Uniform asymptotics for robust location estimates when the scale is unknown

Matias Salibian-Barrera - Carleton University
Ruben H. Zamar - University of British Columbia

June 10, 2002

Abstract

Most asymptotic results for robust estimates rely on regularity conditions that are difficult to verify and that real data sets rarely satisfy. Moreover, these results apply to fixed distribution functions. In the robustness context the distribution of the data remains largely unspecified and hence results that hold uniformly over a set of possible distribution functions are of theoretical and practical interest. In this paper we study the problem of obtaining verifiable and realistic conditions that suffice to obtain uniform consistency and uniform asymptotic normality for location robust estimates when the scale of the errors is unknown. We study M-location estimates calculated with an S-scale and we obtain uniform asymptotic results over contamination neighbourhoods. There is a trade-off between the size of these neighbourhoods and the breakdown point of the scale estimate. We also show how to calculate the maximum size of the contamination neighbourhoods where these uniform results hold.

1 Introduction

Many robust point estimates have been proposed in the last 35 years. Unfortunately, robust inference has not received the same amount of attention in the literature. Since the finite sample distributions of robust estimates are unknown, robust inference typically relies on the asymptotic distributions of these estimates. To construct a satisfactory asymptotic inference theory based on robust estimates we need estimates that:

W.1 are *translation and scale-equivariant*;

W.2 have *high breakdown point* and *high efficiency* when the data are not contaminated;

W.3 are computable with an algorithm that is known to converge under weak regularity conditions;

W.4 have an asymptotic theory that requires *verifiable* and *realistic* regularity assumptions, and

W.5 have asymptotic properties that hold *uniformly* over a relatively large set of distribution functions with *known* size.

There are many asymptotic results available in the literature. However these results are not completely satisfactory and difficult to apply. Typical regularity conditions include: (i) the assumption of symmetry of the distribution of the errors (see for example Bickel, 1975; Maronna and

Yohai, 1981; Huber, 1981; Simpson *et. al*, 1992; Simpson and Yohai, 1998); (ii) the knowledge of the scale of the errors (Huber, 1964; Markatou and Hettmansperger, 1990) or (iii) some conditions that involve the expected value of the estimating equations under the unknown distribution of the data (Huber, 1981). It is clear that (i), (ii) and (iii) violate W.4 above.

Since according to the robustness model one does not know the actual distribution of the data one needs *asymptotic results that hold uniformly* over some set of plausible distributions. Lacking such uniformity makes it impossible, for example, to determine the sample size needed for an acceptable normal approximation for a given data set.

The first reference in the robustness literature to asymptotic distribution results that hold uniformly on a certain set of distribution functions is Huber (1981, pg. 51). See also Fraiman *et al*. (2001). Huber shows that when *the scale of the errors is known* the M-location estimates are asymptotically normal and the approximation is uniform on *the set of symmetric distributions* that have all their mass concentrated on the points where the estimating equation is differentiable. Huber results apply to estimates that do not satisfy W.1 and the resulting asymptotic results violate W.4 and W.5 above.

Hampel (1971) showed that under certain regularity conditions, M-location estimates have uniform asymptotic properties on Prokhorov neighbourhoods. Unfortunately his results apply to non-scale-equivariant estimates and they only guarantee the existence of a neighbourhood with unknown size. In other words, this class of estimates does not satisfy W.1 and the asymptotic results violate W.5.

More recently Davies (1998) constructed M-location estimates with simultaneous scale estimates (Huber's Proposal 2) that are locally asymptotically normal. Davies's results are "locally uniform", that is, for each distribution function there exists a neighbourhood of distributions where the convergence holds uniformly. Unfortunately, the size of these neighbourhoods is unknown, and consequently these results fail W.5. It is also known that simultaneous location-scale estimates do not satisfy W.2 and W.3. Failure to satisfy W.3 (illustrated in Example 1 below) is particularly troubling.

Example 1 *To illustrate the difficulty in calculating simultaneous location and scale estimates, consider the following sample of 10 numbers: 0.67, -0.73, -0.30, 0.55, 0.62, -0.99, 0.45, 10.22, 9.94, and 10.02. There are 3 outliers. We tried to calculate simultaneous location-scale estimates that solve*

$$\frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{x_i-\hat{\mu}_n}{\hat{\sigma}_n}\right)=0\,, \tag{1}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[\chi\left(\frac{x_i-\hat{\mu}_n}{\hat{\sigma}_n}\right)-\frac{1}{2}\right]=0\,, \tag{2}$$

whith $\psi_c(u) = \min(c, \max(-c, u))$ *and* $\chi_d(u) = (u/d)^2$ *for* $|u| \le d$ *and* $\chi_d(u) = 1$ *otherwise. We used* $c = 1.345$ *and* $d = 1.04$ *which corresponds to a scale estimate with 50% breakdown point and a location estimate with 95% efficiency if the errors are normally distributed. We considered two algorithms to solve the above system of equations: the usual Newton-Raphson iterations with initial values* $\mu_0 = \text{median}(x_1, \ldots, x_{10})$ *and* $\sigma_0 = \text{mad}(x_1, \ldots, x_{10})$*, and the following scheme:*

S.1 Let $\mu_0 = \text{median}(x_1, \ldots, x_{10})$*,* $\sigma_0 = \text{mad}(x_1, \ldots, x_{10})$*, and* $i = 0$*;*

S.2 solve (1) for $\hat{\mu}_n$ *with* $\hat{\sigma}_n = \sigma_i$*; let* $\mu_{i+1} = \hat{\mu}_n$*;*

S.3 solve (2) for $\hat{\sigma}_n$ with $\hat{\mu}_n = \mu_{i+1}$ as calculated above; let $\sigma_{i+1} = \hat{\sigma}_n$;

S.4 $i = i + 1$ and repeat from step S.2.

It is easy to see that the Newton-Raphson iterations fail to converge because the matrix of first derivatives becomes non-singular after 7 iterations. The above algorithm however converges to $\hat{\mu}_n = 3.05$ and $\hat{\sigma}_n = 5.53$. But these results are not reliable as can be seen from the following simple exercise. Replace the last 3 observations x_8, x_9 and x_{10} by $x_8 + 30 = 40.22$, $x_9 + 30 = 39.94$ and $x_{10} + 30 = 40.02$. The new limit values are $\hat{\mu}_n = 12.05$ and $\hat{\sigma}_n = 21.63$ which indicate that these "robust" estimates are very sensitive to the outliers in the data. In other words, simultaneous location-scale estimates have serious computational problems and consequently we will concentrate on M-estimates calculated with an auxiliary scale. The MM-location estimates proposed below in this paper give $\hat{\mu}_n = 0.76$ and $\hat{\sigma}_n = 1.22$ for both data sets in this example.

A referee cited the work by Clarke (2000) where it is shown that certain M-location estimates are continuous over full Prokhorov neighbourhoods of the parametric model. It follows that these estimates have uniform asymptotic behaviour over these Prokhorov neighbourhoods. Unfortunately, the class of estimates considered are not scale-equivariant (i.e. they fail W.1), and as in Hampel (1971), only the existence of a neighbourhood of unknown size is shown (i.e. they also fail W.5).

Our results apply to location M-estimates calculated using an S-scale (see Rousseeuw and Yohai, 1984). In this paper we show that these estimates satisfy all the desired properties listed above. In particular, these estimates are scale-equivariant (W.1), have simultaneous high breakdown point and high efficiency at the central model (W.2) and can be easily calculated (W.3). Moreover, we show that under realistic and verifiable regularity conditions (W.4) we obtain uniform asymptotic results (consistency and asymptotic distribution) that hold over a contamination neighbourhood of known size (W.5). We find that the size of these sets depends on the breakdown of the S-scale estimates (the higher the breakdown point the smaller the set of distribution functions where uniformity holds, see Table 1).

Note that the regularity conditions we need in our results depend on two separate aspects of the inference procedure: the parametric model assumed to hold for the "good" data points, and the estimating equations used to calculate the robust estimate. These conditions are verifiable because they do not depend on the unknown distribution of the data. We shall show that a well-known class of estimating equations (namely, scale-equivariant M-estimates calculated with an S-scale) satisfy all our conditions (W.1 to W.5). Moreover, our assumptions do not interfere with the robustness properties of the resulting estimates that can attain simultaneous high breakdown point and high efficiency at the central model.

The rest of the paper is organized as follows. Section 2 contains the definitions of the estimates we consider. Section 3 shows that under mild regularity conditions these estimates are uniformly consistent on contamination neighbourhoods. Section 4 gives additional assumptions under which the above estimates are uniformly asymptotically normal on contamination neighbourhoods. Section 5 contains some concluding remarks and Section 6 contains sketches of the proofs of our main results.

2 MM-location estimates

Consider the following location-scale model: let $x_1, \ldots, x_n$ be n observations on the real line satisfying

$$x_i = \mu + \sigma\, \epsilon_i \qquad i = 1, \ldots n, \tag{3}$$

where ϵ_i, $i = 1, \ldots n$ are independent and identically distributed (i.i.d.) observations with variance equal to 1. The interest is in estimating μ and the scale σ is considered a nuisance parameter.

We will consider *scale-equivariant* M-location estimates $\hat{\mu}_n$ defined as the solution of an estimating equation of the form

$$\frac{1}{n} \sum_{i=1}^{n} \psi\left((x_i - \hat{\mu}_n) / \, \hat{\sigma}_n \right) = 0 \, ; \tag{4}$$

where $\hat{\sigma}_n$ is an S-scale estimate of the residuals (Rousseeuw and Yohai, 1984) and $\psi : \mathbb{R} \to \mathbb{R}$ is a non-decreasing, odd and continuously differentiable real function. An example of such a function is given by

$$\psi_c\,(u) = \text{sign}\,(u) \begin{cases} |\, u / \, c| & \text{if} \quad |u| \leq \; 0.8 \; c \\ p_4\,(|u|\,/c) & \text{if} \quad 0.8 \; c \; < |u| \leq c \\ p_4\,(1) & \text{if} \quad |u| > c \end{cases} \, , \tag{5}$$

where $c > 0$ is a user-chosed tuning constant, and $p_4\,(u) = 38.4 - 175\, u + 300\, u^2 - 225\, u^3 + 62.5\, u^4$ (see Fraiman *et al.* (2001), and also Bednarski and Zontek (1996), for other choices of smooth functions ψ). Following Yohai (1987) we will call these M-location estimates obtained with an S-scale *MM-location* estimates.

The S-scale estimates $\hat{\sigma}_n$ we use in (4) are defined as follows. Let $\rho : \mathbb{R} \to \mathbb{R}_+$ be a bounded, continuous and even function satisfying $\rho\,(0) = 0$ and let $b \in (0, 1)$. The S-scale $\hat{\sigma}_n$ is defined by

$$\hat{\sigma}_n = \inf_{t \in \mathbb{R}} s_n\,(t) \;\; , \tag{6}$$

where, for each $t \in \mathbb{R}$, $s_n\,(t)$ is the solution of

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left((x_i - t) / \, s_n\,(t) \right) = b. \tag{7}$$

Naturally associated with this family are the *S-location* estimates $\tilde{\mu}_n$ given by

$$\tilde{\mu}_n = \arg \inf_{t \in \mathbb{R}} s_n\,(t) \, . \tag{8}$$

Beaton and Tukey (1974) proposed a family of functions ρ_d given by

$$\rho_d\,(u) = \begin{cases} 3\,(u / \, d)^2 - 3\,(u / \, d)^4 + (u / \, d)^6 & \text{if } |u| \leq d \, , \\ 1 & \text{if } |u| > d \, , \end{cases} \tag{9}$$

where the tuning constant d is positive. The above family of functions ρ_d satisfies all the regularity conditions we need to obtain uniform asymptotic properties, and at the same time it yields scale estimates $\hat{\sigma}_n$ with good robustness properties.

Remark 1 – $\boldsymbol{\psi \neq \rho'}$ – *Note that the estimating function ψ in (4) need not be equal to ρ' in (7). Moreover, we will recommend using $\psi = \psi_c$ in (5) and $\rho = \rho_d$ in (9).*

Remark 2 – **High efficiency and breakdown point** – *The robust location estimates* $\hat{\mu}_n$ *defined by (4) with* $\hat{\sigma}_n$ *as in (6) are scale equivariant and can have simultaneously high breakdown and high efficiency at the central model. For example, the choice* $d = 1.548$ *for* ρ_d *in (9),* $b = 0.5$ *in (7), and* $c = 1.525$ *for* ψ_c *in (5) yields a location estimate* $\hat{\mu}_n$ *with 50% breakdown point and 95% efficiency when the errors have a normal distribution.*

The asymptotic properties (consistency and asymptotic normality) of M-location estimates given by (4) are well-known when the distribution of the errors is symmetric (Huber, 1964, 1967, 1981; Boos and Serfling, 1980; Clarke, 1983, 1984). The next two sections establish these properties under more realistic conditions.

3 Uniform consistency

The objective of this section is to determine verifiable conditions under which the *scale equivariant* M-location estimates $\hat{\mu}_n$ given by (4) are uniformly consistent on the contamination "neighbourhood"

$$\mathcal{H}_\epsilon\,(F_0) = \left\{ F \in \mathcal{D} \,:\, F\,(x) = (1-\epsilon)\ F_0\,(\,(x-\mu_0)/\,\sigma_0) + \epsilon\, H\,(x)\, \right\}, \tag{10}$$

where $\mathcal{D}$ denotes the set of all distribution functions, F_0 is a fixed symmetric distribution, μ_0 and σ_0 are the unknown location and scale parameters, $\epsilon \in (0, 1/2)$, and H is an arbitrary distribution function. Since in what follows the central distribution F_0 is fixed, we write $\mathcal{H}_\epsilon$ to denote the set (10) above.

Under certain regularity conditions (see references above) the M-location estimates $\hat{\mu}_n$ and the S- estimates $\hat{\sigma}_n$ and $\tilde{\mu}_n$ are consistent to the functionals $\boldsymbol{\mu}\,(F)$, $\boldsymbol{\sigma}\,(F)$ and $\tilde{\boldsymbol{\mu}}\,(F)$ defined by the following equations. For each $t \in \mathbb{R}$, let $\sigma\,(F, t)$ satisfy

$$E_F\,[\ \rho\,(\ (X-t)\,/\ \sigma\,(F,t)\,)\ ] = b. \tag{11}$$

The asymptotic value of $\hat{\sigma}_n$ is given by

$$\boldsymbol{\sigma}\,(F) = \inf_{t\in\mathbb{R}} \sigma\,(F,t)\,. \tag{12}$$

Similarly, for the S-location estimate $\tilde{\mu}_n$ we have

$$\tilde{\boldsymbol{\mu}}\,(F) = \arg\inf_{t\in\mathbb{R}} \sigma\,(F,t)\,. \tag{13}$$

Finally for the M-location estimate $\hat{\mu}_n$ the corresponding equation is

$$E_F\,[\ \psi\,(\ (X-\boldsymbol{\mu}\,(F))\,/\ \boldsymbol{\sigma}\,(F)\ )\ ] = 0\,. \tag{14}$$

Definition 1 – **Uniform consistency** – *We say that the sequence of estimates* $\hat{\tau}_n$ *is uniformly consistent to the functional* $\tau\,(F)$ *over the contamination neighbourhood* $\mathcal{H}_\epsilon$ *if for all* $\delta > 0$

$$\lim_{m\to\infty} \sup_{F\in\mathcal{H}_\epsilon} P_F \left[\sup_{n\geq m} |\hat{\tau}_n - \tau\,(F)| > \delta \right] = 0\,,$$

where $\tau\,(F)$ *is the a.s. limit of* $\hat{\tau}_n$ *for an i.i.d. sequence of observations with distribution function* F*. We will denote this type of convergence by* $\hat{\tau}_n \xrightarrow{\epsilon} \tau$*.*

Our main result in this section states that if the scale estimate $\hat{\sigma}_n$ in (4) satisfies $\hat{\sigma}_n \xrightarrow{\epsilon} \sigma$ and if ψ is continuously differentiable then $\hat{\mu}_n \xrightarrow{\epsilon} \mu$.

Theorem 1 - Uniform consistency of the M-location estimate with general scale: *Let $x_1, \ldots, x_n$ be i.i.d. observations following the location model (3). Let ψ satisfy*

P.1 $|\psi(u)| \le 1$ *for all* $u \in \mathbb{R}$, *and* $\psi(-u) = -\psi(u)$ *for* $u \ge 0$;

P.2 ψ *is non-decreasing and* $\lim_{u\to\infty} \psi(u) > 0$;

P.3 ψ *is continuously differentiable.*

Suppose that $\hat{\sigma}_n$ in (4) has asymptotic breakdown point ϵ^*. Let $0 \le \epsilon < \epsilon^*$ be such that $\hat{\sigma}_n \xrightarrow{\epsilon} \sigma$, then if $\hat{\mu}_n$ satisfies (4) we have $\hat{\mu}_n \xrightarrow{\epsilon} \mu$.

A sketch of the proof of Theorem 1 is given in the Appendix. A detailed proof can be found in Anonymous (2002).

Remark 3 – Uniform consistency of S-scale estimates – *When $\hat{\sigma}_n$ is an S-scale estimate, Martin and Zamar (1993) showed that if F_0 (the central distribution function in $\mathcal{H}_\epsilon$) has an even and unimodal density, and if the function ρ is even, bounded, continous and non-decreasing in $[0, \infty)$ then $\hat{\sigma}_n$ has asymptotic breakdown point 1/2. They also showed that if in addition F_0 has a positive density on the real line, then for all $0 < \epsilon < 1/2$ we have*

$$\hat{\sigma}_n \xrightarrow{\epsilon} \sigma \, . \tag{15}$$

Theorem 1 and Remark 3 imply that M-location estimates $\hat{\mu}_n$ given by (4) with $\psi = \psi_c$ in the family (5) and scale $\hat{\sigma}_n$ given by (6) with $\rho = \rho_d$ in Tukey's family (9) have high breakdown point, high efficiency and are uniformly consistent over $\mathcal{H}_\epsilon$ for all $0 < \epsilon < 1/2$. In other words, these estimates satisfy W.1, W.2 and W.3 in Section 1. Moreover, their uniform consistency satisfies W.4 since we only need regularity conditions on the central distribution of the contamination neighbourhood. Finally, this uniform consistency is valid over contamination neighbourhoods $\mathcal{H}_\epsilon$ for any $0 \le \epsilon < 1/2$ (W.5).

4 Uniform asymptotic distribution

In this section we show that under certain regularity conditions the MM-location estimates $\hat{\mu}_n$ converge weakly to a normal distribution uniformly over the contamination neighbourhood $\mathcal{H}_\epsilon$. These results are *constructive* and allow us to determine the size of the neighbourhood $\mathcal{H}_\epsilon$ where uniform asymptotic normality holds. The required regularity conditions will be mainly imposed on our estimating equations (4) and (6) and we will show that $\psi = \psi_c$ in (5) and $\rho = \rho_d$ in (9) satisfy these conditions. Hence, our results show that the *scale equivariant* MM-location estimates have simultaneously high breakdown point, high efficiency at the central model and are uniformly asymptotically normal on a contamination neighbourhood of known size (see Remark 2 on page 5).

Asymptotic results for *asymmetric* distributions are not easy to obtain. There are some results in the robustness literature dealing with this problem (Carroll, 1978, 1979; Carroll and Welsh, 1988; Rocke and Downs, 1981). They show that when F is asymmetric the asymptotic distribution of the location estimate depends on that of the scale and that the asymptotic variance calculated with the assumption of symmetry is not correct. Salibian-Barrera (2000) showed that in general the asymptotic distribution of location M-estimates for arbitrary distribution functions when the scale is estimated with an S-scale depends on the behaviour of the S-scale and the corresponding S-location estimate as well. Hence, to obtain uniform asymptotics for these MM-location estimates we need uniform consistency of the *S-scale* and *S-location estimates*.

S-scale estimates are uniformly consistent under relatively weak regularity conditions (see Martin and Zamar (1993) and Remark 3 on page 6).

Uniform consistency of *S-location* estimates requires more assumptions. For a given $0 \leq \epsilon < 1/2$ and an estimating function ρ in (7) let s^+ and s^- satisfy

$$0 < s^- \leq \inf_{F \in \mathcal{H}_\epsilon} \boldsymbol{\sigma}(F) < \sup_{F \in \mathcal{H}_\epsilon} \boldsymbol{\sigma}(F) \leq s^+ < \infty \,. \tag{16}$$

To simplify the notation we will omit the dependence of s^+ and s^- on ϵ. Assume that there exists $t^* \in \mathbb{R}$ such that

$$\inf_{s^- \leq s \leq s^+} \left[E_{F_0} \rho\left(\frac{X-t}{s}\right) - E_{F_0} \rho\left(\frac{X}{s}\right) \right] > \frac{\epsilon}{1-\epsilon}, \qquad \forall\, |t| \geq t^* \,, \tag{17}$$

and

$$\inf_{\substack{-t^* \leq t \leq t^* \\ s^- \leq s \leq s^+}} E_{F_0} \rho''\left(\frac{X-t}{s}\right) > \frac{\epsilon}{1-\epsilon} \sup_x \left[\rho''(x)\right]^- \,, \tag{18}$$

where s^+ and s^- are given in (16).

Condition (18) can be slightly relaxed (see Lemma 7 in Section 6). Assumptions (17) and (18) above do not depend on F (only on F_0, the central distribution of the neighbourhood $\mathcal{H}_\epsilon$) but are tedious to verify and will typically require numerical computations. Note that for a particular ρ these conditions impose an upper bound $\epsilon = \epsilon(\rho)$ on the size of the contamination neighbourhood $\mathcal{H}_\epsilon$. When $\rho = \rho_d$ belongs to Tukey's family (9) and the centre of the contamination neighbourhood is the standard normal distribution Φ we found that there is a trade-off between the breakdown point of the scale estimate and the upper bound $\epsilon(\rho_d)$: the larger the breakdown point the smaller the upper bound $\epsilon(\rho_d)$. Table 1 lists the values of $\epsilon(\rho_d)$ for contamination neighbourhoods of the standard normal distribution and estimating equations that yield estimates with breakdown points between 0.10 and 0.50.

The following theorem states than under these conditions S-location estimates are uniformly consistent. This result will be necessary to obtain uniform asymptotic distribution of the M-location estimate calculated with an S-scale as in (4).

Theorem 2 - Uniform consistency of the S-location estimate*: Suppose that the non-constant function ρ satisfies the following assumptions:*

R.1 $\rho(-u) = \rho(u)$, $u \geq 0$, *and* $\sup_{u \in \mathbb{R}} \rho(u) = 1$*;*

R.2 $\rho(u)$ *is non-decreasing in* $u \geq 0$*;*

R.3 $|\rho'(u)| \leq K < \infty$, $\forall\, u \in \mathbb{R}$*;*

R.4 there exists $0 < c < \infty$ *such that* $\rho(u) = 1$, $\forall\, |u| \geq c$.

Let $b \in (0,1)$, $\tilde{\mu}_n$ *as in (8) and* $\tilde{\boldsymbol{\mu}}(F)$ *as in (13). Let* s^+ *and* s^- *be as in (16) and suppose that* $0 < \epsilon$ *is such that (17) and (18) hold. Then*

$$\lim_{m \to \infty} \sup_{F \in \mathcal{H}_\epsilon} P_F\Big(\sup_{n \geq m} |\tilde{\mu}_n - \boldsymbol{\mu}(F)| > \delta \Big) = 0 \,. \tag{19}$$

BP	d	$\epsilon(d)$
0.50	1.548	0.11
0.45	1.756	0.14
0.40	1.988	0.17
0.35	2.252	0.20
0.30	2.561	0.24
0.25	2.937	0.25

Table 1: Maximum size $\epsilon(d)$ of contamination neighbourhoods around the standard normal distribution where uniform consistency of the *S-location* estimate holds for different breakdown points (BP). The column labeled d contains the tuning constant that yields the respective BP.

A sketch of the proof of Theorem 2 is given in the Appendix. A detailed proof can be found in Anonymous (2002).

We can now state our main result: when the M-location, S-scale and S-location estimates are uniformly consistent, the M-location estimate has an uniformly asymptotically normal distribution.

Theorem 3 *Let $\hat{\mu}_n$ satisfy (4) with a function ψ that satisfies assumptions P.1 and P.2 in Theorem 1 and*

P.4 ψ is twice continuously differentiable; and

P.5 there exists $d > 0$ such that $|\psi(u)| = 1$ for all $|u| \geq d$.

Assume that the S-scale estimate $\hat{\sigma}_n$ in (4) is given by (6) with a function ρ that satisfies R.1 to R.4 in Theorem 2, and

R.5 ρ is twice continuously differentiable.

Suppose that ϵ is such that (17) and (18) hold and that the centre F_0 of the contamination neighbourhood $\mathcal{H}_\epsilon$ has a positive, even and unimodal density. Then

$$\lim_{n\to\infty} \sup_{F\in\mathcal{H}_\epsilon} \sup_{x\in\mathbb{R}} \left| P_F \left\{ \sqrt{n}\, \frac{(\hat{\mu}_n - \mu)}{\sqrt{V}} < x \right\} - \Phi(x) \right| = 0\,,$$

where

$$V = V(\mu, \sigma, F) = \sigma(F)^2 H(F)^2 E_F \left\{ \left[\psi\left(\frac{X - \mu(F)}{\sigma(F)} \right) - J(F) \times \left(\rho\left(\frac{X - \tilde{\mu}(F)}{\sigma(F)} \right) - b \right) \right]^2 \right\}, \quad (20)$$

$$H(F) = 1/E_F \left\{ \psi'\left((X - \mu(F))/\sigma(F) \right) \right\},$$

and

$$J(F) = \frac{E_F\left\{\psi'\left((X - \mu(F))/\sigma(F)\right)(X - \mu(F))/\sigma(F)\right\}}{E_F\left\{\rho'\left((X - \tilde{\mu}(F))/\sigma(F)\right)(X - \tilde{\mu}(F))/\sigma(F)\right\}}\,.$$

A sketch of the proof of Theorem 3 is given in the Appendix. A detailed proof can be found in Anonymous (2002).

Remark 4 – **Regularity conditions** – *The assumptions on F_0 (the centre of the contamination neighbourhood) are needed to show that the S-scale estimate $\hat{\sigma}_n$ is uniformly consistent ($\hat{\sigma}_n \stackrel{\epsilon}{\to} \sigma$). By Theorem 1 we also have that the MM-location estimates are uniformly consistent as well ($\hat{\mu}_n \stackrel{\epsilon}{\to} \mu$). The assumptions on the estimating equation ρ of the S-scale $\hat{\sigma}_n$ and conditions (17) and (18) are needed to obtain uniform consistency of the S-location estimate ($\tilde{\mu}_n \stackrel{\epsilon}{\to} \tilde{\mu}$). See Theorem 2.*

Using Table 1 we find, for example, that scale-equivariant MM-location estimates calculated with $\psi = \psi_{1.525}$ in (5) and an S-scale with $\rho = \rho_{1.548}$ in (9) have simultaneously breakdown point 1/2, are 95% efficiency when the errors are normally distributed, and are uniformly asymptotically normal on a contamination neighbourhood of size at least $\epsilon = 0.11$. If, on the other hand, we use $\rho = \rho_{2.937}$ in (9) we obtain estimates with the same efficiency, lower breakdown point (25%) and that are uniformly asymptotically normal on a contamination neighbourhood of size $\epsilon = 0.25$.

5 Conclusion

We have examined the available asymptotic results for robust location estimates and highlighted their limitations: they apply to estimates that are not scale-equivariant, or to robust estimates that have numerical and theoretical problems; they rely on assumptions which are unrealistic and/or difficult to verify; they are not known to be uniform on a reasonably large set of possible distributions. We identified three key features of robust estimates: translation and scale-equivariance, high breakdown point and efficiency, and a reliable algorithm to compute them. We also indicated two important properties their asymptotic theory should satisfy: be valid under verifiable and realistic regularity assumptions, and hold uniformly over a relatively large set of distribution functions with known size. All the previously available asymptotic results for robust location estimates either violate at least one of the above properties, or they apply to estimates that are not scale-equivariant or that have serious computational limitations (see Example 1).

We propose to use scale-equivariant M-location estimates calculated with a smooth function ψ in the family (5) and with an S-scale estimate calculated with a function ρ in Tukey's class (9). These MM-location estimates have simultaneously high breakdown point and high efficiency at the central model. Moreover, we showed that under realistic and verifiable conditions they are *uniformly consistent* and *uniformly asymptotically normal*. We also showed how to compute the size of the contamination neighbourhood (10) where these uniform results hold. For contamination neighbourhoods centred at the standard normal distribution we found that these values of ϵ range from 11% (for estimates with 50% breakdown point) to 25% (for 25% breakdown point estimates). Hence, in most practical situations where the contamination is below 10% (Hampel, 1986) these estimates have good robustness properties and their uniform asymptotic properties allow for reliable statistical inference based on their asymptotic distribution.

6 Proofs

Proof of Theorem 1: For any $t \in \mathbb{R}$ and $F \in \mathcal{H}_\epsilon$ let

$$\mu_\psi(t, F) = E_F \psi\left(\frac{X - t}{\boldsymbol{\sigma}(F)}\right),$$

and fix an arbitrary $\tilde{\epsilon} > 0$.

Let $\boldsymbol{\sigma} = \boldsymbol{\sigma}(F)$ and $\boldsymbol{\mu} = \boldsymbol{\mu}(F)$. To simplify the notation let $\psi(X, t, s) = \psi((X - t)/s)$. For each t it is easy to see that $Y_i(t) = \psi(X_i, t, \hat{\sigma}_n)$ and $Y(F, t) = E_F \psi(X, t, \boldsymbol{\sigma})$ have the same properties as those in Lemma 6. Let $\overline{\psi}_n(t) = \frac{1}{n}\sum_{i=1}^n Y_i(t)$ and $\mu_\psi(t, F) = E_F(\psi(X, t, \boldsymbol{\sigma}))$. For each $\tau > 0$ and $t \in \mathbb{R}$, the same technique used in the proof of Lemma 6 shows that

$$\lim_{m\to\infty} \sup_{F\in\mathcal{H}_\epsilon} P_F\Big(\sup_{n\geq m} \left|\overline{\psi}_n(t) - \mu_\psi(t, F)\right| > \tau \Big) = 0 \,. \tag{21}$$

For each $m \in \mathbb{N}$, $t \in \mathbb{R}$, $F \in \mathcal{H}_\epsilon$ and $\tau > 0$ let

$$\mathcal{A}_m(F, t, \tau) = \left\{ \sup_{n\geq m} \left| \overline{\psi}_n(t) - \mu_\psi(t, F) \right| > \tau \right\} ;$$

then (21) can be written as

$$\lim_{m\to\infty} \sup_{F\in\mathcal{H}_\epsilon} P_F\Big(\mathcal{A}_m(F, t, \tau) \Big) = 0 \,. \tag{22}$$

Now note that $\mu_\psi(\boldsymbol{\mu}(F), F) = 0$ and that $\mu_\psi(t, F)$ is a non-increasing function in t. We also have

$$\begin{aligned}
\left\{ \hat{\mu}_n < \boldsymbol{\mu} - \tilde{\epsilon} \right\} &\subseteq \left\{ \frac{1}{n}\sum_{i=1}^n \psi(x_i, \boldsymbol{\mu} - \tilde{\epsilon}/2, \hat{\sigma}_n) \leq 0 \right\} \\
&\subseteq \left\{ \left| \overline{\psi}_n(\boldsymbol{\mu} - \tilde{\epsilon}/2) - \mu_\psi(\boldsymbol{\mu} - \tilde{\epsilon}/2, F) \right| > \mu_\psi(\boldsymbol{\mu} - \tilde{\epsilon}/2, F) \right\} \\
&\subseteq \left\{ \left| \overline{\psi}_n(\boldsymbol{\mu} - \tilde{\epsilon}/2) - \mu_\psi(\boldsymbol{\mu} - \tilde{\epsilon}/2, F) \right| > a(\tilde{\epsilon}) \right\} = A_n(F, \tilde{\epsilon}) \ ,
\end{aligned}$$

where $a(\tilde{\epsilon})$ is given by

$$a(\tilde{\epsilon}) = \inf_{F\in\mathcal{H}_\epsilon} \mu_\psi\big(\boldsymbol{\mu}(F) - \tilde{\epsilon}/2, F \big) \,.$$

Similarly

$$\begin{aligned}
\left\{ \hat{\mu}_n > \boldsymbol{\mu} + \tilde{\epsilon} \right\} &\subseteq \left\{ \frac{1}{n}\sum_{i=1}^n \psi(x_i, \boldsymbol{\mu} + \tilde{\epsilon}/2, \hat{\sigma}_n) \geq 0 \right\} \\
&\subseteq \left\{ \left| \overline{\psi}_n(\boldsymbol{\mu} + \tilde{\epsilon}/2) - \mu_\psi(\boldsymbol{\mu} + \tilde{\epsilon}/2, F) \right| > -\mu_\psi(\boldsymbol{\mu} + \tilde{\epsilon}/2, F) \right\} \\
&\subseteq \left\{ \left| \overline{\psi}_n(\boldsymbol{\mu} - \tilde{\epsilon}/2) - \mu_\psi(\boldsymbol{\mu} - \tilde{\epsilon}/2, F) \right| > b(\tilde{\epsilon}) \right\} = B_n(F, \tilde{\epsilon}) \ ,
\end{aligned}$$

where $b(\tilde{\epsilon})$ equals

$$b(\tilde{\epsilon}) = \inf_{F\in\mathcal{H}_\epsilon} -\mu_\psi\big(\boldsymbol{\mu}(F) + \tilde{\epsilon}/2, F \big) \,.$$

We now show that

$$a(\tilde{\epsilon}) = \inf_{F\in\mathcal{H}_\epsilon} \mu_\psi\big(\boldsymbol{\mu}(F) - \tilde{\epsilon}/2, F \big) > 0 \,, \tag{23}$$

and that

$$b(\tilde{\epsilon}) = \inf_{F\in\mathcal{H}_\epsilon} -\mu_\psi\big(\boldsymbol{\mu}(F) + \tilde{\epsilon}/2, F \big) > 0 \,. \tag{24}$$

Equations (23) and (24) can be expressed as: the family of functions $\mu_\psi(t, F)$ has "uniform minimum slope" at $\boldsymbol{\mu}(F)$. Bounding $\partial\mu_\psi/\partial t|$ uniformly over $F \in \mathcal{H}_\epsilon$ will be enough for these conditions to hold. Let $\lambda_F(\delta)$ be

$$\lambda_F(\delta) = E_F \psi\left(\frac{X - \boldsymbol{\mu}(F) + \delta}{\boldsymbol{\sigma}(F)}\right),$$

then $a(\tilde{\epsilon}) = \inf_{F \in \mathcal{H}_\epsilon} \lambda_F(\tilde{\epsilon})$. Note that $\lambda_F(0) = 0$; hence

$$\lambda_F(\tilde{\epsilon}) = \tilde{\epsilon}\, \lambda'_F(\tilde{\epsilon}_F),$$

where $\tilde{\epsilon}_F \in (0, \tilde{\epsilon})$. By assumption there exist s^- and s^+ such that

$$0 < s^- \leq \inf_{F \in \mathcal{H}_\epsilon} \boldsymbol{\sigma}(F) < \sup_{F \in \mathcal{H}_\epsilon} \boldsymbol{\sigma}(F) \leq s^+ < \infty .$$

Then

$$\begin{aligned} \lambda'_F(\tilde{\epsilon}_F) &= E_F \psi'\left(\frac{X - \boldsymbol{\mu}(F) + \tilde{\epsilon}_F}{\boldsymbol{\sigma}(F)}\right) \frac{1}{\boldsymbol{\sigma}(F)} \\ &\geq \frac{1}{s^+}\,(1 - \epsilon_{\mathcal{H}_\epsilon})\, E_{F_0} \psi'\left(\frac{X - \boldsymbol{\mu}(F) + \tilde{\epsilon}_F}{\boldsymbol{\sigma}(F)}\right), \end{aligned}$$

where $\epsilon_{\mathcal{H}_\epsilon}$ is the proportion of contamination in $\mathcal{H}_\epsilon$. It is easy to see that the last term in the above equation is a decreasing function of $\tilde{\epsilon}_F$. Hence $\tilde{\epsilon}_F \leq \tilde{\epsilon}$ implies

$$\lambda_F(\tilde{\epsilon}) = \tilde{\epsilon}\, \lambda'_F(\tilde{\epsilon}_F) \geq \frac{\tilde{\epsilon}}{s^+}\,(1 - \epsilon_{\mathcal{H}_\epsilon})\, E_{F_0} \psi'\left(\frac{X - \boldsymbol{\mu}(F) + \tilde{\epsilon}}{\boldsymbol{\sigma}(F)}\right).$$

The Dominated Convergence Theorem shows that the above expression is continuous as a function of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. It is also positive and hence a sufficient condition to obtain a positive lower bound is that $\boldsymbol{\mu}(F)$ and $\boldsymbol{\sigma}(F)$ be bounded for any $F \in \mathcal{H}_\epsilon$. A similar argument can be applied to show that equation (24) holds.

It follows that $\left\{\left|\hat{\mu}_n - \boldsymbol{\mu}\right| > \tilde{\epsilon}\right\} \subseteq A_n(F, \tilde{\epsilon}) \bigcup B_n(F, \tilde{\epsilon})$. Hence,

$$\bigcup_{n=m}^{\infty} \left\{\left|\hat{\mu}_n - \boldsymbol{\mu}\right| > \tilde{\epsilon}\right\} \subseteq \bigcup_{n=m}^{\infty} A_n(F, \tilde{\epsilon}) \bigcup \bigcup_{n=m}^{\infty} B_n(F, \tilde{\epsilon}) .$$

Immediately

$$\begin{aligned} \mathcal{M}_m(F, \tilde{\epsilon}) &= \left\{\sup_{n \geq m} \left|\hat{\mu}_n - \boldsymbol{\mu}\right| > \tilde{\epsilon}\right\} \\ &\subseteq \left\{\sup_{n \geq m} \left|\overline{\psi}_n(\boldsymbol{\mu} - \tilde{\epsilon}/2) - \mu_\psi(\boldsymbol{\mu} - \tilde{\epsilon}/2, F)\right| > \mu_\psi(\boldsymbol{\mu} - \tilde{\epsilon}/2, F)\right\} \\ &\qquad \bigcup \left\{\sup_{n \geq m} \left|\overline{\psi}_n(\boldsymbol{\mu} + \tilde{\epsilon}/2) - \mu_\psi(\boldsymbol{\mu} + \tilde{\epsilon}/2, F)\right| > -\mu_\psi(\boldsymbol{\mu} + \tilde{\epsilon}/2, F)\right\} \\ &\subseteq \mathcal{A}_m(F, \boldsymbol{\mu} - \tilde{\epsilon}/2, a(\tilde{\epsilon})) \bigcup \mathcal{A}_m(F, \boldsymbol{\mu} + \tilde{\epsilon}/2, b(\tilde{\epsilon})) . \end{aligned}$$

We have

$$P_F[\mathcal{M}_m(F, \tilde{\epsilon})] \leq P_F[\mathcal{A}_m(F, \boldsymbol{\mu} - \tilde{\epsilon}/2, a(\tilde{\epsilon}))] + P_F[\mathcal{A}_m(F, \boldsymbol{\mu} + \tilde{\epsilon}/2, b(\tilde{\epsilon}))] ,$$

and then

$$\sup_{F\in\mathcal{H}_\epsilon} P_F\left[\mathcal{M}_m\left(F,\tilde{\epsilon}\right)\right] \leq \sup_{F\in\mathcal{H}_\epsilon} P_F\left[\mathcal{A}_m\left(F,\boldsymbol{\mu}-\tilde{\epsilon}/2,a\left(\tilde{\epsilon}\right)\right)\right] + \sup_{F\in\mathcal{H}_\epsilon} P_F\left[\mathcal{A}_m\left(F,\boldsymbol{\mu}+\tilde{\epsilon}/2,b\left(\tilde{\epsilon}\right)\right)\right] \,,$$

so that

$$\begin{aligned}\overline{\lim_{m\to\infty}} \sup_{F\in\mathcal{H}_\epsilon} P_F\left[\mathcal{M}_m\left(F,\tilde{\epsilon}\right)\right] &\leq \lim_{m\to\infty} \sup_{F\in\mathcal{H}_\epsilon} P_F\left[\mathcal{A}_m\left(F,\boldsymbol{\mu}-\tilde{\epsilon}/2,a\left(\tilde{\epsilon}\right)\right)\right] \\ &\qquad + \lim_{m\to\infty} \sup_{F\in\mathcal{H}_\epsilon} P_F\left[\mathcal{A}_m\left(F,\boldsymbol{\mu}+\tilde{\epsilon}/2,b\left(\tilde{\epsilon}\right)\right)\right] = 0\,,\end{aligned}$$

and the proof is complete. ■

Proof of Theorem 2: We need to introduce the following notation. Let $\rho\left(x,t,s\right)=\rho\left(\left(x-t\right)/\,s\right)$. Denote the set of positive real numbers $(0,\infty)$ by $\mathbb{R}_+$. For each $t\in\mathbb{R}$ and $s\in\mathbb{R}_+$ let

$$\gamma\left(F,t,s\right) = E_F\rho\left(X,t,s\right)\,, \tag{25}$$

$$\gamma_n\left(t,s\right) = \gamma\left(F_n,t,s\right) = \frac{1}{n}\sum_{i=1}^{n}\rho\left(x_i,t,s\right)\,, \tag{26}$$

where F_n denotes the empirical distribution function of the random sample $x_1,\ldots,x_n$. As in the proof of Lemma 8, equation (17) above implies that

$$\gamma\left(F,0,\boldsymbol{\sigma}\left(F\right)\right) < \gamma\left(F,t,\boldsymbol{\sigma}\left(F\right)\right)\,, \qquad \forall\ |t|\geq t^*\,.$$

Also, because of (18), there exists η independent of F such that

$$\inf_{\substack{-t^*\leq t\leq t^*\\ s^-\leq s\leq s^+}} \gamma''\left(F,t,s\right) \geq \eta > 0\,, \qquad \forall\, F\in\mathcal{H}_\epsilon,$$

where η does not depend on $F\in\mathcal{H}_\epsilon$. Hence the family of functions $\gamma\left(F,t,\boldsymbol{\sigma}\left(F\right)\right)$ with $F\in\mathcal{H}_\epsilon$ has a unique minimum in the fixed interval $(-t^*,t^*)$. For each $F\in\mathcal{H}_\epsilon$ denote this unique minimum by $\tilde{\boldsymbol{\mu}}\left(F\right)$. Now fix an arbitrary neighbourhood $B_\delta\left(\tilde{\boldsymbol{\mu}}\left(F\right)\right)$ of $\tilde{\boldsymbol{\mu}}\left(F\right)$. Let $\tilde{\epsilon}\left(\delta,F\right)$ satisfy

$$\inf_{t\notin B_\delta(\tilde{\ }(F))} \gamma\big(F,t,\boldsymbol{\sigma}\left(F\right)\big) \;\geq\; \gamma\big(F,\tilde{\boldsymbol{\mu}}\left(F\right),\boldsymbol{\sigma}\left(F\right)\big) \;+\; \tilde{\epsilon}\left(\delta,F\right)\,. \tag{27}$$

By Lemma 5 we have that

$$\tilde{\epsilon} = \tilde{\epsilon}\left(\delta\right) = \inf_{F\in\mathcal{H}_\epsilon} \tilde{\epsilon}\left(\delta,F\right) > 0\,. \tag{28}$$

Choose an arbitrary $\tilde{\delta}>0$ and let I_2 and $m_0=m_0\left(\tilde{\delta}\right)$ be as in Lemma 8, i.e.

$$P_F\left[\tilde{\mu}_n\in I_2\,, \quad \forall\, n\geq m\right] > 1-\tilde{\delta}\,, \quad \forall\, m\geq m_0\,. \tag{29}$$

Note that I_2 above does not depend on $F\in\mathcal{H}_\epsilon$. For each $t\in I_2\cap B_\delta\left(\tilde{\boldsymbol{\mu}}\left(F\right)\right)^c$ let $B\left(t\right)$ be a neighbourhood of t small enough so that we have

$$E_F\Big[\inf_{t'\in B(t)}\rho\left(X,t',\boldsymbol{\sigma}\left(F\right)\right)\Big] \geq \gamma\left(F,\tilde{\boldsymbol{\mu}},\boldsymbol{\sigma}\left(F\right)\right)+\tilde{\epsilon}\,.$$

By Lemma 3 we can choose the size of these $B(t)$s independently of t. Hence, their size does not depend on F. Consider a finite coverage $B(t_1), \ldots, B(t_r)$ of $I_2 \cap B_\delta(\tilde{\boldsymbol{\mu}}(F))^c$. Note that this coverage depends on $F \in \mathcal{H}_\epsilon$. For each of these centres t_k let

$$Y_i(t_k) = \inf_{t' \in B(t_k)} \rho\left(X_i, t', \hat{\sigma}_n\right)$$

and

$$Y(F, t_k) = E_F\left[\inf_{t' \in B(t_k)} \rho\left(X, t', \boldsymbol{\sigma}(F)\right)\right] \neq E_F\big[Y_i(t_k)\big] .$$

Consider the events

$$A_m(F, t_k) = \left\{\sup_{n \geq m} \left|\overline{Y}_n(t_k) - Y(F, t_k)\right| \leq \tilde{\epsilon}\right\}, \qquad m \in \mathbb{N} .$$

By Lemma 6 we have that there exists $m_1(\tilde{\delta})$ independent from t_k (i.e. independent from F) such that

$$P_F\Big(A_m(F, t_k)\Big) > 1 - \tilde{\delta}, \qquad \forall\, m \geq m_1(\tilde{\delta}) , \quad \forall F \in \mathcal{H}_\epsilon , \quad \forall\, t_k \in I_2 .$$

Now note that

$$\begin{aligned} A_m(F, t_k) &\subseteq \left\{\frac{1}{n}\sum_{i=1}^{n} \inf_{t \in B(t_k)} \rho(x_i, t, \hat{\sigma}_n) \geq \gamma\big(F, \tilde{\boldsymbol{\mu}}(F), \boldsymbol{\sigma}(F)\big) + 2\tilde{\epsilon}, \forall n \geq m\right\} \\ &\subseteq \left\{\inf_{t \in B(t_k)} \frac{1}{n}\sum_{i=1}^{n} \rho(x_i, t, \hat{\sigma}_n) \geq \gamma\big(F, \tilde{\boldsymbol{\mu}}(F), \boldsymbol{\sigma}(F)\big) + 2\tilde{\epsilon}, \forall n \geq m\right\} = C_m(F, t_k) . \end{aligned}$$

Let

$$D_m(F) = \left\{\frac{1}{n}\sum_{i=1}^{n} \rho\big(x_i, \tilde{\boldsymbol{\mu}}(F), \boldsymbol{\sigma}(F)\big) \;\leq\; \gamma\big(F, \tilde{\boldsymbol{\mu}}(F), \boldsymbol{\sigma}(F)\big) + \tilde{\epsilon}, \quad \forall n \geq m\right\} .$$

Bernstein's Lemma (Inequality) also shows that there exists $m_2 = m_2(\tilde{\delta})$ (independent from F) such that for $m \geq m_2$ we have

$$P_F\big(D_m(F)\big) > 1 - \tilde{\delta} , \quad \forall\, F \in \mathcal{H}_\epsilon .$$

Take $m_3 = \max(m_0, m_1, m_2)$. Note that m_3 does not depend on $F \in \mathcal{H}_\epsilon$. We have

$$P_F\Big[C_m(F) \cap D_m(F)\Big] \geq 1 - 2\tilde{\delta} , \quad \forall\, m \geq m_3 , \quad \forall F \in \mathcal{H}_\epsilon .$$

We also have

$$C_m(F) \cap D_m(F) \quad \subseteq \quad \Big[\tilde{\mu}_m \in B_\delta\big(\tilde{\boldsymbol{\mu}}(F)\big) , \quad \forall\, m \geq m_2\Big] .$$

Hence, for each $\tilde{\delta} > 0$ there exists $m_3(\tilde{\delta})$ such that

$$P_F\Big[\tilde{\mu}_m \in B_\delta\big(\tilde{\boldsymbol{\mu}}(F)\big), \; \forall\, m \geq m_3\Big] \geq 1 - 2\tilde{\delta}, \qquad \forall F \in \mathcal{H}_\epsilon ,$$

that is, for each neighbourhood $B_\delta\big(\tilde{\boldsymbol{\mu}}\,(F)\big)$ we have

$$\lim_{m\to\infty}\inf_{F\in\mathcal{H}_\epsilon} P_F\Big[\,\tilde{\mu}_n \in B_\delta\big(\tilde{\boldsymbol{\mu}}\,(F)\big),\ \forall\, n\geq m\,\Big] = 1\,,$$

or equivalently,

$$\lim_{m\to\infty}\sup_{F\in\mathcal{H}_\epsilon} P_F\Big[\sup_{n\geq m}|\tilde{\mu}_n - \tilde{\boldsymbol{\mu}}\,(F)| > \delta\,\Big] = 0\,.$$

■

To prove Theorem 3 we need uniform versions of the usual "little o in probability" and "big O in probability" definitions. We will also give a formal definition of uniform asymptotic normality.

Definition 2 - Uniform big O in probability: *Let a_n, $n\geq 1$, be a sequence of real numbers and let X_n, $n\geq 1$, be a sequence of random variables. We say that $X_n = UO_P\,(a_n)$ over the set of distribution functions $\mathcal{H}_\epsilon$ if*

$$\lim_{k\to\infty}\sup_{F\in\mathcal{H}_\epsilon}\lim_{n\to\infty} P_F\left[\,\left|\frac{X_n}{a_n}\right| > k\,\right] = 0.$$

Definition 3 - Uniform small o in probability: *Let a_n, $n\geq 1$, be a sequence of real numbers and let X_n, $n\geq 1$, be a sequence of random variables. We say that $X_n = Uo_P\,(a_n)$ over the set of distribution functions $\mathcal{H}_\epsilon$ if $\forall\,\delta > 0$*

$$\lim_{n\to\infty}\sup_{F\in\mathcal{H}_\epsilon} P_F\left[\,\left|\frac{X_n}{a_n}\right| > \delta\,\right] = 0.$$

Definition 4 - Uniformly asymptotically normal: *We say that a sequence X_n, $n\in\mathbb{N}$ is uniformly asymptotically normal (UAN) over the set of distribution functions $\mathcal{H}_\epsilon$ if*

$$\sup_{F\in\mathcal{H}_\epsilon}\sup_{x\in\mathbb{R}}\Big|P_F\,(X_n\leq x) - \Phi\,(x)\Big| = o\,(1)\,. \tag{30}$$

With the above definitions we can show that these "uniform little o", "uniform big O" and "uniform asymptotic distribution" behave similarly to their "non-uniform" counterparts. This is made more precise in the following remark.

Remark 5 - Properties of $UO_p\,(1)$, $Uo_p\,(1)$ and UAN - *In what follows a_n, b_n and X_n, $n\in\mathbb{N}$ denote sequences of random variables. It is easy to see that the following properties hold. Proofs of these results can be found in Salibian-Barrera (2000, Chapter 2).*

Property 1 - *if $a_n = UO_P\,(1)$ and $b_n = Uo_P\,(1)$, then $a_n\times b_n = Uo_P\,(1)$;*

Property 2 - *if $a_n = UO_P\,(1)$ and there exists $b\neq 0$ with $b_n - b = Uo_P\,(1)$, then $a_n/\,b_n = a_n/\,b + Uo_P\,(1)$;*

Property 3 - *if $a_n = Uo_P\,(1)$ and X_n is UAN then $X_n + a_n$ is UAN.*

Proof of Theorem 3: To simplify the notation, in what follows let $\mu = \mu\,(F)$, $\tilde{\mu} = \tilde{\mu}\,(F)$ and $\sigma = \sigma\,(F)$. The idea of the proof is to show that $\sqrt{n}\,(\hat{\mu}_n - \mu)$ can be represented as a linear term plus a uniformly small remainder. We use the Berry Esseen Theorem to show that the linear part is UAN (see Definition 4) and Property 3 above to show that the sum of these terms is also UAN.

First note that by Theorem 2 and 4 we have $\hat{\sigma}_n - \sigma = Uo_P(1)$, $\tilde{\mu}_n - \tilde{\mu} = Uo_P(1)$ and $\hat{\mu}_n - \mu = Uo_P(1)$. We now show that

$$\sqrt{n}\frac{(\hat{\mu}_n - \mu)}{\sqrt{V}} = \sqrt{n}\frac{\overline{W}_n}{\sqrt{V}} + Uo_P(1) \ . \tag{31}$$

where

$$\begin{aligned} W_i &= \Big(\psi\left((x_i - \mu)/\sigma\right) - d\left(\rho\left((x_i - \tilde{\mu})/\sigma\right) - b\right)\Big)\Big/ e\,, \\ d &= \frac{E_F\left\{\psi'\left((X-\mu)/\sigma\right)(X-\mu)/\sigma\right\}}{E_F\left\{\rho'\left((X-\tilde{\mu})/\sigma\right)(X-\tilde{\mu})/\sigma\right\}} \\ e &= E_F\left\{\psi'\left((X-\mu)/\sigma\right)\right\} \ . \end{aligned} \tag{32}$$

To simplify the notation let $\mu = \mu(F)$, $\sigma = \sigma(F)$, $\tilde{\mu} = \tilde{\mu}(F)$ and

$$u_i = (x_i - \mu)/\sigma.$$

A second order Taylor expansion of (4) around the limit values (μ, σ) yields

$$\left\{\frac{1}{n}\sum_{i=1}^n \psi'(u_i)\right\}\frac{1}{\sigma}\sqrt{n}(\hat{\mu}_n - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(u_i) - \sqrt{n}\frac{(\hat{\sigma}_n - \sigma)}{\sigma}\frac{1}{n}\sum_{i=1}^n \psi'(u_i)\,u_i \tag{33}$$

$$+\frac{1}{2}\frac{1}{\sigma^2}\ \sqrt{n}(\hat{\mu}_n - \mu)^2\ \frac{1}{n}\sum_{i=1}^n \psi''(\tilde{u}_i) \tag{34}$$

$$+\frac{1}{2}\frac{1}{\sigma^2}\ \sqrt{n}(\hat{\sigma}_n - \sigma)^2\ \frac{1}{n}\sum_{i=1}^n \Big[\psi''(\tilde{u}_i)\,\tilde{u}_i^2 \tag{35}$$

$$+ 2\psi'(\tilde{u}_i)\,\tilde{u}_i\Big] \tag{36}$$

$$+\frac{1}{n}\frac{1}{\sigma^2}\sum_{i=1}^n \Big[\psi''(\tilde{u}_i)\,\tilde{u}_i + \psi'(\tilde{u}_i)\Big](\hat{\sigma}_n - \sigma)\ (\hat{\mu}_n - \mu) \tag{37}$$

where $\tilde{u}_i = (x_i - \tilde{\mu})/\tilde{\sigma}$ and $(\tilde{\mu}, \tilde{\sigma})$ lies between $(\hat{\mu}_n, \hat{\sigma}_n)$ and (μ, σ). Let

$$B_n = \frac{1}{2}\frac{1}{\sigma}(\hat{\mu}_n - \mu)\frac{1}{n}\sum_{i=1}^n \psi''(\tilde{u}_i)\ , \tag{38}$$

$$C_n = \frac{1}{2}\frac{1}{\sigma}(\hat{\sigma}_n - \sigma)\ \frac{1}{n}\sum_{i=1}^n \left[\psi''(\tilde{u}_i)\,\tilde{u}_i^2 + 2\psi'(\tilde{u}_i)\,\tilde{u}_i\right]\ , \tag{39}$$

and

$$D_n = \frac{1}{n}\frac{1}{\sigma^2}\sum_{i=1}^n \left[\psi''(\tilde{u}_i)\,\tilde{u}_i + \psi'(\tilde{u}_i)\right](\hat{\sigma}_n - \sigma)\ . \tag{40}$$

From (33) to (40) we have

$$\begin{aligned} &\frac{1}{\sigma}\sqrt{n}(\hat{\mu}_n - \mu)\ \left(\frac{1}{n}\sum_{i=1}^n \psi'(u_i) - B_n - D_n\right) \\ &\qquad = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(u_i) - \frac{1}{\sigma}\sqrt{n}(\hat{\sigma}_n - \sigma)\ \left(\frac{1}{n}\sum_{i=1}^n \psi'(u_i)\,u_i - C_n\right). \end{aligned} \tag{41}$$

From (41) and Lemmas 9 and 1 we have

$$\frac{1}{\sigma}\sqrt{n}\left(\hat{\mu}_n-\mu\right)\ \left(\frac{1}{n}\sum_{i=1}^{n}\psi'\left(u_i\right)+Uo_P\left(1\right)\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi\left(u_i\right)-\frac{1}{\sigma}\sqrt{n}\left(\hat{\sigma}_n-\sigma\right)\ \left(\frac{1}{n}\sum_{i=1}^{n}\psi'\left(u_i\right)u_i+Uo_P\left(1\right)\right). \quad (42)$$

It is easy to see that if the function ρ is continuously differentiable, the pair $(\tilde{\mu}_n, \hat{\sigma}_n)$ in (6) and (8) satisfies the following system of equations

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\left(x_i-\tilde{\mu}_n\right)/\,\hat{\sigma}_n\right)=b \quad (43)$$
$$\frac{1}{n}\sum_{i=1}^{n}\rho'\left(\left(x_i-\tilde{\mu}_n\right)/\,\hat{\sigma}_n\right)=0, \quad (44)$$

where ρ' denotes the derivative of ρ.

From equation (43) we get

$$\frac{1}{\sigma}\sqrt{n}\left(\hat{\sigma}_n-\sigma\right)\left[\frac{1}{n}\sum_{i=1}^{n}\rho'\left(v_i\right)v_i-B_n'-D_n'\right]$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\rho\left(v_i\right)-b-\frac{1}{\sigma}\sqrt{n}\left(\tilde{\mu}_n-\tilde{\mu}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\rho'\left(v_i\right)-C_n'\right), \quad (45)$$

where, as before, $B_n'=o_P\left(1\right)$, $C_n'=Uo_P\left(1\right)$ and $D_n'=Uo_P\left(1\right)$. Note that

$$\frac{1}{n}\sum_{i=1}^{n}\rho'\left(v_i\right)=Uo_P\left(1\right),$$

and hence

$$\frac{1}{n}\sum_{i=1}^{n}\rho'\left(v_i\right)-C_n'=Uo_P\left(1\right). \quad (46)$$

From (44) we have

$$\frac{1}{\sigma}\sqrt{n}\left(\tilde{\mu}_n-\tilde{\mu}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\rho''\left(v_i\right)\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\rho'\left(v_i\right)-\frac{1}{\sigma}\sqrt{n}\left(\hat{\sigma}_n-\sigma\right)\left(\frac{1}{n}\sum_{i=1}^{n}\rho''\left(v_i\right)v_i\right)$$
$$=UO_P\left(1\right)-\frac{1}{\sigma}\sqrt{n}\left(\hat{\sigma}_n-\sigma\right)+UO_P\left(1\right). \quad (47)$$

From (45) and (46) we have

$$\frac{1}{\sigma}\sqrt{n}\left(\hat{\sigma}_n-\sigma\right)\left[\frac{1}{n}\sum_{i=1}^{n}\rho'\left(v_i\right)v_i+Uo_P\left(1\right)\right]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\rho\left(v_i\right)-b-\frac{1}{\sigma}\sqrt{n}\left(\tilde{\mu}_n-\tilde{\mu}\right)\times Uo_P\left(1\right).$$

Similarly, from equation (47) we have

$$\frac{1}{\sigma}\sqrt{n}\left(\tilde{\mu}_n-\tilde{\mu}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\rho''\left(v_i\right)\right)=UO_P\left(1\right)-\frac{1}{\sigma}\sqrt{n}\left(\hat{\sigma}_n-\sigma\right)+UO_P\left(1\right).$$

From the last two equations we obtain

$$\frac{1}{\sigma}\sqrt{n}\left(\hat{\sigma}_n-\sigma\right)\Big[a+Uo_P\left(1\right)\Big]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\rho\big(\left(v_i\right)-b\big)+Uo_P\left(1\right)\ , \tag{48}$$

where $a=E_F\left\{\rho'\left(\left(X-\tilde{\mu}\right)/\sigma\right)\left(X-\tilde{\mu}\right)/\sigma\right\}$. From (42) and (48) we have

$$\begin{aligned}&\frac{1}{\sigma}\ \sqrt{n}\left(\hat{\mu}_n-\mu\right)\ \Big[c+Uo_P\left(1\right)\Big]\\&\qquad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi\left(u_i\right)-\Big[\frac{1}{a}\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\rho\left(v_i\right)-b\big)+Uo_P\left(1\right)\Big]\Big[d+Uo_P\left(1\right)\Big]\ ,\end{aligned}$$

where $c=E_F\left\{\psi'\left(\left(X-\mu\right)/\sigma\right)\right\}$, and $d=E_F\left\{\psi'\left(\left(X-\mu\right)/\sigma\right)\left(X-\mu\right)/\sigma\right\}$. Hence,

$$\frac{1}{\sigma}\ \sqrt{n}\left(\hat{\mu}_n-\mu\right)\ \Big[c+Uo_P\left(1\right)\Big]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi\left(u_i\right)-\frac{d}{a}\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\rho\left(v_i\right)-b\big)+Uo_P\left(1\right)\ .$$

From the last equation and Property 2 we obtain (31). Note that $|W_i|$ are bounded (see (32)), and hence their moments are bounded uniformly for $F\in\mathcal{H}_\epsilon$. Let $f:\mathbb{R}^3\to\mathbb{R}$ be any non-negative real function such that

$$E_{F_0}\left[f\left(X,t,s\right)\right]=\int f\left(X,t,s\right)\,dF_0\left(X\right)>0$$

for any $t\in\mathbb{R}$ and $s>0$, where F_0 denotes the central distribution of the contamination neighbourhood $\mathcal{H}_\epsilon$. It is easy to see that if $E_{F_0}\left[f\left(X,t,s\right)\right]$ is a continuous function of (t,s) and $\mathcal{K}_t$ and $\mathcal{K}_s$ are compact sets in the real line such that $\mathcal{K}_s\subset(0,\infty)$ then we have

$$\inf_{F\in\mathcal{H}_\epsilon,\,t\in\mathcal{K}_t,\,s\in\mathcal{K}_s}E_F\left[f\left(X,t,s\right)\right]>0\,.$$

In particular, if $\sigma\left(F\right)$ denotes a scale estimate that satisfies (16) and $\mu\left(F\right)$ is an M-location estimate then

$$\inf_{F\in\mathcal{H}_\epsilon}\mathrm{Var}_F\left[\psi_c\left(\left(X-\mu\left(F\right)\right)/\,\sigma\left(F\right)\right)\right]=\inf_{F\in\mathcal{H}_\epsilon}E_F\left[\psi_c\left(\left(X-\mu\left(F\right)\right)/\,\sigma\left(F\right)\right)\right]^2>0\,. \tag{49}$$

We see that the variance of W_i is bounded away from zero uniformly on $F\in\mathcal{H}_\epsilon$. The Berry Esseen Theorem yields

$$\sup_{F\in\mathcal{H}_\epsilon}\sup_{x\in\mathbb{R}}\left|P_F\left\{\frac{\sqrt{n}\,\overline{W}_n}{\sqrt{V}}<x\right\}-\Phi\left(x\right)\right|=o\left(1\right)\ .$$

Hence we have

$$\sqrt{n}\frac{\left(\hat{\mu}_n-\mu\right)}{\sqrt{V}}=V_n+Uo_P\left(1\right)\ ,$$

where V_n is UAN. Property 3 above completes the proof. ■

Lemma 1 *Let* $\rho : \mathbb{R} \longrightarrow \mathbb{R}_+$ *be a continuous real function that satisfies R.4. Let* $t \in \mathcal{T}$ *and* $s \in \mathcal{S}$*, where* $\mathcal{T}$ *and* $\mathcal{S}$ *are bounded real intervals, with* $\inf \{s \in \mathcal{S}\} > 0$*. Then the function*

$$f(u,t,s) = \rho\left(\frac{u-t}{s}\right), \quad u \in \mathbb{R}, \quad t \in \mathcal{T}, \quad s \in \mathcal{S},$$

is continuous in s *and* t *uniformly in* u*. In other words, for any* $\epsilon > 0$*, there exist* $\delta_t > 0$ *and* $\delta_s > 0$ *such that*

$$|s_1 - s_2| < \delta_s, \quad |t_1 - t_2| < \delta_t \quad \Rightarrow \; |f(u,t_1,s_1) - f(u,t_2,s_2)| < \epsilon, \forall u \in \mathbb{R}.$$

Proof: See Salibian-Barrera (2000).

Lemma 2 *Let* $\rho : \mathbb{R} \to \mathbb{R}_+$ *satisfy R.3 and R.4. Let* $\mathcal{K}_s \subset (0,\infty)$ *be an arbitrary closed set. If the central distribution of* $\mathcal{H}_\epsilon$ *has a bounded density function* ϕ*, then* $\gamma(F,t,s)$ *is continuous in* $s \in \mathcal{K}_s$*, uniformly in* $t \in \mathbb{R}$ *and* $F \in \mathcal{H}_\epsilon$*.*

Proof: See Salibian-Barrera (2000).

Lemma 3 *Let* $\rho : \mathbb{R} \to \mathbb{R}_+$ *be continuous and satisfy R.1 and R.4. Then for any neighbourhood* $B(t_0)$ *we have*

$$E_F\left[\inf_{t' \in B(t_0)} \rho\left(X, t', \boldsymbol{\sigma}\right)\right] \xrightarrow[B(t_0)\searrow\{t_0\}]{} E_F\Big[\rho\left(X, t_0, \boldsymbol{\sigma}\right)\Big],$$

uniformly in $F \in \mathcal{H}_\epsilon$ *as* $B(t_0)$ *shrinks to* $\{t_0\}$*. That is, for every* $\tilde{\epsilon} > 0$ *and* $t_0 \in \mathbb{R}$ *there exists* $\delta = \delta(\tilde{\epsilon}, t_0)$ *such that*

$$\left| E_F\Big[\inf_{t' \in B_\delta(t_0)} \rho\left(X, t', \boldsymbol{\sigma}(F)\right)\Big] - E_F\Big[\rho\left(X, t_0, \boldsymbol{\sigma}(F)\right)\Big]\right| < \tilde{\epsilon}, \quad \forall F \in \mathcal{H}_\epsilon\,,$$

where the ball $B_\delta(t_0)$ *has diameter* δ*. Moreover, let* $\mathcal{K} \subset \mathbb{R}$ *be an arbitrary compact set. Then* $\delta = \delta(\tilde{\epsilon}, \mathcal{K})$ *above can be chosen independently of* t_0*. That is, for every* $\tilde{\epsilon} > 0$ *there exists* $\delta = \delta(\tilde{\epsilon}, \mathcal{K})$ *such that*

$$\left| E_F\Big[\inf_{t' \in B_\delta(t)} \rho\left(X, t', \boldsymbol{\sigma}(F)\right)\Big] - E_F\Big[\rho\left(X, t, \boldsymbol{\sigma}(F)\right)\Big]\right| < \tilde{\epsilon}, \quad \forall F \in \mathcal{H}_\epsilon\,, \forall t \in \mathcal{K}\,,$$

where the ball $B_\delta(t)$ *has diameter* δ*.*

Proof: Fix $\epsilon > 0$. By Lemma 1 there exists $\delta = \delta(\sigma, t_0) > 0$ such that

$$|t - t_0| < \delta \;\Rightarrow\; |\rho(x,t,\sigma) - \rho(x,t_0,\sigma)| < \epsilon \qquad \forall x \in \mathbb{R}\,.$$

Hence, $\rho(x,t,\sigma) < \rho(x,t_0,\sigma) + \epsilon$ for all $x \in \mathbb{R}$ and all t in a sufficiently small neighbourhood $B(t_0)$ of t_0. Immediately we obtain

$$\inf_{t \in B(t_0)} \rho(x,t,\sigma) \leq \rho(x,t_0,\sigma) + \epsilon \qquad \forall x \in \mathbb{R}\,.$$

Similarly we have

$$\inf_{t \in B(t_0)} \rho(x,t,\sigma) \geq \rho(x,t_0,\sigma) - \epsilon\,, \qquad \forall x \in \mathbb{R}\,.$$

Hence

$$\left|\inf_{t \in B(t_0)} \rho(x,t,\sigma) - \rho(x,t_0,\sigma)\right| < \epsilon \qquad \forall x \in \mathbb{R}\,,$$

if $B(t_0)$ is sufficiently small. ■

Lemma 4 *Let $\rho : \mathbb{R} \to \mathbb{R}_+$ be continuous and satisfy R.4. Then $g(x,s) = \inf_{t' \in B(t_0)} \rho(x, t', s)$ is continuous in s uniformly in $x \in \mathbb{R}$.*

Proof: See Salibian-Barrera (2000).

Lemma 5 *Let $\rho : \mathbb{R} \to \mathbb{R}_+$ satisfy (18). Let $\tilde{\boldsymbol{\mu}}(F)$ be the global minimum of $\gamma\big(F, t, \boldsymbol{\sigma}(F)\big)$, and for every $\delta > 0$ let $\tilde{\epsilon}(\delta, F)$ be defined by the property*

$$\inf_{t \notin B_\delta(\tilde{}(F))} \gamma\big(F, t, \boldsymbol{\sigma}(F)\big) \;\geq\; \gamma\big(F, \tilde{\boldsymbol{\mu}}(F), \boldsymbol{\sigma}(F)\big) \,+\, \tilde{\epsilon}(\delta, F)\,.$$

Then

$$\tilde{\epsilon}(\delta) = \inf_{F \in \mathcal{H}_\epsilon} \tilde{\epsilon}(\delta, F) > 0\,. \tag{50}$$

Proof: Equation (18) implies that

$$\inf_{\substack{-t^* \leq t \leq t^* \\ s^- \leq s \leq s^+}} \gamma''(F, t, s) \geq \eta > 0, \qquad \forall\, F \in \mathcal{H}_\epsilon,$$

where η does not depend on F. Hence for any $t \notin B_\delta\big(\tilde{\boldsymbol{\mu}}(F)\big)$ we have

$$\begin{aligned}
\gamma\big(F, t, \boldsymbol{\sigma}(F)\big) - \gamma\left(F, \tilde{\boldsymbol{\mu}}(F), \boldsymbol{\sigma}(F)\right) &= \frac{1}{2}\,\gamma''\left(F, \bar{t}, \boldsymbol{\sigma}(F)\right)\,\left(t - \tilde{\boldsymbol{\mu}}(F)\right)^2 \\
&\geq \eta\;\left(t - \tilde{\boldsymbol{\mu}}(F)\right)^2 \\
&> \eta\;\delta^2 > 0 \qquad\qquad \forall\, F \in \mathcal{H}_\epsilon\,,
\end{aligned}$$

where $\bar{t} \notin B_\delta\big(\tilde{\boldsymbol{\mu}}(F)\big)$ and η does not depend on F. ■

Lemma 6 *Let $\rho : \mathbb{R} \to \mathbb{R}_+$ be continuous and satisfy R.4. Let $t_0 \in \mathbb{R}$ be a fixed real number and let $B(t_0)$ be an arbitrary neighbourhood of t_0. Define*

$$Y_i(t_0) = \inf_{t' \in B(t_0)} \rho\left(X_i, t', \hat{\sigma}_n\right)$$

and

$$Y(F, t_0) = E_F\left[\inf_{t' \in B(t_0)} \rho\left(X, t_0, \boldsymbol{\sigma}(F)\right)\right] \neq E_F\left[Y_i(t_0)\right]\,.$$

Then for any $\delta > 0$

$$\lim_{m \to \infty} \sup_{F \in \mathcal{H}_\epsilon} P_F\Big(\sup_{n \geq m} \left|\overline{Y}_n(t_0) - Y(F, t_0)\right| > \delta\Big) = 0\,. \tag{51}$$

Moreover, let $\mathcal{K} \subset \mathbb{R}$ be an arbitrary bounded interval and assume that $t_0 \in \mathcal{K}$. For any $\tilde{\epsilon} > 0$ and $\delta > 0$ we can choose $m_0 = m_0(\tilde{\epsilon}, \delta, \mathcal{K})$ independently from t_0 such that

$$\sup_{F \in \mathcal{H}_\epsilon} P_F\Big(\sup_{n \geq m} \left|\overline{Y}_n(t_0) - Y(F, t_0)\right| > \delta\Big) < \tilde{\epsilon} \qquad \forall\, m \geq m_0 \quad \forall\, t_0 \in \mathcal{K}\,.$$

$$\lim_{m \to \infty} \sup_{t_0 \in \mathcal{K}} \sup_{F \in \mathcal{H}_\epsilon} P_F\Big(\sup_{n \geq m} \left|\overline{Y}_n(t_0) - Y(F, t_0)\right| > \delta\Big) = 0\,.$$

Proof: Let

$$V_i(F, t_0) = \inf_{t' \in B(t_0)} \rho\left(X_i, t', \boldsymbol{\sigma}(F)\right), \qquad i = 1, \ldots, n.$$

Then $Y(F, t_0) = E_F[V_i(t_0)]$. We have to show that for any $\delta > 0$ and $\tilde{\epsilon} > 0$ there exists $m_0 = m_0(\delta, \tilde{\epsilon})$ independent from $t_0 \in \mathcal{K}$ such that

$$P_F\left[\sup_{n \geq m} \left|\overline{Y}_n(t_0) - Y(F, t_0)\right| > \delta\right] < \tilde{\epsilon}, \quad \forall\, m \geq m_0, \quad \forall\, t_0 \in \mathcal{K}.$$

We cannot use Bernstein's Lemma (also known as Bernstein's inequality) on $Y_i - Y(F)$ because these random variables do not have mean zero nor are they independent. We have

$$P_F\left[\sup_{n \geq m} \left|\overline{Y}_n(t_0) - Y(F, t_0)\right| > \delta\right] \leq P_F\left[\sup_{n \geq m} \left|\overline{V}_n(F, t_0) - Y(F, t_0)\right| > \delta/2\right] + \\ + P_F\left[\sup_{n \geq m} \left|\overline{Y}_n(t_0) - \overline{V}_n(F, t_0)\right| > \delta/2\right]. \quad (52)$$

Also

$$P_F\left[\sup_{n \geq m} \left|\overline{Y}_n(t_0) - \overline{V}_n(F, t_0)\right| > \delta/2\right] \leq P_F\left[\sup_{n \geq m} \left|\hat{\sigma}_n - \boldsymbol{\sigma}(F)\right| > \epsilon'\right], \quad (53)$$

for some $\epsilon' = \epsilon'(\delta)$ that depends on δ but does not depend on t_0 or $B(t_0)$ (although it does depend on $\mathcal{K}$). To prove (53) note that $\overline{Y}_n(t_0) = 1/n \sum_{i=1}^n g(x_i, t_0, \hat{\sigma}_n)$ and $\overline{V}_n(F, t_0) = 1/n \sum_{i=1}^n g(x_i, t_0, \boldsymbol{\sigma}(F))$ with $g(x, t_0, s) = \inf_{t' \in B(t_0)} \rho(x, t', s)$. Note that $g(x, t_0, s)$ is continuous in s uniformly in x and $t_0 \in \mathcal{K}$ (see Lemma 4). Hence, for a given $\delta/2$ there exists a positive $\epsilon' = \epsilon'(\delta)$ that does not depend on $t_0 \in \mathcal{K}$ such that $|\hat{\sigma}_n - \boldsymbol{\sigma}(F)| < \epsilon'$ implies $\left|\overline{Y}_n(t_0) - \overline{V}_n(F, t_0)\right| < \delta/2$ for all $t_0 \in \mathcal{K}$. Hence, for each n we have

$$\left\{\left|\overline{Y}_n(t_0) - \overline{V}_n(F, t_0)\right| > \delta/2\right\} \subset \left\{\left|\hat{\sigma}_n - \boldsymbol{\sigma}(F)\right| > \epsilon'\right\}, \qquad \forall\, t_0 \in \mathcal{K},$$

and then note that for any sequence of random variables $\{X_n\}_{n \in \mathbb{N}}$ if a is a real number, we have $\left\{\sup_{n \geq m} X_n > a\right\} = \bigcup_{n \geq m} \{X_n > a\}$. Together with (53) this bounds the second term in (52). To control the first term, note that the sequence of random variables $W_i(t_0) = V_i(t_0) - E(V_i(t_0)) = V_i(t_0) - Y(F, t_0)$ satisfies the assumptions of Bernstein's Lemma with $c = 2 \sup_u \rho(u)$ and $s_n = n\, \sigma_w^2(t_0)$, where $\sigma_w^2(t_0)$ denotes the variance of $W_i(t_0)$. Hence for any $\delta > 0$ we have

$$\begin{aligned} P_F\Big(\left|\overline{V}_n(t_0) - E(V(t_0))\right| > \delta\Big) &= P_F\Big(\left|\overline{W}_n(t_0)\right| > \delta\Big) \\ &\leq 2 \exp\left(\frac{-n\, \delta^2}{2\, (\sigma_w^2(t_0) + c\, \delta)}\right) \leq 2\ \exp\left(\frac{-n\, \delta^2}{2\, (k^2 + c\, \delta)}\right) \\ &= 2\ \left[\exp(-a(\delta))\right]^n, \end{aligned} \quad (54)$$

where $\sigma_w^2(t_0) \leq k^2 < \infty$ for all $F \in \mathcal{H}_\epsilon$ and for all $t_0 \in \mathcal{K}$. Note that $a(\delta) > 0$ and it does not depend on F nor on t_0. Use (15) to find m_0 large enough such that

$$\sup_{F \in \mathcal{H}_\epsilon} P_F\left[\sup_{n \geq m} \left|\hat{\sigma}_n - \boldsymbol{\sigma}(F)\right| > \epsilon'\right] < \tilde{\epsilon}/2, \quad \forall\, m \geq m_0, \quad (55)$$

and use (54) together with the Borel-Cantelli Lemma and a standard argument to find m_1 large enough (independently from t_0) such that

$$\sup_{F\in\mathcal{H}_\epsilon} P_F\left[\sup_{n\geq m}\left|\overline{V}_n(t_0)-Y(F,t_0)\right|>\delta/2\right]<\tilde{\epsilon}/2\,,\quad \forall\, m\geq m_1\,,\quad \forall\, t_0\in\mathcal{K}\,. \tag{56}$$

Equations (52), (55) and (56) show (51). ■

If $f:\mathbb{R}\to\mathbb{R}$ is a real function, let

$$f^-(x)=\max\left(0,-f(x)\right)\,.$$

Lemma 7 *For each $t\in\mathbb{R}$ define the set*

$$\mathcal{A}(t)=\left\{s_H(t)\;:\;E_H\rho\left(\frac{X-t}{s_H(t)}\right)=b,\; H\in\mathcal{H}_\epsilon\right\},$$

let $s^-(t)=\inf\mathcal{A}(t)$ and $s^+(t)=\sup\mathcal{A}(t)$. Let t^* be the solution of

$$\inf_{s^-(t^*)\leq s\leq s^+(t^*)}\left[-E_\Phi\rho'\left(\frac{X-t^*}{s}\right)\right]=\frac{\epsilon}{1-\epsilon}\sup_x\rho'(x)\,. \tag{57}$$

Assume that for this choice of t^* we have

$$\inf_{\substack{-t^*\leq t\leq t^*\\ s^-\leq s\leq s^+}} E_\Phi\rho''\left(\frac{X-t}{s}\right)>\frac{\epsilon}{1-\epsilon}\sup_x\rho''(x)^-\,. \tag{58}$$

Then $\gamma(F,\cdot,\boldsymbol{\sigma})$ has its unique global minimum in the interval $(-t^*,t^*)$ for any $F\in\mathcal{H}_\epsilon$.

Proof: We will now show that (57) and (58) are sufficient conditions to ensure that $\gamma(F,\cdot,\boldsymbol{\sigma})$ has its unique global minimum in the interval $(-t^*,t^*)$ for any $F\in\mathcal{H}_\epsilon$. The reasoning is as follows. If t is a minimum of $\gamma(F,\cdot,\boldsymbol{\sigma})$ then it solves the equation

$$(1-\epsilon)\,E_\Phi\rho_d'\left(\frac{X-t}{s(t)}\right)+\epsilon E_H\rho_d'\left(\frac{X-t}{s(t)}\right)=0, \tag{59}$$

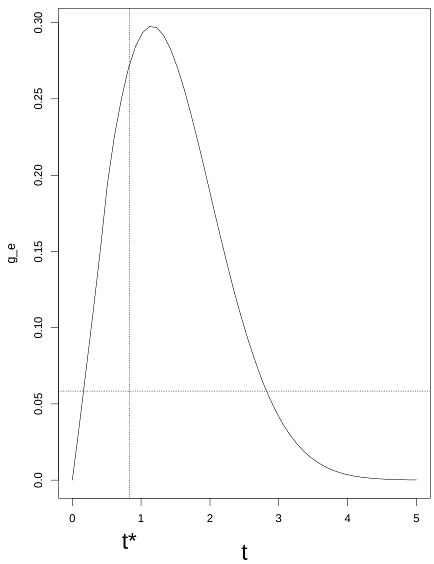
where $s(t)=\boldsymbol{\sigma}$. Hence, t solves

$$-E_\Phi\rho_d'\left(\frac{X-t}{s(t)}\right)=\frac{\epsilon}{(1-\epsilon)}E_H\rho_d'\left(\frac{X-t}{s(t)}\right)\,. \tag{60}$$

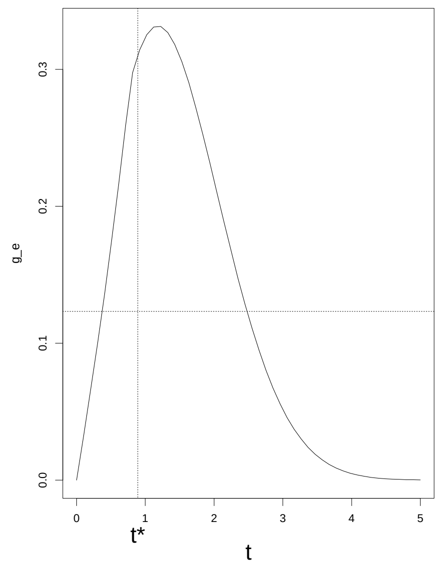
For each $\epsilon\in(0,1/2]$ the largest solution t of (60) is determined by solving

$$g_\epsilon(t)=\inf_{s^-(t)\leq s\leq s^+(t)}\left[-E_\Phi\rho_d'\left(\frac{X-t}{s}\right)\right]=\sup_{H\in\mathcal{H}_\epsilon}\frac{\epsilon}{(1-\epsilon)}E_H\rho_d'\left(\frac{X-t}{s(t)}\right)=\frac{\epsilon}{1-\epsilon}\sup_x\rho_d'(x)\,, \tag{61}$$

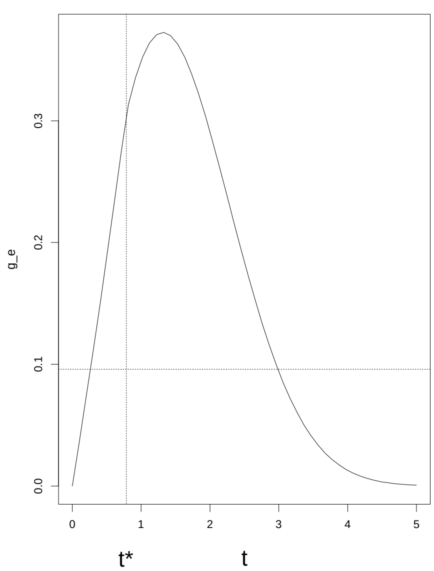
that is, equation (57). In Figure 1 we plot the function $g_\epsilon(t)$ for estimates with breakdown point 50% and 40% and different values of ϵ. We include the threshold t^* obtained in (17). We see that the largest solution of (57) (or 61) is larger than the mentioned threshold, and hence this solution corresponds to a local minimum of $\gamma(F,\cdot,\boldsymbol{\sigma})$. The smallest solution t^* of (57) is then the largest possible value of t satisfying (59) that corresponds to a global minimum. Equation (58) guarantees that every function $\gamma(F,\cdot,\boldsymbol{\sigma})$ is strictly convex in $(-t^*,t^*)$. It follows that there only exists one global minimum, and that it belongs to this interval. ■



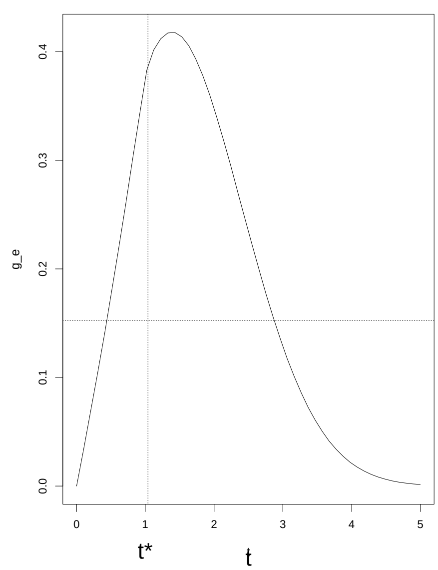
(a) BP = 50%, $\epsilon = 0.05$



(b) BP = 50%, $\epsilon = 0.10$



(c) BP = 40%, $\epsilon = 0.10$



(d) BP = 40%, $\epsilon = 0.15$

Figure 1: Plots of $g_\epsilon(t) = \inf_{s^-(t) \le s \le s^+(t)} \left[-E_\Phi \rho_d' \left(\frac{X-t}{s} \right) \right]$ for estimates with breakdown point 50 and 40%. The threshold t^* is given by (17). The horizontal line is at $\epsilon / (1-\epsilon) \sup_x \rho_d'(x)$.

Lemma 8 *Let $\rho : \mathbb{R} \to \mathbb{R}_+$ satisfy (17) or (57). Then there exists a compact set $I_2 \subset \mathbb{R}$ (independent from $F \in \mathcal{H}_\epsilon$) such that for all $\delta > 0$ there exists m_0 (that only depends on δ, i.e. $m_0 = m_0(\delta)$) such that*

$$P_F\left[\tilde{\mu}_n \in I_2\,, \quad \forall\, n \geq m\right] > 1 - \delta\,, \qquad \forall\, m \geq m_0\,, \qquad \forall\, F \in \mathcal{H}_\epsilon\,. \tag{62}$$

Proof: Let $\gamma(F,t,s) = E_F\rho(X,t,s)$. We will first show that either (17) or (57) imply that

$$\gamma(F, 0, \boldsymbol{\sigma}(F)) < \gamma(F, t, \boldsymbol{\sigma}(F)) \qquad \forall\ |t| \geq t^* \ \forall F \in \mathcal{H}_\epsilon\,. \tag{63}$$

Note that

$$\gamma(F,t,s) = (1-\epsilon)\ E_\Phi\rho(X,t,s) + \epsilon\, E_H\rho(X,t,s)\ ,$$

for some distribution function H. It follows that

$$\gamma(F, 0, \boldsymbol{\sigma}(F)) \leq (1-\epsilon)\ E_\Phi\rho(X, 0, s) + \epsilon\,. \tag{64}$$

From (17) or (57) we have that

$$E_\Phi\rho(X, t, \boldsymbol{\sigma}(F)) - \frac{\epsilon}{1-\epsilon} > E_\Phi\rho(X, 0, \boldsymbol{\sigma}(F))\ , \qquad \forall\ |t| \geq t^*\,. \tag{65}$$

From (64) and (65) we have, for all $|t| \geq t^*$

$$\begin{aligned}
\gamma(F, 0, \boldsymbol{\sigma}(F)) &< (1-\epsilon)\ \left[E_\Phi\rho(X, t, \boldsymbol{\sigma}(F)) - \frac{\epsilon}{1-\epsilon}\right] + \epsilon \\
&= (1-\epsilon)\ E_\Phi\rho(X, t, \boldsymbol{\sigma}(F)) \\
&\leq (1-\epsilon)\ E_\Phi\rho(X, t, \boldsymbol{\sigma}(F)) + \epsilon\, E_H\rho(X, t, \boldsymbol{\sigma}(F)) \\
&= \gamma(F, t, \boldsymbol{\sigma}(F))\ ,
\end{aligned}$$

and that shows (63). Let $I_2 = [-t^*, t^*]$. We will now show that for any $\delta > 0$ there exists $m_0 = m_0(\delta)$ such that for all $m \geq m_0$

$$P_F\Big(\tilde{\mu}_n \in I_2\,, \quad \forall\, n \geq m\Big) > 1 - \delta\,,$$

where neither I_2 nor m_0 depend on $F \in \mathcal{H}_\epsilon$. We will do it by showing that there exists n_0 (independent from $F \in \mathcal{H}_\epsilon$) such that with arbitrarily high probability we have

$$\gamma_n(0, \hat{\sigma}_n) < \gamma_n(t, \hat{\sigma}_n)\ , \quad \forall\, t \notin I_2\,. \tag{66}$$

It is easy to see that (66) implies that for all $n \geq n_0$ and with high probability we have $\tilde{\mu}_n \in I_2$.

Note that the function

$$a(s,t) = E_\Phi\rho\left(\frac{X-t}{s}\right) - E_\Phi\rho\left(\frac{X}{s}\right)$$

is non-decreasing in $|t|$ for each fixed s. Hence, from (17) there exists $\tilde{\delta} > 0$ (independent from t) such that

$$E_\Phi\rho(X, t, s) - E_\Phi\rho(X, 0, s) > \frac{\epsilon}{1-\epsilon} + \tilde{\delta}\,, \quad \forall\ |t| \geq t^*\,, \quad \forall\, s^- \leq s \leq s^+\,.$$

Hence, for any t such that $|t| \geq t^*$ we have

$$\begin{aligned} \gamma(F,t,\boldsymbol{\sigma}) &\geq (1-\epsilon)\, E_\Phi \rho(X,t,\boldsymbol{\sigma}) \\ &> (1-\epsilon)\left(\frac{\epsilon}{1-\epsilon} + \tilde{\delta} + E_\Phi \rho(X,0,\boldsymbol{\sigma})\right) \\ &= (1-\epsilon)\,\tilde{\delta} + \epsilon + E_\Phi \rho(X,0,\boldsymbol{\sigma})\ . \end{aligned}$$

We have

$$\inf_{t \notin I_2} \gamma(F,t,\boldsymbol{\sigma}) \geq (1-\epsilon)\,\tilde{\delta} + \gamma(F,0,\boldsymbol{\sigma})\ , \quad \forall F \in \mathcal{H}_\epsilon\,. \tag{67}$$

Let $\alpha(F) = \gamma(F,0,\boldsymbol{\sigma}(F))$ and $\eta(F) = \inf_{t \notin I_2} \gamma(F,t,\boldsymbol{\sigma}(F))$. Then (67) implies that $\tilde{\alpha} = \inf_{F \in \mathcal{H}_\epsilon}[\eta(F) - \alpha(F)] > 0$. Choose $0 < \tilde{\epsilon} < \tilde{\alpha}/2$.

Note that by Chebychev's inequality, for any $\tau > 0$

$$P_F\Bigg[|\gamma_n(t,s) - \gamma(F,t,s)| > \tau\Bigg] \leq \frac{1}{n}\frac{1}{\tau}\,, \qquad \forall\, t\,, \quad \forall\, s\,, \forall\, F \in \mathcal{H}_\epsilon\,.$$

Hence, for a fixed $\delta > 0$ there exists $n_0 = n_0(\delta)$ (that does not depend on $F \in \mathcal{H}_\epsilon$) such that for all $n \geq n_0$, for all $F \in \mathcal{H}_\epsilon$, and for all t and s we have

$$P_F\Bigg[|\gamma_n(t,s) - \gamma(F,t,s)| < \tilde{\epsilon}/2\Bigg] > 1 - \delta/2\,. \tag{68}$$

We also have that there exists $\tau = \tau(\tilde{\epsilon}) > 0$ such that

$$\Big|\gamma_n(t,s_1) - \gamma(F,t,s_2)\Big| < \tilde{\epsilon}/2\,, \quad \forall\, t\,, \qquad \text{if} \quad |s_1 - s_2| < \tau\,. \tag{69}$$

Because $\hat{\sigma}_n$ converges to $\boldsymbol{\sigma}(F)$ uniformly in $F \in \mathcal{H}_\epsilon$ we have that for each $\tau > 0$ there exists $n_1 = n_1(\tau)$ (independent from F) such that

$$P_F\left[\sup_{m \geq n} |\hat{\sigma}_m - \boldsymbol{\sigma}(F)| > \tau\right] < \delta/2\,, \qquad \forall\, n \geq n_1\,. \tag{70}$$

Equations (69) and (70) show that for $n \geq n_1 = n_1(\tilde{\epsilon})$ we have

$$P_F\Bigg[|\gamma_n(t,\hat{\sigma}_n) - \gamma(F,t,\boldsymbol{\sigma}(F))| < \tilde{\epsilon}/2\Bigg] > 1 - \delta/2\,, \qquad \forall\, t\,. \tag{71}$$

In particular with $t = 0$ in (71) we get

$$\gamma_n(0,\hat{\sigma}_n) < \gamma(F,0,\boldsymbol{\sigma}(F)) + \tilde{\epsilon}/2\,.$$

Similarly we have

$$\begin{aligned} \gamma_n(t,\hat{\sigma}_n) &> \gamma(F,t,\boldsymbol{\sigma}(F)) - \tilde{\epsilon}/2 \\ &> \inf_{t \notin I_2} \gamma(F,t,\boldsymbol{\sigma}(F)) - \tilde{\epsilon}/2\,. \end{aligned}$$

Hence

$$\begin{aligned} \inf_{t \notin I_2} \gamma_n(t,\hat{\sigma}_n) &> \inf_{t \notin I_2} \gamma(F,t,\boldsymbol{\sigma}(F)) - \tilde{\epsilon}/2 \\ &> \gamma(F,0,\boldsymbol{\sigma}(F)) + \tilde{\alpha} - \tilde{\epsilon}/2 \\ &> \gamma_n(0,\hat{\sigma}_n) - \tilde{\epsilon}/2 + \tilde{\alpha} - \tilde{\epsilon}/2 \\ &> \gamma_n(0,\hat{\sigma}_n)\ , \end{aligned}$$

and we see that (66) holds. ■

Lemma 9 *Let $D_1, \ldots, D_n$ be n i.i.d. random variables and let $\overline{D}_n = 1/n \sum_{i=1}^{n} D_i$. Assume that $E_F\left[D_i^2\right] \leq c < \infty$, for all $F \in \mathcal{H}_\epsilon$. Then $\overline{D}_n = UO_P(1)$ and $\overline{D}_n - E_F(D_i) = Uo_P(1)$.*

Proof: Note that the assumption on the second moment of D_i implies that $E_F|D_i| \leq 1 + c$ for all $F \in \mathcal{H}_\epsilon$. To simplify the notation, let $d = 1 + c$. Then we have

$$P_F\left[\left|\overline{D}_n\right| > 2\,d\right] \leq P_F\left[\left|\overline{D}_n - E_F D_i\right| > d\right] \leq \frac{1}{d^2}\frac{1}{n}\operatorname{Var}_F(D_i) \leq \frac{1}{d^2}\frac{1}{n} E_F D_i^2 \,.$$

Hence, $\lim_n P_F\left[\left|\overline{D}_n\right| > 2\,d\right] = 0$ for all $F \in \mathcal{H}_\epsilon$, where d does not depend on F. It follows that

$$\lim_{k\to\infty} \sup_{F\in\mathcal{H}_\epsilon} \lim_{n\to\infty} P_F\left[\left|\overline{D}_n\right| > k\right] = 0\,,$$

that is, $\overline{D}_n = UO_P(1)$. A similar argument shows that $\overline{D}_n - E_F(D_i) = Uo_P(1)$. ■

References

[1] Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, **16**, 147-185.

[2] Bednarski, T. and Zontek, S. (1996). Robust estimation of parameters in mixed unbalanced models. *The Annals of Statistics*, **24**, 1493-1510.

[3] Bickel, P.J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association*, **70**, 428-434.

[4] Boos, D.D. and Serfling, R.J. (1980). A note on differentials and the CLT and LIL for statistical functions, with applications to M-estimates. *The Annals of Statistics*, **8**, 618-624.

[5] Caroll, R. J. (1978). On almost sure expansion for M-estimates. *The Annals of Statistics*, **6**, 314-318.

[6] Carroll, R.J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *Journal of the American Statistical Association*, **74**, 674-679.

[7] Carroll, R. J. and Welsh, A. H. (1988) A Note on Asymmetry and Robustness in Linear Regression. *The American Statistician*, **42**, 285-287.

[8] Clarke, B.R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *The Annals of Statistics*, **11**, 1196-1205.

[9] Clarke, B.R. (1984). Nonsmooth analysis and Fréchet differentiability of M-functionals. Research Report. Murdoch University, Murdoch, Western Australia.

[10] Clarke, B.R. (2000). A remark on robustness and weak continuity of M-estimators. *J. Austral. Math. Soc. (Ser. A)*, **68**, 411-418.

[11] Davies, P.L. (1998). On locally uniformly linearizable high breakdown location and scale functionals. *The Annals of Statistics*, **26**, 1103-1125.

[12] Fraiman, R., Yohai, V. J. and Zamar, R. H. (2001). Optimal robust M-estimates of location. *The Annals of Statistics*. To appear.

[13] Hampel, F.R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, **42**, 1887-1896.

[14] Hampel, F.R., Ronchetti, E.Z., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics. The approach based on influence functions*. New York: Wiley.

[15] Huber, P.J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, **35**, 73-101.

[16] Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, California.

[17] Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.

[18] Markatou, M. and Hettmansperger, T.P. (1990). Robust bounded-influence tests in linear models. *Journal of the American Statistical Association*, **85**, 187-190.

[19] Maronna, R.A. and Yohai, V.J. (1981). Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **58**, 7-20.

[20] Martin, R.D. and Zamar, R. (1993). Bias robust estimation of scale. *The Annals of Statistics*, **21**, 991-1017.

[21] Rocke, D.M. and Downs, G.W. (1981). Estimating the variances of robust estimators of location: influence curve, jackknife and bootstrap. *Communications in Statistics, Part B – Simulation and Computation*, **10**, 221-248.

[22] Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series*. (J. Franke, W. Hardle and D. Martin, eds.). *Lecture Notes in Statist.*, **26** 256-272. Berlin: Springer-Verlag.

[23] Salibian-Barrera, M. (2000). Contributions to the Theory of Robust Inference. Unpublished Ph.D. Thesis. University of British Columbia, Department of Statistics, Vancouver, BC. Available on-line at `ftp://ftp.stat.ubc.ca/pub/matias/Thesis`.

[24] Simpson, D.G., Ruppert, D. and Carroll, R.J. (1992). *Journal of the American Statistical Association*, **87**, 439-450.

[25] Simpson, D.G. and Yohai, V.J. (1998). Functional stability of one-step GM-estimators in approximately linear regression. *The Annals of Statistics*, **26**, 1147-1169.

[26] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, 642-656.