Breakdown points of Cauchy regression-scale estimators

Ivan Mizera
University of Alberta[1]

and

Christine H. Müller
Carl von Ossietzky University of Oldenburg[2]

Abstract. The lower bounds for the explosion and implosion breakdown points of the simultaneous Cauchy M-estimator (Cauchy MLE) of the regression and scale parameters are derived. For appropriate tuning constants, the breakdown point attains the maximum possible value.

Key Words. Breakdown point, explosion, implosion, Cauchy maximum likelihood, M-estimator, simultaneous estimation of regression and scale parameters.

[1]Corresponding author. Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, T6G 2G1, Canada. Email: mizera@stat.ualberta.ca. This work was supported in part by the National Sciences and Engineering Research Council of Canada.

[2]Department of Mathematics, Carl von Ossietzky University of Oldenburg, Postfach 2503, D-26111 Oldenburg, Germany. Email: mueller@math.uni-oldenburg.de.

1. Introduction

Consider a general linear model

$$y = X\beta + \epsilon$$

where $y = (y_1, y_2, \dots, y_N)^\top \in I\!R^N$ is a vector of observations, $\beta \in I\!R^p$ is an unknown parameter vector, $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_N)^\top \in I\!R^N$ a vector of errors, and $X = (x_1, x_2, \dots, x_N)^\top \in I\!R^{N\times p}$ represents the design, the known matrix of regressors—design points.

A well-known robustness measure for studying the behavior of an estimator in the presence of outliers is the finite sample breakdown point, the minimum proportion of outliers which may cause an arbitrary high bias of the estimator; see Huber 1981, Hampel et al. 1986, Rousseeuw and Leroy 1987. The definition that prevails in the literature reflect situations when regressors are prone to outliers—for instance, random. Employing this definition, Maronna et al. (1979) determined that M-estimators have the smallest possible breakdown point $1/N$, a poor behavior exhibited even by estimators generally considered robust—for instance, by the ℓ_1 estimator.

However, as indicated by He et al. (1990) and Ellis and Morgenthaler (1992), this behavior considerably changes under a definition of breakdown point that considers regressors error-free, non-stochastic. Such an approach suites particularly models with qualitative factors like ANOVA, or with regressors fully under experimental control. Let $\mathcal{N}(X) := \max_{\beta\neq 0} \#\{n : x_n^\top\beta = 0\}$, the symbol # hereafter standing for the number of elements of a set. The maximal possible breakdown point within regression equivariant estimators is

$$\frac{1}{N}\left\lfloor \frac{N - \mathcal{N}(X) + 1}{2} \right\rfloor; \tag{1.1}$$

see Müller (1995, 1997). Mizera and Müller (1999) pointed out that this upper bound is attained by M-estimators whose score functions has variation exponent zero, the class containing maximum likelihood estimators (MLE) based on t-distribution with γ degrees of freedom, in particular MLE based on Cauchy ($\gamma = 1$) distribution (hereafter Cauchy estimator).

The results of Mizera and Müller (1999) were proved under the simplifying assumption that the scale is known and equal to one. In applications, however, redescending estimators are very sensitive to the scale factor; if it is too large, their behavior may be fairly non-robust. Therefore, it is of considerable interest whether the same breakdown bounds apply also for the more general setting with unknown scale. In this note, we answer this question affirmatively for the Cauchy estimator. In the context of breakdown point considering outliers in regressors, He et al. (2000) found that the Cauchy estimator has the optimal performance among the MLE based on t-distribution.

In Section 2, we define the Cauchy simultaneous estimator of regression and scale; a choice of a tuning constant for the scale leads to several possible versions. In Section 3, we show that the breakdown point of all versions is reasonable, and that there is a version attaining the upper bound (1.1). We derive the explosion and implosion breakdown point of scale estimators separately; the overall breakdown point is taken to be their minimum. Finally, a short conclusion is given in Section 4.

2. Cauchy regression-scale estimators

The maximization of the likelihood for regression and scale parameters for the errors with the Cauchy distribution leads to the minimization of

$$l(\beta,\sigma,y,X) := \sum_{n=1}^{N} \log\left(1+\left(\frac{y_n - x_n^\top \beta}{\sigma}\right)^2\right) + N\,\log\sigma$$

with respect to $(\beta,\sigma)\in\Theta := I\!R^p \times (0,\infty)$. Let $\varphi(u) := \log(1+u^2)$, and for $K>0$, let

$$D_K(\beta,\sigma,y,X) := \sum_{n=1}^{N} \varphi\left(\frac{y_n - x_n^\top \beta}{\sigma}\right) + K\,\log\sigma.$$

If $K=N$, then $D_K(\beta,\sigma,y,X)$ is equal to $l(\beta,\sigma,y,X)$. If $K\neq N$, we call the likelihood function tuned; its minimization leads to the Cauchy simultaneous regression and scale estimators

$$(\widehat{\beta},\widehat{\sigma}) \in \arg\min\{D_K(\beta,\sigma,y,X):\ (\beta,\sigma)\in\Theta\}. \tag{2.1}$$

where $\Theta = I\!R^p\times(0,\infty)$. A solution of (2.1) in Θ satisfies the system of equations

$$\sum_{n=1}^{N} \psi\left(\frac{y_n - x_n^\top \widehat{\beta}}{\widehat{\sigma}}\right) x_n = 0, \tag{2.2}$$

$$\sum_{n=1}^{N} \chi\left(\frac{y_n - x_n^\top \widehat{\beta}}{\widehat{\sigma}}\right) = \frac{K}{2}, \tag{2.3}$$

where $\psi(u) := \frac{u}{1+u^2}$ and $\chi(u) := \frac{u^2}{1+u^2}$. Since $\chi\left(\frac{u}{\sigma}\right) = 0$ for all $\sigma>0$ if $u=0$, and $\chi\left(\frac{u}{\sigma}\right)\to 0$ if $\sigma\to\infty$ and $u\neq 0$, we have that

$$\sum_{n=1}^{N} \chi\left(\frac{y_n - x_n^\top \beta}{\sigma}\right) \to 0 \qquad \text{if } \sigma\to\infty. \tag{2.4}$$

On the other hand, $\chi\left(\frac{u}{\sigma}\right) \to 1$ if $\sigma \to 0$ and $u \neq 0$; since $\chi\left(\frac{u}{\sigma}\right)$ is decreasing in σ for $u \neq 0$, we have

$$\sum_{n=1}^{N} \chi\left(\frac{y_n - x_n^\top \beta}{\sigma}\right) \to \#\{n: \ y_n \neq x_n^\top \beta\} \qquad \text{if } \sigma \to 0. \tag{2.5}$$

By (2.4) and (2.5), there is no solution of (2.1) in Θ, if

$$\mathcal{N}(\beta, y, X) := \#\{n: \ y_n = x_n^\top \beta\} \geq N - \frac{K}{2}.$$

Let

$$\mathcal{N}(y, X) := \max_{\beta} \mathcal{N}(\beta, y, X).$$

Note that $\mathcal{N}(X) = \max_{\beta \neq 0} \mathcal{N}(\beta, 0, X)$.

Theorem 1. *Suppose that $\mathcal{N}(X) < N - \frac{K}{2}$.*
(i) If $\mathcal{N}(y, X) < N - \frac{K}{2}$, then there is a compact $\Theta_0 \subset \Theta$ and $(\beta_0, \sigma_0) \in \Theta_0$ such that

$$D_K(\beta_0, \sigma_0, y, X) < D_K(\beta, \sigma, y, X)$$

for all $(\beta, \sigma) \in \Theta \setminus \Theta_0$.
(ii) If there exists β_0 such that $\mathcal{N}(\beta_0, y, X) \geq N - \frac{K}{2}$, then

$$\lim_{\sigma \to 0} D_K(\beta_0, \sigma, y, X) < D_K(\beta_0, \sigma, y, X)$$

for all $\sigma \in (0, \infty)$.

Proof. **(i)** Take $\beta_0 = 0$ and $\sigma_0 = 1$. Assume that for every $k \in \mathbb{N}$ there exists (β_k, σ_k) with $D_K(\beta_0, \sigma_0, y, X) \geq D_K(\beta_k, \sigma_k, y, X)$ such that $\|(\beta_k, \sigma_k) - (\beta_0, \sigma_0)\| > k$ or $\frac{1}{\sigma_k} > k$. There are three possible cases:

(a) $(\beta_k)_{k \in \mathbb{N}}$ is bounded and $(\sigma_k)_{k \in \mathbb{N}}$ contains a subsequence convergent to zero—thus, we may w.l.o.g. assume that $\sigma_k \to 0$.
(b) $(\sigma_k)_{k \in \mathbb{N}}$ is bounded and $(\|\beta_k\|)_{k \in \mathbb{N}}$ contains a subsequence tending to infinity—thus, we may w.l.o.g. assume that $\|\beta_k\| \to \infty$.
(c) A subsequence of $(\sigma_k)_{k \in \mathbb{N}}$ converges to ∞; w.l.o.g. we may assume that $\sigma_k \to \infty$.

We will show that $D_K(\beta_k, \sigma_k, y, X)$ is unbounded in all three cases—a contradiction proving (i).
Case (a). We may assume w.l.o.g. that $\beta_k \to \beta$. Then

$$\sum_{n=1}^{N} \log\left(1 + \left(\frac{y_n - x_n^\top \beta_k}{\sigma_k}\right)^2\right) + K \log \sigma_k$$

$$
\begin{aligned}
\geq \quad & \sum_{n: y_n \neq x_n^\top \beta} \log\left(1 + \left(\frac{y_n - x_n^\top \beta_k}{\sigma_k}\right)^2\right) + K \log \sigma_k \\
= \quad & \sum_{n: y_n \neq x_n^\top \beta} \log\left(\frac{1}{\sigma_k^2}\left(\sigma_k^2 + (y_n - x_n^\top \beta_k)^2\right)\right) + K \log \sigma_k \\
= \quad & \sum_{n: y_n \neq x_n^\top \beta} \log\left(\sigma_k^2 + (y_n - x_n^\top \beta_k)^2\right) - 2\,\#\{n:\ y_n \neq x_n^\top \beta\} \log \sigma_k + K \log \sigma_k \\
\longrightarrow \quad & \infty
\end{aligned}
$$

because $\sum_{n: y_n \neq x_n^\top \beta} \log\left(\sigma_k^2 + (y_n - x_n^\top \beta_k)^2\right) > -\infty$ if $\sigma_k \to 0$ and $\beta_k \to \beta$, $\log \sigma_k \to -\infty$ if $\sigma_k \to 0$, and

$$
\begin{aligned}
-2\,\#\{n:\ y_n \neq x_n^\top \beta\} + K &= K - 2N + 2\,\#\{n:\ y_n = x_n^\top \beta\} \\
&\leq K - 2N + 2\mathcal{N}(y, X) < 0.
\end{aligned}
$$

Case (b). W.l.o.g. we may assume $\beta_k / \|\beta_k\| \to \beta$. Then $\sigma_k^2 / \|\beta_k\|^2 \to 0$, $y_n / \|\beta_k\| \to 0$, $x_n^\top \beta_k / \|\beta_k\| \to x_n^\top \beta$. Since σ_k is bounded and

$$
K - 2\,\#\{n:\ x_n^\top \beta \neq 0\} \leq K - 2N + 2\mathcal{N}(X) < 0,
$$

we obtain

$$
(K - 2\,\#\{n:\ x_n^\top \beta \neq 0\}) \log \sigma_k > -\infty.
$$

This implies

$$
\begin{aligned}
& \sum_{n=1}^{N} \log\left(1 + \left(\frac{y_n - x_n^\top \beta_k}{\sigma_k}\right)^2\right) + K \log \sigma_k \\
& \quad \geq \quad \sum_{n: x_n^\top \beta \neq 0} \log\left(1 + \left(\frac{y_n - x_n^\top \beta_k}{\sigma_k}\right)^2\right) + K \log \sigma_k \\
& \quad = \quad \sum_{n: x_n^\top \beta \neq 0} \log\left(\frac{\|\beta_k\|^2}{\sigma_k^2}\left(\frac{\sigma_k^2}{\|\beta_k\|^2} + \left(\frac{y_n}{\|\beta_k\|} - x_n^\top \frac{\beta_k}{\|\beta_k\|}\right)^2\right)\right) + K \log \sigma_k \\
& \quad = \quad \sum_{n: x_n^\top \beta \neq 0} \log\left(\frac{\sigma_k^2}{\|\beta_k\|^2} + \left(\frac{y_n}{\|\beta_k\|} - x_n^\top \frac{\beta_k}{\|\beta_k\|}\right)^2\right) \\
& \qquad\quad + \#\{n : x_n^\top \beta \neq 0\} \log \|\beta_k\|^2 + (K - 2\,\#\{n : x_n^\top \beta \neq 0\}) \log \sigma_k \\
& \quad \longrightarrow \quad \infty.
\end{aligned}
$$

Case (c). If $\sigma_k \to \infty$, then

$$\sum_{n=1}^{N} \log\left(1 + \left(\frac{y_n - x_n^\top \beta_k}{\sigma_k}\right)^2\right) + K\,\log\sigma_k \geq K\,\log\sigma_k \longrightarrow \infty.$$

(ii) For every $\sigma \in (0,\infty)$,

$$\begin{aligned} D_K(\beta_0,\sigma,y,X) &= \sum_{n=1}^{N} \log\left(1 + \left(\frac{y_n - x_n^\top \beta_0}{\sigma}\right)^2\right) + K\,\log\sigma \\ &= \sum_{n:\ y_n \neq x_n^\top \beta_0} \log\left(\frac{1}{\sigma^2}\left(\sigma^2 + (y_n - x_n^\top\beta_0)^2\right)\right) + K\,\log\sigma \\ &= \sum_{n:\ y_n \neq x_n^\top \beta_0} \log\left(\sigma^2 + (y_n - x_n^\top\beta_0)^2\right) + \left(K - 2\,\#\{n:\ y_n \neq x_n^\top\beta_0\}\right)\log\sigma. \end{aligned}$$

The assumption $\mathcal{N}(\beta_0,y,X) \geq N - \frac{K}{2}$ yields

$$K - 2\,\#\{n:\ y_n \neq x_n^\top\beta_0\} = K - 2\,N + 2\,\mathcal{N}(\beta_0,y,X) \geq 0.$$

If $K - 2\,N + 2\,\mathcal{N}(\beta_0,y,X) > 0$ (that is, K is odd), then $\lim_{\sigma\to 0} D_K(\beta_0,\sigma,y,X) = -\infty$. If $K - 2\,N + 2\,\mathcal{N}(\beta_0,y,X) = 0$, then $D_K(\beta_0,\sigma,y,X) = \sum_{n:\ y_n\neq x_n^\top\beta_0} \log\left(\sigma^2 + (y_n - x_n^\top\beta_0)^2\right)$ and (ii) holds with a finite limit. □

In view of Theorem 1, we may formulate the following definition. Let $\overline{\Theta} = I\!R^p \times [0,\infty)$.

Definition 1. A *Cauchy regression-scale estimate with tuning constant* K is, for given y and X, any point (β_0,σ_0) in $\overline{\Theta}$ such that

$$\begin{aligned} (\beta_0,\sigma_0) &\in \arg\min_{(\beta,\sigma)\in\Theta} D_K(\beta,\sigma,y,X) && \text{if } \mathcal{N}(\beta,y,X) < N - \frac{K}{2} \text{ for all } \beta \in I\!R^p; \\ &= (\beta_0, 0), && \text{if } \mathcal{N}(\beta_0,y,X) \geq N - \frac{K}{2}. \end{aligned}$$

For Cauchy regression-scale estimators, we use the notation $C(y,X) = (\widehat{\beta}(y,X), \widehat{\sigma}(y,X))$, suppressing the dependence on K.

3. Breakdown points

There are two types of breakdown of a regression-scale estimator caused by outliers (see Hampel et al. 1986, Rousseeuw and Croux 1992, 1993): explosion, when the estimator for the regression parameters or the scale estimator goes to infinity, or implosion, which means that the scale estimator goes to zero. Assuming that the regressors are error-free and non-stochastic, we allow only for outliers in the observations y_n, in accord with He et al. (1990), Ellis and Morgenthaler (1992), and Müller (1995, 1997).

Definition 2 (Breakdown point)**.** Let $C = (\widehat{\beta}, \widehat{\sigma})$ be an estimator for (β, σ). The explosion breakdown point $\epsilon^\infty(C, y, X)$ of C is defined by

$$\epsilon^\infty(C, y, X) := \frac{1}{N} \min \left\{ M : \sup_{\overline{y} \in B(y,M)} \|C(\overline{y}, X)\| = \infty \right\};$$

the implosion breakdown point $\epsilon^0(C, y, X)$ of C is defined by

$$\epsilon^0(C, y, X) := \frac{1}{N} \min \left\{ M : \inf_{\overline{y} \in B(y,M)} \widehat{\sigma}(\overline{y}, X) = 0 \right\},$$

where

$$B(y, M) := \left\{ \overline{y} \in I\!R^N : \#\{n : \overline{y}_n \neq y_n\} \leq M \right\}.$$

The breakdown point $\epsilon^*(C, y, X)$ of C is defined to be

$$\epsilon^*(C, y, X) = \min\{\epsilon^\infty(C, y, X), \epsilon^0(C, y, X)\}.$$

If the estimator is not unique, then we adopt the least favorable values—those maximizing $\|C(\overline{y}, X)\|$ and minimizing $\widehat{\sigma}(y, X)$.)

Proposition 2 (Explosion without implosion)**.** *Suppose there is a sequence $y^k \in B(y, M)$ such that the Cauchy estimates $C(y^k, X) = (\widehat{\beta}(y^k, X), \widehat{\sigma}(y^k, X))$ satisfy $\|C(y^k, X)\| \to \infty$ and $\frac{1}{\widehat{\sigma}(y^k, X)} = O(1)$. Then $M \geq \min\{\frac{K}{2}, N - \frac{K}{2} - \mathcal{N}(X)\}$.*

Proof. Assume $y_n = y_n^k$ for $n = 1, \ldots, N - M$ for all $k \in I\!N$, and let $\beta_k := \widehat{\beta}(y^k, X)$, $\sigma_k := \widehat{\sigma}(y^k, X)$. Then one of the following cases takes place:
(a) $\sigma_k \to \infty$ and $\|\beta_k\| = O(1)$.
(b) $\sigma_k \to \infty$ and $\sigma_k / \|\beta_k\| \to \infty$.
(c) $\sigma_k \to \infty$, $\sigma_k / \|\beta_k\| = O(1)$, and $\|\beta_k\| / \sigma_k = O(1)$.
(d) $\sigma_k \to \infty$ and $\|\beta_k\| / \sigma_k \to \infty$.
(e) $\|\beta_k\| \to \infty$, $\sigma_k = O(1)$, and $1/\sigma_k = O(1)$.
Cases (a) and (b). If $\sigma_k \to \infty$, then $\mathcal{N}(y^k, X) < N - \frac{K}{2}$ for almost all $k \in I\!N$ so that (β_k, σ_k) is a solution of (2.1), lying in Θ. Since it satisfies (2.3),

$$\frac{K}{2} = \sum_{n=1}^{N-M} \frac{\left(\frac{y_n - x_n^\top \beta_k}{\sigma_k}\right)^2}{1 + \left(\frac{y_n - x_n^\top \beta_k}{\sigma_k}\right)^2} + \sum_{n=N-M+1}^{N} \chi\left(\frac{y_n^k - x_n^\top \beta_k}{\sigma_k}\right)$$

$$\leq \sum_{n=1}^{N-M} \frac{(y_n - x_n^\top \beta_k)^2}{\sigma_k^2 + (y_n - x_n^\top \beta_k)^2} + M.$$

In Case (a) we have

$$\sum_{n=1}^{N-M} \frac{(y_n - x_n^\top \beta_k)^2}{\sigma_k^2 + (y_n - x_n^\top \beta_k)^2} = \sum_{n=1}^{N-M} \frac{1}{\frac{\sigma_k^2}{(y_n - x_n^\top \beta_k)^2} + 1} \to 0$$

because $(y_n - x_n^\top \beta_k)^2$ is bounded for all $n = 1, \ldots, N-M$. It follows that $M \geq \frac{K}{2}$. In Case (b) we may assume w.l.o.g. that $\beta_k / \|\beta_k\| \to \beta_0$. Since $y_n / \|\beta_k\| \to 0$, we have

$$\sum_{n=1}^{N-M} \frac{(y_n - x_n^\top \beta_k)^2}{\sigma_k^2 + (y_n - x_n^\top \beta_k)^2} = \sum_{n=1}^{N-M} \frac{\left(\frac{y_n}{\|\beta_k\|} - x_n^\top \frac{\beta_k}{\|\beta_k\|}\right)^2}{\left(\frac{\sigma_k}{\|\beta_k\|}\right)^2 + \left(\frac{y_n}{\|\beta_k\|} - x_n^\top \frac{\beta_k}{\|\beta_k\|}\right)^2} \to 0,$$

yielding $M \geq \frac{K}{2}$ in Case (b) too.

Case (c): Again, σ_k is away from 0, so that (β_k, σ_k) are solutions of (2.1), lying in Θ. Since φ is symmetric and subadditive, there exists $L > 0$ such that $\varphi(s) - \varphi(t) - L \leq \varphi(t - s)$ and $-\varphi(s) - L \leq \varphi(t - s) - \varphi(t)$ for all $s, t \in I\!R$; see Mizera and Müller (1999). This property implies that

$$\begin{aligned}
0 &\geq D_K(\beta_k, \sigma_k, y^k, X) - D_K(0, 1, y^k, X) \\
&= D_K(\beta_k, \sigma_k, y^k, X) - D_K(0, \sigma_k, y^k, X) + D_K(0, \sigma_k, y^k, X) - D_K(0, 1, y^k, X) \\
&= \sum_{n=1}^{N-M} \varphi\left(\frac{y_n - x_n^\top \beta_k}{\sigma_k}\right) + \sum_{n=N-M+1}^{N} \varphi\left(\frac{y_n^k - x_n^\top \beta_k}{\sigma_k}\right) + K \log \sigma_k \\
&\quad - \sum_{n=1}^{N-M} \varphi\left(\frac{y_n}{\sigma_k}\right) - \sum_{n=N-M+1}^{N} \varphi\left(\frac{y_n^k}{\sigma_k}\right) - K \log \sigma_k \\
&\quad + \sum_{n=1}^{N-M} \varphi\left(\frac{y_n}{\sigma_k}\right) + \sum_{n=N-M+1}^{N} \varphi\left(\frac{y_n^k}{\sigma_k}\right) + K \log \sigma_k \\
&\quad - \sum_{n=1}^{N-M} \varphi(y_n) - \sum_{n=N-M+1}^{N} \varphi(y_n^k) \\
&\geq \sum_{n=1}^{N-M} \varphi\left(\frac{x_n^\top \beta_k}{\sigma_k}\right) - 2 \sum_{n=1}^{N-M} \varphi\left(\frac{y_n}{\sigma_k}\right) - \sum_{n=N-M+1}^{N} \varphi\left(\frac{x_n^\top \beta_k}{\sigma_k}\right) - (N-M)\, L - M\, L \\
&\quad + \sum_{n=1}^{N-M} \varphi\left(\frac{y_n}{\sigma_k}\right) - \sum_{n=1}^{N-M} \varphi(y_n)
\end{aligned}$$

$$+ \sum_{n=N-M+1}^{N} \varphi\left(\frac{y_n^k}{\sigma_k}\right) - \sum_{n=N-M+1}^{N} \varphi(y_n^k) + K\,\log \sigma_k.$$

Setting

$$A_k := \sum_{n=1}^{N-M} \varphi\left(\frac{x_n^\top \beta_k}{\sigma_k}\right) - \sum_{n=N-M+1}^{N} \varphi\left(\frac{x_n^\top \beta_k}{\sigma_k}\right) - N\,L - \sum_{n=1}^{N-M} \varphi\left(\frac{y_n}{\sigma_k}\right) - \sum_{n=1}^{N-M} \varphi(y_n),$$

we have

$$\begin{aligned} 0 &\geq A_k + \sum_{n=N-M+1}^{N} \varphi\left(\frac{y_n^k}{\sigma_k}\right) - \sum_{n=N-M+1}^{N} \varphi(y_n^k) + K\,\log \sigma_k \\ &= A_k + \sum_{n=N-M+1}^{N} \log\left(1 + \left(\frac{y_n^k}{\sigma_k}\right)^2\right) - \sum_{n=N-M+1}^{N} \log\left(1 + (y_n^k)^2\right) \\ &\quad +2\,M\,\log \sigma_k + (K - 2M)\,\log \sigma_k \\ &= A_k + \sum_{n=N-M+1}^{N} \log\left(\frac{\left(1 + \left(\frac{y_n^k}{\sigma_k}\right)^2\right)\sigma_k^2}{1 + (y_n^k)^2}\right) + (K - 2M)\,\log \sigma_k \\ &= A_k + \sum_{n=N-M+1}^{N} \log\left(\frac{\sigma_k^2 + (y_n^k)^2}{1 + (y_n^k)^2}\right) + (K - 2M)\,\log \sigma_k. \end{aligned}$$

Using the fact that $\sigma_k^2 \geq 1$ for large k, we obtain

$$0 \geq A_k + (K - 2M)\,\log \sigma_k;$$

dividing by $\log \sigma_k$ gives

$$0 \geq \frac{1}{\log \sigma_k}\,A_k + (K - 2M). \tag{3.1}$$

We may assume w.l.o.g. that $\beta_k/\|\beta_k\| \to \beta_0$ and $\sigma_k/\|\beta_k\| \to c$. Then together with $y_n/\sigma_k \to 0$ we obtain that

$$A_k = \sum_{n=1}^{N-M} \varphi\left(\frac{x_n^\top \frac{\beta_k}{\|\beta_k\|}}{\frac{\sigma_k}{\|\beta_k\|}}\right) - \sum_{n=N-M+1}^{N} \varphi\left(\frac{x_n^\top \frac{\beta_k}{\|\beta\|}}{\frac{\sigma_k}{\|\beta_k\|}}\right) - N\,L - \sum_{n=1}^{N-M} \varphi\left(\frac{y_n}{\sigma_k}\right) - \sum_{n=1}^{N-M} \varphi(y_n)$$

is bounded. Hence $A_k/\log \sigma_k$ converges to 0 so that the limit of (3.1) gives $M \geq K/2$.

Cases (d) and (e). Since σ_k is greater than 0, we can use (2.3). Assuming w.l.o.g. that

$\beta_k/\|\beta_k\| \to \beta_0$, we obtain with $y_n/\|\beta_k\| \to 0$

$$\frac{K}{2} = \sum_{n=1}^{N-M} \frac{(y_n - x_n^\top \beta_k)^2}{\sigma_k^2 + (y_n - x_n^\top \beta_k)^2} + \sum_{n=N-M+1}^{N} \chi\left(\frac{y_n^k - x_n^\top \beta_k}{\sigma_k}\right)$$
$$\geq \sum_{n=1}^{N-M} \frac{\left(\frac{y_n}{\|\beta_k\|} - x_n^\top \frac{\beta_k}{\|\beta_k\|}\right)^2}{\frac{\sigma_k^2}{\|\beta_k\|^2} + \left(\frac{y_n}{\|\beta_k\|} - x_n^\top \frac{\beta_k}{\|\beta\|}\right)^2} \longrightarrow N - M - \mathcal{N}(X).$$

This implies $M \geq N - \frac{K}{2} - \mathcal{N}(X)$. □

Proposition 3 (Explosion and implosion)**.** *Suppose there is a sequence* $y^k \in B(y, M)$ *such that the corresponding Cauchy estimates* $C(y^k, X) = (\widehat{\beta}(y^k, X), \widehat{\sigma}(y^k, X))$ *satisfy* $\|\widehat{\beta}(y^k, X)\| \to \infty$ *and* $\widehat{\sigma}(y^k, X) \to 0$*. Then* $M \geq N - \frac{K}{2} - \mathcal{N}(X)$*.*

Proof. Again, assume $y_n = y_n^k$ for $n = 1, \ldots, N-M$ for all $k \in I\!N$, and let $\beta_k := \widehat{\beta}(y^k, X)$, $\sigma_k := \widehat{\sigma}(y^k, X)$. There are two possibilities. We have $\sigma_k > 0$ for almost all $k \in I\!N$; then the assertion follows as for the Cases (d) and (e) in the proof of Proposition 2. Or, we have $\sigma_k = 0$ for almost $k \in I\!N$. In the latter case, β_k satisfies

$$\#\{n : \; y_n^k = x_n^\top \beta_k\} = \mathcal{N}(\beta_k, y^k, X) \geq N - \frac{K}{2}.$$

Since $\|\beta_k\| \to \infty$ the equality $y_n = y_n^k = x_n^\top \beta_k$ can be satisfied for $n \in \{1, \ldots, N-M\}$ only if $y_n = 0 = x_n^\top \beta_k$, that means it is satisfied for at most $\mathcal{N}(X)$ elements of $\{1, \ldots, N-M\}$. Hence

$$M + \mathcal{N}(X) \geq \mathcal{N}(\beta_k, y^k, X) \geq N - \frac{K}{2},$$

so that $M \geq N - \frac{K}{2} - \mathcal{N}(X)$ follows. □

Theorem 4 (Explosion breakdown point)**.** *The explosion breakdown point* $\epsilon^\infty(C, y, X)$ *of the Cauchy regression-scale estimator satisfies*

$$\epsilon^\infty(C, y, X) \geq \frac{1}{N} \min\left\{\frac{K}{2}\;,\; N - \frac{K}{2} - \mathcal{N}(X)\right\}.$$

The lower bound is

$$\begin{aligned} &\frac{1}{2} - \frac{\mathcal{N}(X)}{N}, && \text{if } K = N; \\ &\frac{1}{2} - \frac{\mathcal{N}(X)}{2N}, && \text{if } K = N - \mathcal{N}(X). \end{aligned}$$

Proof. Let $(y^k)_{k\in I\!N}$ any sequence in $B(y,M)$ with $\|C(y^k,X)\| = \|(\widehat{\beta}(y^k,X),\widehat{\sigma}(y^k,X))\| \to \infty$. Then we have explosion with or without implosion so that the assertion follows from Proposition 2 and 3. □

The Cauchy regression-scale estimator is regression equivariant: $\widehat{\beta}(y+X\beta,X) = \widehat{\beta}(y,X)+\beta$ for all y, X and β. Hence, the upper bound (1.1) for regression equivariant estimators provides $\epsilon^\infty(C,y,X) \le \frac{1}{N}\left\lfloor \frac{N-\mathcal{N}(X)+1}{2}\right\rfloor$, where equality holds in the case $K = N-\mathcal{N}(X)$. Therefore, the Cauchy regression-scale estimator with tuning constant $K = N-\mathcal{N}(X)$ has the highest explosion breakdown point which is possible within regression equivariant estimators. However, the untuned Cauchy estimator ($K = N$) exhibits similar features if $\mathcal{N}(X)$ is small.

A similar result holds for the implosion breakdown point. Since the Cauchy scale estimator is allowed to be equal to 0, we investigate the implosion only for data points (y,X) with $\widehat{\sigma}(y,X) > 0$, when $\mathcal{N}(y,X) < N - \frac{K}{2}$.

Proposition 5 (Implosion without explosion). *Let $\widehat{\sigma}(y,X) > 0$. If there is a sequence $y^k \in B(y,M)$ such that $C(y^k,X) = (\widehat{\beta}(y^k,X),\widehat{\sigma}(y^k,X))$ satisfy $\|\widehat{\beta}(y^k,X)\| = O(1)$ and $\widehat{\sigma}(y^k,X) \to 0$, then $M \ge N - \frac{K}{2} - \mathcal{N}(y,X)$.*

Proof. There are two possibilities:
(a) For almost all $k \in I\!N$ we have $\widehat{\sigma}(y^k,X) = 0$: there exists β_k with $\mathcal{N}(\beta_k,y^k,X) \ge N-\frac{K}{2}$. This implies $M \ge \mathcal{N}(\beta_k,y^k,X) - \mathcal{N}(y,X) \ge N - \frac{K}{2} - \mathcal{N}(y,X)$.
(b) There is a subsequence with $\sigma_k := \widehat{\sigma}(y^k,X) > 0$. We may w.l.o.g. assume that $\beta_k := \widehat{\beta}(y^k,X) \to \beta_0$. According to (2.3), we obtain

$$\begin{aligned}
\frac{K}{2} &= \sum_{n=1}^{N-M} \frac{(y_n - x_n^\top\beta_k)^2}{\sigma_k^2 + (y_n - x_n^\top\beta_k)^2} + \sum_{n=N-M+1}^{N} \frac{(y_n^k - x_n^\top\beta_k)^2}{\sigma_k^2 + (y_n^k - x_n^\top\beta_k)^2} \\
&\ge \sum_{n=1}^{N-M} \frac{(y_n - x_n^\top\beta_k)^2}{\sigma_k^2 + (y_n - x_n^\top\beta_k)^2} \\
&\ge \sum_{n=1,\dots,N-M:\ y_n\neq x_n^\top\beta_0} \frac{(y_n - x_n^\top\beta_k)^2}{\sigma_k^2 + (y_n - x_n^\top\beta_k)^2} \\
&\longrightarrow \#\{n=1,\dots,N-M:\ y_n \neq x_n^\top\beta_0\} \\
&= N-M-\#\{n=1,\dots,N-M:\ y_n = x_n^\top\beta_0\} \\
&\ge N-M-\mathcal{N}(\beta_0,y,X) \ge N-M-\mathcal{N}(y,X),
\end{aligned}$$

so that $M \ge N - \frac{K}{2} - \mathcal{N}(y,X)$. □

Theorem 6 (Implosion breakdown point)**.** *If $\widehat{\sigma}(y,X) > 0$, then the implosion point $\epsilon^0(C,y,X)$ of the Cauchy regression-scale estimator satisfies*

$$\epsilon^0(C,y,X) \geq \frac{1}{N} \min\left\{ N - \frac{K}{2} - \mathcal{N}(y,X) \ , \ N - \frac{K}{2} - \mathcal{N}(X) \right\}.$$

Proof. The assertion follows from Proposition 3 and 5, for any sequence $(y^k)_{k\in I\!N}$ any sequence in $B(y,M)$ with $\widehat{\sigma}(y^k,X) \to 0$. □

Usually—with probability 1 for data sampled from continuous populations—we have that $\mathcal{N}(y,X) \leq \mathcal{N}(X)$. Theorems 4 and 6 then imply that the breakdown point of the Cauchy regression-scale estimator bounded from below by

$$\frac{1}{N} \min\left\{ \frac{K}{2} \ , \ N - \frac{K}{2} - \mathcal{N}(X), \right\}$$

and attains the maximal value when tuning constant $K = N - \mathcal{N}(X)$.

4. Conclusion

Our results imply, in particular, that the breakdown point of the simultaneous location and scale Cauchy estimator is 1/2. Copas (1975) showed that in this case the solution of the estimating equations is unique, except for the very special case when half of the observations is concentrated at one point and the another half at another point. Although Gabrielsen (1982) showed that this property in general does not hold in regression-scale setting, Lange et al. (1989) indicated that it is true for reasonable well-behaved datasets. Nevertheless, theoretical results that good breakdown behavior of redescending M-estimators requires their appropriate definition: while breakdown point of arbitrary roots may be poor, the breakdown point of the regression-scale Cauchy estimator defined via global minimization is high—when tuned optimally, maximal.

References

Copas, J.B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62**, 701-704.

Ellis, S. P. and Morgenthaler, S. (1992). Leverage and breakdown in L_1 regression. *J. Amer. Statist. Assoc.* **87**, 143-148.

Gabrielsen, G. (1982). On the unimodality of the likelihood for the Cauchy distribution: Some comments. *Biometrika* **69**, 677-678.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics—The Approach Based on Influence Functions.* Wiley, New York.

He, X., Jurečková, J., Koenker, R. and Portnoy, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* 58, 1195-1214.

He, X., Simpson, D.G. and Wang, G. (2000). Breakdown points of t-type regression estimators. *Biometrika* **87**, 675-687.

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

Huber, P. J. (1984). Finite sample breakdown of M- and P-estimators. *Ann. Statist.* **12**, 119-126.

Lange, K.L., Little, R.J.A. and Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *J. Amer. Statist. Assoc.* **84**, 881-896.

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,* eds. T. Gasser and M. Rosenblatt, Lecture Notes in Mathematics **757**, Springer, Berlin, 91-116.

Mizera, I. and Müller, Ch. H. (1999). Breakdown points and variation exponents of robust M-estimators in linear models. *Ann. Statist.* **27**, 1164-1177.

Müller, Ch. H. (1995). Breakdown points for designed experiments. *J. Statist. Plann. Inference.* **45**, 413–427.

Müller, Ch. H. (1997). *Robust Planning and Analysis of Experiments.* Lecture Notes in Statistics **124**, Springer, New York.

Rousseeuw, P.J. and Croux, C. (1992). Explicit scale estimators with high breakdown point. In *L_1-Statistical Analysis and Related Methods*, ed. Y. Dodge, North-Holland, Amsterdam, 77-92.

Rousseeuw, P.J. and Croux, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88**, 1273-1283.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.