Upper and Lower Bounds for Breakdown Points

Christine Müller

Abstract General upper and lower bounds for the finite sample breakdown point are presented. The general upper bound is obtained by an approach of Davies and Gather (2005) using algebraic groups of transformations. It is shown that the upper bound for the finite sample breakdown point has a more simple form than for the population breakdown point. This result is applied to multivariate regression. It is shown that the upper bounds of the breakdown points of estimators of regression parameters, location and scatter can be obtained with the same group of transformations. The general lower bound for the breakdown point of some estimators is given via the concept of d-fullness introduced by Vandev (1993). This provides that the lower bound and the upper bound can coincide for least trimmed squares estimators for multivariate regression and simultaneous estimation of scale and regression parameter.

1 Introduction

The breakdown point of an estimator introduced by Hampel (1971) is a simple and successful measure of the robustness of an estimator against changes of the observations. In particular, it is easy to understand the finite sample version of the breakdown point. Estimators with a high breakdown point are insensitive to a high amount of outlying observations. Moreover, they can be used to detect observations which do not follow the majority of the data. Some estimators have a breakdown of 50% while in other situations the highest possible breakdown point is much smaller than 50%. Therefore it is always important to know what is the highest possible breakdown point. Then it can be checked whether specific estimators can reach this

Christine Müller
TU Dortmund University, Faculty of Statistics, D - 44221 Dortmund, e-mail: `cmueller@statistik.tu-dortmund.de`

upper bound. This can be done by deriving lower bounds for these estimators. Here general upper and lower bounds for the breakdown point are discussed.

Two finite sample breakdown point concepts are given in Section 2. In Section 3, a general upper bound is derived via the approach based on algebraic groups of transformations introduced by Davies and Gather (2005). While Davies and Gather (2005) developed this approach for the population version of the breakdown point, here this approach is transfered to the finite sample version of the breakdown point. This leads to a very simple characterization of the upper bound. Davies and Gather (2005) applied the approach to multivariate location and scatter estimation, univariate linear regression, logistic regression, the Michaelis-Menten model, and time series using different groups of transformations for each case. Regarding multivariate regression in Section 4, linear regression as well as multivariate location and scatter estimation can be treated here with the same approach. In particular the same group of transformations is used for the three cases. In Section 5, a general lower bound for the breakdown of some estimators based on the approach of d-fullness developed by Vandev (1993) is presented. With this approach, Müller and Neykov (2003) derived lower bounds for generalized linear models like logistic regression and log-linear models and Müller and Schäfer (2010) obtained lower bounds for some non-linear models. This approach is used here in Section 6 to provide lower bounds for multivariate regression and simultaneous estimation of the scale and regression parameter in univariate regression. It is shown in particular that least trimmed squares estimators are attaining the upper bounds derived in Section 4.

2 Definitions of Breakdown Points

Let be Θ a finite dimensional parameter space, $\mathbf{z}_1,\ldots,\mathbf{z}_N \in \mathscr{Z}$ a univariate or multivariate sample in $\mathscr{Z}$, and $\widehat{\theta}: \mathscr{Z}^N \longrightarrow \Theta$ an estimator for $\theta \in \Theta$. If $\text{int}(\Theta)$ denotes the interior of the parameter space, then a general definition of the finite sample breakdown is as follows, see e.g. Hampel et al. (1986), p. 97:

Definition 1. The breakdown point of an estimator $\widehat{\theta}: \mathscr{Z}^N \to \Theta$ at $\mathbf{Z} = (\mathbf{z}_1,\ldots,\mathbf{z}_N)^\top \in \mathscr{Z}^N$ is defined as

$$\begin{aligned}\varepsilon^*(\hat{\theta},\mathbf{Z}) := \frac{1}{N}\min\{M;\\ &\text{there exists no compact set } \Theta_0 \subset \text{int}(\Theta) \text{ with } \{\hat{\theta}(\overline{\mathbf{Z}});\ \overline{\mathbf{Z}} \in \mathscr{Z}_M(\mathbf{Z})\} \subset \Theta_0\},\end{aligned}$$

where

$$\mathscr{Z}_M(\mathbf{Z}) := \left\{\overline{\mathbf{Z}} \in \mathscr{Z}^N;\ \text{card}\{n;\ \mathbf{z}_n \neq \overline{\mathbf{z}}_n\} \leq M\right\}$$

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

Often the definition via compact subsets is replaced by a definition via explosion with respect to a special pseudometric d on Θ, see e.g. Donoho and Huber (1983), Davies and Gather (2005).

Definition 2.

$$\varepsilon^*(\hat{\theta}, \mathbf{Z}, d) := \frac{1}{N} \min \left\{ M; \sup_{\overline{\mathbf{Z}} \in \mathscr{Z}_M(\mathbf{Z})} d(\widehat{\theta}(\overline{\mathbf{Z}}), \widehat{\theta}(\mathbf{Z})) = \infty \right\}.$$

Is $\Theta = \mathbb{R}^p$, then the pseudometric can be chosen as the Euclidean metric $\|\cdot\|_p$. Is $\Theta = [0, \infty) \subset \mathbb{R}$, for example for scale parameters, then an appropriate choice for the pseudometric is $d(\theta_1, \theta_2) = |\log(\theta_1 \cdot \theta_2^{-1})|$, see Davies and Gather (2005). This is again a metric but its extension to scatter matrices is only a pseudometric, as is discussed in Section 4.2.

Davies and Gather (2005) used the population version of the breakdown point and not the finite sample version of Definition 2. But they pointed out that the finite sample version is obtained by using the empirical distribution. They provided a general upper bound for the population version of Definition 2 using transformation groups on the sample space $\mathscr{Z}$. Here this approach is given at once in the sample version.

3 A General Upper Bound

For the breakdown point of Definition 2, a general upper bound can be derived if the estimator $\widehat{\theta}$ is equivariant with respect to measurable transformations given by a group

$$\mathscr{G} := \{g;\, g : \mathscr{Z} \longrightarrow \mathscr{Z}\}.$$

Recall that $\mathscr{G}$ is a group in algebraic sense with actions $\circ$ and unit element ι if and only if

- $g_1 \circ g_2 \in \mathscr{G}$ for all $g_1,\ g_2 \in \mathscr{G}$,
- $\iota \circ g = g = g \circ \iota$ for all $g \in \mathscr{G}$,
- for every $g \in \mathscr{G}$ there exists g^{-1} with $g \circ g^{-1} = \iota = g^{-1} \circ g$.

Definition 3. An estimator $\widehat{\theta} : \mathscr{Z}^N \to \Theta$ is called equivariant with respect to a group $\mathscr{G}$ if there exists a group $\mathscr{H}_{\mathscr{G}} = \{h_g;\, g \in \mathscr{G}\}$ of transformations $h_g : \Theta \longrightarrow \Theta$ such that for every $g \in \mathscr{G}$ there exists $h_g \in \mathscr{H}_{\mathscr{G}}$ with

$$\widehat{\theta}((g(\mathbf{z}_1), \ldots, g(\mathbf{z}_n))^\top) = h_g(\widehat{\theta}((\mathbf{z}_1, \ldots, \mathbf{z}_n)^\top))$$

for all samples $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^\top \in \mathscr{Z}^N$.

To derive the upper bound for the breakdown point, the following subset of $\mathscr{G}$ is needed

$$\mathscr{G}_1 := \{g \in \mathscr{G};\ \lim_{k\to\infty} \inf_{\theta\in\Theta} d(\theta, h_{g^k}(\theta)) = \infty\}.$$

If $\mathscr{G}_1 = \emptyset$, then the group $\mathscr{G}$ is too small to produce transformed parameters arbitrarily far away from the original parameter.

Theorem 1. *If the estimator $\widehat{\theta} : \mathscr{Z}^N \to \Theta$ is equivariant with respect to $\mathscr{G}$ and $\mathscr{G}_1 \neq \emptyset$, then*

$$\varepsilon^*(\hat{\theta}, \mathbf{Z}, d) \leq \frac{1}{N} \left\lfloor \frac{N - \Delta(\mathbf{Z}) + 1}{2} \right\rfloor$$

for all $\mathbf{Z} \in \mathscr{Z}^N$, where

$$\Delta((\mathbf{z}_1, \ldots, \mathbf{z}_N)^\top) := \max\{\mathrm{card}\{n;\ g(\mathbf{z}_n) = \mathbf{z}_n\};\ g \in \mathscr{G}_1\}$$

and $\lfloor x \rfloor$ is the largest integer m with $m \leq x$.

Note the more simple form of the quantity $\Delta(\mathbf{Z})$ compared with its form in the population version given by Davies and Gather (2005).

Proof. Regard an arbitrary observation vector $\mathbf{Z}$. Let be $M = \left\lfloor \frac{N-\Delta(\mathbf{Z})+1}{2} \right\rfloor$ and $L = \Delta(\mathbf{Z})$. Then there exists $g \in \mathscr{G}_1$ so that without loss of generality $g(\mathbf{z}_n) = \mathbf{z}_n$ for $n = 1, \ldots, L$. Then we also have $g^k(\mathbf{z}_n) = g^{k-1}(\mathbf{z}_n) = \ldots = g^2(\mathbf{z}_n) = g \circ g(\mathbf{z}_n) = g(g(\mathbf{z}_n)) = g(\mathbf{z}_n) = \mathbf{z}_n$ for all $n = 1, \ldots, L$ and all integer k. Define $\tilde{\mathbf{Z}}^k$ and $\overline{\mathbf{Z}}^k$ for any integer k by

$$\begin{aligned} \tilde{\mathbf{z}}_n^k &= \mathbf{z}_n \quad \text{for} \quad n = 1, \ldots, L \text{ and } L+M+1, \ldots, N, \\ \tilde{\mathbf{z}}_n^k &= g^k(\mathbf{z}_n) \quad \text{for} \quad n = L+1, \ldots, L+M, \end{aligned}$$

and

$$\begin{aligned} \overline{\mathbf{z}}_n^k &= \mathbf{z}_n \quad \text{for} \quad n = 1, \ldots, L+M, \\ \overline{\mathbf{z}}_n^k &= g^{-k}(\mathbf{z}_n) \quad \text{for} \quad n = L+M+1, \ldots, N. \end{aligned}$$

Obviously, $\tilde{\mathbf{Z}}^k \in \mathscr{Z}_M(\mathbf{Z})$. Since $N - (L+M) = N - L - \left\lfloor \frac{N-L+1}{2} \right\rfloor \leq N - L - \frac{N-L}{2} = \frac{N-L}{2} \leq \left\lfloor \frac{N-L+1}{2} \right\rfloor$, we also have $\overline{\mathbf{Z}}^k \in \mathscr{Z}_M(\mathbf{Z})$. Moreover, it holds

$$\begin{aligned} g^k(\overline{\mathbf{z}}_n^k) &= g^k(\mathbf{z}_n) = \mathbf{z}_n = \tilde{\mathbf{z}}_n^k \quad \text{for } n = 1, \ldots, L, \\ g^k(\overline{\mathbf{z}}_n^k) &= g^k(\mathbf{z}_n) = \tilde{\mathbf{z}}_n^k \quad \text{for } n = L+1, \ldots, L+M, \\ g^k(\overline{\mathbf{z}}_n^k) &= g^k(g^{-k}(\mathbf{z}_n)) = g^k \circ g^{-k}(\mathbf{z}_n) = \mathbf{z}_n = \tilde{\mathbf{z}}_n^k \quad \text{for } n = L+M+1, \ldots, N. \end{aligned}$$

Since $g \in \mathscr{G}_1$ and $\widehat{\theta}((g^k(\overline{\mathbf{z}}_1^k), \ldots, g^k(\overline{\mathbf{z}}_n^k))^\top) = h_{g^k}(\widehat{\theta}((\overline{\mathbf{z}}_1^k, \ldots, \overline{\mathbf{z}}_N^k)^\top))$, we obtain

$$\lim_{k\to\infty} d(\widehat{\theta}(\tilde{\mathbf{Z}}^k), \widehat{\theta}(\overline{\mathbf{Z}}^k))$$
$$= \lim_{k\to\infty} d(\widehat{\theta}((g^k(\bar{\mathbf{z}}_1^k), \ldots, g^k(\bar{\mathbf{z}}_n^k))^\top), \widehat{\theta}(\overline{\mathbf{Z}}^k)) = \lim_{k\to\infty} d(h_{g^k}(\widehat{\theta}(\overline{\mathbf{Z}}^k)), \widehat{\theta}(\overline{\mathbf{Z}}^k)) = \infty.$$

Because of $d(\widehat{\theta}(\tilde{\mathbf{Z}}^k), \widehat{\theta}(\overline{\mathbf{Z}}^k)) \le d(\widehat{\theta}(\tilde{\mathbf{Z}}^k), \widehat{\theta}(\mathbf{Z})) + d(\widehat{\theta}(\mathbf{Z}), \widehat{\theta}(\overline{\mathbf{Z}}^k))$, at least one of $d(\widehat{\theta}(\tilde{\mathbf{Z}}^k), \widehat{\theta}(\mathbf{Z}))$ and $d(\widehat{\theta}(\mathbf{Z}), \widehat{\theta}(\overline{\mathbf{Z}}^k))$ must converge to ∞ for $k \longrightarrow \infty$ as well. $\square$

4 Example: Multivariate Regression

The multivariate regression model is given by

$$\mathbf{y}_n^\top = \mathbf{x}_n^\top \mathbf{B} + \mathbf{e}_n^\top,$$

where $\mathbf{y}_n \in I\!R^p$ is the observation vector, $\mathbf{x}_n \in I\!R^r$ the known regression vector, $\mathbf{B} \in I\!R^{r\times p}$ the unknown parameter matrix and $\mathbf{e}_n \in I\!R^p$ the error vector. Set $\mathbf{z} = (\mathbf{x}^\top, \mathbf{y}^\top)^\top \in \mathscr{Z} = I\!R^{r+p}$ and assume that $\mathbf{e}_1, \ldots, \mathbf{e}_N$ are realizations of i.i.d. random variables $E_1, \ldots, E_N$ with location parameter $\mathbf{0}_p$ and scatter matrix Σ, where $\mathbf{0}_p$ denotes the p-dimensional vector of zeros. The interesting aspect of $\mathbf{B}$ shall be the linear aspect $\Lambda = \mathbf{L}\,\mathbf{B}$ with $\mathbf{L} \in I\!R^{s\times r}$. We consider here the problem of estimating Λ in Section 4.1 and of estimating Σ in Section 4.2.

In both cases, we can use the following group of transformations

$$\mathscr{G} = \{g_{\mathbf{A},\mathbf{B}} : \mathscr{Z} \to \mathscr{Z};\ \mathbf{A} \in I\!R^{p\times p} \text{ is regular},\ \mathbf{B} \in I\!R^{r\times p}\}$$

with $g_{\mathbf{A},\mathbf{B}}((\mathbf{x}^\top, \mathbf{y}^\top)^\top) = (\mathbf{x}^\top, \mathbf{y}^\top\mathbf{A} + \mathbf{x}^\top\mathbf{B})^\top$. The unit element of this group is $\iota = g_{\mathbf{I}_p, \mathbf{0}_{r\times p}}$, where $\mathbf{0}_{r\times p}$ is the $r \times p$-dimensional zero matrix and $\mathbf{I}_p$ the p-dimensional identity matrix. The inverse of $g_{\mathbf{A},\mathbf{B}}$ is given by $g_{\mathbf{A}^{-1}, -\mathbf{B}\mathbf{A}^{-1}}$.

4.1 Estimation of a linear aspect of the regression parameters

An estimator $\widehat{\theta} = \widehat{\Lambda} : \mathscr{Z}^N \to I\!R^{s\times p}$ for $\theta = \Lambda = \mathbf{L}\,\mathbf{B} \in I\!R^{s\times p}$ should be scatter equivariant and translation equivariant, i.e. it should satisfy

$$\widehat{\Lambda}((g_{\mathbf{A},\mathbf{B}}(\mathbf{z}_1), \ldots, g_{\mathbf{A},\mathbf{B}}(\mathbf{z}_n))^\top) = h_{g_{\mathbf{A},\mathbf{B}}}(\widehat{\Lambda}(\mathbf{Z}))$$

with $h_{g_{\mathbf{A},\mathbf{B}}}(\Lambda) = \Lambda\,\mathbf{A} + \mathbf{L}\mathbf{B}$ for all $g_{\mathbf{A},\mathbf{B}} \in \mathscr{G}$. With $\mathscr{G}$, also

$$\mathscr{H}_{\mathscr{G}} = \{h_{g_{\mathbf{A},\mathbf{B}}} : I\!R^{s\times p} \to I\!R^{s\times p};\ \mathbf{A} \in I\!R^{p\times p} \text{ is regular},\ \mathbf{B} \in I\!R^{r\times p}\}$$

is a group of transformations.

If $\mathbf{L}\mathbf{B} = \mathbf{0}_{s\times p}$, then $\Lambda = \mathbf{0}_{s\times p}$ satisfies

$$d(\Lambda, h_{g^n_{\mathbf{A},\mathbf{B}}}(\Lambda)) = d(\mathbf{0}_{s\times p}, \mathbf{0}_{s\times p}\mathbf{A}^n) = d(\mathbf{0}_{s\times p}, \mathbf{0}_{s\times p}) = 0$$

for any pseudometric d on $I\!R^{s\times p}$. Hence $\mathbf{LB} \neq \mathbf{0}_{s\times p}$ is necessary for $g_{\mathbf{A},\mathbf{B}} \in \mathscr{G}_1$. Moreover, we have $\Lambda = h_{g_{\mathbf{A},\mathbf{B}}}(\Lambda) = \Lambda\,\mathbf{A} + \mathbf{LB}$ if and only if $\mathbf{LB} = \Lambda\,(\mathbf{I}_p - \mathbf{A})$ so that

$$\mathscr{G}_1 = \{g_{\mathbf{A},\mathbf{B}} \in \mathscr{G};\ \mathbf{LB} \neq \mathbf{0}_{s\times p} \text{ and } \mathbf{LB} \neq \Lambda\,(\mathbf{I}_p - \mathbf{A}) \text{ for all } \Lambda \in I\!R^{s\times p}\}.$$

Set $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^\top$. Now we are going to show that $\Delta(\mathbf{Z})$ is the maximum number of regressors $\mathbf{x}_n$ so that the univariate linear aspect $\mathbf{L}\beta$ with $\beta \in I\!R^r$ is not identifiable at these regressors, i.e. $\Delta(\mathbf{Z})$ is the nonidentifiability parameter $\mathscr{N}_\lambda(\mathbf{X})$ defined in Müller (1995) for univariate regression, see also Müller (1997).

Definition 4. $\mathbf{L}\beta$ is identifiable at $D = \{\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_I}\}$ if and only if for all $\beta \in I\!R^r$

$$\mathbf{x}_{n_i}^\top \beta = 0 \text{ for } i = 1, \ldots, I \text{ implies } \mathbf{L}\beta = 0.$$

If $\mathbf{X}_D = (\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_I})^\top$, then it is well known that $\mathbf{L}\beta$ is identifiable at $D = \{\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_I}\}$ if and only if $\mathbf{L} = \mathbf{K}\mathbf{X}_D$ for some $\mathbf{K} \in I\!R^{s\times I}$, see e.g. Müller (1997), p. 6.

Definition 5. The nonidentifiability parameter $\mathscr{N}_\lambda(\mathbf{X})$ for estimating $\lambda = \mathbf{L}\beta$ in univariate regression, i.e. $\beta \in I\!R^r$, is defined as

$$\begin{aligned}\mathscr{N}_\lambda(\mathbf{X}) &:= \max\{\text{card}\{n;\ \mathbf{x}_n^\top \beta = 0\};\ \beta \in I\!R^r \text{ with } \lambda = \mathbf{L}\beta \neq 0\}\\ &= \max\{\text{card}\, D;\ \lambda = \mathbf{L}\beta \text{ is not identifiable at } D\}.\end{aligned}$$

Theorem 2.

$$\Delta(\mathbf{Z}) = \mathscr{N}_\lambda(\mathbf{X}).$$

Proof. Let be $g_{\mathbf{A},\mathbf{B}} \in \mathscr{G}_1$ and assume that there exists $\mathbf{z}_{n_1}, \ldots, \mathbf{z}_{n_I}$ with $g_{\mathbf{A},\mathbf{B}}(\mathbf{z}_{n_i}) = \mathbf{z}_{n_i}$ for $i = 1, \ldots, I$.

If $\mathbf{A} = \mathbf{I}_p$, then it holds $g_{\mathbf{A},\mathbf{B}}(\mathbf{z}) = \mathbf{z} = (\mathbf{x}^\top, \mathbf{y}^\top)^\top$ if and only if $\mathbf{x}^\top\mathbf{B} = \mathbf{0}_{1\times p}$ so that $\Delta(\mathbf{Z}) \geq \max\{\text{card}\{n;\ \mathbf{x}_n^\top \beta = 0\};\ \beta \in I\!R^p \text{ with } \mathbf{L}\beta \neq 0\}$ since $\mathbf{LB} \neq 0$ for $g_{\mathbf{A},\mathbf{B}} \in \mathscr{G}_1$. In this case, $\mathbf{LB} \neq \Lambda\,(\mathbf{I}_{p\times p} - \mathbf{A})$ is always satisfied for all $\Lambda \in I\!R^{s\times p}$ so that it is no restriction.

Now consider $\mathbf{A} \neq \mathbf{I}_p$. Assume that $\mathbf{L}\beta$ is identifiable at $D = \{\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_I}\}$ with $I = \Delta(\mathbf{Z})$. Then there exists $\mathbf{K} \in I\!R^{s\times I}$ such that $\mathbf{L} = \mathbf{K}\mathbf{X}_D$. Set $\mathbf{Y}_D = (\mathbf{y}_{n_1}, \ldots, \mathbf{y}_{n_I})^\top$. Since $g_{\mathbf{A},\mathbf{B}}(\mathbf{z}_{n_i}) = \mathbf{z}_{n_i}$ if and only if $\mathbf{x}_{n_i}^\top\mathbf{B} = \mathbf{y}_{n_i}^\top(\mathbf{I_p} - \mathbf{A})$, we obtain the contradiction

$$\mathbf{LB} = \mathbf{K}\mathbf{X}_D\mathbf{B} = \mathbf{K}\begin{pmatrix}\mathbf{x}_{n_1}^\top\mathbf{B}\\ \vdots\\ \mathbf{x}_{n_I}^\top\mathbf{B}\end{pmatrix} = \mathbf{K}\begin{pmatrix}\mathbf{y}_{n_1}^\top(\mathbf{I}_p - \mathbf{A})\\ \vdots\\ \mathbf{y}_{n_I}^\top(\mathbf{I}_p - \mathbf{A})\end{pmatrix} = \mathbf{K}\,\mathbf{Y}_D\,(\mathbf{I}_p - \mathbf{A})$$

since $g_{\mathbf{A},\mathbf{B}} \in \mathscr{G}_1$. This means that $\mathbf{L}\beta$ cannot be identifiable at $D = \{\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_I}\}$ so that $\Delta(\mathbf{Z}) = I \leq \max\{\text{card}\{n;\ \mathbf{x}_n^\top \beta = 0\};\ \beta \in I\!R^p \text{ with } \mathbf{L}\beta \neq 0\}$. □

From the proof of Theorem 2 it is clear that the assertion of Theorem 2 holds also without using the scatter equivariance of the estimator $\widehat{\Lambda}$. See also Sections 4.1.1 and 4.1.2.

4.1.1 Location model

A special case of multivariate regression is multivariate location with $\mathbf{x}_n = 1$ for all $n = 1, \ldots, N$, where $\mathbf{B} \in I\!R^{1\times p}$ is the parameter of interest. In this case, identifiability holds always so that $\Delta(\mathbf{Z}) = 0$. Hence we have the highest possible upper point of $\frac{1}{N}\left\lfloor\frac{N+1}{2}\right\rfloor$. This result was obtained by Davies and Gather (2005) using only the translation group

$$\mathscr{G}^L = \{g_{\mathbf{I}_p,\mathbf{B}} : \mathscr{Z} \to \mathscr{Z};\ \mathbf{B} \in I\!R^{1\times p}\}$$

so that

$$\mathscr{G}_1^L = \{g_{\mathbf{I}_p,\mathbf{B}} \in \mathscr{G}^L;\ \mathbf{B} \neq \mathbf{0}_{1\times p}\}.$$

They wrote in their rejoinder that $\mathscr{G}_1^L$ would be empty if scatter transformations are considered as well. But $\mathscr{G}_1^L$ becomes larger if a larger group of transformation is regarded.

For the special case of univariate data, i.e. $p = 1$, with location parameter $\mu \in I\!R$, the condition

$$\mathbf{0}_{s\times p} \neq \mathbf{LB} \neq \Lambda\,(\mathbf{I}_p - \mathbf{A}) \text{ for all } \Lambda \in I\!R^{1\times p} \tag{1}$$

becomes

$$0 \neq b \neq \mu(1-a) \text{ for all } \mu \in I\!R, \tag{2}$$

where $a, b \in I\!R$ replace $\mathbf{A}$ and $\mathbf{B}$. Since condition (2) is only satisfied for $a = 1$ we have

$$\mathscr{G}_1^L = \mathscr{G}_1$$

so that indeed it does not matter if the scatter (here scale) equivariance is additionally demanded.

For univariate data the upper bound $\frac{1}{N}\left\lfloor\frac{N+1}{2}\right\rfloor$ is attained by the median. A multivariate extension of the median, which is scatter and translation equivariant, is Tukey's half space median. But its breakdown point lies only between $\frac{1}{p+1}$ and $\frac{1}{3}$, see Donoho and Gasko (1992). As far as the author knows, there is no scatter and translation equivariant location estimator which attains the upper bound for $p > 1$.

4.1.2 Univariate regression

Another special case of multivariate regression is univariate regression with $p = 1$, where the unknown parameter $\mathbf{B}$ is $\beta \in I\!R^r$. The result

$$\Delta(\mathbf{Z}) = \mathcal{N}_\lambda(\mathbf{X})$$

was obtained by Müller (1995) using only the transformations $g_{\mathbf{b}}((\mathbf{x}^\top, y)^\top) = (\mathbf{x}^\top, y + \mathbf{x}^\top \mathbf{b})^\top$ in a proof similar to that of Theorem 1, see also Müller (1997).

The special case $\mathbf{L}\beta = \beta$ was considered by Davies (1993) who derived the upper bound for the population version of the breakdown point. Davies and Gather (2005) provided this result as an example of the approach via groups.

Using the translation group

$$\mathcal{G}^R = \{g_{\mathbf{b}} : \mathcal{Z} \to \mathcal{Z};\ \mathbf{b} \in \mathbb{R}^r\}$$

with $g_{\mathbf{b}}((\mathbf{x}^\top, y)^\top) = (\mathbf{x}^\top, y + \mathbf{x}^\top \mathbf{b})^\top$, as Davies and Gather (2005) did, leads to

$$\mathcal{G}_1^R = \{g_{\mathbf{b}} \in \mathcal{G}^R;\ \mathbf{b} \neq \mathbf{0}_r\}.$$

But since condition (1) becomes here

$$\mathbf{0}_r \neq \mathbf{b} \neq \beta(1-a) \text{ for all } \beta \in \mathbb{R}^r,$$

with $\mathbf{b} \in \mathbb{R}^r$ and $a \in \mathbb{R}$, it is again only satisfied for $a = 1$ so that

$$\mathcal{G}_1^R = \mathcal{G}_1.$$

Hence as for location estimation, the restriction to translations is no real restriction here.

4.2 Scatter estimation

An estimator $\widehat{\theta} = \widehat{\Sigma} : \mathcal{Z}^N \to \mathcal{S}$ of the scatter matrix $\Sigma \in \mathcal{S} = \{\mathbf{A} \in \mathbb{R}^{p\times p};\ \mathbf{A}$ is symmetric and positive definite$\}$, should be scatter equivariant and translation invariant, i.e. it should satisfy

$$\widehat{\Sigma}((g_{\mathbf{A},\mathbf{B}}(\mathbf{z}_1), \ldots, g_{\mathbf{A},\mathbf{B}}(\mathbf{z}_n))^\top) = h_{g_{\mathbf{A},\mathbf{B}}}(\widehat{\Sigma}(\mathbf{Z}))$$

with $h_{g_{\mathbf{A},\mathbf{B}}}(\Sigma) = \mathbf{A}\,\Sigma\,\mathbf{A}^\top$ for all $g_{\mathbf{A},\mathbf{B}} \in \mathcal{G}$. With $\mathcal{G}$, also

$$\mathcal{H}_{\mathcal{G}} = \{h_{g_{\mathbf{A},\mathbf{B}}} = h_{\mathbf{A}} : \mathcal{S} \to \mathcal{S};\ \mathbf{A} \in \mathbb{R}^{p\times p} \text{ is regular }\}$$

is a group of transformations. An appropriate pseudometric on $\mathcal{S}$ is given by

$$d(\Sigma_1, \Sigma_2) := |\log(\det(\Sigma_1\,\Sigma_2^{-1}))|.$$

It holds $d(\Sigma_1, \Sigma_2) = 0$ if and only if $\det(\Sigma_1\,\Sigma_2^{-1}) = 1$. This is not only satisfied by $\Sigma_1 = \Sigma_2$, since e.g. diagonal matrices like $\operatorname{diag}(1,1)$ and $\operatorname{diag}(\frac{1}{2}, 2)$ are satisfying this as well. Hence d is not a metric. But it is a pseudometric because it is always greater than 0 and it satisfies the triangle inequality. Since $\det(\mathbf{A}\Sigma_1\mathbf{A}^\top \Sigma_2^{-1}) = \det(\Sigma_1 \Sigma_2^{-1})$ as soon as $\det(\mathbf{A}) = 1$, $\mathscr{G}_1$ is given by

$$\mathscr{G}_1 = \{g_{\mathbf{A},\mathbf{B}} \in \mathscr{G};\ \det(\mathbf{A}) \neq 1\}.$$

Since $g_{\mathbf{A},\mathbf{B}}(\mathbf{z}) = \mathbf{z}$ if and only if $\mathbf{x}^\top\mathbf{B} = \mathbf{y}^\top(\mathbf{I}_p - \mathbf{A})$, we have at once the following theorem.

Theorem 3.

$$\Delta(\mathbf{Z}) = \max\{\text{card}\{n;\ \mathbf{x}_n^\top\mathbf{B} = \mathbf{y}_n^\top(\mathbf{I}_p - \mathbf{A})\};\\ \mathbf{B} \in \mathbb{R}^{r\times p},\ \mathbf{A} \in \mathbb{R}^{p\times p} \textit{ is regular with } \det(\mathbf{A}) \neq 1\}.$$

4.2.1 Location model

In the special case of multivariate location with $\mathbf{x}_n = 1$ for all $n = 1,\ldots,N$ and $\mathbf{B} \in \mathbb{R}^{1\times p}$ we have $g_{\mathbf{A},\mathbf{B}}(\mathbf{z}) = \mathbf{z}$ if and only if $\mathbf{B} = \mathbf{y}^\top(\mathbf{I}_p - \mathbf{A})$. Hence $\{\mathbf{y} \in \mathbb{R}^p;\ \mathbf{B} = \mathbf{y}^\top(\mathbf{I}_p - \mathbf{A})\}$ is a hyperplane in $\mathbb{R}^p$. Conversely, if $\{\mathbf{y} \in \mathbb{R}^p;\ \mathbf{c}^\top = \mathbf{y}^\top\mathbf{C}\}$ is an arbitrary hyperplane in $\mathbb{R}^p$, then it can be assumed that $\det(\mathbf{I}_p - \mathbf{C}) \neq 1$ so that $g_{\mathbf{I}_p-\mathbf{C},\mathbf{c}^\top} \in \mathscr{G}_1$. This implies that $\Delta(\mathbf{Z})$ is the maximum number of observations lying in a hyperplane. According to Theorem 1, the upper bound of the breakdown point of an equivariant scatter estimator is given by the maximum number of observations in a hyperplane. It attains its highest value for observations in general position. But since p points are lying in the hyperplane of $\mathbb{R}^p$ spanned by these points, an upper bound for the breakdown point is always $\frac{1}{N}\left\lfloor\frac{N-p+1}{2}\right\rfloor$. The population version of this result was originally given by Davies (1993) and derived by group equivariance in Davies and Gather (2005).

For the one-dimensional case ($p = 1$), it means that the upper bound of the breakdown point of a scale equivariant and translation invariant scale estimator is determined by the maximum number of repeated observations. Here the highest value of the upper bound is given by pairwise different observations. This highest upper bound is for example attained by the median absolute deviation (MAD). However, it can happen that the upper bound is not attained by the median absolute deviation if observations are repeated. Davies and Gather (2007) gave the following example

1.0, 1.8, 1.3, 1.3, 1.9, 1.1, 1.3, 1.6, 1.7, 1.3, 1.3.

The median absolute deviation of this sample is 0.2. But as soon as one observation unequal to 1.3 is replaced by 1.3, the median absolute deviation is 0. Hence the breakdown point of this sample is $\frac{1}{11}$. However, since 1.3 is repeated five times, the upper bound for the breakdown point is

$$\frac{1}{11}\left\lfloor\frac{11-5+1}{2}\right\rfloor = \frac{3}{11}.$$

4.2.2 Univariate regression

In the special case of univariate regression with $p = 1$, the condition

$$\mathbf{x}^\top \mathbf{B} = \mathbf{y}^\top (\mathbf{I}_p - \mathbf{A})$$

becomes

$$\mathbf{x}^\top \beta = y(1-a) \Longleftrightarrow y = \mathbf{x}^\top \tilde{\beta}$$

with $\beta \in I\!R^r$, $1 \neq a \in I\!R$ and $\tilde{\beta} = \frac{1}{1-a}\beta$. This means that $\Delta(\mathbf{Z})$ is the maximum number $\mathscr{E}(\mathbf{X})$ of observations satisfying an exact fit.

Definition 6. The exact fit parameter is defined as

$$\mathscr{E}(\mathbf{X}) := \max\{\text{card}\{n;\, y_n = \mathbf{x}_n^\top \beta\};\, \beta \in I\!R^p\}.$$

Hence we have here

$$\Delta(\mathbf{Z}) = \mathscr{E}(\mathbf{X}).$$

5 A general lower bound for some estimators

Since there are always estimators with a breakdown point of $\frac{1}{N}$ or even 0, a lower bound can be only valid for some special estimators. He we consider estimators of the form

$$\widehat{\theta}(\mathbf{Z}) := \arg\min_{\theta \in \Theta} s(\mathbf{Z}, \theta)$$

with $s : \mathscr{Z}^N \times \Theta \to I\!R$, where $s(\mathbf{Z}, \theta)$ can be bounded from below and above by some quality functions $q : \mathscr{Z} \times \Theta \to I\!R$. These quality functions can be residuals but also some negative loglikelihood functions as considered in Müller and Neykov (2003). Set $q_n(\mathbf{Z}, \theta) = q(\mathbf{z}_n, \theta)$ for $n = 1, \ldots, N$ and $q_{(1)}(\mathbf{Z}, \theta) \leq \ldots \leq q_{(N)}(\mathbf{Z}, \theta)$. Then there shall exists $\alpha,\ \beta \in I\!R$ with $\alpha \neq 0$ and $h \in \{1, \ldots, N\}$ such that

$$\alpha\, q_{(h)}(\mathbf{Z}, \theta) \leq s(\mathbf{Z}, \theta) \leq \beta\, q_{(h)}(\mathbf{Z}, \theta) \tag{3}$$

for all $\mathbf{Z} \in \mathscr{Z}^N$ and $\theta \in \Theta$. In particular, h-trimmed estimators given by

$$\widehat{\theta}_h(\mathbf{Z}) := \arg\min_{\theta \in \Theta} \sum_{n=1}^{h} q_{(n)}(\mathbf{Z}, \theta)$$

are satisfying condition (3). But also S-estimators are satisfying this condition, see e.g. Rousseeuw and Leroy (2003), pp. 135-139.

For deriving the lower bound for the breakdown point, the Definition 1 for the breakdown point is used. This definition is checking whether the estimators are remaining in a compact subset of the parameter space. Via compact sets, Vandev (1993) developed the concept of d-fullness which was used by Vandev and Neykov

(1998) to estimate this breakdown point for trimmed estimators. A modification of this concept, used in Müller and Neykov (2003), bases on the following definitions.

Definition 7. A function $\gamma : \Theta \to I\!R$ is called sub-compact if the set $\{\theta \in \Theta;\ \gamma(\theta) \le c\}$ is contained in a compact set $\Theta_c \subset \text{int}(\Theta)$ for all $c \in I\!R$.

Definition 8. A finite set $\Gamma = \{\gamma_n : \Theta \to I\!R;\ n = 1,\dots,N\}$ of functions is called *d*-full if for every $\{n_1,\dots,n_d\} \subset \{1,\dots,N\}$ the function γ given by $\gamma(\theta) := \max\{\gamma_{n_k}(\theta);\ k = 1,\dots,d\}$ is sub-compact.

At first note the following lemma of Müller and Neykov (2003).

Lemma 1. *If $\{q_n(\mathbf{Z},\cdot);\ n = 1,\dots,N\}$ is d-full, $M \le N-h$, and $M \le h-d$, then $q_{(d)}(\mathbf{Z},\theta) \le q_{(h)}(\overline{\mathbf{Z}},\theta) \le q_{(N)}(\mathbf{Z},\theta)$ for all $\overline{\mathbf{Z}} \in \mathscr{Y}_M(\mathbf{Z})$ and $\theta \in \Theta$.*

Proof. Regard $n_1,\dots,n_h$ with $q_{(k)}(\overline{\mathbf{Z}},\theta) = q_{n_k}(\overline{\mathbf{Z}},\theta)$ for $k = 1,\dots,h$. Since $h \ge M + d$ we have $1 \le k(1) < \dots < k(d) \le h$ with $q_{n_{k(i)}}(\overline{\mathbf{Z}},\theta) = q_{n_{k(i)}}(\mathbf{Z},\theta)$. Then we obtain

$$q_{(h)}(\overline{\mathbf{Z}},\theta) = q_{n_h}(\overline{\mathbf{Z}},\theta) \ge q_{n_{k(d)}}(\overline{\mathbf{Z}},\theta) \ge q_{n_{k(i)}}(\overline{\mathbf{Z}},\theta) = q_{n_{k(i)}}(\mathbf{Z},\theta)$$

for all $i = 1,\dots,d$. This implies $q_{(h)}(\overline{\mathbf{Z}},\theta) \ge q_{(d)}(\mathbf{Z},\theta)$. The other inequality follows similarly. □

Theorem 4 (Müller and Neykov (2003)). *If the estimator $\widehat{\theta}$ satisfies condition (3) and $\{q_n(\mathbf{Z},\cdot);\ n = 1,\dots,N\}$ is d-full, then*

$$\varepsilon^*(\widehat{\theta},\mathbf{Z}) \ge \frac{1}{N}\min\{N-h+1, h-d+1\}.$$

Proof. Let $M = \min\{N-h, h-d\}$. Lemma 1 together with assumption (3) provide that

$$\alpha\, q_{(d)}(\mathbf{Z},\theta) \le s(\overline{\mathbf{Z}},\theta) \le \beta\, q_{(N)}(\mathbf{Z},\theta)$$

for all $\overline{\mathbf{Z}} \in \mathscr{Z}_M(\mathbf{Z})$ and $\theta \in \Theta$. This means

$$\alpha\, q_{(d)}(\mathbf{Z},\widehat{\theta}(\overline{\mathbf{Z}})) \le s(\overline{\mathbf{Z}},\widehat{\theta}(\overline{\mathbf{Z}})) = \min_{\theta} s(\overline{\mathbf{Z}},\theta) \le \beta \min_{\theta} q_{(N)}(\mathbf{Z},\theta)$$

for all $\overline{\mathbf{Z}} \in \mathscr{Z}_M(\mathbf{Z})$. Setting $c_0 := \frac{\beta}{\alpha}\min_\theta q_{(N)}(\mathbf{Z},\theta)$ we have $\{\widehat{\theta}(\overline{\mathbf{Z}});\ \overline{\mathbf{Z}} \in \mathscr{Z}_M(\mathbf{Z})\} \subset \{\theta \in \Theta;\ q_{(d)}(\mathbf{Z},\theta) \le c_0\}$ so that we have only to show that γ given by

$$\begin{aligned}\gamma(\theta) &:= q_{(d)}(\mathbf{Z},\theta) = \max\{q_{(1)}(\mathbf{Z},\theta),\dots,q_{(d)}(\mathbf{Z},\theta)\}\\ &= \max\{q_{n_1(\theta)}(\mathbf{Z},\theta),\dots,q_{n_d(\theta)}(\mathbf{Z},\theta)\}\end{aligned}$$

is sub-compact. Assume that this is not the case. Then there exists $c \in I\!R$ such that $\{\theta;\ \gamma(\theta) \le c\}$ is not contained in a compact set. Hence, there exists a sequence $(\theta_m)_{m\in I\!N} \in \{\theta;\ \gamma(\theta) \le c\}$ such that every subsequence of $(\theta_m)_{m\in I\!N}$ is not converging. Because of $\{n_1(\theta_m),\dots,n_d(\theta_m)\} \subset \{1,\dots,N\}$ we have a subsequence $(\theta_{m(k)})_{k\in\mathbb{N}}$ and $n_1,\dots,n_d$ such that $\{n_1(\theta_{m(k)}),\dots,n_d(\theta_{m(k)})\} = \{n_1,\dots,n_d\}$ for all $k \in \mathbb{N}$. This implies $\gamma(\theta_{m(k)}) = \max\{q_{n_1}(\mathbf{Z},\theta_{m(k)}),\dots,q_{n_d}(\mathbf{Z},\theta_{m(k)})\} \le c$ for all $k \in \mathbb{N}$. However, $\max\{q_{n_1}(\mathbf{Z},\cdot),\dots,q_{n_d}(\mathbf{Z},\cdot)\}$ is sub-compact since $\{q_1(\mathbf{Z},\cdot),\dots,q_N(\mathbf{Z},\cdot)\}$ 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. □

The lower bound of Theorem 4 is maximized if the trimming factor h satisfies $\lfloor \frac{N+d}{2} \rfloor \le h \le \lfloor \frac{N+d+1}{2} \rfloor$. A simple consequence of this fact is the following result concerning trimmed estimators.

Theorem 5. *Assume that* $\{q_n(\mathbf{Z},\cdot);\, n=1,\dots,N\}$ *is* d*-full and* $\lfloor \frac{N+d}{2} \rfloor \le h \le \lfloor \frac{N+d+1}{2} \rfloor$. *Then the breakdown point of any trimmed estimator* $\widehat{\theta}_h$ *satisfies*

$$\varepsilon^*(\widehat{\theta}_h,\mathbf{Z}) \ge \frac{1}{N}\left\lfloor \frac{N-d+2}{2} \right\rfloor.$$

6 Example: Regression

6.1 Multivariate regression

Consider again multivariate regression with $x \in \mathbb{R}^r$, $y \in \mathbb{R}^p$ and unknown matrix $\mathbf{B} \in \mathbb{R}^{r\times p}$ of regression parameters. An appropriate quality function for estimating $\mathbf{B}$ is given by

$$q(\mathbf{z},\theta) = q(\mathbf{x},\mathbf{y},\mathbf{B}) = \|\mathbf{y}-\mathbf{B}^\top\mathbf{x}\|_p^2 = (\mathbf{y}^\top - \mathbf{x}^\top\mathbf{B})\,(\mathbf{y}-\mathbf{B}^\top\mathbf{x}). \tag{4}$$

The h-trimmed estimator $\widehat{\mathbf{B}}$ for $\mathbf{B}$ can be determined by calculating the least squares estimator

$$\widehat{\mathbf{B}}_I(\mathbf{Y}) = (\mathbf{X}_I^\top\mathbf{X}_I)^{-1}\mathbf{X}_I^\top\mathbf{Y}_I$$

for each subsample $I = \{n_1,\dots,n_h\} \subset \{1,\dots,N\}$ for which the inverse of $\mathbf{X}_I^\top\mathbf{X}_I$ exists, where $\mathbf{X}_I = (\mathbf{x}_{n_1},\dots,\mathbf{x}_{n_h})^\top$ and $\mathbf{Y}_I = (\mathbf{y}_{n_1},\dots,\mathbf{y}_{n_h})^\top$. Then $\widehat{\mathbf{B}}(\mathbf{Y})$ is that $\widehat{\mathbf{B}}_{I_*}(\mathbf{Y}_{I_*})$ with

$$I_* = \arg\min\left\{\sum_{j=1}^h \|\mathbf{y}_{n_j} - \widehat{\mathbf{B}}_I(\mathbf{Y}_I)^\top\mathbf{x}_{n_j}\|_p^2;\; I = \{n_1,\dots,n_h\} \subset \{1,\dots,N\}\right\}.$$

However, an exact computation is only for small sample sizes N possible. For larger sample sizes, a genetic algorithm with concentration step like that proposed by Neykov and Müller (2003) can be used, see also Rousseeuw and Driessen (2006).

Note that the inverse of $\mathbf{X}_I^\top\mathbf{X}_I$ always exists as soon as h is greater than the nonidentifiability parameter $\mathcal{N}_\beta(\mathbf{X})$ with $\lambda = \beta$. The subset estimator $\widehat{\mathbf{B}}_I$ is scatter and translation equivariant so that $\widehat{\mathbf{B}}_{I_*}$ is translation equivariant. However $\widehat{\mathbf{B}}_{I_*}$ is only scatter (scale) equivariant if $p = 1$. Otherwise it is only scatter equivariant with respect to orthogonal matrices $\mathbf{A}$ since then

$$\begin{aligned}
&q(\mathbf{x}, \mathbf{A}^\top \mathbf{y} + \mathbf{B}^\top \mathbf{x}, \widehat{\mathbf{B}}_I(\mathbf{YA} + \mathbf{XB})) \\
&= \|\mathbf{A}^\top \mathbf{y} + \mathbf{B}^\top \mathbf{x} - \widehat{\mathbf{B}}_I(\mathbf{YA} + \mathbf{XB})^\top \mathbf{x}\|_p \\
&= \|\mathbf{A}^\top \mathbf{y} + \mathbf{B}^\top \mathbf{x} - (\mathbf{A}^\top \mathbf{Y}_I^\top + \mathbf{B}^\top \mathbf{X}_I^\top)\mathbf{X}_I(\mathbf{X}_I^\top \mathbf{X}_I)^{-1}\mathbf{x}\|_p \\
&= \|\mathbf{A}^\top \mathbf{y} - \mathbf{A}^\top \mathbf{Y}_I^\top \mathbf{X}_I(\mathbf{X}_I^\top \mathbf{X}_I)^{-1}\mathbf{x}\|_p = \|\mathbf{A}^\top \mathbf{y} - \mathbf{A}^\top \widehat{\mathbf{B}}_I(\mathbf{Y})^\top \mathbf{x}\|_p \\
&= (\mathbf{y}^\top - \mathbf{x}^\top \widehat{\mathbf{B}}_I(\mathbf{Y}))\,\mathbf{A}\mathbf{A}^\top\,(\mathbf{y} - \widehat{\mathbf{B}}_I(\mathbf{Y})^\top \mathbf{x}) = q(\mathbf{x}, \mathbf{y}, \widehat{\mathbf{B}}_I(\mathbf{Y}))
\end{aligned}$$

for all $I = \{n_1, \ldots, n_h\} \subset \{1, \ldots, N\}$.

The d-fullness is given here by the nonidentifiability parameter $\mathcal{N}_\beta(\mathbf{X})$. This is an extension of the result in Müller and Neykov (2003) where it was proved for univariate generalized linear models.

Lemma 2. *If the quality function q is given by (4), then $\{q_n(\mathbf{Z}, \cdot);\ n = 1, \ldots, N\}$ is d-full with $d = \mathcal{N}_\beta(\mathbf{X}) + 1$.*

Proof. Consider any $I \subset \{1, \ldots, N\}$ with cardinality $\mathcal{N}_\beta(\mathbf{X}) + 1$. Then the triangle inequality provides for any $c \in I\!R$

$$\begin{aligned}
&\{\mathbf{B} \in I\!R^{r\times p};\ \max_{i\in I} q_i(\mathbf{z}_i, \mathbf{B}) \le c\} = \{\mathbf{B} \in I\!R^{r\times p};\ \max_{i\in I} \|_p \mathbf{y}_i - \mathbf{B}^\top \mathbf{x}_i\|_p \le \sqrt{c}\} \\
&\subset \{\mathbf{B} \in I\!R^{r\times p};\ \max_{i\in I} \|\mathbf{B}^\top \mathbf{x}_i\|_p - \|\mathbf{y}_i\|_p \le \sqrt{c}\} \\
&\subset \{\mathbf{B} \in I\!R^{r\times p};\ \max_{i\in I} \|\mathbf{B}^\top \mathbf{x}_i\|_p \le \sqrt{c} + \max_{i\in I} \|\mathbf{y}_i\|_p\} \\
&= \{\mathbf{B} \in I\!R^{r\times p};\ \max_{i\in I} \|\mathbf{B}^\top \mathbf{x}_i\|_p \le \sqrt{\tilde{c}}\} = \{(\mathbf{b}_1, \ldots, \mathbf{b}_p) \in I\!R^{r\times p};\ \max_{i\in I} \sum_{j=1}^p (\mathbf{b}_j^\top \mathbf{x}_i)^2 \le \tilde{c}\} \\
&\subset \{(\mathbf{b}_1, \ldots, \mathbf{b}_p) \in I\!R^{r\times p};\ \frac{1}{\mathcal{N}_\beta(\mathbf{X}) + 1} \sum_{i\in I} \sum_{j=1}^p \mathbf{b}_j^\top \mathbf{x_i} \mathbf{x}_i^\top \mathbf{b}_j \le \tilde{c}\} \\
&= \{(\mathbf{b}_1, \ldots, \mathbf{b}_p) \in I\!R^{r\times p};\ \frac{1}{\mathcal{N}_\beta(\mathbf{X}) + 1} \sum_{j=1}^p \mathbf{b}_j^\top \sum_{i\in I} \mathbf{x_i} \mathbf{x}_i^\top\, \mathbf{b}_j \le \tilde{c}\}.
\end{aligned}$$

The definition of $\mathcal{N}_\beta(\mathbf{X})$ implies that the matrix $\sum_{i\in I} \mathbf{x_i} \mathbf{x}_i^\top$ is of full rank. Hence the set $\{(\mathbf{b}_1, \ldots, \mathbf{b}_p) \in I\!R^{r\times p};\ \frac{1}{\mathcal{N}_\beta(\mathbf{X})+1} \sum_{j=1}^p \mathbf{b}_j^\top \sum_{i\in I} \mathbf{x_i} \mathbf{x}_i^\top\, \mathbf{b}_j \le \tilde{c}\}$ is bounded and therefore included in a compact subset of $I\!R^{r\times p}$. □

Since the upper bound for the breakdown point given by Theorem 1 and Theorem 2 holds also for estimators which are not scatter equivariant, the combination of these theorems, Theorem 5 and Lemma 2 provides the following result. This result was derived for univariate regression already in Müller (1995).

Theorem 6. *If* $\left\lfloor \frac{N+\mathscr{N}_\beta(\mathbf{X})+1}{2} \right\rfloor \leq h \leq \left\lfloor \frac{N+\mathscr{N}_\beta(\mathbf{X})+2}{2} \right\rfloor$*, then the breakdown point of the trimmed estimator* $\widehat{\mathbf{B}}_h$ *for* $\mathbf{B}$ *with quality function given by (4) satisfies*

$$\varepsilon^*(\widehat{\mathbf{B}}_h, \mathbf{Z}) = \frac{1}{N} \left\lfloor \frac{N - \mathscr{N}_\beta(\mathbf{X}) + 1}{2} \right\rfloor .$$

Müller (1995) showed Theorem 6 not only for estimating β but also for general linear aspects $\lambda = \mathbf{L}\beta$ of univariate regression models. Thereby $\mathscr{N}_\beta(\mathbf{X})$ must be only replaced by $\mathscr{N}_\lambda(\mathbf{X})$ in Theorem 6. However in this case the lower bound cannot be derived via d-fullness. In Müller (1995), the lower bound was proved directly for trimmed estimators, see also Müller (1997). This proof holds also for multivariate regression so that Theorem 6 holds also for linear aspects $\Lambda = \mathbf{LB}$ of multivariate regression.

6.2 Univariate regression with simultaneous scale estimation

If simultaneously the regression parameter $\beta \in I\!R^r$ and the scale parameter $\sigma \in I\!R^+$ in a univariate regression model shall be estimated, then the following quality function can be used

$$q(\mathbf{z}, \beta, \sigma) = q(\mathbf{x}, y, \beta, \sigma) = \frac{1}{2} \left(\frac{y - \mathbf{x}^\top \beta}{\sigma} \right)^2 + \log(\sigma). \tag{5}$$

In Müller and Neykov (2003) a little bit more general quality function were considered. But for simplicity, the quality function (5) shall be used here. The h-trimmed estimator $(\widehat{\beta}, \widehat{\sigma})$ for (β, σ) can be determined by calculating the maximum likelihood estimators

$$\widehat{\beta}_I(\mathbf{y}) = (\mathbf{X}_I^\top \mathbf{X}_I)^{-1} \mathbf{X}_I^\top \mathbf{y}_I$$

and

$$\widehat{\sigma}_I(\mathbf{y}) = \sqrt{\frac{1}{I} \sum_{j=1}^{h} (y_{n_j} - \mathbf{x}_{n_j}^\top \widehat{\beta}_I(\mathbf{y}))^2}$$

for each subsample $I = \{n_1, \ldots, n_h\} \subset \{1, \ldots, N\}$, where $\mathbf{y}_I = (\mathbf{y}_{n_1}, \ldots, y_{n_h})^\top$ and again $\mathbf{X}_I = (\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_h})^\top$. Then $(\widehat{\beta}(\mathbf{y}), \widehat{\sigma}(\mathbf{y}))$ is that $(\widehat{\beta}_{I_*}(\mathbf{y}), \widehat{\sigma}_{I_*}(\mathbf{y}))$ with

$$I_* = \arg\min \left\{ \sum_{j=1}^{h} q(\mathbf{x}_{n_j}, y_{n_j}, \widehat{\beta}_I(\mathbf{y}), \widehat{\sigma}_I(\mathbf{y}));\ I = \{n_1, \ldots, n_h\} \subset \{1, \ldots, N\} \right\}.$$

$\widehat{\beta}_I$ is translation equivariant and scale equivariant and $\widehat{\sigma}_I$ is translation invariant and scale equivariant. Therefore we have

$$
\begin{aligned}
& q(\mathbf{x}, ya + \mathbf{x}^\top \beta, \widehat{\beta}_I(\mathbf{y}a + \mathbf{X}\beta), \widehat{\sigma}_I(\mathbf{y}a + \mathbf{X}\beta)) \\
&= \frac{1}{2}\left(\frac{ya + \mathbf{x}^\top \beta - \mathbf{x}^\top(\widehat{\beta}_I(\mathbf{y})a + \beta)}{\widehat{\sigma}_I(\mathbf{y})a}\right)^2 + \log(\widehat{\sigma}_I(\mathbf{y})a) \\
&= q(\mathbf{x}, y, \widehat{\beta}_I(\mathbf{y}), \widehat{\sigma}_I(\mathbf{y})) + \log(a)
\end{aligned}
$$

for all $I = \{n_1, \ldots, n_h\} \subset \{1, \ldots, N\}$ so that $\widehat{\beta}_{I_*}$ is translation equivariant and scale equivariant and $\widehat{\sigma}_{I_*}$ is translation invariant and scale equivariant.

Since the simultaneous estimator $(\widehat{\beta}, \widehat{\sigma})$ for (β, σ) breaks down when one of its components breaks down, an upper bound of the breakdown point of $(\widehat{\beta}, \widehat{\sigma})$ is $\frac{1}{N}\left\lfloor \frac{N - \max\{\mathscr{N}_\beta(\mathbf{X}), \mathscr{E}(\mathbf{X})\}+1}{2} \right\rfloor$ according to Section 4.1 and Section 4.2.

Deriving a lower bound for the breakdown point, Müller and Neykov (2003) implicitely assumed that the exact fit parameter $\mathscr{E}(\mathbf{X})$ is zero. Here we extend this result for the case that it must be not zero.

Theorem 7. *If the quality function q is given by (5), then* $\{q_n(\mathbf{Z}, \cdot);\ n = 1, \ldots, N\}$ *is d-full with* $d = \max\{\mathscr{N}_\beta(\mathbf{X}), \mathscr{E}(\mathbf{X})\} + 1$.

Proof. We have to show that γ given by

$$
\gamma(\beta, \sigma) := \max_{i \in I} \frac{1}{2}\left(\frac{y_i - \mathbf{x}_i^\top \beta}{\sigma}\right)^2 + \log(\sigma)
$$

is sub-compact for all $I \subset \{1, \ldots, N\}$ with cardinality $\max\{\mathscr{N}_\beta(\mathbf{X}), \mathscr{E}(\mathbf{X})\} + 1$. Take any $c \in \mathbb{R}$ and set $\tilde{\beta}(\sigma) := \arg\min\{\gamma(\beta, \sigma);\ \beta \in \mathbb{R}^r\}$ and $\tilde{\sigma}(\beta) := \arg\min\{\gamma(\beta, \sigma);\ \sigma \in \mathbb{R}^+\}$. Then $\tilde{\beta}(\sigma)$ is independent of σ such that $\tilde{\beta}(\sigma) =: \tilde{\beta}$. Setting

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

we see that γ_1 is a sub-compact function since I has cardinality greater than $\mathscr{E}(\mathbf{X})$. Hence, there exists a compact set $\Theta_1 \subset \mathrm{int}(\mathbb{R}^+)$ such that $\{\sigma;\ \gamma_1(\sigma) \le c\} \subset \Theta_1$. Moreover, we have that with $\eta(\beta) := \max_{i \in I} \left| y_i - \mathbf{x}_i^\top \beta \right|$

$$
\tilde{\sigma}(\beta) = \eta(\beta)
$$

so that

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

The proof of Lemma 2 provides 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 $\Theta_2 \subset \mathrm{int}(\mathbb{R}^r)$. Then we have

$$
\{(\beta, \sigma) \in \mathbb{R}^r \times \mathbb{R}^+;\ \gamma(\beta, \sigma) \le c\}
$$

$$\subset \{(\beta,\sigma) \in \mathbb{R}^r \times \mathbb{R}^+;\ \gamma_1(\sigma) \le c \text{ and } \gamma_2(\beta) \le c\} \subset \Theta_2 \times \Theta_1$$

so that γ is sub-compact. □

Theorem 8. *If* $\left\lfloor \frac{N+\max\{\mathcal{N}_\beta(\mathbf{X}),\mathcal{E}(\mathbf{X})\}+1}{2} \right\rfloor \le h \le \left\lfloor \frac{N+\max\{\mathcal{N}_\beta(\mathbf{X}),\mathcal{E}(\mathbf{X})\}+2}{2} \right\rfloor$*, then the breakdown point of the trimmed estimator* $(\widehat{\beta},\widehat{\sigma})_h$ *for* (β,σ) *with quality function given by (5) satisfies*

$$\varepsilon^*((\widehat{\beta},\widehat{\sigma})_h,\mathbf{Z}) = \frac{1}{N}\left\lfloor \frac{N-\max\{\mathcal{N}_\beta(\mathbf{X}),\mathcal{E}(\mathbf{X})\}+1}{2} \right\rfloor.$$

References

Davies, P.L. (1993). Aspects of robust linear regression. Ann. Statist. **21**, 1843–1899.

Davies, P.L., Gather, U. (2005). Breakdown and groups. With discussion. Ann. Statist. **33**, 977–1035.

Davies, P.L., Gather, U. (2007). The breakdown point – examples and counterexamples. REVSTAT **5**, 1–17.

Donoho, D.L., Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. Ann. Stat. **20**, 1803–1827.

Donoho, D.L., Huber, P.J. (1983). The notion of breakdown point. In: Bickel, P.J., Doksum, K., Hodges, J.L. (eds.) A Festschrift for Erich L. Lehmann, pp. 157–184. Wadsworth, Belmont, CA.

Hampel, F.R. (1971). A general qualitative definition of robustness. Ann. Math. Statist. **42**, 1887–1896.

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

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.

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

Müller, Ch.H., Schäfer, Ch. (2010). Designs with high breakdown point in nonlinear models. In: Giovagnoli, A., Atkinson, A.C., Torsney, B. (eds.) mODa 9 – Advances in Model-Oriented Design and Analysis, pp. 137–144. Physica-Verlag, Heidelberg.

Neykov, N., Müller, Ch.H. (2003). Breakdown point and computation of trimmed likelihood estimators in generalized linear models. In: Dutter, R., Filzmoser, P., Gather, U., Rousseeuw, P.J. (eds.) Developments in Robust Statistics, pp. 277–286. Physica-Verlag, Heidelberg.

Rousseeuw, P.J., Leroy, A.M. (2003). Robust Regression and Outlier Detection. Wiley, New York.

Rousseeuw, P. J., Driessen, K. (2006) Computing LTS regression for large data sets. Data Min. Knowl. Disc. **12**, 29–45.

Vandev, D.L. (1993). A note on breakdown point of the least median squares and least trimmed squares. Statist. Probab. Lett. **16**, 117–119.

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