Hlávka, Zdeněk

Working Paper

Asymptotic properties of robust three-stage procedure based on bootstrap for m-estimator

SFB 373 Discussion Paper, No. 2000,94

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Hlávka, Zdeněk (2000) : Asymptotic properties of robust three-stage procedure based on bootstrap for m-estimator, SFB 373 Discussion Paper, No. 2000,94, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10048202

This Version is available at:
https://hdl.handle.net/10419/62178

ASYMPTOTIC PROPERTIES OF ROBUST THREE-STAGE PROCEDURE BASED ON BOOTSTRAP FOR M-ESTIMATOR

Zdeněk Hlávka

Institute for Statistics and Econometrics, Faculty of Economics
Humboldt University, Spandauerstr. 1, 10178 Berlin, Germany
email: hlavka@wiwi.hu-berlin.de

Key words and Phrases: fixed-width confidence intervals; sequential estimation; bootstrap for M-estimators

ABSTRACT

The paper concerns the fixed-width confidence intervals for location based on M-estimators in the location model. A robust three-stage procedure is proposed and its asymptotic properties are studied. The performance of the procedure depends on some tuning parameters. Their effect on the proposed confidence interval is checked together with the overall behaviour of the procedure in a simulation study.

I INTRODUCTION

The problem of constructing fixed-width confidence intervals is often studied in the literature. There were developed general principles how to construct such confidence intervals. The most well-known is the fully sequential Chow-Robbins procedure which possesses certain optimality properties. Useful for practical purposes is the two-stage Stein procedure and its various modifications and generalizations. While Chow-Robbins procedure requires to check the stopping rule after each observation, the two-stage and, generally, multi-stage procedures require to check the stopping rule only once or few times. This feature of the multi-stage procedures is attractive especially in connection with computationally intensive estimators. For more information see, e.g., Ghosh, Mukhopadhyay, and Sen (1997).

Asymptotic properties of three-stage procedures were investigated e.g. in Hall (1981). An exhaustive overview of the asymptotic theory of multi-stage procedures is given in Ghosh, Mukhopadhyay, and Sen (1997). The generalization of three-stage procedures using bootstrap approximation of critical points was proposed in Aerts and Gijbels (1993).

The above mentioned procedures were originally related to the sample mean. Jurečková (1978) proposed and studied a robust version of the Chow-Robbins procedure, particularly based on M-estimators. Sen (1981) investigated among others sequential methods based on ranks. Some theory concerning sequential procedures for GM-statistics was established in Aerts (1988).

We assume through the paper that $X_1, \ldots, X_n$ are independent identically distributed (iid) random variables with a common distribution function $F(x-\theta)$. Recall that the M-estimator $M_n = M_n(\psi)$ of the parameter θ generated by the score function ψ is defined as a solution of the equation $\sum_{i=1}^n \psi(X_i - t) = 0$. If the score function ψ is monotone and if the solution is not unique we define $M_n = (M_n^+ + M_n^-)/2$, where $M_n^- = \sup\{t : \sum_{i=1}^n \psi(X_i - t) > 0\}$ and $M_n^+ = \inf\{t : \sum_{i=1}^n \psi(X_i - t) < 0\}$.

The fixed-width confidence intervals are typically based on a certain point estimator and its asymptotic distribution.

The properties of the M-estimators in the fixed sample setup were studied rather extensively. For more information we refer to Huber (1981), Jurečková and Sen (1995), Hampel, Ronchetti, Rousseeuw, and Stahel (1990), and Rieder (1994). An introduction to the theory of M-estimators is given e.g. in Serfling (1980).

Using this asymptotic normality, it can be easily shown that the (asymptotically) optimal sample size for the $1-\alpha$ confidence interval of width $2d$ based on the M-estimator is $c_M(d) = \left(u_{1-\alpha/2}\sigma(\psi, F)/d\right)^2$, where $\sigma^2(\psi, F)$ is the variance of the asymptotic Normal distribution of the M-estimator and $u_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of standard Normal distribution.

The objective of this paper is to develop robust three-stage fixed-width confidence intervals and to investigate their basic asymptotic properties. The assumptions and some properties of M-estimators which will be used through the paper are given in Section II. The robust three-stage procedure is introduced and investigated in Section III. The main result is Theorem 2 which says that the procedure is consistent and (first order) asymptotically efficient. Simulations in Section IV allow to assess the small sample behaviour of the robust procedure and the influence of some tuning parameters on the robust fixed-width confidence intervals.

II ASSUMPTIONS AND OVERVIEW OF PROPERTIES OF M-ESTIMATORS

Let us now formulate the regularity conditions on the score function $\psi(.)$ and on the distribution function $F(.) = F_\theta(.) = F_0(. - \theta)$ of the observations X_i. The assumptions on distribution function F in Subsection II.1 are more complicated than the assumptions which are normally used for fixed-width confidence intervals based on sample mean but usual for sequential methods based on M-estimators (see, e.g., Jurečková and Sen (1995)). The assumptions on the score function ψ cover also the score functions with jump discontinuities. The differentiability conditions are given in Subsection II.3. The conditions of skew-symmetry of the score function ψ and of the symmetry of the density f could be relaxed—we keep these conditions for the sake of simplicity of the resulting formulas.

II.1 Assumptions on F

$F_0(.) = F(. + \theta)$ has an absolutely continuous density $f_0(.)$ such that $f_0(x) = f_0(-x)$, $\forall x \in \Re$ and $f_0(x)$ is decreasing in x for $x \geq 0$. Moreover, $F(.)$ has the finite Fisher information, i.e., $0 < I(F) = \int_{-\infty}^{\infty} \{f'(x)/f(x)\}^2 dF(x) < \infty$ and there exists $l > 0$ such that $E|X_1|^l = \int_{-\infty}^{\infty} |x|^l dF(x) < \infty$.

Let $c_l(x) = |x|^l F(x)(1 - F(x))$, $x \in \Re$, and $c_l^* = \sup_{x \in \Re} c_l(x)$. Then $c_l^* < \infty$, $\lim_{x \pm \infty} c_l(x) = 0$, and $\int_{-\infty}^{\infty} \{F(x)(1 - F(x))\}^b dx < \infty \qquad \forall b > \frac{1}{l} > 0$.

II.2 Assumptions on ψ

The score function $\psi(.)$ is nondecreasing and skew-symmetric, and there exists a positive number h such that $\psi(x) = \psi(h)\,\mathrm{sign}(x)$ for $|x| \geq h$.

$\psi(.)$ can be decomposed into $\psi(x) = \psi_1(x) + \psi_2(x)$ for all $x \in \Re$, where $\psi_1(x)$ and $\psi_2(x)$ are respectively the absolutely continuous and the step component.

II.3 Assumptions on $\lambda_F(t)$

The function $\lambda_F(t) = \lambda_{F_\theta}(t) = \int_{-\infty}^{\infty} \psi(x - t) dF(x)$ is differentiable at $t = \theta$, $\lambda_{F_\theta}(\theta) = 0$, $\gamma(\psi, F_\theta) \equiv \lambda'_{F_\theta}(\theta) \neq 0$, and its derivative $\lambda'_{F_\theta}(t)$ is a continuous function of t in some neighbourhood of θ.

II.4 Overview of Basic Properties of M-estimators

An important property of the M-estimator is the following asymptotic linearity result which is generalization of the result of Jurečková and Sen (1981).

LEMMA **1** *Under the assumptions of Section II, for every fixed* $0 < K < \infty$, $\varepsilon > 0$, $\delta > 1/4$, *and* $\eta > 0$, *there exists* n_0 *such that*

$$P\left\{ \sup_{|t| < K \log n} |n^{1-\delta} \left[\lambda_{F_n}(\theta + tn^{-1/2}) - \lambda_{F_n}(\theta)\right] \right.$$

$$\left. - n^{1-\delta} \left[\lambda_F(\theta + tn^{-1/2}) - \lambda_F(\theta)\right] | \geq \varepsilon \right\} \leq cn^{-1-\eta}, \ \forall n > n_0.$$

PROOF: The proof is given in Hlávka (2000). □

Lemma 1 allows to investigate M-estimators using well-known results for sums of iid random variables. One of the corollaries is e.g. the strong consistency of the M-estimator under our assumptions.

In order to define the sequential procedure, we need an estimator of asymptotic variance of the M-estimator. The situation is complicated by the presence of the

jump component of the score function $\psi(.)$. A general estimate of the asymptotic variance could be defined e.g. as

$$\hat{\sigma}_n^2 = \frac{(1/n)\sum_{i=1}^n \psi^2(X_i - M_n)}{\left(\sqrt{n}[\lambda_{F_n}(M_n - tn^{-1/2}) - \lambda_{F_n}(M_n + tn^{-1/2})]/2t\right)^2}. \tag{1}$$

Under our assumptions, using Lemma 1, it can be shown that the estimate (1) is consistent, see also Hlávka (2000), Jurečková and Sen (1981), Aerts (1988).

Lemma 1 proves to be very useful also in the proof of the following Theorem 1 which states that the bootstrap for M-estimators works in our situation. Notice that bootstrap for M-estimators has been already investigated, under different assumptions, e.g. in Karabulut and Lahiri (1997), Lahiri (1992), or Arcones and Giné (1992).

THEOREM **1** *Let the assumptions of Section II be satisfied. Let M_n^* be the M-estimator based on the bootstrap sample from the empirical distribution $F_n(x)$ and let M_n be the M-estimator based on the random sample from the distribution $F(x)$. Then*

$$\lim_{n\to\infty} \sup_{x\in\Re} |P^*\left\{\sqrt{n}(M_n^* - M_n) < x\right\} - \Phi[x/\hat{\sigma}_n]| = 0 \qquad [P]\ a.s., \tag{2}$$

$$\lim_{n\to\infty} \sup_{x\in\Re} |P^*\left\{\sqrt{n}(M_n^* - M_n) < x\right\} - \Phi[x\lambda'_{F_\theta}(\theta)/S_n]| = 0 \qquad [P]\ a.s., \tag{3}$$

and

$$\lim_{n\to\infty} \sup_{x\in\Re} |P^*\left\{\sqrt{n}(M_n^* - M_n) < x\right\} - P\left\{\sqrt{n}(M_n - \theta) < x\right\}| = 0 \qquad [P]\ a.s., \tag{4}$$

where $\Phi(x)$ denotes the distribution function of the standard normal distribution and where $\hat{\sigma}_n^2$ is defined by (1).

PROOF: The proof is given in Hlávka (2000). □

III ROBUST THREE-STAGE PROCEDURE BASED ON BOOTSTRAP

Using bootstrap approximations, we can estimate the quantiles of the unknown distribution of the M-estimator using the (known) distribution of M_n^*, the M-estimator based on observations $X_1^*, \ldots, X_n^* \sim F_n(x)$, where F_n denotes the empirical distribution function of $X_1, \ldots, X_n$.

The standardized bootstrap critical point $\xi_m^A(\alpha)$ is defined as $1-\alpha$ quantile of the (centered and standardized) conditional distribution of $\sqrt{m}|M_m^* - M_m|/\hat{\sigma}_m$, where $\hat{\sigma}_m^2$ is the estimate of the asymptotic variance of the M-estimator.

The studentized bootstrap critical point $\xi_m^U(\alpha)$ is defined as $1-\alpha$ quantile of the (centered and studentized) conditional distribution of $\sqrt{m}|M_m^* - M_m|/\hat{\sigma}_m^*$, where $\hat{\sigma}_m^*$ is the estimate of the asymptotic variance of M_m^* based on the bootstrap sample $X_1^*, \ldots, X_m^*$.

These critical points can be used to introduce the following three-stage procedure.

III.1 The Robust Three-Stage Procedure

In the first stage we fix the parameter $\gamma > 0$ which controls the sample size in the first stage, and we draw

$$m = m(d) = \max\left\{2, \left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^{\circ} + 1\right\} \tag{5}$$

observations. These observations are used to determine $\hat{\sigma}_m^2$ and the standardized bootstrap critical point $\xi_m^A(\alpha)$. The intermediate sample size $N_1(d)$ based on the standardized bootstrap critical points is then given as

$$N_1(d) = \max\left\{m, \left[k\left(\frac{\xi_m^A(\alpha)\hat{\sigma}_m}{d}\right)^2\right]^{\circ} + 1\right\}, \tag{6}$$

where $0 < k < 1$ is the parameter controlling the sample size in this stage. Finally, we draw next $N_2(d) - N_1(d)$ observations, where

$$N_2(d) = \max\left\{N_1(d), \left[\left(\frac{\xi_{N_1}^A(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2\right]^{\circ} + 1\right\} \tag{7}$$

and obtain the $1-\alpha$ confidence interval $(M_{N_2}-d, M_{N_2}+d)$ based on the standardized bootstrap critical points. The construction of fixed-width confidence intervals based on studentized bootstrap critical points is analogous.

III.2 Asymptotic Properties of the Robust Procedure

Basic asymptotic properties of this robust three-stage procedure (5)–(7) based on the standardized bootstrap critical points are stated in the following Theorem 2. Part (*iii*) of the theorem says that the coverage probability of the resulting interval is asymptotically correct, part (*iv*) says that the procedure is first order efficient.

THEOREM **2** *Suppose that the conditions of Section II hold. Then the three-stage procedure based on bootstrapping M-estimators has the following asymptotic properties:*

$$(i) \qquad \lim_{d\to 0+} N_2(d) = \infty \qquad [P]\ a.s., \tag{8}$$

$$(ii) \qquad \lim_{d\to 0+} \frac{N_2(d)}{c_M(d)} = 1 \qquad [P]\ a.s., \tag{9}$$

$$(iii) \qquad \lim_{d\to 0+} P(M_{N_2(d)} - d < \theta < M_{N_2(d)} + d) = 1 - \alpha, \tag{10}$$

$$(iv) \qquad \lim_{d\to 0+} E\left(\frac{N_2(d)}{c_M(d)}\right) = 1, \tag{11}$$

where $c_M(d)$ is the asymptotically optimal number of observations for fixed-width confidence interval based on M-estimator.

For the proof of Theorem 2, we will need the following two lemmas which have also some interest of their own because they provide upper bounds for the difference between the M-estimator M_n and its bootstrap version M_n^*. These lemmas are bootstrap versions of similar results of Jurečková and Sen (1982).

LEMMA **2** *Under the assumptions of Section II, for every* $0 < c_1 < \infty$, $0 < \alpha < 1$, *and* $c_3 < t < \sqrt{m}c_4$, *we have*

$$P[P^*(\sqrt{m}|M_m^* - M_m| > t) > \alpha]$$
$$\leq 2\exp\left\{-\left[-\frac{\sqrt{-\log\alpha}}{\sqrt{2}\psi(h)} - \frac{\sqrt{c_2}}{2}t\right]^2\right\} + 4\exp\{-c_2t^2/4\},$$

where $c_2 = 2[f(h+c_1)]^2 > 0$, *the constant* $\psi(h)$ *is given in Subsection II.2,*

$$c_3 = \frac{\sqrt{-\log\alpha}}{\sqrt{c_2/2}\psi(h)}, \text{ and } c_4 = \min\left\{2c_1, \frac{\sqrt{2}c_1}{\sqrt{c_2}\psi(h)}\right\}.$$

LEMMA **3** *Under assumptions of Section II, there exists* m_0 *such that for all* $m > m_0$ *and* $t > 2hm^{1/2}$

$$P\left[P^*\left(\sqrt{m}|M_m^* - M_m| > t\right) > \alpha\right]$$
$$\leq 3\left[4F(\theta - tm^{-1/2} - h)(1 - F(\theta - tm^{-1/2}/2 - h))\right]^m,$$

where h *is given in Subsection II.2.*

PROOF OF THEOREM 2: Denote by

$$\mathcal{L}_n^A(x) = P_n^*\left\{\sqrt{n}\frac{|M_n^* - M_n|}{S(\psi, F_n)} \leq x\right\} \tag{12}$$

the distribution function of the absolute value of the standardized bootstrap statistics. By Theorem 3.2 in Hlávka (2000) we have that

$$\lim_{n\to\infty}\mathcal{L}_n^A(x) = \begin{cases} 2\Phi(x) - 1, & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \qquad [P] \ a.s. \tag{13}$$

Recall that $\xi_n^A(\alpha)$ denotes the $1-\alpha$ quantile of $\mathcal{L}_n^A$. By the strict monotonicity and continuity of $\Phi(x)$ and by e.g. Lemma 1.5.6 in Serfling (1980) we have for all $t \in (0,1)$

$$\lim_{n\to\infty}\xi_n^A(t) = u_{1-t/2} \qquad [P] \ a.s. \tag{14}$$

Notice that $\lim_{d\to 0+} m(d) = \infty$ and therefore

$$\lim_{d\to 0+}\xi_{m(d)}^A(\alpha) = u_{1-\alpha/2} \qquad [P] \ a.s. \tag{15}$$

This, together with consistence of $\hat{\sigma}_{m(d)}^2$ and the definition of $N_1(d)$ implies that

$$\lim_{d\to 0+} N_1(d) = \infty \qquad [P] \ a.s. \tag{16}$$

which in turn implies that also

$$\lim_{d\to 0+} N_2(d) = \infty \qquad [P]\ a.s. \tag{17}$$

The definition (7) of the stopping time $N_2(d)$ implies the following inequalities.

$$\left[\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right]^2 < N_2(d) \le \left[\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right]^2 + 1$$

$$+N_1(d) I\left[(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 \le k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2\right] \tag{18}$$

Next, by the strong consistency of $\xi^A_n\hat{\sigma}_n$ as an estimator of $u_{1-\alpha/2}\sigma(\psi,F)$ we have for every $\varepsilon > 0$

$$\lim_{m\to\infty} P(|\xi^A_n\hat{\sigma}_n - u_{1-\alpha/2}\sigma| < \varepsilon, \forall n > m) = 1, \tag{19}$$

where we abbreviate $\sigma(\psi,F)$ by σ. Let us fix $\delta > 0$. We can choose $d_0 > 0$ such that

$$P\left\{I\left[(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 \le k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2\right] = 0, \forall d > d_0\right\} \ge 1 - \delta.$$

which implies that there exists d_0 such that for all $d > d_0$

$$I\left[(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 \le k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2\right] = 0 \qquad [P]\ a.s. \tag{20}$$

Combining (18) and (20) leads to part (ii) of the theorem.

Part (iii) of the theorem follows from the Slutzky theorem, the Anscombe theorem and part (ii).

To verify part (iv) of the theorem it is sufficient to prove uniform integrability of the set $\{N_2(d)d^2\}_{d>0}$. It suffices to show that there exists $d_0 > 0$ such that

$$\sum_{l=1}^{\infty} \sup_{0<d<d_0} P\{N_2(d)d^2 > l\} < \infty. \tag{21}$$

We can choose d_0 such that for every $0 < d < d_0$ we have

$$\left(\left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^{\circ} + 1\right) d^2 \le l \tag{22}$$

and

$$2d^2 \le l \tag{23}$$

which implies that

$$P(N_2(d)d^2 > l)$$

$$\le P\left\{\left(\left[\frac{(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2}{d^2}\right]^{\circ} + 1\right) d^2 > l\right\} + P\left\{\left(\left[\frac{k(\xi^A_m(\alpha)\hat{\sigma}_m)^2}{d^2}\right]^{\circ} + 1\right) d^2 > l\right\}$$

$$= P_{1,l}(d) + P_{2,l}(d). \tag{24}$$

Let us first deal with the second probability.

$$P_{2,l}(d) \leq P\left\{\xi_m^A(\alpha)\hat{\sigma}_m > \sqrt{\frac{l-d^2}{k}}\right\} \leq P\left\{\xi_m^A(\alpha)\hat{\sigma}_m > \sqrt{\frac{l}{2k}}\right\}$$

$$= P\left\{\mathcal{L}_m^A\left(\sqrt{\frac{l}{2k\hat{\sigma}_m}}\right) < 1-\alpha\right\} = P\left\{P_m^A\left(|\sqrt{m}(M_m^* - M_m)| \leq \sqrt{\frac{l}{2k}}\right) < 1-\alpha\right\}$$

$$= P\left\{P_m^*\left(|\sqrt{m}(M_m^* - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\} \tag{25}$$

Now we apply Lemma 2 and we obtain, for all $0 < d < d_0$,

$$\sum_{l=1}^{\infty} P_{2,l}(d) = \sum_{l=1}^{\infty} P\left\{P_m^*\left(|\sqrt{m}(M_m^* - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\}$$

$$= \sum_{l=1}^{c_3} P\left\{P_m^*\left(|\sqrt{m}(M_m^* - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\}$$

$$+ \sum_{l=c_3+1}^{\sqrt{m}c_4} P\left\{P_m^*\left(|\sqrt{m}(M_m^* - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\}$$

$$+ \sum_{l=\sqrt{m}c_4}^{\infty} P\left\{P_m^*\left(|\sqrt{m}(M_m^* - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\}, \tag{26}$$

where c_3 and c_4 are given by Lemma 2. The first term on the right hand side of (26) is clearly finite (smaller than c_3), the finiteness of the second term follows from Lemma 2. It remains to investigate the properties of the third term on the right hand side of (26).

We can use Lemma 3 which implies that

$$\sum_{l=\sqrt{m}c_4}^{\infty} P\left\{P_m^*\left(|\sqrt{m}(M_m^* - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\}$$

$$\leq \sum_{l=\sqrt{m}c_4}^{\infty} 3\left[4F\left(\theta - \sqrt{\frac{2l}{km}} - h\right)\left(1 - F\left(\theta - \sqrt{\frac{2l}{km}} - h\right)\right)\right], \tag{27}$$

where h is given in Subsection II.2. Notice that, for $t > 0$,

$$[F(\theta-t)(1-F(\theta-t))]' = f(\theta-t)[2F(\theta-t)-1] \leq 0.$$

Hence, it follows that the term inside the sum in (27) is non-increasing in l. This allows us to show the finiteness of the sum in (27) by showing the finiteness of the following integral for which we have

$$\int_{c_5 m^{1/2}}^{\infty} \left\{4F\left(\theta - t\left(\frac{x}{m}\right)^{1/2} - h\right)\left[1 - F\left(\theta - t\left(\frac{x}{m}\right)^{1/2} - h\right)\right]\right\}^m dx$$

$$= \frac{2m}{t^2} \int\limits_{tc_5^{1/2}m^{1/4}}^{\infty} u \left\{4F\left(\theta-u-h\right)\left[1-F\left(\theta-u-h\right)\right]\right\}^m du$$

$$\leq \frac{2}{t^2}(4c_l^*)^{1/l} m \left\{4F\left(\theta-h\right)\left[1-F\left(\theta-h\right)\right]\right\}^{m-1-b}$$

$$\times \int\limits_{tc_5^{1/2}m^{1/4}}^{\infty} \left\{4F\left(\theta-u-h\right)\left[1-F\left(\theta-u-h\right)\right]\right\}^b du, \tag{28}$$

where c_l^* is given in Subsection II.1 and b is any number satisfying $b > 1/l$. Notice that $F\left(\theta-h\right) < 1/2$ which implies that $4F\left(\theta-h\right)\left[1-F\left(\theta-h\right)\right] < 1$. It follows that $m\left\{4F\left(\theta-h\right)\left[1-F\left(\theta-h\right)\right]\right\}^{m-1-b}$ is uniformly bounded in m and tends to 0 as $m \to \infty$.

Finally, $\int\limits_{tc_5^{1/2}m^{1/4}}^{\infty} \left\{4F\left(\theta-u-h\right)\left[1-F\left(\theta-u-h\right)\right]\right\}^b du < \infty$ by assumptions. Hence, the sum in (27) is finite. It remains to investigate the convergence of $\sum_{l=1}^{\infty} \sup_{0<d<d_0} P_{1,l}(d)$.

Notice that, by definition, $N_1(d) \geq m(d)$ for all $d > 0$ which immediately implies that

$$\sup_{0<d<d_0} P\left(\xi^A_{N_1(d)}(\alpha)\hat{\sigma}_{N_1(d)} > \sqrt{\frac{l}{2}}\right) < \sup_{0<d<d_0} P\left(\xi^A_{m(d)}(\alpha)\hat{\sigma}_m > \sqrt{\frac{l}{2}}\right). \tag{29}$$

Notice that the term on the left hand side is $P_{1,l}(d)$ and the term on the right hand side is exactly $P_{2,l}(d)$, where k is set to 1. The convergence of the term on the right hand side of (29) has already been established. This proves the desired uniform integrability result and concludes the proof of the part (iv) of the theorem.

□

Proof of Lemma 2: Clearly,

$$P^*(\sqrt{m}|M_m^* - M_m| > t) = P^*(M_m^* > M_m + tm^{-1/2}) + P^*(M_m^* < M_m - tm^{-1/2}) \tag{30}$$

We investigate only the first term on the right hand side of (30), the treatment of the other term is similar and it gives the same result. Using the definition of the M-estimator and Theorem 2 of Hoeffding (1963) we have that

$$P^*(M_m^* > M_m + tm^{-1/2})$$

$$\leq P_m^*\left\{\frac{1}{m}\sum_{i=1}^m \psi(X_i^* - M_m^*) \leq \frac{1}{m}\sum_{i=1}^m \psi(X_i^* - M_m - tm^{-1/2})\right\}$$

$$= P_m^*\left\{\frac{1}{m}\sum_{i=1}^m \psi(X_i^* - M_m - tm^{-1/2}) \geq 0\right\}$$

$$= P_m^* \left\{ \frac{1}{m} \sum_{i=1}^{m} \left[\psi(X_i^* - M_m - tm^{-1/2}) - E_m^* \psi((X_i^* - M_m - tm^{-1/2}) \right] \right.$$

$$\left. \geq \left[-\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(X_i - M_m - tm^{-1/2}) \right] / \sqrt{m} \right\}$$

$$\leq \exp \left\{ - \left[\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(X_i - M_m - tm^{-1/2}) \right]^2 \right\},$$

for all $0 \leq t \leq c_1 \sqrt{m}$. This means that

$$P[P^*(M_m^* > M_m + t) > \alpha]$$

$$\leq P \left[\exp \left\{ - \left[\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(X_i - M_m - tm^{-1/2}) \right]^2 \right\} > \alpha \right]$$

$$= P \left[-\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(X_i - M_m - tm^{-1/2}) < \sqrt{-\log \alpha} \right]$$

$$\leq P \left[\left\{ -\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(X_i - \theta - tm^{-1/2}/2) < \sqrt{-\log \alpha} \right\} \bigcap \left\{ \sqrt{m} |M_m - \theta| < t/2 \right\} \right]$$

$$+ P \left[\left\{ -\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(X_i - M_m - tm^{-1/2}) < \sqrt{-\log \alpha} \right\} \bigcap \left\{ \sqrt{m} |M_m - \theta| \geq t/2 \right\} \right]$$

$$\leq P \left[-\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(X_i - \theta - tm^{-1/2}/2) < \sqrt{-\log \alpha} \right] + P \left[\sqrt{m} |M_m - \theta| \geq t/2 \right] \quad (31)$$

Using Lemma 3.1 in Jurečková and Sen (1982), the second probability in (31) is bounded by

$$P \left[\sqrt{m} |M_m - \theta| \geq t/2 \right] \leq 2 \exp \left\{ -c_2 t^2/4 \right\}. \quad (32)$$

for $0 < t < 2c_1 \sqrt{m}$. It remains to investigate the behaviour of the first term in (31). The following inequality was established in the proof of Lemma 3.1 in Jurečková and Sen (1982):

$$-E\psi(X_1 - \theta - \frac{t}{2} m^{-1/2}) \geq \sqrt{2c_2} \psi(h) \frac{t}{2} m^{-1/2}. \quad (33)$$

Using (33) and the Hoeffding inequality, we have that

$$P \left[-\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(X_i - \theta - \frac{t}{2} m^{-1/2}) < \sqrt{-\log \alpha} \right]$$

$$= P \left[\frac{1}{m} \sum_{i=1}^{m} \psi(X_i - \theta - \frac{t}{2} m^{-1/2}) - E\psi(X_i - \theta - \frac{t}{2} m^{-1/2}) \right.$$

$$> -\sqrt{-\frac{\log\alpha}{m}} - E\psi(X_i - \theta - \frac{t}{2}m^{-1/2})\Bigg]$$

$$\leq P\left[\frac{1}{m}\sum_{i=1}^{m}\psi(X_i - \theta - \frac{t}{2}m^{-1/2}) - E\psi(X_i - \theta - \frac{t}{2}m^{-1/2})\right.$$

$$\left. > -\sqrt{-\frac{\log\alpha}{m}} + \sqrt{2c_2}\psi(h)\frac{t}{2}m^{-1/2}\right]$$

$$\leq \exp\left\{\frac{-2\left[-\sqrt{-\log\alpha} + \sqrt{2c_2}\psi(h)t/2\right]^2}{[2\psi(h)]^2}\right\} = \exp\left\{-\left[-\frac{\sqrt{-\log\alpha}}{\sqrt{2}\psi(h)} + \frac{\sqrt{c_2}}{2}t\right]^2\right\} \tag{34}$$

for all t such that

$$0 < t < 2c_1\sqrt{m}$$

and

$$0 < -\sqrt{-\log\alpha} + \sqrt{2c_2}\psi(h)t/2 < c_1\sqrt{m},$$

where h (and $\psi(h)$) are given in Subsection II.2. Combining (30), (31), (32), and (34) leads the desired result.

□

Proof of Lemma 3: Using the technique of proof of Lemmas 3.2 and 3.3. in Jurečková and Sen (1982), we can write

$$P^*\left\{\sqrt{m}|M_m^* - M_m| > t\right\} \leq 2P^*\left\{X_{m,m/2+1}^* \geq M_m + tm^{-1/2} - h\right\}$$

$$\leq 2\left[4F_m(M_m + tm^{-1/2} - h)\{1 - F_m(M_m + tm^{-1/2} - h)\}\right]^m$$

It follows that

$$P\left(P^*\left\{\sqrt{m}|M_m^* - M_m| > t\right\} > \alpha\right)$$

$$\leq P\left(\left[4F_m(M_m + tm^{-1/2} - h)\{1 - F_m(M_m + tm^{-1/2} - h)\}\right]^m > \alpha/2\right)$$

$$\leq P\left[1 - F_m(M_m + tm^{-1/2} - h) > \frac{1}{4}\left(\frac{\alpha}{2}\right)^{1/m}\right]$$

$$\leq P\left[1 - F_m(\theta + tm^{-1/2}/2 - h) > \frac{1}{4}\left(\frac{\alpha}{2}\right)^{1/m}\right] + P\left[\sqrt{n}|M_m - \theta| \geq t/2\right]$$

$$= P\left[F_m(\theta + tm^{-1/2}/2 - h) < 1 - \frac{1}{4}\left(\frac{\alpha}{2}\right)^{1/m}\right] + P\left[\sqrt{n}|M_m - \theta| \geq t/2\right]$$

$$\leq P\left[X_{(m:[m(1-\varepsilon)]^\circ)} > \theta + tm^{-1/2}/2 - h\right] + P\left[\sqrt{n}|M_m - \theta| \geq t/2\right] \tag{35}$$

for all $m > m_0$, where $0 < \varepsilon < 1/4$ and m_0 are such that

$$1 - \frac{1}{4}\left(\frac{\alpha}{2}\right)^{1/m_0} < 1 - \varepsilon.$$

Jurečková and Sen (1982) investigated the behaviour of the second probability in (35). They showed that, for $t > 2hm^{1/2}$, it is bounded by

$$2\left[4F(\theta - tm^{-1/2}/2 - h)(1 - F(\theta - tm^{-1/2}/2 - h))\right]^m. \tag{36}$$

Using their technique, we investigate also the first probability in (35). For the simplicity of notation, we denote

$$a = F(\theta + tm^{-1/2}/2 - h).$$

Notice that $a > 1/2$ for $t > 2hm^{1/2}$. Using Theorem 1 of Hoeffding (1963) with the bounds for $a > 1/2$, we have

$$P\left[X_{(m:[m(1-\varepsilon)]^\circ)} > a\right] \leq P\left[\frac{1}{m}Bi(m,a) - a > 1 - \varepsilon\right]$$

$$\leq \exp\left\{-\frac{m(1-\varepsilon)^2}{2a(1-a)}\right\} \leq [\exp\{-1/4a(1-a)\}]^m \leq [4a(1-a)]^m \tag{37}$$

Combining (35), (36), and (37) yields the lemma.

□

IV SIMULATION STUDY

For the simulations presented in this section we decided to use Huber's $\psi_h(.)$ function which is (for $h > 0$) defined as $\psi_h(x) = \max\{-h, \min(x, h)\}$. This score function has been derived by Huber (1981) as a score function which minimizes the worst possible variance which can be obtained for ε-contaminated Normal distribution. The value of h should be chosen accordingly to the level of contamination ε and it can be obtained as a solution of the equation

$$2\Phi(h) - 1 + \frac{2\varphi(h)}{h} = \frac{1}{1-\varepsilon}, \tag{38}$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ denote the distribution and density function of the Normal distribution, respectively (Antoch and Vorlíčková 1992).

IV.1 Normal Distribution

The asymptotically optimal stopping time for normally distributed data and for the sequential procedure based on M-estimators with Huber's score function is equal to

$$c_M(d) = \left(\frac{u_{1-\alpha/2}}{d}\right)^2 \frac{1 - 2\varphi(h) + (h^2 - 2)(1 - \Phi(h))}{(2\Phi(h) - 1)^2}. \tag{39}$$

We carried out the simulations for the desired length of interval equal to 1, 0.6, and 0.2, i.e., for $d = 0.5$, $d = 0.3$, and $d = 0.1$. The tuning parameters γ and k for the three-stage procedure were chosen as $\gamma = 1/3$ and $k = 1/2$. The calculations

TABLE I

Normal distribution, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 1/2$

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	16	N	15.71	14	0.923
		A	18.86	17	0.934
		U	23.94	21	0.964
0.30	45	N	39.95	37	0.920
		A	48.65	44	0.932
		U	55.88	52	0.945
0.10	399	N	392.74	391.5	0.927
		A	485.46	457	0.948
		U	490.42	465.5	0.943

were repeated 1000 times. The results of simulations are presented in Table I and in Figure I.

The first column of the table contains the value of d which is one half of the desired length $2d$ of the confidence interval. The (asymptotically) optimal number of observations for the standard Normal distribution (39) is given in the second column. The third column specifies the critical points which were used for calculations. "N" denotes the normal critical points, "A" stands for the standardized bootstrap critical points, and "U" denotes the studentized bootstrap critical points. In the fourth, fifth and sixth column, we present the mean, median and the coverage probability estimated from 1000 simulations.

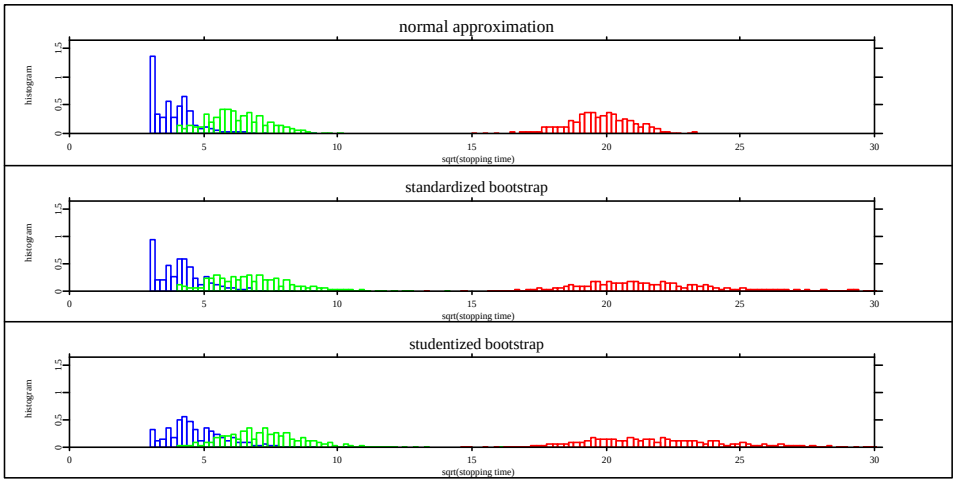
We carried out the simulations for the desired lengths of confidence interval equal to $2d = 1$, $2d = 0.6$, and $2d = 0.2$. As expected, $d = 0.5$ typically leads to a smaller sample size.

Comparing the values of mean and median for all three methods with the optimal sample size given in the tables, we see that the method "N" tends to underestimate the optimal sample size. The coverage probability of 95% was reached only for the method based on bootstrap critical points for $d = 0.5$, other coverage probabilities lie below the value 0.95. Notice that in all cases, the coverage probabilities for the bootstrap based methods are closer to 0.95 than the method based on normal critical points.

In Figure I, you can see the histograms of the square root of the stopping times. The first display on each figure shows histograms for the normal critical points, the second display shows histograms for the procedure based on the standardized bootstrap critical points and on the third display we plot the histograms for the procedure based on studentized bootstrap. The histograms at each display correspond to the stopping times for procedures giving fixed-width confidence intervals of lengths $2d = 1$, $2d = 0.6$, and $2d = 0.2$, respectively. We can clearly see the dependence of the variance on the desired length of the confidence interval.

The difference in the magnitude of the stopping times between these three meth-

FIGURE I
$\sqrt{N_2}$ *for Normal distribution.*



ods is clearly visible for $d = 0.1$.

Also the coverage probabilities and stopping times given in Table I suggest that the method based on normal approximation tends to stop too early. It seems that the methods based on bootstrap correct for this feature. On the other hand, the graphics suggests that the variance of N_2 for the methods based on bootstrap is larger than for the method based on normal critical points.

IV.2 Cauchy Distribution

Even more interesting are simulations for the Cauchy distribution. In this case, the methods based on sample mean can not be used at all, because Cauchy distribution has heavy tails. The possibility to use M-estimators in this situation is therefore of great importance.

The asymptotically optimal stopping time for Cauchy distribution is equal to

$$c_M(d) = \left(\frac{u_{1-\alpha/2}}{d}\right)^2 \frac{h^2 + 2/\pi[h - (h^2+1)\operatorname{arctg} h]}{(2/\pi \operatorname{arctg} h)^2} \tag{40}$$

For this simulations, we decided to modify the rule for choosing the starting sample size. It seems that bootstrap does not have to provide reliable results if the starting sample size is too small. For this reason, we use starting sample size always greater than 20.

It seems that here the asymptotics works better than for the Normal distribution in the previous subsection. The likely reason is that the confidence intervals of the same width require approximately two times more observations for the Cauchy than for the standard Normal distribution

FIGURE II

$\sqrt{N_2}$ *for Cauchy distribution,* $k = 0.5$, $m_0 \geq 20$.

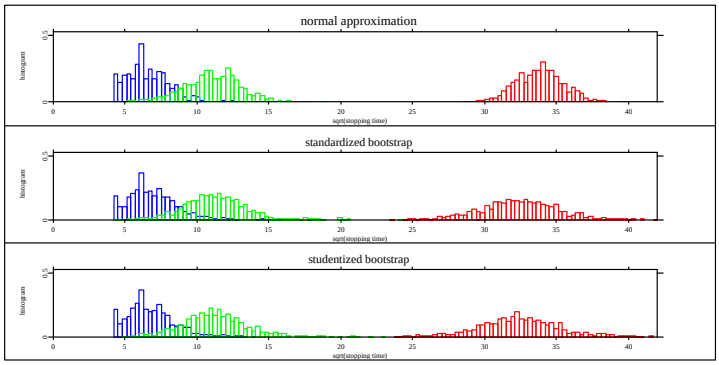


FIGURE III

$\sqrt{N_2}$ *for Cauchy distribution,* $k = 0.7$, $m_0 \geq 20$.

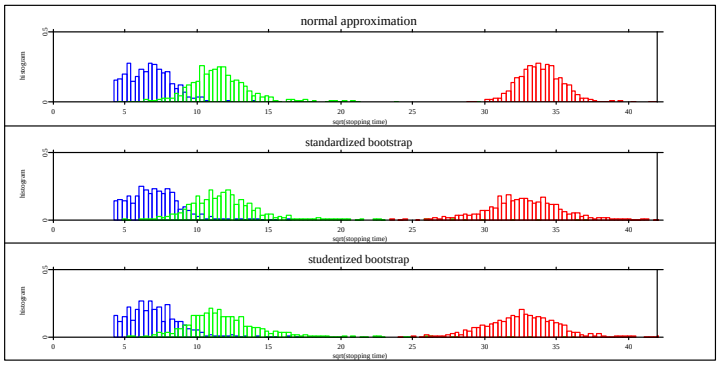


FIGURE IV

$\sqrt{N_2}$ *for Cauchy distribution,* $k = 0.9$, $m_0 \geq 20$.

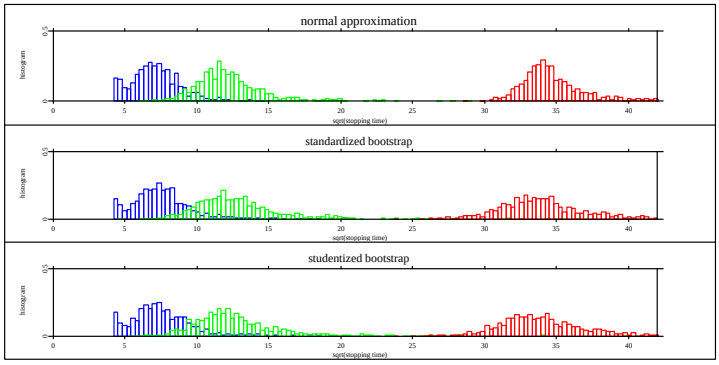


TABLE II

Coverage probabilities for Cauchy distribution, $h = 1.5$, $\gamma = 1/3$, $k = 0.5$, $m_0 \geq 20$

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	46	N	45.39	40	0.944
		A	51.95	45	0.959
		U	49.68	43	0.952
0.30	128	N	126.54	123	0.940
		A	134.34	126	0.950
		U	133.66	124	0.953
0.10	1150	N	1144.9	1147	0.955
		A	1060.2	1049	0.946
		U	1056.2	1043	0.948

TABLE III

Cauchy distribution, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $m_0 \geq 20$

d	critical points	$k = 0.5$	$k = 0.7$	$k = 0.9$
0.50	N	0.944	0.951	0.961
	A	0.959	0.957	0.968
	U	0.952	0.965	0.967
0.30	N	0.940	0.960	0.968
	A	0.950	0.961	0.975
	U	0.953	0.957	0.971
0.10	N	0.955	0.959	0.963
	A	0.946	0.948	0.955
	U	0.948	0.938	0.956

In Table III and Figures II–IV we tried to pursue another interesting phenomenon. As k increases, it seems that the histograms shift slightly to the right. This is confirmed by the values in Table II. We see that the coverage probabilities (together with the final sample size) increase with increasing k. Choosing $k = 0.9$ means that our procedure overestimates the necessary sample size—this effect is more dramatic for longer confidence intervals which result in smaller sample sizes.

We conclude that our method works well also for the Cauchy distribution. We can see that the bootstrap critical points approximate the correct distribution of the M-estimator better than the normal approximation. We recommend to choose the starting sample size to be at least 20 and we recommend to choose $k = 0.5$, or at most $k = 0.7$. Higher values should be chosen only for very small lengths of confidence intervals where the choice of the value of k does not seem to have negative impact on the coverage probability of the resulting confidence interval.

V ACKNOWLEDGMENTS

The research was partly supported by GA ČR 201/00/0769 and by MŠM 113200008. The paper is based on part of author's PhD. thesis.

BIBLIOGRAPHY

Aerts, M. (1988). *Asymptotic Distribution Theory for Random U- and GM-statistics with Application in Sequential Estimation, PhD. thesis.* Limburgs Universitair Centrum Diepenbeek

Aerts, M. & Gijbels, I. (1993). A three-stage procedure based on bootstrap critical points. *Sequential Analysis* 12(2), 93–113

Antoch, J. & Vorlíčková, D. (1992). *Selected Methods of Statistical Data Analysis (in Czech)*, Academia, Praha

Anscombe, F.J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* 48, 600–607.

Arcones, M. & Giné, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In: R. LePage and L. Billard, Eds., *Exploring the Limits of Bootstrap*, John Wiley, New York, 13–48.

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

Dodge, Y. & Jurečková, J. (1995). Estimation of quantile density function based on regression quantiles. *Statistics & Probability Letters* 23, 73–78.

Ghosh, M. & Mukhopadhyay, N. (1981). Consistency and asymptotic efficiency of two-stage and sequential estimation procedures. *Sankhya Ser. A* 43, 220–227.

Ghosh, M., Mukhopadhyay, N. & Sen, P.K. (1997). *Sequential Analysis*, Wiley, New York.

Ghosh, B.K. & Sen, P.K. (1991). *Handbook of Sequential Analysis*, Dekker, New York.

Hall, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.* 9, 1229–1238.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.

Hampel, F.R., Ronchetti, E.R., Rousseeuw, P.J. & Stahel, W.A. (1990). *Robust Statistics - The Approach Based on Influence Functions*, Wiley, New York.

Hlávka, Z. (2000). *Robust Sequential Methods, PhD. thesis.* Faculty of Mathematics and Physics, Charles University, Prague.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Jour. Amer. Statist. Assoc.* 58, 13–29.

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

Jurečková, J. (1978). Bounded-length sequential confidence intervals for regression and location parameters. *The Second Prague Symposium on Asymptotic Statistics*, Charles University, Prague.

Jurečková, J. & Sen, P.K. (1981). Sequential procedures based on M-estimators with discontinuous score functions. *Jour. Statist. Plan. Infer.* 5, 253–266.

Jurečková, J. & Sen, P.K. (1982). M-estimators and L-estimators of location: uniform integrability and asymptotic risk-efficient sequential versions. *Commun. Statist., -Sequential Analysis* 1(1), 27–56.

Jurečková, J. & Sen, P.K. (1995). *Robust Statistical Procedures: Asymptotics and Interrelations*, Wiley, New York.

Karabulut, I. & Lahiri, S.N. (1997). Two-term edgeworth expansion for M-estimators of a linear regression parameter without Cramér-type conditions and an application to the bootstrap. *J. Austral. Math. Soc. (Series A)* 62, 361–370.

Lahiri, S.N. (1992). Bootstrapping M-estimators of a multiple linear regression parameter. *Ann. Statist.* 20, 1548–1570.

Lahiri, S.N. (1992). On bootstrapping M-estimators. *Sankhyā A* 54, 157–170.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1992). *Numerical Recipes in FORTRAN, The Art of Scientific Computing, Second Edition*, Cambridge University Press.

Rieder, H. (1994). *Robust Asymptotic Statistics*, Springer-Verlag, New York.

Sen, P.K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*, Wiley, New York.

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

Shao, J. & Tu, D. (1995). *The Jacknife and Bootstrap*, Springer-Verlag, New York.

Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* 17, 498–499.

Stein, C. (1949). Some problems in sequential estimation (abstract). *Econometrica* 17, 77–78.

Swanepoel, J.W.H, van Wijk, J.W.J. &Venter, J.H. (1984). Fixed-width confidence intervals based on bootstrap procedures. *Commun. in Statist., -Seq. Analys.* 2, 289–310.