

A COMPARISON BETWEEN SEVERAL
ONE-STEP M-ESTIMATORS OF LOCATION AND DISPERSION
IN THE PRESENCE OF A NUISANCE PARAMETER

by

EVE RAINVILLE

B.Sc. University of Ottawa, 1994

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
Department of Statistics

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1996

©Eve Rainville, 1996

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics

The University of British Columbia
Vancouver, Canada

Date August 1st, 1996

Abstract

The idea of one-step estimators has long been used: Le Cam (1956), Neyman (1949) and Fisher (1922) have proposed it in the context of maximum likelihood estimation. More recently, Bickel (1975) adapted this idea to robustness theory when he introduced one-step Huber M-estimators for simple linear models. Huber (1981) and Hampel et al (1986) further investigated the advantages of such one-step M-estimators; while retaining the robustness properties of their initial estimates, one-step M-estimators show increased efficiency, and thus represent a good compromise between robust and parametric estimation.

Different versions of one-step M-estimators, some more numerically stable than others, have been proposed throughout the years. To our knowledge, no thorough comparison of available one-step M-estimators have been done using modern techniques, as in Rousseeuw and Croux (1993a). In this thesis, two versions of one-step M-estimators of location, obtained with the Newton-Raphson method, are studied in the context of unknown dispersion. Their asymptotic efficiencies at Gaussian and non-Gaussian models, as well as their maximum asymptotic bias are compared. We also introduce two new one-step M-estimators of dispersion with unknown location, and challenge the traditional fixed-point method one-step M-estimator of dispersion, originating from Huber (1981) and used by Rousseeuw and Croux (1993a). We identify the optimal situations in which to use any of those three one-step M-estimators of dispersion, using their relative asymptotic efficiency at different models, and their explosion and implosion maximum asymptotic bias curves.

Table of Contents

Abstract	ii
List of Tables	vi
List of Figures	viii
Acknowledgements	x
1 Introduction	1
2 M-Estimators: Maximum Likelihood Type of Estimators	10
2.1 Univariate Problem	10
2.1.1 Qualitative Robustness	11
2.1.2 Infinitesimal Aspects	12
2.1.3 Quantitative Robustness	15
2.2 Nuisance Parameter in the Location-Dispersion Problem	18
3 One-Step M-Estimators	21
3.1 General Idea of One-Step M-Estimators	21
3.2 Estimation of Location With Unknown Dispersion	22
3.3 Estimation of Dispersion With Unknown Location	24
3.4 The First Step Is a Big Step	27
4 Location Estimation with Unknown Dispersion	31

4.1	Introduction	31
4.2	Asymptotic Efficiency of the MOSME of Location	33
4.2.1	Our Choice of Underlying Distributions and of Score Functions	33
4.2.2	The Asymptotic Variance of the Standard One-Step and the Fully-Iterated M-Estimators of Location	35
4.2.3	Asymptotic Variance of the MOSME of Location	36
4.2.4	The Asymptotic Variance of the MLE	38
4.2.5	Asymptotic Efficiency of the MOSME Compared to That of the Standard One-Step (or Fully-Iterated) M-Estimator of Location	40
4.3	Maximum Bias of the MOSME of Location	42
4.3.1	Monotone Non-Decreasing Score Functions	43
4.3.2	Redescending Score Functions	47
4.3.3	Any Type of Score Function	49
4.3.4	Further Work to Be Done	50
4.4	An Example: Hummingbirds	50
4.5	Conclusions	53
5	Dispersion Estimation with Unknown Location	61
5.1	Introduction	61
5.2	Relative Asymptotic Efficiency of the MOSME of Dispersion	63
5.2.1	Our Choice of Underlying Distributions and of Score Functions	63
5.2.2	Asymptotic Value of the MOSME, the One-Step M-Estimator and the τ -Estimator of Dispersion	65
5.2.3	Asymptotic Variance of the MOSME of Dispersion	70

5.2.4	Asymptotic Variance of the Standard One-Step M-Estimator of Dispersion	72
5.2.5	Asymptotic Variance of the τ -Estimator of Dispersion	73
5.2.6	Relative Asymptotic Efficiency of the MOSME Compared to That of the One-Step M-Estimator and the τ -Estimator of Dispersion .	74
5.3	Maximum Bias of the MOSME of Dispersion	78
5.3.1	Further Work to Be Done	85
5.4	Continuation of The Hummingbird Example	86
5.5	Conclusions	88
Bibliography		97
A Normalization of Densities for Calculations of Asymptotic Efficiencies		100
B Derivation of the Influence Function of the MOSME of Location		103
C Derivation of the Influence Function of the MOSME of Dispersion		105
D Derivation of the Influence Function of the One-Step M-Estimator of Dispersion		106
E Derivation of the Influence Function of the τ-Estimator of Dispersion		108

List of Tables

1.1	Results of the Implementation of the Loop in Turbo Pascal	7
4.2	The Value of the Constant Denominator, $E_{\Phi}\psi'(z)$, in the Ratio Defining the Three MOSME's of Location Under Study	35
4.3	Asymptotic Variance $V(\text{MLE}, F)$ of the MLE for Different Underlying Distributions F	39
4.4	Asymptotic Efficiency of the MOSME and the Standard One-Step M-Estimator of Location (Equivalent to That of the Fully-Iterated M-Estimator), Derived from Different Score Functions ψ and for Different Underlying Distributions F . The Asymptotic Efficiency of the Median, the Initial Estimator of Location, is Provided for Comparison Purposes. The Normalized MAD is Used As an Initial Estimator of Dispersion.	40
4.5	Measures of Location of Flying Times of Four Types of Hummingbirds: Adult Females (AF), Adult Males (AM), Junior Females (JF) and Junior Males (JM)	52
5.6	The Value of the Constant Denominator, $E_{\Phi}\chi'(z)z$, in the Ratio Defining the Three MOSME's of Dispersion Under Study	66

5.7	Asymptotic Value of the MLE, the MOSME, the Standard One-Step M-Estimator and the τ -Estimator of Dispersion, As Well As the Normalized MAD and the Standard Deviation (SD), for Different Underlying Distributions F . The Initial Estimators of Dispersion and Location Used to Calculate the One-Step M-estimators Are Respectively the Normalized MAD and the Median.	67
5.8	Relative Asymptotic Variance $RV(\text{MLE}, F)$ of the MLE for Different Underlying Distributions F	76
5.9	Relative Asymptotic Efficiency of the MOSME, the One-Step M-Estimator and the τ -Estimator of Dispersion Derived from Different Score Functions χ , for Different Underlying Distributions F . The Relative Asymptotic Efficiency of the Initial Estimator of Dispersion, the Normalized MAD, Is Provided for Comparison Purposes. The Median is Used as the Initial Estimator of Location	76
5.10	Measures of Dispersion of Flying Times of Four Types of Hummingbirds: Adult Females (AF), Adult Males (AM), Junior Females (JF) and Junior Males (JM). The MAD is the Median Absolute Deviation Multiplied by the Inverse of $\Phi^{-1}(3/4)$	87
A.11	Normalizing Factor d_0 Needed to Standardize the Interquartile Range of Each Distribution F to That of the Standard Normal Distribution	102

List of Figures

4.1	Maximum Bias Function of the MOSME, and a Lower Bound for the Maximum Bias Function of the One-Step M-Estimator and the Fully Iterated M-Estimator of Location, Derived from the $H_{1.345}$ Score Function	56
4.2	Maximum Bias Function of the MOSME, and a Lower Bound for the Maximum Bias Function of the One-Step M-Estimator and the Fully Iterated M-Estimator of Location, Derived From the N_{CDF} Score Function	57
4.3	Lower Bound on the Maximum Bias Function of the MOSME, the Standard One-Step and the Fully-Iterated M-Estimators of Location, Derived from the $T_{4.7}$ Score Function	58
4.4	Maximum Bias Function of the MOSME's of Location Derived from the $H_{1.345}$ and the N_{CDF} Score Functions, and a Lower Bound on the Maximum Bias Function of the MOSME of Location Derived from the $T_{4.7}$ Score Function	59
4.5	Bar Plots of Flying Times of Four Types of Hummingbirds: Adult Females, Adult Males, Junior Females and Junior Males	60
5.6	Explosion Bias Curve of the MOSME, and a Lower Bound for the Explosion Bias Curve of the One-Step M-Estimator of Dispersion, Derived from the $H_{0.975}$ Score Function	90

5.7	Explosion Bias Curve of the MOSME, and a Lower Bound for the Explosion Bias Curve of the One-Step M-Estimator of Dispersion, Derived from the $H_{2.376}$ Score Function	91
5.8	Explosion Bias Curve of the MOSME, and a Lower Bound for the Explosion Bias Curve of the One-Step M-Estimator of Dispersion, Derived from the $T_{3.86}$ Score Function	92
5.9	Explosion Bias Curves of M-Estimators of Dispersion That Are 95% Efficient at the Normal Model, for Small Contamination by Outliers	93
5.10	Implosion Curve of M-Estimators of Dispersion. The MOSME and the One-Step M-Estimator Are Defined Through The $H_{2.376}$ Score Function, and the Tau-Estimator Have the $H_{2.516}$ Score Function. Those M-Estimators Are 95% Efficient at the Normal Model. The Normalized MAD is Provided for Comparison Purposes.	94
5.11	Implosion Curve of M-Estimators of Dispersion. The MOSME and the One-Step M-Estimator Are Defined Through the $T_{3.86}$ Score Function, and the Tau-Estimator, Through the $T_{5.3}$ Score Function. Those M-Estimators Are 95% Efficient at the Normal Model. The Normalized MAD is Provided for Comparison Purposes.	95
5.12	Implosion Curve of MOSME's and τ -Estimators That Are 95% Efficient At the Normal Model	96

Acknowledgements

I am forever grateful to many individuals for helping me complete successfully this thesis.

Thanks to my supervisor, Dr. Ruben Zamar, who has introduced me to robustness theory and guided my thoughts and my research. Although I had no previous experience in robustness, he willingly offered me his financial support and provided me with as much time as I needed to feel comfortable with my thesis topic.

I also wish to thank Dr. Paul Gustafson who found time to revise my thesis in spite of his heavy professional and personal responsibilities.

I must express my gratitude towards Janet Moore who happily provided me with data from her own research about hummingbirds.

Many friends have made my stay in the Statistics department and in Vancouver very enjoyable. Thank you Yulia, Xiaochun, Paige, Grace, Hubert, Shideh, Nancy, Karine, Mike, Didier, Cathy, Terry. I will always remember you.

But most of all, through his constant encouragement and help, my husband Marc Théberge really made a difference. Merci Marc for your availability and contagious curiosity, your great meals, your expertise which you were always eager to share. I couldn't have done it without you.

Chapter 1

Introduction

The importance of point estimation in statistics has long been established. Its usefulness in all the disciplines requiring statistical analysis is increasing more than ever. However, the classical techniques, mostly based on maximum likelihood, still encounter difficulties with less than perfect data, which can sometimes lead to disastrous conclusions. In an attempt to solve the major problem of contamination of the data by outliers, and to adjust to a variety of possible underlying processes generating the data (the true one being impossible to determine), the statistical community developed what is now known as *robustness theory*.

Modest attempts at robustness go back at least as far as two centuries ago. Simple and intuitive robust methods, such as rejection of outliers, have been discussed by Bernoulli (1777) and Bessel and Baeyer (1838). In the 19th and the beginning of the 20th century, other authors considered ways to partly downweight excessive observations, much in the spirit of modern robustness. Tukey (1960) summarized the statistical work of the 1940s and 1950s, demonstrated the nonrobustness of the mean and investigated some robust alternatives. His paper shaped the robustness estimation as a general area of research, and broke the isolation of the early pioneers. See the historical notes by Hampel (1986) (pp. 34-36) for a more complete review of early work in robustness.

But with the first attempt at a reasonably manageable, realistic and comprehensive theory, the robustness theory was officially launched in 1964 by Peter J. Huber through

his famous paper *Robust Estimation of a Location Parameter* (see [18]). In this paper, Huber introduced M-estimators of location, as a generalization of Maximum-likelihood type of estimators, which include the mean and the median, among many others. More specifically, an M-estimator of location is the value θ which satisfies

$$\frac{1}{n} \sum_{i=1}^n \psi(X_i - \theta) = 0, \quad (1.1)$$

where X_1, \dots, X_n is a sample from the population with distribution $F(\theta)$, and ψ is a score function defining the estimator. Huber (1964) found the M-estimator of location which minimizes the maximum asymptotic variance among all location estimators in the symmetric family of ϵ -contaminated distributions

$$V_\epsilon(F_0^\theta) = \{F : F(x) = (1 - \epsilon)F_0(x - \theta) + \epsilon H(x)\},$$

where $0 < \epsilon < 1/2$ is fixed and H is symmetric. This minimax estimator has been called, since then, the Huber M-estimator of location. Considering more generally the asymmetric family $V_\epsilon(F_0^\theta)$, where H is allowed to be asymmetric, Huber(1964) also showed that among all translation equivariant estimators of location, the median minimizes the maximum asymptotic bias. After the publication of this paper, the mean, as an estimator of location, lost without any doubt its momentum.

These two results illustrate two important concerns of robustness theory: the asymptotic efficiency of an estimator versus its asymptotic bias, or more generally, its robustness properties.

Following Huber (1964)'s paper, a variety of robust estimators for dispersion¹, regression, general linear models, and more recently for their multivariate extensions, hypothesis testing, and other more complex statistical models such as time series, were proposed. Through standard techniques, as well as others developed specifically for robustness purposes, it was shown that these estimators offer competitive alternatives to maximum likelihood estimation, especially in the presence of (possibly) corrupted data.

Nevertheless, robustness never appeared to acquire the popularity that would make it become a standard technique of estimation. The following two criticisms about robustness theory may explain why:

- There is generally a trade-off between robustness and efficiency - the more robust an estimator is, the less efficient it is, which affects necessarily the precision of the estimation; and
- Robust estimators represent often a computational challenge, which may require too much computing work and time to overcome.

For example, the popular robust estimators such as the median, the MAD (**M**edian **A**bsolute **D**eviation) and the LMS (**L**east **M**edian of **S**quares) are highly resistant to outliers. However, the Gaussian efficiencies of these estimators is very low: the median and the MAD have respectively 63.7% and 36.7% efficiency (see Hampel et al (1986)), and the LMS regression estimator converges only at the $n^{1/3}$ rate (see Rousseeuw (1984)). Moreover, computing these robust estimators requires more time and memory than the computation of their maximum-likelihood counterparts.

¹Note that the use of the word dispersion in this thesis means what is usually referred to as scale. The dispersion corresponds to the spread of a distribution, whereas the scale is a measure of distance between the center of a distribution and 0, and therefore varies with location. The distinction between the two concepts is just now starting to be made.

In answer to those two criticisms towards robustness, Bickel (1975) adapted an old idea of the statistical literature to robust M-estimation. Le Cam (1956), Neyman (1949) (see [22]) and Fisher (1922), in the early years of modern statistics, had observed that in the univariate estimation of location setup, “if F is known, and $\psi = (-f'/f)$, the estimate obtained by starting with a \sqrt{n} consistent estimate of θ and performing one Gauss-Newton iteration of (1.1) is asymptotically efficient even when the MLE is not and is equivalent to it when it is” (see [5], p. 428). In times when computers were still a dream, less effort to compute an estimator, associated with no apparent loss in its asymptotical properties, represented a major advantage.

Inspired by the observation of Fisher, Neyman and Le Cam, Bickel (1975) proposed a one-step Huber M-estimator to be used in the estimation of simple linear models, such as location and regression through the origin. The author showed that the estimator is asymptotically normal under mild conditions. In his book, Huber (1981) gave explicit expressions for one-step M-estimators of location with unknown dispersion (p. 140 and p. 146), as the first step of Newton’s method, starting with preliminary robust estimates of location and dispersion. The author further showed that if the initial estimate of location was consistent for θ , F was symmetric and the score function defining the estimator was odd, then this one-step M-estimator of location was asymptotically equivalent to the fully-iterated M-estimator of location with preliminary dispersion. In the special case of dispersion estimation with unknown location, Huber (1981) (p. 147) also suggested a fixed-point iterative method for computing the estimate, the first step of which may serve as a one-step M-estimator of dispersion. Hampel et al (1986) (p. 106) further stressed the importance of selecting robust preliminary estimates when computing one-step M-estimators, as otherwise the resulting estimators may not be robust (which was also observed by Andrews et al in [1]).

In all cases, the enthusiasm created by one-step estimators was contagious: those estimators were easy to compute, but most importantly, they represented a good balance between robustness and efficiency. Being one step away from robust initial estimators, they retained the robustness properties of their initial estimators. Simultaneously, one step closer to their fully-iterated version, the one-step estimators were almost as efficient as their fully-iterated version. In the univariate location set up, Andrews et al (1972) (Bickel (1975) has submitted his paper in 1971 and was one of the authors of [1]) showed with some length that for certain score functions, one-step M-estimators were very well behaved. Jurečková and Portnoy (1987) further investigated the use of one-step regression M-estimators based on the LMS to obtain high efficiency.

While one-step M-estimators solve the possible lack of uniqueness of the solution of (1.1) and reduce the computational effort, they can sometimes be numerically unstable. For instance, the standard one-step M-estimator of location with unknown dispersion contains a ratio which denominator can become very small for certain samples and score functions. To address this problem, Hampel et al (1986) suggested different versions of one-step M-estimators of location (pp. 152-153), one of which replaces the denominator by a constant. Similarly, the following example (reproduced from [7], p. 10) illustrates that the fixed-point iterative method can sometimes lead to disastrous results, which raises the question whether the one-step M-estimator of dispersion with unknown location, as introduced by Huber (1981), is stable enough to be trusted.

Example. *The value $x = 1/n$ is a fixed point of the function given by*

$$f(x) = (n + 1)x - 1$$

since

$$f\left(\frac{1}{n}\right) = \frac{n+1}{n} - 1 = \frac{1}{n}.$$

It follows then that iterating this function for any particular value of n using the following loop should, if floating-point arithmetic of the computer were exact, simply result in $x = 1/n$:

```

x := 1/n
for i = 1 to 30
  x := (n + 1) * x - 1

```

We use here the symbol $:=$ for computer assignment, so ' $x := 1/n$ ', means 'Assign the value $1/n$ to the stored variable x '.

Table 1.1 shows the results of implementing such a loop in Turbo Pascal for different values of n . We see that for the various powers of 2 the arithmetic is indeed exact. By contrast, for other values of n the error grows steadily from approximately -3.5×10^5 for $n = 3$ through about -2.3×10^{18} for $n = 10$.

These errors are the direct result of the propagation of the rounding error made in the binary representation of $\frac{1}{3}$ and $\frac{1}{10}$, respectively. \square

Bickel (1975) had most likely foreseen the numerical instability problem. He actually proposed two types of one-step Huber M-estimators of regression. His Type II estimator was the smooth version of his Type I, of which a term was replaced by its asymptotic expectation. As much as robustness is concerned with outliers, it must never ignore important issues such as numerical stability.

With the development of modern techniques, such as influence functions and maximum bias curves, which allow a more complete study of one-step M-estimators, these

<i>n</i>	Final <i>x</i>
1	1.000 000 000 000 00 E + 0000
2	5.000 000 000 000 00 E - 0001
3	1.747 630 000 000 00 E + 0005
4	2.500 000 000 000 00 E - 0001
5	-4.021 311 173 693 75 E + 0010
6	-1.952 324 816 734 00 E + 0012
7	-4.021 071 095 865 60 E + 0013
8	1.250 000 000 000 00 E - 0001
9	-1.616 879 469 807 53 E + 0017
10	-2.308 383 841 816 94 E + 0018
11	-1.962 659 088 425 60 E + 0019
12	-2.443 971 442 609 19 E + 0020
13	-1.935 039 516 698 14 E + 0021
14	-1.328 735 789 941 45 E + 0022
15	7.052 067 281 085 79 E + 0022
16	6.250 000 000 000 00 E - 0002

Table 1.1: Results of the Implementation of the Loop in Turbo Pascal

estimators can be better understood and appreciated. Maximum bias curves occurred briefly in Hampel et al (1986) (pp. 176-177), but were used to their full advantage by Martin and Zamar (1989), Martin et al (1989) and He and Simpson (1993). The influence function as a tool was itself developed by Hampel starting in 1974 and carefully discussed in [15]. Using those techniques, Rousseeuw and Croux (1993a) have studied the bias properties of k -step Huber M -estimators (the k th step of the iterative algorithm) in the univariate location and dispersion setup. They have shown that while the efficiency increases with the number of steps, the bias also increases. This led the authors to recommend the use of one or two-step M -estimators, but not more, especially in multiparameter problems where the dispersion is typically unknown.

Rousseeuw and Croux (1993a) preferred the one-step M -estimator of location suggested by Hampel et al (1986) (p. 153), which is derived from the standard one-step M -estimator, but has a constant denominator in its ratio. To estimate univariate dispersion, the authors used the fixed-point method one-step M -estimator suggested by Huber (1981) (p. 147). However, to our knowledge, no formal comparison have been done of all the one-step M -estimators available, using modern techniques, which would allow one to choose one estimator over another depending on the statistical situation at hand.

The goal of this thesis is to offer such a comparison, in two contexts: estimation of location with unknown dispersion, and estimation of dispersion with unknown location. The standard one-step M -estimator of location, as suggested by Huber (1981) (p. 146), will be compared to the one-step M -estimator used by Rousseeuw and Croux (1993a) in the general setup of unknown dispersion. To estimate dispersion with unknown location, two estimators are derived using Newton-Raphson's method, and compared to the standard, unchallenged, fixed-point one-step M -estimator of dispersion, used for example by Rousseeuw and Croux (1993a).

Chapter 2 presents the general theory of M-estimators as initiated by Huber (1964). Chapter 3 presents formally the one-step M-estimators that will be compared in the following two chapters. In Chapter 4, we study the asymptotic properties of the two one-step M-estimators of location that are of interest; that is, we study their efficiency under Gaussian and non-Gaussian models (following the approach of Rousseeuw and Croux (1993b)), as well as their maximum asymptotic bias. Using the same robustness techniques, the three one-step M-estimators of dispersion with unknown location that are of interest are compared in Chapter 5 in terms of their asymptotic behaviour.

Chapter 2

M-Estimators: Maximum Likelihood Type of Estimators

2.1 Univariate Problem

Let X_1, \dots, X_n be a sample from the population with distribution function $F(x; \theta)$.

Any estimate T_n which minimizes an equation of the form

$$\sum_{i=1}^n \rho(x_i; T_n),$$

or which is defined by the implicit equation

$$\sum_{i=1}^n \psi(x_i; T_n) = 0, \tag{2.2}$$

where ρ is an appropriate loss function and $\psi(x; \theta) = (\partial/\partial\theta)\rho(x; \theta)$ is a score function, is called an M-estimator of θ . This estimator is an extension of the usual maximum likelihood estimator (MLE) of θ : the choice $\rho(x; \theta) = -\log f(x; \theta)$ exactly corresponds to the MLE minimization problem.

For example, when the parameter θ to be estimated is a location parameter, one sets $\psi(x; T_n) = \psi(x - T_n)$. When the parameter σ to be estimated is a dispersion parameter, the function ψ used is $\psi(x; \sigma) = \psi(\frac{x}{\sigma})$.

Note that the M-estimator is not modified when ψ is multiplied by any constant $r > 0$.

2.1.1 Qualitative Robustness

It is easy to study the asymptotic properties of M-estimators when $\psi(x; \theta)$ is monotone in θ . Furthermore, assume

- $\psi(x; \theta)$ is non-increasing in θ ;
- $\psi(x; \theta)$ is measurable in x ;
- $\psi(x; -\infty) > 0$ and $\psi(x; \infty) < 0$;

and define

- $\lambda_F(t) = E_F\{\psi(x; t)\}$;
- $T_n^* = \sup\{t : \sum_{i=1}^n \psi(x_i; t) > 0\}$;
- $T_n^{**} = \inf\{t : \sum_{i=1}^n \psi(x_i; t) < 0\}$.

Consistency

M-estimators are consistent under some conditions. Huber (1981) has shown that if there is a $t_0(F)$ such that

$$\begin{cases} \lambda_F(t) > 0 & \forall t < t_0(F) \\ \lambda_F(t) < 0 & \forall t > t_0(F) \end{cases}$$

then $T_n^* \rightarrow t_0(F)$ and $T_n^{**} \rightarrow t_0(F)$ almost surely $[F]$ and in probability $[F]$. Any value T_n satisfying $T_n^* \leq T_n \leq T_n^{**}$ can serve an M-estimator, which will be consistent since asymptotically, $T_n^* = T_n^{**}$.

Asymptotic Distribution

Assume furthermore that

- there exists at least one $t_0(F)$ such that $\lambda_F(t_0) = 0$;
- $\lambda_F(t)$ is continuously differentiable in a neighborhood of $t_0(F)$ and $\lambda'_F(t_0) < 0$; and
- $\sigma_F^2(t) = E_F \psi^2(x; t) - \lambda_F^2(t)$ is finite, non zero and continuous near $t_0(F)$.

Under those assumptions, M-estimators T_n are asymptotically normally distributed, that is, $\sqrt{n}(T_n - t_0) \xrightarrow{D} N(0, V(\psi, F))$, where $V(\psi, F) = \frac{\sigma_F^2(t_0)}{[\lambda'_F(t_0)]^2}$. See Huber (1981) (pp. 49-50) for the complete details of this proof.

The fact that M-estimators are asymptotically consistent and normally distributed have certainly contributed to their success, since it becomes so easy to make inference under these conditions.

In most cases, T_n is a function of the empirical distribution F_n and derives from a functional T , that is, $T_n = T(F_n)$. If T_n is consistent, then $T_n \rightarrow T(F)$ in probability $[F]$. The discussion that follows will adopt the functional notation. An M-estimator T_n will therefore be referred to as $T(F)$.

2.1.2 Infinitesimal Aspects

The Influence Function

Beyond qualitative properties, it is informative to study the behaviour of M-estimators under infinitesimal changes. In 1968, Hampel introduced *the influence function*, which “describes the effect of an infinitesimal contamination at the point x on the estimate, standardized by the mass of the contamination. One could say it gives a picture of the infinitesimal behavior of the asymptotic value, so it measures the asymptotic bias

caused by contamination in the observations.” (see [15], p. 84) When Hampel first introduced this concept (1968, 1974), he referred to it as the influence curve; however, the term influence function is now widely preferred in view of the generalizations to higher dimensions. More precisely, the influence function of the functional $T(F)$ is

$$IF(x; T, F) = \lim_{t \rightarrow 0} \frac{T(F_{t,x}) - T(F)}{t},$$

where $F_{t,x} = (1 - t)F + t\delta_x$. In other words, it is defined as the derivative with respect to t of the functional $T(F_{t,x})$, evaluated at $t = 0$.

Under the regularity assumptions listed in the previous section, it is easy to show that the influence function of any univariate M-estimator, defined by the score function ψ , has the following form:

$$IF(x; T, F) = \frac{\psi(x; T(F))}{-E_F\{(\partial/\partial\theta)\psi(x; T(F))\}}. \quad (2.3)$$

See Huber (1981) (p. 45) for more details on this derivation. Notice that the influence function of an M-estimator is proportional to ψ .

In the special case of a univariate location problem, the score function defining the M-estimator is $\psi(x; \theta) = \psi(x - \theta)$, and we obtain

$$IF(x; T, F) = \frac{\psi(x - T(F))}{E_F\psi'(x - T(F))}.$$

Similarly, the influence function of a univariate dispersion M-estimator $S(F)$, for which the score function has the form $\chi(x; \sigma) = \chi(\frac{x}{\sigma})$, will be

$$IF(x; S, F) = \frac{\chi(x/S(F))S(F)}{E_F\chi'(x/S(F))(x/S(F))}.$$

Hampel et al (1986) have shown that under regularity conditions, the knowledge of the influence function of an M-estimator is equivalent to the knowledge of its asymptotic

distribution. Indeed, using a Taylor expansion of $T_n(F_n)$ around $T(F)$, the authors showed that the asymptotic variance of an estimator is completely defined by its influence function, since

$$V(T, F) = E_F\{IF(x; T, F)\}^2. \quad (2.4)$$

The influence function is a priori a heuristic tool. It is easy to calculate, and therefore to obtain an expression for the asymptotic variance of an estimator. However, the regularity assumptions for (2.4) are cumbersome to prove; one usually tries to prove normality using another method. However, in all practical cases, the relation (2.4) holds. The influence function approach will be used in the next chapters when deriving an expression for the asymptotic variance of the estimators of interest.

The Gross-Error Sensitivity

The influence function of an estimator can be summarized in many ways other than its expected square. The most important is probably the supremum of its absolute value. Hampel (1968,1974) introduced this notion as the *gross-error sensitivity* of an estimator. More precisely, one defines the gross-error sensitivity of T at F by

$$\gamma^*(T, F) = \sup_x |IF(x; T, F)|,$$

for values of x where the influence function exists. The gross-error sensitivity measures the worst influence which a small amount of contamination of fixed size can have on the value of the estimator. It could therefore be regarded as an upper bound on the standardized asymptotic bias of the estimator.

If $\gamma^*(T, F)$ is finite, we say that T is *B-robust at F* , where the B comes from bias. In view of (2.3), an M-estimator is B-robust at F if and only if $\psi(\cdot, T(F))$ is bounded.

Putting a bound on $\gamma^*(T, F)$ is often the first step in robustifying an estimator. However, in many cases, this will conflict with the goal of asymptotic efficiency. The introduction of one-step M-estimators brings a partial solution to this problem, as will be seen in the next chapter.

2.1.3 Quantitative Robustness

The influence function represents an excellent tool for assessing *local* asymptotic behaviour of an estimator. However, it must be complemented by a measure of *global* reliability, which describes up to what distance from the model distribution the estimator still gives some relevant information. Consider the ϵ -contaminated neighborhood of F_0 :

$$V_\epsilon(F_0) = \{F : F = (1 - \epsilon)F_0 + \epsilon H\},$$

where H is an arbitrary distribution and $0 \leq \epsilon < 1/2$ so it is possible to distinguish between the central model F_0 and the contaminating distribution H . In what follows, we will present two measures of distance from the central model F_0 that should stay as small as possible, and for ϵ as big as possible, for an estimator to be considered robust.

The Breakdown Point

Hampel (1971) introduced the notion of *breakdown point* of an estimator, generalizing a definition by Hodges (1967). Roughly speaking, the breakdown point of an estimator is the maximum ϵ -contamination that an estimator can endure before its value goes to infinity. It gives the limiting fraction of bad outliers the estimator can cope with.

For any estimator, the maximum breakdown point is 50%. Robust M-estimators such as the median or the median absolute deviation (MAD) have 50% breakdown point. On

the other hand, the mean, as an estimator of location, has 0% breakdown point - it is completely intolerant to outliers. In fact, Hampel (1971) and Huber (1981) (section 3.2) have shown that when F is symmetric, one usually chooses odd score functions ψ to estimate location. If ψ is moreover strictly monotone and bounded, the M-estimator defined by ψ is B-robust and has 50% breakdown point. On the other hand, M-estimators are not B-robust and have 0% breakdown point when ψ is strictly monotone but unbounded.

There exists an asymptotic version of the breakdown point, as well as a finite-sample version. In what follows, to accompany the functional (limiting) notation, we will use the asymptotic version, as in Hampel (1971) and Huber (1981). See [9], [10], [17] and [20] for a detailed presentation of the finite-sample version of the breakdown point.

The Maximum Bias Function

A high breakdown point is a necessary condition for a good estimating method, but not a sufficient condition (see [29], p. 877). Many argue that the breakdown point is not as general as it claims it is. To deal with the matter, Hampel et al (1986) (pp. 176-177) briefly introduced maximum bias functions, but this measure of robustness was fully exploited by Martin and Zamar (1989), Martin et al (1989) and He and Simpson (1993). The maximum bias of an estimator (as a function of ϵ) describes how an estimator $T(F)$ can change in $V_\epsilon(F_0)$ due to a given fraction ϵ of contamination.

In the univariate location setup, the maximum bias function of an estimator $T(F)$ is formally defined as

$$B_T(\epsilon) = \sup_{F \in V_\epsilon(F_0)} |(T(F) - T(F_0))|.$$

However, in the univariate dispersion setup, one needs to generalize the concept of maximum bias function. Martin and Zamar (1989) have observed that the presence of

outliers can cause the estimate of dispersion to be very large, but as well, too many inliers may cause the estimate to be dangerously close to 0. They defined the generalized maximum bias as the maximum between the bias due to outliers and the bias due to inliers. Moreover, they allowed (monotone) penalization for inliers and outliers to be independently chosen, because in some setups, one of the two may cause more trouble than the other.

Formally, the generalized maximum bias can be defined by

$$B_S(\epsilon) = \max_{F \in V_\epsilon(F_0)} B[S(F)],$$

where

$$B[S(F)] = \begin{cases} L_1[S(F)/S(F_0)], & \text{if } 0 < S(F) \leq S(F_0), \\ L_2[S(F)/S(F_0)], & \text{if } S(F_0) < S(F) \leq \infty, \end{cases}$$

where L_1 and L_2 are the continuous, nonnegative and monotone (penalization) loss functions, with $L_1(1) = L_2(1) = 0$ and

$$\lim_{t \rightarrow 0} L_1(t) = \lim_{t \rightarrow \infty} L_2(t) = \infty.$$

A popular choice for a loss function is the logarithmic function. From monotonicity of L_1 and L_2 , it follows that

$$B_S(\epsilon) = \max\{L_1[S^-/S(F_0)], L_2[S^+/S(F_0)]\},$$

where S^- and S^+ denote the supremum and the infimum of the functional $S(F)$ as F ranges over $V_\epsilon(F_0)$. Therefore, it is enough to concentrate on S^- and S^+ when studying the generalized maximum bias of an estimator. We shall call $S^+/S(F_0)$ the *explosion maximum bias* of an estimator, and $S^-/S(F_0)$ the *implosion maximum bias* of an estimator, both being functions of ϵ .

Finally, note that the dependence of the explosion and implosion maximum bias functions of an estimator on the ratio $S(F)/S(F_0)$ is justified in terms of dispersion-invariance (see [24], p. 134).

The maximum bias function of an estimator includes its gross-error sensitivity (the slope of the curve at $\epsilon = 0$), and also the breakdown point, where the curve goes to infinity. The maximal bias curve is therefore an additional, and more complete, tool that can help to choose between competing estimators. For a small range of ϵ , some use the gross-error sensitivity of an estimator as a linear approximation to the curve (check Hampel et al (1986) for a rule of thumb on values of ϵ for which the approximation hold).

2.2 Nuisance Parameter in the Location-Dispersion Problem

In many cases, the underlying distribution of the population from which a sample is taken has more than one parameter. For example, the location-dispersion families have typically distributions of the form $F(\frac{x-\theta}{\sigma})$, where $-\infty < \theta < \infty, \sigma > 0$. If the interest is still to estimate only one parameter, then the other becomes a nuisance parameter.

It must also be observed that typical M-estimators of location are in practice location invariant, but not dispersion invariant. Therefore, when estimating location, one needs to provide an M-estimator of dispersion as well. The unknown dispersion is, per say, a nuisance parameter. The same can be said about dispersion M-estimators: they are not location invariant, and thus need an ancillary location estimator.

If ψ and χ are the respective score functions of the M-estimators of location and dispersion, then the simultaneous equations defining them implicitly are

$$\frac{1}{n} \sum_{i=1}^n \psi \left(\frac{x_i - T_n}{S_n} \right) = 0, \quad \text{and} \quad (2.5)$$

$$\frac{1}{n} \sum_{i=1}^n \chi \left(\frac{x_i - T_n}{S_n} \right) = 0. \quad (2.6)$$

From a finite-sample point of view, when estimating location with unknown dispersion, one would compute some M-estimate of dispersion, S_n , and then find the location estimate T_n using (2.5). If, on the other hand, one is interested in estimating dispersion when the location is unknown, then one would solve (2.6) for S_n , with some preliminary estimate of location T_n . In functional terms, the M-estimator of location $T(F)$ with unknown dispersion, and the M-estimator of dispersion $S(F)$ with unknown location can be expressed implicitly by

$$E_F \psi \left(\frac{x - T(F)}{S_0(F)} \right) = 0, \quad \text{and}$$

$$E_F \chi \left(\frac{x - T_0(F)}{S(F)} \right) = 0,$$

where $S_0(F)$ and $T_0(F)$ are the asymptotic values of the initial estimator used for the nuisance parameter.

Fortunately, if the underlying distribution F is symmetric, ψ is odd and χ is even, the location functional $T(F)$ and the dispersion functional $S(F)$ are independent: the asymptotic variance of T depends on S only through its asymptotic value $S(F)$, and that of $S(F)$ does not depend on $T(F)$. More precisely, the influence functions of T and S , respectively $IF(x; T, F)$ and $IF(x; S, F)$ are:

$$IF(x; F, T) = \frac{\psi\left(\frac{x}{S(F)}\right)S(F)}{E_F \psi'\left(\frac{x}{S(F)}\right)}, \quad \text{and}$$

$$IF(x; F, S) = \frac{\chi\left(\frac{x}{S(F)}\right)S(F)}{E_F \chi'\left(\frac{x}{S(F)}\right)\left(\frac{x}{S(F)}\right)}.$$

(Remember that $T(F) = 0$ when F is symmetric.)

Under similar regularity assumptions as listed in the preceding section, the M-estimators of location and dispersion defined by (2.5) and (2.6) have asymptotic normal distributions with variances equal to the expected squares of their influence functions. Therefore, when estimating location with unknown dispersion, one can choose the estimator of dispersion on criteria other than low variability, which enlarge considerably the set of possible candidates. Similarly, the asymptotic variability of an M-estimator of dispersion will not be affected by the estimate of the unknown location parameter, no matter which estimator is chosen.

The maximum bias of an M-estimator of location with unknown dispersion is defined exactly as in the univariate setup. However, one must remember that T is defined implicitly as a function of S , which necessarily causes the bias of T to be affected by the bias of S in an ϵ -contaminated neighborhood such as $V_\epsilon(F_0)$. Clearly, the maximum bias of T will be higher in presence of a nuisance parameter than with no nuisance parameter.

Similarly, the maximum explosion bias of an M-estimator of dispersion with unknown location, defined as in the univariate setup, will be implicitly affected by the estimator of location, and increased accordingly. On the other hand, the maximum implosion bias is not affected by the estimate of the unknown location, compared to the implosion bias of the univariate estimator of dispersion. This is due to the fact that when $S \rightarrow 0$, the location estimate tends to the fixed value $T(F_0)$.

Solving the non-linear equations (2.5) and (2.6) requires a numerical method, which may not be numerically feasible in some situations. One-step M-estimators have been proposed to solve this problem. The next chapter presents these estimators in the context of location and dispersion estimation with a nuisance parameter.

Chapter 3

One-Step M-Estimators

3.1 General Idea of One-Step M-Estimators

The computation of an M-estimator requires an iterative algorithm, since the estimator is defined implicitly by a non-linear equation. Moreover, the presence of a nuisance parameter adds an extra dimension to the problem, being itself the implicit solution of another non-linear equation. Indeed, in multiparameter problems, you can never be sure that a root exists until you find it: there is no bracketing-the-root-by-fancy-algorithms possible. The computation involved often proves to be difficult, as any iterative method can be subject to many problems. For example, the Newton-Raphson algorithm, often used in practice, will diverge if it encounters a local extremum in its search of a root (see [28] p. 362 for a graphical explanation of this problem).

Bickel (1975) suggested, as an alternative to the fully-iterated M-estimator, what he referred to as the *one-step M-estimator*. This estimator corresponds to the first iteration solution of the algorithm used to solve the non-linear equation defining the estimator, as Huber (1981) (section 6.7) further investigated.

In the case where the objective is to estimate the parameter θ of a distribution in the presence of a nuisance parameter, one would first choose preliminary estimates for both parameters. The nuisance parameter would then be considered fixed and equal to its preliminary estimate. To get the one-step M-estimator of θ , one would simply perform one iteration in the algorithm used to solve for the fully iterated M-estimator, starting

from the preliminary estimate of θ and considering the nuisance parameter as fixed.

3.2 Estimation of Location With Unknown Dispersion

Huber (1981) suggested to use the well-known Newton-Raphson method to solve the non-linear equation defining an M-estimator. The one-step M-estimator is therefore obtained by performing one step in the Newton-Raphson method.

Thus, in order to estimate the location parameter of the location-dispersion distribution F , with unknown dispersion, one would choose initial estimates of location and dispersion, with respective asymptotic values $T_0(F)$ and $S_0(F)$. Since the Taylor expansion of

$$E_F \psi \left(\frac{x - T(F)}{S_0(F)} \right) = 0 \quad (3.7)$$

with respect to $T(F)$, around $T_0(F)$, begins with

$$E_F \psi \left(\frac{x - T(F)}{S_0(F)} \right) = E_F \psi \left(\frac{x - T_0(F)}{S_0(F)} \right) - \frac{T(F) - T_0(F)}{S_0(F)} E_F \psi' \left(\frac{x - T_0(F)}{S_0(F)} \right) + \dots,$$

the one-step M-estimator of location is defined, in functional terms, by

$$T(F) = T_0(F) + S_0(F) \frac{E_F \psi \left(\frac{x - T_0(F)}{S_0(F)} \right)}{E_F \psi' \left(\frac{x - T_0(F)}{S_0(F)} \right)}. \quad (3.8)$$

This estimator will be hereafter called the *Standard One-Step M-Estimator of Location*.

The functional notation is used here for simplicity. But in practice, we always deal with finite sample sizes. The finite-sample standard one-step M-estimator of location is given by

$$T^n = T_0^n + S_0^n \frac{\frac{1}{n} \sum_{i=1}^n \psi \left(\frac{x_i - T_0^n}{S_0^n} \right)}{\frac{1}{n} \sum_{i=1}^n \psi' \left(\frac{x_i - T_0^n}{S_0^n} \right)}, \quad (3.9)$$

where T_0^n and S_0^n are the initial finite-sample estimates of location and dispersion, and n is the sample size.

Huber (1981) (p. 146) indicates that (3.9) converges to the exact value of the fully iterated M-estimator in a finite number of steps, provided the underlying distribution F is symmetric and ψ is skew symmetric and piecewise linear. Of course, in practice, we deal with empirical cumulative density functions F_n which are never symmetric. But we assume here that those cumulative density functions converge to symmetric distribution functions. The value of the denominator in the ratio defining the estimator (3.8) is not critical, and can be replaced by any constant greater than $1/2$ as long as $0 \leq \psi' \leq 1$.

However, as Hampel et al (1986) (p. 153) pointed out, for certain score functions ψ (especially the redescending types) and certain samples, it may happen that the denominator $\frac{1}{n} \sum_{i=1}^n \psi'(\frac{x_i - T_0^n}{S_0^n})$ becomes extremely small (approximately equal to 0), which would destabilize the value of the estimator. To avoid this problem, Hampel et al (1986) have suggested replacing the denominator in the ratio by the constant $E_{\Phi} \psi'(z)$, where Φ is the standard normal distribution. This constant is easy to calculate and never becomes 0. It can also be seen as the smooth version of $\frac{1}{n} \sum_{i=1}^n \psi'(\frac{x_i - T_0^n}{S_0^n})$ when the underlying distribution F is normal. The normality assumption is often used in practice, but nothing prevents one from assuming F is otherwise, and calculating the expectation with respect to that different underlying distribution. The modified version of the one-step M-estimator (3.8) is therefore given by:

$$T^*(F) = T_0(F) + S_0(F) \frac{E_F \psi(\frac{x - T_0(F)}{S_0(F)})}{E_{\Phi} \psi'(z)}.$$

This estimator will hereafter be called the *MOSME of Location*, for **Modified One-Step M-Estimator of Location**. Its finite-sample version can be expressed as

$$(T^*)^n = T_0^n + S_0^n \frac{\frac{1}{n} \sum_{i=1}^n \psi(\frac{x_i - T_0^n}{S_0^n})}{E_{\Phi} \psi'(z)},$$

where T_0^n and S_0^n are the initial finite-sample estimates of location and dispersion, and n is the sample size.

Hampel et al (1986) (p. 153) suggest the use of another estimator, the one-step W-estimator, to avoid the problem of a small denominator. This thesis will however focus on the MOSME and the standard one-step M-estimator of location, as they are more commonly used in practice.

3.3 Estimation of Dispersion With Unknown Location

The non-linear equation defining the fully iterated M-estimator of dispersion $S(F)$ with unknown location,

$$E_F \chi \left(\frac{x - T_0(F)}{S(F)} \right) = \beta, \quad (3.10)$$

slightly differs from that of the M-estimator of location with unknown dispersion, in that the right-hand side constant, $\beta = E_{\Phi} \chi(x)$, is greater than 0. It would be undesirable to have $\beta = 0$ since this would force $S(F)$ to be equal to ∞ (assuming $\chi(0) = 0$).

Taking advantage of the form of the expression (3.10), Huber (1981) (p. 147) suggested the use of the fixed-point iterative method to solve it for $S(F)$. Actually, Huber used a particular case of (3.10), the one for which $\chi = \psi'$, where ψ was the score function defining the fully iterated M-estimator of the unknown location. That is, Huber wanted to make the location estimator defined by ψ dispersion invariant. The fixed-point method can however be straightforwardly generalized for any χ .

The one-step M-estimator of dispersion, $S_1^p(F)$, derived from the fixed-point computational algorithm can be expressed as

$$[S_1^p(F)]^2 = \frac{[S_0(F)]^2}{\beta} E_F \chi \left(\frac{x - T_0(F)}{S_0(F)} \right), \quad (3.11)$$

where $T_0(F)$ and $S_0(F)$ are the limiting value of the preliminary estimators of location and dispersion. Note that the standard fixed-point method solves an equation of the form $g(S(F)) = S(F)$, for a certain function g . The reason for solving instead $S(F)g(S(F)) = S(F)^2$ is due to the fact that when $x \rightarrow T_0(F)$, the χ function behaves generally like a second-order polynomial, and therefore $S(F)g(S(F)) = S(F)^2 E_F \chi \left(\frac{x - T_0(F)}{S(F)} \right) \simeq E_F (x - T_0(F))^2$. The latter expression refers directly to the expression for the standard deviation as an estimator of dispersion.

The finite-sample version of (3.11) can be expressed as

$$[(S_1^p)^n]^2 = \frac{[S_0^n]^2}{\beta} \frac{1}{n} \sum_{i=1}^n \chi \left(\frac{x_i - T_0^n}{S_0^n} \right),$$

where T_0^n and S_0^n are the initial finite-sample estimates of location and dispersion, and n is the sample size.

Yohai and Zamar in [34] (p. 407) have obtained τ -estimators of dispersion by letting $S_0(F)$ in (3.11) be an M-estimator of dispersion determined by a smooth function ρ_1 . The estimator $S_1^p(F)$, defined by (3.11), will therefore be hereafter called *One-Step τ -Estimator of Dispersion*.

The fixed-point method for solving (3.10) is not the only one possible, of course. The Newton-Raphson method, as used in the setup of location estimation with unknown dispersion, can also be used. In fact, the fixed-point method is known to converge linearly, whereas the Newton-Raphson method, as a special case with an additional constraint, converges quadratically (see [7], p. 62). To my knowledge, the Newton-Raphson method in the estimation of dispersion has never been fully studied, but it represents clear advantages over the fixed-point method in terms of asymptotic behaviour of the estimator (refer

to chapter 5 for more details). We therefore believe it is worth of some consideration.

Thus, in order to estimate the dispersion parameter of the location-dispersion distribution F , with unknown location, one would choose initial estimates of location and dispersion, respectively referred to as $T_0(F)$ and $S_0(F)$. Since the Taylor expansion of the right-hand side of (3.10), with respect to $S(F)$, around $S_0(F)$, begins with

$$E_F \chi \left(\frac{x-T_0(F)}{S(F)} \right) = E_F \chi \left(\frac{x-T_0(F)}{S_0(F)} \right) - \frac{S(F)-S_0(F)}{S_0(F)} E_F \chi' \left(\frac{x-T_0(F)}{S_0(F)} \right) \left(\frac{x-T_0(F)}{S_0(F)} \right) + \dots$$

the *Standard One-Step M-Estimator of Dispersion* can be defined in functional terms by

$$S_1(F) = S_0(F) + S_0(F) \frac{E_F \chi \left(\frac{x-T_0(F)}{S_0(F)} \right) - \beta}{E_F \chi' \left(\frac{x-T_0(F)}{S_0(F)} \right) \left(\frac{x-T_0(F)}{S_0(F)} \right)}. \quad (3.12)$$

The finite-sample version of the standard one-step M-estimator of dispersion is

$$S_1^n = S_0^n + S_0^n \frac{\frac{1}{n} \sum_{i=1}^n \chi \left(\frac{x_i - T_0^n}{S_0^n} \right) - \beta}{\frac{1}{n} \sum_{i=1}^n \chi' \left(\frac{x_i - T_0^n}{S_0^n} \right) \left(\frac{x_i - T_0^n}{S_0^n} \right)}. \quad (3.13)$$

where T_0^n and S_0^n are the initial finite-sample estimates of location and dispersion, and n is the sample size.

For the same reasons explained in the location setup, it may happen, for certain χ functions and certain samples, that the denominator in the ratio of (3.13) gets dangerously close to 0 and strongly affects the estimate. In this case, the *MOSME of Dispersion* may represent a better choice of estimator, where the MOSME of dispersion is asymptotically defined as

$$S^*(F) = S_0(F) + S_0(F) \frac{E_F \chi \left(\frac{x-T_0(F)}{S_0(F)} \right) - \beta}{E_\Phi \chi'(z) z}. \quad (3.14)$$

Its finite-sample definition is

$$(S^*)^n = S_0^n + S_0^n \frac{\frac{1}{n} \sum_{i=1}^n \chi\left(\frac{x_i - T_0^n}{S_0^n}\right) - \beta}{E_{\Phi} \chi'(z) z},$$

where T_0^n and S_0^n are the initial finite-sample estimates of location and dispersion, and n is the sample size.

3.4 The First Step Is a Big Step

The obvious advantage of one-step M-estimators is their ease of computation, compared to their fully iterated versions. Moreover, the asymptotic properties of one-step M-estimators make them very attractive. For example, it is known that the standard one-step M-estimator of location is asymptotically equivalent to the fully iterated M-estimator, provided the underlying distribution F is symmetric, ψ is odd and most of all, T_0 is consistent, translation invariant and odd. T_0 is said to be translation invariant and odd if

$$T_0(F_{X+c}) = T_0(F_X) + c,$$

and

$$T_0(F_{-X}) = -T_0(F_X).$$

In this thesis, we will focus on the following robustness properties of one-step M-estimators: their breakdown point (the fraction of contamination causing complete disaster), their influence function (which describes the effect of an infinitesimal contamination) and their worst-case bias (which describes how much the estimators can change with a fraction ϵ of contamination).

It can be shown (see Rousseeuw and Croux in [31]) that the breakdown point of one-step M-estimators is equal to that of the initial M-estimators used in their computation.

Selecting very robust initial M-estimators is thus strongly recommended. For example, the median and the median absolute deviation (MAD), as estimates of location and dispersion respectively, have 50% breakdown point, the highest possible. This makes them excellent candidates for preliminary estimates of location and dispersion of a distribution.

On the other hand, many estimators with a high breakdown point have very low asymptotic efficiency (63.7% for the median in the univariate location problem, and 36.7% for the MAD in the univariate scale model). However, one-step M-estimators, while inheriting the breakdown point of their initial estimators, will generally show an improvement over the asymptotic efficiencies of their initial estimators, because they approach the fully iterated M-estimators, which are usually by definition more efficient.

Unfortunately, as Rousseeuw and Croux (1993a) have shown, increasing efficiency does not come without compromise. The authors treat the problems of univariate location estimation and univariate dispersion estimation, that is, estimation of location and dispersion when the nuisance parameter is known. In the former, they use the univariate equivalent of our MOSME of location, while they use the univariate equivalent of our τ -estimator of dispersion in the latter. Rousseeuw and Croux (1993a) show that the maximum bias of these one-step M-estimators is higher than that of their initial estimators. The increase in bias is especially strong in the dispersion setup. However, because only one iteration in the computational algorithm is performed when deriving one-step M-estimators, the increase in worst-case bias is perceived as a good compromise for the increase in efficiency.

Rousseeuw and Croux (1993a) treat in general k -step M-estimators, for $k \geq 1$. They show that it is possible to obtain an arbitrarily high efficiency, while maintaining a 50% breakdown point, through a k -step M-estimator, for a fixed finite k (which corresponds to performing k steps in the algorithm used to solve for the fully-iterated M-estimator,

starting from a preliminary estimator with 50% breakdown point). Notice that under the Gaussian model, the univariate one-step M-estimator of location has already the same asymptotic efficiency as the fully-iterated M-estimator. Therefore, taking further steps will not increase the efficiency of the k -step M-estimator: Rousseeuw and Croux (1993a) show that it is equal to the efficiency of the fully-iterated M-estimator of location, for any k . The authors nevertheless consider k -step M-estimators of location for $k > 1$ in view of possible improvements on their quantitative robustness properties. However, Rousseeuw and Croux (1993a) show that as the efficiency of the estimator goes up (as k increases), its maximum bias will also go up. In the univariate location case, the bias increases only slightly with k . Unfortunately, in the univariate dispersion problem, the maximum bias explodes rapidly with k , which does not justify, in the authors' opinions, the increase in efficiency. Based on these findings, they believe small values of k ($k = 1$ or $k = 2$) are preferable, especially for multiparameter problems (which typically contain a dispersion component). This has driven the choice to focus this thesis on one-step M-estimators.

The purpose of this thesis is to extend the work by Rousseeuw and Croux (1993a). It will consider the more realistic situations with nuisance parameters: estimation of location with unknown dispersion and estimation of dispersion with unknown location. The emphasis will be put on the MOSME's presented earlier in this chapter for the above models, because they have never been studied to the extent they deserve. To follow Rousseeuw and Croux's steps, we will derive the maximum bias of the MOSME for the two models of interest, and compare them to that of the standard one-step M-estimators and the τ -estimators when applicable. To further complete the study, we believe the asymptotic efficiency of an estimator also provides important information about its behaviour, even though Rousseeuw and Croux quickly mentioned it, and only for the normal distribution as the underlying distribution. The asymptotic efficiency of the MOSME will

be derived from its influence function, under some regularity conditions, and for different types (heavy-tailed, normal, light-tailed) of distributions. The asymptotic efficiency of the MOSME will then be compared to that of the other one-step M-estimators. To draw a more complete parallel with Rousseeuw and Croux (1993a), the same preliminary estimators of location and dispersion, the median and the (normalized) MAD, will be used. We believe, in any case, that they represent an excellent choice of preliminary estimators.

The next two chapters treat independently the two problems of interest: estimating location with unknown dispersion, and estimating dispersion with unknown location. Each chapter will first look at the asymptotic efficiency of one-step M-estimators, then at their worst-case bias.

Chapter 4

Location Estimation with Unknown Dispersion

4.1 Introduction

Let X_1, \dots, X_n be a sample from a population with distribution F in the location-dispersion family $\{F(x) : F(x) = F(\frac{x-\theta}{\sigma})\}$. The objective is to estimate the location θ , when the dispersion parameter σ is unknown.

Given a score function ψ , an M-estimator of location is the solution T_n of the equation

$$\frac{1}{n} \sum_{i=1}^n \psi \left(\frac{X_i - T_n}{S_n} \right) = 0, \quad (4.15)$$

where S_n is a robust estimate of the dispersion parameter σ . It can be shown that, under mild regularity conditions, T_n converges a.s. $[F]$ to $T(F)$, the functional implicitly defined as the solution of

$$E_F \psi \left(\frac{X - T(F)}{S(F)} \right) = 0, \quad (4.16)$$

where $S(F)$ is the asymptotic value of S_n . We will therefore adopt the functional notation in the discussion below.

Computing an M-estimate of location requires the use of an iterative method, such as the Newton-Raphson algorithm, as we must solve the nonlinear equation (4.15). Huber (1981) suggested, as an alternative, to use the estimate found by performing only one iteration in the algorithm, starting with initial estimates of location and dispersion. With

the underlying distribution F , the *Standard One-Step M-Estimator of Location*, derived from the score functions ψ , can be formally defined by the functional

$$T_1(F) = T_0(F) + S_0(F) \frac{E_F \psi\left(\frac{x-T_0(F)}{S_0(F)}\right)}{E_F \psi'\left(\frac{x-T_0(F)}{S_0(F)}\right)},$$

where $T_0(F)$ and $S_0(F)$ are asymptotic values of the initial estimators of location and dispersion. By replacing the distribution F by the empirical distribution F_n , it is therefore possible to get a finite-sample estimator for the location parameter θ .

However, for certain score functions ψ and certain samples, it may happen that the finite-sample version of the denominator $E_F \psi'\left(\frac{x-T_0}{S_0}\right)$ becomes 0 or dangerously close to 0. To avoid this problem, Hampel et al (1986) suggested the following modified version of the standard one-step M-estimator of location:

$$T^*(F) = T_0(F) + S_0(F) \frac{E_F \psi\left(\frac{x-T_0(F)}{S_0(F)}\right)}{E_{\Phi} \psi'(z)},$$

which will be denoted hereafter the *MOSME of Location* (**M**odified **O**ne-**S**tep **M**-**E**stimator of Location).

The standard one-step M-estimator of location has essentially the same asymptotic behaviour as its fully-iterated version, at least when F is symmetric, ψ is odd and T_0 is consistent, translation invariant and odd. Our main interest is to study the asymptotic behaviour of the MOSME as a means of comparison with the standard one-step M-estimator of location. More specifically, the asymptotic efficiency at different models and the maximum asymptotic bias of both estimators will be compared.

4.2 Asymptotic Efficiency of the MOSME of Location

4.2.1 Our Choice of Underlying Distributions and of Score Functions

The study developed in section 4.2 will include a set of eleven underlying distributions F and three different score functions ψ .

The discussion in section 4.2 assumes that the underlying distribution F is symmetric. Interesting cases for F include heavy-tailed distributions, which attempt to model samples with outliers. To further make a parallel between the theory related to the asymptotic efficiency of the MOSME and its maximum bias, one case of the contaminated normal distribution will be considered. Finally, two light-tailed distributions will also be used, to illustrate the adaptative property of the MOSME to various situations.

More specifically, the distributions of interest are: the normal distribution, the Student's t distribution with 1, 2, 5, 8, 10, 20 degrees of freedom, the double exponential distribution, a contaminated standard normal distribution, with 5% of outliers with distribution $N(6, 0.01)$ and another 5% with distribution $N(-6, 0.01)$, and two distributions with lighter tails, that is, a symmetrized Beta distribution ($\alpha = \beta = 10$) and the distribution with density $f(x) = 0.5516313254\exp(-x^4)$.

In order to compare the asymptotic variances of our estimators at these distributions, it was decided to normalize the distributions so that their interquartile range would all be equal to the standard normal interquartile range, 1.349. This normalization implied scaling each distribution with a factor d_0 , so that the density associated with the central distribution F is actually $\frac{1}{d_0} f_{\theta=0, \sigma=1}(\frac{x}{d_0})$, and not $f_{\theta=0, \sigma=1}(x)$. See Appendix A for the scale factors d_0 necessary to make each of the above distributions have an interquartile range equal to 1.349.

We are however aware that the normalization is not needed if we compare the efficiencies of the estimators instead of their variances, which we actually do. Indeed, an efficiency is by definition the ratio of asymptotic variances, and the asymptotic variance under $F(\frac{x}{d_0})$ is equal to d_0^2 times the asymptotic variance under $F(x)$. Therefore, the efficiency calculations are independent of the normalization factors. However, the normalization greatly simplifies the calculations when the initial estimator of dispersion in the one-step calculation is chosen to be the MAD, as will later be shown. This is the reason why we have adopted this procedure. If one nevertheless prefers to compare the asymptotic variances of the estimators, instead of their efficiency, the material provided in this thesis will enable that person to do so.

To further illustrate the behaviour of the MOSME, the following three score functions ψ will be used:

$$\psi_{H_{1.345}}(x) = \begin{cases} \frac{x}{1.345} & |x| < 1.345 \\ \text{sign}(x) & |x| \geq 1.345 \end{cases}, \quad (4.17)$$

$$\psi_{N_{CDF}}(x) = 2\Phi(x) - 1, \quad (4.18)$$

and

$$\psi_{T_{4.7}}(x) = \begin{cases} x(4.7^2 - x^2)^2 & |x| \leq 4.7 \\ 0 & |x| > 4.7 \end{cases}. \quad (4.19)$$

They will respectively hereafter be called $H_{1.345}$, N_{CDF} and $T_{4.7}$. The first two score functions are monotone non-decreasing, whereas the third is redescending. The score function $H_{1.345}$ was initially proposed by Huber (1964) who showed that it is asymptotically min-max for $F = \Phi$, within the class of location estimators with general dispersion. The

Score Function ψ	$E_{\Phi}\psi'(z)$
$H_{1.345}$	0.6106876
N_{CDF}	.5641896
$T_{4.7}$	370.4275608

Table 4.2: The Value of the Constant Denominator, $E_{\Phi}\psi'(z)$, in the Ratio Defining the Three MOSME's of Location Under Study

choice of the constant 1.345 makes the estimator $H_{1.345}$ 95% efficient under the normal model. As will be seen later, N_{CDF} is also 95% efficient when the underlying distribution is normal. The score function $T_{4.7}$ is an example of Tukey's biweight function. The choice of the constant 4.7 makes the estimator $T_{4.7}$ 95% efficient under the normal model.

Note that the constant denominator, $E_{\Phi}\psi'(z)$, in the ratio defining the MOSME can be easily calculated when the score functions ψ are determined. Table 4.2 provides those constants. Notice that the score functions (4.17) and (4.18) are normalized to have a maximum of 1. The score function in (4.19) is not. This explains the big difference in the values of the constants in Table 4.2.

4.2.2 The Asymptotic Variance of the Standard One-Step and the Fully-Iterated M-Estimators of Location

In order to study the behaviour of the MOSME of location, it is of interest to compare its asymptotic variance to the asymptotic variance of the standard one-step M-estimator. But the latter is equivalent to the asymptotic variance of the fully iterated M-estimator of location, provided that the initial estimator of location, T_0 , is consistent, translation invariant and odd, and that the underlying distribution F is symmetric and the score function ψ is odd. These last assumptions shall be used throughout this chapter, in view of the simplification they bring to the problem.

Huber (1981) (pp. 140-141) shows that the influence function of the standard one-step M-estimator and of the fully iterated M-estimator of location is, under our assumptions,

$$IF(x; T, F) = \frac{S_0(F)\psi\left(\frac{x}{S_0(F)}\right)}{E_F\psi'\left(\frac{x}{S_0(F)}\right)}, \quad (4.20)$$

where $S_0(F)$ is the asymptotic value of the initial estimator of dispersion. Remember that $T_0(F) = 0$ when F is symmetric. Note that the above influence function (4.20) is directly proportional to the score function ψ defining the estimators.

If we take the initial estimator of dispersion to be the normalized MAD, with asymptotic value

$$S_0(F) = \frac{\text{Med}(|X|)}{\Phi^{-1}(3/4)},$$

then $S_0(F) = 1$ for all the distributions of interest presented in section 4.2.1. This is due to the normalization factor d_0 (see Appendix A) which makes the interquartile range of each distribution equal to 1.349.

With this choice of initial estimator of dispersion, the asymptotic variance of the standard one-step M-estimator of location, and so of the fully iterated M-estimator of location, simplifies to

$$V(T, F) = \frac{E_F\psi^2(x)}{\{E_F\psi'(x)\}^2},$$

under mild regularity conditions.

4.2.3 Asymptotic Variance of the MOSME of Location

Appendix B provides the complete derivation of the influence function of the MOSME of location. In the general non-symmetric case, it is equal to

$$\begin{aligned}
IF(x; T^*, F) &= IF(x; T_0, F) \left\{ 1 - \frac{E_F \psi'(y)}{E_{\Phi} \psi'(z)} \right\} + \\
&IF(x; S_0, F) \left\{ \frac{E_F \psi(y)}{E_{\Phi} \psi'(z)} - \frac{E_F [y \psi'(y)]}{E_{\Phi} \psi'(z)} \right\} + \\
&IF(x; T, F) \left\{ \frac{E_F \psi'(y)}{E_{\Phi} \psi'(z)} \right\} - S_0 \frac{E_F \psi(y)}{E_{\Phi} \psi'(z)},
\end{aligned}$$

where $T_0 = T_0(F)$ and $S_0 = S_0(F)$ are the asymptotic values of the initial estimators of location and dispersion, $y = \frac{y' - T_0}{S_0}$, $IF(x; T_0, F)$ and $IF(x; S_0, F)$ are the influence functions of the initial estimators of location and dispersion, and $IF(x; T, F)$ is the influence function of the standard (or the fully iterated) one-step M-estimator of location. However, it is possible to simplify the above expression with appropriate conditions.

Proposition 1 *Assume T_0 is consistent, translation invariant and odd. Assume that ψ is odd, bounded, differentiable except in at most a finite number of points, and equal to 0 at 0. If F is symmetric, then*

$$IF(x; T^*, F) = (1 - a) IF(x; T_0, F) + a IF(x; T, F),$$

where $a = \frac{E_F \psi'(y)}{E_{\Phi} \psi'(z)}$.

The conditions of Proposition 1 are needed for the standard one-step M-estimator of location to have the same influence function as the fully iterated M-estimator. The most important condition is the symmetry of F , which greatly simplifies our problem since it makes the influence function of the MOSME of location independent of the initial estimator of dispersion. The conditions on the score function ψ are minimal regularity conditions that most score functions used in practice will satisfy.

Note that the influence function of the MOSME illustrates its adaptative behaviour. When the underlying distribution F is approximately normal, the constant a becomes close to 1, and the MOSME behaves like the more efficient standard one-step (fully

iterated) M-estimator. On the other hand, the further away F is from normal, the further away a is from 1, and the more impact the initial (robust) estimator of location has on the behaviour of the MOSME.

One can use a Taylor series expansion, under mild regularity conditions, or the heuristic formula $V(T^*, F) = E_F\{IF(x; T^*, F)\}^2$ to obtain the asymptotic variance of the MOSME, $V(T^*, F)$. In either case, we find that

$$V(T^*, F) = E_F\{(1 - a)IF(x; T_0, F) + aIF(x; T, F)\}^2.$$

Note that the asymptotic value of the initial estimator of location, $T_0(F)$, is 0, since all the distributions are symmetric. It is however important to choose a robust initial estimator of location when using the MOSME or the standard one-step M-estimator. The median is strongly recommended in the literature. Its influence function is

$$IF(x, \text{Med}, F) = \frac{\text{sgn}(x)}{2f(0)},$$

when it is assumed that its asymptotic value is 0. Hampel et al (1986) have moreover shown that the influence function of the median has the sharpest bound for any location estimator, thus the smallest gross-error sensitivity. The asymptotic variance of the median is $1/4f(0)^2$.

As in the preceding section on the standard one-step M-estimator of location, we recommend the use of the normalized MAD as the initial estimator of dispersion. It is asymptotically equal to 1 for all distributions F under study, by the choice of the normalization factors d_0 .

4.2.4 The Asymptotic Variance of the MLE

In order to calculate the efficiency of the MOSME and the one-step M-estimator of location, it is necessary to find the asymptotic variance of the maximum likelihood estimator

Distribution F	$V(\text{MLE}, F)$
double exponential	0.9455161
contaminated normal	0.0713725
t(1)	0.909873
t(2)	1.1356899
t(5)	1.1470014
t(8)	1.1127178
t(10)	1.0962449
t(20)	1.0543198
normal	1.0000000
symmetrized beta	0.9233088
$0.55\exp(-x^4)$	0.5367305

Table 4.3: Asymptotic Variance $V(\text{MLE}, F)$ of the MLE for Different Underlying Distributions F

of location, for each underlying distribution F .

The MLE of location is defined through the score function $\psi_{\text{MLE}}(x) = -f'(x)/f(x)$. The MLE, consistent for the location θ , possesses the smallest asymptotic variance possible, namely the inverse of the Fisher information. That is, the asymptotic variance of the MLE is $V(\text{MLE}, F) = \frac{1}{E_F\{\psi_{\text{MLE}}(x)\}^2}$.

Table 4.3 gives the asymptotic variance of the MLE, $V(\text{MLE}, F)$, for the different distributions under study. Note the strikingly small asymptotic variance of the MLE for the contaminated normal distribution, which is equivalent to a very large value of the Fisher information. When integrating $\{f'(x)/f(x)\}^2$ to obtain this Fisher information, the two contaminations centered on $x = \pm 6$ cause the integral to increase significantly over values of x greater than 5 or smaller than -5. The contaminated normal is the only distribution in Table 4.3 which is not unimodal.

Distribution F	Med	MOSME			Standard One-Step		
		$H_{1.345}$	N_{CDF}	$T_{4.7}$	$H_{1.345}$	N_{CDF}	$T_{4.7}$
dble exp	1.000	0.735	0.742	0.747	0.698	0.718	0.695
cont normal	0.047	0.060	0.058	0.080	0.060	0.058	0.080
t(1)	0.811	0.620	0.609	0.781	0.569	0.571	0.716
t(2)	0.833	0.876	0.870	0.930	0.857	0.856	0.904
t(5)	0.769	0.992	0.993	0.987	0.990	0.992	0.984
t(8)	0.731	0.996	0.999	0.987	0.996	0.999	0.987
t(10)	0.716	0.993	0.996	0.984	0.993	0.997	0.985
t(20)	0.680	0.978	0.983	0.942	0.979	0.983	0.946
normal	0.637	0.950	0.950	0.950	0.950	0.950	0.950
sym beta	0.581	0.902	0.906	0.910	0.901	0.905	0.908
$\exp(-x^4)$	0.300	0.669	0.631	0.666	0.644	0.616	0.643

Table 4.4: Asymptotic Efficiency of the MOSME and the Standard One-Step M-Estimator of Location (Equivalent to That of the Fully-Iterated M-Estimator), Derived from Different Score Functions ψ and for Different Underlying Distributions F . The Asymptotic Efficiency of the Median, the Initial Estimator of Location, is Provided for Comparison Purposes. The Normalized MAD is Used As an Initial Estimator of Dispersion.

4.2.5 Asymptotic Efficiency of the MOSME Compared to That of the Standard One-Step (or Fully-Iterated) M-Estimator of Location

Table 4.4 presents the asymptotic efficiency of the MOSME and the one-step M-estimator of location (equivalent to that of the fully iterated M-estimator), derived from the three score functions (4.17), (4.18) and (4.19), for the different distributions F under study. The asymptotic efficiency of the initial estimator, the median Med, is also given in Table 4.4, as a mean of comparison.

In view of Table 4.4, it can be concluded that the MOSME of location does not have the same asymptotic properties as the fully-iterated M-estimator, contrary to the one-step M-estimator of location under our assumptions.

In fact, for the three score functions $H_{1.345}$, N_{CDF} and $T_{4.7}$, it appears that the efficiency of the MOSME is always greater than, or comparable, to that of the standard one-step M-estimator, and hence to that of the fully-iterated M-estimator, at least for the underlying distributions F selected. This improvement is especially present in the case of very heavy-tailed distributions, which puts the MOSME at an advantage over the other estimators when the sample contains outliers. The MOSME, the standard one-step and the fully-iterated M-estimators of location seem approximately equivalent, in terms of asymptotic efficiency, for underlying distributions F approaching the normal distribution.

Table 4.4 also suggests that a M-estimator derived from the $T_{4.7}$ score function shows a greater efficiency than the one derived from the $H_{1.345}$ and the N_{CDF} score functions, in presence of very heavy-tailed distributions ($t(1)$, $t(2)$ and contaminated normal). The N_{CDF} score function behaves slightly better than the $H_{1.345}$ score function, which itself behaves slightly better than $T_{4.7}$, for distributions approaching the normal distribution ($t(5)$, $t(8)$, $t(10)$ and $t(20)$). For the very light-tailed symmetrized beta distribution, the $T_{4.7}$ score function gives the highest efficiency, followed by N_{CDF} and then $H_{1.345}$. In the case of the light-tailed $0.55\exp(-x^4)$, $H_{1.345}$ and $T_{4.7}$ have comparable efficiencies, slightly higher than N_{CDF} 's. For the double exponential underlying distribution, things are not so clear. The N_{CDF} score function is superior to $H_{1.345}$ and $T_{4.7}$ when using the fully iterated M-estimator. However, the three score functions are roughly equivalent when the MOSME is used.

Therefore, because the presence of outliers is the main problem when estimating location, I would recommend using the MOSME derived from the $T_{4.7}$ score function to estimate the location of a distribution with unknown dispersion.

4.3 Maximum Bias of the MOSME of Location

Let X_1, \dots, X_n be a sample from a population with distribution F in the contamination neighborhood

$$V_\epsilon(F_0^{\theta, \sigma}) = \{F : F = (1 - \epsilon)F_0^{\theta, \sigma} + \epsilon H, H \text{ arbitrary distribution}\}, \quad 0 \leq \epsilon < 1/2,$$

where the central distribution $F_0^{\theta, \sigma}$ belongs to a location-dispersion family, that is,

$$F_0^{\theta, \sigma}(y) = F_0\left(\frac{y - \theta}{\sigma}\right).$$

The arbitrary distribution H generates the outliers that can be present in the sample, and it will invariably affect the estimation of the location parameter. Moreover, we set ϵ less than $1/2$ because it would otherwise be impossible to distinguish between the central, $F_0^{\theta, \sigma}$, and the arbitrary, H , distribution.

The maximum bias function

$$B_{T^*}(\epsilon) = \sup_{F \in V_\epsilon} |T^*(F) - \theta|,$$

as briefly introduced by Hampel et al (1986) (pp. 176-177), can be used to measure the asymptotic robustness of the MOSME of location as a function of the fraction ϵ of contamination. Note that we can take $\theta = 0$ without loss of generality because the estimate $T^*(F)$ is translation invariant.

The following compares the maximum bias of the MOSME, the standard one-step and the fully-iterated M-estimators of location in two distinct situations: when the score function ψ is monotone non-decreasing (as $H_{1.345}$ and N_{CDF} are), and when ψ is redescending (as is $T_{4.7}$).

4.3.1 Monotone Non-Decreasing Score Functions

Let the distribution F_∞ be a point mass contamination at infinity, obtained when $H = \delta_\infty$. We shall hereafter concentrate on the normal central distribution $F_0^{\theta, \sigma} = \Phi$ in the neighborhood V_ϵ .

Proposition 2 *Assume ψ is monotone non-decreasing, bounded and odd. Let $S^-(\epsilon) = \inf_{F \in V_\epsilon} S_0(F)$, $S^+(\epsilon) = \sup_{F \in V_\epsilon} S_0(F)$, and $B(\epsilon) = \sup_{F \in V_\epsilon} |T_0(F)|$. Assume we can interchange integration and derivation, that is $\partial/\partial t \{E_\Phi \psi(\frac{x-t}{s})\} = -\frac{1}{s} E_\Phi \psi'(\frac{x-t}{s})$ and $\partial/\partial s \{E_\Phi \psi(\frac{x-t}{s})\} = -\frac{1}{s} E_\Phi \psi'(\frac{x-t}{s})(\frac{x-t}{s})$. If*

$$\frac{E_\Phi \psi'(\frac{x-t}{s})}{E_\Phi \psi'(x)} \leq \frac{1}{1-\epsilon}, \quad (4.21)$$

$$\forall t \text{ in } [-B(\epsilon), B(\epsilon)], \text{ for fixed } s \text{ in } [S^-(\epsilon), S^+(\epsilon)]$$

and

$$E_\Phi \psi(\frac{x-t}{s}) - E_\Phi \psi'(\frac{x-t}{s})(\frac{x-t}{s}) \geq -\frac{\epsilon}{1-\epsilon}, \quad (4.22)$$

$$\forall s \text{ in } [S^-(\epsilon), S^+(\epsilon)], \text{ for fixed } t \text{ in } [-B(\epsilon), B(\epsilon)],$$

then

$$\sup_{F \in V_\epsilon} T^*(F) = T^*(F_\infty),$$

where $F_\infty = (1 - \epsilon)\Phi + \epsilon\delta_\infty$.

Proof: We clearly always have that $\sup_{F \in V_\epsilon} T^*(F) \geq T^*(F_\infty)$.

Moreover, $\forall \epsilon < 1/2$,

$$\sup_{F \in V_\epsilon} T^*(F) = \sup_{F \in V_\epsilon} \left\{ T_0(F) + S_0(F) \frac{E_F \psi(\frac{x-T_0(F)}{S_0(F)})}{E_\Phi \psi'(x)} \right\}$$

$$\begin{aligned}
&= \sup_{F \in \{(1-\epsilon)\Phi + \epsilon H\}} \left\{ T_0(F) + S_0(F) \frac{(1-\epsilon)E_{\Phi}\psi\left(\frac{x-T_0(F)}{S_0(F)}\right) + \epsilon E_H\psi\left(\frac{x-T_0(F)}{S_0(F)}\right)}{E_{\Phi}\psi'(x)} \right\} \\
&\leq \sup_{F \in \{(1-\epsilon)\Phi + \epsilon H\}} \left\{ T_0(F) + S_0(F) \frac{(1-\epsilon)E_{\Phi}\psi\left(\frac{x-T_0(F)}{S_0(F)}\right) + \epsilon}{E_{\Phi}\psi'(x)} \right\} \\
&= \sup_{\substack{-B(\epsilon) \leq t \leq B(\epsilon) \\ S^-(\epsilon) \leq s \leq S^+(\epsilon)}} \left\{ t + s \frac{(1-\epsilon)E_{\Phi}\psi\left(\frac{x-t}{s}\right) + \epsilon}{E_{\Phi}\psi'(x)} \right\} \\
&= B(\epsilon) + S^+(\epsilon) \frac{(1-\epsilon)E_{\Phi}\psi\left(\frac{x-B(\epsilon)}{S^+(\epsilon)}\right) + \epsilon}{E_{\Phi}\psi'(x)} \\
&= T^*(F_{\infty}).
\end{aligned}$$

The first equality is simply the definition of $T^*(F)$. By definition of V_{ϵ} , it is possible to write $E_F\psi\left(\frac{x-T_0(F)}{S_0(F)}\right)$ as the sum of two terms, as states the second equality. Since ψ is bounded, we can assume without loss of generality that $\sup_x \psi(x) = 1$. Thus $E_H\psi$ is always less or equal to 1, which gives the third line.

The function which is to be maximized on the third line does not depend on the distribution H anymore. In fact, it can be regarded as a function of two arguments, $T_0(F)$ (or t) and $S_0(F)$ (or s). Following the work of Martin and Zamar in [26], if $\epsilon < 1/2$, then $T_0(F)$ and $S_0(F)$ are bounded as in the fourth line. That is, $S^-(\epsilon) \leq S_0(F) = s \leq S^+(\epsilon)$, and $-B(\epsilon) \leq T_0(F) = t \leq B(\epsilon)$.

Assuming the conditions (4.21) and (4.22) hold, the function to be maximized in the fourth line is increasing in $T_0(F) = t$, for all fixed $S^-(\epsilon) \leq S_0(F) = s \leq S^+(\epsilon)$,

when $-B(\epsilon) \leq T_0(F) = t \leq B(\epsilon)$, and it is increasing in $S_0(F) = s$, for all fixed $-B(\epsilon) \leq T_0(F) = t \leq B(\epsilon)$, when $S^-(\epsilon) \leq S_0(F) = s \leq S^+(\epsilon)$. Therefore, we directly get the fifth line, which is by definition $T^*(F_\infty)$.

Hence, we have shown that $\sup_{F \in V_\epsilon} T^*(F) = T^*(F_\infty)$, when the central distribution $F_0^{\theta, \sigma}$ is normal. \square

Analytical derivations, combined with numerical calculations, have shown that the $H_{1.345}$ and the N_{CDF} score functions satisfy the above conditions (4.21) and (4.22), when the median and the (normalized) MAD are used as preliminary estimates of location and dispersion.

The conditions were rewritten for the two specific cases of $H_{1.345}$ and N_{CDF} score functions, and evaluated over a finite and equally-spaced 21x21 grid, covering the range of possible t and s values. For a fixed ϵ , the maximum value of the (normalized) MAD, $S^+(\epsilon)$, is produced by a point mass contamination at infinity, and such contamination also produces the maximum value $B(\epsilon)$ of the location estimator. The minimum value of the (normalized) MAD, $S^-(\epsilon)$, is produced by a point mass contamination at 0, and such contamination also produces the minimum absolute value of the location estimator, 0 (see Martin and Zamar (1989)). More specifically, the value for $B(\epsilon)$ can be explicitly written as $\Phi^{-1}(\frac{1}{2(1-\epsilon)})$. It is the value $T(F)$ which satisfies (4.16) for the median score function $\chi_{Med}(x) = \text{sgn}(x)$ and $F = F_\infty$. The bounds $S^+(\epsilon)$ and $S^-(\epsilon)$ are the implicit solutions of

$$(1 - \epsilon)[\Phi\{B(\epsilon) - S^+(\epsilon)\Phi^{-1}(3/4)\} + 1 - \Phi\{B(\epsilon) + S^+(\epsilon)\Phi^{-1}(3/4)\}] + \epsilon = 1/2,$$

and

$$(1 - \epsilon)[\Phi\{-S^-(\epsilon)\Phi^{-1}(3/4)\} + 1 - \Phi\{S^-(\epsilon)\Phi^{-1}(3/4)\}] = 1/2.$$

That is, $S^+(\epsilon)$ satisfies $E_{F_\infty} \chi_{MAD}\left(\frac{x-B(\epsilon)}{S^+(\epsilon)}\right) = 1/2$ and $S^-(\epsilon)$ satisfies $E_{F_0} \chi_{MAD}\left(\frac{x}{S^-(\epsilon)}\right) = 1/2$, where the MAD score function is $\chi_{MAD}(x) = 1/2\{\text{sgn}(|x| - \Phi^{-1}(3/4)) + 1\}$ (see chapter 5 for more details about score functions of dispersion estimators) and $F_0 = (1 - \epsilon)\Phi + \epsilon\delta_0$.

For all $\epsilon < 1/2$ used, the conditions (4.21) and (4.22) were always met. So, even if those conditions seem somewhat restrictive, it is believed that many often used monotone non-decreasing, bounded and odd score functions ψ satisfy them.

A lower bound for the maximum bias of the fully iterated location M-estimator is $T(F_\infty)$, ie the positive solution t of the non-linear equation

$$-E_\Phi \psi\left(\frac{x-t}{S^+(\epsilon)}\right) = \frac{\epsilon}{1-\epsilon},$$

for a fixed ϵ .

Moreover, if we assume that the score function ψ is monotone non-decreasing, bounded and odd, then we have a lower bound for the bias of the standard one-step location M-estimator, $T_1(F)$, given by $T_1(F_\infty)$, where

$$T_1(F_\infty) = B(\epsilon) + S^+(\epsilon) \frac{(1-\epsilon)E_\Phi \psi\left(\frac{x-B(\epsilon)}{S^+(\epsilon)}\right) + \epsilon}{(1-\epsilon)E_\Phi \psi'\left(\frac{x-B(\epsilon)}{S^+(\epsilon)}\right)}.$$

It is therefore possible to compare, in terms of maximum asymptotic bias, the MOSME with the standard one-step M-estimator, as well as with the fully iterated M-estimator of location.

Figures 4.1 and 4.2 show the maxbias curve of the MOSME and a lower bound for the maxbias curves of the one-step M-estimator and the fully iterated M-estimator of location, derived from the $H_{1.345}$ and the N_{CDF} score functions. In both cases, the MOSME shows a smaller maximum asymptotic bias than the one-step M-estimator and the fully iterated M-estimator, $\forall \epsilon < 1/2$. Note that the improvement by the MOSME

over the one-step and the fully iterated M-estimators of location is especially striking with the $H_{1.345}$ score function.

The maximum bias for the three types of M-estimator was also calculated, using Huber's score function with different values of index, ranging from 0.5 to 1.75. Similar results as those with the $H_{1.345}$ score function were obtained; the bigger the value of the index, the better the improvement by the MOSME, compared to the one-step and the fully iterated M-estimators of location. Hence, the MOSME clearly shows improved asymptotic robustness over the standard one-step M-estimator, as well as the fully iterated M-estimator of location.

4.3.2 Redescending Score Functions

As in the previous section, let's concentrate on the normal central distribution $F_0^{\theta, \sigma} = \Phi$ in the neighborhood V_ϵ .

It is possible to obtain a lower bound on the maximum bias of the MOSME, the standard one-step and the fully iterated M-estimators of location when the score function used in their definition is redescending, as for example the $T_{4.7}$ score function is. Indeed, let $x^* = 2.1$ be the value at which $T_{4.7}$ is maximized. Let F_* be the point mass contamination at x^* , obtained in the neighborhood V_ϵ when $H = \delta_*$. Then, $T^*(F_*)$, $T_1(F_*)$ and $T(F_*)$ are lower bounds on the maximum bias of respectively the MOSME, the standard one-step and the fully iterated M-estimator of location.

To obtain $T^*(F_*)$, $T_1(F_*)$ and $T(F_*)$, we must first determine what the initial estimators of location and dispersion, the median and the normalized MAD, become as a function of ϵ at F_* . For any $\epsilon < 1/2$, the median at F_* is the value $B(\epsilon)$ which satisfies

$$E_{F_*} \psi_{Med} \left(\frac{x - B(\epsilon)}{S^+(\epsilon)} \right) = 1/2,$$

where $S^+(\epsilon)$ is the normalized MAD at F_* and ψ_{Med} is the score function defining the estimator Med. It turns out that we can explicitly write $B(\epsilon) = \Phi^{-1}(\frac{1}{2(1-\epsilon)})$ when ϵ is approximately equal or less than 0.49.

The normalized MAD at F_* is the value $S^+(\epsilon)$ which satisfies

$$E_{F_*} \psi_{MAD} \left(\frac{x - B(\epsilon)}{S^+(\epsilon)} \right) = 1/2,$$

where $B(\epsilon)$ is the median at F_* and ψ_{MAD} is the score function defining the normalized MAD. It turns out that for ϵ approximately equal or less than 0.30, $S^+(\epsilon)$ satisfies more precisely

$$(1 - \epsilon)[\Phi\{B(\epsilon) - S^+(\epsilon)\Phi^{-1}(3/4)\} + 1 - \Phi\{B(\epsilon) + S^+(\epsilon)\Phi^{-1}(3/4)\}] + \epsilon = 1/2.$$

Therefore, for ϵ approximately equal or less than 0.30, a lower bound for the maximum bias of the MOSME defined with the score function $T_{4.7}$ is

$$T^*(F_*) = B(\epsilon) + S^+(\epsilon) \frac{(1 - \epsilon)E_{\Phi} \psi_{T_{4.7}}(\frac{x-B(\epsilon)}{S^+(\epsilon)}) + \epsilon \psi_{T_{4.7}}(\frac{x^*-B(\epsilon)}{S^+(\epsilon)})}{E_{\Phi} \psi'_{T_{4.7}}(z)}.$$

Similarly, we have a lower bound on the maximum bias of the standard one-step M-estimator of location:

$$T_1(F_*) = B(\epsilon) + S^+(\epsilon) \frac{(1 - \epsilon)E_{\Phi} \psi_{T_{4.7}}(\frac{x-B(\epsilon)}{S^+(\epsilon)}) + \epsilon \psi_{T_{4.7}}(\frac{x^*-B(\epsilon)}{S^+(\epsilon)})}{(1 - \epsilon)E_{\Phi} \psi'_{T_{4.7}}(\frac{x-B(\epsilon)}{S^+(\epsilon)}) + \epsilon \psi'_{T_{4.7}}(\frac{x^*-B(\epsilon)}{S^+(\epsilon)})}.$$

Finally, a lower bound on the maximum bias of the fully iterated M-estimator of location is the positive solution t of

$$E_{F_*} \psi_{T_{4.7}} \left(\frac{x - t}{S^+(\epsilon)} \right) = 0,$$

which can be more precisely written as

$$(1 - \epsilon)E_{\Phi} \psi_{T_{4.7}} \left(\frac{x - t}{S^+(\epsilon)} \right) + \epsilon \psi_{T_{4.7}} \left(\frac{x^* - t}{S^+(\epsilon)} \right) = 0.$$

Figure 4.3 shows the lower bounds on the maximum bias of the MOSME, the standard one-step and the fully iterated M-estimators of location, defined through the non-redescending score function $T_{4.7}$. It appears that for small fraction of contamination ϵ ($\epsilon \leq 0.12$), the three estimators are equivalent in terms of maximum bias. For medium ϵ ($0.12 < \epsilon \leq 0.30$), the standard one-step and the fully iterated estimators are non-distinguishable and appear to have a slightly lower maximum bias than the MOSME. However, this cannot be ascertained for a fact, as the figure shows only a lower bound on the maximum bias. However, the redescending nature of the Tukey score function, which discards the contribution of very large values of x , may makes it superflous to change the denominator in $S_1(F)$ to get the MOSME of location. This would explain possible higher bias of the MOSME compared to the standard one-step M-estimator of location.

4.3.3 Any Type of Score Function

It was shown in section 4.3.1 that the maximum bias of the MOSME is uniformly lower than that of the standard one-step or the fully iterated M-estimator of location defined through a monotone non-decreasing score function. When a redescending score function is used, the MOSME is equivalent to the other two estimators for small fraction of contamination ϵ , as section 4.3.2 presented. Therefore, it is possible to now focus our attention to MOSME's only, especially when ϵ is small.

When using the MOSME of location, one needs to decide which score function to select in its definition, so as to minimize its maximum bias. Figure 4.4 shows the maximum bias of the MOSME of location defined through the $H_{1.345}$ and the N_{CDF} score functions, and a lower bound for the maximum bias of the $T_{4.7}$ MOSME of location. It is impossible to conclude anything when $\epsilon > 0.12$, but for small fraction of contamination, Figure 4.4 shows that the Huber score function has the lowest maximum bias, followed by the

Tukey score function, and then N_{CDF} . For larger ϵ , the Huber score function should be preferred to the N_{CDF} score function.

4.3.4 Further Work to Be Done

The work about asymptotic robustness was developed for one specific case of central distribution $F_0^{\theta, \sigma}$ in the neighborhood of V_ϵ : the normal distribution Φ . More calculations need to be done for different central distributions $F_0^{\theta, \sigma}$, as for example, the ones used in the asymptotic efficiency section (4.2) of this text.

Finally, it remains to be seen whether the asymptotic properties of the MOSME and the standard one-step M-estimators studied in this section reflect on their finite-sample performance for small and medium sample sizes. This could be assessed with Monte-Carlo simulations, for example.

4.4 An Example: Hummingbirds

In asymptotic terms, the MOSME of location performs better than the standard one-step M-estimator and the fully iterated M-estimator of location. However, how well do they handle finite data sets?

Figure 4.5 present the bar plots of flying times (in seconds) of four types of hummingbirds: adult females (AF), adult males (AM), junior females (JF) and junior males (JM). During 15 minutes, each bird was put in a cage containing two perches 0.5 meter apart. A red light above the perches alternatively flashed so as to indicate to the bird to fly towards it (the birds were previously trained to react to flashing red lights in that manner). The time for the bird to fly from one perch to the other (in seconds) was recorded. It is believed to be a measure of the agility of the birds. Hummingbirds can

fly for long periods of time without rest. In 15 minutes, a bird typically flew 200 times from one perch to the other.

A close look at Figure 4.5 shows that some bird wandered around before hitting the flashing perch, during some of their flights, introducing extreme outliers in their flying times. On the other hand, for other birds, it is difficult to determine whether a high flying time was due to wandering or if it effectively represents the bird's performance. For that reason, the researcher who provided the data set was interested in using a measure of location to describe the distribution of flying times for each bird that would be resistant enough to possible outliers, while allowing for flexibility if long flying times of a bird are normal. She hesitated between using the median and the mode. Our MOSME of location provides an interesting alternative.

Table 4.5 shows the estimated value of different location measures for the distribution of flying times of all 16 hummingbirds, that is, their means, medians, MOSME's and standard one-step M-estimates of location. The standard one-step M-estimates are shown; note however that the estimates obtained with the fully iterated M-estimator of location are roughly equal, up to the third decimal place, except for the adult female 2, which has an extreme outlier.

Many observations can be drawn from Table 4.5. In general, the MOSME's and the standard one-step M-estimates of location are roughly equal, except in the cases of AF 2, AF 4, JF 1, JF 2 and JM 3 hummingbirds, depending on the score function used. These particular birds have extreme outliers or a very heavy tail. Their MOSME's are closer to the median than the standard one-step M-estimates, which illustrates the special adaptative nature of the MOSME of location.

Notice as well that the Tukey score function appears to be conservative, compared to $H_{1.345}$ or N_{CDF} , in the robustness sense of the term. The MOSME's or the standard

Bird	Mean	Median	MOSME of Location			Standard One-Step		
			$H_{1.345}$	N_{CDF}	$T_{4.7}$	$H_{1.345}$	N_{CDF}	$T_{4.7}$
AF 1	0.707	0.669	0.675	0.676	0.666	0.675	0.676	0.666
AF 2	0.930	0.580	0.607	0.607	0.591	0.610	0.609	0.594
AF 3	0.642	0.609	0.615	0.615	0.608	0.615	0.616	0.608
AF 4	0.751	0.630	0.653	0.655	0.630	0.657	0.659	0.630
AF 5	0.796	0.761	0.770	0.772	0.760	0.771	0.773	0.760
AM 1	0.844	0.812	0.821	0.822	0.814	0.822	0.823	0.815
AM 2	0.886	0.840	0.850	0.851	0.841	0.851	0.852	0.841
AM 3	0.663	0.640	0.644	0.645	0.640	0.644	0.645	0.641
AM 4	0.658	0.620	0.627	0.627	0.621	0.627	0.628	0.621
AM 5	0.673	0.652	0.659	0.660	0.656	0.660	0.660	0.656
JF 1	0.740	0.620	0.653	0.654	0.630	0.657	0.659	0.632
JF 2	0.686	0.561	0.582	0.583	0.564	0.586	0.586	0.565
JF 3	0.657	0.618	0.619	0.620	0.614	0.619	0.620	0.614
JM 1	0.666	0.560	0.571	0.572	0.556	0.573	0.573	0.555
JM 2	0.647	0.570	0.576	0.576	0.568	0.576	0.577	0.568
JM 3	0.733	0.661	0.693	0.693	0.678	0.694	0.694	0.681

Table 4.5: Measures of Location of Flying Times of Four Types of Hummingbirds: Adult Females (AF), Adult Males (AM), Junior Females (JF) and Junior Males (JM)

one-step estimates defined with the $T_{4.7}$ score function are always closer to the median than the estimates defined through the other two score functions. The Tukey score function sometimes even gives smaller estimates than the median. This completes our results of section 4.2 which makes the Tukey score function a better choice in presence of heavy-tailed distributions, or outliers.

Among all bar plots of flying times in Figure 4.5, the one for AM 5 appears to be best approximated by a normal curve. The one-step M-estimates of the location of the flying times of this bird are midway between the median and the mean, which is the closest the estimates get from the mean over among all the birds. The one-step estimates confirms that for AM 5, the MLE may not be a bad estimator of location after all, though some caution in its use is necessary.

It appears that adult males have the worst agility, and that junior males are the most agile birds. Adult and junior females hummingbirds a priori do not show a difference in agility. In order to estimate the mean flying times of one of the four types of birds, and conclude statistically that one type was more agile than the others, one could use a robust analysis which is beyond the scope of this thesis.

4.5 Conclusions

The MOSME of location presented in this chapter is uniformly better than the standard one-step location M-estimator, in the sense that it is easier to compute, it has a comparable and sometimes better asymptotic efficiency under many important symmetric distributions, and has a lower or comparable asymptotic maximum bias when the central distribution is normal. When using a monotone non-decreasing score function, the MOSME has a lower asymptotic bias than the standard one-step for any fraction of

contamination. With a redescending score function, the maximum bias of the MOSME is comparable to that of the standard one-step for small, but realistic, fractions of contamination.

Under our assumptions, the standard one-step M-estimator of location is asymptotically equivalent to the fully iterated M-estimator. This makes again the MOSME better than the fully iterated M-estimator, in terms of asymptotic behaviour. And since finding the solution of a non-linear equation, as is required when computing the fully iterated location M-estimator, can sometimes be problematic, the MOSME still is preferable to the fully iterated M-estimator.

The superiority of the MOSME in terms of asymptotic efficiency is especially strong for very heavy-tailed distributions, which describe the situations of samples with outliers.

Martin and Zamar (1989) have shown through finite sample-size simulations that in practice, the squared bias is at least as large as the variance of M-estimators for rather modest sample sizes. Therefore, the comparison between the MOSME and the standard one-step (fully iterated) M-estimator should give more weight to the maximum bias. And indeed, the MOSME beats uniformly and clearly the other two estimators in terms of maximum asymptotic bias when it is derived from a monotone non-decreasing score function; and it is comparable to the other two estimators for small fractions of contamination, when it is derived from a redescending score function.

It is impossible to single out the best score function to define the MOSME, as all score functions perform optimally in specific, and different, situations. However, when striving for an estimator as accurate (as in low bias) and as precise (as in high efficiency) as possible, and if it is known that the contamination by outliers is small, the Huber score function should be used. If one wants to use a monotone non-decreasing score function, one should prefer the Huber score function to the normal one. When the contamination

by outliers is large, the Tukey score function may represent a better choice than the Huber score function, mainly due to the fact that the efficiency of the Tukey is higher for heavy-tailed distributions than that of the Huber estimator. However, at this point, a formal comparison in terms of maximum bias between the Tukey and the Huber score functions is not possible for large fraction of contamination.

With finite data sets, the example in section 4.4 shows that the MOSME is more robust than the one-step (or the fully iterated) M-estimators of location when it needs to be. Indeed, in the presence of extreme outliers or very heavy tails, the values of the MOSME are closer to the median than that of the other two estimators. The MOSME, by its adaptative nature, becomes more robust and conservative when the data indicates that caution is necessary. On the other hand, the one-step (or the fully iterated) M-estimators do not handle so well data that is far from normally distributed.

We therefore believe that the use of the MOSME should be preferred to that of the standard one-step location M-estimator, or the fully iterated location M-estimator, when estimating the location parameter θ of a distribution.

Furthermore, most of the results presented in this chapter are of an asymptotic nature. It remains to be seen whether the asymptotic superiority of the MOSME reflects on its finite-sample performance for small and medium sample sizes. It would therefore be necessary to make simulations in order to establish for what minimum sample size the results presented in this paper become approximately valid. One could also assess the finite-sample behaviour of one-step M-estimators of location with sensitivity curves and empirical bias curves. If these finite sample-size results agree with the asymptotic results presented in this chapter, and that for small sample sizes, the acceptance of the MOSME as a tool would be greatly facilitated.

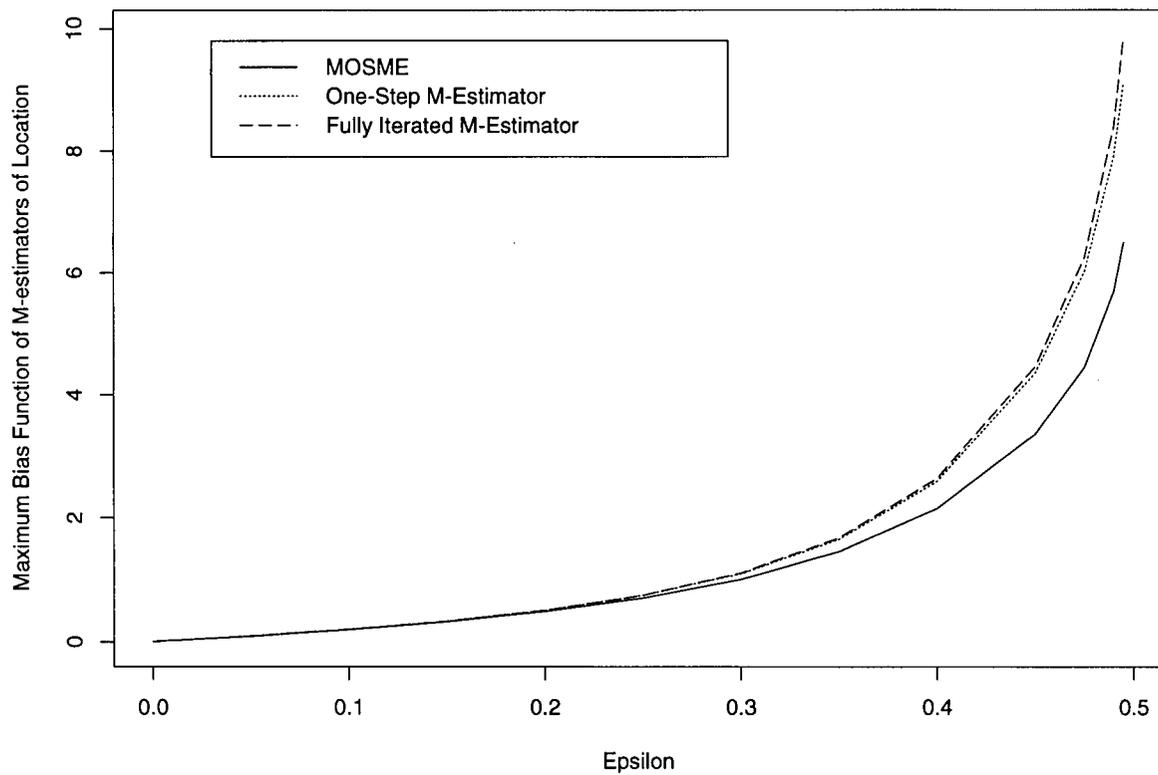
Maximum Bias Function of M-estimators Using $H_{1.345}$ Score Function

Figure 4.1: Maximum Bias Function of the MOSME, and a Lower Bound for the Maximum Bias Function of the One-Step M-Estimator and the Fully Iterated M-Estimator of Location, Derived from the $H_{1.345}$ Score Function

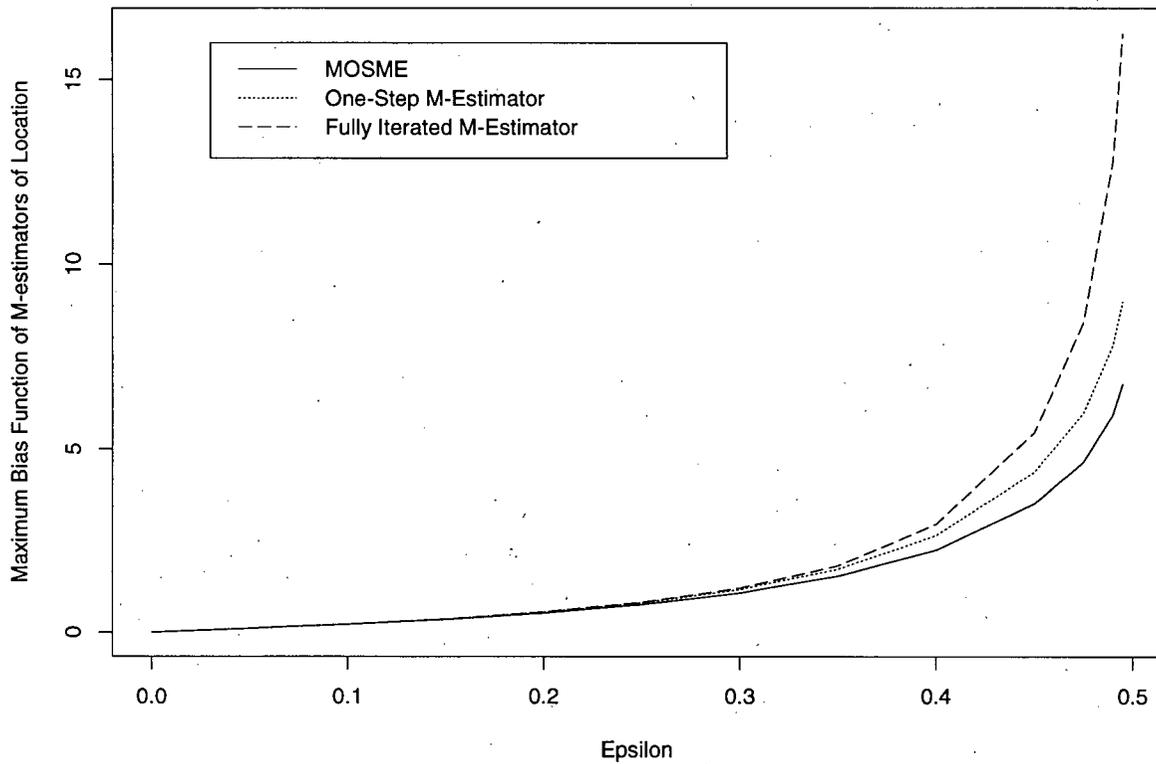
Maximum Bias Function of M-estimators Using N_{CDF} Score Function

Figure 4.2: Maximum Bias Function of the MOSME, and a Lower Bound for the Maximum Bias Function of the One-Step M-Estimator and the Fully Iterated M-Estimator of Location, Derived From the N_{CDF} Score Function

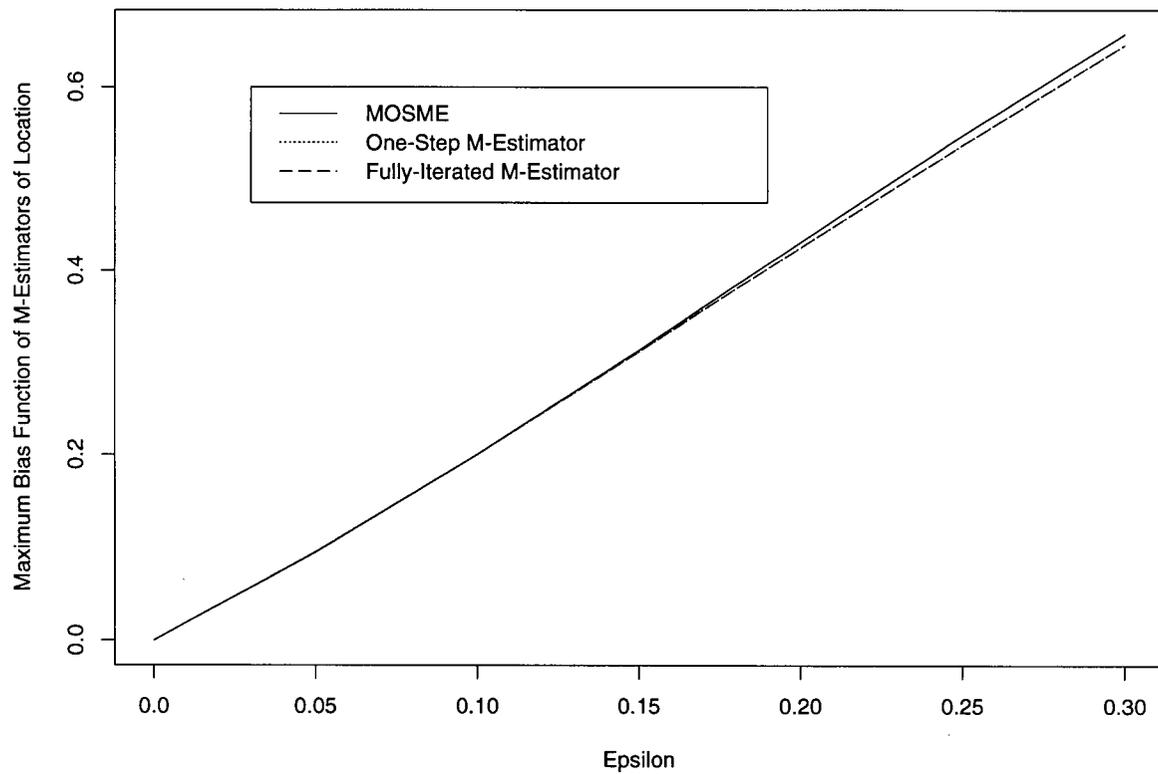
Maximum Bias Function of M-Estimators Using $T_{4.7}$ Score Function

Figure 4.3: Lower Bound on the Maximum Bias Function of the MOSME, the Standard One-Step and the Fully-Iterated M-Estimators of Location, Derived from the $T_{4.7}$ Score Function

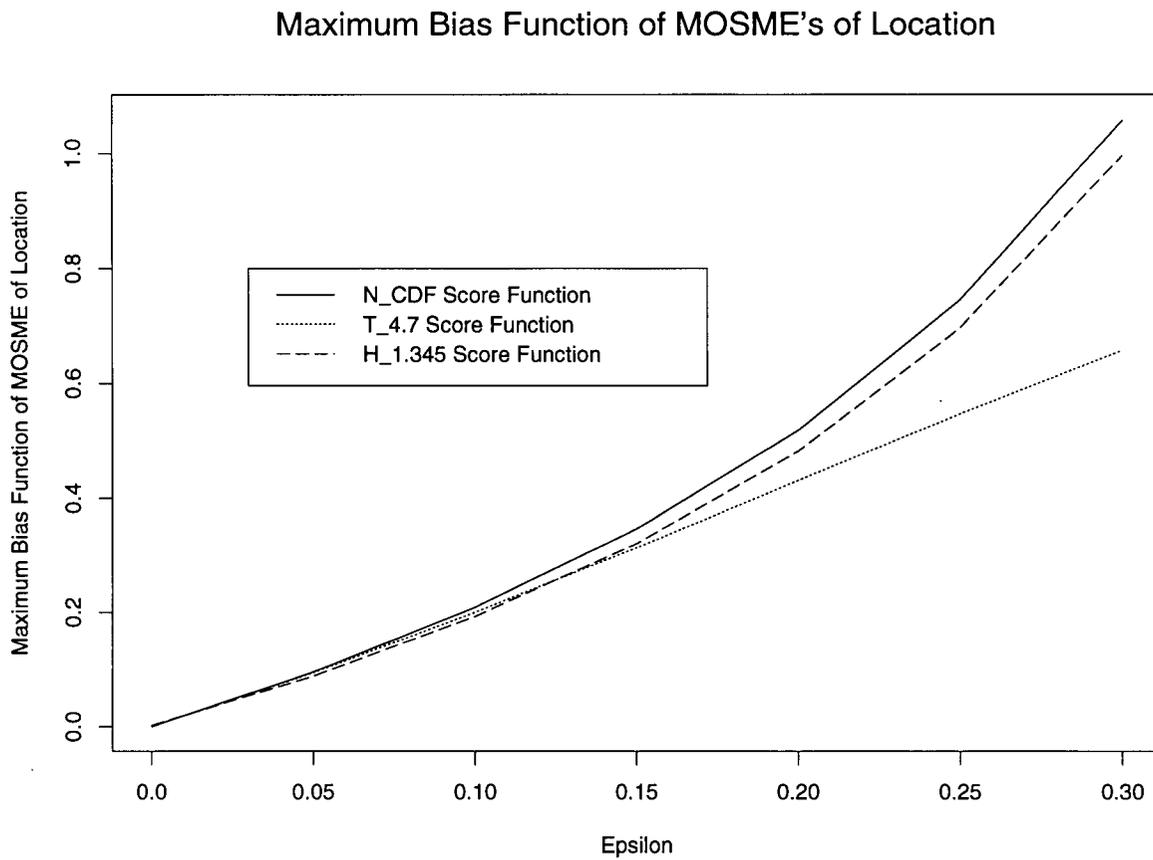


Figure 4.4: Maximum Bias Function of the MOSME's of Location Derived from the $H_{1.345}$ and the N_{CDF} Score Functions, and a Lower Bound on the Maximum Bias Function of the MOSME of Location Derived from the $T_{4.7}$ Score Function

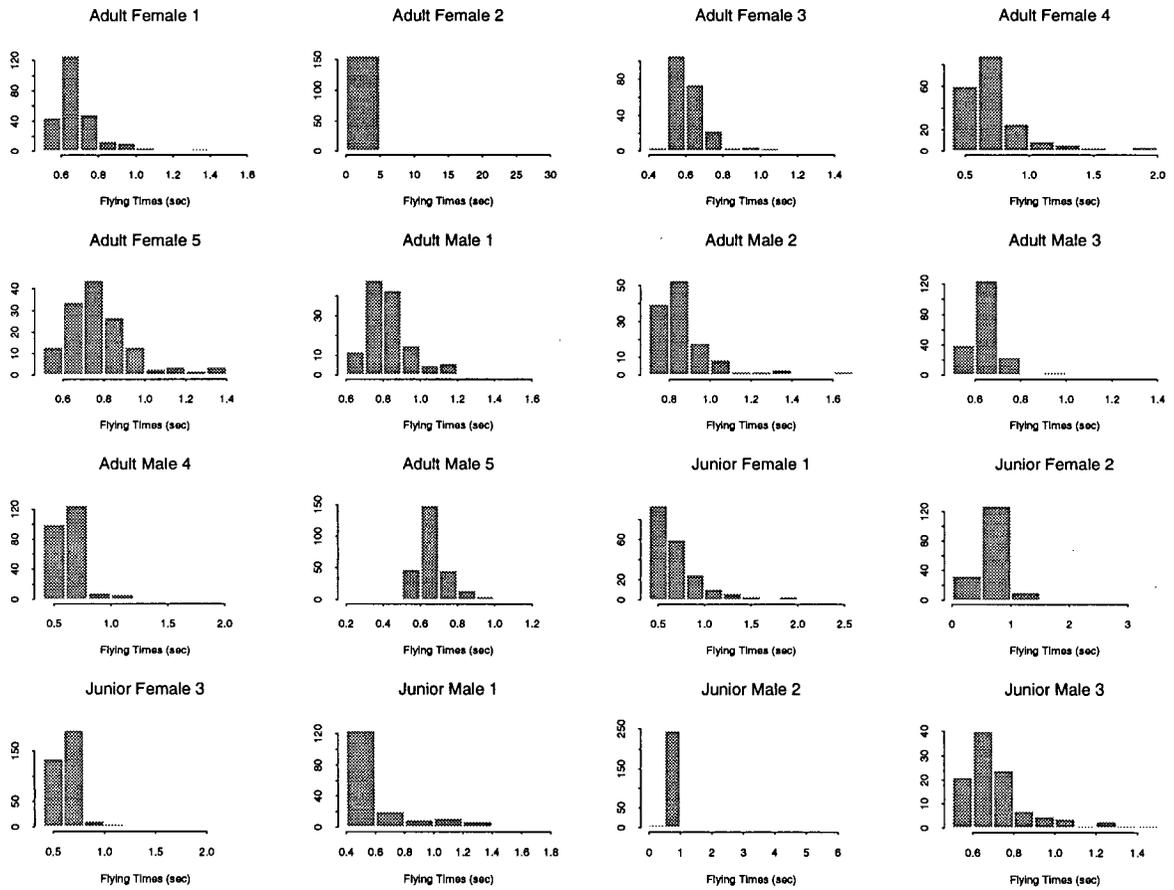


Figure 4.5: Bar Plots of Flying Times of Four Types of Hummingbirds: Adult Females, Adult Males, Junior Females and Junior Males

Chapter 5

Dispersion Estimation with Unknown Location

5.1 Introduction

Let X_1, \dots, X_n be a sample from a population with distribution F in the location-dispersion family $\{F(x) : F(x) = F(\frac{x-\theta}{\sigma})\}$. The objective is to estimate the dispersion σ , when the location parameter θ is unknown.

Given a score function χ , an M-estimator of dispersion is the solution S_n of the equation

$$\frac{1}{n} \sum_{i=1}^n \chi \left(\frac{X_i - T_n}{S_n} \right) = 0, \quad (5.23)$$

where T_n is a robust estimate of the location parameter θ . It can be shown that, under mild regularity assumptions, S_n converges a.s. $[F]$ to $S(F)$, the functional implicitly defined as the solution of

$$E_F \chi \left(\frac{X - T(F)}{S(F)} \right) = 0,$$

where $T(F)$ is the asymptotic value of T_n . We will therefore adopt the functional notation in the discussion below.

Computing an M-estimator of dispersion requires the use of an iterative method, as we must solve the nonlinear equation (5.23). Because the score function χ is of the form $\chi(x) = g(x) - \beta$, where g is an even function with $g(0) = 0$, Huber (1981) suggested, as an alternative, to use the estimate found by performing one iteration in the fixed-point

algorithm (starting with some initial estimates of location and dispersion). With the underlying distribution F , the τ -Estimator of Dispersion, $S_1^p(F)$, derived from the score functions χ , can be formally defined by the functional

$$S_1^p(F) = \sqrt{\frac{S_0(F)^2}{\beta} E_F \chi \left(\frac{x - T_0(F)}{S_0(F)} \right)},$$

where $T_0(F)$ and $S_0(F)$ are the asymptotic values of the initial estimators of location and dispersion. By replacing the distribution F by the empirical distribution F_n , it is therefore possible to get a finite-sample estimator for the dispersion parameter σ . Note that the subscript “p” in S_1^p stands for “p”oint, as in fixed-point.

Another iterative algorithm widely used to estimate location with an M-estimator, but not so much dispersion, is the Newton-Raphson method. The *Standard One-Step M-Estimator of Dispersion*, $S_1(F)$, can be defined as the first iteration of this algorithm, starting from initial estimates of location and dispersion. More precisely,

$$S_1(F) = S_0(F) + \frac{S_0(F) E_F \chi \left(\frac{x - T_0(F)}{S_0(F)} \right)}{E_F \chi' \left(\frac{x - T_0(F)}{S_0(F)} \right) \left(\frac{x - T_0(F)}{S_0(F)} \right)},$$

where $T_0(F)$ and $S_0(F)$ are the asymptotic values of the initial estimators of location and dispersion.

However, for certain score functions χ and certain samples, it may happen that the finite-sample version of the denominator $E_F \chi' \left(\frac{x - T_0(F)}{S_0(F)} \right) \left(\frac{x - T_0(F)}{S_0(F)} \right)$ in $S_1(F)$ becomes 0 or dangerously close to 0. To avoid this problem, we suggest replacing the denominator by $E_{\Phi} \chi'(z)z$, following Hampel et al (1986)'s idea on p. 153 in the context of location estimation with unknown dispersion. This defines the *MOSME (Modified One-Step M-Estimator) of Dispersion*:

$$S^*(F) = S_0(F) + S_0(F) \frac{E_F \chi \left(\frac{x - T_0(F)}{S_0(F)} \right)}{E_{\Phi} \chi'(z)z}.$$

Our main interest is to evaluate the asymptotic behaviour of the MOSME of dispersion. It is therefore necessary to compare it with the estimators it is aimed at improving: the standard one-step M-estimator and the τ -estimator of dispersion. We believe it has adaptive properties which the τ -estimator or the standard one-step M-estimator lack.

The M-estimators of dispersion, other than the MLE's, are not consistent for distributions other than the standard normal, Φ , which makes their asymptotic values differ, as well as their asymptotic properties. For this reason, the usual notion of asymptotic efficiency of an estimator will not apply directly. Instead, the relative asymptotic efficiency will be used to assess the asymptotic performance of the MOSME of dispersion as compared with the standard one-step M-estimator and the τ -estimator. The worst-case asymptotic bias of the three estimators will also be evaluated and commented on in the following sections.

5.2 Relative Asymptotic Efficiency of the MOSME of Dispersion

5.2.1 Our Choice of Underlying Distributions and of Score Functions

To assess the asymptotic efficiency of the one-step estimators of dispersion, the eleven underlying distributions F used in chapter 4 (see section 4.2.1) will again be considered.

To further illustrate the behaviour of the MOSME, the following three score functions χ will be used:

$$\chi_{H_{0.975}}(x) = \begin{cases} x^2 - 0.5 & |x| < 0.975 \\ 0.451 & |x| \geq 0.975 \end{cases}, \quad (5.24)$$

$$\chi_{H_{2.376}}(x) = \begin{cases} x^2 - 0.9686 & |x| < 2.376 \\ 4.6768 & |x| \geq 2.376 \end{cases}, \quad (5.25)$$

and

$$\chi_{T_{3.86}}(x) = \begin{cases} \frac{x^6}{3.86^6} - \frac{3x^4}{3.86^4} + \frac{3x^2}{3.86^2} - 0.165 & |x| < 3.86 \\ 0.835 & |x| \geq 3.86 \end{cases} \quad (5.26)$$

The first two score functions, (5.24) and (5.25), are two cases of the general score function

$$\chi_c(x) = \begin{cases} x^2 & |x| < c \\ c^2 & |x| \geq c \end{cases} \quad (5.27)$$

suggested by Huber (1981) (p. 109), which satisfies $E_{\Phi} \chi_c(x) = \beta(c)$. Notice that it is the square of Huber's score function ψ_{H_c} which defines a fully iterated M-estimator that is variance minimax in the location setup. For c not too small, the fully iterated M-estimator of dispersion defined through (5.27) is also optimally B-robust (see [15], p. 122).

The MOSME and the standard one-step M-estimator derived from (5.24) will be hereafter denoted by $H_{0.975}$. The score function $\chi_{H_{0.975}}$ was chosen by tradition for the value $\beta(c) = \beta(0.975) = 1/2$. The MOSME and the standard one-step M-estimator derived from (5.25) will be hereafter denoted by $H_{2.376}$. The value of the constant 2.376 was chosen because it makes the MOSME and the standard one-step M-estimator of dispersion 95% efficient under the normal model. Finally, the score function (5.26) is derived from Tukey's redescending score function in the location setup. It will define a MOSME and a standard one-step M-estimator denoted by $T_{3.86}$. Both the MOSME and the standard one-step M-estimator of dispersion $T_{3.86}$ are approximately 95% efficient under the normal model.

The τ -estimators derived from the score functions (5.25) and (5.26) are less than 95% efficient at the normal model. To make a fair comparison between the τ -estimator and

the MOSME or the standard one-step M-estimator of dispersion, we must ensure that all estimators are 95% efficient at the normal model. The two following score functions:

$$\chi_{H_{2.516}}(x) = \begin{cases} x^2 & |x| < 2.516 \\ 6.33 & |x| \geq 2.516 \end{cases}, \quad (5.28)$$

with $\beta(2.516) = E_{\Phi}\chi_{H_{2.516}}(x) = 0.9785$, and

$$\chi_{T_{5.3}}(x) = \begin{cases} \frac{x^6}{5.3^6} - \frac{3x^4}{5.3^4} + \frac{3x^2}{5.3^2} & |x| < 5.3 \\ 1 & |x| \geq 5.3 \end{cases}. \quad (5.29)$$

with $\beta(5.3) = E_{\Phi}\chi_{T_{5.3}}(x) = 0.096$, define two τ -estimators that are 95% efficient at the normal model. They will respectively be denoted by $H_{2.516}$ and $T_{5.3}$. Note that the value of the non-differentiable point 5.3 in the Tukey score function, which makes the τ -estimator 95% efficient at the normal model, is 40% higher than that for the MOSME. This will invariably affect the worst-case bias of the τ -estimator of dispersion. The increase in the non-differentiable point c in the Huber score function is not as pronounced.

The constant denominator, $E_{\Phi}\chi'(z)z$, in the ratio defining the MOSME can be easily calculated when the score functions χ are determined. Table 5.6 provides those constants. The slightly higher constant for $H_{2.516}$ is due to the fact that none of the three score functions are normalized to have a maximum of 1. The score function $H_{2.516}$ in fact has a much larger maximum than the other two.

5.2.2 Asymptotic Value of the MOSME, the One-Step M-Estimator and the τ -Estimator of Dispersion

To draw a parallel between the estimation of location discussed in the previous chapter and the estimation of dispersion, a first attempt consists in providing the efficiency for the estimators of dispersion. However, it is in a sense meaningless to compare the maximum

Score Function χ	$E_{\Phi}\chi'(z)z$
$H_{0.975}$	0.3736064
$H_{2.376}$	1.7396048
$T_{3.86}$	0.2677105

Table 5.6: The Value of the Constant Denominator, $E_{\Phi}\chi'(z)z$, in the Ratio Defining the Three MOSME's of Dispersion Under Study

likelihood estimator of dispersion to one-step M-estimators, because they do not estimate the same quantity.

The maximum likelihood estimator of dispersion is consistent for σ . Without loss of generality, we can assume that $\sigma = 1$. Therefore, the MLE takes asymptotically the value 1. Indeed, for a given distribution F , the score function defining the MLE is

$$\chi_{\text{MLE}}(x) = -x \frac{f'(x)}{f(x)} - 1.$$

Assuming F is symmetric, we must solve the equation $E_F \chi_{\text{MLE}}(\frac{x}{\sigma}) = 0$ when calculating the asymptotic value of the MLE of dispersion, σ . But $E_F \left\{ -x \frac{f'(x)}{f(x)} \right\} = 1$, assuming we can interchange derivation and integration. Therefore, $E_F \left\{ -\frac{x}{1} \frac{f'(\frac{x}{1})}{f(\frac{x}{1})} \right\} - 1 = 0$, and the MLE takes asymptotically the value 1.

On the other hand, one-step M-estimators will estimate a function of σ , function that depends on the type of M-estimator (standard one-step M-estimator, MOSME or τ -estimator), the score function χ used in its definition and the initial estimates of location and dispersion. Table 5.7 provides the asymptotic value of the MLE, the MOSME, the standard one-step M-estimator and the τ -estimator of dispersion, as well as the normalized MAD (median absolute deviation from the median, multiplied by the inverse of $\Phi^{-1}(3/4)$) and the standard deviation, for the different underlying distributions F of interest. The initial estimators of dispersion and location used to calculate the one-step

F	MLE	MOSME			Standard One-Step			τ -Estimator		MAD	SD
		$H_{0.975}$	$H_{2.376}$	$T_{3.86}$	$H_{0.975}$	$H_{2.376}$	$T_{3.86}$	$H_{2.516}$	$T_{5.3}$		
dbl exp	1.00	1.00	1.21	1.21	1.01	1.21	1.21	1.19	1.23	1.00	1.38
cont norm	1.00	1.00	1.17	1.21	1.01	1.22	1.27	1.16	1.33	1.00	2.07
t(1)	1.00	1.01	1.39	1.43	1.02	1.52	1.53	1.34	1.50	1.00	∞
t(2)	1.00	1.01	1.22	1.23	1.01	1.25	1.25	1.20	1.28	1.00	∞
t(5)	1.00	1.00	1.09	1.09	1.00	1.09	1.09	1.09	1.11	1.00	1.20
t(8)	1.00	1.00	1.06	1.06	1.00	1.06	1.06	1.06	1.06	1.00	1.10
t(10)	1.00	1.00	1.04	1.05	1.00	1.05	1.05	1.04	1.05	1.00	1.08
t(20)	1.00	1.00	1.02	1.02	1.00	1.02	1.02	1.02	1.02	1.00	1.03
normal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
sym beta	1.00	1.00	0.98	0.98	1.00	0.98	0.98	0.98	0.98	1.00	0.97
$\exp(-x^4)$	1.00	0.99	0.87	0.88	0.99	0.84	0.86	0.87	0.88	1.00	0.86

Table 5.7: Asymptotic Value of the MLE, the MOSME, the Standard One-Step M-Estimator and the τ -Estimator of Dispersion, As Well As the Normalized MAD and the Standard Deviation (SD), for Different Underlying Distributions F . The Initial Estimators of Dispersion and Location Used to Calculate the One-Step M-estimators Are Respectively the Normalized MAD and the Median.

M-estimators are respectively the normalized MAD and the median.

Table 5.7 demonstrates that the term *dispersion of a distribution F* in itself is not clearly defined or even meaningful. Indeed, the standard deviation is commonly used as a measure of dispersion. When the underlying distribution is normal, the standard deviation can easily be interpreted. For example, if the standard deviation of a normal sample is 5, then the dispersion of the sampled normal population is estimated to be 5 times that of the standard normal distribution Φ . Approximately 68% of the sampled values fall within ± 5 of the sample mean. Notice also that the MLE of dispersion in the normal case is exactly the standard deviation. On the other hand, what can be said about the dispersion of a Cauchy or a $t(2)$ family, for which the sample standard deviation is, for example, 5? No matter how spread or dispersed the Cauchy or the $t(2)$

distribution are, their standard deviations always equal infinity. The sample standard deviation cannot therefore be a meaningful measure of dispersion for these families.

As an alternative, one could use the MLE to estimate the parameter σ and refer to it as the dispersion of a distribution. However, calculating the MLE assumes that the distribution F generating the data is known, which is rarely the case in practice. Besides, even though the MLE possesses the analytical interpretation of σ , its geometric interpretation remains obscure. Indeed, remember that the distributions presented in Table 5.7 have been normalized so that their interquartile range equals 1.349. The normalization factor d_0 is now part of the definition of their density ($f(x; \theta, \sigma) = f_{d_0}(x; \theta, \sigma)$) and dispersion estimation is concerned with the problem of identifying the distribution F in the family $\{F : F = F_{d_0}(\frac{x-\theta}{\sigma})\}$, from which the sample is drawn. Therefore, assume for example that the MLE of dispersion from a sample is found to be equal to 5. Does this mean that $d_0 = 5$ and $\sigma = 1$, or rather that $d_0 = 0.5$ and $\sigma = 10$, or some other combination? The first possibility would imply that the distribution of interest has a light-tailed, whereas the second possibility indicates that F is heavy-tailed. Somehow, it is hard to grasp what kind of geometric dispersion the MLE actually estimates.

For a clear geometric interpretation of dispersion, the MAD offers a good deal. It is easy to picture the interquartile range of a sample: it is the range, centered around the middle of the sample, which contains half of the data points. If for example a sample has a normalized MAD of 5, then the dispersion of the sampled population can be estimated as 5 times larger than the dispersion of the same distribution with interquartile range equal to 1.349.

However, maybe because of its ease of interpretation, the MAD does not define the complete picture. Indeed, all the distributions presented in Table 5.7 have the same MAD, even though some F are heavy-tailed, others are light-tailed and one is normal.

Clearly, the centre half of mass of these distributions is contained in the exact same range. But, we are given no indication regarding the shape of their tails. Are they elongated, dispersed, or rather compact, short?

With these questions in mind, one can easily conclude from Table 5.7 that the one-step M-estimators of dispersion do indeed estimate dispersion in its most meaningful way. For any distribution in the table, the estimates of dispersion are finite, and roughly close to 1. The closer the estimates are from 1, the closer the distributions are from normal. All estimates of dispersion over 1 come from heavy-tailed distributions, whereas all estimates under 1 refer to light-tailed distributions. In fact, by their intrinsic definitions, the one-step M-estimators of dispersion are a linear function of the MAD; this function differs more or less from the identity as the underlying distribution differs from the standard normal distribution.

The one-step M-estimators therefore provide information about dispersion which is more complete than the MAD, while staying geometrically interpretable and meaningful. A look at the one-step estimates in Table 5.7 tells us that all distributions have roughly the same interquartile range, but that their nature differ greatly, since the estimates are all close to 1, but still different from each other. Finally, note that computing a one-step M-estimator of dispersion doesn't require the knowledge of the distribution generating the data, since the empirical distribution is used: this is a clear advantage over the MLE, which is defined through the joint density of the data points.

But, which of the MOSME, the standard one-step or the τ -estimator should be used to estimate dispersion? It appears that the MOSME or the τ -estimator represent the best choice, depending on the situation. Both estimators are relatively close to 1, and always more than the standard one-step M-estimator of dispersion, especially in the presence a heavy-tailed distribution. When using the Huber score function, the τ -estimator is

slightly closer to 1 than the MOSME for heavy-tailed distributions. In situations where a smoother score function is used, such as the Tukey score function, the MOSME is closer to 1 than the τ - estimator with heavy-tailed distributions. Any choice of estimator and score function is good for light-tailed distributions. Finally the choice of the Huber score function which is 95% efficient at the normal model seems more appropriate than $H_{0.975}$ score function. $H_{2.376}$ better reflects the nature of the tails of the underlying distribution, by diverging more from 1.

5.2.3 Asymptotic Variance of the MOSME of Dispersion

Appendix C provides the complete derivation of the influence function of the MOSME of dispersion. In the general non-symmetric case, it is equal to

$$\begin{aligned} IF(x; S^*, F) = & IF(x; S_0, F) \left\{ 1 + \frac{E_F \chi(y)}{E_\Phi \chi'(z)z} - \frac{E_F \chi'(y)y}{E_\Phi \chi'(z)z} \right\} - \\ & IF(x; T_0, F) \left\{ \frac{E_F \chi'(y)}{E_\Phi \chi'(z)z} \right\} + \\ & \frac{S_0}{E_\Phi \chi'(z)z} \left\{ \chi\left(\frac{x-T_0}{S_0}\right) - E_F \chi(y) \right\}, \end{aligned}$$

where $T_0 = T_0(F)$, $S_0 = S_0(F)$, $y = \frac{y'-T_0}{S_0}$, and $IF(x; S_0, F)$ and $IF(x; T_0, F)$ are the respective influence functions of the initial estimates of location and dispersion. However, it is possible to simplify the above expression with appropriate conditions.

Proposition 3 *Assume the score function χ is even, bounded and twice differentiable everywhere except in at most a finite number of points. If F is symmetric, then*

$$\begin{aligned} IF(x; S^*, F) = & IF(x; S_0, F) \left\{ 1 + \frac{E_F \chi(y)}{E_\Phi \chi'(z)z} - \frac{E_F \chi'(y)y}{E_\Phi \chi'(z)z} \right\} + \\ & \frac{S_0}{E_\Phi \chi'(z)z} \left\{ \chi\left(\frac{x}{S_0}\right) - E_F \chi(y) \right\}, \end{aligned}$$

where $y = \frac{y'}{S_0(F)}$.

Note that $T_0(F) = 0$ when F is symmetric. The most important condition in Proposition 3 is the symmetry of F , because it makes the MOSME of dispersion independent of the

initial estimator of location. The conditions on the score function χ are minimal regularity conditions that most score functions used in practice will satisfy.

Under mild regularity assumptions, a Taylor series expansion gives an expression for the asymptotic variance of the MOSME of dispersion:

$$V(S^*, F) = E_F\{IF(x; S^*, F)\}^2.$$

In view of Proposition 3, it is clear that the choice of the initial estimator of dispersion, $S_0(F)$, will greatly affect the efficiency of the MOSME of dispersion. The literature strongly recommends the use of the robust normalized MAD

$$S_n = \frac{\text{Med}|x - \text{Med}(x)|}{\Phi^{-1}(3/4)}$$

as an initial estimate of dispersion, which has the following influence function

$$IF(x; S_0, F) = \frac{\text{sgn}(|x| - \Phi^{-1}(3/4))}{4f(\Phi^{-1}(3/4))\Phi^{-1}(3/4)}$$

when F is symmetric and properly scaled so that $S_0(F) = 1$.

The MAD was first promoted by Hampel (1974). Other possibilities for starting estimators of dispersion are presented by Rousseeuw and Croux (1991), which do not assume a symmetric underlying distribution. Indeed, one must note that the MAD is aimed at symmetric distributions and has low gaussian efficiency, but its influence function has the sharpest bound among all possible dispersion estimators for symmetric F , and therefore, the MAD has the lowest possible gross-error sensitivity for symmetric F (see [15], p. 142). Huber (1981) concluded that "the MAD has emerged as the single most useful ancillary estimate of scale" (p. 107). The SHORTH (the shortest half of the data), equivalent to the MAD for symmetric distributions, is also becoming more and more popular because it was found to be approximately minimax bias robust within the class of M-estimators of dispersion with general location (see Martin and Zamar (1993)).

On the other hand, because only symmetric distributions are studied, the initial estimator of location will asymptotically have the value 0 ($T_0(F) = 0$), and therefore, does not affect directly the asymptotic variance of the MOSME of dispersion. However, in practice, because any sample size is finite, one should carefully choose the initial estimate of location. The median is recommended in the literature, because of its robustness properties.

5.2.4 Asymptotic Variance of the Standard One-Step M-Estimator of Dispersion

Appendix D provides the complete derivation of the influence function of the standard one-step M-estimator of dispersion. In the general non-symmetric case, it is equal to

$$\begin{aligned} IF(x; S_1, F) = & IF(x; S_0, F) \left\{ 2 \frac{E_{FX}(y)}{E_{FX}'(y)y} + \frac{E_{FX}(y) E_{FX}''(y)y^2}{[E_{FX}'(y)]^2} \right\} + \\ & IF(x; T_0, F) \left\{ \frac{E_{FX}(y) E_{FX}''(y)y}{[E_{FX}'(y)]^2} + \frac{E_{FX}(y) E_{FX}'(y)}{[E_{FX}'(y)]^2} - \frac{E_{FX}'(y)}{E_{FX}'(y)y} \right\} + \\ & \frac{S_0}{E_{FX}'(y)y} \left\{ \chi\left(\frac{x-T_0}{S_0}\right) - \frac{E_{FX}(y) \chi'\left(\frac{x-T_0}{S_0}\right)\left(\frac{x-T_0}{S_0}\right)}{E_{FX}'(y)y} \right\}, \end{aligned}$$

where $T_0 = T_0(F)$, $S_0 = S_0(F)$, $y = \frac{y' - T_0(F)}{S_0(F)}$ and $IF(x; S_0, F)$ and $IF(x; T_0, F)$ are the respective influence functions of the initial estimators of dispersion and location. However, it is possible to simplify the above expression with appropriate conditions.

Proposition 4 *Assume the score function χ is even, bounded and twice differentiable everywhere except in at most a finite number of points. If F is symmetric, then*

$$\begin{aligned} IF(x; S_1, F) = & IF(x; S_0, F) \left\{ 2 \frac{E_{FX}(y)}{E_{FX}'(y)y} + \frac{E_{FX}(y) E_{FX}''(y)y^2}{[E_{FX}'(y)]^2} \right\} + \\ & \frac{S_0}{E_{FX}'(y)y} \left\{ \chi\left(\frac{x}{S_0}\right) - \frac{E_{FX}(y) \chi'\left(\frac{x}{S_0}\right)\left(\frac{x}{S_0}\right)}{E_{FX}'(y)y} \right\}, \end{aligned}$$

where $y = \frac{y'}{S_0(F)}$.

Notice that $T_0(F) = 0$ when F is symmetric.

The influence function of the standard one-step M-estimator is the sum of two terms, similarly to the influence function of the MOSME: one is a multiple of $IF(x; S_0, F)$ and the other is a multiple of $S_0(F)$. It is obvious that the initial estimator of dispersion holds great importance in both the definition of the MOSME and the standard one-step M-estimator. The influence function for the one-step M-estimator contains a $\chi''(y)$ and many ratios with denominator $E_F \chi'(y)y$, which may cause computational problems. On the other hand, the influence function of the MOSME contains no $\chi''(y)$ and only includes ratios with denominator $E_{\Phi} \chi'(z)z$, which are known to be more stable.

Under mild regularity assumptions, it is possible to express the asymptotic variance of the standard one-step M-estimator as

$$V(S_1, F) = E_F \{IF(x; S_1, F)\}^2.$$

5.2.5 Asymptotic Variance of the τ -Estimator of Dispersion

Appendix E provides the complete derivation of the influence function of the τ -estimator of dispersion, $S_1^p(F)$. In the general non-symmetric case, it is equal to

$$IF(x; S_1^p, F) = \frac{1}{2S_1^p} \{IF(x; S_0, F) [\frac{2}{S_0} (S_1^p)^2 - \frac{S_0}{\beta} E_F \chi'(y)y] \\ - IF(x; T_0, F) \frac{S_0}{\beta} E_F \chi'(y) - (S_1^p)^2 + \frac{S_0^2}{\beta} \chi(\frac{x-T_0}{S_0})\},$$

where $T_0 = T_0(F)$, $S_0 = S_0(F)$, $S_1^p = S_1^p(F)$, $y = \frac{y'-T_0(F)}{S_0(F)}$ and $IF(x; T_0, F)$ and $IF(x; S_0, F)$ are the influence functions of the initial estimators of location and dispersion. However, it is possible to simplify the above expression with appropriate conditions.

Proposition 5 Assume the score function χ is even, bounded and twice differentiable everywhere except in at most a finite number of points. If F is symmetric, then

$$IF(x; S_1^p, F) = IF(x; S_0, F) \left\{ \frac{S_1^p(F)}{S_0(F)} - \frac{S_0(F)}{2S_1^p(F)\beta} E_F \chi'(y) y \right\} + \frac{[S_0(F)]^2}{2S_1^p(F)\beta} \chi\left(\frac{x-T_0(F)}{S_0(F)}\right) - \frac{S_1^p(F)}{2},$$

where $y = \frac{y'}{S_0(F)}$.

Notice that $T_0(F) = 0$ when F is symmetric. The value of the τ -estimator appears in its influence function, contrary to the MOSME or the standard one-step M-estimator of dispersion. The assumption of symmetry for F makes the influence function of the τ -estimator of dispersion with unknown location equal to the influence function of the τ -estimator of dispersion with no location parameter (see [30]). The problem of calculating the influence functions for symmetric F is therefore greatly simplified.

Under mild regularity assumptions, it is thus possible to express the asymptotic variance of the τ -estimator of dispersion as

$$V(S_1^p, F) = E_F \{ IF(x; S_1^p, F) \}^2.$$

5.2.6 Relative Asymptotic Efficiency of the MOSME Compared to That of the One-Step M-Estimator and the τ -Estimator of Dispersion

The asymptotic efficiency of an M-estimator of dispersion is not a good measure of its asymptotic variability, because as shown in section (5.2.2), the asymptotic value of an M-estimator depends strongly on its score function as well as on its type. Huber (1981) (p. 108) suggests rather the use of the *relative asymptotic variance*, $RV(S, F)$, that is, the asymptotic variance of $\sqrt{(n)} \log(S(F_n)/S(F))$, which is defined as

$$RV(S, F) = V(\log S, F) = \frac{V(S, F)}{S^2},$$

where $S = S(F)$. This suggestion is in accordance with Bickel and Lehmann (1976) who observed that the relative variance instead of the variance for dispersion estimators is needed to obtain a natural measure of accuracy.

Let the *relative asymptotic efficiency* of an M-estimator of dispersion be defined as the ratio of the relative asymptotic variance of the MLE over its relative asymptotic variance. The relative asymptotic efficiency of the MOSME will be compared to that of the one-step M-estimator and the τ -estimator of dispersion to assess the asymptotic performance of the MOSME in comparison with that of the other two estimators.

The Fisher information, $I(\sigma)$, for the location-dispersion family of distributions $\{F_{\theta,\sigma} : F_{\theta,\sigma} = F(\frac{x-\theta}{\sigma})\}$, is equal to

$$I(\sigma) = \frac{1}{\sigma^2} E_{F_{0,1}} \left\{ x \frac{f'_{0,1}(x)}{f_{0,1}(x)} + 1 \right\}^2$$

Because the MLE of dispersion is consistent for σ , its asymptotic variance is $V(\text{MLE}, F) = \frac{1}{I(\sigma)}$. Its asymptotic value being 1 for any distribution F , as shown in section (5.2.2), the relative asymptotic variance of the MLE of dispersion is thus $RV(\text{MLE}, F) = \frac{1}{I(1)}$. Table 5.8 gives the relative asymptotic variance of the MLE, $RV(\text{MLE}, F)$, for the different distributions under study.

Finally, Table 5.9 shows the relative asymptotic efficiency of the MOSME, the one-step M-estimator and the τ -estimator of dispersion, for the score functions $H_{0.975}$, $H_{2.376}$, $H_{2.516}$, $T_{3.86}$, and $T_{5.3}$, and different underlying distributions F . The relative asymptotic efficiency of the normalized MAD is also provided, as a mean of comparison, since it is used as the initial dispersion estimator. The asymptotic variance of the normalized MAD is equal to

$$V(\text{MAD}, F) = \frac{1}{16f^2(\Phi^{-1}(3/4))\{\Phi^{-1}(3/4)\}^2}$$

Distribution F	$RV(\text{MLE}, F)$
double exponential	1
contaminated normal	0.5
t(1)	2
t(2)	1.25
t(5)	0.8
t(8)	0.6875
t(10)	0.65
t(20)	0.5750084
normal	0.5
symmetrized beta	0.4210503
$0.55\exp(-x^4)$	0.25

Table 5.8: Relative Asymptotic Variance $RV(\text{MLE}, F)$ of the MLE for Different Underlying Distributions F

Distribution F	MAD	MOSME			Standard One-Step			τ -Estimator	
		$H_{0.975}$	$H_{2.376}$	$T_{3.86}$	$H_{0.975}$	$H_{2.376}$	$T_{3.86}$	$H_{2.516}$	$T_{5.3}$
dbble exp	0.481	0.575	0.873	0.910	0.573	0.534	0.918	0.844	0.935
cont normal	0.336	0.409	0.352	0.311	0.405	0.219	0.211	0.346	0.246
t(1)	0.811	0.947	0.916	0.913	0.884	0.371	0.788	0.902	0.880
t(2)	0.703	0.837	0.959	0.963	0.812	0.585	0.922	0.955	0.929
t(5)	0.534	0.660	0.977	0.993	0.653	0.813	0.987	0.970	0.974
t(8)	0.476	0.597	0.976	0.992	0.593	0.879	0.989	0.970	0.985
t(10)	0.456	0.574	0.977	0.989	0.571	0.899	0.986	0.970	0.987
t(20)	0.413	0.524	0.966	0.975	0.524	0.932	0.973	0.964	0.980
normal	0.368	0.470	0.950	0.947	0.470	0.950	0.946	0.950	0.953
sym beta	0.317	0.410	0.919	0.891	0.410	0.942	0.898	0.920	0.892
$\exp(-x^4)$	0.233	0.328	0.825	0.780	0.335	0.735	0.833	0.841	0.769

Table 5.9: Relative Asymptotic Efficiency of the MOSME, the One-Step M-Estimator and the τ -Estimator of Dispersion Derived from Different Score Functions χ , for Different Underlying Distributions F . The Relative Asymptotic Efficiency of the Initial Estimator of Dispersion, the Normalized MAD, Is Provided for Comparison Purposes. The Median is Used as the Initial Estimator of Location

Many observations can be made from Table 5.9. The relative asymptotic efficiency of the MOSME, the one-step M-estimator and the τ -estimator of dispersion is improved over that of the MAD, their initial estimator of dispersion, for almost all distributions. The very heavy-tailed distributions like $t(1)$ and $t(2)$ seem to cause some problems to the one-step M-estimator with the $H_{2.376}$ score function, and the contaminated normal distribution is in general not so well handled by estimators derived from $H_{2.376}$, $H_{2.516}$, $T_{3.86}$ and $T_{5.3}$ score functions. When compared to the score function $H_{2.376}$, the relative efficiency of the MOSMEs derived from $H_{0.975}$ does not in general show a significant improvement over the relative efficiency of the MAD. The $H_{0.975}$ MOSME's are probably not worth considering, at least in a relative asymptotic efficiency context. However, for very heavy-tailed ($t(1)$, $t(2)$, contaminated normal and double exponential) distributions, the one-step M-estimator derived from the $H_{2.376}$ score function actually performs worse than the one derived from $H_{0.975}$: a sign that sacrificing robustness for efficiency with the standard one-step M-estimator can sometimes be a bad strategy.

The $T_{3.86}$ score function, used either with the MOSME or the one-step M-estimator, appears to be a better or a comparable choice for estimating dispersion, with heavy-tailed distributions. The conclusion is however reversed for the τ -estimator: $H_{2.516}$ generally performs better than $T_{5.3}$ for heavy-tailed distributions (with the exception of the double exponential distribution). However, the MLE beats any of the estimators presented in Table 5.9, and that, for all distributions F , since all relative efficiencies presented are below 1.000. But using the MLE for estimating dispersion requires that the entire sample be drawn from F , with no outlier allowed: this is too restricted a situation in many cases. Besides, the distribution F must be specified for the MLE to be defined.

The relative asymptotic efficiency of the MOSME is higher than that of the one-step M-estimator or the τ -estimator of dispersion for heavy-tailed distributions (t , double

exponential and contaminated normal), and that, for all score functions considered, except for the Tukey score function used with the double exponential distribution. The improvement appears more significant in very heavy-tailed distributions (Cauchy, $t(2)$, double exponential and contaminated normal) used with efficient score functions ($H_{2.376}$, $H_{2.516}$, $T_{3.86}$ and $T_{5.3}$). The improvement by the MOSME over the standard one-step M-estimator is also generally greater than over the τ -estimator.

For the light-tailed distributions symmetrized beta and $\exp(-x^4)$, however, the one-step M-estimator of dispersion performs generally better than the MOSME or the τ -estimator, but not significantly better.

5.3 Maximum Bias of the MOSME of Dispersion

Let X_1, \dots, X_n be a sample from a population with distribution F in the contamination neighborhood

$$V_\epsilon(F_0^{\theta, \sigma}) = \{F : F = (1 - \epsilon)F_0^{\theta, \sigma} + \epsilon H, H \text{ arbitrary distribution}\},$$

where $0 \leq \epsilon < 1/2$. Refer back to section 4.3 for a more detailed description of the nature of this contamination neighborhood.

The worst-case bias of an estimator gives a measure of its asymptotic robustness as a function of the fraction ϵ of the contamination. Since outliers, as well as inliers, can possibly affect the estimator, we need to consider both cases separately. The *explosion bias curve* of the MOSME of dispersion is defined by

$$B_{S^*}^+(\epsilon) = \sup_{F \in V_\epsilon} \frac{S^*(F)}{\sigma}$$

and describes the behaviour of the estimator in presence of outliers. On the other hand, the *implosion bias curve* of the MOSME of dispersion is

$$B_{S^*}^-(\epsilon) = \inf_{F \in V_\epsilon} \frac{S^*(F)}{\sigma}$$

and describes the MOSME in presence of inliers.

Let the distribution F_∞ be obtained when $H = \delta_\infty$, the distribution which puts its total mass at infinity. Let the distribution F_0 be obtained when $H = \delta_0$, the point mass contamination at 0. We shall hereafter concentrate on the normal central distribution $F_0^{\theta, \sigma} = \Phi$ in the neighborhood V_ϵ , as in section 4.3.

Proposition 6

Assume the score function χ is odd and bounded.

Let $S^-(\epsilon) = \inf_{F \in V_\epsilon} S_0(F)$, $S^+(\epsilon) = \sup_{F \in V_\epsilon} S_0(F)$, and $B(\epsilon) = \sup_{F \in V_\epsilon} |T_0(F)|$. Assume $E_\Phi \chi'(z)z > 0$ and assume that we can interchange derivation and integration, that is $\partial/\partial t \{E_\Phi \chi(\frac{x-t}{s})\} = -\frac{1}{s} E_\Phi \chi'(\frac{x-t}{s})$ and $\partial/\partial s \{E_\Phi \chi(\frac{x-t}{s})\} = -\frac{1}{s} E_\Phi \chi'(\frac{x-t}{s})(\frac{x-t}{s})$. If

$$E_\Phi \chi'(\frac{x-t}{s}) \geq 0, \quad (5.30)$$

$$\forall t \text{ in } [0, B(\epsilon)], \text{ for fixed } s \text{ in } [S^-(\epsilon), S^+(\epsilon)]$$

and

$$E_\Phi \chi(\frac{x-t}{s}) - E_\Phi \chi'(\frac{x-t}{s})(\frac{x-t}{s}) \geq \frac{-E_\Phi \chi'(z)z - \epsilon}{1-\epsilon}, \quad (5.31)$$

$$\forall s \text{ in } [S^-(\epsilon), S^+(\epsilon)], \text{ for fixed } t \text{ in } [0, B(\epsilon)],$$

then

$$\sup_{F \in V_\epsilon} S^*(F) = S^*(F_\infty), \quad \text{where } F_\infty = (1 - \epsilon)\Phi + \epsilon\delta_\infty,$$

and

$$\inf_{F \in V_\epsilon} S^*(F) = S^*(F_0), \quad \text{where } F_0 = (1 - \epsilon)\Phi + \epsilon\delta_0.$$

Proof: We will prove the explosion bias result only. The proof for the implosion bias of the MOSME follows the exact same idea.

We clearly always have that $\sup_{F \in V_\epsilon} S^*(F) \geq S^*(F_\infty)$.

Moreover, $\forall \epsilon < 1/2$,

$$\begin{aligned}
\sup_{F \in V_\epsilon} S^*(F) &= \sup_{F \in V_\epsilon} \left\{ S_0(F) + S_0(F) \frac{E_F \chi\left(\frac{x-T_0(F)}{S_0(F)}\right)}{E_\Phi \chi'(x)x} \right\} \\
&= \sup_{F \in \{(1-\epsilon)\Phi + \epsilon H\}} \left\{ S_0(F) + S_0(F) \frac{(1-\epsilon)E_\Phi \chi\left(\frac{x-T_0(F)}{S_0(F)}\right) + \epsilon E_H \chi\left(\frac{x-T_0(F)}{S_0(F)}\right)}{E_\Phi \chi'(x)x} \right\} \\
&\leq \sup_{F \in \{(1-\epsilon)\Phi + \epsilon H\}} \left\{ S_0(F) + S_0(F) \frac{(1-\epsilon)E_\Phi \chi\left(\frac{x-T_0(F)}{S_0(F)}\right) + \epsilon}{E_\Phi \chi'(x)x} \right\} \\
&= \sup_{-B(\epsilon) \leq t \leq B(\epsilon)} \left\{ s + s \frac{(1-\epsilon)E_\Phi \chi\left(\frac{x-t}{s}\right) + \epsilon}{E_\Phi \chi'(x)x} \right\} \\
&\quad S^-(\epsilon) \leq s \leq S^+(\epsilon) \\
&= S^+(\epsilon) + S^+(\epsilon) \frac{(1-\epsilon)E_\Phi \chi\left(\frac{x-B(\epsilon)}{S^+(\epsilon)}\right) + \epsilon}{E_\Phi \chi'(x)x} \\
&= S^*(F_\infty).
\end{aligned}$$

The first equality is simply the definition of $S^*(F)$. By definition of V_ϵ , it is possible to write $E_F \chi\left(\frac{x-T_0(F)}{S_0(F)}\right)$ as the sum of two terms, as states the second equality. Since χ is bounded, we can assume without loss of generality that $\sup_x \chi(x) = 1$. Thus $E_H \chi$ is always less or equal to 1, which gives the third line.

The function which is to be maximized on the third line does not depend on the distribution H anymore. In fact, it can be regarded as a function of two arguments, $T_0(F)$ (or t) and $S_0(F)$ (or s). Following the work of Martin and Zamar in [26], if $\epsilon < 1/2$, then $T_0(F)$ and $S_0(F)$ are bounded as in the fourth line. That is, $S^-(\epsilon) \leq S_0(F) = s \leq S^+(\epsilon)$, and $-B(\epsilon) \leq T_0(F) = t \leq B(\epsilon)$.

Notice that since χ is odd, $E_{\Phi}(\frac{x-t}{s}) = E_{\Phi}(\frac{x+t}{s})$. We can therefore assume without loss of generality that $t \geq 0$. Assuming the conditions (5.30) and (5.31) hold, the function to be maximized in the fourth line is increasing in $T_0(F) = t$, for all fixed $S^-(\epsilon) \leq S_0(F) = s \leq S^+(\epsilon)$, when $0 \leq T_0(F) = t \leq B(\epsilon)$, and is increasing in $S_0(F) = s$, for all fixed $0 \leq T_0(F) = t \leq B(\epsilon)$, when $S^-(\epsilon) \leq S_0(F) = s \leq S^+(\epsilon)$. Therefore, we directly get the fifth line, which is by definition $S^*(F_{\infty})$.

Hence, we have shown that $\sup_{F \in \mathcal{V}_{\epsilon}} S^*(F) = S^*(F_{\infty})$, when the central distribution $F_0^{\theta, \sigma}$ is normal. \square

Analytical derivations, combined with numerical calculations, have shown that the MOSME's with $H_{0.975}$, $H_{2.376}$ and the $T_{3.86}$ score functions satisfy the above conditions (5.30) and (5.31), when the median and the normalized MAD are used as preliminary estimates of location and dispersion.

The conditions were rewritten for the three specific cases of $H_{0.975}$, $H_{2.376}$ and $T_{3.86}$ score functions, and evaluated over a finite and equally-spaced 21x21 grid, covering the range of possible t and s values. For a fixed ϵ , the maximum value of the (normalized) MAD, $S^+(\epsilon)$, is produced by a point mass contamination at infinity, and such contamination also produces the maximum value $B(\epsilon)$ of the location estimator. The minimum value of the (normalized) MAD, $S^-(\epsilon)$, is produced by a point mass contamination at 0, and such contamination also produces the minimum absolute value of the location

estimator, 0 (see Martin and Zamar (1993)). More specifically, the value for $B(\epsilon)$ can be explicitly written as $\Phi^{-1}(\frac{1}{2(1-\epsilon)})$. It is the value $T(F)$ which satisfies (4.16) for the median score function $\chi_{Med}(x) = \text{sgn}(x)$ and $F = F_\infty$. The bounds $S^+(\epsilon)$ and $S^-(\epsilon)$ are the implicit solutions of

$$(1 - \epsilon)[\Phi\{B(\epsilon) - S^+(\epsilon)\Phi^{-1}(3/4)\} + 1 - \Phi\{B(\epsilon) + S^+(\epsilon)\Phi^{-1}(3/4)\}] + \epsilon = 1/2,$$

and

$$(1 - \epsilon)[\Phi\{-S^-(\epsilon)\Phi^{-1}(3/4)\} + 1 - \Phi\{S^-(\epsilon)\Phi^{-1}(3/4)\}] = 1/2.$$

That is, $S^+(\epsilon)$ satisfies $E_{F_\infty}\chi_{MAD}(\frac{x-B(\epsilon)}{S^+(\epsilon)}) = 1/2$ and $S^-(\epsilon)$ satisfies $E_{F_0}\chi_{MAD}(\frac{x}{S^-(\epsilon)}) = 1/2$, where the MAD score function is $\chi_{MAD}(x) = 1/2\{\text{sgn}(|x| - \Phi^{-1}(3/4)) + 1\}$ and $F_0 = (1 - \epsilon)\Phi + \epsilon\delta_0$.

For all $\epsilon < 1/2$ used, the conditions (5.30) and (5.31) were always met. So, even if those conditions seem somewhat restrictive, it is believed that many often used bounded and even score functions χ satisfy them.

Assuming that the score function χ is bounded, we have a lower bound for the explosion bias curve of the standard one-step M-estimator of dispersion, given by $S_1(F_\infty)$, where:

$$S_1(F_\infty) = S^+(\epsilon) + S^+(\epsilon) \frac{(1 - \epsilon)E_\Phi\chi(\frac{x-B(\epsilon)}{S^+(\epsilon)}) + \epsilon\chi(\infty)}{(1 - \epsilon)E_\Phi\chi'(\frac{x-B(\epsilon)}{S^+(\epsilon)})(\frac{x-B(\epsilon)}{S^+(\epsilon)})}.$$

Similarly, it is possible to get an upper bound on the implosion curve of the standard one-step M-estimator of dispersion. It is given by

$$S_1(F_0) = S^-(\epsilon) + S^-(\epsilon) \frac{(1 - \epsilon)E_\Phi\chi(\frac{x}{S^-(\epsilon)}) - \epsilon\chi(0)}{(1 - \epsilon)E_\Phi\chi'(\frac{x}{S^-(\epsilon)})(\frac{x}{S^-(\epsilon)})}.$$

It is also possible to get exact expressions for the explosion and implosion bias curves of the τ -estimator. Rousseeuw and Croux (1993a) gives them in the context of estimation of dispersion with known location. The following proposition improves their results.

Proposition 7 *If $s \rightarrow s^2 E_{\Phi} \chi(\frac{x-t}{s})$ is increasing in the range $[S^-(\epsilon), S^+(\epsilon)]$ for fixed t in $[-B(\epsilon), B(\epsilon)]$, and if $t \rightarrow E_{\Phi} \chi(\frac{x-t}{s})$ is increasing in the range $[0, B(\epsilon)]$ for fixed s in $[S^-(\epsilon), S^+(\epsilon)]$, then*

$$\sup_{F \in F_{\epsilon}} S_1^p(F) = S_1^p(F_{\infty})$$

and

$$\inf_{F \in F_{\epsilon}} S_1^p(F) = S_1^p(F_0).$$

Proof: The proof follows the exact same steps as in Proposition 6. \square

Analytical derivations, combined with numerical calculations similar to those performed for the MOSME of dispersion, have shown that the τ -estimators with $H_{2.516}$ and the $T_{5.3}$ score functions satisfy the above conditions in Proposition 7, when the median and the normalized MAD are used as preliminary estimates of location and dispersion. Note that the condition necessitating $s^2 E_{\Phi} \chi(\frac{x-t}{s})$ to be increasing in s can be explicitly written as $\Phi(t - cs) \geq \Phi(t + cs) - 1$ for the H_c score function, which is clearly always true for any s, t and c .

It is therefore possible to compare, in terms of maximum asymptotic explosion and implosion bias, the MOSME with the standard one-step M-estimator and the τ -estimator of dispersion.

Figures 5.6, 5.7 and 5.8 show the explosion bias curve of the MOSME and a lower bound for the explosion bias curve of the one-step M-estimator of dispersion, derived from the $H_{0.975}$, $H_{2.376}$ and $T_{3.86}$ score functions. In all three cases, the MOSME shows a smaller maximum asymptotic bias than the one-step M-estimator, $\forall \epsilon < 1/2$. The

improvement is however better in the case of the more efficient estimators $H_{2.376}$ and the $T_{3.86}$.

Figure 5.9 shows the explosion curves of the MOSME and the τ -estimator of dispersion for the Huber and Tukey score functions, for small, but realistic, contamination by outliers ($\epsilon < 0.30$). We notice that for small values of ϵ , the MOSME has a lower maximum bias than the τ -estimator. However, the situation is reversed when ϵ gets bigger. As the figure shows, this happens with the Huber score function when $\epsilon \geq 0.075$. (It also happens for the Tukey score function with $\epsilon > 0.47$.) Nevertheless, notice that situations with small fraction of contaminations are representative of many real data set with outliers.

When defining the MOSME or the τ -estimator of dispersion, one needs to choose between different score functions χ . As Figure 5.9 shows, the Huber score function is preferable to the Tukey score function with the MOSME, as well as with the τ -estimator.

Finally, Figure 5.10 shows the implosion curves of the MOSME, the standard one-step M-estimator and the τ -estimator derived from the Huber score function. Figure 5.11 shows the implosion curves of the MOSME, the standard one-step M-estimator and the τ -estimator derived from the Tukey score function. In both Figures, the implosion curve of the normalized MAD is also given for comparison purposes. We notice that for smaller ϵ ($\epsilon < 0.27$ with the Huber score function and $\epsilon < 0.33$ for the Tukey score function), the MOSME outperforms the other estimators. For larger ϵ , the MOSME uniformly beats the τ -estimator, but is outperformed by the standard one-step M-estimator. The latter does not implode at $\epsilon = 0.5$. Depending on the situation, this may or not be desirable: when half the data coincides with a single point, the standard one-step M-estimator does not become 0 (whereas the MOSME, the τ -estimator and the MAD do). Rousseeuw and Croux (1993a) mention that the fully iterated M-estimator derived from the $H_{2.376}$ score function explodes at $\epsilon = 0.9686/5.645 = 0.17$ and implodes at $\epsilon = 1 - 0.17 = 0.83$.

Similarly, the fully iterated M-estimator derived from the $T_{3.86}$ score function explodes at $\epsilon = 0.165$ and implodes at $\epsilon = 1 - 0.165 = 0.835$. At the normal model, the standard one-step M-estimator is consistent for σ and so has the same asymptotic properties as the fully iterated M-estimator of dispersion.

When using a one-step M-estimator of dispersion in a situation of inliers, one needs to decide which score function to select in its definition, so as to minimize its implosion bias. Figure 5.12 shows that the MOSME's, the standard one-step M-estimators and the τ -estimators derived from the Huber and Tukey score function are roughly equivalent in terms of implosion bias for small, but realistic, contamination ($\epsilon < 0.3$). The $H_{2.376}$ MOSME appears however to slightly beat the other estimators. If one expects a large fraction of contamination of inliers ($\epsilon > 0.3$), then the standard one-step M-estimator of dispersion with the $H_{2.376}$ score function would probably be the best choice. As discussed, though, this choice may lead to catastrophic results in the presence of special samples. In that case, the MOSME with the $T_{3.86}$ score function may be a wiser choice.

5.3.1 Further Work to Be Done

The work about asymptotic robustness was developed for one specific case of central distribution $F_0^{\theta, \sigma}$ in the neighborhood of V_ϵ : the normal distribution, Φ . More calculations need to be done for different central distributions $F_0^{\theta, \sigma}$, as for example, the ones used in the asymptotic efficiency section (5.2) of this text.

One could also define the MOSME, the standard one-step M-estimator and the τ -estimator of dispersion using other types of score functions, and compare the different estimators in the hope that one proves better than all the others in most situations.

Finally, it remains to be seen whether the asymptotic properties of the three one-step M-estimators studied in this section reflect on their finite-sample performance for small

and medium sample sizes. This could be assessed with Monte-Carlo simulations, for example.

5.4 Continuation of The Hummingbird Example

To illustrate the use of the MOSME as an alternative to the standard one-step M-estimator or the τ -estimator of dispersion, we shall estimate the dispersion for the distributions of the flying times of the sixteen hummingbirds presented in the previous chapter, section 4.4. Remember that there are four types of hummingbirds: adult females (AF), adult males (AM), junior females (JF) and junior males (JM). Refer to Figure 4.5 for the bar plots of the flying times of all 16 birds.

The researcher who provided the data was interested not only in finding a measure of location within each type of birds, but also in describing more generally the features of the distribution of their flying times, such as for example their dispersion. The use of one-step M-estimators of dispersion represent good alternatives to the popular measure of dispersion: the standard deviation.

Table 5.10 presents the values of the MOSME, the standard one-step M-estimator and the τ -estimator of dispersion, as well as the standard deviation and the normalized MAD, for the sixteen birds, assuming an underlying normal distribution. We notice that contrary to the location setup, the Tukey score function is less conservative in terms of robustness than the Huber score function, since the Tukey estimates are further from the MAD than the Huber estimates. As expected, the $H_{0.975}$ estimates stay very close to the MAD, and do not bring much more light than the MAD: the score function is almost as robust as the MAD. On the other hand, the MOSME and the τ -estimates of dispersion are roughly comparable when using a Huber score function which is 95% efficient at the

Bird	SD	MAD	MOSME of Dispersion			Standard One-Step			τ -Estimator	
			$H_{0.975}$	$H_{2.376}$	$T_{3.86}$	$H_{0.975}$	$H_{2.376}$	$T_{3.86}$	$H_{2.516}$	$T_{5.3}$
AF 1	0.151	0.075	0.079	0.091	0.093	0.079	0.095	0.097	0.090	0.099
AF 2	2.430	0.097	0.096	0.131	0.133	0.095	0.147	0.140	0.127	0.140
AF 3	0.143	0.073	0.073	0.080	0.082	0.073	0.082	0.085	0.080	0.089
AF 4	0.275	0.074	0.076	0.111	0.117	0.077	0.151	0.145	0.106	0.124
AF 5	0.176	0.120	0.124	0.139	0.143	0.125	0.141	0.144	0.138	0.149
AM 1	0.132	0.076	0.083	0.100	0.100	0.083	0.097	0.099	0.097	0.102
AM 2	0.160	0.074	0.078	0.093	0.096	0.078	0.096	0.098	0.092	0.100
AM 3	0.103	0.044	0.048	0.052	0.054	0.047	0.054	0.054	0.052	0.056
AM 4	0.167	0.073	0.072	0.078	0.080	0.072	0.080	0.082	0.079	0.086
AM 5	0.105	0.064	0.066	0.075	0.077	0.066	0.078	0.078	0.075	0.080
JF 1	0.314	0.135	0.144	0.186	0.191	0.143	0.200	0.201	0.180	0.199
JF 2	0.387	0.090	0.094	0.123	0.128	0.095	0.133	0.137	0.119	0.135
JF 3	0.179	0.042	0.042	0.049	0.050	0.042	0.050	0.052	0.049	0.054
JM 1	0.237	0.044	0.048	0.063	0.067	0.048	0.077	0.083	0.061	0.073
JM 2	0.449	0.044	0.043	0.053	0.055	0.043	0.058	0.059	0.053	0.061
JM 3	0.182	0.104	0.101	1.124	0.129	0.100	0.138	0.137	0.124	0.138

Table 5.10: Measures of Dispersion of Flying Times of Four Types of Hummingbirds: Adult Females (AF), Adult Males (AM), Junior Females (JF) and Junior Males (JM). The MAD is the Median Absolute Deviation Multiplied by the Inverse of $\Phi^{-1}(3/4)$.

normal model. When using a Tukey score function with that 95% Gaussian efficiency, the τ -estimator compares favourably with the standard one-step M-estimator of dispersion.

Interpreting those measures of dispersion represents an important issue of dispersion estimation. For example, the flying times of the junior female bird 1 has a high standard deviation, compared with its other measures of dispersion, which is due to a very heavy tail (see Figure 4.5 for a visual assessment of the outlier). The normalized MAD of the data is equal to 0.135; in other words, the interquartile range of this sample is 13.5% that of the normal distribution (which MAD equals 1). Because the values of the different one-step M-estimators for this specific bird are significantly higher than the MAD, we know that the sample is very heavy-tailed, without even looking at Figure 4.5.

The researcher was interested in determining a general measure of dispersion in flying times within the four types of hummingbirds. A robust analysis could be used, but it is beyond the scope of this thesis. However, a simple look at the estimates of dispersion suggests no type of hummingbird has a dispersion in flying times clearly different from the other types.

5.5 Conclusions

The MOSME of dispersion presented in this chapter is in general better than the standard one-step M-estimator of dispersion: it is easier to compute, it has an asymptotic value closer to that of the MLE, it has a higher relative asymptotic efficiency in presence of many heavy-tailed distributions, it has a lower asymptotic explosion bias and a lower implosion bias for small ϵ -contaminations by inliers with central normal distribution. However, if one is concerned mainly about bias, and the data possibly contain a large contamination of inliers but not many repetitions, then the standard one-step M-estimator may be a better estimator of dispersion. In that situation, the choice of the Huber score function would be advisable. For light-tailed underlying distributions, the one-step M-estimator of dispersion performs slightly better in terms of asymptotic relative efficiency than the MOSME (and the τ -estimator). However, heavy-tailed distributions, an attempt to model samples with outliers, are of more interest and importance to robust theory.

On the other hand, the τ -estimator offers some competition to the MOSME, in the sense that it is also easy to calculate, it estimates more accurately dispersion of heavy-tailed distributions when it is derived from the Huber score function, and it presents a lower explosion bias for large contamination with central normal distribution. The

MOSME derived from the Tukey score function has nevertheless a higher relative efficiency than the τ -estimator for heavy-tailed distributions. Moreover, the MOSME outperforms the τ -estimator in terms of explosion bias for small (and more realistic) percentages of contamination, and beats the τ -estimator in the presence of any percentage of inliers.

As the hummingbirds example studied in this chapter indicates, there is similarly no obvious winner in terms of dispersion estimator with finite sample sizes. The behaviour of the MOSME, the standard one-step or the τ -estimator of dispersion compares to any of the other, depending on the score function used. The estimates however provides information about the tails of the distribution generating the data, which the MAD or the standard deviation cannot do.

If one is mainly interested in the accuracy (as in “small bias”) of the estimate of dispersion of normal data with outliers present, and didn’t expect a large ϵ -contamination, then we would recommend the use of the MOSME of dispersion derived from the Huber score function. A large portion of outliers would however call for the τ -estimator with the Huber score function. However, if one was mainly concerned with the precision (as in “high relative efficiency”) of the estimates of dispersion, then the MOSME defined by the Tukey score function would provide the best estimator of dispersion.

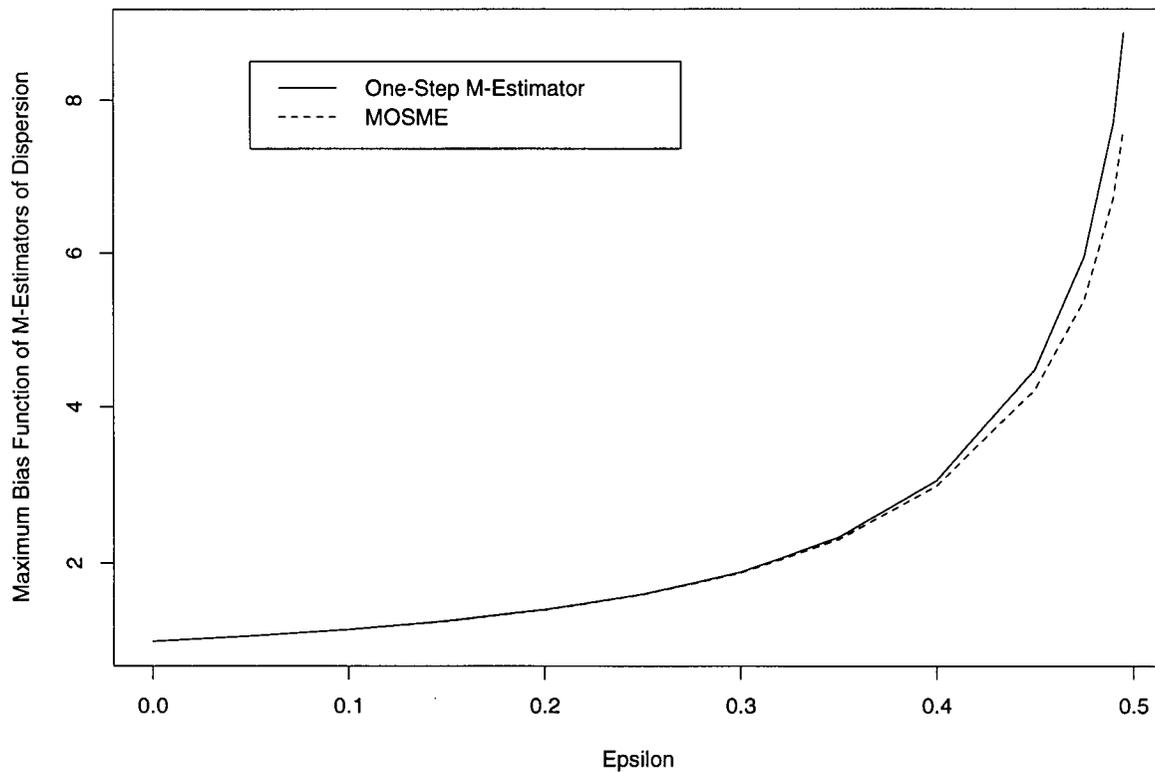
Maximum Bias Function of M-Estimators Using $H_{0.975}$ Score Function

Figure 5.6: Explosion Bias Curve of the MOSME, and a Lower Bound for the Explosion Bias Curve of the One-Step M-Estimator of Dispersion, Derived from the $H_{0.975}$ Score Function

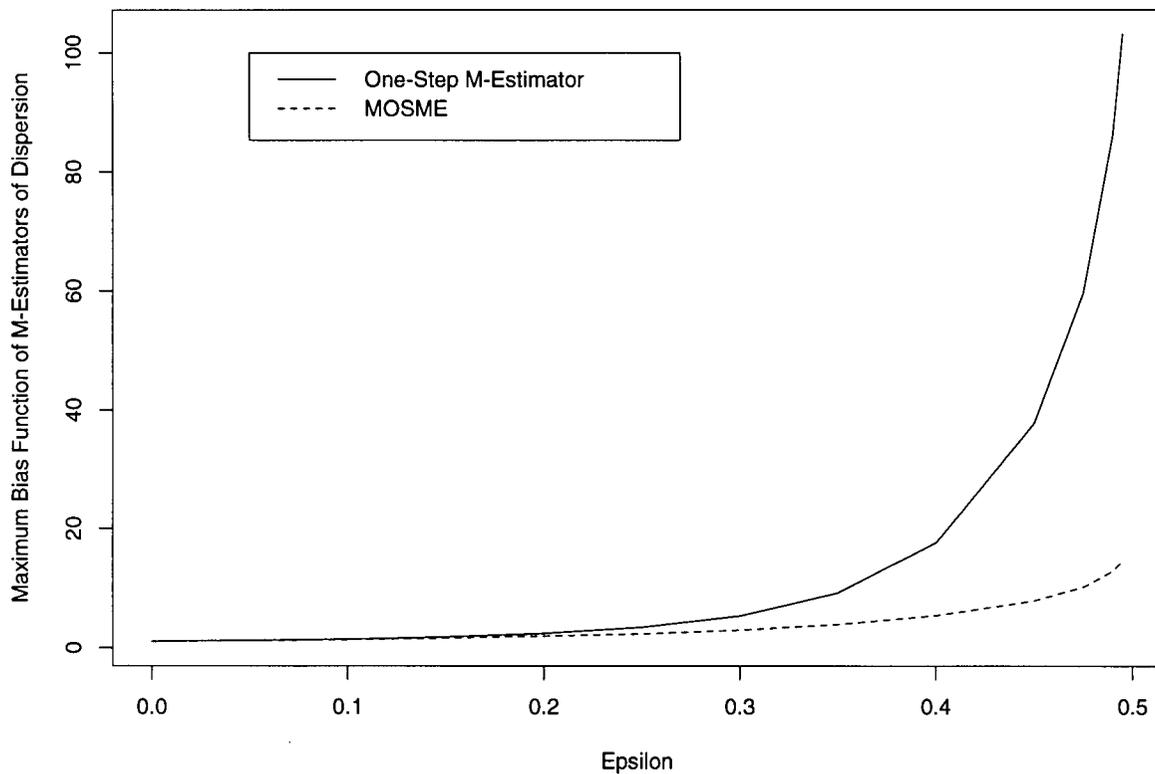
Maximum Bias Function of M-Estimators Using Huber_{2.376} Score Function

Figure 5.7: Explosion Bias Curve of the MOSME, and a Lower Bound for the Explosion Bias Curve of the One-Step M-Estimator of Dispersion, Derived from the $H_{2.376}$ Score Function

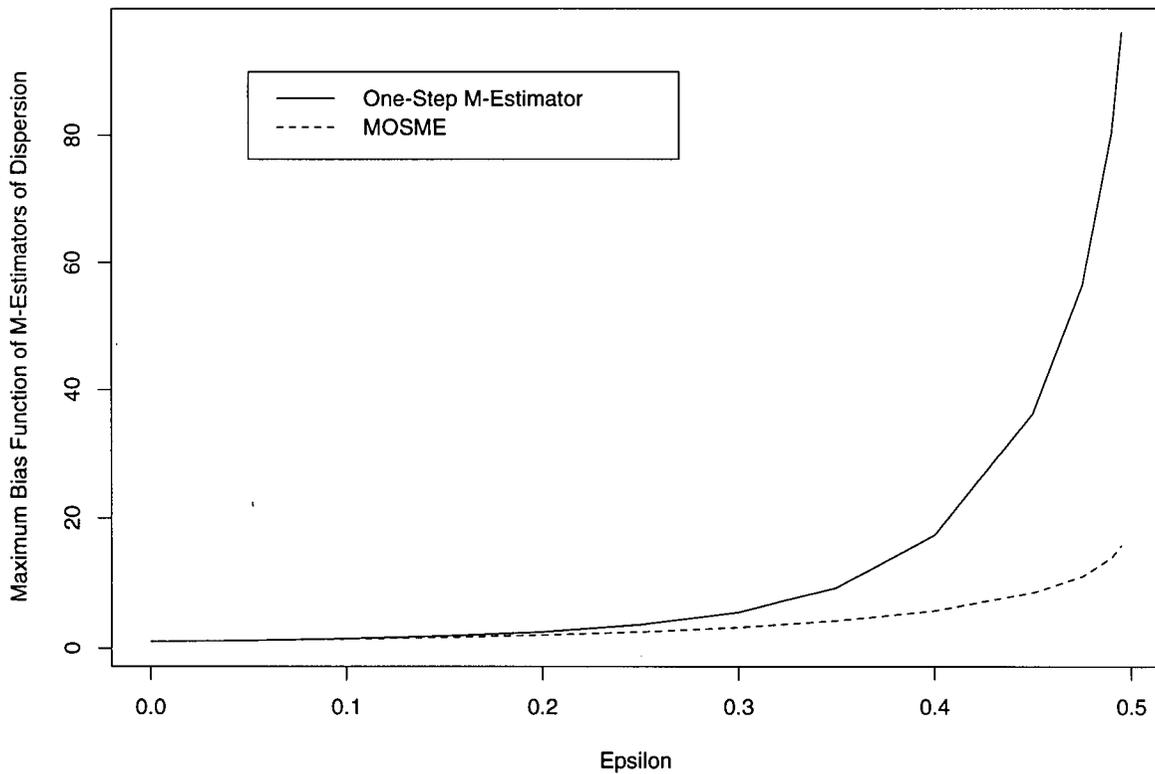
Maximum Bias Function of M-Estimators Using $T_{3.86}$ Score Function

Figure 5.8: Explosion Bias Curve of the MOSME, and a Lower Bound for the Explosion Bias Curve of the One-Step M-Estimator of Dispersion, Derived from the $T_{3.86}$ Score Function

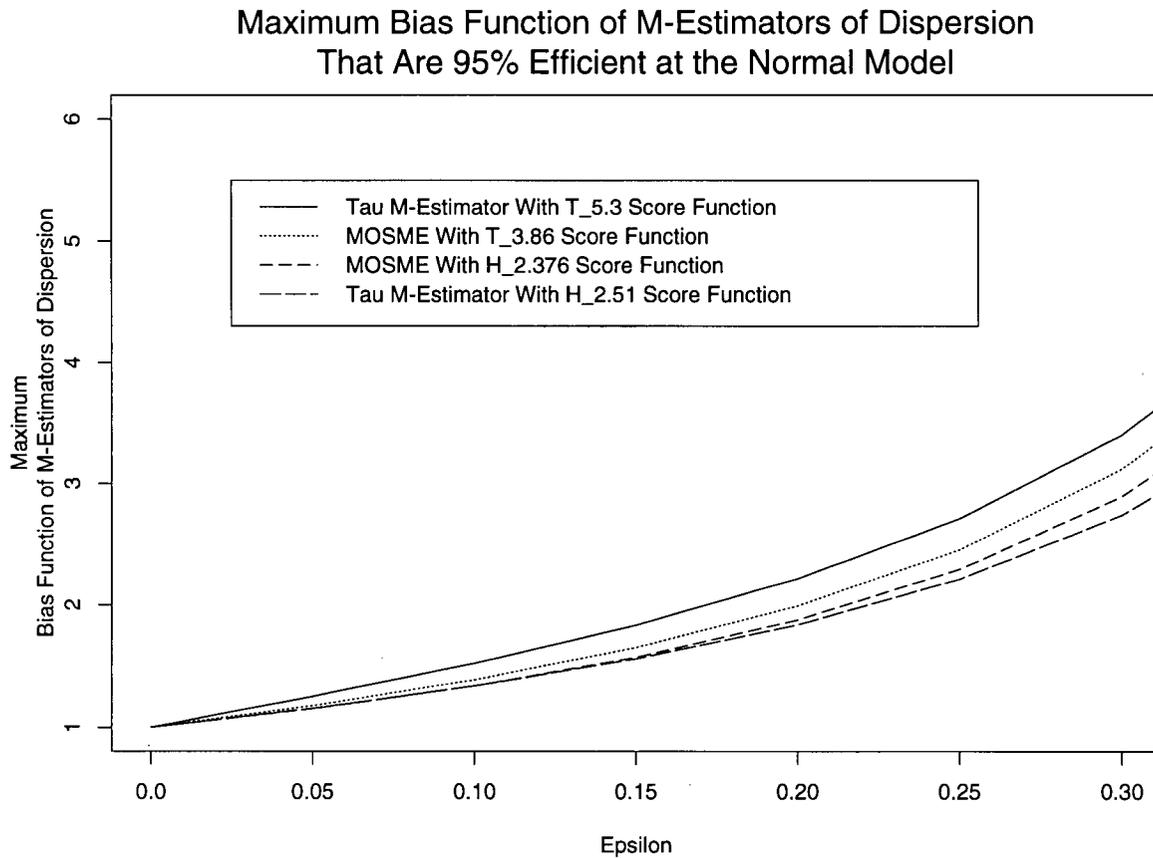


Figure 5.9: Explosion Bias Curves of M-Estimators of Dispersion That Are 95% Efficient at the Normal Model, for Small Contamination by Outliers

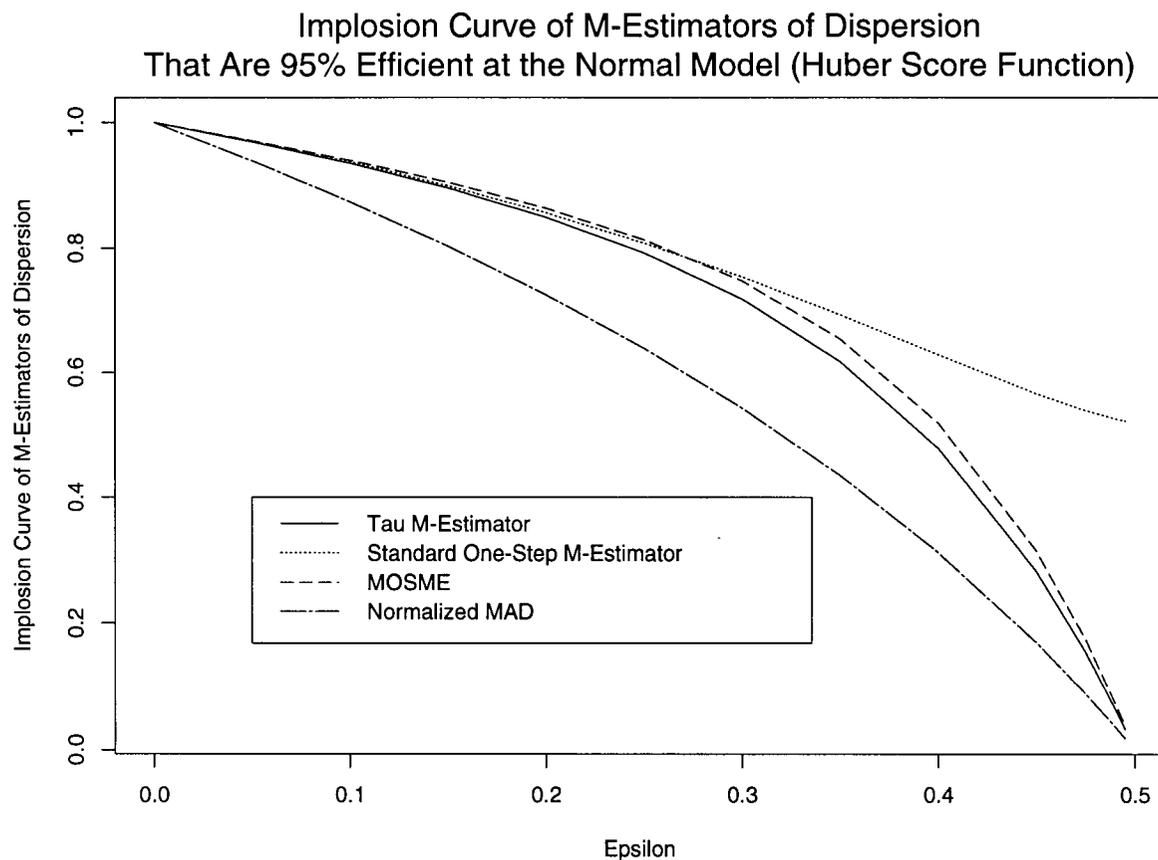


Figure 5.10: Implosion Curve of M-Estimators of Dispersion. The MOSME and the One-Step M-Estimator Are Defined Through The $H_{2.376}$ Score Function, and the Tau-Estimator Have the $H_{2.516}$ Score Function. Those M-Estimators Are 95% Efficient at the Normal Model. The Normalized MAD is Provided for Comparison Purposes.

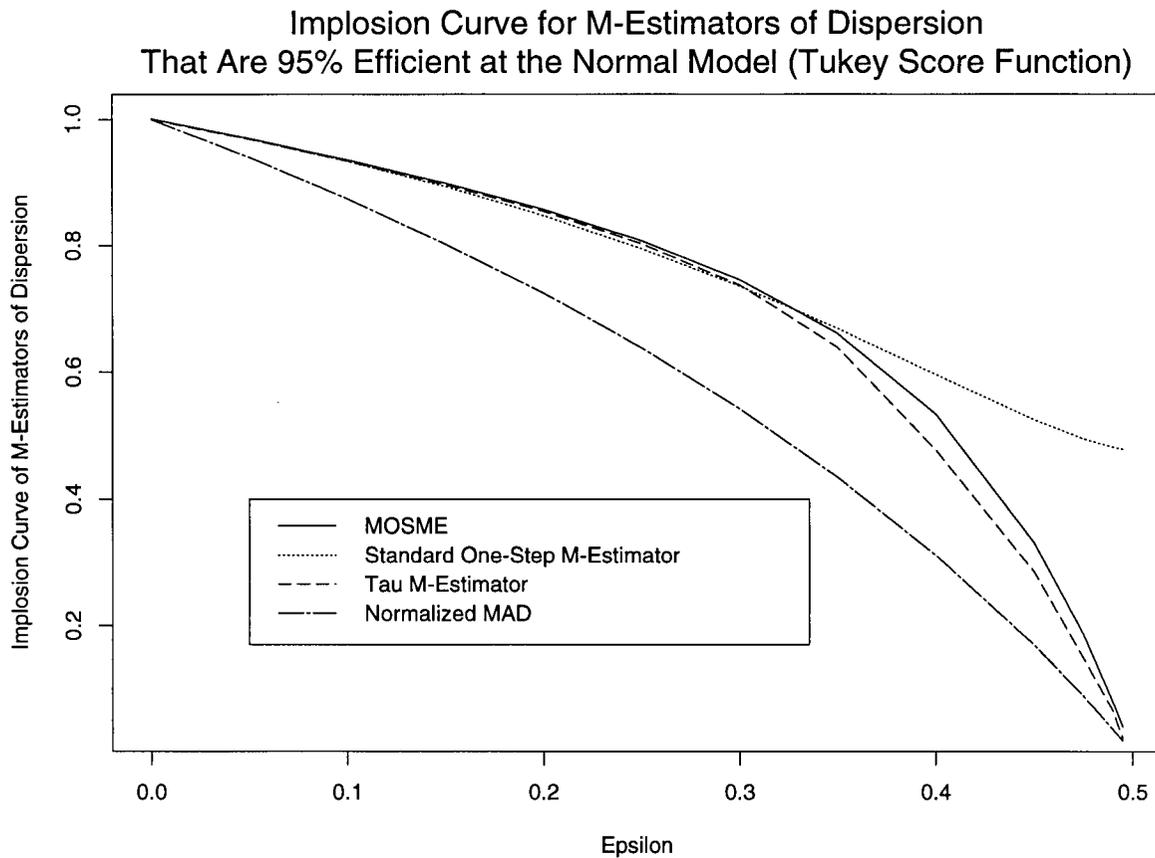


Figure 5.11: Implosion Curve of M-Estimators of Dispersion. The MOSME and the One-Step M-Estimator Are Defined Through the $T_{3.86}$ Score Function, and the Tau-Estimator, Through the $T_{5.3}$ Score Function. Those M-Estimators Are 95% Efficient at the Normal Model. The Normalized MAD is Provided for Comparison Purposes.

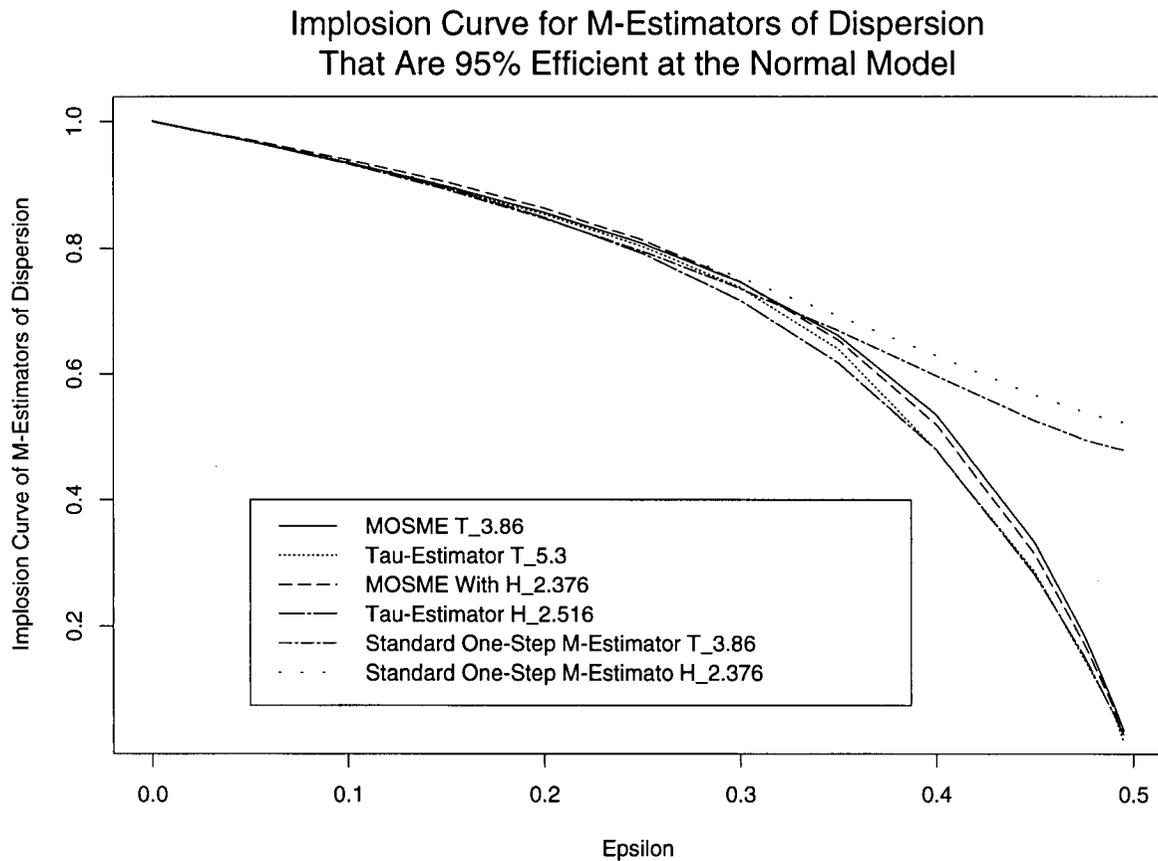


Figure 5.12: Implosion Curve of MOSME's and τ -Estimators That Are 95% Efficient At the Normal Model

Bibliography

- [1] Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, J.W. Tukey and W.H. Rogers (1972). *Robust Estimates of Location: Survey and Advances*, Princeton, NJ: Princeton University Press.
- [2] Bernoulli, D. (1777). *Dijudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inde formand*, *Acta Acad. Sci. Petropolit* **1**, 3-33. (English translation by C. G. Allen (1961). *Biometrika*, **48**, 3-13.)
- [3] Berrendero, J.R. (1995). *A Note on One Step M-Estimates of Location*, Spain: Universidad Carlos III de Madrid (Working Paper).
- [4] Bessel, F.W. and J.J. Baeyer (1838). *Gradmessung in Ostpreussen und ihre Verbindung mit Preussischen und Russischen Dreiecksketten*, Druckerei der Königlichen Akademie der Wissenschaften Berlin. (Reprinted in part in *Abhandlungen von F. W. Bessel*, R. Engelmann (ed.). W. Engelmann, Leipzig, 1876, Vol. 3, pp. 62-138.)
- [5] Bickel, P.J. (1975). *One-Step Huber Estimates In the Linear Model*, *J. Amer. Statist. Ass.* **70**, 428-434.
- [6] Bickel, P.J. and E.L. Lehmann (1976). *Descriptive Statistics for Non-Parametric Models III: Dispersion*, *Ann. Statist.* **4**, 1139-1158.
- [7] Buchanan, J.L. and P.R. Turner (1992). *Numerical Methods and Analysis*, New York: McGraw-Hill, 751 p.
- [8] Casella G. and R.L. Berger (1990). *Statistical Inference*, Belmont, CA: Wadsworth & Brooks/Cole Advanced Books and Softwares, 650 p.
- [9] Donoho, D.L. (1982). *Breakdown Properties of Multivariate Location Estimators*, unpublished manuscript, Harvard University, Dept. of Statistics.
- [10] Donoho, D.L. and P.J. Huber (1983). *The Notion of Breakdown Point*, in *Festschrift für Erich L. Lehmann*, eds. P.J. Bickel, K. Doksum and J.L. Hodges, Jr., Belmont, CA: Wadsworth, 157-184.
- [11] Fisher R.A. (1922). *On the Mathematical Foundations of theoretical Statistics*, reprinted in *Contributions to mathematical Statistics* (1950) by F.J. Wiley, New York.

- [12] Hampel, F.R. (1968). *Contributions to the Theory of Robust Estimation*, Ph. D. Thesis, University of California, Berkeley.
- [13] Hampel, F.R. (1971). *A General Definition of Robustness*, *Ann. Math. Statist.* **42**, 1887-1896.
- [14] Hampel, F.R. (1974). *The Influence Curve and Its Role in Robust Estimation*, *J. Amer. Statist. Ass.* **62** 1179-1186.
- [15] Hampel, F.R., E. M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986). *Robust Statistics - The Approach Based on Influence Functions*, New York: John Wiley & Sons.
- [16] He, X. and D.G. Simpson (1993). *Lower Bounds for Contamination Bias: Globally Minimax Versus Locally Linear Estimation*, *Ann. Statist.* **21**, 314-327.
- [17] Hodges, J.L. Jr. (1967). *Efficiency in Normal Samples and tolerance to Extreme values for Some Estimates of Location*, *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1)*, 163-168.
- [18] Huber, P.J. (1964). *Robust Estimation of a Location Parameter*, *Ann. Math. Statist.* **35**, 73-101.
- [19] Huber, P.J. (1981). *Robust Statistics*, New York: John Wiley & Sons.
- [20] Huber, P.J. (1984). *Finite sample Breakdown of M- and P- Estimators*, *Ann. Statist.* **12**, 119-126.
- [21] Jurečková, J. and S. Portnoy (1987). *Asymptotics for One-Step M-estimators in Regression With Application to Combining Efficiency and High Breakdown Point*, *Comm. Statist. Theory Methods* **16**, 2187-2200.
- [22] Kiefer, J.C. (1987). *introduction to Statistical Inference*, New York: Springer-Verlag.
- [23] Le Cam, L. (1956). *On the Asymptotic Theory of Estimation and testing Hypotheses*, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley: University of California Press, 129-56.
- [24] Lehmann, E.L. (1983). *The Theory of Point Estimation*, New York: John Wiley.
- [25] Martin, R.D. and R.H. Zamar (1989). *Asymptotically Min-Max Bias Robust M-Estimates of Scale for Positive Random Variables*, *Ann. Statist.* **84**, 494-501.
- [26] Martin, R.D. and R.H. Zamar (1993). *Bias Robust Estimation of Scale*, *Ann. Math. Statist.* **21**, 991-1017.

- [27] Martin, R.D., V.J. Yohai and R.H. Zamar (1989). *Min-Max Bias Robust Regression*, *Ann. Statist.* **17**, 1608-1630.
- [28] Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, (1992). *Numerical Recipes in C*, 2nd edition, New York: Cambridge University Press, 994 p.
- [29] Rousseeuw, P.J. (1984). *Least Median of Squares Regression*, *J. Amer. Statis. Assoc.* **79**, 871-880.
- [30] Rousseeuw, P.J. and C. Croux (1991). *Alternatives to the Median Absolute Deviation*, *J. Amer. Statis. Assoc.* **88**, 1273-1283.
- [31] Rousseeuw, P.J. and C. Croux (1993a). *The Bias of k -Step M -estimators*, Belgium: University of Antwerp, UIA, Department of Mathematics & Computer Science (Report no. 93-06).
- [32] Rousseeuw, P.J. and C. Croux (1993b). *Alternatives to the Median Absolute Deviation*, *J. Amer. Statis. Assoc.* **88**, 1273-1283.
- [33] Tukey (1960). *A Survey of Sampling from Contaminated Distributions*, *Contributions to Probability and Statistics*, I. Olkin (ed.). Stanford, CA: Stanford University Press, 448-485.
- [34] Yohai, V.J. and R.H. Zamar (1988). *High Breakdown-Point Estimates of Regression by Means of the Minimization of an Efficient Scale*, *J. Amer. Statis. Assoc.* **83**, 406-413.

Appendix A

Normalization of Densities for Calculations of Asymptotic Efficiencies

Many distributions F are used as examples in the study of asymptotic efficiency of estimators of location and of relative asymptotic efficiency of estimators of dispersion. They are defined as the central distribution in their respective distribution family $\{F(x) : F(x) = F(\frac{x-\theta}{\sigma}), -\infty < x < \infty, -\infty < \theta < \infty, \sigma > 0\}$. The following lists the expression for the density associated with any member of their distribution family:

double exponential : $f_{d_0}(x; \theta, \sigma) = \frac{1}{2d_0\sigma} \exp(-\frac{|x-\theta|}{d_0\sigma})$

contaminated normal : $f_{d_0}(x; \theta, \sigma) = 0.9 \frac{\exp(-\frac{(x-\mu)^2}{2d_0^2\sigma^2})}{\sqrt{2\pi\sigma d_0}} + 0.05 \frac{\exp(-\frac{(x+6-\mu)^2}{2(0.01)d_0^2\sigma^2})}{\sqrt{2(0.01)\pi\sigma d_0}} + 0.05 \frac{\exp(-\frac{(x-6-\mu)^2}{2(0.01)d_0^2\sigma^2})}{\sqrt{2(0.01)\pi\sigma d_0}}$

t(1) : $f_{d_0}(x; \theta, \sigma) = \frac{1}{\pi d_0\sigma} \frac{1}{1 + (\frac{x-\theta}{d_0\sigma})^2}$

t(2) : $f_{d_0}(x; \theta, \sigma) = \frac{\Gamma(3/2)}{\Gamma(1)} \frac{1}{d_0\sigma\sqrt{2\pi}} \frac{1}{\left(1 + \frac{1}{2} \left(\frac{x-\theta}{d_0\sigma}\right)^2\right)^{3/2}}$

t(5) : $f_{d_0}(x; \theta, \sigma) = \frac{\Gamma(3)}{\Gamma(5/2)} \frac{1}{d_0\sigma\sqrt{5\pi}} \frac{1}{\left(1 + \frac{1}{5} \left(\frac{x-\theta}{d_0\sigma}\right)^2\right)^3}$

t(8) : $f_{d_0}(x; \theta, \sigma) = \frac{\Gamma(9/2)}{\Gamma(4)} \frac{1}{d_0\sigma\sqrt{8\pi}} \frac{1}{\left(1 + \frac{1}{8} \left(\frac{x-\theta}{d_0\sigma}\right)^2\right)^{9/2}}$

t(10) : $f_{d_0}(x; \theta, \sigma) = \frac{\Gamma(11/2)}{\Gamma(5)} \frac{1}{d_0\sigma\sqrt{10\pi}} \frac{1}{\left(1 + \frac{1}{10} \left(\frac{x-\theta}{d_0\sigma}\right)^2\right)^{11/2}}$

t(20) : $f_{d_0}(x; \theta, \sigma) = \frac{\Gamma(21/2)}{\Gamma(10)} \frac{1}{d_0\sigma\sqrt{20\pi}} \frac{1}{\left(1 + \frac{1}{20} \left(\frac{x-\theta}{d_0\sigma}\right)^2\right)^{21/2}}$

$$\begin{aligned} \text{normal} & : f_{d_0}(x; \theta, \sigma) = \frac{1}{\sqrt{2\pi d_0 \sigma}} \exp\left(-\frac{1}{2} \left(\frac{x-\theta}{d_0 \sigma}\right)^2\right) \\ \text{symmetrized beta} & : f_{d_0}(x; \theta, \sigma) = \frac{\Gamma(20)}{d_0 \sigma \Gamma(10)^2} \left(-\left(\frac{x-\theta}{d_0 \sigma}\right)^2 + 1/4\right)^9, \\ & \qquad \qquad \qquad \sigma(\theta - 1/2) < x < \sigma(\theta + 1/2) \\ 0.55 \exp(x^4) & : f_{d_0}(x; \theta, \sigma) = \frac{0.5516313254}{d_0 \sigma} \exp\left(-\left(\frac{x-\theta}{d_0 \sigma}\right)^4\right) \end{aligned}$$

Notice that the densities are function of a certain factor d_0 . Indeed, in order to make the different central distributions comparable in terms of spread, it was decided to normalize the associated density in their families by a factor d_0 . The specific choice of d_0 makes the interquartile range of each central distribution equal to the standard normal interquartile range, 1.349. Because these central distributions are symmetric around 0, their interquartile range is equal to $2r$, where r satisfies

$$\int_{-r}^r dF_{d_0}(x; \theta = 0, \sigma = 1) = 0.50. \tag{A.32}$$

By fixing the value of r to 0.6744897502 (which can be obtained by solving the above equation (A.32) with $F_{d_0}(x; \theta, d_0) = \Phi(x)$, where Φ is the standard normal distribution function), it is therefore possible to find the normalizing factor d_0 that satisfies (A.32). Table A.11 shows the factor d_0 necessary to make each of the distributions of interest have an interquartile range equal to 1.349.

In summary, the distributions F used in the calculations of asymptotic efficiency of estimators of location, and of relative asymptotic efficiency of estimators of dispersion are the central distributions in their family, and so, have densities $f_{d_0}(x; \theta = 0, \sigma = 1)$.

Distribution F	Normalization Factor d_0
double exponential	0.9723764576
contaminated normal	0.8820206848
t(1)	0.6744897502
t(2)	0.8254780431
t(5)	0.9274971822
t(8)	0.9541517180
t(10)	0.9631157239
t(20)	0.9811421330
normal	1.0000000000
symmetrized beta	8.8849776417
$0.55\exp(-x^4)$	1.475435073

Table A.11: Normalizing Factor d_0 Needed to Standardize the Interquartile Range of Each Distribution F to That of the Standard Normal Distribution

Appendix B

Derivation of the Influence Function of the MOSME of Location

The influence function of the functional $T(F)$, as defined by Hampel et al in [15] (first introduced by Hampel in 1968), is

$$IF(x; T, F) = \lim_{t \rightarrow 0} \frac{T(F_{t,x}) - T(F)}{t},$$

where $F_{t,x} = (1-t)F + t\delta_x$. In other words, it is defined as the derivative with respect to t of the functional $T(F_{t,x})$, evaluated at $t = 0$.

Therefore, to derive the influence function of the MOSME of dispersion $T^*(F)$, we first need to write the expression for $T^*(F_{t,x})$, and then evaluate its derivative with respect to t at $t = 0$.

The expression for $T^*(F_{t,x})$ is

$$T^*(F_{t,x}) = T_0(F_{t,x}) + S_0(F_{t,x}) \frac{\{(1-t)E_F\psi(\frac{y-T_0(F_{t,x})}{S_0(F_{t,x})}) + t\psi(\frac{x-T_0(F_{t,x})}{S_0(F_{t,x})})\}}{E_{\Phi}\psi'(z)}.$$

Derivating this expression with respect to t , and evaluating it at $t = 0$, gives

$$\begin{aligned} IF(x; T^*, F) &= IF(x; T_0, F) + \frac{IF(x; S_0, F) \cdot E_F\psi(y)}{E_{\Phi}\psi'(z)} + \frac{S_0}{E_{\Phi}\psi'(z)} \{-E_F\psi(y) + \\ &E_F\psi'(y)(\frac{-IF(x; T_0, F)S_0 - (y' - T_0)IF(x; S_0, F)}{S_0^2}) + \psi(\frac{x-T_0}{S_0})\} \\ &= IF(x; T_0, F) \left\{1 - \frac{E_F\psi'(y)}{E_{\Phi}\psi'(z)}\right\} + \\ &IF(x; S_0, F) \left\{\frac{E_F\psi(y)}{E_{\Phi}\psi'(z)} - \frac{E_F\psi'(y)y}{E_{\Phi}\psi'(z)}\right\} + \\ &S_0 \frac{\psi(\frac{x-T_0}{S_0})}{E_{\Phi}\psi'(z)} - S_0 \frac{E_F\psi(y)}{E_{\Phi}\psi'(z)}, \end{aligned} \tag{B.33}$$

where $T_0 = T_0(F)$, $S_0 = S_0(F)$, $y = \frac{y - T_0(F)}{S_0(F)}$, $IF(x; S_0, F)$ and $IF(x; T_0, F)$ are the respective influence functions of the initial estimates of dispersion and of location.

But if T_0 is consistent, then $T_0(F) = \theta = T(F)$, where $T(F)$ is the fully-iterated M-estimator of location θ . Moreover, in that case, $E_F(\frac{y - T_0(F)}{S_0(F)}) = 0$, since by definition $E_F(\frac{y - T(F)}{S_0(F)}) = 0$ and $T_0(F) = T(F)$. If we also assume that ψ is odd and F is symmetric, then $E_F\psi'(\frac{y - T_0(F)}{S_0(F)})(\frac{y - T_0(F)}{S_0(F)}) = 0$, since $T_0(F) = T(F) = \theta$. Therefore, provided that T_0 is consistent, ψ is odd and F is symmetric, equation (B.33) reduces to.

$$IF(x; T^*, F) = IF(x; T_0, F) \left\{ 1 - \frac{E_F\psi'(\frac{y}{S_0(F)})}{E_{\Phi}\psi'(z)} \right\} + S_0(F) \frac{\psi(\frac{x}{S_0(F)})}{E_{\Phi}\psi'(z)} - S_0(F) \frac{E_F\psi(\frac{y}{S_0(F)})}{E_{\Phi}\psi'(z)}.$$

Notice that the middle term on the right-hand side of the above expression can be rewritten as

$$\begin{aligned} S_0(F) \frac{\psi(\frac{x}{S_0(F)})}{E_{\Phi}\psi'(z)} &= \frac{E_F\psi'(\frac{y}{S_0(F)})}{E_{\Phi}\psi'(z)} S_0(F) \frac{\psi(\frac{x}{S_0(F)})}{E_F\psi'(\frac{y}{S_0(F)})} \\ &= \frac{E_F\psi'(\frac{y}{S_0(F)})}{E_{\Phi}\psi'(z)} IF(x; T, F), \end{aligned}$$

where $IF(x; T, F)$ is the influence function of the one-step location M-estimator, which is equivalent to the influence function of the fully-iterated location M-estimator, provided that T_0 is consistent, odd and translation invariant (see [19] for more details). If these last conditions for T_0 hold, and if ψ is odd and F is symmetric, the expression for $IF(x; T^*, F)$ finally simplifies to

$$IF(x; T^*, F) = (1 - a)IF(x; T_0, F) + aIF(x; T, F),$$

where $a = \frac{E_F\psi'(\frac{y}{S_0(F)})}{E_{\Phi}\psi'(z)}$.

Appendix C

Derivation of the Influence Function of the MOSME of Dispersion

As explained in Appendix B, to derive the influence function of the MOSME of dispersion $S^*(F)$, we first need to write the expression for $S^*(F_{t,x})$, and then evaluate its derivative with respect to t at $t = 0$.

The expression for $S^*(F_{t,x})$ is

$$S^*(F_{t,x}) = S_0(F_{t,x}) + \frac{S_0(F_{t,x}) \left\{ (1-t) E_F \chi \left(\frac{y - T_0(F_{t,x})}{S_0(F_{t,x})} \right) + t \chi \left(\frac{x - T_0(F_{t,x})}{S_0(F_{t,x})} \right) \right\}}{E_{\Phi} \chi'(z) z}.$$

Derivating this expression with respect to t , and evaluating it at $t = 0$, gives

$$\begin{aligned} IF(x; S^*, F) &= IF(x; S_0, F) + \frac{1}{E_{\Phi} \chi'(z) z} \left\{ IF(x; S_0, F) E_F \chi(y) + S_0 [-E_F \chi(y) + \right. \\ &\quad \left. E_F (\chi'(y) \left\{ \frac{-IF(x; T_0, F) S_0 - (y' - T_0) IF(x; S_0, F)}{S_0^2} \right\}) + \chi \left(\frac{x - T_0}{S_0} \right)] \right\} \\ &= IF(x; S_0, F) \left\{ 1 + \frac{E_F \chi(y)}{E_{\Phi} \chi'(z) z} - \frac{E_F \chi'(y) y}{E_{\Phi} \chi'(z) z} \right\} + \\ &\quad IF(x; T_0, F) \left\{ \frac{E_F \chi'(y)}{E_{\Phi} \chi'(z) z} \right\} + \frac{S_0}{E_{\Phi} \chi'(z) z} \left\{ \chi \left(\frac{x - T_0}{S_0} \right) - E_F \chi(y) \right\}, \end{aligned} \tag{C.34}$$

where $T_0 = T_0(F)$, $S_0 = S_0(F)$, $y = \frac{y' - T_0(F)}{S_0(F)}$, $IF(x; S_0, F)$ and $IF(x; T_0, F)$ are the respective influence functions of the initial estimates of dispersion and of location.

But since $E_F \chi' \left(\frac{y - T_0(F)}{S_0(F)} \right) = 0$ when χ is even and F is symmetric, and since $T_0(F) = 0$ when F is symmetric, equation (C.34) reduces to

$$\begin{aligned} IF(x; S^*, F) &= IF(x; S_0, F) \left\{ 1 + \frac{E_F \chi \left(\frac{y}{S_0(F)} \right)}{E_{\Phi} \chi'(z) z} - \frac{E_F \chi' \left(\frac{y}{S_0(F)} \right) \left(\frac{y}{S_0(F)} \right)}{E_{\Phi} \chi'(z) z} \right\} + \\ &\quad \frac{S_0(F)}{E_{\Phi} \chi'(z) z} \left\{ \chi \left(\frac{x}{S_0(F)} \right) - E_F \chi \left(\frac{y}{S_0(F)} \right) \right\}. \end{aligned}$$

Appendix D

Derivation of the Influence Function of the One-Step M-Estimator of Dispersion

As explained in Appendix B, to derive the influence function of the one-step M-estimator of dispersion, $S_1(F)$, we first need to write the expression for $S_1(F_{t,x})$, and then evaluate its derivative with respect to t at $t = 0$.

The expression for $S_1(F_{t,x})$ is

$$S_1(F_{t,x}) = S_0(F_{t,x}) + \frac{S_0(F_{t,x}) \left\{ (1-t) E_F \chi \left(\frac{y-T_0(F_{t,x})}{S_0(F_{t,x})} \right) + t \chi \left(\frac{x-T_0(F_{t,x})}{S_0(F_{t,x})} \right) \right\}}{(1-t) E_F \chi' \left(\frac{y-T_0(F_{t,x})}{S_0(F_{t,x})} \right) \left(\frac{y-T_0(F_{t,x})}{S_0(F_{t,x})} \right) + t \chi' \left(\frac{x-T_0(F_{t,x})}{S_0(F_{t,x})} \right) \left(\frac{x-T_0(F_{t,x})}{S_0(F_{t,x})} \right)}$$

Derivating this expression with respect to t (using the derivation formula for a ratio a two functions), and evaluating it at $t = 0$, gives

$$\begin{aligned} IF(x; S_1, F) &= IF(x; S_0, F) + \frac{1}{[E_{FX'}(y)y]^2} \left\{ (IF(x; S_0, F) E_F \chi(y) + \right. \\ & S_0 [-E_F \chi(y) + E_F \chi'(y) \left\{ \frac{-IF(x; T_0, F) S_0 - (y-T_0) IF(x; S_0, F)}{S_0^2} \right\}] E_F \chi'(y) y - S_0 E_F \chi(y) (-E_F \chi'(y) y + \\ & E_F [(\chi''(y) y + \chi'(y)) \left(\frac{-IF(x; T_0, F) S_0 - (y-T_0) IF(x; S_0, F)}{S_0^2} \right)]) + \\ & \left. \chi' \left(\frac{x-T_0}{S_0} \right) \left(\frac{x-T_0}{S_0} \right) \right\} \tag{D.35} \\ &= IF(x; S_0, F) \left\{ 2 \frac{E_{FX}(y)}{E_{FX'}(y)y} + \frac{E_{FX}(y) E_{FX}''(y)y^2}{[E_{FX'}(y)y]^2} \right\} + \\ & IF(x; T_0, F) \left\{ \frac{E_{FX}(y) E_{FX}''(y)y}{[E_{FX'}(y)y]^2} + \frac{E_{FX}(y) E_{FX}'(y)}{[E_{FX'}(y)y]^2} - \frac{E_{FX}'(y)}{E_{FX'}(y)y} \right\} + \\ & E_F \chi(y) \left\{ -\frac{S_0 \chi' \left(\frac{x-T_0}{S_0} \right) \left(\frac{x-T_0}{S_0} \right)}{[E_{FX'}(y)y]^2} \right\} + \frac{S_0 \chi \left(\frac{x-T_0}{S_0} \right)}{E_{FX'}(y)y}, \end{aligned}$$

where $T_0 = T_0(F)$, $S_0 = S_0(F)$, $y = \frac{y - T_0(F)}{S_0(F)}$, $IF(x; S_0, F)$ and $IF(x; T_0, F)$ are the respective influence functions of the initial estimates of dispersion and of location.

But since $E_F \chi'(\frac{y - T_0(F)}{S_0(F)}) = 0$ and $E_F \chi''(\frac{y - T_0(F)}{S_0(F)}) = 0$ when χ is even and F is symmetric, and since $T_0(F) = 0$ when F is symmetric, equation (D.35) reduces to

$$IF(x; S_1, F) = IF(x; S_0, F) \left\{ 2 \frac{E_{FX}(\frac{y}{S_0(F)})}{E_{FX}'(\frac{y}{S_0(F)}) (\frac{y}{S_0(F)})} + \frac{E_{FX}(\frac{y}{S_0(F)}) E_{FX}''(\frac{y}{S_0(F)}) (\frac{y}{S_0(F)})^2}{[E_{FX}'(\frac{y}{S_0(F)}) (\frac{y}{S_0(F)})]^2} \right\} + \frac{S_0(F)}{E_{FX}'(\frac{y}{S_0(F)}) (\frac{y}{S_0(F)})} \left\{ \chi\left(\frac{x}{S_0(F)}\right) - \frac{\chi'(\frac{x}{S_0(F)}) (\frac{x}{S_0(F)}) E_{FX}(\frac{y}{S_0(F)})}{E_{FX}'(\frac{y}{S_0(F)}) (\frac{y}{S_0(F)})} \right\}.$$

Appendix E

Derivation of the Influence Function of the τ -Estimator of Dispersion

As explained in Appendix B, to derive the influence function of the τ -estimator of dispersion, $S_1^p(F)$, we first need to write the expression for $S_1^p(F_{t,x})$, and then evaluate its derivative with respect to t at $t = 0$.

The expression for $S_1^p(F_{t,x})$ is

$$S_1^p(F_{t,x}) = \left\{ \frac{(S_0(F_{t,x}))^2}{\beta} \left[(1-t) E_F \chi \left(\frac{y - T_0(F_{t,x})}{S_0(F_{t,x})} \right) + t \chi \left(\frac{x - T_0(F_{t,x})}{S_0(F_{t,x})} \right) \right] \right\}^{1/2},$$

where $\beta = E_{\Phi} \chi(x)$.

Derivating this expression with respect to t , and evaluating it at $t = 0$ gives

$$\begin{aligned} IF(x; S_1^p, F) &= \frac{1}{2\sqrt{\frac{S_0^2}{\beta} E_F \chi(y)}} \left\{ \frac{2S_0 IF(x; S_0, F)}{\beta} E_F \chi(y) + \right. \\ &\quad \left. \frac{S_0^2}{\beta} [-E_F \chi(y) + E_F \chi'(y) \left(\frac{-S_0 IF(x; T_0, F) - (y - T_0) IF(x; S_0, F)}{S_0^2} \right) + \chi \left(\frac{x - T_0}{S_0} \right)] \right\} \\ &= \frac{1}{2S_1^p} \left\{ IF(x; S_0, F) \left[\frac{2}{S_0} (S_1^p)^2 - \frac{S_0}{\beta} E_F \chi'(y) y \right] - \right. \\ &\quad \left. IF(x; T_0, F) \frac{S_0}{\beta} E_F \chi'(y) - (S_1^p)^2 + \frac{S_0^2}{\beta} \chi \left(\frac{x - T_0}{S_0} \right) \right\}, \end{aligned} \tag{E.36}$$

where $T_0 = T_0(F)$, $S_0 = S_0(F)$, $S_1^p = S_1^p(F)$, $y = \frac{y' - T_0}{S_0}$, $IF(x; T_0, F)$ and $IF(x; S_0, F)$ are the respective influence functions of the initial estimators of location and dispersion.

But since $E_F \chi'(y) = 0$ when χ is even and F is symmetric, and since $T_0(F) = 0$ when F is symmetric, $y = \frac{y'}{S_0(F)}$ and equation (E.36) reduces to

$$IF(x; S_1^p, F) = IF(x; S_0, F) \left\{ \frac{S_1^p(F)}{S_0(F)} - \frac{S_0(F)}{2S_1^p(F)\beta} E_F \chi'(y) y \right\} + \frac{[S_0(F)]^2}{2S_1^p(F)\beta} \chi \left(\frac{x - T_0(F)}{S_0(F)} \right) - \frac{S_1^p(F)}{2}.$$