta(z) \right)_{\theta = \theta_0} - a_{\theta_0} \right]^{+b_{\theta_0}} - b_{\theta_0},
\]

for some constants \( a_{\theta_0} \) and \( b_{\theta_0} \).

For details, see Ronchetti(1979).

Further examples are given in Rousseeuw and Ronchetti(1979, 1981).
2.3 CONNECTIONS WITH OTHER INFLUENCE FUNCTIONS FOR TESTING

In this section we present the connections between our influence function for tests (defined in Section 2.2) and other influence functions for testing. In order to define them, we first have to summarize the basic concepts of Bahadur and Pitman efficiency and the notion of $P$-value. Consider the situation of Subsection 2.2a.

2.3a Bahadur and Pitman efficiency

Bahadur (1960) defines $\{T_n; n \in \mathbb{N}\}$ to be a standard sequence if the following conditions are satisfied:

(2.3.BA1) There exists a continuous probability distribution function $F$ such that, for each $x \in \mathbb{R}$

$$\lim_{n \to \infty} P_{\theta_0} \left\{ T_n < x \right\} = F(x).$$

(2.3.BA2) There exists a constant $a$, $0 < a < \infty$, such that

$$\log(1-F(x)) = \frac{1}{2} a \cdot x^2 (1+o(1)), \text{ as } x \to \infty.$$  

(2.3.BA3) There exists a function $b(\theta)$ on $\Omega \setminus \{\theta_0\}$, with $0 < b(\theta) < \infty$, such that for each $\theta \in \Omega \setminus \{\theta_0\}$

$$\lim_{n \to \infty} P_{\theta} \{ |n^{-1/2} T_n - b(\theta)| > x \} = 0 \text{ for every } x > 0.$$  

For any standard sequence Bahadur shows that

$$-2n^{-1} \log(1-F(T_n)) + a \cdot b(\theta) = c(\theta), \forall \theta \in \Omega \setminus \{\theta_0\}, \text{ as } n \to \infty.$$
\[ c(\theta) \text{ is the approximate slope of the sequence } \{ T_n : n \in \mathbb{N} \} \text{ and} \]
\[ e_{1,2}^{(B)}(\theta) = c_{(1)}(\theta)/c_{(2)}(\theta) \quad (2.3.1) \]

is the approximate relative efficiency of two standard sequences \( \{ T_n^{(1)} : n \in \mathbb{N} \} \), \( \{ T_n^{(2)} : n \in \mathbb{N} \} \).

The idea behind Bahadur's comparison of tests is to look "at the rate of convergence (against 0) of the level of the test". For details, see Bahadur(1960,1967).

Pitman considers tests at level \( \alpha \) and a sequence of alternatives which converges to the hypothesis at a certain rate (typically \( n^{-1/2} \)). In this way he obtains a limiting power which is different from 1 (as \( n \to \infty \)): this is the basis for the comparison between different tests.

Using the formalism of Noether(1955) we can compute the \textit{asymptotic Pitman efficacy} \( E \) (which is a monotone increasing function of the asymptotic power) and the \textit{Pitman relative efficiency}

\[ e_{1,2}^{(P)} = E_1/E_2 \quad (2.3.2) \]

(see Subsection 2.2b).

Furthermore, we can express \( E \) in terms of expectation and variance of the test statistic using (2.2.1).

Wieand(1976, p.1005) gives a condition under which Bahadur and Pitman efficiencies coincide, namely

\[ \lim_{\theta \to \theta_0} e_{1,2}^{(B)}(\theta) = \lim_{\alpha \to 0} e_{1,2}^{(P)}(\alpha, \beta), \quad (2.3.3) \]
\( \tilde{c}_L(H) = d(T(H)) \),

where \( d \) is a differentiable function,

then

\[
IF_L(z; \{P_n\}, F_\theta) = c^i_L(\theta) \cdot IF_{test}(z; T, F_\theta).
\]

**Proof.**

Applying the definitions of \( IF_{test} \) and \( \xi \) we have:

\[
IF_L(z; \{P_n\}, F_\theta) = (\partial \partial \epsilon) \tilde{c}_L((1-\epsilon)F_\theta + \epsilon z) \bigg|_{\epsilon=0}
\]

\[
= (\partial \partial \epsilon) d(T((1-\epsilon)F_\theta + \epsilon z)) \bigg|_{\epsilon=0}
\]

\[
= d'(\xi(\theta)) \cdot IF(z; T, F_\theta)
\]

\[
= (\partial \partial \theta) d(\xi(\theta)) \cdot IF_{test}(z; T, F_\theta).
\]

**REMARKS**

1. Proposition 2.3 shows that \( IF_L \), \( IF_{test} \) and the influence function of the test statistic are proportional. Therefore, they have the same qualitative behaviour, as far as boundedness and continuity properties (if \( d' \) is continuous) are concerned.

2. From Proposition 2.3 we obtain

\[
\int (IF_L(z; \{P_n\}, F_\theta))^2 dF_\theta(z) = (c^i_L(\theta))^2 \cdot \int (IF_{test}(z; T, F_\theta))^2 dF_\theta(z).
\]

If

\[
\int (IF_L(z; \{P_n\}, F_\theta))^2 dF_\theta(z) = \tau^2(\theta) \quad \text{(Lambert, 1981, p. 651),}
\]

and
\[
\lim_{\theta \to \theta_0} \int (IF_{\text{test}}(z; T, F_\theta))^2 dF_\theta(z) = \\
\int (IF_{\text{test}}(z; T, F_\theta))^2 dF_\theta(z) = (E(T; F_\theta))^2,
\]

we can rewrite the asymptotic Pitman efficacy \( E \) in terms of \( c_L \) and \( \tau \):

\[
E(T; F_{\theta_0}) = \lim_{\theta \to \theta_0} \left( c_L(\theta) / \tau(\theta) \right)^2.
\]

(2.3.8)

3. Note that (2.3.C2) is not satisfied for permutation tests.

2.3d Eplett's influence function for two-sample rank tests

Let \( u_1, \ldots, u_m \) and \( v_1, \ldots, v_n \) be \( N := m + n \) independent observations, where the \( u \)'s are taken from a population with distribution function \( F(u) \) and the \( v \)'s from a population with distribution function \( G(v) \). Assume that \( F \) and \( G \) are continuous (see Eplett, 1980).

The location model asserts that

\[
G(u) = F(u+\theta),
\]

(2.3.9)

for some \( \theta \in \mathbb{R} \). We want to test the hypothesis

\[
H_0 : \theta = 0.
\]

We consider two-sample rank tests defined by the test statistic

\[
T_N(u_1, \ldots, u_m; v_1, \ldots, v_n) = \sum_{i=1}^m a_N(R_{i_1}),
\]

where \( R_{i_1} \) is the rank of \( u_{i_1} \) in the ordered combined sample and \( a_N(1), \ldots, a_N(N) \) are the scores.

We suppose that there exists a score generating function
\[ J : [0,1] \to \mathbb{R} \]

which satisfies the following conditions:

(2.3.J1) \( J \) is odd (meaning \( J(1-u) = -J(u) \)), nondecreasing and
square integrable;

(2.3.J2) \[ \lim_{N \to \infty} \int_0^1 (a_N(1+\lfloor uN \rfloor) - J(u))^2 \, du = 0, \]
where \( \lfloor uN \rfloor \) is the greatest integer which is not
greater than \( uN \).

Furthermore, assume (Eplett, 1980, p. 65)

(2.3.E1) There exists \( \lambda \in ]0,1[ \) such that \( m/n \to \lambda \), when \( N \to \infty \).

(2.3.E2) \( F \) has an absolutely continuous density \( f \). Denote by
\( f' \) its derivative.

With the notation of Section 2.2, Eplett (1980, p. 65) proves
that
\[ N^{-1/2} (T_N - \varepsilon_N(\theta))/\sigma_N(\theta) \to (\lambda(1-\lambda))^{1/2} \beta(J;F) \cdot \theta + o(\theta), \quad (2.3.10) \]
in probability (under the model (2.3.9)),

where
\[ \beta(J;F) = -\int_0^\infty J(F(u)) \cdot f'(u) \, du / \left( \int_0^1 (J(u))^2 \, du \right)^{1/2} \]
\[ = (\lambda(1-\lambda))^{-1/2} (E(J;F))^{1/2}. \]

Moreover, the Bahadur slope at the model equals
\[ c(\theta) = (E(J;F))^{1/2} \cdot \theta. \quad (2.3.11) \]
DEFINITION 2.8 (Eplett, 1980, p. 65)

The E-influence function for a sequence of two-sample rank tests is defined by

$$IF_E(z; J, F) := \lim_{\epsilon \to 0} \frac{\beta(J; (1-\epsilon)F + \epsilon \delta_z) - \beta(J; F)}{\epsilon},$$

for all $z \in \mathbb{R}$ where the limit exists.

The influence function $IF_E(z; J, F)$ describes the infinitesimal behaviour of the Bahadur slope at the model when the distribution function generating the model lies in the vicinity of $F$. Moreover, $IF_E(z; J, F)$ is, roughly speaking, the derivative of the square root of the asymptotic Pitman efficacy which can be written as

$$E(J; F) = \left( \int IF_{test}(z; J, F)^2 dF(z) \right)^{-1/2} \lambda(1-\lambda),$$

where $IF_{test}$ is the extension of our influence function (Section 2.2) to the two-sample case (see Rousseeuw and Ronchetti, 1979, 1981).

Since $(E(J; F))^{1/2}$ can be viewed as the inverse of the asymptotic standard deviation of the standardized test statistic (see Interpretation, Subsection 2.2a and 2.2b), it turns out that $IF_E$ is the natural counterpart of the concept of change-of-variance function which is defined in the same way for estimators (see Subsection 2.4a and 3.2c).

Therefore, we think that the name change-of-power function is more appropriate for Eplett's influence function (see Subsection 2.4b).
2.4 FURTHER INFINITESIMAL CONCEPTS

2.4a The change-of-variance function

The concept of change-of-variance function is a relatively new tool for robustness theory. It was discovered in 1972 by Hampel who used it to construct the so-called hyperbolic tangent estimators of location. The function as well as the estimators were sometimes referred to briefly (Hampel, 1973a, p.98 and 1974, p.393). Later these ideas were developed for the location case more rigorously by Rousseeuw (1981) and Hampel, Rousseeuw and Ronchetti (1980, 1981).


Consider the model of Subsection 2.1a. Let T be an estimator and \( V(T; F) \) its asymptotic variance at F. Put

\[
F_{\varepsilon, z} := (1 - \varepsilon) F + \varepsilon z.
\]

Define the change-of-variance function of the estimator T at F by means of the expression

\[
\text{CVF}(z; T, F) := (\partial_{\varepsilon} \partial_{\varepsilon}) V(T; F_{\varepsilon, z}) \bigg|_{\varepsilon = 0}.
\]

(2.4.1)

defined in all \( z \in \mathbb{Z} \) where the right member exists.

Further, define the change-of-variance sensitivity as

\[
\kappa^*(T, F) := \sup\{ \text{CVF}(z; T, F) / V(T; F) : z \in \mathbb{Z} \}.
\]

CVF describes the infinitesimal increment of the asymptotic
variance in the vicinity of \( F \). Therefore, it is a measure of
the stability of the asymptotic variance under contamination.
Let us apply this concept to M-estimators of location.

Consider the location model with a model distribution
\( F_{\theta}(z) = F(z - \theta) \ (z \in \mathbb{R}) \) which satisfies the following condi-

(2.4.F1) \( F \) has a twice continuously differentiable density \( f \)
which is symmetric with regard to zero and satisfies
\( f(z) > 0, \ z \in \mathbb{R} \).

(2.4.F2) The score-function \(-f'/f = (-\log f)'\) is strictly
monotone increasing and
\[
\int (-f'/f)' \psi = -\int (-f'/f)f' < \infty.
\]

Let \( \{\psi\} \) be the family of M-estimators for location defined by
equation (1.2.2). We assume that \( \psi \) satisfies the following
conditions (Rousseeuw, 1981, p. 15):

(2.4.PSI1) \( \psi \) is defined and continuous on \( \mathbb{R}, \ \psi(-z) = -\psi(z), \forall z \in \mathbb{R} \)
and \( \psi(z) \geq 0, \) for \( z \geq 0 \).

(2.4.PSI2) The set of points in which \( \psi' \) is not defined or not
continuous is finite.

(2.4.PSI3) \( \int \psi^2 dF < \infty \); \( 0 < \int \psi'dF = \int (-f'/f) \psi dF < \infty \).

The asymptotic variance of an M-estimator corresponding to \( \psi \)
at the symmetric distribution \( H \) equals the expression

\[
V(T; H) = \int \psi^2 dH / (\int \psi' dH)^2
\]

(Huber, 1967).
In order to preserve the symmetry, we substitute \((1-\varepsilon)F + \epsilon \delta_z\) by a symmetric contamination \((1-\varepsilon)F + \epsilon (\delta_z + \delta_{-z})/2\). (Formally, we obtain the same result for the CVF.)

It follows that the change-of-variance function of an M-estimator for location corresponding to \(\psi\) at \(F\) is given by

\[
\text{CVF}(z; \psi, F) = V(\psi; F) \left( 1 + \frac{\psi^2(z)}{\int \psi^2 dF} - 2\psi'(z)/\int \psi' dF \right) \tag{2.4.2}
\]

in the points where the right member is defined.

Now, we can state the following Hampel type optimality problem.

Find an M-estimator which minimizes the asymptotic variance at \(F\) within the class \(\{\psi\}\), under the side condition of a bounded change-of-variance sensitivity, that is \(\kappa^*(\psi, F) \leq b\), for a given positive constant \(b\). \tag{2.4.3}

**REMARKS**

1. We standardize \(\text{CVF}\) (by dividing by \(V\)) for mathematical convenience.

2. We don't use the absolute value of \(\text{CVF}\) because we do not have to worry about large negative values of it: these values would mean a decrease in the asymptotic variance.

3. By means of the change-of-variance sensitivity \(\kappa^*\) we can find an approximation to the supremum of the asymptotic variance of an estimator (defined by \(\psi\)) in a \(\varepsilon\)-neighbourhood of a distribution \(F\):
\[ \sup \{ V(\psi; (1-\varepsilon)F+\varepsilon G) : G \text{ symmetric, } 0<\varepsilon<1 \} \]
\[ \simeq V(\psi; F)(1 + \varepsilon \kappa^*(\psi, F)). \]  
(2.4.4)

4. Change-of-variance functions for R-estimators can be found in Ronchetti (1979). On the other hand, it seems difficult to solve (2.4.3) within the class of R-estimators.

5. A related notion is the concept of change-of-bias function which is, roughly speaking, the derivative (with respect to \( \varepsilon \)) of the gross-error-sensitivity (computed at a contaminated distribution \( F_\varepsilon \)). This notion is referred to briefly in Hampel (1973a, p. 98).

The solution to the above problem (2.4.3) for \( F=\Phi \) is given by the Huber estimator defined by \( \psi=\psi_c \) (Rousseeuw, 1981). If we restrict ourselves to redescending M-estimators, that is M-estimators for location defined by a function \( \psi \) that satisfies (besides (2.4.PSI1), (2.4.PSI2), (2.4.PSI3))

(2.4.PSI4) \( \psi(z) = 0 \) for \( |z| > c \),

the solution to the above problem (2.4.3) for \( F=\Phi \) is given by the so-called hyperbolic tangent estimators (Hampel, Rousseeuw and Ronchetti, 1980, 1981). They are defined by the following function

\[ \psi(z) = z \quad \text{for } 0 \leq |z| \leq a \]
\[ = \text{atanh} \left[ \beta \cdot (c - |z|) \right] \text{sign}(z) \quad \text{for } a \leq |z| \leq c \]
\[ = 0 \quad \text{for } c \leq |z| \]  
(2.4.5)

for some constants \( a, \alpha, \beta \).
Figure 2.4.1: $\psi$ defining hyperbolic tangent estimators

(The corresponding $\rho$-function, $\rho(z)=\int_0^z \psi(t)dt$, is plotted in Figure 8.2.1.)

Similar hyperbolic tangent estimators solve a certain minimax problem (see Collins, 1976; Huber, 1977c).

In Subsection 3.2c we give a generalization of the change-of-variance function to the regression situation and we solve the related Hampel type problem.
2.4b The change-of-power function

**DEFINITION 2.10** (Ronchetti, 1979)

Consider a parametric model generated by a distribution $F$. Let $T$ be a test statistic and $E(T; F)$ the asymptotic Pitman efficacy of the test defined by $T$ (at $F$).

Put $F_{\varepsilon, z} := (1-\varepsilon)F + \varepsilon \delta_z$.

Define the *change-of-power function* of the test defined by $T$ (at $F$) by

$$\text{CPF}(z; T, F) := (\exists \varepsilon) E(T; F_{\varepsilon, z}) \bigg|_{\varepsilon=0} \ ,$$

(2.4.6)

defined in all $z \notin Z$ whenever the right member exists.

Note that Eplett's influence function for two-sample rank tests (Subsection 2.3d) is proportional to the change-of-power function defined here and so is the change-of-variance function of R-estimators derived from rank tests.

In this situation too we can state a Hampel type problem with respect to the change-of-power function.

In the class of M-tests the solution is given by the test defined by the Huber function. In the case of rank tests it seems rather difficult to solve the corresponding problem (Ronchetti, 1979).
3. LINEAR MODELS: ROBUST ESTIMATION AND TESTING

3.1 THE MODEL AND THE CLASSICAL LEAST SQUARES ESTIMATORS

Consider the following linear model.

Let \( \{(x_i, y_i) : i=1, \ldots, n\} \) be a sequence of independent and identically distributed random variables such that

\[
y_i = x_i^T \theta + e_i \quad i=1, \ldots, n,
\]

where

- \( y_i \in \mathbb{R} \) is the ith observation,
- \( x_i \in \mathbb{R}^p \) the ith row of the design matrix \( X \) with \( X \in \mathbb{R}^{n \times p} \)
- \( \theta \in \Omega \subset \mathbb{R}^p \) a p-vector of unknown parameters and
- \( e_i \in \mathbb{R} \) the ith error.

Suppose that \( e_i \) is independent of \( x_i \) and has a symmetric distribution \( G(e/\sigma) \), where \( \sigma > 0 \) is a scale parameter, with density \( g \) with respect to the Lebesgue measure.

Let \( K(x) \) be the distribution function of \( x_i \), with the density \( k \) with respect to a \( \sigma \)-finite measure \( \nu \).

Estimation and test procedures of the classical statistic in linear models are based on the well-known method of least squares.
This method was first published by Legendre in 1805, but the problem of the priority between Gauss and Legendre over this discovery is still open (Stigler, 1981).

The use of least squares estimators is justified mostly by the Gauss-Markov Theorem.

**DEFINITION 3.1** (see Scheffé, 1959, p.13)

Let \( l(\theta) = b^T \theta \) be a linear function of the unknown parameters. \( l(\theta) \) is called estimable if there exist constants \( a_1, \ldots, a_n \) such that

\[
E_\theta (\sum_{i=1}^{n} a_i y_i) = l(\theta).
\]

The *least squares (LS) estimate* \( T_n^{(LS)} \) of \( \theta \) is any statistic that minimizes the Euclidean norm of the residuals:

\[
\Gamma(\theta) = \sum_{i=1}^{n} ((y_i - x_i^T \theta)/\sigma)^2,
\]

that is

\[
\Gamma(T_n^{(LS)}) = \min\{\Gamma(\theta) : \theta \in \Omega}\.
\]

Its optimality is stated by the following result.

**PROPOSITION 3.1** (Gauss-Markov Theorem; see Scheffé, 1959, p.14)

Under the assumptions

(3.1.GM1) \( E e_i = 0 \) \( i=1,\ldots,n \),

(3.1.GM2) \( \text{cov}(e_1, \ldots, e_n) = \sigma^2 I \),

every estimable function \( b^T \theta \) has a unique linear estimate which has minimum variance in the class of all unbiased linear estimates. It is given by \( b^T T_n^{(LS)} \), where \( T_n^{(LS)} \) is any set of least
squares estimates.
If in addition the errors are normally distributed, then this estimator has minimum variance among all unbiased estimates.

Some remarks about this result are in order.
Linearity is a drastic restriction: many maximum likelihood estimators (for example, under the Cauchy distribution) are not linear.
If we want to show that the least squares estimator is optimal in the class of all unbiased estimators, we have to assume that the errors are normally distributed. Therefore, the restriction to linear estimators can be justified only by normality (or simplicity).
The normal model is never exactly true and in the presence of small departures from the normality assumption on the errors, the least squares procedures (estimators and tests) lose efficiency drastically (Huber, 1973; Hampel, 1973a, 1978a, 1980; Schrader and Hettmansperger, 1980). Thus, one would prefer to have procedures which are only nearly optimal at the normal model but which behave well in a certain neighbourhood of it. Such procedures are commonly called robust.

In Section 3.2 we shall review the robust estimators for linear models and in the following chapters (4, 5, 7) we shall develop new robust test procedures for linear models using the infinitesimal approach.
3.2 THE ROBUST ESTIMATION THEORY IN LINEAR MODELS

In this section we assume the linear model of Section 3.1 with normal errors \( (\varepsilon = \Phi) \). (Note that \( \Phi \) may be replaced by a distribution function satisfying the conditions (2.4.F1) and (2.4.F2) of Subsection 2.4a.)

3.2a Huber estimators

In 1973 Huber extended his results on robust estimation of a location parameter to the case of linear models. Basically, he proposed to compute weighted least squares estimates with weights (redefined iteratively) of the form

\[
    w_i = \min(1, c/|r_i|) ,
\]

where \( r_i \) is the \( i \)th residual and \( c \) is a positive constant.

More generally, Huber proposed M-estimators \( T_n \) defined by

\[
    \Gamma(T_n) = \min\{\Gamma(\theta) : \theta \in \Theta\} ,
\]

where

\[
    \Gamma(\theta) = \sum_{i=1}^{n} \rho\left((y_i - x_i^T\theta)/\sigma\right) ,
\]

for some function \( \rho : \mathbb{R} \rightarrow \mathbb{R}^+ \).

If \( \rho \) has a derivative \( \rho' = \psi \), \( T_n \) satisfies the system of equations

\[
    \sum_{i=1}^{n} \psi((y_i - x_i^T T_n)/\sigma) x_i = 0. \tag{3.2.4}
\]
REMARKS

1. Actually one has to estimate the scale parameter $\sigma$.
   A popular (and very robust) estimate is the median deviation (see Hampel, 1974, p. 388); another possible choice is the Proposal 2 estimate of Huber (see Huber, 1981, p. 137).

2. The least squares estimator is defined by the function $\rho(r) = r^2/2$. Moreover, the Huber estimator defined by the weights (3.2.1) can be obtained putting $\rho(r) = \rho_C(r)$ in (3.2.3) ($\psi(r) = \psi_C(r)$).

3. Note that $M$-estimation based on a dispersion function of the form (3.2.3) is related to a metric.

Let

$$v(u,v) := \sum_{i=1}^{n} \rho\left(\frac{(u(i) - v(i))/\sigma}{\sigma}\right),$$

$$u = (u(1), \ldots, u(n))^T; \; v = (v(1), \ldots, v(n))^T.$$

McKean and Schrader (1980) show that in most cases there exists a monotone function $f$ such that $f(v(u,v))$ is a metric on $\mathbb{R}^n$. This $f$ exists for $\rho = \rho_C$ (Huber estimator) and for $\rho$ defining Hampel three-part-reddescending estimator (Hampel, 1974).

In order to compare the robustness properties of the Huber estimator we compute its influence function at the model distribution $H_\theta(x,y)$ with density (ignoring the scale)

$$h_\theta(x,y) = \phi(y - x^T \theta) \kappa(x) \quad \text{(Section 3.1)}.$$

Using the formula (2.1.11) with
z=(x,y), F_0(z)=H_0(x,y), \zeta(z,\theta)=\psi_c(y-x^T\theta)x

we obtain

\[ \text{IF}(x,y;\psi_c,H_0) = \psi_c(y-x^T\theta)M^{-1}x, \]

(3.2.5)

where the matrix \( M \) is defined by

\[ M = (E_{\psi_c')}(E_{xx^T}) . \]

We see that this influence function depends on \( y \) only through \( r:=y-x^T\theta \). Following Hampel (1973a, 1978b) we can rewrite IF as a product of two factors, namely the (scalar) influence of the residual (IR) and the (vector-valued) influence of position in factor space (IP):

\[ \text{IF}(x,y;\psi_c,H_0) = \text{IT}(x,r;\psi_c,H_0) = \text{IR}(r;\psi_c,H_0) \cdot \text{IP}(x;\psi_c,H_0), \]

(3.2.6)

where

\[ \text{IR}(r;\psi_c,H_0) = \psi_c(r)/(E_{\psi_c}) \]

\[ \text{IP}(x;\psi_c,H_0) = (Exx^T)^{-1}x . \]

The factorization (3.2.6) is unique if we define IR as the influence function of the corresponding M-estimator of location (defined by \( \psi_c \)).

IT is called the total influence on the estimator at \( H_0 \).

Although \( \text{IR}(r;\psi_c,H_0) \) is bounded (this means a bounded influence of the residual on the estimator and represents an improvement of least squares estimators from the robustness point of view), the influence of position in factor space is unbounded. Thus, a single \( x_i \), which is an outlier in the
factor space, will almost completely determine the fit (Hampel, 1973a).
In this sense the Huber estimator is only the first step in the robustification of an estimator. In order to cope with problems caused by outlying points in factor space, we need more refined estimators.

3.2b General M-estimators

General M-estimators for linear models are defined implicitly by the following system of equations (see Maronna and Yohai, 1981)

\[ \sum_{i=1}^{n} \eta(x_i, (y_i - x_i^T \hat{\beta})/\sigma) x_i = 0, \]  

(3.2.7)

for an appropriate class of functions \( \eta: \mathbb{R}^p \times \mathbb{R} \to \mathbb{R} \).
\( \eta(x, \cdot) \) is typically odd for all \( x \in \mathbb{R}^p \) and positive on \( \mathbb{R}^+ \).
Note that (3.2.7) is an extension of (3.2.4).
There have been several proposals for choosing \( \eta \). For a stimulating discussion we refer to the papers by Krasker and Welsch (1979), Welsch (1979). The important choices of \( \eta \) are of the form

\[ \eta(x, r) = w(x) \cdot \psi(r \cdot s(x)), \]

for appropriate functions \( \psi: \mathbb{R} \to \mathbb{R} \) and \( w: \mathbb{R}^p \to \mathbb{R}^+ \),
\( s: \mathbb{R}^p \to \mathbb{R}^+ \) (weight functions).
Huber (1973) uses \( w(x) = 1 \), \( s(x) = 1 \).
Mallows (see Hill, 1977) uses \( s(x) = 1 \).
Andrews (see Hill, 1977) uses \( w(x) = 1 \).
Hill and Ryan (see Hill, 1977) use \( s(x) = w(x) \).

Schweppe (see Handschin, Schweppe, Kohlas, Piechter, 1975) uses \( s(x) = 1/w(x) \).

Now, suppose for simplicity, \( \sigma = 1 \).

The influence function of a general M-estimator defined by \( \eta \) at a distribution \( H \) (on \( \mathbb{R}^p \times \mathbb{R} \)) is given by

\[
IF(x, y; \eta, H) = \eta(x, y - x^T \cdot T(H)) \cdot B^{-1}(\eta, H) x,
\]

(3.2.8)

where \( T(H) \) is the functional defined by

\[
E_H \eta(x, y - x^T \cdot T(H)) x = 0
\]

and

\[
B(\eta, H) := E_H \eta'(x, y - x^T \cdot T(H)) xx^T,
\]

\[
\eta'(x, r) := (\partial \eta / \partial r) \eta(x, r).
\]

Maronna and Yohai (1981) show, under certain conditions, that these estimators are consistent and asymptotically normal with asymptotic covariance matrix

\[
V(\eta; H) = E_H IF(x, y; \eta, H) \cdot IF^T(x, y; \eta, H)
\]

\[
= B^{-1}(\eta, H) \cdot D(\eta, H) \cdot B^{-1}(\eta, H),
\]

(3.2.9)

where

\[
D(\eta, H) := E_H \eta^2(x, y - x^T \cdot T(H)) xx^T.
\]

At the model distribution \( H = H_0 \) we obtain:

\[
IF(x, y; \eta, H_0) = \eta(x, y - x^T \theta) M^{-1} x
\]

(3.2.10)

\[
V(\eta; H_0) = M^{-1} Q M^{-1},
\]

(3.2.11)

where
\[ M := E \eta'(x,r)x x^T = \int \eta'(x,r)x x^T d\phi(r) dK(x) \]
\[ Q := E \eta^2(x,r)x x^T = \int \eta^2(x,r)x x^T d\phi(r) dK(x) . \]

(The exact conditions will be discussed in detail in the framework of the test problem, see Chapter 4.)

There are two important definitions of sensitivity. Hampel (1974, 1978b) considers the supremum of the Euclidean norm of IF (see (2.1.3) and (3.2.10)):

\[ \gamma^*(\eta, H_0) = \sup \{ |IF(x,y; \eta, H_0)| : x,y \} \]
\[ = \sup \{ |\eta(x,y - x_0^T)| \cdot |M^{-1}x| : x,y \} . \quad (3.2.12) \]

Krasker and Welsch (1979) and Stahel (1981) propose to measure the IF with respect to the norm of the inverse of the asymptotic covariance matrix of the estimator:

\[ \gamma^*_\text{self}(\eta, H_0) := \sup \{ \| IF(x,y; \eta, H_0) \|_{V^{-1}} : x,y \} \]
\[ = \sup \{ (IF^T(x,y; \eta, H_0) \cdot V^{-1}(\eta; H_0) \cdot IF(x,y; \eta, H_0))^{1/2} : x,y \} \]
\[ = \sup \{ |\eta(x,y - x_0^T)| \cdot (x^T Q^{-1} x)^{1/2} : x,y \} . \quad (3.2.13) \]

\[ \gamma^*_\text{self} \] is called the self-standardized sensitivity. Note that \[ \gamma^*_\text{self} \] is invariant to the coordinate system.

Now consider Hampel's optimality criterion (2.1.12). To obtain an optimally robust estimator in the class of general M-estimators, we look for an estimator that minimizes the trace of the asymptotic covariance matrix at the model, under the side condition of a bounded sensitivity.
If the gross-error-sensitivity is measured by the Euclidean norm ($\gamma^*$, see (3.2.13)), it turns out that the optimally robust estimator is the **Hampel-Krasker estimator** which is defined by a $\eta$-function of Schwepe's form

$$\eta_{HK}(x,r) = |Ax|^{-1} \psi_c(r|Ax|), \quad (3.2.14)$$

where the matrix $A$ is defined implicitly by

$$A^{-1} = E((2\psi(c/|Ax|)-1)xx^T) \quad (3.2.15)$$

$$= En_{HK}^1(x,r)xx^T)$$

and $c$ is a positive constant (see Krasker, 1980).

If the gross-error-sensitivity is measured by $\gamma_{self}^*$ (see (3.2.13)), Krasker and Welsch (1979) show that the estimator defined by

$$\eta_{self}(x,r) = |A_{self}^{-1}x| \psi_c(r|A_{self}^{-1}x|), \quad (3.2.16)$$

where $A_{self}$ is any root of

$$Q_{self}^{-1} = (En_{self}^2(x,r)xx^T)^{-1} = A_{self}^TA_{self}$$

(defined implicitly) and $c$ is a positive constant, satisfies the first-order necessary condition for efficiency, that is, if there is an estimator that minimizes the asymptotic covariance matrix under the condition of a bounded self-standardized sensitivity, this is of the form $\eta_{self}^*$.

On the other hand, Stahel (1981, p.32 and 137) proves that $\eta_{self}$ is admissible with respect to the asymptotic covariance matrix, subject to a bound on $\gamma_{self}^*$. 
3.2c The change-of-variance function of general M-estimators

In this Subsection we extend the definition of change-of-variance function (see 2.4a) to the regression situation and we solve a related Hampel's problem.

Consider the general M-estimators given in Subsection 3.2b.

**DEFINITION 3.2**

Let $\mathcal{V}(\eta; H)$ be defined by (3.2.9). We define the *change-of-variance function* of an M-estimator corresponding to $\eta$ at the model distribution $H_0$ by the matrix

$$CVF(x, y; \eta, H_0) := (\partial \partial_\epsilon)\mathcal{V}(\eta; (1-\epsilon)H_0 + \epsilon \delta(x, y))|_{\epsilon=0},$$

for all $x, y$ where this expression exists.

Let $D(x; \eta) \subset \mathbb{R}$ be the finite set of points where $\eta'(x, \cdot)$ is not defined. After some elementary calculations, applying Definition 3.2, we obtain the following results

$$CVF(x, y; \eta, H_0) = M^{-1}Q M^{-1} + \eta^2(x, y-x^T \theta) \cdot M^{-1}xx^T M^{-1} - \eta'(x, y-x^T \theta) \cdot (M^{-1}xx^T M^{-1}Q M^{-1} + M^{-1}Q M^{-1}xx^T M^{-1})$$

(3.2.17)

and

$$\text{tr} \ CVF(x, y; \eta, H_0) = \text{tr} \ \mathcal{V}(\eta; H_0) + \eta^2(x, y-x^T \theta) \cdot |M^{-1}x|^2 - 2\eta'(x, y-x^T \theta) \cdot (M^{-1}x)^T M^{-1}Q(M^{-1}x),$$

(3.2.18)

for all $(x, y)$ such that $y-x^T \theta \notin D(x; \eta)$.

In analogy with $\gamma^*$ we propose the following measures of sensitivity.
DEFINITION 3.3

\[ \kappa^*(\eta, H_0) := \sup \{ \text{tr CVF}(x, y; \eta, H_0)/\text{tr } V(\eta; H_0) : y - x^T \theta \notin D(x; \eta) \} \]

change-of-variance sensitivity  \hfill (3.2.19)

and

\[ \kappa_{\text{self}}^*(\eta, H_0) := \sup \{ \text{tr} [\text{CVF}(x, y; \eta, H_0) \cdot V^{-1}(\eta; H_0)] : y - x^T \theta \notin D(x; \eta) \}
\]

self-standardised change-of-variance sensitivity. \hfill (3.2.20)

Applying (3.2.19) and (3.2.20) we obtain

\[ \kappa^*(\eta, H_0) = \sup \{ 1 + \eta^2 (x, y - x^T \theta) \cdot |M^{-1} x|^2 /\text{tr } V(\eta; H_0) \]

\[ - 2 \eta' (x, y - x^T \theta) \cdot (M^{-1} x) \cdot (M^{-1} Q(M^{-1} x)) /\text{tr } V(\eta; H_0) : y - x^T \theta \notin D(x; \eta) \} \]

\hfill (3.2.21)

and

\[ \kappa_{\text{self}}^*(\eta, H_0) = \sup \{ p + \eta^2 (x, y - x^T \theta) \cdot x^T Q^{-1} x - 2 \eta' (x, y - x^T \theta) \cdot x^T M^{-1} x : y - x^T \theta \notin D(x; \eta) \} \]

\hfill (3.2.22)

Note that both definitions (3.2.19) and (3.2.20) extend Definition 2.9 to the regression case and that \( \kappa_{\text{self}}^* \) is invariant to coordinate system.

On the other hand, \( \kappa^*(\eta, H_0) \) describes the infinitesimal behaviour of the asymptotic mean square error:

\[ \kappa^*(\eta, H_0) = \sup \{ (\partial / \partial \epsilon) \log \text{tr } V(\eta; (1-\epsilon) H_0 + \epsilon \delta(x, y)) \big|_{\epsilon=0} : x, y \}
\]

\[ = \sup \{ (\partial / \partial \epsilon) \log E|IP|^2 \big|_{\epsilon=0} : x, y \}
\]

(see Subsection 2.1c).
Let us now consider the following Hampel type problem.

Find an $M$-estimator (defined by $\eta$) which minimizes
\[
\text{tr } V(\eta; H_0) \text{ within the class of Fisher-consistent general } M\text{-estimators, under the side condition}
\]
\[
\kappa^*(\eta; H_0) \leq \kappa_0. \tag{3.2.23}
\]

The next proposition shows that, under certain conditions, the Hampel-Krasker estimator solves the problem (3.2.23).

**PROPOSITION 3.2**

Assume either $p=1$ or

\[
\begin{cases}
(i) \text{ the distribution of } x \text{ is radially symmetric, and} \\
(ii) \text{ we consider } \eta's \text{ that depend on } x \text{ only through } |x|.
\end{cases}
\]

Then, the Hampel-Krasker estimator minimizes the trace of the asymptotic covariance matrix within the class of Fisher-consistent general $M$-estimators, under the side condition of a bounded change-of-variance sensitivity.

**Proof.**

We prove the assertion under the conditions (i) and (ii).

The proof is similar under $p=1$.

Without loss of generality, suppose $\theta=0$.

We have to solve the following problem.

For a given $\kappa$ minimize

\[
\tilde{V}(\eta) = \text{tr } (M^{-1}Q M^{-1}),
\]

under the side condition
\[ 1 + \eta^2(x,r) |z|^2 / \tilde{V}(\eta) - 2\eta^1(x,r) \cdot z^T M^{-1} Q z / \tilde{V}(\eta) \leq \kappa \] (3.2.24)

for all \( x, r \in \mathcal{P}(x; \eta) \), where
\[ M = M(\eta) = E n^1(x,r) xx^T, \]
\[ Q = Q(\eta) = E n^2(x,r) xx^T, \]
\[ \tilde{V}(\eta) = tr (M^{-1} Q M^{-1}) = E n^2(x,r) |z|^2, \]
\[ z = z(\eta) = M^{-1} x. \]

Let \( c > 0 \) be such that \( \kappa^*(\eta_{HK}, H_0) = \kappa \).

(This \( c \) exists for sufficiently large \( \kappa \), since
\[ \kappa^*(\eta_{HK}, H_0) = 1 + c^2 / E \psi c^2(r |z|) \).

The \( M \)-estimator corresponding to \( \eta \) is defined only up to the multiplication of the defining equation by an arbitrary positive constant. Since \( M \) and \( Q \) are proportional to the identity matrix under (i) and (ii), we may assume without loss of generality
\[ M = M(\eta) \equiv A^{-1} = M(\eta_{HK}) \quad \forall \eta, \] (3.2.25)
and
\[ z(\eta) = M^{-1}(\eta) x = Ax = z \quad \forall \eta, \]

where \( A \) is the defining matrix of the Hampel-Krasker estimator (see (3.2.15)).

Moreover,
\[ \tilde{V}(\eta) = E n^2(x,r) |z|^2 \]
\[ = E (\eta(x,r) |z| - r |z|)^2 - E r^2 |z|^2 + 2E n(x,r)x \cdot |z|^2, \]
and, by integration by parts,
\[ E n(x,r)x |z|^2 = \text{tr} A \quad (\text{independently of } \eta!), \] (3.2.26)
Thus, the original problem is equivalent to minimizing
\[ E(\eta(x,r) - r|z|)^2 \]
under the conditions (3.2.24) and (3.2.25).

Suppose there exists \( \eta_1(x,r) \) which satisfies conditions (3.2.24) and (3.2.25) and such that
\[ E(\eta_1(x,r) - r|z|)^2 < E(\eta_{HK}(x,r) - r|z|)^2. \tag{3.2.27} \]
Then, there exist \( r_0, x_0 \) such that
\[ (\eta_1(x_0,r_0) - r_0|z_0|)^2 < (\eta_{HK}(x_0,r_0) - r_0|z_0|)^2, \]
with \( z_0 := Ax_0 \).
Now the proof continues exactly in the same way as in the location case (see Rousseeuw, 1981, p. 47). We give a sketch of it.

Without loss of generality, suppose \( r_0|z_0| > 0 \).
Then, \( r_0|z_0| > c \) and \( \eta_1(x_0,r_0) - r_0|z_0| > \psi_c(r_0|z_0|) = c \).
Moreover, \( \eta_1(x_0,r_0) > 0 \) (otherwise we obtain a contradiction with (3.2.24)) and one can prove that
\[ \eta_1(x_0,r) - r|z_0| > \eta_1(x_0,r_0) - r_0|z_0| > c, \forall r > r_0. \tag{3.2.28} \]
Combining (3.2.24) and (3.2.28) we get for sufficiently large \( r \), say \( r > R_0 \),
\[ \eta_1(x_0,r)/\eta_1^2(x_0,r) - c^2/|z_0|^2 > \frac{1}{2}|z_0|^2/\beta_0(\eta_1) > 0, \tag{3.2.29} \]
where \( \beta_0(\eta_1) := z_0^T M_n^{-1}(\eta_1) Q(\eta_1) z_0 \).
Now define for \( r > R_0 \)
\[ P(x_0,r) := -(|z_0|/c) \cdot \coth^{-1}(\eta_1(x_0,r_0) - r_0|z_0|/c) \).
Because of (3.2.29) \( P'(x_0,r) > 0 \), hence
\[ P(x_0, r) - P(x_0, R_0) \geq \frac{1}{2}|z_0|^2 \cdot (r-R_0)/\beta_0(\eta_1) . \]

From this inequality we obtain the final contradiction:

\[ 0 < \coth^{-1}(\eta_1(x_0, R_0) |z_0|/c) + \frac{1}{2} R_0 \cdot c |z_0|/\beta_0(\eta_1) \]
\[ - \frac{1}{2} c \cdot r |z_0|/\beta_0(\eta_1) \to -\infty , \text{ as } r \to \infty . \]

\[
\]

REMARK

The proof of Proposition 3.2 goes through if we substitute \( G=\phi \) by a distribution function which satisfies the conditions (2.4.F1) and (2.4.F2) of Subsection 2.4a.

We conjecture that a similar result can be proved for \( \eta_{\text{self}} \) (with respect to \( \kappa^*_\text{self} \)).

CONJECTURE

For each \( \kappa_\theta > 2p \), there exists a unique positive constant \( c \) such that \( \kappa^*(\eta_{\text{self}}, \theta) = \kappa_\theta \) and \( \eta_{\text{self}} \) (defined by (3.2.16)) satisfies the first-order necessary condition for efficiency, under the side condition \( \kappa^*(\eta, \theta) \leq \kappa_\theta \).

The existence of such a positive constant \( c \) can be proved very easily.

Let \( \kappa_\theta > 2p \) and define \( c:= (\kappa_\theta - p)^{1/2} \).

Then, \( c > p^{1/2} \) and by Theorem 2 of Krasker and Welsch (1979)
(see also formula 5.10 of that paper) \( \eta_{\text{self}} \) exists.

Moreover, by (3.2.16) and (3.2.22)
\[ \kappa^*(n_{\text{self}}, R_\theta) = \sup \{ p + \psi_c^2 (r(x^T_{Q_{\text{self}}} x)^{1/2}) \]
\[ - 2 \cdot 1_{\{r(x^T_{Q_{\text{self}}} x)^{1/2} \leq c\}} x^T_{M_{\text{self}}} x : x, r \} \]
\[ = \sup \{ p + \psi_c^2 (r(x^T_{Q_{\text{self}}} x)^{1/2}) : x, r \} \]
\[ = p + c^2 \]
\[ = \kappa_\theta \]. \]
3.3 THE TEST PROBLEM IN LINEAR MODELS

Consider the general linear model of Section 3.1 and suppose we want to test the linear hypothesis

\[ \theta_j(\theta) = 0, \quad j=q+1, \ldots, p, \]

where \( \theta_{q+1}, \ldots, \theta_p \) are \((p-q)\) independent estimable functions and \(0<q<p\).

Through a transformation of the parameter space we can reduce this hypothesis to

\[ H_0 : \theta^{(q+1)} = \ldots = \theta^{(p)} = 0. \tag{3.3.1} \]

Let \( \omega \) be the subspace of \( \Omega \) obtained imposing the condition \( H_0 \). If the errors are normally distributed, the classical \( F \)-test for testing \( H_0 \) is optimal in several senses (see Lehmann, 1959 and for a summary Seber, 1980, pp. 34-37). It is defined by the critical region

\[ \{ F > F_{p-q,n-p;1-\alpha} \}, \]

where \( \alpha \) is the level of the test, \( F_{p-q,n-p;1-\alpha} \) is the \((1-\alpha)\) quantile of the \( F \)-distribution with \((p-q)\) and \((n-p)\) degrees of freedom and \( F \) is the test statistic. \( F \) is defined as follows

\[
F = \left| \frac{(X(T_\Omega)_n - X(T_\omega)_n)/\delta}{(p-q)} \right|^2 \\
= \sum_{i=1}^{n} \left[ \frac{(y_i - x_i^T(T_\omega)_n)/\delta}{(p-q)} \right]^2 - \left( \frac{(y_i - x_i^T(T_\Omega)_n)/\delta}{(p-q)} \right)^2 \\
\]

where \((T_\Omega)_n\) and \((T_\omega)_n\) are the least squares estimates of \( \theta \).
in the full (\( \Omega \)) and reduced model (\( \omega \)) respectively and

\[
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^T \hat{\Theta}_n)^2 / (n-p)
\]

is the least squares (unbiased) estimate of \( \sigma^2 \).

The F-test is equivalent to the corresponding likelihood ratio test (see Section 5.1 for details). Since the likelihood ratio criterion is based on a very general and natural principle (for the theoretical motivations see, for example, Akaike, 1973), we shall work with it in order to construct a robust version of the F-test.

From a robustness point of view, the classical test procedures based on the least squares estimates in the full and reduced model have similar problems as the least squares estimators. Although the F-test (for testing \( H_0 \)) is relatively robust with respect to the level (being an approximation of the permutation test, see Scheffé, 1959, p.313), it does lose power rapidly in the presence of departures from the normality assumption on the errors (Hampel, 1973a, 1980; Schrader and Hettmansperger, 1980).

Whereas robust estimation theory in linear models has recently received more and more attention, the test problem has been somewhat neglected. On the other hand, the need for robust test procedures is obvious because one cannot estimate the parameters robustly and apply unmodified classical test procedures for testing hypotheses about them.
Only recently, Schrader and McKean(1977) and Schrader and Hettmansperger(1980) proposed a new class of tests for linear models; we discuss these in Section 4.4a and 5.1. These tests are based on Huber estimates in the full and reduced model. Nevertheless, this is only the first step for finding a robust version of the F-test. Like Huber estimates, these tests do not overcome problems caused by situations where the fit is mostly determined by outlying points in the factor space. In order to construct tests which fit bounded-influence estimation, we introduce in Chapter 4 new (extended) classes of tests and we study their asymptotic properties. Moreover, we derive in Chapter 5 optimally bounded influence tests which are the counterpart to optimally bounded influence estimators.

**REMARK**

Our work is devoted to distributional robustness and we do not investigate other departures from the assumptions in the linear model.

For instance, Andersen, Jensen and Schou(1981) study the behaviour of the test procedures in a two-way analysis of variance model with correlated errors.
4. A NEW GENERAL CLASS OF TESTS
FOR LINEAR MODELS

In this chapter we introduce a new general class of tests and we study their asymptotic properties (influence function and asymptotic distribution). These tests are the natural counterpart to bounded-influence estimators. Assume throughout this chapter the general linear model of Section 3.1 with normal errors (G=Φ) and the corresponding test problem (Section 3.3).
4.1 **THE $\tau$-TEST**

The aim of this section is to define a class of tests that can be viewed as an extention of the log-likelihood ratio test for linear models. To do this let us first define the following class of functions

$$\tau : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^+ , \ (x,r) \mapsto \tau(x,r)$$

with the following properties:

(4.1.TAU) $\forall x \in \mathbb{R}^p , \ r \in \mathbb{R} \ , \ \tau(x,r) \not= 0 , \ \tau(x,r) \geq 0 , \ \tau(x,0) = 0.$

(4.1.ETA1) $\forall x \in \mathbb{R}^p , \ \tau(x,\cdot) \text{ is differentiable. Let}$

$$\eta(x,r) := (\partial \partial r) \tau(x,r) \text{ and assume}$

(i) $\forall x \in \mathbb{R}^p , \ \eta(x,\cdot) \text{ is continuous and odd},$

(ii) $\forall x \in \mathbb{R}^p , \ r \in \mathbb{R}^+ , \ \eta(x,r) \geq 0.$

(4.1.ETA2) $\forall x \in \mathbb{R}^p , \ \eta(x,\cdot) \text{ is differentiable on } \mathbb{R} \setminus D(x;\eta), \text{ where}$

$D(x;\eta)$ is a finite set. Let

$$\eta'(x,r) := (\partial \partial r) \eta(x,r) \quad \text{if} \ x \in \mathbb{R}^p , \ r \in \mathbb{R} \setminus D(x;\eta)$$

$$:= 0 \quad \text{otherwise}.$$

Assume:

(i) $\forall x \in \mathbb{R}^p , \ \sup \{ |\eta'(x,r)| : r \in \mathbb{R} \} < \infty,$

(ii) $M := \text{E} \eta'(x,r) xx^T$ is nonsingular.
DEFINITION 4.1

We introduce the class of tests \( \{ \tau \} \) by means of the test statistics

\[
S_n^2(x_1, \ldots, x_n; y_1, \ldots, y_n) := 2^{-1}(p-q)(q-p) \sum_{i=1}^{n} \tau(x_i, r_{wi}) - \tau(x_i, r_{\tau i}) ,
\]

(4.1.1)

where \( \tau \) satisfies the conditions (4.1.TAU), (4.1.ETA1), (4.1.ETA2),

\[
x_{wi} := (y_i - x_i^T(T_n) / \sigma , \quad r_{\tau i} := (y_i - x_i^T(T_n)^\tau / \sigma .
\]

(\( T_n \)) and (\( T_n^\tau \)) are the M-estimators in the reduced and full model, that is

\[
\Gamma((T_n \omega)_{n}) = \min \{ \Gamma(\theta) : \theta \in \omega \} ,
\]

(4.1.2)

\[
\Gamma((T_n^\tau)_{n}) = \min \{ \Gamma(\theta) : \theta \in \Omega \}
\]

(4.1.3)

with

\[
\Gamma(\theta) := \sum_{i=1}^{n} \tau(x_i, (y_i - x_i^T\theta) / \sigma) .
\]

(4.1.4)

"Large" values of \( S_n^2 \) are significant.

(In order to give a critical region we shall compute the asymptotic distribution of \( S_n^2 \) under \( H_0 \), see Section 4.3.)

(\( T_n \)) and (\( T_n^\tau \)) fulfill the equations

\[
\sum_{i=1}^{n} (x_i, r_{wi}) x_i^* = 0
\]

(4.1.5)

\[
\sum_{i=1}^{n} (x_i, r_{\tau i}) x_i = 0
\]

(4.1.6)

Note that \( x^* = (x^{(1)}, \ldots, x^{(p-q)}, 0, \ldots, 0)^T \) (see Notation) and

(\( T_n \)) = ((\( T_n \))^{(1)}, \ldots, (\( T_n \))^{(q)}, 0, \ldots, 0)^T \) (under \( H_0 \) the last (\( p-q \)) components of \( \theta \) equal 0!).
EXAMPLES

Define the following functions
\[ w : \mathbb{R}^{d} \rightarrow \mathbb{R}^{+} \]
\[ \rho : \mathbb{R} \rightarrow \mathbb{R}^{+} ; \psi : \mathbb{R} \rightarrow \mathbb{R} , r \mapsto \psi(r) := (\partial \psi / \partial r) \rho(r) \]

Some choices of \( \tau \) are of the form \( \tau(x,r) = \tilde{w}(x) \rho(r \cdot s(x)) \) for certain functions \( \tilde{w}(x) \) and \( s(x) \). They correspond to the estimators given in Subsection 3.2b.

<table>
<thead>
<tr>
<th>( \tau(x,r) )</th>
<th>( \eta(x,r) )</th>
<th>estimator corresp. to ( \eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r^2/2 )</td>
<td>( r )</td>
<td>least squares</td>
</tr>
<tr>
<td>( \rho(r) )</td>
<td>( \psi(r) )</td>
<td>Huber</td>
</tr>
<tr>
<td>( w(x) \rho(r) )</td>
<td>( w(x) \psi(r) )</td>
<td>Mallows</td>
</tr>
<tr>
<td>( w(x) \rho(r/w(x)) )</td>
<td>( \psi(r/w(x)) )</td>
<td>Andrews</td>
</tr>
<tr>
<td>( \rho(r \cdot w(x)) )</td>
<td>( w(x) \psi(r \cdot w(x)) )</td>
<td>Hill and Ryan</td>
</tr>
<tr>
<td>( w^2(x) \rho(r/w(x)) )</td>
<td>( w(x) \psi(r/w(x)) )</td>
<td>Schweppete</td>
</tr>
</tbody>
</table>

The conditions (4.1.TAU), (4.1.ETA1), (4.1.ETA2) can then be written in terms of \( w \) and \( \rho \).

REMARK

Actually, one usually has to estimate the scale parameter \( \sigma \).

A suitable way is to estimate \( \sigma \) in the full model, taking the median deviation (see Hampel, 1974, p.388) or the Proposal 2 estimate of Huber (see Huber, 1981, p.137). More precisely, for a given real function \( \chi \), one has to solve (4.1.5), (4.1.6) and

\[ \sum_{i=1}^{n} \chi(|r_{\Omega i}|) = 0 \]  

(4.1.7)
with respect to $(T_0)_n$, $(T_\omega)_n$ and $\sigma$.

The median deviation corresponds to $\chi(|r|) = \text{sign}(|r|-c)$, where $c$ is a constant. The Proposal 2 estimate of Huber can be obtained choosing $\chi(|r|) = (\psi_c(r))^2 - \beta$, where $c$ and $\beta$ are certain constants.

From now on we put for simplicity $\sigma = 1$. 
4.2 THE INFLUENCE FUNCTION

The next proposition gives us the influence function of the test statistic $S_n$ at the null hypothesis. (Note that, under the hypothesis, $\theta = \theta^*$, so $H_{\theta^*}$ is the model distribution at the null hypothesis.)

PROPOSITION 4.1

Let $S_n$ be the test statistic defining the $r$-test. Assume (4.1.TAU), (4.1.ETA1), (4.1.ETA2) and the following conditions

(4.2.IF1) $\forall \alpha \in \Omega \cap \mathbb{R}^D$, $\exists J(\alpha) := \int \eta(x, y-x^T \alpha) x dH_{\theta^*}(x, y)$

$\exists (\exists \delta \alpha) J(\alpha)$, continuous.

(4.2.IF2) $J(\theta^*) = 0$.

Then, the influence functions of $(T_\omega)_n$, $(T_\Omega)_n$, $S_n$ (viewed as functionals $T_\omega$, $T_\Omega$, $S$) exist and equal

(i) $IF(x, y; T_\omega, H_{\theta^*}) = \eta(x, y-x^T \theta^*) \cdot (M^*)^+ x$,

(ii) $IF(x, y; T_\Omega, H_{\theta^*}) = \eta(x, y-x^T \theta^*) \cdot M^{-1} x$,

(iii) $IF(x, y; S, H_{\theta^*}) = \left| \eta(x, y-x^T \theta^*) \right| \cdot ((x^T (M^{-1} - (M^*)^+) x) / (p-q))^{1/2}$.

Proof.

In order to compute the influence function of the test statistic, we define the following functional $S$ on the space of the distribution functions on $\mathbb{R}^D \times \mathbb{R}$:

\[ S(H) = (2(p-q)^{-1} \int \left[ T(u,v-u^T T_\omega (H)) - T(u,v-u^T T_\Omega (H)) \right] dH(u,v) \right)^{1/2} \]

(4.2.1)

where \( H \) is an arbitrary distribution function on \( \mathbb{R}^p \times \mathbb{R} \) and \( T_\omega, T_\Omega \) fulfill the system of equations

\[
\int \eta(u,v-u^T T_\omega (H)) u^* \, dH(u,v) = 0, \\
\int \eta(u,v-u^T T_\Omega (H)) u \, dH(u,v) = 0.
\]

(Note that \( \forall H, T_\omega^{(j)}(H) = 0 \), for \( j=q+1, \ldots, p \).)

Let \( H_\theta(x,y) \) be the distribution function of \((x_1,y_1)\) in the given linear model and let \( H^{(n)} \) be the empirical distribution function. Then, we have

\[
S_n(x_1, \ldots, x_n; y_1, \ldots, y_n) = S(H^{(n)})
\]

and, by (4.2.IF2), (Fisher-consistency)

\[
S(H_{\theta*}) = 0.
\]

Now, assertions (i) and (ii) of this proposition follow from Theorem G11.1 of Stahel(1981, p.116), with \( P=H_{\theta*} \). His conditions "a", "b", "c" follow from (4.2.IF1), (4.2.IF2) and condition "d" from (4.1.ETA2)(i). Finally, condition "e" follows from (4.1.ETA2), since \( \{ (x,y) : \gamma - x^T \theta \in \mathcal{E}(x; \eta) \} \) is a regular hyperplane in his sense (Stahel,1981, p.12).

To show (iii), denote by \( \delta(x,y) \) the distribution on \( \mathbb{R}^p \times \mathbb{R} \) that puts mass 1 at \((x,y)\) and define the following \( \varepsilon \)-contaminated distribution

\[
H_{\theta*; \varepsilon}(x,y) := (1-\varepsilon)H_{\theta*} + \varepsilon \delta(x,y).
\]

After a straightforward computation we get
\[ (\partial \partial \varepsilon) S^2(\mathcal{H}_{\theta^*}; \varepsilon, (x, y)) \bigg|_{\varepsilon=0} = 0 \]

and
\[ (\partial^2 \partial \varepsilon^2) S^2(\mathcal{H}_{\theta^*}; \varepsilon, (x, y)) \bigg|_{\varepsilon=0} = \]
\[ 2 (p-q)^{-1} \left[ - \eta(x, y-x^T \theta^*) x^T \cdot IF(x, y; T_\omega, \mathcal{H}_{\theta^*}) \right. \]
\[ + \left. \eta(x, y-x^T \theta^*) x^T \cdot IF(x, y; T_\Omega, \mathcal{H}_{\theta^*}) \right] = \]
\[ 2 (p-q)^{-1} \cdot \eta^2(x, y-x^T \theta^*) (x^T M^{-1} x - x^T (M^*)^+ x) \]

Using l'Hôpital's rule twice, we obtain
\[
IF(x, y; S, \mathcal{H}_{\theta^*}) = \lim_{\varepsilon \to 0} \frac{(S(\mathcal{H}_{\theta^*}; \varepsilon, (x, y)) - S(\mathcal{H}_{\theta^*}))}{\varepsilon} = \]
\[ = \left( \lim_{\varepsilon \to 0} \frac{S^2(\mathcal{H}_{\theta^*}; \varepsilon, (x, y))}{\varepsilon^2} \right)^{1/2} \]
\[ = \left( \frac{1}{2} \cdot (\partial^2 \partial \varepsilon^2) S^2(\mathcal{H}_{\theta^*}; \varepsilon, (x, y)) \bigg|_{\varepsilon=0} \right)^{1/2} \cdot \]

From this (iii) follows and this ends the proof.

In order to show the properties of the influence function of \( S \), we first prove the following lemma we shall use also in the next section.
**LEMMA 4.2**

Let $A$, $B$ be $(p \times p)$ symmetric, positive definite matrices. Then

(i) $A^{-1} - (A^*)^+$ is positive semidefinite of rank $p-q$.

(ii) There exists a nonsingular $(p \times p)$ matrix $V$ such that

$$V^T B^{-1} V = I \quad \text{and} \quad V^T (A^{-1} - (A^*)^+) V = \Lambda,$$

where

$$\Lambda = \text{diag}(0, \ldots, 0, \lambda_{q+1}, \ldots, \lambda_p), \quad \lambda_{q+1} \geq \lambda_{q+2} \geq \cdots \geq \lambda_p > 0$$

are the $(p-q)$ positive eigenvalues of $B(A^{-1} - (A^*)^+)$

and * and $^+$ are defined in the Notation at the beginning.

**Proof.**

(i) follows easily noting that

$$\max\{x^T (A^*)^+ x / x^T A^{-1} x : x \neq 0\} =$$

$$\{\text{greatest eigenvalue of } A(A^*)^+\} = 1,$$

hence

$$x^T (A^{-1} - (A^*)^+) x \geq 0, \quad \forall x \in \mathbb{R}^p$$

and

$$\text{rank}(A^{-1} - (A^*)^+) = \text{rank}(I - A(A^*)^+) = p-q.$$

(ii) is the simultaneous diagonalization of $B^{-1}$ and

$(A^{-1} - (A^*)^+)$, see Bellman(1970), pp.58-59.
PROPOSITION 4.3

Let $M$ be given by (4.1.ETA2)(ii).

Then

(i) There exists a nonsingular lower triangular $(p \times p)$ matrix $U$ with positive diagonal elements such that

$$UU^T = M \quad \text{and} \quad U^T(M^{-1} - (M^*)^+)U = I^{**}.$$ 

Moreover, $M_{22} = U_{22}U_{22}^T$.

(ii) The function

$$\tilde{n} : \mathbb{R}^P \rightarrow \mathbb{R}^+, \quad x \mapsto \tilde{n}(x) := (x^T(M^{-1} - (M^*)^+)x)^{1/2}$$

is pseudonorm in $\mathbb{R}^P$, that is $\tilde{n}$ satisfies the conditions

$$\forall a \in \mathbb{R}, \ x,y \in \mathbb{R}^P,$$

$$\tilde{n}(0) = 0 \quad ; \quad \tilde{n}(ax) = |a|\tilde{n}(x) \quad ; \quad \tilde{n}(x+y) \leq \tilde{n}(x) + \tilde{n}(y).$$

Proof.

(i) First observe that the matrix $M$ is positive definite, since using (4.1.ETA1)(i),(ii) and performing an integration by parts we have

$$M = \int \eta'(x,r) d\Phi(r) xx^T dK(x) = \int \eta(x,r) \cdot r \cdot d\Phi(r) xx^T dK(x),$$

where the integral

$$\int \eta(x,r) \cdot r \cdot d\Phi(r)$$

is positive for all $x \in \mathbb{R}^P$.

Therefore, we can apply part(ii) of Lemma 4.2 with $A=B=M$.

There exists a regular $(p \times p)$ matrix $U$ such that

$$U^T M^{-1} U = I \quad \text{and} \quad U^T(M^{-1} - (M^*)^+)U = I^{**}. \quad (4.2.2)$$

(The positive eigenvalues of $M(M^{-1} - (M^*)^+)$ equal 1.)

(4.2.2) is equivalent to
\[ UU^T = M \quad , \quad U^T(M^*)^+U = I^* \quad . \] (4.2.3)

From (4.2.3) we obtain in the partition's notation

\[ U_{11}^T U_{11} = M_{11} \]
\[ U_{11}^T M_{11}^{-1} U_{12} = 0 \]
\[ U_{12}^T M_{11}^{-1} U_{12} = 0 \ . \]

It follows that \( U \) is the solution of the equation \( UU^T = M \), under the condition \( U_{12} = 0 \).

The Choleski decomposition of \( M \) solves this problem.

Moreover, we have

\[ M_{22,1} = M_{22} - M_{12}^T M_{11}^{-1} M_{12} \]
\[ = U_{21}^T U_{21} + U_{22}^T U_{22} - (U_{11}^T U_{21})^T (U_{11}^T U_{11})^{-1} (U_{11}^T U_{21}) \]
\[ = U_{21}^T U_{21} + U_{22}^T U_{22} - U_{21}^T U_{11}^{-1} U_{11}^T U_{21} \]
\[ = U_{22}^T U_{22} \ . \]

(ii) follows easily noting that \( M^{-1} - (M^*)^+ \) is positive semidefinite and from (i)

\[ x^T (M^{-1} - (M^*)^+) x = |(U^{-1} x)^*|^2 \ . \]
COROLLARY 4.4

If the conditions of Proposition 4.1 hold, then

(i) \( \text{IF}(x,y;S,H_{\theta^*}) = (p-q)^{-1/2} |\eta(x,y-x^{T \theta^*})_*(U^{-1}x)^*| \) ;

(ii) \( f(\text{IF}(x,y;S,H_{\theta^*}))^2 dH_{\theta^*}(x,y) = (p-q)^{-1} \text{tr} \left[ Q(M^{-1} - (M^*)^+) \right] \)
    \[ = (p-q)^{-1} \text{E}_{\eta^2}(x,r) |(U^{-1}x)^*|^2 \]

if \( Q := \text{E}_{\eta^2}(x,r)xx^T \) exists.

Proof.

(i) follows directly from Proposition 4.1(iii) and Proposition 4.3.

(ii): We have

\[
\begin{align*}
\text{f(\text{IF}(x,y;S,H_{\theta^*}))^2 dH_{\theta^*}(x,y) =} \\
(p-q)^{-1} \int_{\eta^2(x,y-x^{T \theta^*})} x^T (M^{-1} - (M^*)^+) \cdot \eta(x,y) \\
(p-q)^{-1} \text{tr} \left[ \eta^2(x,y-x^{T \theta^*})xx^T (M^{-1} - (M^*)^+) \right] \cdot dH_{\theta^*}(x,y) \\
(p-q)^{-1} \text{tr} \left[ Q(M^{-1} - (M^*)^+) \right].
\end{align*}
\]

REMARKS

1. \( \hat{\eta}(\cdot) \) defines a pseudodistance in the space of the \( x \)'s.

   It can be viewed as a measure of (pseudo)distance between the full and the reduced model.

2. We shall see later that if \( n\hat{\eta}^2/n \) has asymptotically a \( \chi^2 \) distribution under the hypothesis, \( f(\text{IF}(x,y;S,H_{\theta^*}))^2 dH_{\theta^*}(x,y) \)
   is its standardization factor (see Section 4.4).
4.3 ASYMPTOTIC DISTRIBUTIONS

In order to compute the asymptotic distribution of the test statistic defining the $t$-test, we first perform a von-Mises expansion of the functional $S^2$ defined by (4.2.1). Following von Mises (1947), we get

$$S^2(F_1) = S^2(H_{\theta*}) + (\partial \partial \varepsilon)S^2(F_{\varepsilon}) \bigg|_{\varepsilon=0} + \frac{1}{2}(\partial^2 \partial \varepsilon^2)S^2(F_{\varepsilon}) \bigg|_{\varepsilon=0} + \ldots,$$

where $F_{\varepsilon} = (1-\varepsilon)H_{\theta*} + \varepsilon F_1$.

We perform the same kind of calculations as in the proof of Proposition 4.1 and we obtain

$$S^2(H_{\theta*}) = 0,$$

$$(\partial \partial \varepsilon)S^2(F_{\varepsilon}) \bigg|_{\varepsilon=0} = 0,$$

$$(\partial^2 \partial \varepsilon^2)S^2(F_{\varepsilon}) \bigg|_{\varepsilon=0} =$$

$$2(p-q)^{-1} \int (\zeta^T(u,v) \cdot (M^{-1}-(M^*)_+) \cdot (\zeta(x,y) dF_1(x,y)) d(F_1-H_{\theta*})(u,v)$$

$$2(p-q)^{-1} (\int \zeta^T(u,v) dF_1(u,v)) \cdot (M^{-1}-(M^*)_+) \cdot (\int \zeta(u,v) dF_1(u,v)),$$

where $\zeta(u,v) := \eta (u, v-u^T \theta*)u$.

Thus, we have

$$S^2(F_1) = (p-q)^{-1} (\int \zeta^T(u,v) dF_1(u,v)) \cdot (M^{-1}-(M^*)_+) \cdot (\int \zeta(u,v) dF_1(u,v))$$

$$+ \ldots,$$

and evaluating $S^2$ at $F_1 = H^{(n)}$ = empirical distribution function, we obtain the following expansion
\[ (p-q)n S_n^2 = (p-q)n S^2(H(n)) = \]

\[ \nu_n^T (\theta^*) (M^{-1} - (M^*)^+) \nu_n (\theta^*) + \ldots, \tag{4.3.1} \]

where \( \nu_n (\theta) := n^{-1/2} \sum_{i=1}^{n} \eta(x_i, y_i - x_i^T \theta) x_i. \)

The next proposition shows that under some regularity conditions the statistic \((p-q)n S_n^2\) and the quadratic forms

\[ \nu_n^T (\theta^*) (M^{-1} - (M^*)^+) \nu_n (\theta^*) \quad \text{and} \quad n^{1/2} (\tilde{T}_\Omega)_{n}^T M_{22,1} n^{1/2} (\tilde{T}_\Omega)_{n} \]

have the same asymptotic distribution.

Recall that \( \tilde{T}_\Omega = (T_{\Omega}^{(q+1)}, \ldots, T_{\Omega}^{(p)})^T. \)

**PROPOSITION 4.5**

Consider the model \( H_\theta \), let \( S_n \) be the test statistic defining the \( \tau \)-test and define

\[ \nu_n (\theta) := n^{-1/2} \sum_{i=1}^{n} \eta(x_i, y_i - x_i^T \theta) x_i. \]

Besides (4.1.TAU), (4.1.ETA1), (4.2.1F1), (4.2.2F2) assume

(4.3.ETA3) \( \forall x \in \mathbb{R}^p, \eta(x, \cdot) \) is differentiable and

(i) \( \forall x \in \mathbb{R}^p, \sup \{ |\eta'(x, r)| : r \in \mathbb{R} \} < \infty, \)

(ii) \( M := E\eta'(x, x)x^T \) is nonsingular,

(iii) \( \forall x \in \mathbb{R}^p, y \in \mathbb{R}, \)

\[ |\eta'(x, y - x^T \alpha_1) - \eta'(x, y - x^T \alpha_2)| \leq b(x) |\alpha_1 - \alpha_2| \quad \text{and} \]

\[ E b(x) |x|^2 < \infty. \]

(4.3.AS1) \( (T_{\Omega})_{n} \rightarrow \theta^* \), \( (T_{\Omega})_{n} \rightarrow \theta^* \)

in probability (under \( H_{\theta^*} \)), as \( n \rightarrow \infty. \)
(4.3.AS2) Let

\[ v(x, y, \alpha, d) := \sup \{ |\eta(x, y - x^T \alpha_1) - \eta(x, y - x^T \alpha) | \mid x : \alpha_1, \alpha, |\alpha_1 - \alpha| \leq d \} . \]

There exist positive numbers a, b, c, d such that

(i) \[ |J(\alpha)| \geq a |\alpha - \theta^*| \text{ for } |\alpha - \theta^*| \leq d \],

(ii) \[ \text{E}v(x, y, \alpha, d') \leq \text{b} \cdot d' \quad \forall d' \geq 0, \forall \alpha, |\alpha - \theta^*| + d' \leq d \],

(iii) \[ \text{E}v^2(x, y, \alpha, d') \leq \text{c} \cdot d' \quad \forall d' \geq 0, \forall \alpha, |\alpha - \theta^*| + d' \leq d \].

(4.3.AS3) \[ E \eta^2(x, r) |x|^2 < \infty \]

Then the test statistics

\[ \frac{n \cdot S_n^2}{n} , \]

\[ \tilde{W}_n^2 := (p-q)^{-1} n^{1/2} (T_n^\Omega)^T n M_{22} n^{1/2} (T_n^\Omega)^n , \]

\[ \tilde{W}_n^2 := (p-q)^{-1} \cdot V_n^T (\theta^*) (M^{-1} - (M^*)^+) V_n (\theta^*) \]

have the same asymptotic distribution. (This will be given in the next Proposition 4.6.)

Proof.

Set

\[ L_n(\theta) := - \sum_{i=1}^{n} (x_i' y_i - x_i' 0) , \]

\[ C_n := n^{1/2} (T_n^\Omega - \theta^*) . \]

First, note that the given conditions guarantee that asymptotically, under \( H_{\theta^*} \), \( C_n \) has a multivariate normal distribution with mean 0 and covariance matrix \( M^{-1} Q M^{-1} \), where
Q:=E_n^2(x,r)xx^T (see Huber, 1967).

On the other hand, we obtain from (4.2.IF2) and (4.3.AS3) that

\[ L_n (V_n(\theta*)) \rightarrow N(0,Q), \]

as \( n \rightarrow \infty \) (central limit theorem).

Therefore, the distribution of \( V_n(\theta*) \) equals \textit{asymptotically} the distribution of \( M \cdot C_n \):

\[ V_n(\theta*) \xrightarrow{L} M \cdot C_n \]

Now, we expand \( L_n(\theta) \) using a Taylor series expansion up to the quadratic term. The linear term vanishes because of (4.1.6) and we get

\[ 2(L_n((T_n)_n) - L_n(\theta*)) = C_n^T(M + R_n)C_n, \]

where \( R_n \rightarrow 0 \) in probability by (4.3.ETA3), as \( n \rightarrow \infty \).

Thus,

\[ 2(L_n((T_n)_n) - L_n(\theta*)) \xrightarrow{L} C_n^T \cdot M \cdot C_n \xrightarrow{L} V_n(\theta*)M^{-1}V_n(\theta*). \] (4.3.2)

We can perform the same kind of calculations for the reduced model and we have

\[ 2(L_n((T_n)_n) - L_n(\theta*)) \xrightarrow{L} (V_n(\theta*))^T M^{-1}_{ll} V_n(\theta*). \] (4.3.3)

Define a \((p \times q)\) matrix \( R \) by

\[ r_{ij} = 1 \quad \text{if} \quad i=j \]
\[ = 0 \quad \text{otherwise}. \]

Then,
\[ R^T V_n(\theta^*) = V_n(\theta^*) \]  
(4.3.4)

and combining (4.3.2), (4.3.3) and (4.3.4), we obtain asymptotically

\[
(p-q) \cdot n \cdot S^2_n = 2 \left( L_n \left( (T_n) L_n \right) - L_n \left( (T_\omega) L_n \right) \right)
\]

\[
\overset{L}{=} V_n^T(\theta^*) M^{-1} V_n(\theta^*) = (V_n(\theta^*))^T M_{11}^{-1} V_n(\theta^*)
\]

\[
= V_n^T(\theta^*) M^{-1} V_n(\theta^*) - V_n^T(\theta^*) R M_{11}^{-1} R^T V_n(\theta^*)
\]

\[
\overset{L}{=} C_n \left( M - M R M_{11}^{-1} R^T M \right) C_n
\]

\[
= C_n^T Y C_n ,
\]

where \( Y \) is a \((p \times p)\) matrix defined as follows

\[ Y_{11} = Y_{12} = Y_{21} = 0 , \quad Y_{22} = M_{22} .1 \]

Finally, we have asymptotically

\[
(p-q) \cdot n \cdot S^2_n \overset{L}{=} C_n^T Y C_n = n^{1/2} (\Omega_n^T M_{22} \cdot 1 \cdot n^{1/2} (\Omega_n^T
\]

\[
\overset{L}{=} V_n^T(\theta^*) (M^{-1} - (M^*)^+) V_n(\theta^*) ,
\]

which completes the proof. \( \square \)

**REMARK**

The consistency condition (4.3.A3.1) for M-estimators in the full and reduced model can be checked using a Theorem by Huber(1967).
PROPOSITION 4.6

Let assumptions (4.1.TAU), (4.1.ETA1), (4.1.ETA2), (4.3.AS3) hold.

Replace in the conditions (4.2.IF1), (4.2.IF2), (4.3.AS2) the parameter \( \theta^* \) by a general parameter \( \theta \) and assume these conditions. Moreover, assume

\((4.3.AS1') (T_n) \rightarrow 0 \) in probability (under \( H_\theta \)), as \( n \rightarrow \infty \).

Then, under the sequence of alternatives

\[ H_{(n)} : \theta^{(j)} = n^{-1/2} \xi^{(j)} \quad j=q+1, \ldots, p, \]

where \( \Delta=(\Delta^{(1)}, \ldots, \Delta^{(p)})^T \),

the test statistic \( (p-q)W_n^2 \) has the asymptotic distribution

\[ \sum_{j=q+1}^{p} \left( \lambda_j^{1/2} N_j + (C_T \xi)_{(j)} \right)^2, \]

where

\( N_{q+1}, \ldots, N_p \) are independent univariate standard normal variables,

\( \lambda_{q+1} \geq \ldots \geq \lambda_p > 0 \) are the \( (p-q) \) positive eigenvalues of \( Q(M^{-1} - (M^*)^+) \) and

\( C \) is a Choleski root of \( M\) defined by

\[ CC^T = M_{22.1} \quad (4.3.6) \]

\[ C^T(M^{-1}Q M^{-1})_{22} C = A_{22} = \text{diag}(\lambda_{q+1}, \ldots, \lambda_p). \quad (4.3.7) \]
Proof.

We have

\[(p-q)w_n^2 = n^{1/2} (\bar{\Omega}_n - \bar{\Omega})^T M_{22.1} n^{1/2} (\bar{\Omega}_n - \bar{\Omega}) + 2n^{1/2} (\bar{\Omega}_n - \bar{\Omega})^T M_{22.1} \tilde{\Omega} + \tilde{\Omega}^T M_{22.1} \tilde{\Omega}, \tag{4.3.8}\]

where \( \tilde{\Omega} = n^{-1/2} \bar{\Omega} \).

Since \( n^{1/2} (\bar{\Omega}_n - \bar{\Omega}) \) has asymptotically a multivariate normal distribution with mean 0 and covariance matrix \( (M^{-1}Q M^{-1})_{22} \), this implies that we have to find a simultaneous diagonalization of the symmetric, positive definite matrices \( (M^{-1}Q M^{-1})_{22} \) and \( M_{22.1} \).

There exists a nonsingular \((p-q) \times (p-q)\) matrix \( \tilde{V} \) such that

\[ \tilde{V}^T ((M^{-1}Q M^{-1})_{22})^{-1} \tilde{V} = I_{22} \quad \text{and} \]
\[ \tilde{V}^T M_{22.1} \tilde{V} = \Lambda_{22}, \tag{4.3.9} \]

where \( \Lambda_{22} = \text{diag}(\lambda_{q+1}, \ldots, \lambda_p) \) and \( \lambda_{q+1} \geq \cdots \geq \lambda_p > 0 \) are the \((p-q)\) positive eigenvalues of \((M^{-1}Q M^{-1})_{22} \cdot M_{22.1} \).

Noting that

\[ M^{-1} \cdot M^{-1} = M^{-1} - (M^*)^T, \]

with \( Y_{11} = Y_{12} = Y_{21} = 0 \) , \( Y_{22} = M_{22.1} \),

we have

\{eigenvalues of \((M^{-1}Q M^{-1})_{22} \cdot M_{22.1}\) =
\{positive eigenvalues of \((M^{-1}Q M^{-1})_{22}\) =
\{positive eigenvalues of \((M^{-1}Q M^{-1})_{22} \cdot M_{22.1}\) =
\{positive eigenvalues of \((M^{-1}Q M^{-1})_{22} \cdot M_{22.1}\) =

\{positive eigenvalues of \((M^{-1}Q (M^{-1} - (M^*)^T))\) ,

and, defining \( C := (\Lambda^{-1/2} \tilde{V}^{-1})^T \), equations (4.3.6) and (4.3.7)
are satisfied.

Finally, from (4.3.8) we obtain the following asymptotic distribution of \((p-q)\hat{W}_n^2\)

\[
\sum_{j=q+1}^{p} \lambda_j N_{j}^2 + 2 \sum_{j=q+1}^{p} \lambda_j^{1/2} N_{j}(C(T)_{j}) + \hat{\sigma}_{T} C C \hat{\sigma}_{T},
\]

\[\square\]

**REMARKS**

1. In Subsection 8.1b we shall discuss the computational aspects of the distribution given by Proposition 4.6.

2. It was pointed out by a referee that a related result on the distribution of the likelihood ratio test statistic when the data do not come from the parametric model under consideration was obtained by Kent(1982).

3. The test defined by the test statistic \(\hat{W}_n^2\) is invariant with respect to linear transformations of the parameter space which leave invariant the hypothesis \(H_0\), that is \(\theta \mapsto D\theta\), where \(D\) is a \((p \times p)\) nonsingular matrix with \(D_{21} = 0\).

**COROLLARY 4.7**

Let be \(c > 0\). Then, under the given conditions, the following equivalence holds

\[
\sup\{ |IF(x, y; S, H_0)| : x, y \} = \\
\sup\{ |IF(x, y; W, H_0)| : x, y \} < c \\
\sup\{ ||IF(x, y; T, H_0)||_{M_{22, 1}} : x, y \} < c(p-q)^{1/2},
\]

where \(|| \cdot ||_{M_{22, 1}}\) denotes the norm in \(R^{p-q}\) generated by \(M_{22, 1}\).
The proof follows easily from the definitions of $S_n$, $W_n$, $(\Omega_n)_n$.

The expression

$$\| IF(x, y; T_\Omega, H_{\theta^*}) \|_{M_{22.1}}^2 = IF^T(x, y; \overline{T}_\Omega, H_{\theta^*}) M_{22.1} IF(x, y; \overline{T}_\Omega, H_{\theta^*})$$

$$IF^T(x, y; \overline{T}_\Omega, H_{\theta^*})*((M^{-1})_{22})^{-1} IF(x, y; \overline{T}_\Omega, H_{\theta^*})$$

defines a standardized sensitivity for the estimator $(\overline{T}_\Omega)_n$.
This should be compared with other standardizations for estimators, see Krasker and Welsch (1979), Stahel (1981) and Subsection 3.2b).
4.4 SPECIAL CASES

In this section we study the cases where the distribution of the test statistic under the sequence of alternatives (4.3.5) reduces to

\[ \lambda \cdot \chi^2_{p-q}(\delta^2), \]

where \( \chi^2_{p-q}(\delta^2) \) denotes the noncentral \( \chi^2 \)-distribution with \((p-q)\) degrees of freedom and noncentrality parameter \( \delta^2 \). In each case we shall give the standardization factor \( \lambda \) and the noncentrality parameter \( \delta^2 \). Note that the asymptotic power of the corresponding test is a monotone increasing function of \( \delta^2 \).

The conditions and the notation are the same as in Section 4.3.

4.4a \( \tau(x,r) = \rho(r) \)

In this case we have:

\[ \lambda_j = A/B = (p-q)^{-1} \text{tr}[Q \{M^{-1} - (M^*)^\perp\}] \]

\[ = \int (IF(x,y;S,H_{0*}))^2 dH_{0*}(x,y) =: \lambda, \quad j=q+1, \ldots, p, \]

\[ \delta^2 = (B^2/A) \Delta^\top (E\!x^\top)^{22.1} \Delta, \]

where

\[ Q = A \cdot E\!x^\top, \quad A = E\!\psi^2 \]

\[ M = B \cdot E\!x^\top, \quad B = E\!\psi', \quad \rho' = \psi. \]
This class of tests was proposed by Schrader and McKean (1977) and Schrader and Hettmansperger (1980) and carries out in a natural way M-estimation.

Let us look at the influence function of these tests. Using (iii) of Proposition 4.1, we have

\[
\begin{align*}
\text{IF}(x,y;S,H_{\theta*}) &= \\
&= \left( |\psi(y-x^T\theta*)| B^{1/2} \right) \left( x^T((Exx^T)^{-1} - (Ex^*x^*T)^+)^x / (p-q) \right)^{1/2}.
\end{align*}
\]

(4.4.1)

Using again the notions of influence of the residual and influence of position in factor space (Section 3.2a), we can rewrite (4.4.1) in the following way

\[
\text{IF}(x,y;S,H_{\theta*}) = \text{IT}(x,y;S,H_{\theta*}) = \text{IR}(r;S,H_{\theta*}) \text{IP}(x;S,H_{\theta*}),
\]

(4.4.2)

with

\[
\begin{align*}
\text{IR}(r;S,H_{\theta*}) &= |\psi(r)| / B^{1/2}, \\
\text{IP}(x;S,H_{\theta*}) &= \left( x^T((Exx^T)^{-1} - (Ex^*x^*T)^+)^x / (p-q) \right)^{1/2}.
\end{align*}
\]

REMARKS

1. One can standardize the test statistic S by dividing it by $B^{1/2}$. This operation does not change the test; on the other hand, IR becomes the absolute value of the influence function of a location M-estimator defined by $\psi$.

2. If we choose a bounded function $\psi$, the influence of the residual is bounded too, but the total influence IT is not because $\text{IP}(x;S,H_{\theta*}) \to \infty$, when $|x| \to \infty$. This justifies...
the consideration of the general class of tests \( \tau \).

3. Let \( SSI(x,r;T,H_{\theta*}) \) be the self-standardized influence of an M-estimator \( T \) (defined by the function \( \psi \)) at \( H_{\theta*} \), that is

\[
SSI(x,r;T,H_{\theta*}) = \left( IT^T(x,r;T,H_{\theta*}) \cdot V(T;H_{\theta*})^{-1} IT(x,r;T,H_{\theta*}) \right)^{1/2}, \tag{4.4.3}
\]

where \( V(T;H_{\theta*}) \) is the asymptotic covariance matrix of the estimator \( T \) at \( H_{\theta*} \) (see Subsection 3.2b).

Then, we have

\[
(IT(x,r;S,H_{\theta*}))^2 = \lambda(p-q)^{-1}[(SSI(x,r;T_\Omega,H_{\theta*}))^2 - (SSI(x,r;T_\omega,H_{\theta*}))^2].
\]

4.4b \( q = p-1 \)

In this case we obtain the following parameters

\[
\lambda = \lambda_p = \text{En}^2(x,r) \cdot (U^{-1}x)^{(p)} |^2,
\]

\[
\delta^2 = (u_{pp})^2/\lambda_p^\sigma^2 = (M_{22.1}/\lambda_p)^\sigma^2,
\]

where the matrix \( U \) is given by Proposition 4.3.

In this case we want to find a standardization of \( IF(x,y;S,H_{\theta*}) \) that extends in a natural way Definition 2.4 to the \( \tau \) test. Although it is not necessary to standardize the influence function of the test statistic in order to solve the optimality problem (2.2.9) for infinitesimal robustness of tests (see Subsection 2.2c), we shall show that in this
case such a standardization exists.

With the notation of Section 4.1 and 4.2 define

$$\xi_T(\theta) := S^2(H_\theta) = 2(p-q)^{-1}\int (\tau(x,y-x^T\theta^*) - \tau(x,y-x^T\theta)) h_\theta(x,y) \mu(x) dy,$$

where $h_\theta(x,y) = \phi(y-x^T\theta) k(x)$.

Assume that both estimators $T_\omega, T_\Omega$ are Fisher-consistent at the reduced and full model

$$T_\omega(H_{\theta^*}) = \theta^*, \quad T_\Omega(H_\theta) = \theta$$

and denote by a dot the differentiation with respect to $\theta$.

Then, after some straightforward calculations we obtain the vector

$$\dot{\xi}_T(\theta^*) = 0$$

and the matrix

$$\ddot{\xi}_T(\theta^*) = 2(p-q)^{-1}\int \eta(x,y-x^T\theta^*) h_\theta^T(x,y) \mu(x) dy$$

and, finally in the partition's notation

$$\dot{\xi}_T(\theta^*) = 0$$

$$\begin{align*}
(\ddot{\xi}_T(\theta^*))_{11} &= (\ddot{\xi}_T(\theta^*))_{12} = (\ddot{\xi}_T(\theta^*))_{21} = 0 \\
(\ddot{\xi}_T(\theta^*))_{22} &= M_{22,1} = 2(p-q)^{-1} U_{22} U_{22}^T.
\end{align*}$$

If $q=p-1$, $U_{22} = u_{pp}$, hence

$$\begin{align*}
(\ddot{\xi}_T(\theta^*))_{22} &= 2u_{pp}^2.
\end{align*}$$

Therefore, we propose in this case standardizing the influence function of the test statistic as follows
\[
IF_{test}(x, y; \tau, H_{0*}) := \frac{IF(x, y; S, H_{0*})}{\frac{1}{2} \xi_\tau (\theta^*)}^{1/2}_{pp} \frac{\xi_\tau (\theta^*)}{pp} = \frac{IF(x, y; S, H_{0*})}{u_{pp}}.
\]

(4.4.6)

This definition will become clear in Section 7.2. Note that (4.4.6) is consistent with Definition 2.4 in the sense that it has the following properties:

(i) \( IF_{test}(x, y; \tau, H_{0*}) = \lim_{\theta \to \theta^*} \frac{IF(x, y; S^2, H_{0})}{(\xi_\tau (\theta))^2} \),

so (4.4.6) is the natural extension of Definition 2.4.

(ii) \( IF_{test}(x, y; \tau, H_{0*}) \) is invariant to monotone differentiable transformations of the test statistic.

(iii) \( \delta^2 = (\frac{u_{pp}}{\bar{u}_p})^2 = \left( \int IF_{test}(x, y; \tau, H_{0*}) \right)^2 \frac{dH_{0*}}{(x, y)} \).

4.4c The density of \( x \) is "radially symmetric" with respect to \( \bar{x} \)

If

\[
\begin{align*}
L(x) &= \tilde{L}(\bar{x}, |\bar{x}|) \\
\tau(x, r) &= \tilde{\tau}(\bar{x}, |\bar{x}|, r)
\end{align*}
\]

(4.4.RAD)

then

\[
\lambda_j = \frac{(\lambda^{(Q)}/\lambda^{(M)})}{j = q+1, \ldots, p}
\]

\[
\delta^2 = \left( \frac{(\lambda^{(M)})^2}{\lambda^{(Q)}} \right) \frac{dH_{0*}}{\bar{x}^2},
\]

where

\[
\lambda^{(Q)} := (p-q)^{-1} E_{\eta} (x, r) |\bar{x}|^2,
\]

\[
\lambda^{(M)} := (p-q)^{-1} E_{\eta} (x, r) |\bar{x}|^2.
\]
In this case we have

\[ U_{22} = (\lambda^{(M)})^{1/2} \cdot I_{22}. \]

Thus, a natural standardization of the influence function of the test statistic is as follows

\[
\text{IF}_{\text{test}}(x, y; T, H_{\theta^*}) = \text{IF}(x, y; S, H_{\theta^*}) / (\lambda^{(M)})^{1/2}
\]

\[ (= \text{IF}(x, y; S, H_{\theta^*}) / u_{PP}). \]
5. OPTIMALLY BOUNDED INFLUENCE TESTS

In this chapter we solve the optimality problem (2.2.9) for infinitesimal robustness of tests (see Subsection 2.2c), that is we find a test (defined by a function $\tau$) which maximizes the asymptotic power, subject to a bound on the influence function of the test statistic at the null hypothesis. Using the results of Chapter 4, we shall give the solution to this problem for Huber's regression (Section 5.1), Mallows' regression (Section 5.2) and in the general case (Section 5.3). Throughout this chapter we assume the linear model of Section 3.1 with normal errors ($G=\Phi$) and the corresponding test problem of Section 3.3. We use the notation of Chapter 4.
5.1 THE $\rho_c$-TEST

PROPOSITION 5.1

Assume: $\tau(x,r) = \rho(r)$.

Then the test which maximizes the asymptotic power, under the side condition of a bounded influence of the residual

$$\sup\{IR(r; S, H_{0_1}): r \} \leq b_0^*,$$

is given by Huber's function $\rho_c(r)$.

Proof. (Ronchetti and Rousseeuw, 1980)

Since the asymptotic power is a monotone increasing function of the noncentrality parameter

$$\delta^2 = (B^2/A)\overline{\Delta}^T(Exx^T)_{22,1}\overline{\Delta}$$

$$= [(E\psi')^2/E\psi^2]\overline{\Delta}^T(Exx^T)_{22,1}\overline{\Delta}$$

(see Subsection 4.4a), the assertion follows easily using Hampel's Lemma 5 (see Hampel, 1968, 1974).

REMARKS

1. The $\rho_c$-test can be viewed as a likelihood ratio test when the error distribution has a density $\varrho_0$ proportional to $\exp(-\rho_c(r))$. This distribution minimizes the Fisher-information within the gross-error model $P_c(\theta)$ ("least favorable distribution").

Note that a test of the same type (a likelihood ratio test
under a least favorable distribution) is used by Carroll (1980, p.73) who proposes a robust method for testing transformations to achieve approximate normality.

2. The $\rho_c$-test is asymptotically equivalent to a test proposed by Bickel (1976, p.167) who applies the classical $F$-test to transformed observations (see Schrader and Hettmansperger, 1980).
5.2 THE OPTIMAL TEST FOR MALLOWS’ TYPE REGRESSION

In this section we derive the optimally robust test in the case of Mallows' type regression. Although this test is less efficient than the test we shall give in Section 5.3, there is a general interest in Mallows' regression since it provides good diagnostic information (Welsch, 1981).

PROPOSITION 5.2

Consider the class of tests defined by \( \tau(x,r) = w(x)\phi(r) \).

Assume either \( q = p-1 \) or (4.4.RAD) (see Section 4.4).

Then, the test which maximizes \( M_{22,1}^{1/\lambda} \) (respectively \( (\lambda(M^2)/\lambda(Q) \) under (4.4.RAD)) and therefore the asymptotic power, under the side condition of a bounded influence function

\[
\sup \left( \left| \frac{n(x,y-x^T\theta^*)}{\|u_p\|} \right| \cdot (x^T(M^{-1}-(M^*)^+)x)/(p-q)^{1/2} : x,y \in \delta_0 \right),
\]

is defined by the function

\[
\tau_M(x,r) = w_C(\frac{\|z\|}{z}(p-q)^{-1/2}) \cdot \rho_C(r),
\]

where

\[
w_C(t) := \psi_C(t)/t, \quad z = U^{-1}_M x,
\]

\( U_M \) is a lower triangular matrix defined implicitly by

\[
E w_C(\frac{\|z\|}{z}(p-q)^{-1/2})zz^T = I
\]

and \( c, c' \) are some positive constants.

Under (4.4.RAD) we have: \( \|z\| = \|z\|(p-q)^{1/2} \).
Proof.

We prove the assertion under the condition (4.4.RAD). The proof under the condition \((q = p-1)\) is similar to the proof for the general case that will be given in Proposition 5.4.

Define

\[
\psi(r) := (\partial \partial r) \rho(r)
\]

\[
A(\psi) := E\psi^2, \quad W_Q(w) := (p-q)^{-1} E w^2(x)|x|^{2}
\]

\[
B(\psi) := E\psi', \quad W_M(w) := (p-q)^{-1} E w(x)|x|^{2}.
\]

Under the condition (4.4.RAD) we have

\[
\tau(x,r) = w(x)\rho(r) = \widehat{w}(x,|x|)\rho(r),
\]

\[
\left( |\eta(x,r)/u_{pp}|^2 \right) \left( (x^T(M^{-1} - (M^*)^+)x) / (p-q) \right)^{1/2} =
\]

\[
\left( |\psi(r)/B(\psi)|^2 \right) \left( (p-q)^{-1/2} (w(x)/W_M(w)) |x| \right), \quad (5.2.1)
\]

\[
\left( \lambda^M \right)^2 / \lambda^Q = \left( (B(\psi))^2 / A(\psi) \right) \left( (W_M(w))^2 / W_Q(w) \right), \quad (5.2.2)
\]

where \(r := y - x^T \theta^*\).

Without loss of generality, assume \(\theta^* = 0\).

Then, it is obvious from (5.2.1) and (5.2.2) that the original extremal problem is separable, that is, the original problem is equivalent to the following.

Given \(b' > 0\), \(b'' > 0\), \(b'b'' \leq b_0\),

find \(\psi\) which minimizes \(A(\psi)/(B(\psi))^2\), subject to

\[\sup\{ |\psi(r)|/B(\psi) : r \leq b' \} \leq b'\]; \quad (5.2.3)

find \(w\) which minimizes \(W_Q(w)/(W_M(w))^2\), subject to

\[\sup\{ (p-q)^{-1/2}(w(x)/W_M(w)) |x| : x \leq b'' \} \leq b'' \]. \quad (5.2.4)
The problem (5.2.3) is solved by Proposition 5.1: the solution is $\psi_{c'}$, where the constant $c'$ depends on $b'$.

We have to solve (5.2.4).

Let $c > 0$ be such that $(b'')^{-1} = \alpha(c)/c$, where

$$\alpha(c) := \mathbb{E} W_c(|\bar{\mathcal{X}}|)|\bar{\mathcal{X}}|^2.$$  

$(c > 0$ exists if $b'' \geq (E|\bar{\mathcal{X}}|)^{-1}$, because $\alpha(c)/c$ is continuous on $]0,\infty[\,$, $0 < \alpha(c)/c < E|\bar{\mathcal{X}}|$, $\alpha(c)/c \to E|\bar{\mathcal{X}}|$ when $c \to 0$ and $\alpha(c)/c \to 0$, when $c \to \infty$.)

Since the multiplication of the test statistic by a positive constant does not change the test, we may assume without loss of generality

$$(p-q)^{1/2}W_M(w) = \mathbb{E} W_c(|\bar{\mathcal{X}}|)|\bar{\mathcal{X}}|^2 = \alpha(c) .$$  \hspace{1cm} (5.2.5)

Then (5.2.4) reduces to minimizing $\mathbb{E} \tilde{w}^2(\bar{x}, |\bar{\mathcal{X}}|)|\bar{\mathcal{X}}|^2$, subject to (5.2.5) and

$$\sup \{ \tilde{w}(\bar{x}, |\bar{\mathcal{X}}|)|\bar{\mathcal{X}}| : x \} \leq c .$$  \hspace{1cm} (5.2.6)

Now, by (5.2.5)

$$\mathbb{E} \tilde{w}^2(\bar{x}, |\bar{\mathcal{X}}|)|\bar{\mathcal{X}}|^2 = \mathbb{E} [ |\bar{\mathcal{X}}| - \tilde{w}(\bar{x}, |\bar{\mathcal{X}}|)|\bar{\mathcal{X}}| ]^2 - \mathbb{E} |\bar{\mathcal{X}}|^2$$

$$+ 2(p-q)^{1/2} \alpha(c) ,$$

and the problem reduces to minimizing $\mathbb{E} [ |\bar{\mathcal{X}}| - \tilde{w}(\bar{x}, |\bar{\mathcal{X}}|)|\bar{\mathcal{X}}| ]^2$,

under the side condition (5.2.6).

It is easy to check that

$$\tilde{w}(\bar{x}, |\bar{\mathcal{X}}|) = (\min(|\bar{\mathcal{X}}|, c))/|\bar{\mathcal{X}}| = w_c(|\bar{\mathcal{X}}|)$$

solves this minimizing problem.
5.3 THE OPTIMAL TEST FOR THE GENERAL M-REGRESSION

In order to state the optimality result, we first have to show the existence of a certain matrix.

PROPOSITION 5.3

Let be $c > 0$. If $E x^T$ is nonsingular, then

(i) for sufficiently large $c > 0$ there exists a symmetric and positive definite $(p \times p)$ matrix $M_S(c,p-q)$ which satisfies the equation

$$E((2\phi(c(p-q))^{1/2}/|x^T(M^{-1}-(M^*)^+x)^{1/2}) - 1)xx^T) = M ;$$

(5.3.1)

(ii) $M_S$ converges to $E x^T$, when $c \to \infty$;

(iii) Denote by $U_S$ the lower triangular matrix with positive diagonal elements such that $U_S U_S^T = M_S$ and define

$$\eta_S(x,r) := (|z|/(p-q)^{1/2})^{-\frac{1}{2}} \psi(c|z|/(p-q)^{1/2}) ,$$

with $z = U_S^{-1}x$.

Then, $M_S = E \eta_S'(x,r)xx^T$.

Proof.

Assertions (i) and (ii) can be shown using the same techniques as in Krasker(1981, Proposition 1, p.1338), noting that

$$\|M\| \leq \|M^*\| , \text{ where } \|M\| := \sum_{i,j} m_{ij}^2.$$
(iii) follows from Proposition 4.3 using (5.3.1). □

REMARK
The subscript $S$ for $M_S$ indicates that $M_S$ is the matrix $M$ corresponding to $\eta_S(x,r)$; this function is of the Schweppe form (see Subsection 3.2b).

PROPOSITION 5.4
Assume either $(q = p-1)$ or (4.4.RAD).
Then, the test that solves the given extremal problem (2.2.9) within the class $\{\tau\}$, is defined, by a function of the form

$$
\tau_S(x,r) = (|z|/(p-q)^{1/2})^{-2} \cdot \rho_c(r |z|/(p-q)^{1/2}) = \rho_\zeta(x)(r),
$$

where $\zeta(x) := c(p-q)^{1/2}/|z|$, $z = U^{-1}_S x$ and $U_S$ is the lower triangular matrix given by Proposition 5.3(iii) and defined implicitly by

$$
E(2\Phi(c(p-q)^{1/2}/|z|)-1)zz^T = I. \tag{5.3.2}
$$

Proof.
We show the assertion under the condition $(q = p-1)$.
The proof is similar under (4.4.RAD).

Put

$$
M(\eta) = E\eta'(x,r)xx^T, \quad Q(\eta) = E\eta^2(x,r)xx^T
$$

and let $\lambda_p(\eta)$ be the positive eigenvalue of

$$
Q(\eta) \cdot (M^{-1}(\eta) - (M^*)^+(\eta)).
$$
Moreover, denote by $\text{U}(n)$ the Choleski decomposition of $\text{M}(n)$ which is given by Proposition 4.3: $\text{U}(n)$ is a lower triangular matrix with positive diagonal elements.

We have to solve the following problem.

For a given $b > 0$, find a test which maximizes $\text{M}_{22.1}/\lambda_p$, under the side condition

$$\sup\{(|\eta(x,r)|/u_{pp})^2 : (x^T (\text{M}^{-1} - (\text{M}^*)^+) x)^{1/2} : x, r \leq b \}.$$  

(5.3.3)

Since $\text{U}^T (\text{M}^{-1} - (\text{M}^*)^+) \text{U} = \text{I}^{**}$ (Proposition 4.3),

we obtain

$$\lambda_p(n) = (\text{U}^{-1} \text{Q} \text{U}^{-T})_{pp} = \text{E} n^2 (x, r) |(\text{U}^{-1} x)^{(p)}|^2,$$

$$\text{M}_{22.1}(n) = (u_{pp}(n))^2.$$  

(5.3.4)  

(5.3.5)

Moreover, (5.3.3) becomes

$$\sup\{(|\eta(x,r)|/u_{pp})^2 : (\text{U}^{-1} x)^{(p)} : x, r \leq b \}.$$  

(5.3.6)

Choose $c > 0$ such that $b = c/(U_S)_{pp}$, where $U_S$ is defined by Proposition 5.3(iii), (this $c$ exists because

$$c^2/(U_S)_{pp}^2 \geq c^2/\text{tr} M_S \geq c^2/\text{E}|x|^2 \to \infty,$$

when $c \to \infty$) and assume, without loss of generality,

$$u_{pp}(n) = (U_S')_{pp}.$$  

(5.3.7)

(The multiplication of the test statistic by a positive constant does not change the test !)

Combining (5.3.4), (5.3.5), (5.3.6) and (5.3.7), the original problem reduces to the following:
minimize $E \| (U^{-1}x) \|_p^2$, under the conditions (5.3.7) and

$$\sup \{ \| \eta(x,r) \|_p \| (U^{-1}x) \|_p \mid : x,r \leq C \} \leq c. \quad (5.3.8)$$

Now,

$$E \| (U^{-1}x) \|_p^2 = -E r^2 \| (U^{-1}x) \|_p^2 + 2$$

$$+ E(\eta(x,r) \cdot (U^{-1}x) \cdot r(U_S^{-1}x) \cdot (p) \cdot r(U_S^{-1}x) \cdot (p))^2,$$

since

$$E(\eta(x,r) \cdot r(U^{-1}x) \cdot (p) \cdot (U_S^{-1}x) \cdot (p))$$

$$= (U^{-1}E(\eta(x,r) \cdot r \cdot xx^T) \cdot U_S^{-1})_{pp},$$

and integrating by parts,

$$= (U^{-1}E(\eta'(x,r) \cdot xx^T) \cdot U_S^{-1})_{pp} = (U^{-1}M \cdot U_S^{-1})_{pp}$$

$$= (U^{-1}U \cdot U^{-1}U_S^{-1})_{pp} = (U^T \cdot U_S^{-1})_{pp} = \psi_{pp}/(U_S) = 1,$$

where in the last equalities we have used (5.3.7) and $U_{12} = 0$. Thus, minimizing $E \| (U^{-1}x) \|_p^2$, under the conditions (5.3.7) and (5.3.8) is equivalent to minimizing

$$E(\eta(x,r) \cdot (U^{-1}x) \cdot (p) - r \cdot (U_S^{-1}x) \cdot (p))^2,$$

subject to (5.3.8).

Clearly, the optimal $\tilde{\eta}$ must satisfy

$$\tilde{\eta}(x,r) \cdot (U^{-1}(\tilde{\eta})x) \cdot (p) = \psi_{c}(r(U_S^{-1}x) \cdot (p)).$$

Therefore,
\[ \eta_S(x,r) = |\tilde{z}|^{-1} \psi_c(r|\tilde{z}|), \]

where \( z = u_S^{-1}x \), solves this extremal problem.

Any other solution defines a test which has the same influence function and the same asymptotic power and in this sense is equivalent to \( \eta_S \).

**Remarks**

1. Interpretation of the optimality results from the estimation point of view

We know that the \((p-q)\) positive eigenvalues of 

\[ Q(M^{-1} - (M^*)^+) \]

are the eigenvalues of \((M^{-1}Q M^{-1})_{22} \cdot M_{22}.1\)

(see the proof of Proposition 4.6). Therefore, under the conditions of Section 4.4, the asymptotic power is a monotone decreasing function of \( \text{tr}(M^{-1}Q M^{-1})_{22} \) which equals the trace of the asymptotic covariance matrix of the estimator \( \hat{\Theta}_\tilde{z} \).

On the other hand, Corollary 4.7 establishes the equivalence between the gross-error-sensitivity of the test statistic \( S \) (and \( W \)) and that of the estimator \( \hat{\Theta}_\tilde{z} \).

It follows that, under the conditions of Section 4.4, the following problems are equivalent:

**Test problem**

maximize the asymptotic power of the test, subject to 

\[ \sup \{ |IF(x,y;S,H_{0*}|) : x,y \leq c \}. \]
Estimation problem

minimize the trace of the asymptotic covariance matrix of the estimator $\hat{\sigma}^2_{\Omega}$, under the condition

$$\sup\{|| IF(x,y;T_{\Omega},H_0^*) ||_{M_{2,1}} : x,y \} \leq c(p-q)^{1/2}.$$

2. Finite sample approximations to the optimal weights

The $\eta$-functions defining optimally bounded influence tests and optimally bounded influence estimators are of the same form (see Proposition 5.2 and 5.4 and 3.2.14); there is a difference only in the weights. The optimal weights for the test take into account that (after standardization) only the last $(p-q)$ components are of interest for the testing problem.

Welsch(1981) provides finite sample approximations to the optimal weights for estimators. He takes $n-1$ observations $((x_i,y_i)$ omitted), computes an approximate influence function for $(x_i,y_i)$ and in this way obtains the matrix

$$(n-1)^{-1}x^TX(i)x(i)^T$$

that can be viewed as an approximation to the matrix $M$. ($X(i)$ denotes the design matrix $X$ without the $i$th row $x_i^T$.) Using this approximation to the matrix $M$ we can obtain finite sample approximations to the optimal weights for the tests.
Mallows' regression: \( w_c(v_i^{-1}) \cdot \rho_c(x_i) \)

General case: \( \rho_c \cdot v_i(x_i) \),

where

\( v_i := ((n-1)(p-q))^{-1}(h(\Omega) / (1-h(\Omega)) - h(\omega) / (1-h(\omega)))^{-1/2} \)

and \( h_{ii}^{(*)} \) is the element \((i,i)\) of the hat matrix \( X(X^TX)^{-1}X^T \) of the corresponding model.

3. Note that from these optimally robust tests one can easily derive robust confidence regions for the parameters and a robust version of stepwise regression.

These procedures can be easily implemented in a package for robust regression. Especially, we plan to integrate them in ROBETH, a package of robust linear regression programs, which has been written at the ETH Zürich and is still under development (see Marazzi,1980).
6. A ROBUST SELECTION PROCEDURE

6.1 THE AKAIKE INFORMATION CRITERION AND THE C_p STATISTIC

6.1a Generalities

Let \( z_1, \ldots, z_n \) be \( n \) observations and suppose we want to fit
a model with \( p \) parameters. Some examples are the linear model
in the regression analysis, the AR and ARMA models in time
series. A natural question which arises is the following:
which parameters are really needed in the model?
Akaike (1973) proposed the following Information Criterion
for the selection of the most important factors.

Choose the model such that

\[
AIC(p) := -2 \cdot \log(f_{T(ML)}^{n,p}(z_1, \ldots, z_n)) + 2p = \min_{n; p} \tag{6.1.1}
\]

where \( f_0(z_1, \ldots, z_n) \) is the joint density of \( z_1, \ldots, z_n \) and
\( T(ML) \) is the maximum likelihood estimate of the vector \( \theta \) in
a model with \( p \) parameters.

First, we remark that the Akaike Criterion has a clear inter-
pretation: the first term of \( AIC(p) \) is a measure for the
quality of the fit and the second one indicates the increased unreliability due to the increased number of parameters. The Akaike Criterion looks for a compromise between these two components.

Secondly, this procedure may be viewed as an extension of the maximum likelihood principle and it is based on a general information theoretic criterion. In fact \(-\text{AIC}(p)\) is a suitable estimate of the expected entropy of the model: by the Akaike procedure the entropy will, at least approximately, be maximized (see Akaike, 1973).

Bhansali and Downham (1977) proposed to generalize the Akaike Criterion as follows:

For a given \(\alpha, \ 0 < \alpha < 4\), define

\[
\text{AIC}_\alpha(p) := -2 \cdot \log(f_{\text{ML}}^{(z_1, \ldots, z_n)}_{T_n;p}) + \alpha p
\]  

(6.1.2)

and choose the model that minimizes \(\text{AIC}_\alpha(p)\) with respect to \(p\).
6.1b Relation between the generalized Akaike statistic and
the likelihood ratio test in linear models

Let us apply the generalized Akaike Criterion to the linear model of Section 3.1. Let \( \hat{\sigma}_0 \) be a "good estimate" of \( \sigma \) (for example, take the estimate of \( \sigma \) in the model with all parameters) and consider the \( x_i \)'s as given.

Then we have

\[
\log(f_{T^{(ML)}}(z_1, \ldots, z_n)) = \\
\log(\prod_{i=1}^{n} \frac{1}{\hat{\sigma}_0^2} g((y_i - x_i^T \hat{T}^{(ML)} n; p) / \hat{\sigma}_0^2)) k(x_i)) = \\
\sum_{i=1}^{n} \log(\hat{\sigma}_0^{-1} k(x_i)) + \sum_{i=1}^{n} \log g((y_i - x_i^T \hat{T}^{(ML)} n; p) / \hat{\sigma}_0^2),
\]

(6.1.3)

and using (6.1.2)

\[
AIC_\alpha(p) = -2\sum_{i=1}^{n} \log(\hat{\sigma}_0^{-1} k(x_i)) - 2\sum_{i=1}^{n} \log g((y_i - x_i^T \hat{T}^{(ML)} n; p) / \hat{\sigma}_0^2) \\
+ \alpha \cdot p.
\]

(6.1.4)

In the classical case (G=\( \Phi \)) we obtain

\[
AIC_\alpha(p) = C(n, \hat{\sigma}_0) + |y - x \cdot \hat{T}^{(ML)} n; p|^2 / \hat{\sigma}_0^2 + \alpha \cdot p,
\]

where \( X \) is the design matrix, \( y=(y_1, \ldots, y_n)^T \) and

\[
C(n, \hat{\sigma}_0) := -2\sum_{i=1}^{n} \log(\hat{\sigma}_0^{-1} k(x_i)) .
\]

For the regression model Mallows(1964,1973) introduced the \( C_p \) procedure based on the minimization of the \( C_p \) statistic

\[
C_p = |y - x \cdot \hat{T}^{(ML)} n; p|^2 / \hat{\sigma}_0^2 + 2p - n .
\]
In this case the procedures based on AIC and \( C_p \) are equivalent. Now we want to investigate the relation between the generalized Akaike statistic and the likelihood ratio test.

Let

\[
H_0 : \theta^{(q+1)} = \ldots = \theta^{(p)} = 0
\]

be the null hypothesis in the given linear model and let \( \Lambda \) be the likelihood ratio test statistic. Define

\[
L_{q,p} := -2(p-q)^{-1} \log \Lambda.
\]

From (6.1.4) we get

\[
L_{q,p} = -2(p-q)^{-1} \sum_{i=1}^{n} \log g((y_i - x_i^T T_{niq}) / \hat{\sigma}_0) - \log g((y_i - x_i^T T_{nip}) / \hat{\sigma}_0)
\]

\[
= (p-q)^{-1} ((\text{AIC}_\alpha (q) - q) - (\text{AIC}_\alpha (p) - p)).
\]

Therefore, the generalized Akaike statistic and the likelihood ratio test statistic are related by the formula

\[
L_{q,p} = \alpha - (p-q)^{-1} (\text{AIC}_\alpha (p) - \text{AIC}_\alpha (q)). \quad (6.1.5)
\]

Moreover, putting \( \alpha = 2 \)

\[
L_{q,p} = 2 - (p-q)^{-1} (C_p - C_q). \quad (6.1.6)
\]
6.2 A ROBUST PROCEDURE

In this section we propose a robust selection procedure for linear models in the case that the \( x_i \)'s are given.

6.2a Definition

The equation (6.1.5) relates the generalized Akaike procedure and the test in a very natural way. Moreover, the log-likelihood ratio statistic plays a central role in the construction of the Akaike Information Criterion. Therefore, a reasonable robust selection procedure must be related to the robust test through a similar equation.

We propose the following robust selection procedure.

Choose the model such that

\[
AIC^{(rob)}_{C\alpha}(p) = 2\sum_{i=1}^{n} \rho \left( \frac{y_i - x_i^T T_n}{\hat{\sigma}_0} \right) + \alpha p = \min \tag{6.2.1}
\]

or equivalently

\[
C^{(rob)}_{p;C\alpha} = 2\sum_{i=1}^{n} \rho \left( \frac{y_i - x_i^T T_n}{\hat{\sigma}_0} \right) + \alpha p - n = \min \tag{6.2.2}
\]

where \( T_n \) is the Huber estimate in the model with \( p \) parameters, \( \hat{\sigma}_0 \) is a "good estimate" of \( \sigma \) (for example, use the Proposal 2 estimate of Huber in the full model) and \( c \) is a suitable positive constant (for example \( c=1.5 \)).

The choice of \( \alpha \) is discussed in the next Subsection 6.2b.
6.2b Properties of $\text{AIC}^{\text{rob}}_\alpha (p)$ and $C^{\text{rob}}_{p;\alpha}$

**PROPERTY 1**

$\text{AIC}^{\text{rob}}_\alpha (p)$ is essentially the generalized Akaike statistic defined by (6.1.4) under the least favorable distribution with density

$$g_0(r) = (1-e)(2\pi)^{-1/2}\exp(-\rho_0(r))$$  \hspace{1cm} (6.2.3)

(see Section 1.2 and 5.1).

**PROPERTY 2**

The likelihood ratio test under $g_0$ is a test of the class $\{\tau\}$ (Section 4.1), with $\tau(x;r) = \rho_c(r)$. Let $S^2_{n;c,q,p}$ be the test statistic defining this test. Then

$$nS^2_{n;c,q,p} = \alpha - (p-q)^{-1}(\text{AIC}^{\text{rob}}_\alpha (p) - \text{AIC}^{\text{rob}}_\alpha (q))$$  \hspace{1cm} (6.2.4)

and equation (6.1.5) holds for the robust procedures.

**Choice of $\alpha$**

In order to propose a choice for $\alpha$ in (6.2.1), we need the following general asymptotic equivalence.

Stone (1977) shows that the Akaike statistic is asymptotically equivalent to

$$-2 \log(f_{\text{ML}}(z_1, \ldots, z_n)) + 2 \cdot \text{tr}(I^{-1}_2 L_1),$$  \hspace{1cm} (6.2.5)

where $-I_2$ is the matrix of the second derivatives (with
respect to the parameter $\theta$) of the log-likelihood function and $L_1$ is the matrix of the products of the first derivatives.

Now, from Property 1 we know that $AIC^{(rob)}_a(p)$ can be viewed as the Akaike statistic with respect to the least favorable distribution $g_0$ of the errors. Therefore, in our case the log-likelihood function equals

$$\Sigma_{i=1}^{n} \log (\sigma_0^{-1} k(x_i)) + \Sigma_{i=1}^{n} \log g_0((y_i - x_i^T \hat{T}_{n+p}^{(ML)})/\sigma_0)$$

and

$$L_1 = A \cdot E_k xx^T, \quad L_2 = B \cdot E_k xx^T,$$

where $A = E\psi^2_C$, $B = E\psi'_C$, $\psi_C(r) = (\partial/\partial r)\rho_C(r)$, hence

$$2 \text{tr}(L_2^{-1} L_1) = 2(A/B)p. \quad (6.2.6)$$

In view of (6.2.5) and (6.2.6) we propose the following choice of $a$ in (6.2.1)

$$a = a_c = 2 \cdot A/B = 2 \cdot E\psi^2_C/E\psi'_C.$$

Note that $a_{\infty} = 2$ and $AIC^{(rob)}_{a_{\infty}}(p) = AIC(p)$.

**REMARK**

An alternative robust procedure for the selection of the variables can be obtained replacing the F-test in the classical stepwise regression by the optimally robust test derived in Chapter 5. This procedure can be viewed as a robust version of stepwise regression.
Consider the linear model of Section 3.1 with normal errors $(G=\Phi), \sigma=1$, and the related test problem (Section 3.3). Put $q = p-1$.

We define a new class of tests $(C(a)-type tests)$ and we compute their influence function and their asymptotic power.

As far as these properties are concerned, these tests are equivalent to the tests which have been introduced in Chapter 4.

Therefore, the test which solves Hampel's problem in this class is equivalent to the optimally robust test we have derived in Proposition 5.4.

Moreover, we shall show an interesting connection between our optimal test and an asymptotically minimax test proposed by Wang(1981) in a general framework.
7.1 DEFINITION OF A C(α)-TYPE TEST

The class of tests we want to define will be based on a function \( \eta \) which satisfies the same conditions as in Chapter 4. Let us recall the definition.

The function

\[
\eta : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}, \ (x,r) \mapsto \eta(x,r)
\]

satisfies the following conditions:

(7.1.ETA1) (i) \( \forall x \in \mathbb{R}^p, \ \eta(x,*) \) is odd,

(ii) \( \forall x \in \mathbb{R}^p, \ r \in \mathbb{R}^+, \ \eta(x,r) \geq 0 \).

(7.1.ETA2) \( \forall x \in \mathbb{R}^p, \ \eta(x,*) \) is differentiable on \( \mathbb{R} \setminus D(x;\eta) \),

where \( D(x;\eta) \) is a finite set.

Let \( \eta'(x,r) := (\exists \gamma \in \mathbb{R}) \eta(x,r) \) if \( x \in \mathbb{R}^p, \ r \in \mathbb{R} \setminus D(x;\eta) \)

\[ := 0 \quad \text{otherwise}. \]

Assume:

(i) \( \forall x \in \mathbb{R}^p, \ \sup\{|\eta'(x,r)| : r \in \mathbb{R}\} < \infty \),

(ii) \( M := E\eta'(x,r)xx^T \) is nonsingular.

**DEFINITION 7.1**

The class of C(α)-type tests for linear models is defined by means of the test statistics

\[
Z_n(\theta^*;\eta) := n^{-1/2} \sum_{i=1}^{n} \eta(x_i, y_i - x_i^T \theta^*) (U^{-1} x_i)^{(p)}
\]

\[
= n^{-1/2} u_{pp}^{-1} \sum_{i=1}^{n} \eta(x_i, y_i - x_i^T \theta^*) [x_i^{(p)} - U_{21} U_{11}^{-1} \bar{x}_i],
\]

(7.1.1)
where \( n \) satisfies the conditions (7.1.ETA1), (7.1.ETA2) and \( U \) is the lower triangular matrix with positive diagonal elements such that \( UU^T = M \) (see Proposition 4.3).
"Large" values of \( Z_n \) are significant.

REMARKS

1. \( C(\alpha) \) tests were introduced by Neyman(1958) in a general framework and extended to the robust testing problem by Wang(1981, Remarks 2 and 3, p.1100) who was able to derive an (asymptotically minimax) robust version of the optimal \( C(\alpha) \) test. We shall see in Section 7.3 that our optimally bounded influence test has a strong connection with Wang's asymptotically minimax test.

2. If we put in (7.1.1) \( \eta(x,r) = -g'(r)/g(r) \), where \( g \) is the density of the error distribution, \( Z_n \) becomes the test statistic of the optimal \( C(\alpha) \) test obtained by Neyman(1958, p.228). In this sense the tests defined by (7.1.1) can be called \( C(\alpha) \)-type tests.

Moreover, they can be viewed as an extension of the M-tests to linear models (cf. Subsection 2.2d).

3. The test statistic \( Z_n \) depends on the unknown nuisance parameters \( \theta^{(1)}, \ldots, \theta^{(p-1)} \). In Section 7.2 we shall discuss the properties of a studentized version \( Z_n((T_{\omega})_n;\eta) \), \( (T_{\omega})_n \) being a suitable estimate of \( \theta^* \).
4. In order to determine asymptotic critical regions of the test, we shall compute in Section 7.2 the asymptotic distribution of $Z_n$. 
7.2 THE INFLUENCE FUNCTION AND THE ASYMPTOTIC POWER OF C(α)-TYPE TESTS

PROPOSITION 7.1

Let $Z_n$ be the test statistic defining a C(α)-type test and let $H_{θ*}$ be the model distribution under the null hypothesis.

If (7.1.ETA1) and (7.1.ETA2) hold, then

(i) the influence function of $Z_n$ (standardized as functional $Z$) is given by

$$IF(x,y;Z,H_{θ*}) = η(x,y-θ^T)(U^{-1}x)(p) ;$$

(ii) $|IF(x,y;Z,H_{θ*})| = IF(x,y;S,H_{θ*}) = IF(x,y;W,H_{θ*})$,

where $S$ and $W$ are the test statistics (functionals) defining the tests introduced in Chapter 4 (see Proposition 4.5 and Corollary 4.7).

Proof.

Assertion (i) follows applying the definition of influence function and noting that by (7.1.ETA1)(i) $Eη(x,r)x = 0$.

Assertion (ii) follows from (i) noting that

$$|(U^{-1}x)(p)|^2 = x^T(M^{-1} - (M^*)^+)^x .$$

Let us now compute the asymptotic power of a C(α)-type test defined by $η$. 
PROPOSITION 7.2

Besides (7.1.ETA1) and (7.1.ETA2) assume 

\[(7.2.AS3) \, \mathbb{E} \eta^2(x, r) \cdot |x|^2 < \infty .\]

Put \( \lambda_p := \mathbb{E} \eta^2(x, r) \cdot (U^{-1}x)(p)^2 \).

Then, under the sequence of alternatives 

\[ H(n) : \theta = \theta^* + n^{-1/2} \Delta \]

(7.2.1)

where \( \Delta = (0, \ldots, 0, \Delta(p)) \),

the test statistic \( Z_n \) has asymptotically a normal distribution with mean \( \mu_{\theta^*}(p) \) and variance \( \lambda_p \).

Moreover, the asymptotic power of the \( C(\alpha) \)-type test defined by \( Z_n \) is given by 

\[ 1 - \Phi(\Phi^{-1}(1-\alpha) - \Delta(p) \mu_{\theta^*}/\lambda_p^{1/2} ). \]

Proof.

It suffice to compute \( \mathbb{E}_\theta Z_n \) under the sequence (7.2.1).

Define:

\[ \xi(\theta) := \int \eta(x, y - x^T \theta^*) \cdot (U^{-1}x)(p) \cdot h_\theta(x, y) d\mu(x) dy, \]

where \( h_\theta(x, y) = \phi(y - x^T \theta)k(x) \).

Then, for \( j=1, \ldots, p-1 \), we have

\[ (\partial \xi(\theta) \partial \theta^*) \xi(\theta) = \int \eta' \cdot (x, y - x^T \theta^*) \cdot x^T \cdot (U^{-1}x)(p) \cdot dH_{\theta^*}(x, y) \]

\[ + \int \eta(x, y - x^T \theta^*) \cdot (U^{-1}x)(p) \cdot (\partial \xi(\theta) \partial \theta^*) h_\theta(x, y) \bigg|_{\theta^*} d\mu(x) dy \]

(7.2.2)

Using
\[(U^{-1}x)(p) = u_{pp}^{-1}(-U_{21} U_{11}^{-1} x + x(p)) , \quad (7.2.3)\]

we get

\[En'(x,r)\overline{x}^T(U^{-1}x)(p) = \]
\[u_{pp}^{-1}(-U_{21} U_{11}^{-1} En'(x,r)\overline{x}^T + En'(x,r)\overline{x}^T x(p)) = \]
\[u_{pp}^{-1}(-U_{21} U_{11}^{-1} U_{11} U_{11}^T + M_{12}) = 0 , \quad (7.2.4)\]

and the first term of the right hand side of (7.2.2) vanishes.

In a similar way one can prove that also the second term of the right hand side of (7.2.2) equals 0.

Moreover,

\[\frac{\partial \xi(\theta)(p)}{\partial \theta} (\theta) \bigg|_{\theta^*} = \]
\[\int_{\gamma(x,y-x^T\theta^*)}(U^{-1}x)(p), (\partial \xi(\theta)(p))_{\gamma}(x,y) \bigg|_{\theta^*} \partial \mu(x) dy = \]
\[En(x,r)(U^{-1}x)(p),r \cdot x(p) = \]

and, by integrating by parts,

\[En'(x,r)(U^{-1}x)(p),x(p) = \]
\[u_{pp}^{-1}(-U_{21} U_{11}^{-1} M_{12} + m_{pp}) = u_{pp}^{-1} (u_{pp}^2) = u_{pp} . \quad (7.2.5)\]

Finally, denoting by a dot the differentiation with respect to \( \theta \), we obtain

\[(\xi(\theta^*)(j))(p) = 0 , \quad \text{for } j=1,\ldots,p-1 \quad (7.2.6)\]

\[(\xi(\theta^*)(p)) = u_{pp} \quad (7.2.7)\]

and by (7.1.ETAL)(i)

\[\xi(\theta^*) = 0 . \]
Therefore, using a Taylor expansion we get
\[ E_n \mathcal{Z}_n = n^{1/2} \xi(\bar{\theta}) = n^{1/2} \xi^{T}(\bar{\theta}) \ast (\bar{\theta} - \bar{\theta}^*) + o(|\bar{\theta} - \bar{\theta}^*|) \]
\[ = u_{pp}^\Delta(p) + o(n^{-1/2}). \]

This completes the proof. \(\square\)

**REMARKS**

1. As in Subsection 4.4b we want to give a suitable standardization of the influence function. Using (7.2.6) and (7.2.7) we can standardize the influence function in the following way
\[ IF_{test}(x,y;\eta C(\alpha),H_{\theta}^\ast) := IF(x,y;Z,H_{\theta}^\ast)/(\xi(\bar{\theta}^*))^{(p)} \]
\[ = IF(x,y;Z,H_{\theta}^\ast)/u_{pp}. \]  

(7.2.8)

Note that \( \xi(\bar{\theta}) = E_n \mathcal{Z}_n^{-1/2} \mathcal{Z}_n = \mathcal{Z}(H_{\theta}). \) Thus, (7.2.8) extends the one dimensional definition in a natural way.

Again from Proposition 7.2 it follows that the asymptotic power of the \( C(\alpha) \)-type test defined by \( \eta \) is a monotone increasing function of \( (IF_{test}(x,y;\eta C(\alpha),H_{\theta}^\ast))^2 d_{H_{\theta}^\ast}(x,y))^{-1} \).

Formula (7.2.8) should be compared with (4.4.6).

2. The test defined by the test statistic \( \mathcal{Z}_n \) depends on the unknown nuisance parameter \( \theta^* \). By techniques similar to those used by Wang(1981) one can show that the result of Proposition 7.2 holds if we substitute \( \theta^* \) by a suitable \( (n^{1/2}\text{-consistent}) \) estimate \( (T_\omega)_n \) (see Wang,1981, p.1099).
Moreover, the result of Proposition 7.1 still holds. Namely, suppose that the influence function of $T_\omega$ exists and define

$$Z(H) = \int \eta(x,y - x^T \cdot T_\omega(x))(U^{-1}x)(p) \, dH(x,y).$$

Then,

$$\text{IF}(x,y; Z, H_{\theta^*}) = (\exists \delta \in \mathbb{C}) \int ((1-\varepsilon)H_{\theta^*} + \varepsilon \delta(x,y)) \bigg|_{\varepsilon=0}$$

$$= \eta(x,y - x^T \cdot T_\omega(x))(U^{-1}x)(p) + \int \eta(s,v - s^T \cdot T_\omega(x))(U^{-1}s)(p) \, dH_{\theta^*}(s,v)$$

$$- \int \eta'(s,v - s^T \cdot T_\omega(x))(U^{-1}s)(p) s^T dH_{\theta^*}(s,v) \, \text{IF}(x,y; T_\omega, H_{\theta^*}).$$

(7.2.9)

Now, the second term of the right hand member of (7.2.9) is equal 0 in view of (7.2.3) and (7.1.ETA1)(i), and so is the third term by (7.2.4).
7.3 OPTIMALLY ROBUST C(α)-TYPE TESTS

7.3a The optimal C(α)-type test

From the point of view of the asymptotic properties (influence function and asymptotic efficiency) C(α)-type tests defined by η, τ-tests and tests based on the quadratic form \( W_n^2 \) are equivalent (see Sections 4.3, 4.4b and Section 7.2). The optimally robust test can be found in each class minimizing \( \lambda_p/u_{pp}^2 \), subject to a bound on the influence function. Applying Proposition 5.4 we have the following result.

**COROLLARY 7.3**

The test which solves the optimality problem (2.2.9) within the class of C(α)-type tests is defined by a function of the form:

\[
\eta_S(x,r) = \left| z^{(p)} \right|^{-1} \psi_c(r | z^{(p)} |) = \psi_{c/|z^{(p)}|}(r),
\]

where \( z = U_S^{-1} x \), \( U_S \) is the lower triangular matrix defined implicitly by

\[
E(2\phi(c/|z^{(p)}|)-1)zz^T = I
\]

and \( c \) is a given positive constant.
7.3b Connection with an asymptotically minimax test

proposed by Wang

In this subsection we want to investigate the connections between the optimally robust C(α)-type test defined by ηS and a test introduced by Wang(1981).

Wang studies the testing problem using minimax techniques. He considers the following situation.

Suppose we are given a parametric model (Ω, A, {F_θ : θ ∈ Ω ⊂ R^p}) (see Subsection 2.1a) and suppose we want to test an hypothesis on one component of the parameter θ, say θ(p) = 0, the other (p-1) components being nuisance parameters. Then, using the technique of shrinking neighbourhoods, Wang is able to derive an asymptotically minimax test and to extend the work of Rieder(1978) to the case where the model distribution is indexed by nuisance parameters.

Let me write this result in the situation of the linear model.

Define:

\[ r := y - x^T \theta^* \quad , \quad a^T := (a^{(1)}, \ldots, a^{(p)}) , \]

\[ d(x, \bar{a}) := x^{(p)} + \bar{a}^T x = x^{(p)} + \sum_{j=1}^{p-1} a^{(j)} x^{(j)} , \]

\[ \Lambda(\bar{a}, \theta^*) = \tilde{\Lambda}(x, y; \bar{a}, \theta^*) := r \cdot d(x, \bar{a}) . \quad (7.3.1) \]

For given \( \varepsilon > 0 , \delta_1 > 0 \), define \( V_0(\bar{a}, \theta^*) \) and \( V_1(\bar{a}, \theta^*) \) implicitly by means of the equations

\[ E[\max\{(\Lambda(\bar{a}, \theta^*) - V_1(\bar{a}, \Lambda^*)) , 0\}] = \varepsilon / \delta_1 , \quad (7.3.2) \]
\[ E\left[ \max\{ (V_0(\overline{\alpha},\theta^*) - \Lambda(\overline{\alpha},\theta^*)), 0 \} + \Lambda(\overline{\alpha},\theta^*) \right] = \epsilon/\delta_1. \quad (7.3.3) \]

Moreover, let \( \overline{\alpha} = \overline{\alpha}(\theta^*) \) be the solution to the equation
\[ E\left( \frac{\Lambda(\overline{\alpha},\theta^*)}{V_0(\overline{\alpha},\theta^*)} \cdot \mathbf{x} \cdot \overline{\mathbf{x}} \right) = 0, \quad (7.3.4) \]

and define
\[ \nu^2(\theta^*) := E\left( \frac{\Lambda(\overline{\alpha},\theta^*)}{V_0(\overline{\alpha},\theta^*)} \right)^2. \quad (7.3.5) \]

**PROPOSITION 7.4** (Wang, 1981, p.1099, p.1104)

The test defined by the critical region
\[ \{ Y_n(\theta^*) \geq \Phi^{-1}(1-\alpha) + \epsilon V_1(\overline{\alpha}(\theta^*),\theta^*)/\nu(\theta^*) \} , \quad (7.3.6) \]

where
\[ Y_n(\theta^*) := n^{-1/2}(\nu(\theta^*))^{-1} \sum_{i=1}^{n} \left[ \Lambda(x_i, y_i; \overline{\alpha}(\theta^*), \theta^*) \right] V_1(\overline{\alpha}, \theta^*) , \]

is an asymptotically minimax test at level \( \alpha \).

The test defined by (7.3.6) will be called Wang test.

In order to have a better performance at the model, Wang (1981, p.1105) proposes a modification of the test (7.3.6) and defines a test by means of the following critical region
\[ \{ Y_n(\theta^*) \geq \Phi^{-1}(1-\alpha) \} . \quad (7.3.7) \]

The test defined by (7.3.7) will be called modified Wang test.

Then we have the following result.
PROPOSITION 7.5

The modified Wang test is a C(α)-type test. It is equivalent to the optimally robust C(α)-type test defined by \( \eta_S \) (given by Corollary 7.3), that is, both tests have the same influence function and the same asymptotic power.

Proof.

We prove that \( Y_n(\theta^*) = \lambda_p^{-1/2}(\eta_S) \cdot Z_n(\theta^*; \eta_S) \).

Then the assertion follows using Proposition 7.2 and Remark 1. of Section 7.2.

First, consider equation (7.3.4). Using (7.3.1) we have

\[
0 = E(A)_{V_1, \tilde{\alpha}} = E(A_{\tilde{\alpha}} V_1, \tilde{\alpha})
\]

\[
= \int d(x, \tilde{\alpha}) \cdot \tilde{\alpha} \cdot \left[ \int \frac{V_1}{|d(x, \tilde{\alpha})|} \cdot r \cdot d\phi(r) \right] dK(x),
\]

and (7.3.4) becomes

\[
\int \left[ \Phi(V_1/|d(x, \tilde{\alpha})|) - \Phi(V_0/|d(x, \tilde{\alpha})|) \right] \cdot d(x, \tilde{\alpha}) \cdot \tilde{\alpha} dK(x) = 0.
\]

(7.3.8)

Now, combining (7.3.2) and (7.3.3) and noting that \( E[A] = 0 \), we obtain

\[
\int (A - V_1) \cdot 1_{\{A \geq V_1\}} + \int (A - V_0) \cdot 1_{\{A \leq V_0\}} = 0,
\]

and performing these integrations we get finally by (7.3.8)

\[
V_0 = -V_1 \quad (=: -V).
\]

(7.3.9)

For a given positive constant \( c \), define \( \eta_S \) and \( U_S \) as in Corollary 7.3. Moreover, choose
\[ V = c \cdot (U_S)_{pp} \cdot \]  

(7.3.10)

Then

\[ \bar{d}_T^\top = -(U_S)_{21} \cdot ((U_S)_{11})^{-1} \]  

(7.3.11)

and

\[ d(x, \bar{a}) = x^{(p)} - (U_S)_{21} \cdot ((U_S)_{11})^{-1} x \]

\[ = (U_S^{-1} x)^{(p)} \cdot (U_S)_{pp} \]  

(7.3.12)

(The left member of (7.3.8) becomes

\[ E \left( (2\Phi(c/|U_S^{-1} x)^{(p)}|) - 1 \right) \cdot (U_S^{-1} x)^{(p)} \cdot (U_S)_{pp} \cdot \bar{x} \]

\[ = [E_{n_S}^r(x, r) \cdot (U_S^{-1} x)^{(p)} \cdot \bar{x}] \cdot (U_S)_{pp} = 0 \]  

by (7.2.4).)

Moreover, by (7.3.5), (7.3.9), (7.3.10) and (7.3.12) we have

\[ v^2(\theta*) = E \left( \frac{r \cdot d}{-\psi_v} \right)^2 = E(d^2(x, \bar{a}) \cdot \psi_v^2 / |d|)(x) \]

\[ = (U_S)_{pp}^2 \cdot E(\psi^2 / |z^{(p)}|(x) \cdot |z^{(p)}|^2) \]

\[ = (U_S)_{pp}^2 \cdot \lambda_p(\eta_S) \]  

(7.3.13)

where

\[ z^{(p)} = (U_S^{-1} x)^{(p)} \text{ and } \lambda_p(\eta_S) = E_{n_S}(x, r) \cdot |z^{(p)}|^2 \]

Finally we get

\[ y_n(\theta*) = n^{-1/2} \cdot ((U_S)_{pp} \cdot \lambda_p^{1/2}(\eta_S))^{-1} \cdot \sum_{i=1}^n \left[ r \cdot d \right]_{-\psi_v} \]
\[ = n^{-1/2} \cdot \frac{1}{\eta_S} \cdot \prod_{i=1}^{n} \left( (U_S)^{pp} \right)^{-1} \psi_c (U_S)^{pp} \cdot (r \cdot z(p), (U_S)^{pp}) \]
\[ = n^{-1/2} \cdot \frac{1}{\eta_S} \cdot \prod_{i=1}^{n} \psi_c (r \cdot z(p)) \]
\[ = n^{-1/2} \cdot \frac{1}{\eta_S} \cdot \prod_{i=1}^{n} \psi_c (r \cdot z(p)) \cdot c / |z(p)| \]
\[ = \frac{1}{\eta_S} \cdot Z_n (\theta_star; \eta_S) . \]

PROPOSITION 7.6

The Wang test is always less efficient at the model than the modified Wang test.

Proof.
The modified Wang procedure has the same asymptotic power as the optimally robust C(\(\alpha\))-type test defined by \(\eta_S\). Therefore, the square of its efficacy equals \((U_S)^{pp} / \eta_S^2\) .

On the other hand, the square of the efficacy of the Wang test is given by the formula (see Wang, 1981, p.1104)

\[ s^2(\theta*) = (v(\theta*) - (\epsilon / \delta_1) \cdot (\nu / \nu(\theta*)))^2 . \]

Thus, the relative efficiency of the Wang test with respect to the modified Wang test can be computed as

\[ \text{eff(Wang test , modified Wang test) = } s^2(\theta*) / ((U_S)^{pp} / \eta_S^2) \]

hence, using (7.3.10) and (7.3.13)
\[
\left[ \frac{\lambda_p(n_s)}{(v_s)_{pp}} \right] \cdot \left[ (v_s)_{pp} \lambda_p^{1/2}(n_s) + \frac{(v_s)_{pp}}{(v_s)_{pp} \lambda_p^{1/2}(n_s)} \right]^2
\]

\[
= \left( \lambda_p(n_s) + \frac{(v_s)_{pp}}{(v_s)_{pp} \lambda_p^{1/2}(n_s)} \right)^2.
\]

Using (7.3.2) we get finally

\[
\text{eff}[\text{Wang test, modified Wang test}]
\]

\[
= 1 - c^2 \mathbb{E} \left[ \left( \frac{|z^{(P)}|}{c} \right) \phi'(\frac{c}{|z^{(P)}|}) - \phi(-\frac{c}{|z^{(P)}|}) \right] < 1.
\]

\[
\square
\]
8. COMPUTATIONAL ASPECTS AND COMPARISON BETWEEN DIFFERENT TESTS

8.1 COMPUTATIONAL ASPECTS

In this section we discuss some computational aspects involved in the calculations of optimally bounded influence tests: the computation of the matrix $U_S$ and the computation of the asymptotic distribution.

8.1a Algorithm for the computation of the matrix $U_S$

Our problem is to determine the lower triangular matrix with positive diagonal elements $U = U_S$ which solves the following equation (see Proposition 5.4)

$$E \left[ (2z(c(p-q)^{1/2}/|\bar{z}|) - 1)zz^T \right] = I,$$  \hspace{1cm} (8.1.1)

where $z = U^{-1}x$.

If the $x_i$'s are given and fixed, one has to replace the expectation over $K$ with an average over $\{x_1, \ldots, x_n\}$.

Define:

$$\tilde{M}(U) := E \left[ (2z(c(p-q)^{1/2}/|\bar{z}|) - 1)xx^T \right].$$
Then (8.1.1) is equivalent to

$$\tilde{M}(U) = UU^T.$$  \hspace{1cm} (8.1.2)

We can find the solution $U_S$ of (8.1.2) using the following algorithm.

**ALGORITHM**

1. Define: $M_0 := Exx^T$.
   
   Let $U_0$ be the Choleski decomposition of $M_0$ (that is, $U_0 U_0^T = M_0$ and $U_0$ lower triangular with positive elements). $U := U_0$.

2. Define: $\tilde{M}_1 := \tilde{M}(U)$ and let $U_1$ be the Choleski decomposition of $\tilde{M}_1$.

3. IF $\| U_1 - U \| < \varepsilon$, GOTO 4; otherwise $U := U_1$ and GOTO 2.

4. END.

Empirical experience shows a rather quick convergence.

We use this algorithm for computing the matrix $U$ in Sections 8.2 and 8.3.
8.1b Computation of the asymptotic distribution of the 
\( T \)-test statistic

In order to determine \( P \)-values we have to compute the 
distribution function of \( \sum_{j=1}^{m} \lambda_j N_j^2 \), where the \( N_j \)'s are \( m \) independent normal distributed random variables and 
\( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m > 0 \) are \( m \) positive real numbers (see Proposition 4.6 with \( m := p-q \)).

Let us denote this distribution function by \( F_m(z; \lambda_1, \ldots, \lambda_m) \).

Johnson and Kotz (1970, p.149 ff.) discuss different approximations to \( F_m \), especially a power series expansion, a representation of \( F_m \) as a mixture of \( \chi^2 \)-distributions and an 
expansion in terms of Laguerre polynomials.

All these expansions seem to converge slowly (see Grenander, Pollak and Slepian, 1959, p. 377). Therefore, we prefer to 
follow an approach due to Grenander, Pollak and Slepian (1959).

It is based on the following integral equation for \( F_m \):

\[
z \cdot F_m(z; \lambda_1, \ldots, \lambda_m) = \int_0^z \alpha(z-y) F_m(y; \lambda_1, \ldots, \lambda_m) dy,
\]

(8.1.3)

where \( \alpha(t) := 1 + \frac{1}{2} \sum_{j=1}^{m} \exp(-t/(2\lambda_j)) \).

(8.1.3) can be discretized and one obtains

\[
F_m(\ell \Delta; \lambda_1, \ldots, \lambda_m) = (\ell - \frac{1}{2} \alpha(0))^{-1} \cdot \prod_{\nu=1}^{\ell-1} \alpha((\ell-\nu) \cdot \Delta) F_m(\nu \Delta; \lambda_1, \ldots, \lambda_m),
\]

(8.1.4)

for \( \ell = 2, 3, \ldots \).

For details, see Grenander, Pollak and Slepian (1959).
REMARK

It would be interesting to try and use small-sample asymptotic techniques in order to find a better approximation to the distribution function $F_m$. These methods are very promising and have been used successfully in other situations, for example, for getting an approximation to the distribution of M-estimators (see Hampel, 1973; Field and Hampel, 1982).
8.2 ASYMPTOTIC BEHAVIOUR OF DIFFERENT TESTS

In this section we compare the asymptotic behaviour of different tests for different distributions. In our exposition we apply the methods which have been used by Maronna, Bustos and Yohai (1979) for comparing different estimators.

We consider the following linear model

\[ y = x^T \theta + \varepsilon, \]

where \( x^T = (1, x^{(2)}) \), \( \theta^T = (\theta^{(1)}, \theta^{(2)}) \).

Let

\[ H_0 : \theta^{(2)} = 0 \]

be the null hypothesis and let \( G \) and \( K \) be the distributions of \( \varepsilon \) and \( x^{(2)} \), respectively.

For \( x^{(2)} \) and \( \varepsilon \) we choose the following contaminated normal distributions

\[ (1-\varepsilon)\Phi(t) + \varepsilon \Phi(t/s) \]

with the following parameters

\[
\begin{array}{c|cccc}
\varepsilon & 0 & 0.1 & 0.1 & 0.05 \\

s & 1 & 3 & 5 & 10 \\
\end{array}
\]

We want to study the asymptotic behaviour of the optimally bounded influence tests:

\[ \tau_M(x, r) = w_{c_1}(|z^{(2)}|), \rho_{c_2}(r) \quad \text{(see Proposition 5.2)} \]

\[ \tau_S(x, r) = \rho_{c_3/|z^{(2)}|}(r) \quad \text{(see Proposition 5.4, Corollary 7.3)} \]
Moreover, in order to investigate the behaviour of a test defined by means of a redescending $\eta$-function 
$$(\eta(x,r) = (\partial\partial r)\tau(x,r))$$, we consider also the following procedure:

$$\tau_{MT}(x,r) = w_{c_{6}}(|z^{(2)}|) \cdot \tilde{\rho}(r;\kappa^{*},c_{5},c_{6},A_{5},B_{5})$$,

where $\tilde{\rho}$ is the $\rho$-function defining the hyperbolic tangent estimator (see Subsection 2.4a).

$\tilde{\rho}$ is defined by

$$\tilde{\rho}(r;\kappa^{*},c_{5},c_{6},A_{5},B_{5}) = \begin{cases} 
\frac{x^{2}}{2} & \text{if } 0 \leq |r| \leq c_{6} \\
\left(c_{6}\right)^{2}/2 + b(c_{6}) - b(r) & c_{6} < |r| \leq \kappa^{*} \\
\left(c_{5}\right)^{2}/2 + b(c_{5}) - b(r) & c_{5} < |r| \leq c_{6} 
\end{cases}$$,

where

$$b(r) = 2(A_{5}/B_{5}) \cdot \log \left[ 2 \cdot \cosh \left( \left( (k^{*}-1)B_{5}^{2}/A_{5} \right)^{1/2}/2 \right), (c_{5} - |r|) \right]$$.

![Diagram](image)

**Figure 8.2.1**: the $\rho$-function defining hyperbolic tangent estimators ($\tilde{\rho}$) and Huber's $\rho$-function ($\rho_{\rho_{c_{6}}}$)

Note that $c_{5}$, $A_{5}$, $B_{5}$ are computed implicitly in terms of $c_{5}$ and $\kappa^{*}$.
For each test we compute the standardized sensitivities 
(at G=Φ)
\[ u_{22}^{-1} \cdot \sup \{|η(x, r)| \cdot |z^{(2)}| : x, r \} \]
and the efficacies
\[ M_{22.1}/λ_2 = u_{22}^{-1}/λ_2 \]
under different distributions.
Note that the efficacy of the tests defined by τ_M and τ_MT factorizes; in these cases we have
\[ M_{22.1}/λ_2 = DX \cdot DR \]
where DX depends only on w_c and K and DR = B^2/A, with
\[ A = Eψ_c^2, \quad B = Eψ'_c, \quad ψ_c(r) = (∂ψ/∂r)φ_c(r) . \]
The constants are chosen so that all the tests have the same asymptotic efficiency, when \( x^{(2)} \sim N(0, 1) \) and \( e \sim N(0, 1) \).
Table 8.2.2 describes the calibration as well the standardized sensitivities of the tests.
From Table 8.2.3 one can obtain the following conclusions:
1) τ_MT is better than τ_M for all distributions under consideration.
2) τ_S is better than τ_MT when the distribution of e has moderate tails.
3) τ_S has the better standardized sensitivities (computed at G=Φ).
Table 8.2.4 gives the factors for the τ_S and τ_MT-test.
<table>
<thead>
<tr>
<th>TEST</th>
<th>CONSTANTS</th>
<th>DX</th>
<th>DR</th>
<th>EFF</th>
<th>$u_{22}$</th>
<th>A</th>
<th>B</th>
<th>ST. SENS.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_M$</td>
<td>$c_1 = 1.80$ \hspace{1em} $c_2 = 1.58$</td>
<td>.98</td>
<td>.97</td>
<td>.95</td>
<td>.900</td>
<td>.809</td>
<td>.886</td>
<td>3.16</td>
</tr>
<tr>
<td>$\tau_{MT}$</td>
<td>$c_4 = 1.80$ \hspace{1em} $c_5 = 4.68$ \hspace{1em} $\kappa^* = 4.50$</td>
<td>.98</td>
<td>.97</td>
<td>.95</td>
<td>.910</td>
<td>.842</td>
<td>.905</td>
<td>3.37</td>
</tr>
<tr>
<td></td>
<td>$c_6 = 1.70$ \hspace{1em} $A_5 = .84$ \hspace{1em} $B_5 = .90$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_S$</td>
<td>$c_3 = 2.67$</td>
<td>-</td>
<td>-</td>
<td>.95</td>
<td>.930</td>
<td>-</td>
<td>-</td>
<td>2.67</td>
</tr>
</tbody>
</table>

**Table 8.2.2**: Calibration and standardized sensitivities

at $x^{(2)} \sim N(0, 1)$, $e \sim N(0, 1)$
<table>
<thead>
<tr>
<th>K</th>
<th>TEST</th>
<th>$G_{s=0}$</th>
<th>$G_{s=1}$</th>
<th>$G_{s=3}$</th>
<th>$G_{s=5}$</th>
<th>$G_{s=10}$</th>
<th>$ST_{SENS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon=0$</td>
<td>$s=1$</td>
<td>$\tau_M$</td>
<td>.95</td>
<td>.75</td>
<td>.71</td>
<td>.80</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>$\tau_{MT}$</td>
<td>.95</td>
<td>.77</td>
<td>.77</td>
<td>.87</td>
<td>3.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_S$</td>
<td>.95</td>
<td>.76</td>
<td>.72</td>
<td>.78</td>
<td>2.87</td>
<td></td>
</tr>
<tr>
<td>$\varepsilon=1$</td>
<td>$s=3$</td>
<td>$\tau_M$</td>
<td>1.52</td>
<td>1.21</td>
<td>1.13</td>
<td>1.28</td>
<td>2.70</td>
</tr>
<tr>
<td></td>
<td>$\tau_{MT}$</td>
<td>1.53</td>
<td>1.23</td>
<td>1.23</td>
<td>1.40</td>
<td>2.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_S$</td>
<td>1.54</td>
<td>1.27</td>
<td>1.16</td>
<td>1.25</td>
<td>2.34</td>
<td></td>
</tr>
<tr>
<td>$\varepsilon=1$</td>
<td>$s=5$</td>
<td>$\tau_M$</td>
<td>2.44</td>
<td>1.94</td>
<td>1.91</td>
<td>2.04</td>
<td>2.33</td>
</tr>
<tr>
<td></td>
<td>$\tau_{MT}$</td>
<td>2.44</td>
<td>1.97</td>
<td>1.97</td>
<td>2.24</td>
<td>2.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_S$</td>
<td>2.51</td>
<td>2.06</td>
<td>1.86</td>
<td>2.00</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>$\varepsilon=.05$</td>
<td>$s=10$</td>
<td>$\tau_M$</td>
<td>2.99</td>
<td>2.38</td>
<td>2.23</td>
<td>2.51</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>$\tau_{MT}$</td>
<td>3.00</td>
<td>2.42</td>
<td>2.42</td>
<td>2.76</td>
<td>2.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_S$</td>
<td>3.24</td>
<td>2.61</td>
<td>2.32</td>
<td>2.50</td>
<td>1.85</td>
<td></td>
</tr>
</tbody>
</table>

Table 8.2.3: Asymptotic efficacies
<table>
<thead>
<tr>
<th>( \varepsilon = 0 )</th>
<th>( \varepsilon = .1 )</th>
<th>( \varepsilon = .1 )</th>
<th>( \varepsilon = .05 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s = 1 )</td>
<td>( s = 3 )</td>
<td>( s = 5 )</td>
<td>( s = 10 )</td>
</tr>
</tbody>
</table>

\[ \begin{array}{cccc}
\tau_M & 0.970 & 0.771 & 0.722 & 0.814 \\
\tau_{MT} & 0.973 & 0.785 & 0.785 & 0.894 \\
\end{array} \]

\[ \begin{array}{cccc}
D_X & 979 & 1.589 & 2.512 & 3.087 \\
\end{array} \]

Table 8.2.4: Factorization for \( \tau_M \) and \( \tau_{MT} \)-test.
8.3 NUMERICAL EXAMPLES

8.3a Example 1

The data to this example are taken from Ezekiel and Fox (1959, p.57-58). The water flow at two different points (Libby, Mont. and Newgate, B.C.) on Kootenay River in January was measured. Figure 8.3.1 and Figure 8.3.2a) show the data.

<table>
<thead>
<tr>
<th>Year</th>
<th>Newgate</th>
<th>Libby</th>
</tr>
</thead>
<tbody>
<tr>
<td>1925</td>
<td></td>
<td>42.0</td>
</tr>
<tr>
<td>26</td>
<td>24.0</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>38.0</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>49.4</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>24.6</td>
<td></td>
</tr>
<tr>
<td>1930</td>
<td></td>
<td>24.2</td>
</tr>
<tr>
<td>31</td>
<td>19.7</td>
<td>27.1</td>
</tr>
<tr>
<td>32</td>
<td>18.0</td>
<td>20.0</td>
</tr>
<tr>
<td>33</td>
<td>26.1</td>
<td>33.4</td>
</tr>
<tr>
<td>34</td>
<td>44.9</td>
<td>77.6</td>
</tr>
<tr>
<td>1935</td>
<td>26.1</td>
<td>37.0</td>
</tr>
<tr>
<td>36</td>
<td>19.9</td>
<td>21.6</td>
</tr>
<tr>
<td>37</td>
<td>15.7</td>
<td>17.6</td>
</tr>
<tr>
<td>38</td>
<td>27.6</td>
<td>35.1</td>
</tr>
<tr>
<td>39</td>
<td>24.9</td>
<td>32.6</td>
</tr>
<tr>
<td>1940</td>
<td>23.4</td>
<td>26.0</td>
</tr>
<tr>
<td>41</td>
<td>23.1</td>
<td>27.5</td>
</tr>
<tr>
<td>42</td>
<td>31.3</td>
<td>38.7</td>
</tr>
<tr>
<td>43</td>
<td>23.8</td>
<td>27.6</td>
</tr>
</tbody>
</table>

Figure 8.3.1: Water flow at two points on Kootenay River, in January [Units of cfs]
Figure 8.3.2

Water flow at Newgate vs. Water flow at Libby.
Let
\[ Y = \text{Water flow at Newgate} , \quad Z = \text{Water flow at Libby} , \]
and consider the following linear model ("straight line")
\[ y_i = \alpha + \beta z_i + \epsilon_i , \quad i = 1931, 1932, \ldots, 1943 . \]

We want to test the hypothesis
\[ H_0 : \beta = 0 . \]

From Figure 8.3.2a) we see that \( z_{1934} \) is an outlier in the factor space. We want to study the behaviour of different tests (classical \( F \)-test, \( \rho \)-test, optimal \( \tau \)-test) when the observation \( y_{1934} \) (corresponding to \( z_{1934} \)) varies between 0 and its actual value 44.9. The tests under study are defined by the following functions

<table>
<thead>
<tr>
<th>test</th>
<th>( \tau(x, r) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F )</td>
<td>( r^2 / 2 )</td>
</tr>
<tr>
<td>( \rho_{c_1} )</td>
<td>( \rho_{c_1}(r) )</td>
</tr>
<tr>
<td>optimal ( \tau )</td>
<td>( \rho_{c_2} /</td>
</tr>
</tbody>
</table>

The scale parameter \( \sigma \) was estimated in the full model using Huber's Proposal 2.
Choice of the constants

1. $c_1$ and $c_2$ can be chosen such that the corresponding tests have a given efficiency, say .95, at the normal model (that is, when $z_i$ and $e_i$ are normally distributed).

We obtain the following values:

$$c_1 = 1.345 \quad c_2 = 2.67.$$ 

2. For the optimal $\tau$-test there is another possibility for choosing the constant $c_2$. Consider the "location case" in our situation

$$q = p-1, \quad k(x) = k(\bar{x},1) \quad (\bar{x} = 1)$$

and compute the constant $c_2$ corresponding to an efficiency of .95 in this case. A similar method is used by Krasker and Welsch (1979) for computing the constants of bounded-influence estimators for regression.

In this case equation (5.3.2) becomes

$$2\Phi(c(\lambda^{(M)}_S)^{1/2}) - 1 = \lambda^{(M)}_S,$$ \hfill (8.3.1)

which defines $\lambda^{(M)}_S$ implicitly. Moreover, the efficiency equals the expression

$$\text{efficiency} = \left[ -c^2 + (c^2+1)/\lambda^{(M)}_S - (c/\lambda^{(M)}_S)(2/\pi)^{1/2} \exp\left( - \frac{1}{2} c^2 \lambda^{(M)}_S \right) \right]^{-1}.$$ \hfill (8.3.2)
Given \( c \), we can compute \( \lambda_S^{(M)} \) by means of (8.3.1); then, (8.3.2) gives us the efficiency. Table 8.3.3 shows some values of \( c \) and the corresponding efficiencies.

\[
\begin{array}{ccc}
\hline
\text{c} & \lambda_S^{(M)} & \text{efficiency} \\
\hline
1.7 & 0.891 & 0.95 \\
1.9 & 0.933 & 0.97 \\
2.3 & 0.977 & 0.99 \\
\hline
\end{array}
\]

Table 8.3.3: Efficiencies and bending constant

Figure 8.3.2b) shows the \( \log_{10} \) P-values (computed by means of the asymptotic distribution \( \chi^2_1 \)) for the F-test and the \( \rho_c \)-test, for \( \text{0}_c \leq y \leq 40 \). The \( \log_{10} \) P-values corresponding to the optimal \( \tau \)-test (with constant \( c_2 = 2.67 \)) was less than -13 for all \( y > 0 \). (One gets the same result using \( c_2 = 1.7 \).)

From Figure 8.3.2b) we see that the F-test and the \( \rho_c \)-test becomes significant only for \( y > 26 \), while the optimal \( \tau \)-test is already strong significant for \( y > 0 \). This shows the excellent robustness properties of the optimal \( \tau \)-test in this situation.
8.3b Example 2
(computed in collaboration with W. Stahel)

The data to this example are taken from Draper and Smith (1966, p.104 ff.)

We have the following variables:

- \( Y \) = response or number of pounds of steam used per month,
- \( X_8 \) = average atmospheric temperature in the month (in °F),
- \( X_6 \) = number of operating days in the month.

Figure 8.3.4 shows the data.

<table>
<thead>
<tr>
<th>Obs.No.</th>
<th>( X_8 )</th>
<th>( X_6 )</th>
<th>( Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35.3</td>
<td>20</td>
<td>10.98</td>
</tr>
<tr>
<td>2</td>
<td>29.7</td>
<td>20</td>
<td>11.13</td>
</tr>
<tr>
<td>3</td>
<td>30.8</td>
<td>23</td>
<td>12.51</td>
</tr>
<tr>
<td>4</td>
<td>58.0</td>
<td>20</td>
<td>8.40</td>
</tr>
<tr>
<td>5</td>
<td>61.4</td>
<td>21</td>
<td>9.27</td>
</tr>
<tr>
<td>6</td>
<td>71.3</td>
<td>22</td>
<td>8.73</td>
</tr>
<tr>
<td>7</td>
<td>74.4</td>
<td>11</td>
<td>6.36</td>
</tr>
<tr>
<td>8</td>
<td>76.7</td>
<td>23</td>
<td>8.50</td>
</tr>
<tr>
<td>9</td>
<td>70.7</td>
<td>21</td>
<td>7.82</td>
</tr>
<tr>
<td>10</td>
<td>57.5</td>
<td>20</td>
<td>9.14</td>
</tr>
<tr>
<td>11</td>
<td>46.4</td>
<td>20</td>
<td>8.24</td>
</tr>
<tr>
<td>12</td>
<td>28.9</td>
<td>21</td>
<td>12.19</td>
</tr>
<tr>
<td>13</td>
<td>28.1</td>
<td>21</td>
<td>11.88</td>
</tr>
<tr>
<td>14</td>
<td>39.1</td>
<td>19</td>
<td>9.57</td>
</tr>
<tr>
<td>15</td>
<td>46.8</td>
<td>23</td>
<td>10.94</td>
</tr>
<tr>
<td>16</td>
<td>46.5</td>
<td>20</td>
<td>9.58</td>
</tr>
<tr>
<td>17</td>
<td>59.3</td>
<td>22</td>
<td>10.09</td>
</tr>
<tr>
<td>18</td>
<td>70.0</td>
<td>22</td>
<td>8.11</td>
</tr>
<tr>
<td>19</td>
<td>70.0</td>
<td>11</td>
<td>6.83</td>
</tr>
<tr>
<td>20</td>
<td>74.5</td>
<td>23</td>
<td>8.88</td>
</tr>
<tr>
<td>21</td>
<td>72.1</td>
<td>20</td>
<td>7.88</td>
</tr>
<tr>
<td>22</td>
<td>58.1</td>
<td>21</td>
<td>8.47</td>
</tr>
<tr>
<td>23</td>
<td>44.6</td>
<td>20</td>
<td>8.86</td>
</tr>
<tr>
<td>24</td>
<td>33.4</td>
<td>20</td>
<td>10.36</td>
</tr>
<tr>
<td>25</td>
<td>28.6</td>
<td>22</td>
<td>11.08</td>
</tr>
</tbody>
</table>

*Figure 8.3.4*
We consider the linear model

\[ Y = \alpha + \beta_8 X_8 + \beta_6 X_6 + \epsilon, \]

and we want to test the hypothesis

\[ H_0 : \beta_6 = 0. \]

The factor space is given by Figure 8.3.5 and the observations are plotted in Figure 8.3.6a).

![Figure 8.3.5: The variables $X_8, X_6$](image)

From Figure 8.3.5 we see that there exist two outliers in the factor space. As in the previous example, we study the behaviour of the $\log_{10} p$-values of the $F$, $\rho$- and optimal $\tau$-test when the observation ($y_7$) corresponding to the point ($X_8 = 74.4$, $X_6 = 11$) varies between 0 and 20. (Its actual value is 6.36.)
Figure 8.3.6b) shows the overall excellent behaviour of the \( \tau \)-test (strongly significant for all \( y \! \)), the good behaviour of the \( \rho_c \)-test (at least for \( y \geq 8 \)) which is still significant (at the 5% level) in the region \( y \geq 8 \) and the bad behaviour of the \( F \)-test which becomes even not significant for \( 8.7 \leq y \leq 18.7 \).
REFERENCES


Huber, P.J. (1968), "Robust Confidence Limits", Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 10, 269-278.


INDEX

admissibility of estimators, 57
Akaike, 66,106,109,150
    criterion, 108-109
    generalized criterion, 109
Andersen, 67,150
Andrews, 54,71
asymptotic
    mean square error, see mean square error
    minimax test, see minimax
    power of a test, see power
    relative efficiency, see efficiency
    testing model, see model

Bahadur, 34,35,150
    efficiency, see efficiency
Bodnarski, 14,150
Bellman, 76,150
Beran, 12,150
Bhansali, 109,150
Bickel, 12,87,150
breakdown point, 9,11-12
Bustos, 135,154

C(α) tests, 117
C(α)-type tests for linear models, II,7,115
    connections with an asymptotically minimax test, 127-129
    definition, 116-118
    influence function and asymptotic power, 119-123
    optimally robust, 124,125,127,129
capacities, 2-alternating, 14
Carroll, 97,151
censored likelihood ratio test, 13-14
change-of-bias function, 45
change-of-power function, I,6,10,41,47
change-of-variance function, 1,6,10,41
of general M-estimators, 58-64
of M-estimators for location, 42-46
of R-estimators, 45,47
change-of-variance sensitivity, see sensitivity
self-standardized, see sensitivity
χ²-distribution, 79,80,145
mixture, 133
Choleski decomposition, 78,132
Collins, 8,46,151
contamination
c-, 13,28,29,74,135
free, 12
continuity of estimators, 11
covariance matrices, 9
derivative of a functional, see functional
design matrix, 48
diagonalization of symmetric and positive definite matrices, 78
derdifferentiability
compact, 18
Fréchet, 18
Gâteaux, 17
dispersion function, 52
Downham, 109,150
Draper, 146,151
efficacy, 137
asymptotic Pitman, 26,28,33,35,39,41,47
efficiency
asymptotic relative, 26
Bahadur, 34
Pitman, 34,35
relative, 35
empirical distribution, 16,19,23,32,80
entropy, 109
Eplett, 8,39,40,41,151
estimable function, 49,65
Ezekiel, 141,151
F-distribution, 65
Fiechter, 55,152
Field, 134,151
Filippova, 18,19,151
Fisher
  consistency, 17,22,23,24,25,28,74
  information, 9,27,96
Fox, 141,151
Fréchet differentiability, see differentiability
F-test for linear models, I,II,5,23,65-66,66,97,114,142-143,147-149
  functional
    statistic viewed as, 4,17,18,24,73
    derivative of, 17,18
    von Mises, 18,23,32

Gâteaux differentiability, see differentiability
Gauss-Markov theorem, 49
Grenander, 133,151
gross-error, 10
  -model, see model
  -sensitivity, see sensitivity
  grouping effects, 4,10

Hall, 36,154
Hampel, 4,5,9,10,11,12,16,17,22,33,42,45,50,52,53,56,65,66,71,96,134,151,152
  approach, 8; see also infinitesimal approach
criterion, see optimality criterion for infinitesimal robustness of
  estimators
  -Krasker estimator, 6,57,60,61
  lemma 5, 22,33,96
Hannschn, 55,152
Hettmansperger, 50,66,67,90,97,155
Hill, 54,55,71,152
Hodges, 25,152
Hôpital's rule, 75
Huber, 2,5,6,9,10,13,14,21,27,43,46,50,51,52,54,71,83,84,152,153
  approach, 8; see also minimax approach
  proposal 2, 52,71-72,112,143
  regression, 7,95
Huber estimators
  for linear models, 52-54
  for location, 7
  influence function of, 52-53
Huber function

\[ \psi^{-}, 22 \]

\( p^{-}, 2, 96, 136 \)

Huber-Carl, 14, 28, 153

hyperbolic tangent estimators, 42, 45-46, 136

infinitesimal

approach, I, 5, 15, 16

robustness, see robustness

influence function, I, 4, 6, 7, 9

level, 29-30

power, 29-30

influence function for estimators, 5, 10, 16-20, 25

properties, 18-19

influence function for M-estimators, see M-estimators

influence function for testing, 5, 6, 10, 31

connections with other, 34-41

definition, 23-25

properties, 25-30

influence function of \( C(\alpha) \)-type tests, see \( C(\alpha) \)-type tests

influence function of \( \tau \)-tests, see \( \tau \)-tests

influence of position in factor space, II, 53, 90

influence of the residual, II, 53, 90, 96

Jensen, 67, 150

Johnson, 133, 153

Kent, 87, 153

Kohlas, 55, 152

Kotz, 133, 153

Krasker, 6, 22, 54, 56, 57, 63, 88, 101, 144, 153

estimator, see Hampel-Krasker estimator

Lambert, 36, 37, 38, 153, 154

least favorable

distribution, 9, 96

pair of distributions, 14

least squares

estimators, 48, 49, 65, 71

procedures, 48, 50, 66-67

weighted, 51
Lehmann, 25, 65, 152, 154
lemma 5, see Hampel’s Lemma 5
level influence function, see influence function
likelihood ratio test, 86, 88, 87, 96, 110
linear models, see model
local-shift-sensitivity, see sensitivity
location model, see model

Mallows, 6, 54, 71, 95, 110, 154
Marazzi, 107, 154
Maronna, 54, 55, 135, 154
maximum likelihood estimator, 9, 108–110
McKean, 52, 67, 90, 154, 155
mean, 8
mean square error
asymptotic, 21, 59
median deviation, 52, 71–72
M-estimators
asymptotic normality, 21, 43, 55
change-of-variance function, see change-of-variance function
for linear models, 54–57, 70
for location, 8, 32, 43, 45, 90
general, 6, 20, 134
influence function, 20, 21
receding, 45, 136; see also hyperbolic tangent estimators
three-part receding, 52
minimax
approach, 5, 9
asymptotically test, II, 7, 14, 15, 115, 117, 125–126
criterion, 8, 9, 13, 46, 125, 126
robust, 9
model
asymptotic testing, 14
full, 65, 67, 70, 71, 79, 84
gross-error, 9, 96
linear, I, 5, 6, 48, 68
location, I, 8, 15, 39
M-tests, 30, 32–33, 47, 117

neighbourhood, 8, 44, 50
Prohorov, 13
shrinking, 14, 125
total variation, 13, 14
Newton's method, 8
Neyman, 117,154
Noether, 26,28,35,154
nuisance parameters, 15,117,122,125

optimality criterion for infinitesimal robustness of estimators
with respect to the change-of-variance function, 44-45,60-64
with respect to the influence function, 5,20-22,56
optimality criterion for infinitesimal robustness of tests, I,5,6,23,
30-31,115,124
optimally bounded influence estimators, II
choice of the constants, 144
optimally bounded influence tests, I,II,5,6,7,15,30,67,95,124,143-145,
147-149
asymptotic behaviour, 135-140
computation, 131-134
choice of the constants, 144-145
outlier, 53
in the factor space, 53-54,67,143,147

permutation tests, 39,66
Pitman
approach, 30
asymptotic efficacy, see efficacy
efficiency, see efficiency
theorem, 26
Pollak, 133,151
power
asymptotic of a test, 26,27,28,30,31,35,95,96,105,120
influence function, see influence function
Prohorov distance, 10
proposal 2, see Huber's proposal 2
pseudodistance, 79
pseudoinverse, 1
Ψ-function, see Huber Ψ-function
P-values, 34,36,37,133,142,147,149
asymptotic distribution, 36

Reeds, 17,18,19,155
rejection point, 18-19
residual, see influence of the residual
R-estimators, 32,47
change-of-variance function, see change-of-variance function
Robust

estimation in linear models, 5, 6, 51-64
first-step, 6
score function, 22
second-step, 6
selection procedure, II, 7, 108-112, 113-114
stepwise regression, 107, 114
testing, I, 5, 9, 13-15

Robustness

correction, 8
qualitative, 9-11
stability aspects, 10
distributional, 67
infinitesimal, I, 5, 23, 26, 30, 95
\( \rho \)-test, II, 6, 7, 96-97, 142-145, 147-149
\( \rho \)-function, see Huber \( \rho \)-function
Ronchetti, 24, 25, 26, 28, 29, 30, 32, 33, 41, 42, 45, 47, 96, 152, 155
roundoff effects, 4, 10
Rousseeuw, 24, 25, 26, 29, 30, 33, 41, 42, 45, 52, 96, 152, 155
R-tests, 14, 32, 47
two-sample, 39, 47
Ryan, 55, 71

Scale parameter, 52, 71, 143
Scheffe, 49, 66, 155
Schou, 67, 150
Schroeder, 50, 52, 66, 67, 90, 97, 154, 155
Schweppe, 55, 57, 71-102, 152
score generating function, 39-40
Seber, 65, 156
self-standardized sensitivity, see sensitivity
sensitivity
change-of-variance, 42, 44, 58
gross-error, 18-19, 21, 22, 45, 56, 57, 105, 137, 138, 139
local-shift, 18-19
self-standardized, 56, 91
self-standardized change-of-variance, 59
Slepian, 133, 151
slope of a sequence of tests, 35, 36, 37, 40, 41
small-sample asymptotics, 134
Smith, 146, 151
Stahel, 20, 22, 56, 57, 74, 88, 158
standardization of the test statistic, 90, 91-93, 94, 122
standard sequence, 34
Stigler, 49,156
Stone, 113,156
Strassen, 14,153
symmetry radial, 60,93

T-tests, 7,124
asymptotic distribution, 6,80-94,133-134
definition,69-71
elements, 71
influence function, 73-79
optimally robust for Huber’s type regression, 96-97
optimally robust for Mallows’type regression, 98-100,107
optimally robust for the general case, 101-105,107
Taylor expansion, 83
tests based on quadratic forms, 82-87, 124
test problem in linear models, 65-68
three-part redescending M-estimators, see M-estimators

von Mises, 18,19,80,156
functional, see functional

Wang, 7,15,115,117,122,125,126,127,129,156
weak topology, 10
Welsch, 6,54,56,57,63,88,98,106,144,153,156
Wieand, 35,36,156

Yohai, 54,55,135,154
CURRICULUM VITAE

I was born on November 4, 1955 in Mendrisio (TI, Switzerland) as the first child of Flavio and Giuseppina Ronchetti-Caverzasio. My full name is Elvezio Mauro Danilo Ronchetti and I am citizen of Monte (TI, Switzerland).

From 1961 to 1966 I visited the primary school and from 1966 to 1971 the "Ginnasio" in Mendrisio. In 1971 I entered the "Liceo Cantonale" in Lugano, where I obtained the "Maturità" (Highschool Diploma) in 1974, with the best grades (winning the Maraini-prize).

From 1974 to 1979 I studied mathematics and physics at the Swiss Federal Institute of Technology (ETH) in Zürich, where I passed in 1979 my Diploma in Mathematics best (with distinction and obtaining the Polya-prize). My Diploma Thesis, written under the direction of Prof. F.R. Hampel, was on "Robustness Properties of Tests".

Since spring 1979 I am working as assistant at the Statistics Group of the ETH, where I wrote the present work.

My main field of research is robust statistics, especially infinitesimal methods in robust estimation and testing.