Z. Wahrscheinlichkeitstheorie verw. Gebiete
61, 541–551 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Most Robust *M*-Estimators in the Infinitesimal Sense

Peter J. Rousseeuw

Mathematics Department, F7, VUB, Pleinlaan 2, B-1050 Brussels, Belgium

Summary. The change-of-variance curve (CVC) is generalized to M-estimators with piecewise continuous ψ-functions, in which case it becomes a Schwartz distribution. An M-estimator is called *most B-robust* when it minimizes Hampel's gross-error sensitivity γ^*, and *most V-robust* when it minimizes the change-of-variance sensitivity κ^*. In the general case, the median is most B-robust and most V-robust. If consideration is restricted to redescending M-estimators, then the skipped median is most B-robust and the median-type tanh-estimator is most V-robust. By means of these results, complete solutions of the problems of optimal infinitesimal robustness are obtained.

1. Introduction

M-estimators of location were introduced by Huber (1964), who studied their robustness properties by means of a minimax theorem for the asymptotic variance. The infinitesimal robustness of the asymptotic value was described in an intuitively appealing way by Hampel's (1974) influence curve (IC). In order to investigate the infinitesimal robustness of the asymptotic variance, Rousseeuw (1981) defined the change-of-variance curve (CVC), rediscovering an idea of F. Hampel going back to 1972. With this tool a new class of redescending M-estimators was constructed by Hampel, Rousseeuw and Ronchetti (1981). However, these papers imposed the severe restriction that the underlying ψ-function be continuous, which excluded many interesting estimators such as the median and the Huber-type skipped mean. In the present paper this restriction is dispensed with, and the general CVC becomes a Schwartz distribution.

The aim of this paper is to study and to compare various robustness concepts which may be associated with the IC and the CVC. To the first we assign the prefix "B-" from "bias", and for the latter we use "V-" from "variance". It was already proven in (Rousseeuw 1981) that V-robustness

0044-3719/82/0061/0541/$02.20

implies B-robustness, and that Huber's minimax solutions are optimal B-robust as well as optimal V-robust. In Sect. 3 the notions of *most B-robust* and *most V-robust* estimators are introduced. In the general case, the median satisfies both properties (Theorems 1 and 2). This enables us to give complete solutions to the problems of optimal infinitesimal robustness. In Sect. 4, attention is focused on redescending M-estimators. In this subclass the *skipped median* is most B-robust, whereas the *median-type tanh-estimator* is most V-robust. At the normal distribution, the corresponding problems of optimal robustness lead to skipped Huber estimators and to tanh-estimators.

2. General Definition of the CVC and Basic Results

Consider Huber's (1964) framework of robust estimation. Let $X_1, \ldots, X_n$ be i.i.d. observations with distribution function $G(x-\theta)$, where G satisfies certain regularity conditions. An M-estimator of the location parameter θ is given by an equation of the form

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

There are several ways to select a solution, such as taking the root nearest to the sample median or using Newton's method starting with the median as in Collins (1976). Under regularity conditions on ψ and G (Huber 1967), the sequence (T_n) is consistent and $n^{1/2}(T_n-\theta)$ is asymptotically normal with asymptotic mean zero and asymptotic variance

$$V(\psi, G) = \int \psi^2 \, dG / (\int \psi' \, dG)^2$$

The model distribution F (identified with its cumulative distribution function) is fixed, and satisfies

(F1) F has a twice continuously differentiable density f which is symmetric and strictly positive;

(F2) the function $\Lambda = -f'/f = (-\ln f)'$ satisfies $\Lambda'(x) > 0$ for all x, and $\int \Lambda'(x) f(x)\, dx = -\int \Lambda(x) f'(x)\, dx < \infty$.

The function Λ, which is continuously differentiable by (F1), is the ψ-function corresponding to the maximum likelihood estimator (MLE). The assumption $\Lambda'(x) > 0$ for all x entails convexity of $-\ln f$ and unimodality of f. Condition (F2) implies that the Fisher information

$$I(F) = \int \Lambda^2 \, dF$$

satisfies $0 < I(F) = \int \Lambda' \, dF < \infty$. Note that F also satisfies Huber's conditions (1964, page 80), so his minimax asymptotic variance theorem is applicable at F.

Examples. Our favourite choice for F will be the standard normal Φ with density $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ and $I(\Phi) = 1$. The MLE is the arithmetic mean, given by $\Lambda(x) = x$. The logistic distribution provides another example, with $F(x) = (1 + \exp(-x))^{-1}$, $\Lambda(x) = \tanh(x/2) = 2F(x) - 1$ and $I(F) = 1/3$.

The class Ψ consists of all real functions ψ satisfying:
(i) ψ is well-defined and continuous on $\mathbb{R} \setminus C(\psi)$, where $C(\psi)$ is finite. In each point of $C(\psi)$, there exist finite left and right limits of ψ which are different. Also $\psi(-x) = -\psi(x)$ if $\{x, -x\} \subset \mathbb{R} \setminus C(\psi)$ and $\psi(x) \geqq 0$ for $x \geqq 0$ not belonging to $C(\psi)$;
(ii) the set $D(\psi)$ of points in which ψ is continuous but in which ψ' is not defined or not continuous, is finite;
(iii) $\int \psi^2 \, dF < \infty$;
(iv) $0 < \int \psi' \, dF = -\int \psi(x) f'(x) \, dx = \int \Lambda \psi \, dF < \infty$.
Clearly, $C(\psi)$ and $D(\psi)$ are symmetric about zero. We say the functions ψ_1 and ψ_2 in Ψ are *equivalent* if and only if $C(\psi_1) = C(\psi_2)$ and all x not in this set satisfy $\psi_1(x) = r\psi_2(x)$ $(r > 0)$.

Note that this class Ψ generalizes those of Rousseeuw (1981) and Hampel, Rousseeuw and Ronchetti (1981), which only contain continuous ψ-functions. The latter restriction made mathematical manipulations easier, while many estimators were covered. But on the other hand some important examples with discontinuous ψ were excluded, such as the median and the Huber-type skipped mean. To the author's knowledge, the present class Ψ covers all ψ-functions ever used for this estimation problem.

What is the meaning of expressions of the type $\int \psi' \, dG$ in this general case? The answer is given by Huber (1964, page 78) who states that ψ' may be interpreted as a Schwartz distribution, so

$$\int \psi' \, dG = \int_{\mathbb{R} \setminus (C(\psi) \cup D(\psi))} \psi' \, dG + \sum_{i=1}^{m} [\psi(c_i+) - \psi(c_i-)] \, g(c_i)$$

where the first term is a classical integral, $c_1 < \ldots < c_m$ are the points of $C(\psi)$ and g is the density of G. Therefore, ψ' formally denotes the Schwartz distribution

$$\psi' \, 1_{\mathbb{R} \setminus (C(\psi) \cup D(\psi))} + \sum_{i=1}^{m} [\psi(c_i+) - \psi(c_i-)] \, \delta_{c_i},$$

which is the sum of a "regular" part and a linear combination of Dirac delta "functions" δ_{c_i}. With this in mind, put

$$A(\psi) = \int \psi^2 \, dF \quad \text{and} \quad B(\psi) = \int \psi' \, dF.$$

From the definition of Ψ it follows that $0 < A(\psi) < \infty$ and $0 < B(\psi) < \infty$. Hampel's (1974) influence curve (IC) describes the infinitesimal behaviour of the asymptotic value of the estimator corresponding to ψ, and is given by

$$\Omega(\psi, F, x) = \psi(x)/B(\psi)$$

on $\mathbb{R} \setminus C(\psi)$. The gross-error sensitivity equals

$$\gamma^*(\psi) = \sup_{x \in \mathbb{R} \setminus C(\psi)} |\Omega(\psi, F, x)|.$$

Let us now generalize the change-of-variance curve (Rousseeuw 1981) to the class Ψ. Given some ψ in Ψ, we want to investigate the infinitesimal stability of $V(\psi, G)$ in the vicinity of F. Consider a distribution G which has a symmetric density g and satisfies $0<\int\psi^2\,dG<\infty$ and $0<\int\psi'\,dG<\infty$. Keeping in mind the interpretation of ψ', we may verify that

$$\frac{\partial}{\partial\varepsilon}[\ln V(\psi,(1-\varepsilon)F+\varepsilon G)]_{\varepsilon=0}=\int\left[1+\frac{\psi^2(x)}{A(\psi)}-2\frac{\psi'(x)}{B(\psi)}\right]dG(x).$$

This motivates the following definition.

Definition. The change-of-variance curve $\Xi(\psi,F,x)$ of $\psi\in\Psi$ at F is the Schwartz distribution consisting of the sum of the regular part

$$\left[1+\frac{\psi^2(x)}{A(\psi)}-2\frac{\psi'(x)}{B(\psi)}\right]1_{\mathbb{R}\setminus(C(\psi)\cup D(\psi))}(x)$$

which is continuous on $\mathbb{R}\setminus(C(\psi)\cup D(\psi))$, and

$$-\frac{2}{B(\psi)}\left[\sum_{i=1}^{m}(\psi(c_i+)-\psi(c_i-))\,\delta_{c_i}(x)\right].$$

Now $\Omega(\psi,F,x)$ is skew-symmetric, $\Xi(\psi,F,x)$ is symmetric and $\int\Omega(\psi,F,x)\,dF(x)=0=\int\Xi(\psi,F,x)\,dF(x)$ for all ψ in Ψ.

Examples. The median corresponds to $\psi_{\text{med}}(x)=\text{sign}(x)$, so $V(\psi_{\text{med}},F)=1/(2f(0))^2$ and

$$\Xi(\psi_{\text{med}},F,x)=2\left[1_{\mathbb{R}\setminus\{0\}}(x)-\frac{1}{f(0)}\delta_0(x)\right].$$

On the other hand a Huber-type skipped mean, given by $\psi_{\text{sk}(c)}(x)=x\,1_{[-c,c]}(x)$ where $0<c<\infty$, yields a CVC containing two delta "functions" with *positive* factor, because of the downward jumps of $\psi_{\text{sk}(c)}$ at c and $-c$.

Definition. The change-of-variance sensitivity $\kappa^*(\psi)$ is defined as $+\infty$ if a delta "function" with positive factor occurs in the CVC, and otherwise as

$$\kappa^*(\psi)=\sup\{\Xi(\psi,F,x);\ x\in\mathbb{R}\setminus(C(\psi)\cup D(\psi))\}.$$

This definition generalizes that of (Rousseeuw 1981), where a rationale can be found. On a heuristic level, $\kappa^*(\psi)$ may be compared with the robustness measure $\sup\{V(\psi,F);F\in\mathfrak{F}\}$ that was recently studied in detail by Collins (1977) and Collins and Portnoy (1981).

Clearly, upward jumps of ψ (adding only a negative delta function to the CVC) do not involve $\kappa^*(\psi)$, but any downward jump makes it infinite. For example, we have $\kappa^*(\psi_{\text{med}})=2$ and $\kappa^*(\psi_{\text{sk}(c)})=\infty$. An M-estimator is called *B-robust* (from "bias") *when* $\gamma^*(\psi)$ is finite, and *V-robust* (from "variance") when $\kappa^*(\psi)$ is finite.

Lemma 1. *V-robustness implies B-robustness for all ψ in Ψ, and they are equivalent for monotone ψ.*

Proof. The proof is a straightforward generalization of Theorem 1 and its corollary in (Rousseeuw 1981). In the second part we note that $\psi(c_i+)-\psi(c_i-)>0$ for all c_i in $C(\psi)$, hence $\kappa^*(\psi)=\sup\{\Xi(\psi,F,x);\ x\in\mathbb{R}\setminus(C(\psi)\cup D(\psi))\}$. □

Remark. In general we do not have equivalence, as is exemplified by the Huber-type skipped mean.

Examples. The median is both B-robust and V-robust. The same holds for the MLE if Λ is bounded, as is the case when F is the logistic distribution, where $\gamma^*(\Lambda)=3$ and $\kappa^*(\Lambda)=4$. However, if Λ is unbounded (e.g. at $F=\Phi$), then the MLE is neither B-robust nor V-robust.

3. Most Robust M-Estimators

The main purpose of this paper is to determine those estimators which do not only have finite sensitivities, but which even possess the smallest sensitivities possible. An estimator minimizing $\gamma^*(\psi)$ we call *most B-robust*, and when it minimizes $\kappa^*(\psi)$ we say it is *most V-robust.*

Theorem 1. *The median is the most B-robust estimator in Ψ. For all ψ in Ψ we have $\gamma^*(\psi)\geqq(2f(0))^{-1}$, and equality holds if and only if ψ is equivalent to ψ_{med}.*

Proof. Assume that ψ is bounded, otherwise $\gamma^*(\psi)=\infty$ and there is nothing left to prove. Clearly $\sup_{x\in\mathbb{R}\setminus C(\psi)} |\psi(x)|>0$, or else we would have $A(\psi)=0$. Then

$$B(\psi)=\int|\Lambda||\psi|\,dF\leqq \sup_{x\in\mathbb{R}\setminus C(\psi)} |\psi(x)|\int|\Lambda|\,dF.$$

It holds that $\int|\Lambda|\,dF=2\int_0^\infty -f'(x)\,dx=2f(0)$, hence $\gamma^*(\psi)=\sup_{x\in\mathbb{R}\setminus C(\psi)}|\psi(x)|/B(\psi)\geqq(2f(0))^{-1}=\gamma^*(\psi_{\text{med}})$.

For the uniqueness part, suppose that some ψ in Ψ satisfies $\gamma^*(\psi)=(2f(0))^{-1}$. This implies $\int|\Lambda||\psi|\,dF=\int|\Lambda|\sup_{x\in\mathbb{R}\setminus C(\psi)}|\psi(x)|\,dF$, where $f(y)>0$ for all y and $|\Lambda(y)|>0$ for $y\neq 0$. By means of some elementary analysis it follows for all $y\in\mathbb{R}\setminus C(\psi)$ that $|\psi(y)|=\sup_{x\in\mathbb{R}\setminus C(\psi)}|\psi(x)|$, which is a strictly positive finite constant. Suppose w.l.o.g. that this constant equals 1. But from $\psi(x)\geqq 0$ for positive x in $\mathbb{R}\setminus C(\psi)$ and skew symmetry of ψ, this implies $\psi(x)=\operatorname{sign}(x)$ for all $x\neq 0$ not belonging to $C(\psi)$. Therefore $C(\psi)=\{0\}$ and $\psi=\psi_{\text{med}}$, which ends the proof. □

Remark. P. Huber (1981, page 74) already showed in a different setting that the median gives the smallest asymptotic bias.

The M- estimators corresponding to

$$\begin{aligned}\psi_t(x)&=\Lambda(x) & |x|&\leqq t\\ &=\Lambda(t)\operatorname{sign}(x) & |x|&>t\end{aligned}$$

(where $0<t<\infty$) were introduced by P. Huber (1964, page 80) as solutions to his minimax asymptotic variance problem. In the special case $F=\Phi$, they correspond to $\min(t,\max(x,-t))$ and are generally called *Huber estimators*. We say that an M-estimator is *optimal B-robust* if it minimizes $V(\psi,F)$ under the side condition of an upper bound on $\gamma^*(\psi)$. Hampel (1974) proved that the ψ_t

are optimal B-robust. Now that the range of γ^* is known from Theorem 1, it becomes possible to list the complete solution of the problem of optimal B-robustness.

Corollary 1. *The only optimal B-robust M-estimators are (up to equivalence) given by* $\{\psi_{\text{med}}, \psi_t (0<t<\infty)\}$ *if* Λ *is unbounded, and by* $\{\psi_{\text{med}}, \psi_t (0<t<\infty), \Lambda\}$ *otherwise.*

Proof. From an extension of Lemmas 1 and 2 of (Rousseeuw 1981) it follows that for each constant g in $((2f(0))^{-1}, \gamma^*(\Lambda))$ there exists a unique t in $(0,\infty)$ such that $\gamma^*(\psi_t)=g$, and that this ψ_t minimizes $V(\psi, F)$ among all ψ in Ψ which satisfy $\gamma^*(\psi)\leqq g$; moreover, any other solution is equivalent to ψ_t. Theorem 1 implies that no solution can exist for any $g<(2f(0))^{-1}=\gamma^*(\psi_{\text{med}})$. If one puts $g=(2f(0))^{-1}$, then only functions equivalent to ψ_{med} can satisfy $\gamma^*(\psi)\leqq g$, and therefore the median itself is automatically optimal B-robust. Let us now consider the upper extremity $\gamma^*(\Lambda)$. If $\gamma^*(\Lambda)=\infty$, the MLE is not *B-robust. If* $\gamma^*(\Lambda)<\infty$, the consideration of any $g\geqq\gamma^*(\Lambda)$ always yields Λ itself, because Λ minimizes $V(\psi, F)$ in Ψ (Cauchy-Schwarz). □

Theorem 2. *The median is also the most V-robust estimator in* Ψ. *For all* ψ *in* Ψ *we have* $\kappa^*(\psi)\geqq 2$, *and equality holds if and only if* ψ *is equivalent to* ψ_{med}.

Proof. We start by proving the desired inequality. In case ψ is unbounded, it holds that $\gamma^*(\psi)=\infty$ which implies $\kappa^*(\psi)=\infty$ by Lemma 1. We may therefore assume that ψ is bounded, so $0<\sup_{x\in\mathbb{R}\setminus C(\psi)}|\psi(x)|<\infty$. We put this supremum equal to 1 w.l.o.g., so $A(\psi)\leqq 1$. We also assume that in each point of $C(\psi)$ (if any) the jump $\psi(c_i+)-\psi(c_i-)$ is positive, because otherwise again $\kappa^*(\psi)=\infty$.

Case A. First suppose that $\sup_{x\in\mathbb{R}\setminus C(\psi)}|\psi(x)|$ is reached by ψ or by one of the left and right limits of ψ at points of $C(\psi)$. We restrict our attention to $[0,\infty)$, where $\psi(x)\geqq 0$ for $x\in\mathbb{R}\setminus C(\psi)$. *Case A.1* Suppose there exists $y\in[0,\infty)$ such that $y\notin(C(\psi)\cup D(\psi))$ and $|\psi(y)|=1$. Then $\psi(y)=1$, and $y>0$ because of skew-symmetry of ψ. Clearly $\psi'(y)=0$. Finally $\Xi(\psi, F, y)=1+\psi^2(y)/A(\psi)\geqq 1+\psi^2(y)=2$, hence $\kappa^*(\psi)\geqq 2$. *Case A.2.* Suppose there exists $y\in[0,\infty)$ such that $y\in D(\psi)$ and $\psi(y)=1$, or $y\in C(\psi)$ and $\psi(y+)=1$. Then there is some $\delta>0$ such that $(y, y+\delta)\subset\mathbb{R}\setminus(C(\psi)\cup D(\psi))$. For all ε in $(0,\delta)$ there exists a point z in $(y, y+\varepsilon)$ for which $\psi'(z)\leqq 0$, or else the supremum would become strictly larger than 1. We can therefore construct a sequence $x_n\downarrow y$, $x_n\in(y, y+\delta)$ such that $\psi'(x_n)\leqq 0$ for all n, hence $\Xi(\psi, F, x_n)\geqq 1+\psi^2(x_n)/A(\psi)$. Now $\lim_{n\to\infty}(1+\psi^2(x_n)/A(\psi))=1+A(\psi)^{-1}\geqq 2$, hence $\kappa^*(\psi)\geqq\sup\{\Xi(\psi, F, x_n);\, n>0\}\geqq\sup\{1+\psi^2(x_n)/A(\psi);\, n>0\}\geqq 2$.

Case B is the negation of case *A*. Now $|\psi(x)|<1$ for all $x\in\mathbb{R}\setminus C(\psi)$, so $A(\psi)<1$, hence $1/A(\psi)=1+3\varepsilon$ where $\varepsilon>0$. There exists $0<K<\infty$ such that $(C(\psi)\cup D(\psi))\subset(-K, K)$. Clearly, $\sup_{[0,\infty)\setminus C(\psi)}\psi(x)=1$. However, $\sup_{[0,K]}\psi(x)<1$ by the Weierstrass theorem, because ψ is piecewise continuous and $[0,K]$ is compact. Hence, $\sup_{(K,\infty)}\psi(x)=1$ and on this set ψ is continuously differentiable. There exists $x>K$ such that $\psi(x)>(A(\psi)(1+2\varepsilon))^{1/2}$. *Case B.1.* Suppose

$\psi'(x) \leqq 0$. Then $\Xi(\psi, F, x) \geqq 1 + \psi^2(x)/A(\psi) \geqq 2 + 2\varepsilon$, hence $\kappa^*(\psi) > 2$. *Case B.2.* Suppose $\psi'(x) > 0$ and there exists $y > x$ such that $\psi'(y) \leqq 0$. By continuity of ψ', $x < z = \inf\{t \in [x, y];\ \psi'(t) \leqq 0\} \leqq y$ and $\psi'(z) = 0$. But now $\psi(z) > \psi(x)$ or we could find $u \in (x, z)$ with $\psi'(u) \leqq 0$ by Lagrange's mean value theorem. Thus $\Xi(\psi, F, z) = 1 + \psi^2(z)/A(\psi) > 2 + 2\varepsilon$, hence $\kappa^*(\psi) > 2$. *Case B.3.* Suppose $\psi'(y) > 0$ for all $y \geqq x$. It is clearly impossible that $\psi'(y) > B(\psi)\varepsilon/2 > 0$ for all $y \geqq x$, because then ψ would ultimately grow larger than 1. Therefore there exists $z \geqq x$ for which $\psi'(z) \leqq B(\psi)\varepsilon/2$, and $\psi(z) \geqq \psi(x)$. Hence $\Xi(\psi, F, z) \geqq 1 + \psi^2(x)/A(\psi) - \varepsilon > 2 + \varepsilon$, so $\kappa^*(\psi) > 2$. The inequality $\kappa^*(\psi) \geqq 2$ is hereby proven in all cases.

In the second part of the proof we suppose that some ψ in Ψ satisfies $\kappa^*(\psi) = 2$. This implies that ψ is bounded (Lemma 1), and again we put $\sup_{x \in \mathbb{R} \setminus C(\psi)} |\psi(x)| = 1$. Revisiting the different cases of the first part of the proof, we see that case B is excluded because there always $\kappa^*(\psi) > 2$. In case A, we can only have equality if $A(\psi) = 1$. By means of some elementary analysis it follows that $|\psi(y)| = 1$ for all $y \in \mathbb{R} \setminus C(\psi)$. The proof then proceeds as in Theorem 1. □

If we now replace $\gamma^*(\psi)$ by $\kappa^*(\psi)$ in the definition of optimal B-robustness, then we can speak of *optimal V-robustness.*

Corollary 2. *The optimal V-robust M-estimators coincide with the estimators listed in Corollary 1.*

Proof. Lemma 1 implies that $\kappa^*(\Lambda)$ is finite if and only if Λ is bounded. From an extension of Lemma 1 and Theorem 2 of (Rousseeuw 1981) it follows that for each k in $(2, \kappa^*(\Lambda))$ there is a unique t in $(0, \infty)$ for which $\kappa^*(\psi_t) = k$, and that this ψ_t minimizes $V(\psi, F)$ among all ψ in Ψ which satisfy $\kappa^*(\psi) \leqq k$; any other solution is equivalent to ψ_t. Theorem 2 implies that no solution can exist for any $k < 2$. The result then follows as in Corollary 1. □

4. Redescending M-Estimators

In this section we focus our attention to the subclass

$$\Psi_c = \{\psi \in \Psi;\ \psi(x) = 0 \text{ for all } |x| > c\}$$

of Ψ, where $0 < c < \infty$ is a fixed constant. This implies that Hampel's (1974) rejection point $\rho^*(\psi)$ is not larger than c, which means that all observations farther away than c are rejected. Such estimators are called *redescending*. They became well-known after their appearance in Andrews et al. (1972); for a recent survey, see Huber (1981, Sect. 4.8). On the other hand, estimators of this type go back to Daniel Bernoulli (Stigler 1980).

In Ψ_c, Lemma 1 becomes useless since each ψ is already B-robust by the Weierstrass theorem (being bounded on $[-c, c]$ because of piecewise continuity), and this class does not contain any monotone functions. On the other hand, there exist elements of Ψ_c which are not V-robust, such as the Huber-type skipped mean. The lower bounds for γ^* and κ^* as given by Theorems 1 and 2 remain valid, but they can no longer be reached because the median is not redescending. Exact bounds will be given in the present section. Inspired by Theorem 1, we introduce

$$\psi_{\text{med}(c)}(x) = \text{sign}(x)\, 1_{[-c,c]}(x).$$

We call this estimator a *skipped median*, because observations farther away than c are skipped. (Here "skipped" refers to the Huber-type skipped mean, and not to Tukey's skipping procedures (Andrews et al., 1972).)

Theorem 3. *The skipped median is the most B-robust estimator in Ψ_c. For all ψ in Ψ_c we have $\gamma^*(\psi) \geqq [2(f(0)-f(c))]^{-1}$, and equality holds if and only if ψ is equivalent to $\psi_{\text{med}(c)}$.*

Proof. The proof is a simple adaptation of that of Theorem 1, where the real line is replaced by $[-c, c]$. □

The problem of optimal B-robustness also has to be solved anew, because the mappings ψ_t do not belong to Ψ_c. For all t in $(0, c)$ we define:

$$\psi_{c,t} = \psi_t\, 1_{[-c,c]}.$$

(At $F = \Phi$, we could say that $\psi_{c,t}$ determines a *skipped Huber estimator.*) If we let t equal c, then the horizontal part disappears and we obtain

$$\tilde{\psi}_c = \Lambda\, 1_{[-c,c]}.$$

We also introduce the notation

$$J(c) = \int_{[-c,c]} \Lambda^2\, dF,$$

and it follows that $A(\tilde{\psi}_c) = B(\tilde{\psi}_c) = J(c)$. (If $F = \Phi$, then $\tilde{\psi}_c$ determines the Huber-type skipped mean, and $J(c) = 2\Phi(c) - 1 - 2c\phi(c)$.)

Corollary 3. *The only optimal B-robust estimators in Ψ_c are (up to equivalence) given by $\{\psi_{\text{med}(c)}, \psi_{c,t} (0 < t < c), \tilde{\psi}_c\}$.*

Proof. By means of a reasoning analogous to Lemma 1 of (Rousseeuw 1981), we verify that $t \to \gamma^*(\psi_{c,t})$ is an increasing continuous bijection from $(0, c)$ onto $(\gamma^*(\psi_{\text{med}(c)}), \gamma^*(\tilde{\psi}_c))$. The rest of the proof mimics that of Corollary 1, keeping in mind that the upper endpoint $\gamma^*(\tilde{\psi}_c) = \Lambda(c)/J(c)$ is always finite. □

Unfortunately, all ψ-functions of Corollary 3 possess downward jumps at c and $-c$, hence they all have infinite change-of-variance sensitivities. Therefore, the most V-robust and the most B-robust estimators can no longer be the same.

Lemma 2. *There exist constants κ_c and B_c such that the function*

$$\chi_c(x) = (\kappa_c - 1)^{1/2} \tanh[\tfrac{1}{2} B_c (\kappa_c - 1)^{1/2} (c - |x|)]\, \text{sign}(x)\, 1_{[-c,c]}(x)$$

belongs to Ψ_c and satisfies $A(\chi_c) = 1$ and $B(\chi_c) = B_c$. Moreover, $\kappa^*(\chi_c) = \kappa_c > 2F(c)/(2F(c) - 1) > 2$ and $0 < V(\chi_c, F)^{-1} = B_c^2 < J(c) < I(F)$.

Proof. This is a straightforward generalization of Lemma 3.1 of Hampel, Rousseeuw and Ronchetti (1981), where χ_c was merely considered as an auxiliary construction because it is not continuous. To prove $\kappa^* = \kappa_c$, note that $\Xi(\chi_c, F, x) = \kappa_c$ if $0 < |x| < c$. □

In Table 1 some values of κ_c and B_c can be found at $F = \Phi$. Note that the asymptotic efficiency e equals B_c^2 because $A(\chi_c) = 1 = I(\Phi)$. Uniqueness of κ_c and B_c will follow from Theorem 4. Looking at Table 1, it appears that the estimator corresponding to χ_c yields an acceptable alternative to the median (corresponding to $c = \infty$) provided c is not too small. Apart from a finite ρ^* and γ^* it also possesses a finite κ^*, so χ_c is "more robust" than $\psi_{\mathrm{med}(c)}$. In fact, Theorem 4 states that χ_c is most V-robust in Ψ_c, a property shared by the median in Ψ. Moreover, $C(\chi_c) = \{0\} = C(\psi_{\mathrm{med}})$ and in both cases the jump is upward, so both estimators show the same behaviour at the center. Because of all this, we call the estimator corresponding to χ_c a *median-type tanh-estimator.*

Theorem 4. *The median-type tanh-estimator is the most V-robust estimator in* Ψ_c. *For all* ψ *in* Ψ_c *we have* $\kappa^*(\psi) \geq \kappa_c$, *and equality holds if and only if* ψ *is equivalent to* χ_c.

Proof. Assume that $\kappa^*(\psi) < \infty$. Consider the function ζ given by $\zeta(x) = \psi(x)/\chi_c(x)$ for $x \in (0, c) \setminus C(\psi)$ and $\zeta(0) = \psi(0+)/\chi_c(0+)$, which is continuous on $[(0, c) \setminus C(\psi)] \cup \{0\}$, satisfies $\zeta(x) \geq 0$ and is continuously differentiable on $(0, c) \setminus [C(\psi) \cup D(\psi)]$.

Case A. Suppose $\sup_{[0,c) \setminus C(\psi)} \zeta(x)$ is reached by ζ on $[0, c)$ or by a left or right limit of ζ in a point of $C(\psi)$. Suppose w.l.o.g. that this supremum equals 1. *Case A.1.* Suppose there exists y in $[0, c) \setminus [C(\psi) \cup D(\psi)]$ with $\zeta(y) = 1$. Then $y > 0$, $\chi_c(y) = \psi(y)$ and $\zeta'(y) = 0$, hence $\psi'(y) = \chi_c'(y) < 0$. We have $A(\psi) \leq 1$ and $B(\psi) \leq B_c$, so $\Xi(\psi, F, y) \geq \kappa_c$. *Case A.2.* Suppose there exists y in $[0, c)$ such that $y \in D(\psi)$ and $\zeta(y) = 1$, or $y \in C(\psi)$ and $\zeta(y+) = 1$. Constructing $x_n \downarrow y$ such that $\psi'(x_n) \leq \chi_c'(x_n)$ (see case A.2 of Theorem 2), we again verify that $\kappa^*(\psi) \geq \kappa_c$.

Table 1. Values of κ_c and B_c at $F = \Phi$

c	κ_c	B_c	e	γ^*
2.0	4.457305	0.509855	0.2600	2.6946
2.5	3.330328	0.604034	0.3649	2.0688
3.0	2.796040	0.668619	0.4471	1.7491
3.5	2.505102	0.711310	0.5060	1.5694
4.0	2.331507	0.739426	0.5468	1.4610
4.5	2.221654	0.758161	0.5748	1.3922
5.0	2.149604	0.770809	0.5941	1.3471
5.5	2.101379	0.779423	0.6075	1.3168
6.0	2.068765	0.785313	0.6167	1.2964
7.0	2.031553	0.792091	0.6274	1.2731
8.0	2.014392	0.795236	0.6324	1.2623
10.0	2.002953	0.797340	0.6358	1.2552
∞	2.000000	0.797885	0.6366	1.2533

Case B is the negation of case *A*. From $\kappa^*(\psi)<\infty$ it follows that $\psi(c)=0$, and there exists $M\geqq -\psi'(x)$ for all x. Choose $0<\delta<c$ such that $[c-\delta, c) \subset \mathbb{R}\setminus[C(\psi)\cup D(\psi)]$. Then $0\leqq\psi(x)\leqq M(c-x)$ for all x in $[c-\delta, c]$, hence $\sup_{[c-\delta,c)} \zeta(x)<\infty$; put this supremum equal to 1. Because $\sup_{[0,c-\delta]\setminus C(\psi)} \zeta(x)<1$, we have $A(\psi)<1$ and $B(\psi)<B_c$. *Case B.1.* Suppose there exists $0<\delta'<\delta$ such that ζ is nondecreasing on $(c-\delta', c)$, hence $\zeta(x)\uparrow 1$ for $x\uparrow c$. There exists $0<\delta''<\delta'$ such that $\zeta^2(y)>A(\psi)$ on $[c-\delta'', c)$. Choose $\varepsilon>0$ such that $(1-\varepsilon)/B(\psi)>1/B_c$. There exists z in $[c-\delta'', c)$ such that $\psi'(z)\leqq(1-\varepsilon)\chi_c'(z)$, or else $\zeta(x)<1-\varepsilon$ on $[c-\delta'', c)$, a contradiction. Finally, $\Xi(\psi, F, z)>\kappa_c$. *Case B.2.* Suppose that for every $0<\delta'<\delta$ the function ζ is not nondecreasing on $(c-\delta', c)$. By some elementary analysis, a sequence $(c-\delta)<x_n\uparrow c$ can be constructed such that $\zeta(x_n)\uparrow 1$ and $\zeta'(x_n)=0$ for all n. This implies $\psi'(x_n)=\zeta(x_n)\chi_c'(x_n)$, so $\lim_{n\to\infty}\psi'(x_n) =\chi_c'(c-)<0$. Moreover, $\lim_{n\to\infty}\psi^2(x_n)/A(\psi)=0=\lim_{n\to\infty}\chi_c^2(x_n)$. Finally, $\kappa^*(\psi)\geqq\lim_{n\to\infty}\Xi(\psi, F, x_n)>\kappa_c$.

The uniqueness part is proven as in Theorem 2. □

Lemma 3. *For each $k>\kappa_c$ there exist A, B and p such that*

$$\begin{aligned} \chi_{c,k}(x) &= A(x) & 0\leqq|x|\leqq p \\ &=(A(k-1))^{1/2}\tanh[\tfrac{1}{2}((k-1)B^2/A)^{1/2}(c-|x|)]\operatorname{sign}(x) & p\leqq|x|\leqq c \\ &=0 & c\leqq|x| \end{aligned}$$

where $0<p<c$ satisfies

$$A(p)=(A(k-1))^{1/2}\tanh[\tfrac{1}{2}((k-1)B^2/A)^{1/2}(c-p)],$$

belongs to Ψ_c and satisfies $A(\chi_{c,k})=A$, $B(\chi_{c,k})=B$ and $\kappa^*(\chi_{c,k})=k$. Moreover, $0<A<B<J(c)<I(F)<\infty$ and $V(\tilde{\psi}_c, F)<V(\chi_{c,k}, F)<V(\chi_c, F)$.

Proof. This generalizes Theorem 3.1 of Hampel, Rousseeuw and Ronchetti (1981). The inequality $V(\tilde{\psi}_c, F)<V(\chi_{c,k}, F)$ follows from $B^2<J(c)A$ (Cauchy-Schwarz) and $V(\tilde{\psi}_c, F)=J(c)^{-1}$. □

The estimators determined by these $\chi_{c,k}$ are called *tanh-estimators. A* discussion of their properties and a table of values of A, B and p at $F=\Phi$ can be found in Hampel, Rousseeuw and Ronchetti (1981). Uniqueness of A, B and p follows from Corollary 4.

Corollary 4. *The only optimal V-robust estimators in Ψ_c are (up to equivalence) given by $\{\chi_c, \chi_{c,k}(k>\kappa_c)\}$.*

Proof. This follows from Theorem 4, combined with a straightforward generalization of Theorem 4.1 of Hampel, Rousseeuw and Ronchetti (1981). Note that the upper endpoint $\kappa^*(\tilde{\psi}_c)$ is always infinite. □

Remarks. This last situation completes the picture. Corollaries 1 and 2 depended on the upper extremities $\gamma^*(\Lambda)$ and $\kappa^*(\Lambda)$ being finite or infinite. In Corollary 3 we saw that $\gamma^*(\tilde{\psi}_c)$ was always finite regardless of F, so $\tilde{\psi}_c$ is itself optimal B-robust. In the last corollary $\kappa^*(\tilde{\psi}_c)$ is always infinite, so $\tilde{\psi}_c$ is never V-robust. We further remark that optimal B-robustness and optimal V-robustness lead in Ψ_c to markedly different solutions, because the ψ-functions of Corollary 3 have abrupt downward jumps at c and $-c$, whereas those of Corollary 4 all redescend in a smooth way.

In the case where also a scale parameter σ occurs, it is possible to estimate σ using the median deviation (Hampel 1974), given by $\mathrm{MAD}_n = \mathrm{median}\{|X_i - \mathrm{median}\{X_j\}|\}$ (usually multiplied with a tuning constant). Then one can estimate the location parameter θ by means of $\sum_{i=1}^{n} \psi((X_i - T_n)/\mathrm{MAD}_n) = 0$.

The sensitivities γ^* and κ^* are local concepts, which may be compared with global ones. In fact, for M-estimators there is a relation between B-robustness ($\gamma^* < \infty$) and a nonzero breakdown point (Huber 1981, p. 54 and p. 114, Donoho and Huber 1982, Sect. 1.2). Moreover, there also seems to be a connection between V-robustness ($\kappa^* < \infty$) and a nonzero variance breakdown point (Donoho and Huber 1982, Sec. 4.2).

Acknowledgement. This paper summarizes the author's Ph.D. dissertation, which was written under the generous guidance of F. Hampel. The research was supported by the NFWO Belgium.

References

1. Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J.W.: Robust estimates of location: survey and advances. Princeton: University Press 1972
2. Collins, J.R.: Robust estimation of a location parameter in the presence of asymmetry. Ann. Statist. **4**, 68–85 (1976)
3. Collins, J.R.: Upper bounds on asymptotic variances of M-estimators of location. Ann. Statist. **5**, 646–657 (1977)
4. Collins, J.R., Portnoy, S.L.: Maximizing the variance of M-estimators using the generalized method of moment spaces. Ann. Statist. **9**, 567–577 (1981)
5. Donoho, D.L., Huber, P.J.: The notion of breakdown point. To appear (1982)
6. Hampel, F.R.: The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. **69**, 383–393 (1974)
7. Hampel, F.R., Rousseeuw, P.J., Ronchetti, E.: The change-of-variance curve and optimal redescending M-estimators. J. Amer. Statist. Assoc. **76**, 643–648 (1981)
8. Huber, P.J.: Robust estimation of a location parameter. Ann. Math. Statist. **35**, 73–101 (1964)
9. Huber, P.J.: The behaviour of maximum likelihood estimates under nonstandard conditions. Proc. Fifth Berkeley Sympos. Math. Statist. Probab. **1**, 221–233 Univ. California Press 1967
10. Huber, P.J.: Robust Statistics. Wiley, New York 1981
11. Rousseeuw, P.J.: A new infinitesimal approach to robust estimation. Z. Wahrscheinlichkeitstheorie verw. Gebiete **56**, 127–132 (1981)
12. Stigler, S.M.: Studies in the history of probability and statistics XXXVIII: R.H. Smith, a Victorian interested in robustness. Biometrika **67**, 217–221 (1980)

Received December 23, 1981; in final form July 30, 1982