Designs with High Breakdown Point in Nonlinear Models

Christine H. Müller and Christina Schäfer

Abstract Least trimmed squares estimators are outlier robust since they have a high breakdown point because of trimming large residuals. But the breakdown point depends also on the design. In generalized linear models and nonlinear models, the connection between breakdown point and design is given by the fullness parameter defined by Vandev and Neykov (1998). As Müller and Neykov (2003) have shown, this fullness parameter is given in generalized linear models by the largest subdesign where the interesting parameter is not identifiable. In this paper, we show that this connection does not hold for all nonlinear models. This means that the identifiability at subdesigns cannot be used for finding designs which provide high breakdown points. Instead of this, the fullness parameter itself must be determined. For some nonlinear models with two parameters, the fullness parameter is derived here. It is shown that the fullness parameter and thus a lower bound for the breakdown point depends heavily on the design and the parameter space.

1 Introduction

We assume a nonlinear model given by

$$Y_n = g(t_n, \theta) + Z_n,$$

where $Y_1, \ldots, Y_N$ are independent observations, $Z_1, \ldots, Z_N$ are independent errors, $\theta \in \Theta \subset \Re^r$ is an unknown parameter, $t_1, \ldots, t_N \in \mathscr{T} \subset \Re^q$ are nonrandom exper-

Christine H. Müller
Fakultät Statistik, Technische Universität Dortmund, D-44221 Dortmund, Germany, e-mail: cmueller@statistik.tu-dortmund.de

Christina Schäfer
Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg i.O., Germany, e-mail: christina-schaefer@hotmail.de

imental conditions, and $g : \mathscr{T} \times \Theta \to \Re$ is a known function which is nonlinear in θ. Set $Y = (Y_1, \ldots, Y_N)^\top$ with realization $y = (y_1, \ldots, y_N)^\top$ and let $D = (t_1, \ldots, t_N)^\top$ be the design. The density of Y_n is given by

$$f(y_n, t_n, \theta) = h((y_n - g(t_n, \theta))^2) \tag{1}$$

where h is a monotone decreasing function which is known. Then the negative log-likelihood function given by $l_n(y, D, \theta) = -log(f(y_n, \theta))$ is a monotone increasing function of $|y_n - g(t_n, \theta)|$.

Here we will consider the breakdown point behavior of estimators $\widehat{\theta} : \Re^N \to \Theta$. The breakdown point of an estimator $\widehat{\theta}$ is defined according to Rousseeuw and Leroy (1987) by

$$\varepsilon^*(\hat{\theta}, y) := \frac{1}{N} \min \{M;$$
$$\text{there exists no compact set } \Theta_0 \subset \Theta \text{ with } \left\{\hat{\theta}(\overline{y});\ \overline{y} \in \mathscr{Y}_M(y)\right\} \subset \Theta_0 \right\},$$

where $\mathscr{Y}_M(y) := \left\{\overline{y} \in \Re^N;\ \text{card}\{n;\ y_n \neq \overline{y}_n\} \leq M\right\}$ is the set of contaminated samples corrupted by at most M observations. Often the condition $\Theta_0 \subset \Theta$ is replaced by

$$\Theta_0 \subset \text{int}(\Theta) \tag{2}$$

to include also the implosion point for restricted parameter spaces. To facilitate the task here, we will consider only $\Theta_0 \subset \Theta$ which means that the breakdown point is only an explosion point.

There are several approaches to high breakdown point estimators for nonlinear models. See e.g. Stromberg and Ruppert (1992), Vandev (1993), Sakata and White (1995), Vandev and Neykov (1998). High breakdown point estimators are in particular obtained by trimming large residuals. See e.g. the least trimmed squares estimators in Rousseeuw and Leroy (1987), Procházka (1988), or Jurečková and Procházka (1994). However, in generalized linear models or nonlinear models it is more appropriate to trim the smallest likelihood functions or the largest negative loglikelihood functions as Vandev (1993) and Hadi and Luceño (1997) proposed.

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 estimator $TL_h(y)$ given by

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

where $N - h$ observations are trimmed and $l_{(1)}(y, D, \theta) \leq \ldots \leq l_{(N)}(y, D, \theta)$. Vandev (1993) and Vandev and Neykov (1998) studied the breakdown point behavior of trimmed likelihood estimators and showed a relation between the breakdown point and the fullness parameter d of $\{l_n(y, D, \cdot);\ n = 1, \ldots, N\}$. They defined the d-fullness as follows.

Definition 1. A finite set $\Psi = \{\psi_n : \Theta \to \Re;\ n = 1,\ldots,N\}$ of functions is called d-full if for every $\{n_1,\ldots,n_d\} \subset \{1,\ldots,N\}$ the function ψ given by $\psi(\theta) := \max\{\psi_{n_k}(\theta);\ k = 1,\ldots,d\}$ is sub-compact. If $\psi(\theta) := \max\{\psi_n(\theta);\ n = 1,\ldots,N\}$ is not subcompact, then the fullness parameter of $\Psi = \{\psi_n : \Theta \to \Re;\ n = 1,\ldots,N\}$ is defined as $N+1$.

Thereby a function $\psi : \Theta \to \Re$ is called sub-compact if the set $\{\theta \in \Theta;\ \psi(\theta) \le C\}$ is contained in a compact set $\Theta_C \subset \Theta$ for all $C \in \Re$.

Again we use here for simplicity $\Theta_C \subset \Theta$ instead of $\Theta_C \subset \mathrm{int}(\Theta)$ in the the original definition of Vandev and Neykov.

The relation between breakdown point and d-fullness was worked out in more detail by Müller and Neykov (2003). In particular they showed the following theorem.

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

$$\varepsilon^*(TL_h, y) \ge \frac{1}{N} \left\lfloor \frac{N-d+2}{2} \right\rfloor.$$

Theorem 1 means in particular that the fullness parameter d should be as small as possible to achieve a high breakdown point. Müller and Neykov (2003) also proved that the fullness parameter of $\{l_n(y,\cdot);\ n = 1,\ldots,N\}$ in linear models and in many generalized linear models satisfies $d = \mathscr{N}(D) + 1$ where the so called identifiability parameter

$$\mathscr{N}(D) := \max\left\{ \sum_{n=1}^{N} 1_{\mathscr{D}}(t_n);\ \mathscr{D} \subset \{t_1,\ldots,t_N\} \text{ where } \theta \text{ is not identifiable at } \mathscr{D} \right\}$$

was introduced by Müller (1995). In linear models and generalized linear models, where $g(t_n,\theta) = x(t_n)^\top \theta$ is satisfied, we have that

$$\mathscr{N}(D) = \max_{0 \ne \theta \in \Re^p} \mathrm{card}\left\{ n \in \{1,\ldots,N\};\ x(t_n)^\top \theta = 0 \right\},$$

so that $\mathscr{N}(D)$ provides the maximum number of explanatory variables lying in a subspace. This means in particular for linear and generalized linear models that $d = \mathscr{N}(D) + 1$ is the smallest number so that every subset of the design with this number of points provides identifiability of θ.

Although Theorem 1 holds also for nonlinear models and identifiability can be defined also for nonlinear models, there is no simple relation between identifiability and the fullness parameter d which holds for all nonlinear models. This is shown in Section 2. Section 3 and Section 4 treat the determination of the fullness parameter for two special nonlinear models with two parameters. Thereby, nonlinear models with unrestricted parameter space are considered in Section 3, and nonlinear models with restricted parameter space are studied in Section 4. It is shown that the fullness parameter and thus the lower bound for the breakdown point depends heavily on the design and the parameter space. In Section 5, extensions of the results are discussed.

2 Identifiability and d fullness

If the density satisfies (1), then the monotony of h and the logarithm implies with the triangle inequality

$$\begin{aligned} &\max\{l_{n_k}(y,D,\theta);\ k=1,\ldots,d\} \\ &\quad = \max\{-log(h((y_{n_k}-g(t_{n_k},\theta))^2));\ k=1,\ldots,d\} \leq C \\ &\Longleftrightarrow \max\{|g(t_{n_k},\theta)|;\ k=1,\ldots,d\} \leq C_2, \end{aligned}$$

where the constants C, C_1, and C_2 are independent of θ, but depend on y. Hence the following theorem holds.

Theorem 2.

$$\{l_n(y,D,\cdot);\ n=1,\ldots,N\} \quad \textit{is d-full} \Leftrightarrow \{|g(t_n,\theta)|;\ n=1,\ldots,N\} \quad \textit{is d-full}.$$

Identifiability in nonlinear models is defined as follows.

Definition 2. θ is identifiable at D with respect to g if and only if

$$g(t_n,\theta) = g(t_n,\tilde{\theta}) \text{ for all } n=1,\ldots,N \implies \theta = \tilde{\theta}$$

for all $\theta,\ \tilde{\theta} \in \Theta$.

Identifiability in nonlinear models with more than two unknown parameters is often difficult to verify. Therefore, only a simple nonlinear model is regarded, namely $g(t,\theta) = \alpha \cdot \exp(\beta t)$. Then the following result holds.

Theorem 3. *If* $g(t,\theta) = \alpha \cdot \exp(\beta t)$ *with* $\theta = (\alpha,\beta)^\top \in \Theta = [a,\infty) \times [b,\infty)$ *and* $0 < a < \frac{1}{\exp(bt)}$ *and* $D = t$ *with* $t > 0$ *then we have*

- θ *is not identifiable at* D,
- $|g(t,\cdot)|$ *is subcompact and thus* $\{l_1(y,D,\cdot)\}$ *is 1-full.*

Theorem 3 means for all designs $D = (t_1,\ldots,t_N) \in \Re^N$ with $t_n \neq 0$ for $n = 1,\ldots,N$ that the fullness parameter d of $\{l_n(y,D,\cdot);\ n=1,\ldots,N\}$ is 1 while the identifiability parameter satisfies $\mathscr{N}(D) \geq 1$. Hence $d = \mathscr{N}(D)+1$, which holds in linear and many generalized models, is not satisfied.

Proof of Theorem 3. Since $a < \frac{1}{\exp(bt)}$, there exists $\alpha, \tilde{\alpha}$ with $a < \alpha < \tilde{\alpha} < \frac{1}{\exp(bt)}$. Set $\beta = \ln\left(\frac{1}{\alpha}\right)\frac{1}{t}$ and $\tilde{\beta} = \ln\left(\frac{1}{\tilde{\alpha}}\right)\frac{1}{t}$. Then we have $\beta, \tilde{\beta} > \ln(\exp(bt))\frac{1}{t} = b$ and

$g(t,(\alpha,\beta)) = \alpha \exp\left(\ln\left(\frac{1}{\alpha}\right)\frac{1}{t}t\right) = 1 = \tilde{\alpha}\exp\left(\ln\left(\frac{1}{\tilde{\alpha}}\right)\frac{1}{t}t\right) = g(t,(\tilde{\alpha},\tilde{\beta}))$, so that $\theta = (\alpha,\beta)^\top$ is not identifiable at $D = t$. Furthermore, $|g(t,(\alpha,\beta))| = \alpha \cdot \exp(\beta t) \leq C$ implies $a \leq \alpha \leq \frac{C}{\exp(\beta t)} \leq \frac{C}{\exp(bt)}$ and $\exp(\beta t) \leq \frac{C}{\alpha} \leq \frac{C}{a}$ so that $\theta = (\alpha,\beta)^\top \in \left[a, \frac{C}{\exp(bt)}\right] \times \left[b, \frac{1}{t}\cdot\ln\left(\frac{C}{a}\right)\right]$ which is a compact set. Hence $\{l_1(y,D,\cdot)\}$ is 1-full. □

A restricted parameter space like $\Theta = [a,\infty)\times[b,\infty)$ used in Theorem 3 is typical for nonlinear models based on exponential functions with high breakdown point. This is discussed in more detail in the following two sections.

3 Nonlinear models with unrestricted parameter space

In this section, nonlinear models based on the exponential function with two parameters are studied. If the design consists of one negative and one positive value then no restriction of the parameter space is necessary.

Theorem 4. *If* $t_1 < 0 < t_2$ *and* $g(t,\theta) = \alpha + \exp(\beta t)$ *or* $g(t,\theta) = \alpha t + \exp(\beta t)$ *with* $\theta = (\alpha,\beta)^\top \in \Theta = \Re \times \Re$ *then*

$$\max\{|g(t_1,\cdot)|, |g(t_2,\cdot)|\}$$

is subcompact.

Proof. Consider at first $g(t,\theta) = \alpha + \exp(\beta t)$ and let be $C \geq 0$ arbitrary. Then $\max\{|g(t_1,\theta)|,|g(t_2,\theta)|\} \leq C$ implies $-C \leq \alpha + \exp(\beta t_i) \leq C$ for $i = 1,2$ so that $-C - \exp(\beta t_i) \leq \alpha \leq C - \exp(\beta t_i) \leq C$ for $i = 1,2$. Since $t_1 < 0 < t_2$, it holds $\beta t_1 \leq 0$ or $\beta t_2 \leq 0$ for any β so that $\alpha \geq -C - \exp(0) = -C - 1$. Hence $\alpha \in [-C-1, C]$.

Moreover, $\exp(\beta t_i) \leq C - \alpha \leq 2C+1$ for $i = 1,2$ so that $\beta t_i \leq \ln(2C+1)$ for $i = 1,2$ which implies $\beta \in \left[\frac{\ln(2C+1)}{t_1}, \frac{\ln(2C+1)}{t_2}\right]$. Hence $\max\{|g(t_1,\cdot)|,|g(t_2,\cdot)|\}$ is subcompact for $g(t,\theta) = \alpha + \exp(\beta t)$.

Now consider $g(t,\theta) = \alpha t + \exp(\beta t)$. Again, let be $C \geq 0$ arbitrary. Then

$$\alpha t_i + e^{\beta t_i} \leq C \;\text{ for }\; i = 1,2 \tag{3}$$

implies $\alpha t_i \leq C - e^{\beta t_i} \leq C$ for $i = 1,2$ so that $\alpha \geq \frac{C}{t_1},\;\alpha \leq \frac{C}{t_2}$. Hence there exists $k \geq 0$ with $-k \leq \alpha \leq k$. With this k we obtain $-kt_1 \geq \alpha t_1 \geq kt_1$, and $-kt_2 \leq \alpha t_2 \leq kt_2$ so that

$$kt_1 \leq -\alpha t_1 \leq -kt_1,\;\; kt_2 \geq -\alpha t_2 \geq -kt_2. \tag{4}$$

Inequality (3) also implies $e^{\beta t_i} \leq C - \alpha t_i$ for $i = 1,2$. With (4) we obtain $e^{\beta t_1} \leq C - \alpha t_1 \leq C - kt_1$ and $e^{\beta t_2} \leq C - \alpha t_2 \leq C + kt_2$ so that $\beta t_1 \leq \ln(C - kt_1)$, $\beta t_2 \leq \ln(C + kt_2)$ and $\beta \geq \frac{\ln(C-kt_1)}{t_1}$, $\beta \leq \frac{\ln(C+kt_2)}{t_2}$. Hence, there exists $k' \geq 0$ with $-k' \leq$

$\beta \le k'$ so that $(\alpha,\beta)^\top \in [-k,k]\times[-k',k']$. This means that $\max\{|g(t_1,\cdot)|,|g(t_2,\cdot)|\}$ is subcompact for $g(t,\theta)=\alpha t+\exp(\beta t)$ as well.□

As soon as all experimental conditions are either negative or positive, then no subcompactness is possible. With loss of generality, we can consider only the case where all experimental conditions are positive.

Theorem 5. *If* $0 \le t_1 \le t_2 \le \ldots \le t_N$ *and* $g(t,\theta)=\alpha+\exp(\beta t)$ *or* $g(t,\theta)=\alpha t+\exp(\beta t)$ *with* $\theta=(\alpha,\beta)^\top \in \Theta = \Re\times\Re$ *then*

$$\max\{|g(t_n,\cdot)|;\ n=1,\ldots,N\}$$

is not subcompact. In particular, the fullness parameter of $\{l_n(y,D,\cdot);\ n=1,\ldots,N\}$ *is* $N+1$.

Proof. Set $\alpha=0$. Then $\exp(\beta t_n)\le C$ for all $n=1,\ldots,N$ is satisfies by $\beta \le \frac{\ln(C)}{t_N}$. Hence

$$\{0\}\times\left(-\infty,\frac{\ln(C)}{t_N}\right)\subset\{\theta;\ \max\{|g(t_n,\cdot)|;\ n=1,\ldots,N\}\le C\}$$

so that $\max\{|g(t_n,\cdot)|;\ n=1,\ldots,N\}$ is not subcompact.□

Now define for any design $D=(t_1,\ldots,t_N)\in\Re^N$

$$N^+(D):=\operatorname{card}\{t_n;\ t_n>0\}\ \text{ and }\ N^-(D):=\operatorname{card}\{t_n;\ t_n<0\}.$$

Corollary 1. *If* $g(t,\theta)=\alpha+\exp(\beta t)$ *or* $g(t,\theta)=\alpha t+\exp(\beta t)$ *with* $\theta=(\alpha,\beta)^\top\in\Theta=\Re\times\Re$ *and* $\min\{N^+(D),N^-(D)\}>0$ *then the fullness parameter of* $\{l_n(y,D,\cdot);\ n=1,\ldots,N\}$ *is given by*

$$\max\{N-N^+(D)+1,N-N^-(D)+1\}.$$

Since the fullness parameter should be as small as possible to maximize the lower bound for the breakdown point according to Theorem 1, a breakdown point maximizing design for the setup of Corollary 1 is a design with $N^+(D)=N^-(D)=\frac{N}{2}$. In this case, the lower bound for the breakdown point is approximately $\frac{1}{4}$.

However, in most applications, a nonnegative design region is assumed for a model like $g(t,\theta)=\alpha+\exp(\beta t)$ or $g(t,\theta)=\alpha t+\exp(\beta t)$. Then a fullness parameter less than $N+1$ and thus a lower bound for the breakdown point greater than 0 is only achieved if the parameter space is restricted. This situation is studied in the next section.

4 Nonlinear models with restricted parameter space

Considering $g(t,\theta) = \alpha + \exp(\beta t)$ or $g(t,\theta) = \alpha t + \exp(\beta t)$, it is enough to restrict the parameter space of β.

Theorem 6. *If* $0 \le t_1 < t_2$ *and* $g(t,\theta) = \alpha + \exp(\beta t)$ *or* $g(t,\theta) = \alpha t + \exp(\beta t)$ *with* $\theta = (\alpha,\beta)^\top \in \Theta = \Re \times [b,\infty)$ *and* $b \ge 0$ *then*

$$\max\{|g(t_1,\cdot)|, |g(t_2,\cdot)|\}$$

is subcompact.

Proof. Consider at first $g(t,\theta) = \alpha t + \exp(\beta t)$ and let be $C \in [0,\infty)$ arbitrary. Then

$$-C \le \alpha t_i + e^{\beta t_i} \le C \ \text{ for } \ i = 1,2 \tag{5}$$

implies $\alpha t_i \le C - e^{\beta t_i} \le C$, $\ \alpha t_i \ge -C - e^{\beta t_i}$ so that $\alpha \le \frac{C}{t_i}$ and

$$\alpha \ge \frac{1}{t_i}(-C - e^{\beta t_i}). \tag{6}$$

(6) means $-\alpha \le \frac{1}{t_i}(C + e^{\beta t_i})$ so that with (5) we obtain $e^{\beta t_j} \le C - \alpha t_j \le C + \frac{1}{t_i}(C + e^{\beta t_i}) t_j = C\left(1 + \frac{1}{t_i}\right) + \frac{t_j}{t_i} e^{\beta t_i}$. Dividing by $e^{\beta t_i} \ge 1$ ($t_i \ge 0, \beta \ge 0$) yields $e^{\beta (t_j - t_i)} \le \frac{C}{e^{\beta t_i}}\left(1 + \frac{1}{t_i}\right) + \frac{t_j}{t_i} \le C\left(1 + \frac{1}{t_i}\right) + \frac{t_j}{t_i}$ so that $\beta(t_j - t_i) \le \ln\left(C\left(1 + \frac{1}{t_i}\right) + \frac{t_j}{t_i}\right)$. With $t_j = t_2, t_i = t_1$ we obtain $\beta \le \frac{1}{t_2 - t_1} \ln\left(C\left(1 + \frac{1}{t_1}\right) + \frac{t_2}{t_1}\right) =: K_1$ because of $t_2 - t_1 > 0$. Inequality (6) provides then $\alpha \ge \frac{1}{t_i}(-C - e^{\beta t_i}) \ge \frac{1}{t_i}(-C - e^{K_1 t_i})$. Hence there exists $K_2 \ge 0$ such that $(\alpha,\beta)^\top \in [-K_2, K_2] \times [b, K_1]$.
The assertion for $g(t,\theta) = \alpha + \exp(\beta t)$ follows similarly.□

Theorem 6 means that $\{l_n(y,D,\cdot);\ n = 1,\ldots,N\}$ is 2-full if $0 \le t_1 < t_2 < \ldots < t_N$. In this case the lower bound for the breakdown point is approximately $\frac{N}{2}$ which is the maximum possible value for the lower bound. It is also obvious that repeated observation at the same experimental condition would reduce the breakdown point as Müller (1995) showed for linear models.

5 Discussion

Extension to nonlinear models with more than two parameters are possible and will be published elsewhere. However, all these results concern only the explosion point and not the implosion point, where condition (2) would be necessary in the definition of the breakdown point. In particular for restricted parameter spaces, the implosion point is of interest, in particular when the bound is 0. But using the condition (2) in the definition of subcompactness, as Vandev (1993) and Vandev and Neykov (1998) did, would not help. It seems that the d-fullness criterion is only useful for the explosion point.

References

[Hadi and Luceño (1997] Hadi, A. S. and A. Luceño (1997). Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Comput. Stat. Data Anal. 25*(3), 251–272.

[Jurečková and Procházka (1994] Jurečková, J. and B. Procházka (1994). Regression quantiles and trimmed least squares estimator in nonlinear regression model. *J. Nonparametric Stat. 3*(3-4), 201–222.

[Müller and Neykov (2003] Müller, Ch. H. and N. Neykov (2003). Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *J. Statist. Plann. Inference 116*, 503–519.

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

[Procházka (1988] Procházka, B. (1988). Regression quantiles and trimmed least squares estimator in the nonlinear regression model. *Comput. Stat. Data Anal. 6*(4), 385–391.

[Rousseeuw and Leroy (1987] Rousseeuw, P. J. and A. M. Leroy (1987). *Robust regression and outlier detection*. New York: John Wiley.

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

[Stromberg and Ruppert (1992] Stromberg, A. J. and D. Ruppert (1992). Breakdown in nonlinear regression. *J. Am. Stat. Assoc. 87*(420), 991–997.

[Vandev (1993] Vandev, D. (1993). A note on the breakdown point of the least median of squares and least trimmed squares estimators. *Stat. Probab. Lett. 16*(2), 117–119.

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