The University of British Columbia
Department of Statistics
Technical Report #176

On Globally Robust Confidence Intervals for Regression Coefficients

Matías Salibián-Barrera

March 1998

1 Abstract

We will develop confidence intervals for linear regression coefficients when the parametric model is violated by the presence of a fraction ϵ of outliers and high leverage data points. Our method will be based on a robustified bootstrap technique. Unlike the classical bootstrap, our robust bootstrap does not produce unduly heavy tails or extreme re-sampled statistics when the original sample is contaminated. Classical bootstrap has not been widely applied in robust regression due to the heavy computation involved in the calculation of robust estimators. Our method greatly simplifies the calculations making robust inference a real alternative for the practitioner concerned with the validity of the underlying regression model.

2 Current state of knowledge

We will work in an ϵ-contamination neighborhood of a central distribution F_0. We will assume that the distribution F that generates the data satisfies

$$F \in \mathcal{Q}_\epsilon(F_0) = \{H : H = (1-\epsilon)F_0 + \epsilon G,\ G \text{ an arbitrary distribution}\} \qquad (1)$$

for a fixed $0 \leq \epsilon < 1/2$. Roughly speaking, ϵ is the amount of contamination that affects our distribution of interest F_0.

The theory of robust point estimation is fairly well developed. In what follows I am going to make a short review of some proposals for robust point estimates that have appeared in the literature. In the next section I will comment on the published results and reports on robust inference.

2.1 Robust Point Estimation

2.1.1 M-estimates for location and scale

Huber (1964) introduced the class of M-estimates. Suppose that $x_1, \ldots, x_n$ are independent and identically distributed with density $f(x, \theta)$. The M-estimate of θ is

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

where ρ is a user-chosen function. When $\rho(x,\theta) = -\log f(x,\theta)$, T_n is the maximum likelihood estimate of θ. We can also define T_n implicitly by the equation

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

where $\psi = \partial\rho/\partial\theta$.

Assume the following location model. Let $x_1, \ldots, x_n$ be a random sample satisfying

$$x_i = \mu + \epsilon_i \quad i = 1, \ldots, n \qquad (2)$$

where ϵ_i are independent errors with common distribution F_0 and finite variance σ^2. The parameter of interest is μ. The cumulative distribution function F of the x_i's satisfies $F(x) = F_0((x-\mu)/\sigma)$. In this case it is natural to take $\psi(\cdot,\cdot)$ to be a function of a single argument, namely the difference $x_i - \theta$. The defining equation becomes

$$\sum_{i=1}^{n} \psi(x_i - T_n) = 0. \tag{3}$$

One important drawback of this definition is that in general, if b denotes a real number,

$$T_n(b\ x_1, \ldots, b\ x_n) \neq b\ T_n(x_1, \ldots, x_n).$$

That is, the resulting estimate might not be scale equivariant.

To overcome this problem we couple our equation with a scale estimate $S_n = S_n(x_1, \ldots, x_n)$ that satisfies

$$S_n(ax_1, \ldots, ax_n) = aS_n(x_1, \ldots, x_n)$$

and instead of using equation (3) we solve

$$\frac{1}{n}\sum_{i=1}^{n} \psi((x_i - T_n)/S_n) = 0. \tag{4}$$

Possible choices for S_n are the M-estimates of scale

$$\frac{1}{n}\sum_{i=1}^{n} \chi((x_i - T_n)/S_n) = b \tag{5}$$

where b is a constant and χ is again a user-chosen function.

The procedure of simultaneously solving equations (4) and (5) is known in the literature as Huber's Proposal 2 (see Huber, 1964). However, on top of the numerical inconvenience of solving this system of two non-linear equations, the robustness properties of the resulting estimates are not satisfactory.

Another choice for the scale estimator is to solve the following equation independently of (4)

$$\frac{1}{n}\sum_{i=1}^{n}\chi\left(\frac{x_i-\tilde{T}_n}{S_n}\right)=b \tag{6}$$

where $\tilde{T}_n$ is an auxiliary location estimate that is generally different from T_n in (4). For example, we can use $\tilde{T}_n = \text{median}\,(x_i)$.

Under very mild conditions the estimates T_n and S_n are strongly consistent, i.e., $T_n \to \mu$ and $S_n \to \sigma$ almost surely. It is well known (see for example Huber, 1981) that if ψ is odd, χ is even and the underlying distribution is symmetric, S_n and T_n are asymptotically independent and T_n satisfies

$$\sqrt{n}\,(T_n-\mu) \longrightarrow N\,(0,\ \mathrm{AV}\,(F,T))\,,$$

with asymptotic variance given by

$$\mathrm{AV}\,(F,T)=\sigma^2\ \frac{E_F\psi^2\left(\left(X-\mu\right)/\sigma\right)}{\left(E_F\psi'\left(\left(X-\mu\right)/\sigma\right)\right)^2}. \tag{7}$$

2.1.2 GM-estimates for Regression

In the following we will assume a regression model with random explanatory variables. Let $(y_i, \mathbf{x}_i)$ be random vectors in $\mathrm{I\!R}^{p+1}$ satisfying the model

$$y_i=\mathbf{x}_i'\theta+\sigma\epsilon_i, \qquad i=1,\ldots,n, \tag{8}$$

where $\mathbf{x}_i \in \mathrm{I\!R}^p$. We will assume that the errors ϵ_i are independent of $\mathbf{x}_i$ and that they have a symmetric distribution around zero, with variance $\sigma^2 < \infty$. The parameter of interest is θ.

If we denote the density of the ϵ_i by g, and k is the density (in $\mathrm{I\!R}^p$) of the explanatory variables, then for each θ the joint distribution of the vector $(y, \mathbf{x})$ is given by

$$f_\theta\,(y,\mathbf{x})=\sigma^{-1}\ g\left(\left(y-\mathbf{x}'\theta\right)/\sigma\right)\ k\,(\mathbf{x})\,.$$

For each vector $\beta \in \mathbb{R}^p$, denote the corresponding residuals by

$$r_i(\beta) = y_i - \mathbf{x}_i'\beta. \tag{9}$$

Define the GM-estimate of θ (see Krasker and Welsch, 1982, Maronna and Yohai, 1981 and Huber *et. al.*, 1986) as the solution for $\mathbf{T}_n$ of

$$\sum_{i=1}^{n} \phi\left(\mathbf{x}_i, \frac{y_i - \mathbf{T}_n'\mathbf{x}_i}{S_n}\right)\mathbf{x}_i = 0 \tag{10}$$

where the function $\phi(\cdot,\cdot) : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}$ satisfies:

- for all $\mathbf{x} \in \mathbb{R}^p$ $\phi(\mathbf{x}, \cdot)$ is odd, uniformly continuous and
- $\phi(\mathbf{x}, r) \geq 0$ for $r \geq 0$.

As before, S_n is an estimate of σ. It can be defined by an equation of the form

$$\frac{1}{n}\sum_{i=1}^{n} \chi\left(\frac{y_i - \mathbf{T}_n'\mathbf{x}_i}{S_n}\right) = E_G\chi(e). \tag{11}$$

By varying ϕ we obtain different types of estimates. For example, if ϕ only depends on the residuals r_i, i.e. $\phi(\mathbf{x}, r) = \psi(r)$ for some function ψ, we get the class of M-estimates for regression. More generally, all the proposals can be written as

$$\phi(\mathbf{x}, r) = w(\mathbf{x}) \cdot \psi(r\ v(\mathbf{x}))$$

for different functions $w : \mathbb{R}^p \to \mathbb{R}$ and $v : \mathbb{R}^p \to \mathbb{R}$. The main idea is to penalize not only those observations with large residuals $r_i(\mathbf{T}_n) = y_i - \mathbf{T}_n'\mathbf{x}_i$ but also the ones with high leverage (see Weisberg, 1985, page 111).

Mallow's and Andrews' proposal corresponds to $v(\mathbf{x}) = 1$ (see Hill, 1977). Schweppe's function $\phi(\mathbf{x}, r) = w(\mathbf{x})\ \psi(r/w(\mathbf{x}))$ is obtained when $v(\mathbf{x}) = 1/w(\mathbf{x})$ (see Merrill and Schweppe, 1971). See also Hampel *et. al.* (1986) for a more detailed discussion.

Maronna and Yohai (1981) showed that if the system of equations

$$E_F \phi \left(\mathbf{x}, \frac{y - \beta' \mathbf{x}}{\sigma} \right) \mathbf{x} = \mathbf{0} \tag{12}$$

$$E_F \chi \left(\frac{y - \beta' \mathbf{x}}{\sigma} \right) = E_G \chi \left(e \right) \tag{13}$$

has a unique solution $\eta_0 = (\beta_0, \sigma_0)$, then the GM estimates defined by (10) and (11) are consistent for η_0 and asymptotically normal, with covariance matrix

$$A^{-1} \; E_F \left(\Psi \left(\mathbf{x}, y, \eta_0 \right) \Psi \left(\mathbf{x}, y, \eta_0 \right)' \right) \; \left(A^{-1} \right)^t, \tag{14}$$

where

$$A = E_F J \left(\mathbf{x}, y, \eta_0 \right),$$

Ψ is a $p+1$ vector with the functions involved in the equations (12) and (13):

$$\Psi \left(\mathbf{x}, y, \eta \right) = \begin{pmatrix} \phi \left(\mathbf{x}, r \left(\eta \right) / \sigma \right) x_1 \\ \phi \left(\mathbf{x}, r \left(\eta \right) / \sigma \right) x_2 \\ \vdots \\ \phi \left(\mathbf{x}, r \left(\eta \right) / \sigma \right) x_p \\ \chi \left(r \left(\eta \right) / \sigma \right) \end{pmatrix}$$

and J is the $(p+1) \times (p+1)$ matrix of derivatives of Ψ with respect to $\eta = (\beta, \sigma)$. This formula simplifies when the distribution G of the errors is symmetric. In this case, $\mathbf{T}_n$ and S_n are asymptotically independent and the covariance matrix (14) is

$$B^{-1} E \left(\phi \left(\mathbf{x}, \frac{y - \beta_0' \mathbf{x}}{\sigma_0} \right)^2 \mathbf{x}\mathbf{x}' \right) B^{-t},$$

where

$$B = E \left(\phi' \left(\mathbf{x}, \frac{Y - \theta_0' \mathbf{x}}{\sigma_0} \right) \mathbf{x}\mathbf{x}' \right)$$

and $\phi' \left(\mathbf{x}, r \right) = \partial \phi / \partial r \left(\mathbf{x}, r \right)$.

I want to stress that the uniqueness of the solution to equations (12) and (13) is a strong condition. Two sufficient conditions for this property to hold are that the distribution of the errors is symmetric and that $\phi(\mathbf{x}, \cdot)$ is increasing for each $\mathbf{x}$ (Maronna and Yohai, 1981). See also Yohai and Maronna (1979) for the case when the explanatory variables are fixed (symmetry of the distribution of the errors is also needed to obtain consistency here).

A global measure of robustness is the breakdown-point (BP). Donoho and Huber (1983) gave the following definition for finite samples. Let $Z_n = (z_1, \ldots, z_n)$ be a random sample and $T_n(Z_n)$ be the estimate calculated with the sample Z_n. For each integer m let

$$b(m, T, Z_n) = \sup |T_{m+n}(Z_n, W_m) - T_n(Z_n)|$$

where the supremum is calculated over all the samples W_m of size m. Define the breakdown-point of T for the sample Z_n as

$$\epsilon_n^*(T, Z_n) = \min\{m/(m+n) \ : \ b(m, T, Z_n) = \infty\}.$$

The breakdown point is the smallest proportion of arbitrary observations that the estimator can resist without becoming unbounded.

We have the following result for the breakdown of the estimators defined by (10) and (11): BP > 0 if and only if $\sup_{\mathbf{x},r} |\mathbf{x}\phi(\mathbf{x}, r)| < \infty$. In this case the breakdown point is positive but it decreases to zero as the number of predictors increases (roughly as $1/(p+1)$) (see Maronna, Bustos, and Yohai, 1979). We see then that the BP of the GM-estimates cannot be 0.5 when $p > 1$. It is of interest to have a class of regression estimates with high breakdown point independent of the number of explanatory variables. The S-estimates for regression have this property.

2.1.3 S-estimates for Regression

The S-estimates for regression were introduced by Rousseeuw and Yohai (1984). The idea is to define $\mathbf{T}_n$ as the set of parameters β that produces residuals with the smallest dispersion. More precisely, suppose that $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ satisfy model (8) and let $r_i(\beta)$ be defined by (9). The M-scale, $s_n(\beta)$, of the residuals $(r_1(\beta), \ldots, r_n(\beta))$ is defined as the solution to the equation

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{s_n(\beta)}\right) = E_G(\rho(\epsilon)).$$

The corresponding regression S-estimate is now defined by the property of minimizing $s_n(\beta)$, that is,

$$\mathbf{T}_n = \arg\min_{\beta \in \mathbb{R}^p} s_n(\beta).$$

S-estimates for regression are consistent for the true parameter θ and asymptotically normal when the distribution of the errors is symmetric around zero. But these estimates cannot achieve high efficiency and high breakdown-point at the same time. In fact, there exists a trade-off between efficiency and robustness (see Hössjer, 1992).

MM- and τ-estimates were introduced in order to obtain estimators with high efficiency and breakdown-point simultaneously.

2.1.4 MM-estimates for Regression

In 1987 Yohai introduced the MM-estimates for regression. Assume the regression model (8) and recall that G denotes the distribution of ϵ_i. Consider two functions $\rho_0, \rho_1 : \mathbb{R} \to \mathbb{R}$ such that

$$E_G(\rho_0(\epsilon)) = 0.5, \qquad \rho_1(u) \le \rho_0(u) \;\; \forall u \in \mathbb{R} \qquad \text{and} \qquad \sup_{x \in \mathbb{R}} \rho_0(x) = \sup_{x \in \mathbb{R}} \rho_1(x)$$

The MM-estimate $\mathbf{T}^1$ is defined in three steps as follows:

1. Compute an initial regression estimate $\mathbf{T}^0$
2. Compute the M-scale of the residuals $r_i\left(\mathbf{T}^0\right)$, S_n, using the function ρ_0.
3. Define $\mathbf{T}^1$ as any solution to the equation

$$\sum_{i=1}^{n} \rho_1'\left(r_i\left(\mathbf{T}^1\right)/S_n\right)\mathbf{x}_i = 0$$

that also satisfies

$$\sum_{i=1}^{n} \rho_1\left(r_i\left(\mathbf{T}^1\right)/S_n\right) \leq \sum_{i=1}^{n} \rho_1\left(r_i\left(\mathbf{T}^0\right)/S_n\right).$$

These estimates have high breakdown point and high efficiency at the central model. Under the usual regularity conditions, including that the distribution of the errors is symmetric, these estimates are strongly consistent and asymptotically normal, with variance depending only on the limiting value of the scale estimator S_n.

2.1.5 τ-estimates for Regression

In 1988, Yohai and Zamar introduced the class of τ-estimates. Assume the linear model (8) and consider two functions $\rho_1, \rho_2 : \mathbb{R} \to \mathbb{R}$. For each $\beta \in \mathbb{R}^p$ define the τ estimate of the scale of the residuals $\left(r_1(\beta), \ldots, r_n(\beta)\right)$ by

$$\tau_n^2(\beta) = s_n^2(\beta)\, \frac{1}{n}\sum_{i=1}^{n} \rho_2\left(\frac{r_i(\beta)}{s_n(\beta)}\right)$$

where $s_n(\beta)$ is the M-scale of the residuals calculated using the function ρ_1.

The τ-estimates for regression are then defined as

$$\hat{\theta} = \arg\min_{\beta} \tau_n(\beta).$$

In particular, $\tau_n(\hat{\theta})$ is an efficient and robust estimate of σ. As before, under certain regularity conditions (which include symmetry of the distribution of the residuals) we get strong consistency of these estimates and asymptotic normality.

2.2 Robust Inference

Robust inference has received little attention in the literature beyond point estimation. I will first comment on some proposals for robust testing and then on studies of the variability of robust estimates.

2.2.1 Robust Tests

Consider the location model (2) with known σ. We can therefore assume that the cumulative distribution function of our observations F satisfies $F(x) = F_0(x - \mu)$ for some fixed distribution function F_0 and location parameter μ. We want to test the null hypothesis $H_0: \ \mu = \mu_0$ versus a one- or two-sided alternative. If T_n is the estimate defined by the equation (3) it follows that under H_0 the statistic

$$E_n = \sqrt{n}\,(T_n - \mu_0)$$

is asymptotically normal. We can use a test of the form: reject H_0 when $E_n > C$ (or when $|E_n| > C$ if the alternative hypothesis is two-sided). The problem with this approach is that there is no known formula for the asymptotic variance of E_n under H_0 when the scale has to be estimated and F_{μ_0} is asymmetric.

For the same location model with known scale, other proposed test statistics (Ronchetti, 1979, and Sen, 1982) are of the form

$$E_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\,(X_i - \mu_0)$$

where ψ satisfies $\int \psi\,(u - \mu_0)\ dF_{\mu_0} = 0$. Then, under the null hypothesis, we have

$$E_n \longrightarrow N\left(0, \int \psi^2\,(u - \mu_0)\ dF_{\mu_0}\right)$$

and we can proceed as in the previous case. The same criticism with respect to the of estimation of σ under asymmetric distributions applies here.

For the linear model (8), Ronchetti (1982) introduced the class of τ-tests. They are defined as follows. Let $\tau(\cdot,\cdot) : \mathcal{R}^p \times \mathcal{R} \to \mathcal{R}$ be a function such that $\tau(\mathbf{x}, r) \geq 0$, $\tau(\mathbf{x}, 0) = 0$, and $\tau(\mathbf{x}, \cdot)$ is differentiable for all $\mathbf{x} \in \mathcal{R}^p$. Let $\theta \in \mathrm{I\!R}^p$ be the vector of regression parameters, and let $\theta = (\theta_1, \theta_2)$ with $\theta_1 \in \mathrm{I\!R}^{p-q}$ and $\theta_2 \in \mathrm{I\!R}^q$. Consider the null hypothesis $H_0 :\ \theta_2 = 0$. Let $r_i(\beta)$ be the residuals as defined in (9). The test statistic is

$$S_n^2 = \frac{2}{q}\frac{1}{n}\sum_{i=1}^{n}\left\{\tau\left(\mathbf{x}_i, r_i\left(T_{\omega n}\right)/\sigma\right) - \tau\left(\mathbf{x}_i, r_i\left(T_{\Omega n}\right)/\sigma\right)\right\}$$

where $T_{\omega n}$, $T_{\Omega n}$ are the GM-estimates calculated with $\phi(\mathbf{x}, r) = \partial\tau(\mathbf{x}, r)/\partial r$ (see section (2.1.2)) for the reduced and full models respectively. This test statistic can also be thought as the robust equivalent to the classical F test. To see the relationship note that instead of minimizing the residual sum of squares, $T_{\omega n}$ and $T_{\Omega n}$ satisfy

$$T_{\omega n} = \arg\min_{\beta\in\Theta_\omega}\sum_{i=1}^{n}\tau\left(\mathbf{x}_i, r_i(\beta)/\sigma\right) \quad \text{and} \quad T_{\Omega n} = \arg\min_{\beta\in\Theta}\sum_{i=1}^{n}\tau\left(\mathbf{x}_i, r_i(\beta)/\sigma\right)$$

where Θ_ω is the reduced parameter space.

This test assumes that the scale parameter σ is known. As before, under symmetric distributions, the simultaneous estimation of σ does not change the asymptotic results. No suggestion is made on how to proceed when the distribution of the errors is not symmetric.

The asymptotic distribution of the statistic S_n^2 under the null hypothesis is that of a linear combination of χ^2 random variables,

$$\sum_{j=p-q+1}^{p}\lambda_j N_j^2$$

where N_j are independent standard normal random variables; $\lambda_{p-q+1} \geq \lambda_{p-q+2} \geq$

$\ldots\lambda_p > 0$ are the q positive eigenvalues of the matrix

$$Q\left[M^{-1} - \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}\right],$$

M_{11}^{-1} is the inverse of the upper $(p-q+1)\times(p-q+1)$ part of the matrix

$$M = E\left(\phi'\left(\mathbf{x}, r\right)\mathbf{x}\mathbf{x}'\right), \tag{15}$$

Q is defined as

$$Q = E\left(\phi^2\left(\mathbf{x}, r\right)\mathbf{x}\mathbf{x}'\right) \tag{16}$$

and $\phi' = \partial\phi\left(\mathbf{x}, r\right)/\partial r$.

Other proposals to test the same null hypothesis $H_0:\ \theta_2 = 0$ are the aligned GM tests (see Markatou and Hettmansperger, 1990). They are defined as follows. Assume that σ^2 is known and hence, without loss of generality, let $\sigma = 1$. Consider the statistic

$$W_n^2 = \left[n^{-1/2}\sum_{i=1}^{n}\phi\left(\mathbf{x}_i, y_i - \mathbf{x}_i'\hat{\beta}_0\right)\mathbf{x}_i^*\right]'\ \hat{U}^{-1}\ \left[n^{-1/2}\sum_{i=1}^{n}\phi\left(\mathbf{x}_i, y_i - \mathbf{x}_i'\hat{\beta}_0\right)\mathbf{x}_i^*\right]$$

where $\hat{\beta}_0$ is the reduced model GM estimator, $\mathbf{x}_i' = (\mathbf{x}_i^{**\,\prime}, \mathbf{x}_i^{*\,\prime})$ with $\mathbf{x}_i^{**} \in \mathcal{R}^{p-q}$ and $\mathbf{x}_i^* \in \mathcal{R}^q$,

$$\hat{U}^{-1} = \hat{Q}_{22} - \hat{M}_{21}\hat{M}_{11}^{-1}\hat{Q}_{12} - \hat{Q}_{21}\hat{M}_{11}^{-1}\hat{M}_{12} + \hat{M}_{21}\hat{M}_{11}^{-1}\hat{Q}_{11}^{-1}\hat{M}_{11}^{-1}\hat{M}_{12}$$

and Q and M are defined in (16) and (15).

Markatou, Stahel and Ronchetti (1991) study two other types of tests (see also Markatou and He, 1994). The Wald-type tests are based on statistics of the form

$$T^2 = \hat{\beta}_2'V_{22}^{-1}\hat{\beta}_2$$

where $\hat{\beta}_2$ is a robust estimate of β_2 and V_{22} is the $q\times q$ sub matrix of $\hat{V}$, an estimator of the asymptotic covariance matrix. This matrix will depend on the design.

The scores (Rao) type test is based on the statistic

$$R^2 = Z_n' M_{(22.1)} V_{22}^{-1} M_{(22.1)} Z_n$$

where

$$Z_n = \frac{1}{n} \sum_{i=1}^{n} \phi\left(\mathbf{x}_i, \frac{y_i - \mathbf{x}_i' T_{\omega n}}{\hat{\sigma}}\right) \mathbf{x}_i^*$$

and

$$M_{(22.1)} = M_{22} - M_{21} M_{11}^{-1} M_{12}$$

where the matrices M_{ij} correspond to the same partition as before.

All these tests have good local robustness behavior (see the papers cited above). To study their global behavior when the fraction of contamination is relatively large He, Simpson and Portnoy (1990) introduced the concept of breakdown-point for tests. Denote Θ_0 and Θ_1 the null and alternative hypotheses respectively. The power breakdown function of a test statistic T is (see (1) for the definition of $\mathcal{Q}_\epsilon(F)$)

$$\epsilon_\theta^*(T) = \inf\left\{\epsilon > 0 : T\left(F_{\theta_0}\right) \in T\left(\mathcal{Q}_\epsilon\left(F_\theta\right)\right) \text{ for some } \theta_0 \in \Theta_0\right\}$$

similarly, the level breakdown function is

$$\epsilon_\theta^{**}(T) = \inf\left\{\epsilon > 0 : T\left(F_\theta\right) \in T\left(\mathcal{Q}_\epsilon\left(F_{\theta_0}\right)\right) \text{ for some } \theta_0 \in \Theta_0\right\}$$

The interpretation of these definitions is that for each θ, ϵ_θ^* is the minimum proportion of outliers that will cause the null hypothesis to be confounded with the alternative θ. On the other hand, ϵ_θ^{**} is the smallest amount of contamination that will result in a failure to reject H_0 when the parameter is θ.

Markatou and He (1994) studied some proposals to obtain simultaneously high breakdown points and good local behavior.

Note that with this definition of breakdown point, if the boundaries of Θ_0 and Θ_1 intersect the resulting breakdown point is zero. This is due to the fact that if the asymptotic bias is positive, for any proportion of outliers there exists a contamination that will drive the estimator towards the wrong region.

This last remark can be restated in terms of confidence intervals. If we construct a classical type confidence interval centered at a robust estimate T_n

$$\left(T_n - a\ \mathrm{SE}/\sqrt{n}, T_n + a\ \mathrm{SE}/\sqrt{n}\right)$$

its asymptotic level will be zero, because as $n \to \infty$ the confidence interval converges to the point T_∞ which is typically different from the parameter of interest μ_0.

2.2.2 Robust Confidence Intervals and Estimation of Variance

Gross (1976) studied confidence intervals robustness for heavy tailed symmetric distributions, following an extensive Monte Carlo study, carried out at Princeton University, and published by Andrews *et. al.* (1972). He found disappointing results for jackknifed estimates.

Carroll (1979) compared several robust estimates and techniques to estimate their variances under asymmetric distributions. He showed that the empirical approximation to the asymptotic variance severely underestimates it. He found that jackknifing yielded satisfactory results to estimate the variances in the location model, but it did not seem promising for the regression setup.

Rocke and Downs (1981) compared the empirical asymptotic variance, the jackknife and the bootstrap under a variety of different distributions both symmetric and asymmetric. In a numerical study on the location-scale model they noted a poor performance of the bootstrap variance estimator, even under symmetric distributions with slightly heavy tails. For example, they got variance estimates as high as twice the right value for samples of size 20 consisting of 19 independent observations from a N(0,1) distribution and one from a N(0,100) distribution.

This "heavy tails problem" of the bootstrap variance estimator had been studied by Gosh *et. al.* (1984) two years earlier. They noted that in estimating the variance of quantiles, a moment condition has to be satisfied to get convergence of the bootstrap variance estimators. In other words, we can have that the sequence of estimates T_n converges in distribution to a random variable Z with $\mathrm{Var}\,(Z) < \infty$, that the sequence of bootstrap estimates T_n^* converges in distribution to the same limit, but that the sequence $Var_*\,(T_n^*) \to \infty$. We can illustrate this with the following diagram

$$\begin{array}{ccl} \mathcal{L}\,(T_n) & & \\ & \searrow & \\ & & Z \quad \text{and } \mathrm{Var}\,(Z) < \infty \text{ but } \mathrm{Var}_*\,(T_n^*) \to \infty, \\ & \nearrow & \\ \mathcal{L}_*\,(T_n^*) & & \end{array}$$

where $\mathcal{L}\,(T)$ denotes the distribution of the statistic T and $\mathcal{L}_*$ and Var_* denote the bootstrap distribution and variance, respectively.

This disturbing fact can pose practical problems for other estimates, as we will see later. In light of this result it becomes clear that the problem faced by Rocke and Downs' bootstrap estimate was due to the heavy tails of that particular sampling scheme (19 standard normal plus one N(0,100)).

More recently Shao (1990, 1992) studied the estimation of the variance of Frechet-differentiable statistical functionals (see Fernholz, 1983) when the distribution generating the data has heavy tails. He proposed a modified bootstrap variance estimate to overcome the "heavy tails problem" mentioned above. This modified bootstrap variance estimator truncates the values of the re-sampled estimates to down-weight its tails. Even though this estimate is asymptotically consistent, the author does not give any guideline on how to proceed in a practical situation. Moreover, the proposed truncation limits are fixed and do not depend on the distribution of the data. Asymptotically this is not relevant, but for a finite sample application it can be of critical importance.

Another serious problem with the use of bootstrap is the number of times we need to calculate the estimator. Recently Hu and Zidek (1995) and Hu and Kalbfleisch (1997) proposed to resample the equations defining the estimates. In this way, if we are interested in a one dimensional parameter, the scale is known and the estimating equation is monotone then we only need to solve the estimating equation twice to obtain the percentiles for the confidence intervals. This is a very interesting idea to reduce effectively the computing effort to perform bootstrap estimation. Unfortunately our regression model does not satisfy those conditions.

An intuitively appealing alternative for robust inference is the following. First, based on an objective or subjective rule determine which observations are outliers and discard them. Then apply classical analysis methods to the "cleaned" data set. The main objection to this approach is that the resulting confidence level and efficiency have not been determined yet (see Hampel *et. al.*, 1986, pages 62-71 for a detailed discussion). This seems to be a problem too complex to be tackled in the near future. The resulting confidence intervals have unknown actual coverage (one must consider the variability of the whole procedure, including the rejection of outliers) and the resulting confidence intervals are optimistically short. Finally, the bias of the resulting estimate that centers the interval is also unknown and cannot be sharply bounded.

In summary, we reach the following conclusions. First, the simple method of approximating the asymptotic variances with formulae derived under symmetry assumptions might not work. Valid formulae are very involved, difficult to obtain when at all, and difficult to estimate. Second, special attention should be paid to problems arising from heavy tailed distributions. Finally, the computing effort needed to re-sample robust estimates for regression is a serious concern.

We note that a reliable, theoretically justified and computationally feasible technique to calculate robust confidence intervals is missing in the literature.

3 Objectives

Most of the published results on robust inference only apply under the assumption that the distribution is symmetric. This assumption is unrealistic if we suspect that the data contain outliers. On the other hand, re-sampling methods like the classical bootstrap do not rely on that assumption but need a computational effort that makes their use unfeasible. Moreover, they might not perform well when the distribution of the data has heavy tails.

We wish to fill this gap in the robust literature by providing feasible robust confidence intervals for regression coefficients when the scale is unknown and the underlying distribution is not necessarily symmetric. These confidence intervals will be relatively easy to compute. They will be robust under global contamination neighborhoods. We will also pay attention to their efficiency and length. To be of practical relevance they should also be roughly comparable with the classical methods when applied to clean data.

4 Significance

We will provide a robust inference method that can be used in practice under reasonable assumptions. The implementation of our method will be feasible in terms of computing time. It will provide conservative confidence intervals with reasonable coverage levels, and at the same time will be efficient when no contamination is present in the data.

We expect our method to be widely used in practice. We will provide a new way of performing robust inference in linear regression models. The theoretical results justifying our approach will be valid under mild assumptions. Unlike the classical bootstrap, our technique will be feasible in practical applications. We will develop a reliable, justified and usable tool to investigate the effect of atypical data points on linear regression analyses.

5 Approach

One way to define robust confidence intervals is as follows. For a fixed significance level α and a robust estimate T_n we need to find the percentile q_n such that

$$1-\alpha = P\left(\left|T_n-\mu_0\right|<q_n\right) \tag{17}$$

where μ_0 is the parameter of interest. Assume that the sequence T_n converges almost surely to a limit T_∞. This limit need not be equal to μ_0 due to the possible asymptotic bias of T_n. Denote this bias by $B\left(T,F\right)=T_\infty-\mu_0$. Assume that T_n has a normal asymptotic distribution

$$\sqrt{n}\ \left(T_n-T_\infty\right)\longrightarrow N\left(0,AV\left(T,F\right)\right).$$

Equation (17) is equivalent to

$$P\left(-\sqrt{n}\left(\frac{q_n+B\left(T,F\right)}{\sqrt{AV\left(T,F\right)}}\right)<\sqrt{n}\left(\frac{T_n-T_\infty}{\sqrt{AV\left(T,F\right)}}\right)<\sqrt{n}\left(\frac{q_n-B\left(T,F\right)}{\sqrt{AV\left(T,F\right)}}\right)\right).$$

Then q_n satisfies

$$1-\alpha=\Phi\left(\sqrt{n}\left(\frac{q_n-B\left(T,F\right)}{\sqrt{AV\left(T,F\right)}}\right)\right)-\Phi\left(-\sqrt{n}\left(\frac{q_n+B\left(T,F\right)}{\sqrt{AV\left(T,F\right)}}\right)\right)+o\left(1\right)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable.

On the other hand, define the asymptotic quantile Q_n as the solution to

$$1-\alpha=\Phi\left(\sqrt{n}\left(\frac{Q_n-B(T,F)}{\sqrt{AV(T,F)}}\right)\right)-\Phi\left(-\sqrt{n}\left(\frac{Q_n+B(T,F)}{\sqrt{AV(T,F)}}\right)\right).$$

The relationship between the desired quantile, q_n, and the asymptotic one, Q_n, is given by

$$|q_n-Q_n|=o\left(1/\sqrt{n}\right)$$

(see Fraiman *et. al.*, 1998).

To use these confidence intervals in practice we need to estimate the asymptotic bias $B(T,F)$ and asymptotic variance $AV(T,F)$. Unfortunately there is no method available to estimate the asymptotic bias. Moreover, we conjecture that this bias is not estimable for ϵ-contamination neighborhoods.

A conservative solution to this problem is to bound $B(T,F)$ by the maximum asymptotic bias on the neighborhood: $B(\epsilon)=\sup_{F\in\mathcal{Q}_\epsilon} B(T,F)$. Formulae to calculate $B(\epsilon)$ are available for a wide range of robust estimators (see for example Martin *et. al.*, 1989; Martin and Zamar, 1993; Croux *et. al.*, 1996 and Barrendero and Zamar, 1998).

Fraiman *et. al.* (1998) proposed to bound $AV(T,F)$ by its maximum in the neighborhood in order to find optimal estimates. Our approach will be based on the estimation of the asymptotic variance.

In order to estimate this variance we can consider at least three different approaches: 1) we can use the empirical approximation to the asymptotic variance; 2) we can use classical bootstrap; and 3) we can use our robust bootstrap.

5.1 The Empirical Approximation of the Asymptotic Variance

In order to use the empirical approximation we need, of course, a formula for the asymptotic variance. If the distribution that generates the data is not symmetric, using the asymptotic variance derived under the assumption of symmetry typically underestimates the actual asymptotic variance (Rocke and Downs, 1986; Carroll, 1979).

It is not trivial to derive a formula for the asymptotic variance when we do not assume symmetry. In this case the asymptotic distribution of the scale estimator becomes entangled with that of the parameter of interest. In turn, the distribution of the scale estimate will typically depend on that of an auxiliary position estimate. The asymptotic distribution of this other auxiliary estimator will also be relevant. In this way one enters a "loop" of dependence relations with no clear ending.

We found that one can stop this recursion by using S-estimates. I will discuss our findings in the context of the location-scale model (2). Let $(x_1, \ldots, x_n)$ be a random sample. Define the S-estimate of location, $\tilde{T}_n$, as

$$\tilde{T}_n = \arg\min_{t \in \mathcal{R}} S(t)$$

where $S(t)$ satisfies

$$\frac{1}{n} \sum_{i=1}^{n} \chi \left(\frac{x_i - t}{S(t)} \right) = b$$

for each t. The S-scale, S_n, is then given by

$$S_n = S\left(\tilde{T}_n\right). \tag{18}$$

See also section (2.1.3) for the definition of S-scales in the regression model.

In the following theorem we establish a formula for the asymptotic variance of the location estimate defined by equation (4) when S_n is given by (18). This formula applies to any distribution F in the neighborhood. The natural extension to the regression model (8) holds. For simplicity, in the next theorem I only present the result for the location-scale model.

Theorem 1 *Let ψ and χ be functions with finite first derivative everywhere. Let $\tilde{T}_n$ be the S-estimate of position and S_n the corresponding S-estimate of scale, as defined in (18). Let T_n be the M-estimate of location given by (4). Assume that the almost sure finite limits of the sequences T_n, $\tilde{T}_n$ and S_n are μ_0, $\tilde{\mu}_0$ and σ respectively. Then*

$$\sqrt{n}\left(T_n - \mu_0\right) \longrightarrow N\left(0, \sigma^2 E_F\left(Z^2\right)\right),$$

where

$$Z = H\left(F\right)\psi\left(\frac{X-\mu_0}{\sigma}\right) - J\left(F\right)\left(\chi\left(\frac{X-\tilde{\mu}_0}{\sigma}\right) - b\right),$$

$$H\left(F\right) = 1/E_F\left(\psi'\left(\left(X-\mu_0\right)/\sigma\right)\right),$$

and

$$J\left(F\right) = \frac{E_F\left(\psi'\left(\left(X-\mu_0\right)/\sigma\right)\left(X-\mu_0\right)/\sigma\right)}{E_F\left(\chi'\left(\left(X-\tilde{\mu}_0\right)/\sigma\right)\left(X-\tilde{\mu}_0\right)/\sigma\right)} \times H\left(F\right).$$

The natural estimate for $\sigma^2 E_F\left(Z^2\right)$ is $S_n^2 E_{F_n}\left(Z^2\right)$. A serious problem we found with the use of this formula is that its denominators can become close to zero yielding extremely large estimates for the asymptotic variance. This happens relatively often for small sample sizes $(n \leq 30)$. We have performed a small Monte Carlo simulation study, and our results confirmed this remark. The following tables contain descriptive summaries of some results of our numerical study. We generated 10,000 random samples of size 30 and 10,000 samples of size 50 for each of the following distributions: standard normal, χ_3 and log-normal. For each of the six possible combinations of distribution and sample size we estimated the asymptotic variance using the formula suggested by the theorem above. We report the minimum, first quartile, median, third quartile and the maximum of these 10,000 variance estimates. Note that, especially for $n = 20$, the maximum values are much larger than the true value given by the theorem.

Normal - Asymp. Var.: 1.027

	Asymptotic Variances				
n	Min	Q1	Q2	Q3	Max
20	0.38	0.86	1.00	1.14	2.41
50	0.61	0.94	1.02	1.10	1.61

χ_3 - Asymp. Var.: 2.410

	Asymptotic Variances				
n	Min	Q1	Q2	Q3	Max
20	0.68	1.82	2.23	2.76	9.66
50	1.10	2.06	2.34	2.67	4.75

Log-Normal - Asymp. Var.: 1.297

	Asymptotic Variances				
n	Min	Q1	Q2	Q3	Max
20	0.32	0.90	1.18	1.57	18.07
50	0.50	1.06	1.26	1.51	3.94

5.2 Bootstrap

Since its introduction by Efron (1979), the bootstrap method has received a lot of attention in the literature (see Efron 1979, 1982; Efron and Tibshirani, 1993; Hall, 1986a, 1990, 1992 and Wu, 1986). Given a random sample $\mathbf{x}' = (x_1, \ldots, x_n)$ and a statistic $R(\mathbf{x}, F)$ that depends on the sample and possibly on the underlying distribution F, the aim is to estimate the distribution of R,

$$\mathcal{L}_F\left(R\left(\mathbf{x}, F\right)\right),$$

by that of

$$\mathcal{L}_{F_n}\left(R\left(\mathbf{x}^*, F_n\right)\right),$$

where $\mathbf{x}^*$ denotes a random sample taken from the distribution F_n. The replacement of F by F_n is called the "plug-in" principle (see Efron and Tibshirani, 1993).

Efron (1979) showed that $\mathcal{L}_{F_n}(R(\mathbf{x}^*, F_n))$ is a reasonable estimate in some simple cases and established the validity of the principle for a general class of statistics when the sampling space is finite.

When applying this principle to a particular statistic $R(\mathbf{x}, F)$ we should check whether this approximation is "good". The definition of "goodness" might depend on the particular problem at hand. For example, the criteria could be different if we want to estimate moments or percentiles of the distribution $\mathcal{L}_F(R(\mathbf{x}, F))$.

In some cases we are interested in the asymptotic distribution of

$$n^\alpha (R(\mathbf{x}, F) - \theta(F)) \tag{19}$$

where α is a real number and $\theta(F)$ denotes a parameter of the distribution F. If we want to estimate the limiting distribution of this sequence by the plug-in principle, we should show that the sequence

$$n^\alpha (R(\mathbf{x}^*, F_n) - \theta(F_n))$$

has the same limiting distribution as (19).

Bickel and Freedman (1981) give some general asymptotic theory to answer this question. They prove that under certain regularity conditions the bootstrap approximation to the asymptotic distribution works for sample means, von Mises functionals (Fernholz, 1983), quantiles, and trimmed means among others. In particular they establish the two following theorems for means and smooth functions of means respectively.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be n independent vector-valued random variables in $\mathbb{R}^p$ with common cumulative distribution function F. Let $\bar{\mathbf{x}}_n$ denote their sample mean and F_n

their empirical distribution function. Similarly, let $\mathbf{x}_1^*, \ldots, \mathbf{x}_m^*$ be m independent vector-valued random variables in $\mathbb{R}^p$ with common cumulative distribution function F_n. Let $\bar{\mathbf{x}}_m^*$ denote their sample mean. Assume that $E\left[\|\mathbf{x}_1\|^2\right] < \infty$ and let Σ be the covariance matrix of $\mathbf{x}_1$. We have the following theorem (see Bickel and Freedman, 1981):

Theorem 2 (Bickel and Freedman) *For almost all sample sequences, given* $\mathbf{x}_1, \ldots, \mathbf{x}_n$, *as* n *and* m *tend to infinity the conditional distribution of* $\sqrt{m}\left(\bar{\mathbf{x}}_m^* - \bar{\mathbf{x}}_n\right)$ *converges weakly to* $N(0, \Sigma)$.

To state the second theorem we need some notation. Let $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $(\mathbf{x}_1^*, \ldots, \mathbf{x}_m^*)$ be as in the previous theorem. Let

$$S_n = \sum_{i=1}^{n} h(\mathbf{x}_i) \qquad \text{and} \qquad \mu = E_F\left[h(\mathbf{x}_1)\right],$$

where $h : \mathbb{R}^p \to \mathbb{R}^k$. Let

$$\tilde{S}_n = \sum_{i=1}^{n} h(\mathbf{x}_i^*),$$

and let $g(\mathbf{v}) : \mathcal{R}^k \to \mathcal{R}$ be a real valued function with finite differential

$$\dot{g}(\mathbf{v}) = \left(\partial g / \partial v_1(\mathbf{v}), \ldots, \partial g / \partial v_k(\mathbf{v})\right)$$

at μ. The following theorem states that under regularity conditions bootstrapping commutes with smooth functions.

Theorem 3 (Bickel and Freedman) *Let* S_n, $\tilde{S}_n$, μ *and* g *be as in the previous paragraphs. If* $E\left[\|h(\mathbf{x}_1)\|^2\right] < \infty$,

$$\sqrt{n}\left\{g\left(\frac{\tilde{S}_n}{n}\right) - g\left(\frac{S_n}{n}\right)\right\} = \dot{g}(\mu)' \sqrt{n}\left(\frac{\tilde{S}_n}{n} - \frac{S_n}{n}\right) + o_p(1)$$

Many papers in the literature deal with the problem of determining the accuracy and order of coverage level of the bootstrap confidence intervals (see for example, Hall 1986a, 1986b, 1988a and 1990). At this stage of our work we are only interested in the asymptotic validity of our methods. Further refinements will be considered in future work (see item 6 in section (5.4)).

There are results in the literature showing that the bootstrap also works to approximate the asymptotic distribution of robust estimates. See, for example, Shorack (1982), Parr (1985), Yang (1985), Lohse (1987), Shao (1990), Cheng (1991), Arcones *et. al.* (1992). See also Dumbgen (1993) and Cuevas *et. al.* (1993).

Bootstrap can be used to calculate confidence intervals via the estimation of the asymptotic variance or by getting approximate percentiles for the limiting distribution (see Hall 1988a for a comprehensive discussion). Two serious problems arise in either case: first, since bootstrap samples are taken with replacement, the proportion of outliers in the bootstrap sample may be higher than in the original one; second, the computational complexity of the robust estimates imposes an upper bound on the number of recalculations that are feasible.

We will call the first problem "lack of robustness" of the classical bootstrap. In agreement with Shao (1990), we found that the bootstrap distribution has heavy tails that produce inflated variance estimators and unduly long confidence intervals.

Between 2,000 and 3,000 bootstrap samples are needed to estimate the percentiles for a confidence interval (Efron and Tibshirani, 1993). That many recalculations of a robust regression estimate are in practice unfeasible with today's technology.

5.3 Robust Bootstrap

Our robust bootstrap deals with the two problems mentioned above. On the one hand, each evaluation of our bootstrapped estimator is just a linear combination of the re-sampled points so that the computing requirement is minimal. On the other hand, we sample the data paired with weights so that outlying observations are down-weighted. The resulting estimate will not breakdown as long as the bootstrap sample contains at least one "good" point.

For simplicity, I will first present our robust bootstrap in the location model with known scale. The more general location-scale setup is presented immediately after.

5.3.1 Robust Bootstrap for Location

Assume $x_1, \ldots, x_n$ are n observations following the location model (2) and let T_n be defined by (3). Assume that the scale is known, an hence without loss of generality let $\sigma = 1$. Assume that

$$\sqrt{n}\left(T_n - T_\infty\right) \longrightarrow N\left(0,\ \mathrm{AV}\left(F, T\right)\right), \tag{20}$$

with

$$\mathrm{AV}\left(F, T\right) = \frac{E_F \psi^2\left(X - T_\infty\right)}{\left(E_F \psi'\left(X - T_\infty\right)\right)^2}. \tag{21}$$

Let

$$\omega_j\left(T_n\right) = \left(\frac{\psi\left(x_j - T_n\right)}{x_j - T_n}\right) \Big/ \sum_{i=1}^{n} \frac{\psi\left(x_i - T_n\right)}{x_i - T_n} \quad j = 1, \ldots, n\ ,$$

similarly

$$\omega_j\left(T_\infty\right) = \left(\frac{\psi\left(x_j - T_\infty\right)}{x_j - T_\infty}\right) \Big/ \sum_{i=1}^{n} \frac{\psi\left(x_i - T_\infty\right)}{x_i - T_\infty} \quad j = 1, \ldots, n\ ,$$

and

$$\hat{T}_n = \sum_{j=1}^{n} \omega_j\left(T_\infty\right) x_j\ . \tag{22}$$

Let $x_1^*, \ldots, x_n^*$ be a bootstrap sample from $x_1, \ldots, x_n$. Our bootstrapped $\hat{T}_n$ is explicitly given by

$$\hat{T}_n^* = \sum_{j=1}^{n} \omega_j^* \left(T_n\right) x_j^*, \tag{23}$$

where

$$\omega_j^* \left(T_n\right) = \left(\frac{\psi\left(x_j^* - T_n\right)}{x_j^* - T_n} \right) \Big/ \sum_{i=1}^{n} \frac{\psi\left(x_i^* - T_n\right)}{x_i^* - T_n}.$$

In what follows I will motivate our method. We have the identity

$$T_n = \sum_{j=1}^{n} \omega_j \left(T_n\right) x_j. \tag{24}$$

A Taylor expansion around T_∞ gives

$$T_n = \sum_{j=1}^{n} \omega_j \left(T_\infty\right) x_j + \left(T_n - T_\infty\right) \sum_{j=1}^{n} \omega_j' \left(\tilde{T}_n\right) x_j, \tag{25}$$

where $\tilde{T}_n$ is a point in the interval between T_n and T_∞, and $\omega'(x) = \partial\omega/\partial x$.

From equation (25) we get

$$\left(1 - \sum_{j=1}^{n} \omega_j' \left(\tilde{T}_n\right) x_j\right) \sqrt{n} \left(T_n - T_\infty\right) = \sqrt{n} \left(\sum_{j=1}^{n} \omega_j \left(T_\infty\right) x_j - T_\infty \right). \tag{26}$$

Recall the definition of $\hat{T}_n$ in (22) and rewrite equation (26) as

$$\sqrt{n} \left(T_n - T_\infty\right) = a_n \left(\tilde{T}_n\right) \sqrt{n} \left(\hat{T}_n - T_\infty\right), \tag{27}$$

where

$$a_n \left(\tilde{T}_n\right)^{-1} = 1 - \sum_{j=1}^{n} \omega_j' \left(\tilde{T}_n\right) x_j.$$

Our approach is motivated by equation (24). We can, in principle, bootstrap the right hand side of this representation keeping the weights $\omega_j \left(T_n\right)$ fixed. Because we are fixing the weights we expect the variability of our bootstrapped statistic to be smaller than the variance of the full boostrapped T_n. Equation (27) suggests the correction we need to compensate this loss on variability.

To justify our bootstrap we can prove that the sequence $a_n \left(\tilde{T}_n\right)^{-1}$ converges almost surely to a non-zero constant. Unfortunately the regularity conditions we need to carry on the proof are very restrictive. We would need the following functions to be uniformly continuous:

$$\psi'(t), \quad \psi(t), \quad \text{and} \quad \psi(t)/t^2 - \psi'(t)/t.$$

The last condition is particulary stringent. A function ψ such that $\psi(t) \propto t^\alpha$ for t in a neighborhood of zero, with $1 < \alpha < 2$ does not satisfy it.

Fortunately, we have the following theorem that justifies our bootstrap under milder conditions.

Theorem 4 *Let the sequence T_n be defined by (3). Assume that (20) and (21) hold, and that $0 < |E_F[\psi(X - T_\infty)/(X - T_\infty)]| < +\infty$. We have the following:*

1. *Let $\hat{T}_n$ be defined by (22) then*

$$\sqrt{n}\left(\hat{T}_n - T_\infty\right) \longrightarrow N\left(0,\ AV\left(F, \hat{T}\right)\right),$$

with

$$AV\left(F, \hat{T}\right) = \frac{E_F \psi^2 (X - T_\infty)}{\left(E_F \frac{\psi(X - T_\infty)}{X - T_\infty}\right)^2}.$$

2. *There exists a non-zero and finite constant $a \in \mathbb{R}$ such that*

$$a^2\ AV\left(F, \hat{T}\right) = AV(F, T), \tag{28}$$

where $AV(F,T)$ is defined in (21).

3. *If $\psi'(t)$ and $\psi(t)/t$ are uniformly continuous on $\mathbb{R}$, then there exists a sequence $\{a_n\}_{n\geq 1} = \{a_n(T_n)\}_{n\geq 1} \in \mathbb{R}$ such that*

$$a_n \longrightarrow a \quad \textit{almost surely}, \tag{29}$$

where a satisfies (28).

Proof of theorem (4):

1. Simple calculations yield

$$\sqrt{n}\left(\hat{T}_n - T_\infty\right) = \sqrt{n}\left(\frac{\sum_{i=1}^n \psi\left(x_i - T_\infty\right)}{\sum_{j=1}^n \frac{\psi(x_j - T_\infty)}{x_j - T_\infty}}\right)$$

   which converges in distribution to a normal random variable with mean zero and variance

$$\frac{E_F \psi^2\left(X - T_\infty\right)}{\left(E_F\left(\frac{\psi(X-T_\infty)}{X-T_\infty}\right)\right)^2}. \tag{30}$$

2. From equation (7) we find that the asymptotic variance of $\sqrt{n}\left(T_n - T_\infty\right)$ is

$$\frac{E_F \psi^2\left(X - T_\infty\right)}{\left(E_F \psi'\left(X - T_\infty\right)\right)^2}.$$

   Comparing with (30) we find the needed constant

$$a = \frac{E_F\left(\frac{\psi(X-T_\infty)}{X-T_\infty}\right)}{E_F \psi'\left(X - T_\infty\right)}. \tag{31}$$

3. Let the sequence $a_n\left(T_n\right)$ be

$$a_n\left(T_n\right) = \frac{\sum_{i=1}^n \frac{\psi(x_i - T_n)}{x_i - T_n}}{\sum_{i=1}^n \psi'\left(x_i - T_n\right)},$$

   Lemma A in page 253 of Serfling (1980) asserts the almost sure convergence of a_n to a.

There are several questions to be addressed. For example, it is not immediately clear that we can actually bootstrap $\sqrt{n}\left(\hat{T}_n - T_\infty\right)$ (see section 5.2). We should also show that replacing T_∞ by T_n in the statistic being bootstrapped does not upset our results. These questions will be discussed further after the following extension of our method to the location-scale model.

5.3.2 Robust Bootstrap for Location and Scale

Assume the location model (2) with unknown scale, and let T_n and S_n be defined by (4) and (5). Under mild assumptions (see Maronna and Yohai, 1981) we have

$$\sqrt{n}\left(\left(\begin{array}{c} T_n \\ S_n \end{array}\right) - \left(\begin{array}{c} T_\infty \\ S_\infty \end{array}\right)\right) \longrightarrow N_2\left(\mathbf{0}, \Sigma\right), \tag{32}$$

where N_2 denotes a normal distribution in $\mathbb{R}^2$, and the asymptotic variance-covariance matrix is

$$\Sigma \;\; = \;\; D^{-1}\, C\, \left(D^{-1}\right)', \tag{33}$$

where

$$C = \left(\begin{array}{cc} E_F \psi^2\left(Y\right) & E_F\left[\psi\left(Y\right)\chi\left(Y\right)\right] \\ E_F\left[\psi\left(Y\right)\chi\left(Y\right)\right] & E_F \chi^2\left(Y\right) - b^2 \end{array}\right), \tag{34}$$

$$D = -\frac{1}{S_\infty}\left(\begin{array}{cc} E_F \psi'\left(Y\right) & E_F\left[\psi'\left(Y\right)Y\right] \\ E_F \chi'\left(Y\right) & E_F\left[\chi'\left(Y\right)Y\right] \end{array}\right), \tag{35}$$

and $Y = (X - T_\infty)/S_\infty$.

Let the weights $\omega_i\left(T_n, S_n\right)$ and $v_i\left(T_n, S_n\right)$ be defined as

$$\omega_i\left(T_n, S_n\right) \;\; = \;\; \frac{\psi\left(\frac{x_i - T_n}{S_n}\right)}{x_i - T_n} \Bigg/ \sum_{j=1}^{n} \frac{\psi\left(\frac{x_j - T_n}{S_n}\right)}{x_j - T_n} \qquad 1 \le i \le n,$$

$$v_i\left(T_n, S_n\right) \;\; = \;\; \chi\left(\frac{x_i - T_n}{S_n}\right) \Big/ \left(\frac{x_i - T_n}{S_n}\right) \qquad 1 \le i \le n.$$

Similarly define $\omega_i\left(T_\infty, S_\infty\right)$ and $v_i\left(T_\infty, S_\infty\right)$ for $1 \le i \le n$. We have

$$\left(\begin{array}{c} T_n \\ S_n \end{array}\right) = \left(\begin{array}{c} \sum_{i=1}^n \omega_i\left(T_n, S_n\right) x_i \\ \frac{1}{bn}\sum_{i=1}^n v_i\left(T_n, S_n\right)\left(x_i - T_n\right) \end{array}\right).$$

To simplify the notation call

$$\theta_n \equiv \left(\begin{array}{c} \theta_{n1} \\ \theta_{n2} \end{array}\right) = \left(\begin{array}{c} T_n \\ S_n \end{array}\right), \qquad \theta_\infty = \left(\begin{array}{c} T_\infty \\ S_\infty \end{array}\right),$$

and

$$\hat{\theta}_n \equiv \begin{pmatrix} \hat{\theta}_{n1} \\ \hat{\theta}_{n2} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \omega_i \left(T_\infty, S_\infty\right) x_i \\ \\ \frac{1}{bn} \sum_{i=1}^{n} v_i \left(T_\infty, S_\infty\right) \left(x_i - T_\infty\right) \end{pmatrix}. \tag{36}$$

The motivation for theorem 5 below is as follows. We have the identity

$$\theta_n = \begin{pmatrix} \theta_{n1} \\ \theta_{n2} \end{pmatrix} = \begin{pmatrix} T_n \\ S_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \omega_i \left(T_n, S_n\right) x_i \\ \\ \frac{1}{bn} \sum_{i=1}^{n} v_i \left(T_n, S_n\right) \left(x_i - T_n\right) \end{pmatrix}. \tag{37}$$

Rearranging the terms after a Taylor expansion around $\theta_\infty = (T_\infty, S_\infty)'$ we get

$$\sqrt{n} \left(\theta_n - \theta_\infty\right) = \left[I - A_n\left(\tilde{\theta}_n\right)\right]^{-1} \sqrt{n}\left(\hat{\theta}_n - \theta_\infty\right) \tag{38}$$

where

$$A_n\left(\tilde{\theta}_n\right) = \begin{pmatrix} \partial\theta_{n1}/\partial T_n\left(\tilde{\theta}_n\right) & \partial\theta_{n1}/\partial S_n\left(\tilde{\theta}_n\right) \\ \\ \partial\theta_{n2}/\partial T_n\left(\tilde{\theta}_n\right) & \partial\theta_{n2}/\partial S_n\left(\tilde{\theta}_n\right) \end{pmatrix},$$

and $\tilde{\theta}_n$ is a point between θ_n and θ_∞. Equation (37) motivates our bootstrap, and (38) suggests the correction needed to account for the use of fixed weights.

As before, we can complete our argument by proving that the sequence of matrices $\left[I - A_n\left(\tilde{\theta}_n\right)\right]$ converges almost surely to a non-singular matrix. The necessary regularity conditions for this proof are very restrictive.

The next theorem is the location-scale extension of theorem 4.

Theorem 5 *Assume that (32) and (33) to (35) hold, and that* $0 < \left|E_F\left(\psi\left(Y\right)/Y\right)\right| < +\infty$*, with* $Y = \left(X - T_\infty\right)/S_\infty$*.*

1. *Let* $\hat{\theta}_n$ *be defined by (36). Then*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_\infty\right) \longrightarrow N_2\left(\mathbf{0}, \Sigma_2\right),$$

where

$$\Sigma_2 = \begin{pmatrix} \frac{E_F\psi^2(Y)}{\left(E_F[\psi(Y)/Y]\right)^2} & \frac{E_F[\psi(Y)\chi(Y)]}{b\ E_F[\psi(Y)/Y])} \\ \\ \frac{E_F[\psi(Y)\chi(Y)]}{b\ E_F[\psi(Y)/Y]} & \frac{E_F\chi^2(Y)-b^2}{b^2} \end{pmatrix}, \tag{39}$$

and $Y = (X - T_\infty)/S_\infty$*.*

2. *There exists a matrix* $A \in I\!R^{2\times 2}$ *such that*

$$\Sigma = A\ \Sigma_2\ A^t \tag{40}$$

where Σ *is given by (33) to (35) and* Σ_2 *is given by (39).*

3. *If the functions* $\psi'(t)$, $\psi'(t)\,t$, $\chi'(t)$, $\chi'(t)\,t$ *and* $\psi(t)/t$ *are uniformly continuous, then there exists a sequence of matrices* $\{A_n\}_{n\geq 1} = \{A_n(\theta_n)\}_{n\geq 1}$ *such that*

$$A_n \longrightarrow A \text{ almost surely,}$$

where A *satisfies (40).*

Proof of theorem 5:

1. The following argument will show that there exists a function $g : I\!R^3 \to I\!R^2$, and n independent and identically distributed random vectors $\mathbf{y}_i \in I\!R^3$, $1 \leq i \leq n$, with mean $\mu = E_F(\mathbf{y}_i)$, such that

$$\sqrt{n}\left(\hat{\theta}_n - \theta_\infty\right) = \sqrt{n}\left(g(\bar{\mathbf{y}}) - g(\mu)\right).$$

This representation will allow us to calculate the asymptotic variance-covariance matrix of $\sqrt{n}\left(\hat{\theta}_n - \theta_\infty\right)$. By Slutzky's Theorem it will be given by

$$\nabla g(\mu)\ \Sigma_{\mathbf{y}}\ \nabla g(\mu)'$$

where $\nabla g(\mu)$ denotes the matrix of first derivatives of g at μ, and $\Sigma_{\mathbf{y}}$ is the matrix of variance-covariance of the random vectors $\mathbf{y}_i$.

Note that

$$\begin{aligned}
\hat{\theta}_{n1} &= \sum_{i=1}^n x_i \frac{\Psi\left(\frac{x_i - T_\infty}{S_\infty}\right)/(x_i - T_\infty)}{\sum_{j=1}^n \Psi\left(\frac{x_j - T_\infty}{S_\infty}\right)/(x_j - T_\infty)} \\
&= \frac{E_{F_n}\left(\Psi\left(\frac{X - T_\infty}{S_\infty}\right)\right)}{E_{F_n}\left(\Psi\left(\frac{X - T_\infty}{S_\infty}\right)/(X - T_\infty)\right)} + T_\infty
\end{aligned}$$

and

$$\begin{aligned}\hat{\theta}_{n2} &= \frac{1}{nb}\sum \chi\left(\frac{x_i - T_\infty}{S_\infty}\right) S_\infty \\ &= S_\infty \;\; E_{F_n}\left(\chi\left(\frac{x - T_\infty}{S_\infty}\right)\right) \Big/ \; b\,.\end{aligned}$$

For each $i = 1, \ldots, n$ define the random vectors $\mathbf{y}_i \in \mathbb{R}^3$ as

$$\mathbf{y}_i = \left(\Psi\left(\frac{x_i - T_\infty}{S_\infty}\right),\;\; \Psi\left(\frac{x_i - T_\infty}{S_\infty}\right) \Big/ \;(x_i - T_\infty)\,,\;\; \chi\left(\frac{x_i - T_\infty}{S_\infty}\right)\right).$$

Their sample mean is

$$\bar{\mathbf{y}} = \left(E_{F_n}\Psi\left(\frac{X - T_\infty}{S_\infty}\right), E_{F_n}\left(\Psi\left(\frac{X - T_\infty}{S_\infty}\right)\Big/ (X - T_\infty)\right), E_{F_n}\chi\left(\frac{X - T_\infty}{S_\infty}\right)\right).$$

Define $g : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$g\,(x,\; y,\; z) \;=\; (x/y + T_\infty, \;\; S_\infty\; z/b)\,,$$

and note that

$$\hat{\theta}_n = g\,(\bar{\mathbf{y}})\,.$$

Now let $\mu \equiv (\mu_1, \mu_2, \mu_3) = E_F\,(\mathbf{y}_1)$. We have $\mu_1 = 0$, $0 < |\mu_2| < \infty$, and $\mu_3 = b$. This implies

$$g\,(\mu) = \begin{pmatrix} T_\infty \\ S_\infty \end{pmatrix} = \theta_\infty.$$

We have then expressed $\hat{\theta}_n - \theta_\infty$ as a smooth function of means:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_\infty\right) = \sqrt{n}\,(g\,(\bar{\mathbf{y}}) - g\,(\mu))\,.$$

It is easy to check that for any $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ with $x_2 \neq 0$ we have

$$\nabla g\,(\mathbf{x}) = \begin{pmatrix} 1/x_2 & -x_1/x_2^2 & 0 \\ 0 & 0 & S_\infty/b \end{pmatrix},$$

so that

$$\nabla g(\mu) = \begin{pmatrix} 1/\mu_2 & 0 & 0 \\ 0 & 0 & S_\infty/b \end{pmatrix}$$
$$= \begin{pmatrix} \left(E_F\left[\psi(Y)\,Y\right]\right)^{-1} & 0 & 0 \\ 0 & 0 & S_\infty/b \end{pmatrix},$$

where $y = (X - T_\infty)/S_\infty$. From here, part 1 of the theorem follows.

2. Consider the matrix A such that

$$A^{-1} = \begin{pmatrix} E_F\psi'(Y)\ /\ E_F\left[\psi(Y)/Y\right] & E_F\left[\psi'(Y)\,Y\right]\ /\ E_F\left[\psi(Y)/Y\right] \\ E_F\chi'(y)/\,b & E_F\left[\chi'(Y)\ \ Y\right]/\,b \end{pmatrix},$$

where $Y = (X - T_\infty)/S_\infty$. Note that A is non-singular if and only if D in equation (35) is non-singular. It is easy to see that A satisfies (40).

3. From Lemma A in page 253 of Serfling (1980) and the assumed regularity conditions we get the result with

$$A_n^{-1} = \begin{pmatrix} E_{F_n}\psi'\left(\tilde{Y}\right)\ /\ E_{F_n}\left[\psi\left(\tilde{Y}\right)/\tilde{Y}\right] & E_{F_n}\left[\psi'\left(\tilde{Y}\right)\tilde{Y}\right]\ /\ E_{F_n}\left[\psi\left(\tilde{Y}\right)/\tilde{Y}\right] \\ E_{F_n}\chi'\left(\tilde{Y}\right)/\,b & E_{F_n}\left[\chi'\left(\tilde{Y}\right)\ \tilde{Y}\right]/\,b \end{pmatrix},$$

where $\tilde{Y} = (X - T_n)/S_n$.

This finishes the proof of the theorem.

The questions that remain to be answered are:

- Can we bootstrap $\sqrt{n}\left(\hat{\theta}_n - \theta_\infty\right)$?
- When we replace $(T_\infty, S_\infty)'$ by $(T_n, S_n)'$, is the bootstrap distribution close to the one we want to estimate? This question can be seen as a matter of "continuity" or qualitative robustness of the bootstrap (see Cuevas, *et. al.*, 1993).

The representation

$$\sqrt{n}\left(\hat{\theta}_n - \theta_\infty\right) = \sqrt{n}\left(g\left(\bar{\mathbf{y}}\right) - g\left(\mu\right)\right),$$

we get from the proof of Part 1 in Theorem 5 together with Theorem 3 positively answer the first question.

We are still working on the second question. Empirical results obtained are very encouraging.

5.4 Challenges

We anticipate finding the following challenges in our work:

a) The theoretical justification of the robust bootstrap is yet to be completed. The empirical results we have obtained are promising. We have only found one reference in the literature referring to the continuity of the bootstrap (Cuevas, *et. al.*, 1993).

b) The extension to linear regression of our approach to location-scale. With some modifications we expect to be able to adapt our methods to this case. Computer coding will of course be more delicate.

c) We have to develop guidelines on the optimal choice of estimators to get the "best" confidence interval, i.e., for a given level α, the one that yields minimum length. We think we can tackle this problem in a local manner by calculating the Influence Function of the quantiles and then finding the estimate that minimizes their Gross Error Sensitivity.

d) We understand that stand-alone computer codes for new techniques are usually difficult to adopt. We will try to develop, for example, a R / Splus function that will allow the statistician to routinely integrate these robust confidence intervals in his or her every day work. We will also need good computer code to perform our simulation studies.

A simple sketch or the steps of our work follows.

1. The location-scale model. This step is almost completed. Empirical evidence for the remaining work is very encouraging.

2. The linear regression model.

3. The choice of the estimator in the location-scale model. This problem has already been studied by Fraiman *et. al.* (1998) (see page 19).

4. The choice of the estimator in the linear regression model. A general approach as in 3 above might not work here. In this case we will proceed with the local linearization to determine an optimal estimator for small contamination proportions.

5. The development of a good computer code. It has to be powerful enough to be able to handle large simulation studies, but it should also fit easily in a friendly interface (for example, a R / Splus function).

6. The study of the speed of convergence of the robust bootstrap percentiles to its asymptotic values. We can investigate other proposals for determining confidence intervals from bootstrap methods. It is often the case that "studentized" or bootstrap-t intervals have better coverage level (see for example the papers by Hall listed in the bibliography). The lack of an estimate for $\hat{\sigma}$ might not be an important problem when using bootstrap-t intervals (Yohai, 1998, in a personal communication). These potential improvements should be considered.

7. Further extensions to Generalized Linear Models, Non Linear Parametric Models, Multivariate Location and Scatter, Principal Components, Errors in Variables, etc.

6 Bibliography

- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances.* Princeton University Press, Princeton, N.J.

- Arcones, M. A. and Giné, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the Limits of Bootstrap*, Wiley series in Probability and Mathematical Statistics - Applied Probability and Statistics Section, John Wiley & Sons, New York.

- Barrendero, J. R. and Zamar, R. (1998). The maxbias curve of robust regression estimates. Work in progress.

- Bickel, P. J. and Freedman, D. A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196-1217.

- Carroll, R. J. (1979). On estimating variances of Robust Estimators when the errors are asymmetric. *J. Amer. Statist. Assoc.* **74**, 674-679.

- Cheng, K. F. (1991). M-Estimator using Jackknife Pseudovalues. *Scand. J. Statist.*, **18**, 51-61.

- Croux, C., Rousseeuw, P. J., Van Bael, A. (1996). Positive-breakdown regression by minimizing nested scale estimators. *J. Statist. Plann. Inference*, **53**, 197–235.

- Cuevas, A. and Romo, J. (1993) On robustness properties of bootstrap approximations. *J. Statist. Plann. Inference*, **37**, 181-191.
- Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown-point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Docksum and J. L Hodges, Jr., eds.) 157-184. Wadsworth, Belmont, Calif.
- Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, **95**, 125-140.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM. Philadelphia.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Monographs on Statistics and Applied Probability, 57. Chapman & Hall, New York.
- Fernholz, L. T. (1983). Von Mises Calculus for Statistical Functionals. Springer Lecture Notes in Statistics, **19**.
- Fraiman, R., Yohai, V. J., and Zamar, R. (1998). Optimal robust M-estimates of location. Work in progress.
- Gosh, M., Parr, W. C., Singh, K. and Babu, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.*, **12**, 1130-1135.

- Gross, A. M. (1976). Confidence Interval Robustness with Long-Tailed Symmetric Distributions. *Journal of the American Statistical Association*, **71**, 409-416.

- Hall, P. (1986a). On the bootstrap and confidence intervals. *Ann. Statist.*, **14**, 1431-1452

  – (1986b). On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.*, **14**, 1453-1462.

  – (1988a). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.*, **16**, 927-953.

  – (1988b). Reply to comments on "Theoretical comparison of bootstrap confidence intervals". *Ann. Statist.*, **16**, 981-985.

  – (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions. *Ann. Probability*, **18**, 1342-1360.

  – (1992). *The bootstrap and Edgeworth expansion.* Springer-Verlag. Berlin-New York.

- Hampel, F. R., Ronchetti, E. Z., Rousseeuw, P. J. and Stahel, W. A. *Robust Statistics. The approach based on influence functions.* 1986. Wiley Series in Probability and Mathematical Statistics, New York.

- He, X. (1991). A local breakdown property of robust tests in linear regression. *J. Mult. Anal.*, **38**, 294-305.

- He, X., Simpson, D. G. and Portnoy, S. L. (1990). Breakdown robustness of tests. *Journal of the American Statistical Association*, **85**, 446-452.

- Hill, R. W. (1977). Robust regression when there are outliers in the carriers. Ph. D. thesis. Harvard University, Cambridge, Masss.

- Hössjer, O. (1992). On the optimality of S-estimators. *Statist. Prob. Lett.*, **14**, 413-419.

- Hu, F. and Kalbfleisch, J. D. (1997). Estimating equations and the bootstrap. Manuscript.

- Hu, F. and Zidek, J. V. (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika*, **82**, 263-275.

- Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73-101.

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

- Krasker, W. S. and Welsch, R. E. (1982). Efficient Bounded Influence Regression Estimation. *Journal of the American Statistical Association*, **77**, 595-604.

- Lohse, K (1987). Consistency of the bootstrap. *Stat. Decis*, **5**, 353-366.

- Markatou, M. and He, X. (1994). Bounded influence and high breakdown point testing procedures in linear models. *Journal of the American Statistical Association*, **89**, 543-549.

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

- Markatou, M., Stahel, W. A., and Ronchetti, E. (1991). Robust M-type testing procedures for linear models. In *Directions in robust statistics and diagnostics*, Part I, Stahel, W., Weisberg, S., eds. Springer-Verlag. pp. 201-220.

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

- Maronna, R. A., Bustos, O. H. and Yohai, V. J. (1979). Bias- and efficiency-robustness of general M-estimators for regression with random carriers. In *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblatt (eds.). Lecture Notes in Mathematics 757. Springer, Berlin, pp. 91-116.

- Maronna, R. A. and Yohai, V. J. (1981). Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Z. Wahrsch. verw. Geb.*, **58**, 7-20.

- Martin, R. D. and Zamar, R. H. (1993). Efficiency-constrained bias-robust estimation of location. *Ann. Statist.*, **21**, 338-354.

- Martin, R. D., Yohai, V. J. and Zamar, R. H. (1989). Min-max bias robust regression. *Ann. Statist.*, **17**, 1608-1630.

- Merrill, H. M. and Schweppe, F. C. (1971). Bad data suppression in power system static state estimation. *IEEE Trans. Power App. Syst.*, **PAS-90**, 2718-2725.

- Parr, W. C. (1985) The bootstrap: some large sample theory and connections with robustness. *Statist. Prob. Lett.*, **3**, 97-100.

- Rocke, D. M. and Downs, G. W. (1981). Estimating the variances of robust estimators of location: influence curve, jackknife and bootstrap. *Comm. Stat. B.*, **10**, 221-248.

- Ronchetti, E. (1979). Robustheitseigenschaften von Tests. Diploma thesis. ETH, Zurich.

- Ronchetti, E. (1982). Robust Testing in Linear Models: The Infinitesimal Approach. Unpublished Ph. D. thesis, Swiss Federal Institute of Technology, Zurich.

- 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. Springer, New York.

- Sen, P. K. (1982). On M tests in linear models. *Biometrika*, **52**, 591-611.

- Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics.* Wiley Series in Probability and Mathematical Statistics.

- Shao, Jun (1990) Bootstrap estimation of the Asymptotic variances of statistical functionals. *Ann. Inst. Statist. Math.*, **42**, 737-752.

- Shao, Jun (1992) Bootstrap variance estimators with truncation. *Statist. Prob. Lett.*, **15**, 95-101.

- Shorack, G. R. (1982). Bootstrapping robust regression. *Comm. Stat. A.*, **11**, 961-972.

- Weisberg, S. (1985). *Applied Linear Regression.* John Wiley & Sons, New York.

- Wu, C. F. J. (1986) Jackknife, bootstrap and other re-sampling methods in regression analysis. *Ann. Statist.*, **14**, 1261-1295.

- Yang, S-S (1985) On bootstrapping a class of differentiable statistical functionals with applications to L- and M-estimates. *Statistica Neerlandica*, **39**, 375-385.

- Yohai, V. J. (1985) High Breakdown-Point and High Efficiency Robust Estimates for Regression. Technical Report No. 66, Dept. Statistics, Univ Washington, Seattle.

- Yohai, V. J. (1987) High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Ann. Statist.*, **15** 642-656.

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

- Yohai, V. J. and Zamar, R. H. (1988). High breakdown point estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.*, **83**, 406-413.