Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models

By CHRISTINE H. MÜLLER
Carl von Ossietzky University, Department of Mathematics, Postfach 2503, D - 26111 Oldenburg, Germany
mueller@math.uni-oldenburg.de

and NEYKO NEYKOV
Bulgarian Academy of Sciences, Institute of Meteorology and Hydrology, 66 Tsarigradsko Chaussee, 1784 Sofia, Bulgaria
neyko.neykov@meteo.bg

March 2001

Summary

Lower bounds for breakdown points of trimmed likelihood (TL) estimators in a general setup are expressed by the fullness parameter of Vandev (1993), where results of Vandev and Neykov (1998) are extended. A special application of the general result are the breakdown points of TL estimators and related estimators as the S estimators in generalized linear models. For the generalized linear models, a connection between the fullness parameter and the quantity $\mathcal{N}(X)$ of Müller (1995) is derived for the case that the explanatory variables may be not in general position which happens in particular in designed experiments. These results are in particular applied to logistic regression and log-linear models where also upper bounds for the breakdown points are derived.

Key words: Breakdown point, d-fullness, trimmed likelihood estimator, S estimator, linear model, logistic regression, log-linear model, exponential model, scale estimation, designed experiments

AMS subject classification: Primary 62J12, secondary 62F35, 62G35

Running title: Breakdown points of trimmed likelihood estimators

1 Introduction

Assume that the distribution of an observation Y_n has the density $f_n(y_n, \theta)$ and that the observations $Y_1, \ldots, Y_N$ are independent. Let $y := (y_1, \ldots, y_N)^\top$ the vector of all realized observations, $l_n(y, \theta) := -\log f_n(y_n, \theta)$ the log-likelihood, and $l(y, \theta) := (l_1(y, \theta), \ldots, l_N(y, \theta))^\top$. Maximum likelihood (ML) estimators are maximizing the likelihood, i.e. minimzing $\sum_{n=1}^{N} l_n(y, \theta)$. Trimming the least likely observations, i.e. the observations with the largest $l_n(y, \theta)$, leads to trimmed likelihoods. Maximizing the trimmed likelihood provides the trimmed likelihood estimators $TL_h(y)$ given by

$$TL_h(y) := \arg\min_{\theta} \sum_{n=1}^{h} l_{(n)}(y, \theta),$$

where $N - h$ observations are trimmed and $l_{(1)}(y, \theta) \leq \ldots \leq l_{(N)}(y, \theta)$. These estimators can be also extended to weighted trimmed likelihood estimators WTL_h defined by

$$WTL_h(y) := \arg\min_{\theta} \sum_{n=1}^{h} w_n \, l_{(n)}(y, \theta),$$

where the weights satisfy $w_n \geq 0$ for $n = 1, \ldots, h$ and $w_h > 0$. See e.g. Hadi and Luccño (1997) and Vandev and Neykov (1998).

In the case of normal distribution with known variance, the trimmed likelihood estimators coincide with the least trimmed squares (LTS) estimators of Rousseeuw (1984, 1985) and Rousseeuw and Leroy (1987). Breakdown points of LTS estimators for linear regression were derived in Rousseeuw (1984, 1985), Rousseeuw and Leroy (1987), Vandev (1993), Vandev and Neykov (1993), Coakley and Mili (1993), Hössjer (1994), Müller (1995, 1997), Mili and Coakley (1996) and Hössjer (1994) showed also consistency and asymptotic normality. Trimmed likelihood estimators for normal distribution with unknown variance were regarded in Bednarski and Clarke (1993) who derived their asymptotic properties like Fisher consistency, asymptotic normality and compact differentiability.

Up to now, not much is known about trimmed likelihood estimators for distributions different from the normal distribution. There are approaches on robust and in particular high breakdown point estimators for logistic regression and other nonlinear models given by Stefanski, Carroll, and Ruppert (1986), Copas (1988), Künsch, Stefanski and Carroll (1989), Stromberg and Ruppert (1992), Carroll and Pederson (1993), Wang and Carroll (1993, 1995), Christmann (1994), Sakata and White (1995), Hubert (1997), Christmann and Rousseeuw (1999). But these approaches do not concern trimmed likelihood estimators.

Only Vandev and Neykov (1998) derived breakdown points of trimmed likelihood estimators for logistic regression and exponential linear models with unknown dispersion. Their approach

bases on the concept of d-fullness developed by Vandev (1993). However, they could only derive breakdown points under the restriction that the explanatory variables $x_1, \ldots, x_N$ of the logistic regression and the exponential linear model are in general position. This restriction was also used in the approaches of Rousseeuw (1984, 1985) and Rousseeuw and Leroy (1987) concerning LTS estimators. Müller (1995, 1997) and Mili and Coakley (1996) dropped this restriction and showed that then the breakdown point of LTS estimators is determined by $\mathcal{N}(X)$ defined as

$$\mathcal{N}(X) \;:=\; \max_{0 \neq \beta \in I\!R^p} \operatorname{card} \left\{ n \in \{1, \ldots, N\};\; x_n^\top \beta = 0 \right\},$$

where $X := (x_1, \ldots, x_N)^\top \in I\!R^{N \times p}$. Hence $\mathcal{N}(X)$ provides the maximum number of explanatory variables lying in a subspace. If the explanatory variables are in general position then $\mathcal{N}(X) = p - 1$ which is the minimum value for $\mathcal{N}(X)$. In other cases $\mathcal{N}(X)$ is much higher. These other cases appear mainly when the explanatory variables are not random but fixed and this happens in particular if they are given by an experimenter in a designed experiment.

In this paper we are showing that the quantity $\mathcal{N}(X)$ determines the breakdown point not only of LTS estimators in linear models but also of any trimmed likelihood estimator and related estimators as the S estimators in generalized linear models. In particular, we will show how the fullness parameter of Vandev (1993) is connected with $\mathcal{N}(X)$. This leads to a general approach about lower bounds for breakdown points in generalized linear models with and without dispersion parameters. Although the approach is a generalization and combination of that in Müller (1995, 1997), Mili and Coakley (1996) and Vandev and Neykov (1998) it is much simpler and the proofs are shorter. In particular, restrictions of the sample size and the trimming factor h which are used in Vandev and Neykov (1998) can be dropped.

In Section 2, the most general result concerning a lower bound for breakdown points of trimmed likelihood estimators in general models is presented. The first application of the general result is given in Section 3 for generalized linear models without dispersion parameter. Here it is shown how the fullness parameter d of Vandev (1993) is connected with the quantity $\mathcal{N}(X)$. From these results, lower bounds for the breakdown points in linear models, in logistic regression models and in log-linear models appear as simple examples. Since also upper bounds for the breakdown points are derived for the logistic regression and the log-linear models by special considerations, the logistic regression model and the log-linear model are treated separately in Section 4 and Section 5, respectively. The second application of the general result of Section 2 concerns generalized linear models with dispersion parameter and is presented in Section 6. Here we also derive breakdown points of S estimators by completing a proof of Rousseeuw and Yohai (1984) and Rousseeuw and Leroy (1987).

2 Breakdown points of trimmed likelihood estimators in general models

Let Θ be an topological space. For example, $\Theta = [0,1]$ for binomial experiments, $\Theta = [0,\infty)$ for variance estimation, $\Theta = I\!R^p$ for regression experiments, or $\Theta = I\!R \times I\!R^+$. In such general setting breakdown points of an estimator for $\theta \in \Theta$ are defined as follows, where $\text{int}(\Theta)$ denotes the interior of Θ. Compare e.g. Hampel et al. (1986), p. 97.

Definition 1. *The breakdown point of an estimator $\widehat{\theta} : \mathcal{Y}^N \to \Theta$ at $y \in \mathcal{Y}^N$ is defined as*

$$\epsilon^*(\widehat{\theta}, y) := \frac{1}{N} \min \{M; \\ \text{there exists no compact set } \Theta_0 \subset \text{int}(\Theta) \text{ with } \{\widehat{\theta}(\overline{y});\ \overline{y} \in \mathcal{Y}_M(y)\} \subset \Theta_0\Big\},$$

where

$$\mathcal{Y}_M(y) := \big\{\overline{y} \in \mathcal{Y}^N;\ card\{n;\ y_n \neq \overline{y}_n\} \leq M\big\}$$

is the set of contaminated samples corrupted by at most M observations.

In the case $\Theta = I\!R^p$, we have that $N \cdot \epsilon^*(\widehat{\theta}, y)$ is the smallest number M of contaminated observations so that $\{\widehat{\theta}(\overline{y});\ \overline{y} \in \mathcal{Y}_M(y)\}$ is unbounded.

In some situations the breakdown point satisfies $\epsilon^*(\widehat{\theta}, y) = 0$, which can only happen if $\widehat{\theta}(y)$ is not uniquely defined and given by values not lying in a compact subset of $\text{int}(\Theta)$. Then the estimator is not identifiable. Typically the values of a nonidentifiable estimator are lying in a whole subspace of Θ, and this happens in particular in complex models, for example, in models where the observations depend on several explanatory variables. Maximum likelihood estimators are not identifiable if the maximum of $\prod_{n=1}^N f_n(y_n, \theta)$, or equivalently the minimum of $\sum_{n=1}^N l_n(y, \theta)$, is attained at several θ. Then, setting $\gamma(\theta) = \sum_{n=1}^N l_n(y, \theta)$, a breakdown point equal to zero means that the set $\{\theta \in \Theta;\ \gamma(\theta) \leq C\}$ is not contained in a compact subset of $\text{int}(\Theta)$ for all $C \geq \min_\theta \gamma(\theta)$. Since we want to extend these considerations to trimmed likelihood estimators, we make the following definition.

Definition 2. *A function $\gamma : \Theta \to I\!R$ is called sub-compact if the set $\{\theta \in \Theta;\ \gamma(\theta) \leq C\}$ is contained in a compact set $\Theta_C \subset \text{int}(\Theta)$ for all $C \in I\!R$.*

This definition of sub-compactness is similar to the definition of Vandev and Neykov (1998) but not the same since Vandev and Neykov demanded that $\{\theta;\ \gamma(\theta) \leq C\}$ itself is a compact set. But this is for our purposes too restrictive. With Definition 2 we have the following conclusion.

Lemma 1. *The breakdown point of a maximum likelihood estimator satisfies* $\epsilon^*(\widehat{\theta}, y) > 0$ *if* γ *given by* $\gamma(\theta) = \sum_{n=1}^{N} l_n(y, \theta)$ *is sub-compact.*

Since $\max_{n=1,\ldots,N} l_n(y,\theta) \le \sum_{n=1}^{N} l_n(y,\theta) \le N \max_{n=1,\ldots,N} l_n(y,\theta)$ Lemma 1 holds also if γ is given by $\gamma(\theta) = \max_{n=1,\ldots,N} l_n(y,\theta)$. To study the breakdown behavior of maximum likelihood estimators at subsamples, we use the definition of d-fullness of Vandev and Neykov (1998).

Definition 3. *A finite set* $\Gamma = \{\gamma_n : \Theta \to I\!R;\ n = 1, \ldots, N\}$ *of functions is called* d*-full if for every* $\{n_1, \ldots, n_d\} \subset \{1, \ldots, N\}$ *the function* γ *given by* $\gamma(\theta) := \max\{\gamma_{n_k}(\theta);\ k = 1, \ldots, d\}$ *is sub-compact.*

The d-fullness of the log-likelihood functions $\{l_n(y,\cdot);\ n = 1, \ldots, N\}$ provides positive breakdown points of the maximum likelihood estimators at any subsample with d observations. Moreover, as Vandev and Neykov (1998) showed, d-fullness is also related to the breakdown points of TL and WTL estimators. Here we extend this result and show that the proof of this extension is even much shorter and simpler than that of Vandev and Neykov. For this proof, we use in particular the fact that the definition of d-fullnes is based on the maximum of $\gamma_{n_k}(\theta)$ instead of the sum. The extension concerns any estimator S of the form

$$S(y) := \arg\min_{\theta \in \Theta} s(y, \theta)$$

with $s : \mathcal{Y}^N \times \Theta \to I\!R$, where $s(y,\theta)$ can be estimated by $l_{(h)}(y,\theta)$ such that there exists $\alpha,\ \beta \in I\!R$ with $\alpha \neq 0$ and

$$\alpha\, l_{(h)}(\tilde{y}, \theta) \le s(\tilde{y}, \theta) \le \beta\, l_{(h)}(\tilde{y}, \theta) \tag{1}$$

for all $\tilde{y} \in \mathcal{Y}^N$ and $\theta \in \Theta$. It is obvious that $s(y,\theta)$ of the TL and WTL estimators satisfies condition (1). But there are also other estimators which fall under condition (1). One of these estimators, the S-estimator, is treated in Section 6.

Theorem 1. *If the estimator* S *satisfies condition (1) and* $\{l_n(y,\cdot);\ n = 1, \ldots, N\}$ *is* d*-full, then*

$$\epsilon^*(S, y) \ge \frac{1}{N} \min\{N - h + 1, h - d + 1\}.$$

The proof of Theorem 1 bases on the following lemma.

Lemma 2. *If* $M \le N - h$ *and* $M \le h - d$ *then* $l_{(d)}(y,\theta) \le l_{(h)}(\overline{y},\theta) \le l_{(N)}(y,\theta)$ *for all* $\overline{y} \in \mathcal{Y}_M(y)$ *and* $\theta \in \Theta$.

Proof. Regard $n_1, \ldots, n_h$ with $l_{(k)}(\overline{y}, \theta) = l_{n_k}(\overline{y}, \theta)$ for $k = 1, \ldots, h$. Since $h \geq M + d$ we have $1 \leq k(1) < \ldots < k(d) \leq h$ with $l_{n_{k(i)}}(\overline{y}, \theta) = l_{n_{k(i)}}(y, \theta)$. Then we obtain

$$l_{(h)}(\overline{y}, \theta) = l_{n_h}(\overline{y}, \theta) \geq l_{n_{k(d)}}(\overline{y}, \theta) \geq l_{n_{k(i)}}(\overline{y}, \theta) = l_{n_{k(i)}}(y, \theta)$$

for all $i = 1, \ldots, d$. This implies $l_{(h)}(\overline{y}, \theta) \geq l_{(d)}(y, \theta)$. The other inequality follows similarly. ■

Proof of Theorem 1. Let $M = \min\{N - h, h - d\}$. Lemma 2 together with assumption (1) provide that

$$\alpha\, l_{(d)}(y, \theta) \leq s(\overline{y}, \theta) \leq \beta\, l_{(N)}(y, \theta)$$

for all $\overline{y} \in \mathcal{Y}_M(y)$ and $\theta \in \Theta$. This means

$$\alpha\, l_{(d)}(y, S(\overline{y})) \leq s(\overline{y}, S(\overline{y})) = \min_\theta s(\overline{y}, \theta) \leq \beta \min_\theta l_{(N)}(y, \theta)$$

for all $\overline{y} \in \mathcal{Y}_M(y)$. Setting $C_0 := \frac{\beta}{\alpha} \min_\theta l_{(N)}(y, \theta)$ we have $\{S(\overline{y});\ \overline{y} \in \mathcal{Y}_M(y)\} \subset \{\theta \in \Theta;\ l_{(d)}(y, \theta) \leq C_0\}$ so that we have only to show that γ given by

$$\gamma(\theta) := l_{(d)}(y, \theta) = \max\{l_{(1)}(y, \theta), \ldots, l_{(d)}(y, \theta)\} = \max\{l_{n_1(\theta)}(y, \theta), \ldots, l_{n_d(\theta)}(y, \theta)\}$$

is sub-compact. Assume that this is not the case. Then there exists $C \in \mathbb{R}$ such that $\{\theta;\ \gamma(\theta) \leq C\}$ is not contained in a compact set. Hence, there exists a sequence $(\theta_m)_{m \in \mathbb{N}} \in \{\theta;\ \gamma(\theta) \leq C\}$ such that every subsequence of $(\theta_m)_{m \in \mathbb{N}}$ is not converging. Because of $\{n_1(\theta_m), \ldots, n_d(\theta_m)\} \subset \{1, \ldots, N\}$ we have a subsequence $(\theta_{m(k)})_{k \in \mathbb{N}}$ and $n_1, \ldots, n_d$ such that $\{n_1(\theta_{m(k)}), \ldots, n_d(\theta_{m(k)})\} = \{n_1, \ldots, n_d\}$ for all $k \in \mathbb{N}$. This implies $\gamma(\theta_{m(k)}) = \max\{l_{n_1}(y, \theta_{m(k)}), \ldots, l_{n_d}(y, \theta_{m(k)})\} \leq C$ for all $k \in \mathbb{N}$. However, $\max\{l_{n_1}(y, \cdot), \ldots, l_{n_d}(y, \cdot)\}$ is sub-compact since $\{l_1(y, \cdot), \ldots, l_N(y, \cdot)\}$ is d-full. This provides that $(\theta_{m(k)})_{k \in \mathbb{N}}$ contains a convergent subsequence which is a contradiction. Hence γ is sub-compact. ■

Note that Theorem 1 of Vandev and Neykov (1998) provides a lower bound of the breakdown point of weighted trimmed likelihood estimators which is $(N - h + 1)/N$. However this lower bound is derived under the additional assumptions of $N \geq 3\,d$ and $(N + d)/2 \leq h \leq N - d$. Since $(N+d)/2 \leq h$ implies $h - d \geq (N-d)/2 \geq N - h$ the lower bound of Vandev and Neykov is not better than that of Theorem 1. Hence Theorem 1 is not only an extension of Theorem 1 of Vandev and Neykov to other estimators but also provides the lower bound without additional assumptions on N and h.

Note also that the lower bound of Theorem 1 is maximized if the trimming factor h satisfies $\left\lfloor \frac{N+d}{2} \right\rfloor \leq h \leq \left\lfloor \frac{N+d+1}{2} \right\rfloor$ where $\lfloor z \rfloor := \max\{n \in \mathbb{N};\ n \leq z\}$. A simple consequence of this fact is the following result concerning trimmed likelihood estimators.

Theorem 2. *Assume that* $\{l_n(y, \cdot);\ n = 1, \ldots, N\}$ *is* d*-full and* $\left\lfloor \frac{N+d}{2} \right\rfloor \leq h \leq \left\lfloor \frac{N+d+1}{2} \right\rfloor$. *Then the breakdown point of any weighted trimmed likelihood estimator* WTL_h *satisfies*

$$\epsilon^*(WTL_h, y) \geq \frac{1}{N} \left\lfloor \frac{N - d + 2}{2} \right\rfloor.$$

In the next sections, we derive the fullness parameter d - and thus the lower bound for the breakdown point - for special models.

3 Application on generalized linear models without dispersion parameter

Assume that the distribution of the observations Y_n have densities $f(y_n, x_n, \beta)$ given by a linear exponential family, that is

$$f(y_n, x_n, \beta) = \exp\{T(y_n)^\top g(x_n^\top \beta) + c(x_n^\top \beta) + h(y_n)\},$$

where $T : \mathcal{Y} \to I\!R^r$, $g : I\!R \to I\!R^r$, $c : I\!R \to I\!R$, and $h : \mathcal{Y} \to I\!R$ are known functions with $\mathcal{Y} \subset I\!R^q$, $x_n \in \mathcal{X} \subset I\!R^p$, $n = 1, \ldots, N$, are known explanatory variables and $\beta \in I\!R^p$ is unknown. Then the log-likelihood functions are given by

$$l_n(y, X, \beta) = -T(y_n)^\top g(x_n^\top \beta) - c(x_n^\top \beta) - h(y_n),$$

where $X = (x_1, \ldots, x_n)^\top$. For estimating β we can use again trimmed or weighted trimmed likelihood estimators. The breakdown point of these estimators is determined according to Theorem 2 by the fullness parameter of $\{l_1(y, X, \cdot), \ldots, l_N(y, X, \cdot)\}$. We will now show that this fullness parameter depends on the quantity $\mathcal{N}(X)$ of Müller (1995) defined in the introduction. Intuitively it is clear that we only can expect identifiability of β with d observations and thus d-fullness if d explanatory variables always span the whole $I\!R^p$. This is just satisfied by $d \geq \mathcal{N}(X)+1$ by definition of $\mathcal{N}(X)$. We even have $d = \mathcal{N}(X)+1$ for a lot of generalized linear models, however sometimes with some restrictions on the sample space. The formal proof of the relation between the fullness parameter and the quantity $\mathcal{N}(X)$ is based on the following lemma.

Lemma 3. *Let $X \in I\!R^{N\times p}$ and $I \subset \{1, \ldots, N\}$ with cardinality $\mathcal{N}(X) + 1$. Then the set $\{\beta \in I\!R^p;\ \max_{i\in I} |x_i^\top \beta| \leq D\}$ is bounded for all $D \in I\!R$.*

Proof. We have the following inclusion

$$\begin{aligned}
&\{\beta \in I\!R^p;\ \max_{i\in I} |x_i^\top \beta| \leq D\} \\
&\subset \left\{\beta \in I\!R^p;\ \frac{1}{\mathcal{N}(X)+1} \sum_{i\in I} (x_i^\top \beta)^2 \leq D^2\right\} \\
&= \left\{\beta \in I\!R^p;\ \frac{1}{\mathcal{N}(X)+1}\, \beta^\top \sum_{i\in I} x_i\, x_i^\top \beta \leq D^2\right\}.
\end{aligned}$$

Because I is of cardinality $\mathcal{N}(X)+1$ the definition of $\mathcal{N}(X)$ implies that the matrix $\sum_{i\in I} x_i\, x_i^\top$ is of full rank. Hence the set $\left\{\beta \in I\!R^p;\ \frac{1}{\mathcal{N}(X)+1}\, \beta^\top \sum_{i\in I} x_i\, x_i^\top \beta \leq D^2\right\}$ is bounded. ■

Theorem 3. *If the function γ_z given by $\gamma_z(\theta) = -T(z)^\top g(\theta) - c(\theta) - h(z)$ is sub-compact for all $z \in \mathcal{Y}$ then the family $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is $\mathcal{N}(X)+1$-full for all $y \in \mathcal{Y}^N$ and all $X \in \mathcal{X}^N$.*

Proof. Regard any $C \in I\!R$ and any $I \subset \{1, \ldots, N\}$ with cardinality $\mathcal{N}(X)+1$. Because of the sub-compactness of γ_z there exists D_i, $i \in I$, such that

$$\begin{aligned}
&\left\{\beta \in I\!R^p;\ \max_{i\in I} l_i(y, X, \beta) \le C\right\} \\
&= \bigcap_{i\in I}\{\beta \in I\!R^p;\ l_i(y, X, \beta) \le C\} = \bigcap_{i\in I}\left\{\beta \in I\!R^p;\ \gamma_{y_i}(x_i^\top \beta) \le C\right\} \\
&\subset \bigcap_{i\in I}\left\{\beta \in I\!R^p;\ |x_i^\top \beta| \le D_i\right\} \subset \left\{\beta \in I\!R^p;\ \max_{i\in I} |x_i^\top \beta| \le \max_{i\in I} D_i\right\}.
\end{aligned}$$

The last set is contained in compact set because of Lemma 3. ■

Example 1 (Linear models).

In a linear model where the errors have normal distribution with known variance the log-likelihood function is

$$\begin{aligned}
&l_n(y, X, \beta) \\
&= \frac{1}{2}\frac{(y_n - x_n^\top \beta)^2}{\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2) \\
&= -y_n \frac{1}{\sigma^2} x_n^\top \beta + \frac{(x_n^\top \beta)^2}{2\sigma^2} + \frac{(y_n)^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2).
\end{aligned}$$

Since $\gamma_z(\theta) = -z\,\frac{1}{\sigma^2}\,\theta + \frac{\theta^2}{2\sigma^2} + \frac{z^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2)$ is sub-compact the condition of Theorem 3 is satisfied so that the set $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is $\mathcal{N}(X)+1$-full. Hence Theorem 2 provides that any weighted trimmed likelihood estimator with $\left\lfloor \frac{N+\mathcal{N}(X)+1}{2} \right\rfloor \le h \le \left\lfloor \frac{N+\mathcal{N}(X)+2}{2} \right\rfloor$ has a breakdown point not less than $\frac{1}{N}\left\lfloor \frac{N-\mathcal{N}(X)+1}{2} \right\rfloor$. This result was already obtained by Müller (1995, 1997) since the trimmed likelihood estimators coincide with the least trimmed squares estimators in this case. In Müller (1995, 1997) it was also shown that $\frac{1}{N}\left\lfloor \frac{N-\mathcal{N}(X)+1}{2} \right\rfloor$ is an upper bound for regression equivariant estimators as well. Since also weighted trimmed likelihood estimators are regression equivariant we even have that the breakdown point of the WTL estimators is exactly $\frac{1}{N}\left\lfloor \frac{N-\mathcal{N}(X)+1}{2} \right\rfloor$.

Note that Vandev and Neykov (1998) also treated this linear model and the least trimmed squares estimators. But they made the assumption that $x_1, \ldots, x_N$ are in general position, that is $\mathcal{N}(X) = p - 1$. Under this assumption, they showed only that the set $\{l_n(y, X, \beta);\ n = 1, \ldots, N\}$ is $p+1$-full although it is p-full.

Theorem 3 together with Theorem 1 provide only a lower bound for the breakdown points. Since regression equivariance makes only sense for linear models but not for other generalized

linear models an upper bound cannot derived by regression equivariance as it was shown for linear models by Müller (1995, 1997). In other generalized linear models it also can happen that even the maximum likelihood estimators never breaks down. However, as soon as the ML estimator has a breakdown point less than or equal to $\frac{1}{N}$ it is obvious that the following upper bound for the breakdown point holds.

Lemma 4. *If the breakdown point of the maximum likelihood estimator satisfies*

$$\epsilon^*(ML, y, X) \leq \frac{1}{N}$$

for all $y \in \mathcal{Y}^N$ and $X \in \mathcal{X}^N$, then we have for any weighted trimmed likelihood estimator WTL_h

$$\epsilon^*(WTL_h, y, X) \leq \frac{1}{N}(N - h + 1).$$

The assumption $\epsilon^*(ML, y, X) \leq \frac{1}{N}$ of Lemma 4 must be shown for each generalized linear model separately. Moreover, the upper bound given by Lemma 4 is only useful if h is large enough. In particular for $h \geq \left\lfloor \frac{N+\mathcal{N}(X)+1}{2} \right\rfloor$ we obtain $\epsilon^*(WTL_h, y, X) \leq \frac{1}{N} \left\lfloor \frac{N-\mathcal{N}(X)+2}{2} \right\rfloor$. This is very similar to the upper bound for regression equivariant estimators in linear models. However, for $h < \left\lfloor \frac{N+\mathcal{N}(X)+1}{2} \right\rfloor$ no reasonable upper bound is provided by Lemma 4. Upper bounds for small h can be only derived by special considerations for each generalized linear model separately. Therefore we treat in the following two sections as examples the logistic regression model and the log-linear model.

4 Logistic regression

Let t_n the total number of observations and $s_n \in \{0, \ldots, t_n\}$ the number of successes under condition x_n. In a logistic regression model, it is assumed that the number of successes S_n has a binomial distribution with parameters t_n and π_n where $\pi_n = \exp(x_n^\top \beta)/(1 + \exp(x_n^\top \beta))$ is the probability of success explained by the explanatory variable x_n. Setting $y = (s, t)$ with $t = (t_1, \ldots, t_N)^\top$ and $s = (s_1, \ldots, s_N)^\top$, the log-likelihood function is

$$\begin{aligned} l_n(y, X, \beta) &= l_n(s, t, X, \beta) \\ &= -s_n\, x_n^\top \beta + t_n \log(1 + \exp(x_n^\top \beta)) - \log(\tbinom{t_n}{s_n}) \\ &= \gamma_{s_n, t_n}(x_n^\top \beta). \end{aligned} \tag{2}$$

The function $\gamma_{u,v}$ given by $\gamma_{u,v}(\theta) = -u\theta + v \log(1 + \exp(\theta)) - \log(\binom{v}{u})$ is sub-compact as soon as $0 < u < v$ so that according to Theorem 3 the set $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is $\mathcal{N}(X)+1$-full for all $y = (s, t)$ satisfying $0 < s_n < t_n$ for $n = 1, \ldots, N$. Hence, Theorem 1 provides $\frac{1}{N} \min\{N - h + 1, h - \mathcal{N}(X)\}$ as an lower bound for the breakdown point of any

weighted trimmed likelihood estimator WTL_h. This lower bound is also an upper bound as the following theorem shows. For that let be $\mathcal{Y}^*$ the set of all $y = (s,t)$ with $0 < s_n < t_n$ for $n = 1,\ldots,N$. Here we exclude the case $s_n = 0$ or $s = t_n$ to avoid problems described in Christmann and Rousseeuw (1999) concerning missing overlap. A combination of our results and those of Christmann and Rousseeuw would provide a result for $0 \leq s_n \leq t_n$ but this is beyond this paper.

Theorem 4. *The breakdown point of any weighted trimmed likelihood estimator WTL_h for logistic regression satisfies*

$$\min_{y\in\mathcal{Y}^*} \epsilon^*(WTL_h, y, X) = \frac{1}{N}\,\min\{N-h+1, h-\mathcal{N}(X)\}.$$

Proof. After the remarks above we have only to show that $\frac{1}{N}\,\min\{N-h+1, h-\mathcal{N}(X)\}$ is an upper bound. If $h > \left\lfloor \frac{N+\mathcal{N}(X)+1}{2} \right\rfloor$ the upper bound follows from Lemma 4 if we can show $\epsilon^*(ML, y, X) \leq \frac{1}{N}$ for all y for the maximum likelihood estimator. Hence regard any $y = (s,t)$. If $ML(y,X)$ is not contained in a compact subset of $\mathbb{R}^p$ then $\epsilon^*(ML, y, X) = 0$. Otherwise, since the second derivative of $\sum_{n=1}^N l_n(y, X, \beta)$ with respect to β is a positive semidefinite matrix, it is sufficient and necessary for $\widehat{\beta} = ML(y,X)$ that $X^\top s = X^\top e(t, \widehat{\beta})$ holds where

$$e(t,\widehat{\beta}) := \left(t_1 \frac{\exp(x_1^\top\widehat{\beta})}{1+\exp(x_1^\top\widehat{\beta})}\ ,\ \ldots\ ,\ t_N \frac{\exp(x_N^\top\widehat{\beta})}{1+\exp(x_N^\top\widehat{\beta})} \right)^\top.$$

Regard the sequence $y^k = (s^k, t^k) \in \mathcal{Y}_1((s,t))$ with $s_1^k = 1$ and $t_1^k = k$ for all $k \in \mathbb{N}$. Assume that $\widehat{\beta}^k = ML(y^k, X)$ is bounded. Then

$$X^\top e(t^k, \widehat{\beta}^k) = x_1\, t_1^k \frac{\exp(x_1^\top\widehat{\beta}^k)}{1+\exp(x_1^\top\widehat{\beta}^k)} + x_2\, t_2 \frac{\exp(x_2^\top\widehat{\beta}^k)}{1+\exp(x_2^\top\widehat{\beta}^k)} + \ldots + x_N\, t_N \frac{\exp(x_N^\top\widehat{\beta}^k)}{1+\exp(x_N^\top\widehat{\beta}^k)}$$

is unbounded while $X^\top s^k$ is bounded which is a contradiction to $X^\top s^k = x^\top e(t^k, \widehat{\beta}^k)$. Hence $ML(y^k, X)$ is not bounded so that $\epsilon^*(ML, y, X) \leq \frac{1}{N}$.

Now regard the case $h \leq \left\lfloor \frac{N+\mathcal{N}(X)+1}{2} \right\rfloor$. W.l.o.g. we can assume that there exists β_0 such that $x_n^\top\beta_0 = 0$ for $n = 1,\ldots,\mathcal{N}(X)$. Then by definition of $\mathcal{N}(X)$ we have $x_n^\top\beta_0 \neq 0$ for $n = \mathcal{N}(X)+1,\ldots,N$. At least $M_0 = \left\lfloor \frac{N-\mathcal{N}(X)+1}{2} \right\rfloor$ of these n satisfy $x_n^\top\beta_0 < 0$ otherwise we can regard $-\beta_0$. W.l.o.g. let $x_n^\top\beta_0 < 0$ for $n = N-M_0+1,\ldots,N$. Setting $M = h-\mathcal{N}(X)$ we have $M \leq M_0$. Now regard the following special sample $y = (s,t)$ with $s_n = 1$ for $n = 1,\ldots,N$, $t_n = 2$ for $n = 1,\ldots,\mathcal{N}(X)$ and $t_n = u > 2$ for $n = \mathcal{N}(X)+1,\ldots,N$. As corrupted sample we use $\overline{y} = (\overline{s}, \overline{t})$ with $\overline{s}_n = 0$, $\overline{t}_n = t_n$ for $n = N-M+1,\ldots,N$ and $\overline{s}_n = s_n$, $\overline{t}_n = t_n$ for $n = 1,\ldots,N-M$. Then $y \in \mathcal{Y}^*$ and $\overline{y} \in \mathcal{Y}_M(y)$. Moreover, we have

$$\min_\beta l_n(\overline{s}, \overline{t}, X, \beta) = \min_\beta l_n(s,t,X,\beta) = l_n(s,t,X,k\beta_0) = \log(2)$$

for $n = 1, \ldots, \mathcal{N}(X)$ and all $k \in I\!R$, and

$$\min_{\beta} l_n(s, t, X, \beta) \geq \min_{\mu} \left(-\mu + u \log(1 + \exp(\mu)) - \log(u)\right) > \log(2)$$

for $n = \mathcal{N}(X) + 1, \ldots, N$. This implies

$$\min_{\beta} \sum_{n=1}^{h} w_n \, l_{(n)}(\overline{s}, \overline{t}, X, \beta) \geq \sum_{n=M+1}^{h} w_n \, \log(2).$$

Since we have with the property $x_n^\top \beta_0 < 0$ for $n = N - M + 1, \ldots, N$ for k large enough

$$\begin{aligned}
&\sum_{n=1}^{h} l_{(n)}(\overline{s}, \overline{t}, X, k\beta_0) \\
&\quad = \quad \sum_{n=M+1}^{h} w_n \, l_{n-M}(s, t, X, k\beta_0) + \sum_{n=1}^{M} w_n \, t_n \, \log\left(1 + \exp(x_{N-n+1}^\top \, k \, \beta_0)\right) \\
&\quad \stackrel{k \to \infty}{\longrightarrow} \quad \sum_{n=M+1}^{h} w_n \, \log(2),
\end{aligned}$$

the estimator $WTL_h(\overline{y}, X)$ is not contained in a bounded subset of $I\!R^p$ so that $\epsilon^*(WTL_h, y, X) \leq \frac{1}{N} M = \frac{1}{N} \left(h - \mathcal{N}(X)\right)$. ■

The proof of Theorem 4 shows that for deriving the upper bound for $h > \left\lfloor \frac{N + \mathcal{N}(X) + 1}{2} \right\rfloor$ it is necessary to assume that also the total numbers t_n can be contaminated by outliers. Without this assumption even the maximum likelihood estimator need not to break down by one or more corrupted observations. The assumption of possibly contaminated total numbers makes sense as soon as the total numbers are given by random which is often the case (see the example below). If the total numbers are given by random then the log-likelihood function (2) should have additional terms. But since these terms are additive they do not influence the determination of the maximum likelihood estimator if these terms are independent of β. However they can influence the trimmed likelihood estimators by changing the order of the $l_n(y, X, \beta)$ and would make the second step of the proof of Theorem 4 (for $h \leq \left\lfloor \frac{N + \mathcal{N}(X) + 1}{2} \right\rfloor$) more complicated though it also would work. Thus here we regarded for simplicity the simple trimmed likelihood estimators based on the log-likelihood functions given by (2).

Theorem 4 in particular shows that the maximum breakdown point for logistic regression is attained for h satisfying $\left\lfloor \frac{N + \mathcal{N}(X) + 1}{2} \right\rfloor \leq h \leq \left\lfloor \frac{N + \mathcal{N}(X) + 2}{2} \right\rfloor$ and equals $\frac{1}{N} \left\lfloor \frac{N - \mathcal{N}(X) + 1}{2} \right\rfloor$. Hence we have the same maximum breakdown point value and the same optimal trimming proportion h as for linear models.

Note, that for the special case that $x_1, \ldots, x_N$ are in general position, that is $\mathcal{N}(X) = p - 1$, a lower bound for the breakdown point similar to that in Theorem 4 was already obtained by Vandev and Neykov (1998) under the additional restriction of $N \geq 3(p + 1)$. Thereby, they

showed again that the set $\{l_n(y, X, \beta);\ n = 1, \ldots, N\}$ is only $p+1$-full although it is p-full.

Example 2 (Toxicological experiment with fish eggs).

This example involves data which resulted from a toxicological experiment conducted at the University of Waterloo, Canada, and are presented in O'Hara Hines and Carter (1993, p.13). Six different concentrations of the toxicant potassium cyanate (KSCN) were applied to 48 vials of trout fish eggs. Each vial contained between 61 and 179 eggs. The eggs in half the vials were allowed to water harden for several hours before the toxicant was applied (this is a process in which the surface of a fish eggs becomes toughened after a few hours in water). For the remaining vials, the toxicant was applied immediately after fertilization. After 19 days of the start of the experiment the number of dead eggs in each vial was counted.

	WH	Concen-tration	No Eggs	No Dead		WH	Concen-tration	No Eggs	No Dead
1	1	90	111	8	25	0	90	130	7
2	1	90	97	10	26	0	90	179	25
3	1	90	108	10	27	0	90	126	5
4	1	90	122	9	28	0	90	129	3
5	1	180	68	4	29	0	180	114	12
6	1	180	109	6	30	0	180	149	4
7	1	180	109	11	31	0	180	121	4
8	1	180	118	6	32	0	180	105	0
9	1	360	98	6	33	0	360	102	4
10	1	360	110	5	34	0	360	145	21
11	1	360	129	9	35	0	360	61	1
12	1	360	103	17	36	0	360	118	3
13	1	720	83	2	37	0	720	99	29
14	1	720	87	3	38	0	720	109	53
15	1	720	118	16	39	0	720	99	40
16	1	720	100	9	40	0	720	70	0
17	1	1440	140	60	41	0	1440	100	14
18	1	1440	114	47	42	0	1440	127	10
19	1	1440	103	49	43	0	1440	132	8
20	1	1440	110	20	44	0	1440	113	3
21	1	2880	143	79	45	0	2880	145	113
22	1	2880	131	85	46	0	2880	103	84
23	1	2880	111	78	47	0	2880	143	105
24	1	2880	111	74	48	0	2880	102	78

Treating the number of dead eggs in each vial as the response, a logistic regression model was fitted to the data with covariates for water hardening (0 if the toxicant was applied before water hardening and 1 after), and for a linear and quadratic term in log-concentration of toxicant. The quadratic term in log-concentration is used to describe a sharp increase in mortality caused by the two highest concentrations. Thus the logistic regression model is

$$\text{logit}(p/(1-p)) = \beta_1 + \beta_2 * WH + \beta_3 * \log_{10}(Concentration) + \beta_4 * \log_{10}(Concentration)^2$$

The maximum likelihood estimator for $(\beta_1, \beta_2, \beta_3, \beta_4)^\top$ based on all observations is $ML(y, X) = (10.28, 0.03, -11.4, 2.50)^\top$.

O' Hara Hines and Carter (1993) pinpoint the observations 38, 39 and 26 as possible outliers. They also reported that Pregibon's influence diagnostics indicated that the observations 38 and 39 were pinpointed as potential outliers. The maximum likelihood estimator without the observations 38 and 39 is $(15.40, 0.27, -15.53, 3.26)^\top$ and without the observations 26, 38 and 39 is $(14.04, 0.32, -14.64, 3.11)^\top$.

Markatou et al. (1997) analyzed the same data. They identified the observations 38 and 39 as potential outliers, whilst their methods gave a weight nearly 1 to observations 26 by means of the negative exponential RAF (Residual Adjustment Function) downweight function. When the Hellinger RAF was used for the construction of the weights, observations 13, 32, 40, 43 and 44 received a weight of 0. They reported that examination of those observations revealed that observations 32 and 40 had a 0 response, while observations 43 and 44 had the lowest mortality at concentration levels 720 and 1440, respectively, at the same water-hardening level. The maximum likelihood estimate without the observations 13, 32, 40, 43 and 44 is $(6.49, -0.23, -8.42, 1.97)^\top$.

For satisfying the assumption $0 < s_n < t_n$ of Theorem 4, we dropped the observations 32 and 40 for our calculations so that only 46 observations are available. Since 24 observations satisfy WH=1, we have $\mathcal{N}(X) = 24$. Hence, according to Theorem 4, the maximum breakdown point is 11/46 and is attained by any weighted trimmed likelihood estimator with $h = 35$ or $h = 36$. Since an exact algorithm for calculating the trimmed likelihood estimator with weights $w_n = 1$ and $h = 36$ had run too long, we used a genetic algorithm and obtained $TL_{36}(y, X) = (7.36, -0.12, -9.29, 2.16)^\top$. The trimmed observations were 13, 14, 20, 21, 38, 39, 41, 42, 43, 44. Hence there is some coincidence with the results of Markatou et al. (1997) with respect to the estimate and the trimmed observations.

5 Log-linear models

The response variables Y_n of a log-linear model have a Poisson distribution with parameter $\lambda_n = \exp(x_n^\top \beta)$ so that the log-likelihood function is

$$l_n(y, X, \beta) = -y_n\, x_n^\top \beta + \exp(x_n^\top \beta) + \log(y_n!).$$

The function γ_z given by $\gamma_z(\theta) = -z\theta + \exp(\theta) + \log(z!)$ is sub-compact as soon as $z > 0$ so that according to Theorem 3 the set $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is $\mathcal{N}(X)+1$-full for all y satisfying $y_n > 0$ for all $n = 1, \ldots, N$. Hence, Theorem 1 provides $\frac{1}{N} \min\{N - h + 1, h - \mathcal{N}(X)\}$ as a lower bound for the breakdown point of any weighted trimmed likelihood estimator WTL_h. This lower bound is also an upper bound as the following theorem shows. For that let be $\mathcal{Y}^*$ the set of all y with $y_n > 0$ for $n = 1, \ldots, N$.

Theorem 5. *The breakdown point of any weighted trimmed likelihood estimator WTL_h for a log-linear model satisfies*

$$\min_{y \in \mathcal{Y}^*} \epsilon^*(WTL_h, y, X) = \frac{1}{N} \min\{N - h + 1, h - \mathcal{N}(X)\}.$$

Proof. The proof is similar to that for the logistic regression model. The first step is to show $\epsilon^*(ML, y, X) \leq \frac{1}{N}$ for using Lemma 4. Again a sufficient and necessary condition for $\widehat{\beta} = ML(y, x)$ lying in a bounded subset of $\mathbb{R}^p$ is $X^\top y = X^\top e(\widehat{\beta})$ where here $e(\widehat{\beta}) := (\exp(x_1^\top \widehat{\beta}), \ldots, \exp(x_1^\top \widehat{\beta}))^\top$. Then, as soon as $y^k \in \mathcal{Y}_1(y)$ is unbounded also the corresponding $\widehat{\beta}^k = ML(y^k, X)$ cannot be bounded.

The second step is to construct a sample $y \in \mathcal{Y}^*$ and a corrupted sample $\overline{y} \in \mathcal{Y}_M(y)$ with $M = h - \mathcal{N}(X)$ such that $WTL_h(\overline{y}, X)$ is not contained in a bounded subset of $\mathbb{R}^p$. For that, let $x_1, \ldots, x_n$ and β_0 as in the proof of Theorem 4. Set $y_n = 1$ for $n = 1, \ldots, \mathcal{N}(X)$, $y_n = z > e^2 - \frac{1}{2}$ for $n = \mathcal{N}(X) + 1, \ldots, N$, $\overline{y}_n = y_n$ for $n = 1, \ldots, N - M$, and $\overline{y}_n = 0$ for $n = N - M + 1, \ldots, N$. Then we have

$$\min_\beta l_n(\overline{y}, X, \beta) = \min_\beta l_n(y, X, \beta) = l_n(y, X, k\beta_0) = 1$$

for $n = 1, \ldots, \mathcal{N}(X)$ and all $k \in \mathbb{R}$, and

$$\begin{aligned}
&\min_\beta l_n(y, X, \beta) \\
&\quad \geq \min_\mu \left(-z\mu + \exp(\mu) + \log(z!)\right) \\
&\quad = -z\, \log(z) + z + \log(z!) \\
&\quad \geq -z\, \log(z) + z + \left(z + \frac{1}{2}\right) \log\left(z + \frac{1}{2}\right) - z - \frac{1}{2} \log\left(\frac{1}{2}\right) \\
&\quad \geq \frac{1}{2} \log\left(z + \frac{1}{2}\right) > 1
\end{aligned}$$

for $n = \mathcal{N}(X) + 1, \ldots, N$. The rest follows as in the proof of Theorem 4. ■

Again the maximum breakdown point for log-linear models is $\frac{1}{N}\left\lfloor \frac{N-\mathcal{N}(X)+1}{2} \right\rfloor$ and coincides with the maximum breakdown point for linear models. This maximum breakdown point is also attained by the same trimming proportion h.

6 Application on exponential linear models with dispersion parameter

Assume that the observations Y_n are distributed with q-th power exponential distribution, i.e., the density function is given by

$$f(y_n, x_n, \beta, \sigma) = \frac{q\,(1/2)^{(1+1/q)}}{\sigma\,\Gamma(1/2)} \exp\left(-\frac{1}{2}\left|\frac{y_n - x_n^\top \beta}{\sigma}\right|^q\right\},$$

where Γ is here the gamma function. Special cases of this distribution are the normal ($q = 2$), the Laplace ($q = 1$), the double exponential ($0 < q < 2$), the leptokurtic ($1 < q < 2$), the platikurtic ($q > 2$) and the rectangular distribution ($q \to \infty$). The fullness parameter of $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$, where

$$l_n(y, X, \beta, \sigma) = \frac{1}{2}\left|\frac{y_n - x_n^\top \beta}{\sigma}\right|^q + \log(\sigma) - \log\left(\frac{q\,(1/2)^{(1+1/q)}}{\Gamma(1/2)}\right) \tag{3}$$

with $\beta \in I\!R^p$ and $\sigma \in I\!R^+$, was derived in Vandev and Neykov (1998) under the assumption that $x_1, \ldots, x_N$ are in general position, i.e. $\mathcal{N}(X) = p - 1$. They showed in Lemma 3 that the fullness parameter is $p + 1$. Here we show that the fullness parameter is even p and that the fullness parameter can be also determined in the case where $x_1, \ldots, x_N$ are not in general position.

Lemma 5. *If the log-likelihood function is given by (3) with $q > 0$ then the set $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is $\mathcal{N}(X)+1$-full.*

Proof. We have to show that γ given by

$$\gamma(\beta, \sigma) := \max_{i \in I} \frac{1}{2}\left|\frac{y_i - x_i^\top \beta}{\sigma}\right|^q + \log(\sigma) - K$$

with $K \in I\!R$ is sub-compact for all $I \subset \{1, \ldots, N\}$ with cardinality $\mathcal{N}(X) + 1$. We will do it with the same trick as in Vandev and Neykov (1998) but with a shorter proof. Take any $C \in I\!R$ and set $\tilde{\beta}(\sigma) := \arg\min\{\gamma(\beta, \sigma);\ \beta \in I\!R^p\}$ and $\tilde{\sigma}(\beta) := \arg\min\{\gamma(\beta, \sigma);\ \sigma \in I\!R^+\}$. Then $\tilde{\beta}(\sigma)$ is independent of σ such that $\tilde{\beta}(\sigma) =: \tilde{\beta}$. Setting

$$\gamma_1(\sigma) := \gamma(\tilde{\beta}(\sigma), \sigma) = \max_{i \in I} \frac{1}{2}\left|\frac{y_i - x_i^\top \tilde{\beta}}{\sigma}\right|^q + \log(\sigma) - K$$

we see that γ_1 is a sub-compact function. Hence, there exists a compact set $\Theta_1 \subsetneqq I\!R^+$ such that $\{\sigma;\ \gamma_1(\sigma) \le C\} \subset \Theta_1$. Moreover, we have that with $\eta(\beta) := \max_{i\in I} \left|y_i - x_i^\top \beta\right|$

$$\tilde{\sigma}(\beta) = \eta(\beta)\ \left(\frac{q}{2}\right)^{1/q}$$

so that

$$\gamma_2(\beta) := \gamma(\beta, \tilde{\sigma}(\beta)) = \frac{1}{q} + \log\left(\eta(\beta)\right) + \frac{1}{q}\ \log\left(\frac{q}{2}\right) - K.$$

Example 1 implies that η is sub-compact. Since the logarithm is monoton also γ_2 is sub-compact so that $\{\beta;\ \gamma_2(\beta) \le C\} \subset \Theta_2$ for some compact set $\Theta_2 \subsetneqq I\!R^p$. Then we have

$$\begin{aligned}
&\{(\beta,\sigma) \in I\!R^p \times I\!R^+;\ \gamma(\beta,\sigma) \le C\} \\
&\quad\subset\ \ \{(\beta,\sigma) \in I\!R^p \times I\!R^+;\ \gamma_1(\sigma) \le C \text{ and } \gamma_2(\beta) \le C\} \subset \Theta_2 \times \Theta_1. \ \blacksquare
\end{aligned}$$

Theorem 1 and Lemma 5 immediately imply that any weighted trimmed likelihood estimator WTL_h for (β,σ) has a breakdown point not less than $\frac{1}{N}\min\{N-h+1, h-\mathcal{N}(X)\}$ and that the lower bound of the breakdown point attains its maximum value of $\frac{1}{N}\left\lfloor\frac{N-\mathcal{N}(X)+1}{2}\right\rfloor$ if $\left\lfloor\frac{N+\mathcal{N}(X)+1}{2}\right\rfloor \le h \le \left\lfloor\frac{N+\mathcal{N}(X)+2}{2}\right\rfloor$. This maximum lower bound is also the upper bound for the breakdown point since the estimator for β is regression equivariant so that the upper bound follows from Müller (1995, 1997).

However, also other robust estimators can be used for distributions with unknown dispersion parameter. Estimators with good breakdown properties are the S-estimators. An S-estimator S_c is defined by (see Rousseeuw and Yohai 1984, Rousseeuw and Leroy 1987, p. 135)

$$S_c(y,X) := \arg\min_\beta s_c(y,X,\beta),$$

where $s_c(y,X,\beta)$ is given as solution of

$$b_c(y,X,\beta,s_c(y,X,\beta)) := \frac{1}{N}\sum_{n=1}^{N} \rho_c\left(\frac{|y_n - x_n^\top\beta|}{s_c(y,X,\beta)}\right) = K.$$

Usually K is chosen as the expectation $E_{\beta,\sigma}(b_c(Y,X,\beta,\sigma))$ to get consistency under the model distribution. If ρ_c is strictly increasing on $[0,c]$ and constant on $[c,\infty)$ then S-estimators have high breakdown points. This was shown in Rousseeuw and Yohai (1984) for $x_1,\ldots,x_N$ in general position and in Mili and Coakley (1996) for general $x_1,\ldots,x_N$. However they showed only the inequality

$$\alpha\, l_{(h)}(y,X,\beta) \le s_c(y,X,\beta) \le \beta\, l_{(h)}(y,X,\beta) \tag{4}$$

for all $y,\ X,\ \beta$, and c satisfying $\rho_c(c) = K\,N/(N-h+1)$, where $l_n(y,X,\beta) = |y_n - x_n^\top\beta|$. A detailed proof of the inequality (4) for $h = \left\lfloor\frac{N+1}{2}\right\rfloor$ was given in the book of Rousseeuw

and Leroy (1987), p. 136-139. Also in this book, they conclude from (4) without additional arguments that the breakdown points of $\arg\min_\beta l_{(h)}(y, X, \beta)$ and $S_c(y, X)$ coincide. But the proof of Theorem 1 shows that indeed additional arguments are necessary and that they base on the concept of d-fullness. Hence only now a complete proof of the breakdown points of the S-estimators is possible.

Theorem 6. *Any S-estimator S_c with $\rho_c(c) = K\,N/(N-h+1)$ has a breakdown point satisfying $\epsilon^*(S_c, y, X) \geq \frac{1}{N}\min\{N-h+1, h-\mathcal{N}(X)\}$. If $\left\lfloor \frac{N+\mathcal{N}(X)+1}{2} \right\rfloor \leq h \leq \left\lfloor \frac{N+\mathcal{N}(X)+2}{2} \right\rfloor$ then $\epsilon^*(S_c, y, X) = \frac{1}{N}\left\lfloor \frac{N-\mathcal{N}(X)+1}{2} \right\rfloor$.*

Proof. It follows from Example 1 that the set $\{l_n(y, X, \cdot);\ n = 1, \ldots, N\}$ is $\mathcal{N}(X)+1$-full. Hence, Theorem 1 and inequality (4) provide the lower bounds. That $\frac{1}{N}\left\lfloor \frac{N-\mathcal{N}(X)+1}{2} \right\rfloor$ is also an upper bound follows from the result of Müller (1995, 1997) concerning regression equivariant estimators. ■

References

Bednarski, T. and Clarke, B. R. (1993). Trimmed likelihood estimation of location and scale of the normal distribution. *Austral. J. Statist.* **35**, 141-153.

Carroll, R.J. and Pederson, S. (1993). On robustness in the logistic regression model. *J. R. Statist. Soc. B* **55**, 693-706.

Christmann, A. (1994). Least median of weighted squares in logistic regression with large strata. *Biometrika* **81**, 413-417.

Christmann, A. and Rousseeuw, P.J. (1999). Measuring overlap in logistic regression. *Technical Report, University of Antwerp, submitted.*

Coakley, C.W. and Mili, L. (1993). Exact fit points under simple regression with replication. *Statist. Probab. Lett.* **17**, 265-271.

Copas, J. B. (1988). Binary regression models for contaminated data (with discussion). *J. R. Statist. Soc. B* **50**, 225-265.

Hadi, A.S. and Luccño, A. (1997). Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics & Data Analysis* **25**, 251-272.

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.

Hand, D.J., Daly, F., Lunn, A.D., Mc Conway, K.J. and Ostrowski, E. (1994). *A Handbook of Small Data Sets.* Chapman & Hall, London.

Hössjer, O. (1994). Rank-based estimates in the linear model with high breakdown point. *J. Amer. Statist. Assoc.* **89**, 149-158.

Hubert, M. (1997). The breakdown value of the L_1 estimator in contingency tables. *Statistics and Probability Letters.* **33**, 419-425.

Künsch, H. R., Stefanski, L. A. and Carroll, R. J. (1989). Conditionally unbiased bounded influence estimation in general regression models, with applications to generalized linear models. *J. Amer. Statist. Assoc.* **84**, 460-466.

Markatou, M., Basu, A. and Lindsay, B. (1997). Weighted likelihood estimating equations: The discrete case with applications to logistic regression. *J. Statist. Plann. Inference.* **57**, 215-232.

Mili, L. and Coakley, C. W. (1996). Robust estimation in structured linear regression. *Ann. Statist.* **15**, 2593-2607.

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.

O'Hara Hines, R.J. and Carter, E.M. (1993). Improved added variable and partial residual plots for the detection of influential observations in generalized linear models. *Appl. Statist.* **42**, 3-20.

Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79**, 851-857.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Vol. B*, eds. W. Grossman, G. Pflug, I. Vincze and W. Wertz. Reidel, Dordrecht, 283-297.

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

Rousseeuw, P. J. and Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis*, eds. J. Franke, W. Härdle, R. D. Martin. *Lecture Notes in Statistics* **26**, Springer, New York, 256-272.

Sakata, S. and White, H. (1995). An alternative definition of finite-sample breakdown point with applications to regression model estimators. *J. Amer. Statist. Assoc.* **90**, 1099-1106.

Stefanski, L.A., Carroll, R.J. and Ruppert, D. (1986). Optimally bounded score functions for generalized linear models with applications to logistic regression. *Biometrika* **73**, 413-424.

Stromberg, A. J. and Ruppert, D. (1992). Breakdown in nonlinear regression. *J. Amer. Statist. Assoc.* **87**, 991-997.

Vandev, D. L.(1993). A note on breakdown point of the least median squares and least trimmed squares. *Statistics and Probability Letters.* **16**, 117-119.

Vandev, D. and Neykov, N.(1998). About regression estimators with high breakdown point. *Statistics.* **32**, 111-129.

Wang, C.Y. and Carroll, R.J. (1993). On robust estimation in logistic case-control studies. *Biometrika* **80**, 237-241.

Wang, C.Y. and Carroll, R.J. (1995). On robust logistic case-control studies with response-dependent weights. *J. Statist. Plann. Inference* **43**, 331-340.