Breakdown and groups [1]

By P. L. Davies and U. Gather

Universität Essen, Universität Dortmund

The concept of breakdown point was introduced by Hodges (1967) and Hampel (1968, 1971) and still plays an important though at times a controversial role in robust statistics. It has proved most successful in the context of location, scale and regression problems. In this paper we argue that this success is intimately connected to the fact that the translation and affine groups act on the sample space and give rise to a definition of equivariance for statistical functionals. For such functionals a nontrivial upper bound for the breakdown point can be shown. In the absence of such a group structure a breakdown point of one is attainable and this is perhaps the decisive reason why the concept of breakdown point in other situations has not proved as successful. Even if a natural group is present it is often not sufficiently large to allow a nontrivial upper bound for the breakdown point. One exception to this is the problem of the autocorrelation structure of time series where we derive a nontrivial upper breakdown point using the group of realizable linear filters. The paper is formulated in an abstract manner to emphasize the role of the group and the resulting equivariance structure.

AMS 2000 subject classifications. Primary 62G07; secondary 65D10, 62G20.

Key words and phrases. equivariance, breakdown point, robust statistics

[1]Research supported in part by Sonderforschungsbereich 475, University of Dortmund.

1. Introduction.

1.1. *Breakdown points and equivariance.* The notion of breakdown point was introduced by Hodges (1967) and Hampel (1968, 1971). Their definition was based on a sequence of estimators $T_n, n = 1, \ldots$ with T_n applied to a sequence of samples $(X_1(F), \ldots, X_n(F))$ of random variables distributed according to F. Huber (1981) took a more functional analytical approach and considered statistical functionals T defined on the space of distributions. In this framework important properties of statistical functionals can be phrased in terms of boundedness, continuity and differentiability. Breakdown is related to the boundedness of the functional and the breakdown point is defined in terms of metrics or the sizes of neighbourhoods on the space of distributions. A simple and intuitive definition of breakdown point but one restricted to finite samples, the finite sample breakdown point, was introduced by Donoho (1982) and Donoho and Huber (1983). Successful applications of the concept of breakdown point have been to the location, scale and regression problems in $\mathbb{R}^k$ or to problems which are intimately related to these (see for example Ellis and Morgenthaler (1992), Davies and Gather (1993), Hubert (1997), Terbeck and Davies (1998), He and Fung (2000), Müller and Uhlig (2001)). The reason for this is that such problems have a rich equivariance structure deriving from the translation or linear group operating on $\mathbb{R}^k$. By restricting the class of statistical functionals to those with the appropriate equivariance structure one can prove the existence of nontrivial highest breakdown points which in many cases can be achieved at least locally (Huber (1981), Davies (1993)). The simplest example is perhaps that of the median. If we use the replacement finite sample breakdown point of Donoho and Huber (1983) then the median has a breakdown point of $\lfloor (n+1)/2 \rfloor / n$ and this is known to be the highest possible value for translation equivariant location functionals. If we consider scale functionals then the situation is somewhat different. The statistical folklore is that the highest possible finite sample breakdown point for any affine equivariant scale functional is $\lfloor n/2 \rfloor /n$ and that this is attained by the median absolute deviation, MAD. Some authors (Croux and Rousseeuw (1992), Davies (1993)) are aware that this is not correct, as is also shown by the following example. For the sample

$$\boldsymbol{x}_{11} = (1.0, 1.8, 1.3, 1.3, 1.9, 1.1, 1.3, 1.6, 1.7, 1.3, 1.3) \tag{1.1}$$

the MAD has a finite sample breakdown point of only 1/11. This can be seen by replacing the data point 1.0 by 1.3: the altered data set has a MAD of zero and this is generally regarded as breakdown. If the sample has no repeated observations then the MAD has a finite sample breakdown point of $\lfloor n/2 \rfloor /n$ and this is indeed the highest possible finite sample breakdown point for a scale functional. The difference between the maximal finite sample breakdown points for location functionals ($\lfloor (n+1)/2 \rfloor /n$) and scale functionals ($\lfloor n/2 \rfloor /n$) is explained by our main theorem below. Clearly if no restrictions are placed on the statistical functionals under consideration then the highest possible breakdown point is one and this is attained by any constant functional. We claim that the existence of nontrivial upper bounds is an essential component of the concept of breakdown point and that such nontrivial bounds are linked to a sufficiently rich equivariance structure. Another way of looking at the problem is the following. If the structure imposed by the model is very high then this will restrict the size of the group operating on the parameter space. It will consequently be easier to find equivariant functionals with a breakdown point of one. In other words the more highly structured the subset of interest the easier it is to find it in a sea of noise.

1.2. *Previous work.* The success of the concept of breakdown point in location, scale, linear regression, and related problems has lead many authors to develop definitions applicable in other situations. We mention nonlinear regression (Stromberg and Ruppert (1992)), time series (Martin and Jong (1977), Papantoni-Kazakos (1984), Tatum and Hurvich (1993), Lucas (1997), Mendes (2000), Ma and Genton (2000)), radial data (He and Simpson (1992)) and more general situations as in Sakata and White (1995) and He and Simpson (1993), the latter one restricting contamination to gross-error models. None of the above articles with the exception of He and Simpson (1993) mentions a group structure or a corresponding equivariance structure for the class of functionals under consideration. In particular constant functionals are not excluded and have the highest possible breakdown point of one. In these more general situations this may be the reason for the general lack of acceptance of the proposed definitions of breakdown points.

1.3. *Two examples.* In some cases there is a canonical group acting on the sample space but it is too small to be of use. The first example is the problem considered by Ruckstuhl and Welsh (2001) which is formulated in terms of estimating the parameter θ of a binomial distribution $\mathrm{Bin}(k, \theta)$ based on different samples. Breakdown occurs when the estimator takes values arbitrarily close to the boundary of the parameter space $\Theta = [0, 1]$. Ruckstuhl and Welsh show that the asymptotic breakdown point of the maximum likelihood estimator is one. A simple estimator based on a sample $x_1, \ldots, x_n \in \{0, \ldots, k\}$ with a breakdown point of one is

$$T_w(\boldsymbol{x}_n) = \frac{1}{n} \sum \min\{\max\{x_j, 1\}, k-1\}/k. \tag{1.2}$$

Rousseeuw (personal communication) has pointed out that there is a canonical group acting on the sample space which consists of the identity and the mapping g

defined by $g(x) = k - x$. A functional T is equivariant with respect to this group if $T(P^g) = 1 - T(P)$ where $P^g(B) := P(g^{-1}(B))$ for any (Borel) set B. The maximum likelihood estimator is equivariant in this sense as is the functional (1.2). In other words, even equivariant functionals can have a breakdown point of one.

The second example is taken from time series. We use the basic model of a stationary autoregressive process of order one $x_{n+1} = \theta x_n + r_{n+1}$ so that the parameter space is $\Theta = (-1, 1)$. We take the sample space $\mathcal{X}$ to be the set of all doubly infinite sequences $\mathbb{R}^{\mathbb{Z}}$. The only group we know of which is compatible with the model is the multiplicative group $\mathbb{R}\backslash\{0\}$ which multiplies each component of a point $x \in \mathcal{X}$ by a nonzero number. This leaves the structure of the autoregressive process unchanged and hence any constant functional with a breakdown point of one is equivariant with respect to this group.

1.4. *Fisher and asymptotic consistency.* As already mentioned if no group structure exists then constant functionals are not excluded and these have a breakdown point of one. It may be thought that the absence of a group structure could be compensated by imposing additional restrictions on the set of allowable functionals in the hope of attaining a nontrivial upper bound for the breakdown point. Two restrictions which are plausible are Fisher consistency and asymptotic consistency. They at least have the advantage of excluding the constant functionals. Given a parametric family of distributions P_θ Fisher consistency is defined by $T(P_\theta) = \theta$ for all $\theta \in \Theta$. This will in general not help as the parametric family P_θ is too sparse in the set of all distributions P, if we define a functional T by

$$T(P) = \begin{cases} \theta & \text{if } P = P_\theta, \quad \theta \in \Theta, \\ \theta_0 & \text{otherwise.} \end{cases} \tag{1.3}$$

$$((\mathcal{X},\mathcal{B}(\mathcal{X}),\mathcal{P}),d) \;\underset{T}{\rule{10em}{0.4pt}}\; (\Theta,D)$$

FIG. 1. *Connections 1*

As an example we consider the family of normal distributions on $\mathbb{R}$ with $P_\theta = N(\theta,1)$ and $\theta_0 = 0$. The functional defined by (1.3) has a breakdown point of one but is Fisher consistent at the family of normal distributions. A similar argument shows that asymptotic consistency does not help: there exist functionals which are Fisher consistent and asymptotically consistent and still have a breakdown point of one.

2. A general definition of breakdown point. The approach we adopt is the functional analytic one of Huber (1981) which goes back at least to von Mises (1937, 1947). We will consider a measurable sample space $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ and the set $\mathcal{P}$ of all non-degenerate probability measures on this space. This clearly includes all non-degenerate empirical measures. In particular in scale problems we do not assume that the data points are "in general position". This latter point is not just of theoretical interest as problems of collinearity or the rounding of observations can occasionally be so severe as to necessitate a modification of the scale functionals used. The problem also occurs in the existence of joint M-estimators of location and scale (Huber (1981), Chapter 6, Kent and Tyler (1991)) and in structured regression problems (see the remarks of Huber (1995)). Such problems are rarely discussed in the literature but see Davies (1993) and Dietel (1993).

The situation we describe is shown in Figure 1. On the left we have a measurable sample space $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ and the family $\mathcal{P}$ of probability measures defined on $\mathcal{B}(\mathcal{X})$.

Moreover $\mathcal{P}$ is equipped with some pseudometric $d : \mathcal{P} \times \mathcal{P} \to [0, \infty)$ which satisfies

$$\sup_{P,Q \in \mathcal{P}} d(P, Q) = 1. \tag{2.1}$$

On the right we have a parameter space Θ which is equipped with a pseudometric D on $\Theta \times \Theta$ which satisfies

$$\sup_{\theta_1, \theta_2} D(\theta_1, \theta_2) = \infty. \tag{2.2}$$

The two are connected by a functional T

$$T : \mathcal{P} \to \Theta \tag{2.3}$$

which associates to every point $P \in \mathcal{P}$ a point $T(P) \in \Theta$. The breakdown point $\varepsilon^*(T, P, d, D)$ of the functional T at the distribution P with respect to the pseudometrics d and D is defined by

$$\varepsilon^*(T, P, d, D) = \inf\{\varepsilon > 0 : \sup_{d(P,Q) < \varepsilon} D(T(P), T(Q)) = \infty\}. \tag{2.4}$$

We note that the breakdown point is a local concept.

The replacement finite sample breakdown point of a functional T is defined as follows. If $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ is a sample of size n we denote its empirical distribution by

$$P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

Let $\boldsymbol{y}_{n,k}$ be a sample obtained from $\boldsymbol{x}_n$ by altering k of the x_i such that the two samples differ in exactly k points. Denote the empirical distribution of $\boldsymbol{y}_{n,k}$ by $Q_{n,k}$. The finite sample breakdown point (fsbp) of T at the sample $\boldsymbol{x}_n$ (or P_n) is then defined by

$$\text{fsbp}\,(T, \boldsymbol{x}_n, D) = \frac{1}{n} \min\{k \in \{0, 1, \ldots, n\} : \sup_{Q_{n,k}} D(T(P_n), T(Q_{n,k})) = \infty\}. \tag{2.5}$$

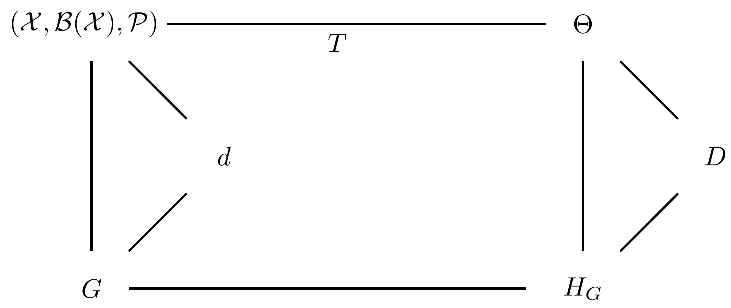

FIG. 2. *Connections 2*

3. Groups and equivariance.

3.1. *An upper bound for the breakdown point.* As mentioned in Section 1.1 most extensions of the concept of breakdown point have considered the situation described by Figure 1. We claim that a reasonable definition of breakdown point requires more structure and this is shown in Figure 2. On the left we still have the sample space $\mathcal{X}$, the family of probability measures $\mathcal{P}$ and the pseudometric d but we now have the additional structure of a group G of measurable transformations g of $\mathcal{X}$ onto itself, $g : (\mathcal{X}, \mathcal{B}(\mathcal{X})) \to (\mathcal{X}, \mathcal{B}(\mathcal{X}))$. For any $P \in \mathcal{P}$ and any $g \in G$ we define P^g by

$$P^g(B) = P(g^{-1}(B)), \quad B \in \mathcal{B}. \tag{3.1}$$

Finally we impose a technical condition on d and require

$$d(\alpha P + (1-\alpha)Q_1, \alpha P + (1-\alpha)Q_2) \le 1 - \alpha, \quad P, Q_1, Q_2 \in \mathcal{P}, 0 < \alpha < 1. \tag{3.2}$$

On the right we also have an additional structure namely a group of transformations of Θ into itself which is parameterized by the group G on the left. Specifically we suppose that the group G induces a group $H_G = \{h_g : g \in G\}$ of transformations

$h_g : \Theta \to \Theta$ which describes the equivariance structure of the problem. A functional $T : \mathcal{P} \to \Theta$ is called equivariant with respect to G if

$$T(P^g) = h_g(T(P)) \quad \text{for all } g \in G,\ P \in \mathcal{P}. \tag{3.3}$$

We now define

$$G_0 = \{g \in G : D(\theta, h_g(\theta)) = 0 \text{ for all } \theta \in \Theta\}. \tag{3.4}$$

The restriction of $g \in G$ to a set $B \in \mathcal{B}$ will be denoted by $g_{|B}$ and the unit element of G by ι. Given this we define

$$\Delta(P) = \sup\{P(B) : B \in \mathcal{B},\ g_{|B} = \iota_{|B} \text{ for some } g \notin G_0\}. \tag{3.5}$$

Theorem 3.1.

With the above notation and under the assumption

$$\lim_{n\to\infty} \inf_{\theta} D(\theta, h_{g^n}(\theta)) = \infty \quad \textit{for all } g \notin G_0 \tag{3.6}$$

we have

$$\varepsilon^*(T, P, d, D) \leq (1 - \Delta(P))/2 \tag{3.7}$$

for all G-equivariant functionals T, for all $P \in \mathcal{P}$, for all pseudometrics d and D satisfying (2.1), (2.2), and (3.2).

Proof: The proof of the theorem follows the lines of Rousseeuw (1983, 1984) and Davies (1987, 1993). Let B_0 and $g \notin G_0$ be such that $g_{|B_0} = \iota_{|B_0}$. Consider the measures Q_1, Q_2 and Q_n defined by

$$Q_1(B) = P(B \cap B_0), \quad B \in \mathcal{B}$$

$$Q_2(B) = P(B) - Q_1(B), \quad B \in \mathcal{B}$$

$$Q_n(B) = (Q_2(B) + Q_2^{g^n}(B))/2 + Q_1(B), \quad B \in \mathcal{B}.$$

As $Q_1^g = Q_1^{g^{-1}} = Q_1$ we have

$$Q_n^{g^{-n}} = (Q_2^{g^{-n}} + Q_2)/2 + Q_1,$$

which is indeed an element of $\mathcal{P}$. From this it follows on using (3.2)

$$d(Q_n^{g^{-n}}, P) \le (1 - P(B_0))/2 \tag{3.8}$$

$$d(Q_n, P) \le (1 - P(B_0))/2. \tag{3.9}$$

By the equivariance of T we have

$$T(Q_n^{g^{-n}}) = h_{g^{-n}}(T(Q_n))$$

from which it follows

$$D(T(Q_n^{g^{-n}}), T(Q_n)) \le D(T(P), T(Q_n^{g^{-n}})) + D(T(P), T(Q_n)).$$

From (3.6) we have

$$\lim_{n\to\infty} D(T(Q_n^{g^{-n}}), T(Q_n)) = \infty$$

and hence

$$\lim_{n\to\infty} (D(T(P), T(Q_n^{g^{-n}})) + D(T(P), T(Q_n))) = \infty.$$

Both $D(T(P), T(Q_n^{g^{-n}}))$ and $D(T(P), T(Q_n))$ cannot remain bounded and we conclude that for any $\varepsilon > (1 - P(B_0))/2$

$$\sup_{d(P,Q)<\varepsilon} D(T(P), T(Q)) = \infty.$$

As this holds for any B_0 such that $g_{|B_0} = \iota_{|B_0}$ for some $g \notin G_0$ the claim of the theorem follows. $\square$

We can prove a similar result for the finite sample breakdown point.

THEOREM 3.2.

With the above notation and under assumption (3.6) we have

$$fsbp\,(T, \boldsymbol{x}_n, D) \leq \left\lfloor \frac{n - n\Delta(P_n) + 1}{2} \right\rfloor / n. \tag{3.10}$$

Proof: Firstly we note that there are exactly $n\Delta(P_n)$ points in $\boldsymbol{x}_n$ for which $g(x_i) = x_i$ for some $g \notin G_0$. We assume without loss of generality that these are the sample points $x_1, \ldots, x_{n\Delta(P_n)}$. If $\Delta(P_n) = 0$ there are no such points and some obvious alterations to following proof are required. To ease the notation we write

$$l(n) = \left\lfloor \frac{n - n\Delta(P_n) + 1}{2} \right\rfloor.$$

We consider the sample $\boldsymbol{y}^*_{n,k}$ given by

$$\boldsymbol{y}^*_{n,k} = (x_1, \ldots, x_{n\Delta(P_n)}, \ldots, x_{n-l(n)}, g^m(x_{n-l(n)+1}), \ldots, g^m(x_n))$$

for some $m \geq 1$ and some $g \notin G_0$. We denote its empirical distribution by $Q^*_{n,k}$. The sample $\boldsymbol{y}^*_{n,k}$ contains at least $n - l(n)$ points of the original sample $\boldsymbol{x}_n$. The transformed sample $g^{-m}(\boldsymbol{y}^*_{n,k})$ is equal to

$$(x_1, \ldots, x_{n\Delta(P_n)}, g^{-m}(x_{n\Delta(P_n)+1}), \ldots, g^{-m}(x_{n-l(n)}), x_{n-l(n)+1}, \ldots, x_n)$$

It contains at least $n\Delta(P_n) + l(n)$ points of the original sample $\boldsymbol{x}_n$ and as

$$n\Delta(P_n) + l(n) \geq n - l(n)$$

it contains at least $n - l(n)$ points of $\boldsymbol{x}_n$. By the equivariance of T we have

$$T(Q^{*g^{-m}}_{n,k}) = h_{g^{-m}}(T(Q^*_{n,k}))$$

from which it follows

$$D(h_{g^{-m}}(T(Q^*_{n,k})), T(Q^*_{n,k})) \leq D(T(P_n), T(Q^*_{n,k})) + D(T(P_n), T(Q^{*g^{-m}}_{n,k}))$$

From (3.6) we have

$$\lim_{m\to\infty} D(h_{g^{-m}}(T(Q^*_{n,k})), T(Q^*_{n,k})) = \infty$$

and hence $D(T(P_n), T(Q^*_{n,k}))$ and $D(T(P_n), T(Q^{*g^{-m}}_{n,k}))$ cannot both remain bounded and we conclude that for any $k \geq \lfloor \frac{n-n\Delta(P_n)+1}{2} \rfloor$

$$\sup_{Q_{n,k}} D(T(P_n), T(Q_{n,k})) = \infty$$

from which the claim of the theorem follows. □

There is in fact a direct connection between the two theorems. We consider the total variation metric d_{tv} defined by

$$d_{tv}(P, Q) = \sup_{B\in\mathcal{B}(\mathcal{X})} |P(B) - Q(B)|.$$

If $\mathcal{B}(\mathcal{X})$ "shatters" every finite set of points in $\mathcal{X}$ then

$$d_{tv}(P_n, P_n^*) = k/n$$

where P_n denotes the empirical measure deriving from $(x_1, \ldots, x_n)$ and P_n^* that deriving from $(x_1^*, \ldots, x_n^*)$ with the two samples differing in exactly k points. Suppose now that $\varepsilon^*(T, P_n, d_{tv}, D) = (1-\Delta(P_n))/2$. If $k < n(1-\Delta(P_n))/2$ then breakdown in the sense of finite sample breakdown point cannot occur and we see that

$$\text{fsbp}\,(T, \boldsymbol{x}_n, D) \geq \left\lfloor \frac{n - n\Delta(P_n)}{2} \right\rfloor /n. \tag{3.11}$$

Unfortunately the inequality of Theorem 3.2 seems not to be provable in the same manner.

3.2. *The pseudometric d.* The definition of breakdown point in (2.4) is framed partly in terms of the pseudometric d over the space $\mathcal{P}$ of probability measures. The pseudometric d is required to satisfy only (2.1) and (3.2). This does not mean that the breakdown point will be the same for all such d. A simple counterexample is provided by the scale problem in $\mathbb{R}$. If we use the Kolmogoroff metric then the breakdown point of the MAD at an atomless distribution is 1/4 (Huber (1981), page 110). However if we use the Kuiper metric then the breakdown point is 1/2 in spite of the fact that both metrics fulfill the restrictions. More generally if d' and d'' are two metrics satisfying (2.1) and (3.2) and such that $d' \leq d''$ then

$$\varepsilon^*(T, P, d', D) \leq \varepsilon^*(T, P, d'', D) \leq (1 - \Delta(P))/2. \tag{3.12}$$

In particular if $\varepsilon^*(T, P, d', D) = (1 - \Delta(P))/2$ then $\varepsilon^*(T, P, d'', D) = (1 - \Delta(P))/2$. A class ordered metrics which are of use for one-dimensional scale problems is provided by the generalized Kuiper metrics $d_{m/2}, m \in \mathbb{N}$, defined as follows. We set

$$\begin{aligned} d_{m/2}(P,Q) = \sup\big\{|P(J) - Q(J)| : J = \cup_{j=1}^{M} I_j, M = \lfloor (m+1)/2 \rfloor,& \\ \lfloor m/2 \rfloor \quad = \quad \#\{I_j \text{ finite interval}\},& \\ \lceil m/2 \rceil - \lfloor m/2 \rfloor \quad = \quad \#\{I_j \text{ infinite interval}\}\big\}.& \qquad (3.13) \end{aligned}$$

Thus $d_{0.5}$ is the ordinary Kolmogoroff metric, d_1 the ordinary Kuiper metric. For $d_{4.5}$ the supremum in (3.13) is taken over the union of four finite intervals and one infinite interval. We note in passing that d_9 is the default value for the metric in Davies and Kovac (2002) in the context of the modality of densities. We have

$$d_{0.5} \leq d_1 \leq \ldots \leq d_{m/2} \leq d_{(m+1)/2}. \tag{3.14}$$

Furthermore all the metrics satisfy (2.1) and (3.2).

3.3. *The pseudometric D, a canonical choice.* As we have seen in the case of d in Section 3.2 there seems to be no canonical choice: different choices of d can lead to different breakdown points. A similar problem exists with respect to the pseudometric D on Θ. We now indicate a possibility of making D dependent on d. The idea is that two parameter values θ_1 and θ_2 are far apart with respect to D if and only if the corresponding distributions are far apart with respect to d. We illustrate the idea using the location problem in $\mathbb{R}$. Suppose we have data with empirical distribution P_n and two values of the location parameter θ_1 and θ_2. We transform the data using the translations θ_1 and θ_2 which gives rise to two further distributions $P_n(\cdot - \theta_1)$ and $P_n(\cdot - \theta_2)$. If these two distributions are clearly distinguishable then $d(P_n(\cdot - \theta_1), P_n(\cdot - \theta_2))$ will be almost one. An opposed case is provided by an autoregressive process of order one. The parameter space is $\Theta = (-1, 1)$ and this may be metricized in such a manner that $D(\theta_1, \theta_2)$ tends to infinity for fixed θ_1 as θ_2 tends to the boundary. However values of θ close to, on or even beyond the boundary, may not be empirically distinguishable from values of θ in the parameter space. A sample of size $n = 100$ generated with $\theta_1 = 0.95$ is not easily distinguishable from a series generated with $\theta_2 = 0.9999$ even though $D(\theta_1, \theta_2)$ is large.

We now give a choice of D in terms of d and such that (2.2) is satisfied. We set

$$G(\theta_1, \theta_2) = \{g \in G : h_g(\theta_1) = \theta_2\}$$

and then define D by

$$D(\theta_1, \theta_2) = D_P(\theta_1, \theta_2) = \inf_{g \in G(\theta_1, \theta_2)} |\log(1 - d(P^g, P))|. \tag{3.15}$$

The interpretation is that we associate P with the parameter value θ_1 and P^g with the parameter value θ_2. The requirement (2.2) will only hold if $d(P^g, P)$ may be arbitrarily close to one so that the distributions associated with θ_1 and θ_2 are as far apart as possible. It is easily checked that D defines indeed a pseudometric on Θ; namely $D_P \geq 0$, D_P is symmetric and satisfies the triangle inequality. In some situations it seems reasonable to require that d and D be invariant with respect to the groups G and H_G respectively. If d is G–invariant, i.e.

$$d(P,Q) = d(P^g, Q^g), \quad \text{for all } P,Q \in \mathcal{P}, g \in G,$$

then D, defined by (3.15), inherits the invariance, i.e.

$$D(\theta_1, \theta_2) = D(h_g(\theta_1), h_g(\theta_2)), \quad \text{for all } \theta_1, \theta_2 \in \Theta, g \in G.$$

The G–invariance of d can often be met.

4. Examples.

4.1. *Location functionals and the translation group.* We take $\mathcal{X}$ to be k-dimensional Euclidean space $\mathbb{R}^k$ and G is the translation group with elements $g(x) = x + a$ with $a \in \mathbb{R}^k$. The parameter space Θ is $\mathbb{R}^k$ and the group H_G is again the translation group, with $h_g = g$. The pseudometric D on Θ is now a metric with $D(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|_k$. It clearly satisfies (2.2). G_0 consists only of the unit element of G which clearly implies (3.6). The pseudometric d is not so important. One possibility is

$$d(P,Q) = \sup_{B \in \mathcal{C}} |P(B) - Q(B)| \tag{4.1}$$

where

$$\mathcal{C} = \{C : C = \{x : x^t b + a \leq 0\}\},$$

with b a point in $\mathbb{R}^k$ and a a real number. This is a weak metric defined over a class of subsets with polynomial discrimination. The pseudometric d satisfies (3.2). As there is no set B which satisfies the definition of Δ in (3.5) we have $\Delta(P) = 0$ for all P. Theorem 3.1 now states that $\varepsilon^*(T, P, d) \leq 1/2$ for any translation equivariant functional.

4.2. *Scatter functionals and the affine group.* $\mathcal{X}$ is again the k-dimensional Euclidean space $\mathbb{R}^k$ but G is now the group of affine transformations

$$g(x) = Ax + a, \quad x \in \mathbb{R}^k, \tag{4.2}$$

where A is a nonsingular $k \times k$-matrix and a is a point in $\mathbb{R}^k$. The parameter space Θ is the space Σ_k of nonsingular symmetric $k \times k$-matrices. The group H_G of transformations of the parameter space is given by

$$h_g(\sigma) = A\sigma A^t, \quad \sigma_k \in \Sigma_k, \tag{4.3}$$

where g is given by (4.2). The pseudometric on Σ_k is given by

$$D(\sigma_1, \sigma_2) = |\log(\det(\sigma_1\sigma_2^{-1}))| \quad \sigma_1, \sigma_2 \in \Sigma_k. \tag{4.4}$$

It is easily checked that D satisfies (2.2). G_0 is given by

$$G_0 = \{g : g(x) = Ax + a, \det(A) = 1\}$$

from which (3.6) follows. The pseudometric d is again not so important but we define it by

$$d(P, Q) = \sup_{B \in \mathcal{C}} |P(B) - Q(B)| \tag{4.5}$$

where

$$\mathcal{C} = \{C : C = \{x \in \mathbb{R}^k : x^t cx + x^t b + a \leq 0\}\},$$

with c is a non-negative definite $k \times k$-matrix, b a point in $\mathbb{R}^k$ and a a real number. This is a weak metric defined over a class of subsets with polynomial discrimination, and it clearly satisfies (3.2). We now show that

$$\Delta(P) = \sup\{P(B) : B \text{ is a hyperplane of dimension } \leq k-1\}. \tag{4.6}$$

Suppose $g_{|B} = \iota_{|B}$ with $g(x) = Ax+a$. Then for $x \in B$ we have $Ax+a = I_k x$ where I_k denotes the $k \times k$ identity matrix. This implies $(A - I_k)x + a = 0$ for all $x \in B$.

As $g \neq \iota$ this cannot hold for all x and hence B is contained in the hyperplane $\{x : (A - I_k)x + a = 0\}$ which is of dimension at most $k - 1$. On the other hand if B is a lower dimensional hyperplane given by

$$B = \{x : Ax + a = 0\}$$

then we may choose $\alpha \neq 0$ such that $\alpha A + I_k$ is nonsingular. Define g by

$$g(x) = (\alpha A + I_k)x + \alpha a$$

Then g is a nonsingular affine transformation and $g(x) = x$ if and only if $x \in B$. Theorem 3.1 is now Theorem 3.2 of Davies (1993).

4.3. *Regression functionals and the translation group.* $\mathcal{X}$ is now the $k+1$-dimensional Euclidean space $\mathbb{R}^k \times \mathbb{R}$ where the first k components define the design points and the $k+1$ component is the corresponding value of y. The group G consists of all transformations

$$g((x^t, y)^t) = (x^t, y + x^t a)^t, \quad (x^t, y)^t \in \mathbb{R}^k \times \mathbb{R}, \tag{4.7}$$

with $a \in \mathbb{R}^k$. The space Θ is $\mathbb{R}^k$ and a functional $T : \mathcal{P} \to \Theta$ is equivariant with respect to the group if

$$T(P^g) = T(P) - a$$

with g as in (4.7). The group H_G is the translation group

$$h_g(\theta) = \theta + a \tag{4.8}$$

with g as in (4.7). The pseudometric D is simply the Euclidian distance $D(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|_k$ which satisfies (2.2). G_0 consists only of the identity and it is clear from (4.8) that (3.6) holds. Again the pseudometric d is not of great importance and we define it by

$$d(P, Q) = \sup_{B \in \mathcal{C}} |P(B) - Q(B)| \tag{4.9}$$

where

$$\mathcal{C} = \{C : C = \{(x^t, y)^t \in \mathbb{R}^k \times \mathbb{R} : |x^t\theta + y| \le c\}\},$$

with $\theta \in \mathbb{R}^k$ and $c \ge 0$. The metric d satisfies (3.2). We now show that

$$\Delta(P) = \sup\{P(C \times \mathbb{R}) : C \subset \mathbb{R}^k \text{ is a plane of dimension } \le k-1\}. \tag{4.10}$$

Suppose $g_{|B} = \iota_{|B}$ with $g((x^t, y)^t) = (x^t, y + x^t a)^t$. Then for $(x^t, y)^t \in B$ we must have $x^t a = 0$ so that $B \subset C \times \mathbb{R}$ with $C = \{x; x^t a = 0\}$ a plane. As $g((x^t, y)^t) = (x^t, y + x^t a)^t \ne (x^t, y)^t$ for all $(x^t, y)^t \in \mathbb{R}^k \times \mathbb{R}$ it follows that $a \ne 0$ and so the plane C is indeed one of dimension at most $k-1$. On the other hand if C is a plane of dimension at most $k-1$ then we can express it as $C = \{x : x^t a = 0\}$ for some $a \in \mathbb{R}^k, a \ne 0$. On setting $B = C \times \mathbb{R}$ we see that $g_{|B} = \iota_{|B}$ with $g((x^t, y)^t) = ((x^t, y + x^t a)^t)$ proving (4.10). The result is now Theorem 3.1 of Davies (1993).

4.4. *Time series and realizable linear filters.* As far as we know there are no results corresponding to the location, scale and linear regression problems for time series. We are aware of some work corresponding to a breakdown point concept for time series, namely Martin and Jong (1977), Papantoni-Kazakos (1984), Tatum and Hurvich (1993), Lucas (1997), Mendes (2000), Ma and Genton (2000), de Luna and Genton (2001), and references given in these papers. In none of these papers are any equivariance properties or group structures imposed. We now present what would seem to be a first result in this direction. The problem is that of specifying an autocorrelation matrix or a Toeplitz form for a given probability measure. We set

$$\mathcal{X} = \mathcal{X}_\delta = \{x \in \mathbb{C}^{\mathbb{Z}} : \sum_{j=0}^{\infty} |x_{n-j}|(1+\delta)^{-j} < \infty \text{ for all } n \in \mathbb{Z}\} \tag{4.11}$$

for some $\delta > 0$ and equip $\mathcal{X}$ with the usual Borel σ-algebra. We suppose that $\mathcal{P}$ is the set of all probability measures on $\mathcal{X}$. The group G is given by

(4.12)

$$G = \{g : g : \Gamma_{1+\varepsilon} \to \mathbb{C}, \text{ homomorphic and bounded with } \inf_{z \in \Gamma_{1+\varepsilon}} |g(z)| > 0\}$$

where Γ_r denotes the open disc in $\mathbb{C}$ of radius r. Each such g has a power series expansion

$$g(z) = \sum_{j=0}^{\infty} g_j z^j \tag{4.13}$$

with $g_0 \neq 0$ and

$$|g_j| \leq K(g, \eta)(1+\eta)^{-j} \tag{4.14}$$

for each $\eta, 0 < \eta < \varepsilon$ and some constant $K = K(g, \eta)$. If $\varepsilon > \delta$ then each g in G defines a linear filter on $\mathcal{X}$, which by an abuse of notation, we also denote by g. It is defined by

$$(g(x))_n = \sum_{j=0}^{\infty} x_{n-j} g_j, n \in \mathbb{Z}. \tag{4.15}$$

We define

$$\mathcal{A} = \{\alpha \in \mathbb{C}^{\mathbb{Z}} : \alpha_j = 0, j \geq n \text{ for some } n, |\alpha_j| \leq K(1+\varepsilon)^j, j \to -\infty\} \tag{4.16}$$

with ε as in (4.12). The inner product of $\alpha \in \mathcal{A}$ with an $x \in \mathcal{X}$ is defined by

$$\alpha^t x := \sum_{j=-\infty}^{\infty} x_j \alpha_j. \tag{4.17}$$

The conditions placed on $\mathcal{A}$ guarantee that $\alpha^t x$ is well defined. The pseudometric d on $\mathcal{P}$ is defined by

$$d(P, Q) = \sup_{\alpha \in \mathcal{A}, b \in \mathbb{R}} |P(\{x \in \mathcal{X} : \alpha^t x \leq b\}) - Q(\{x \in \mathcal{X} : \alpha^t x \leq b\})|. \tag{4.18}$$

It can easily be checked that d satisfies the conditions of Theorem 3.1.

We take the parameter space Θ to be the space of finite distribution functions F on $(-\pi, \pi]$ which may be identified with the corresponding Toeplitz form

$$r(n) = \int_{(-\pi,\pi]} \exp(in\lambda) dF(\lambda), n \in \mathbb{Z}. \tag{4.19}$$

The group H_G is defined as follows. For $F \in \Theta$ and $g \in G$ we define $h_g(F)$ by

$$h_g(F) = F_g \text{ where } dF_g(\lambda) = |g(\exp(i\lambda))|^2 dF(\lambda). \tag{4.20}$$

Finally the pseudometric D on Θ is defined by

$$D(F_1, F_2) = \begin{cases} \int_{-\pi}^{\pi} \left|\log\left(\frac{dF_1}{dF_2}\right)\right| d\lambda & F_1 \asymp F_2 \\ \infty & \text{otherwise} \end{cases} \tag{4.21}$$

where dF_1/dF_2 denotes the Radon-Nikodym derivative of F_1 with respect to F_2 and $F_1 \asymp F_2$ means that the two measures are absolutely continuous with respect to each other. The conditions placed on the group G imply that

$$\inf_{\lambda \in (-\pi,\pi]} |g(\exp(i\lambda))| > 0$$

from which it follows that F_g and F are mutually absolutely continuous for any $g \in G$ and any $F \in \Theta$. Furthermore

$$dF_g/dF = |g(\exp(i\lambda))|^2$$

from which it easily follows that

$$D(F, h_g(F)) = 2 \int_{-\pi}^{\pi} |\log(g(\exp(i\lambda)))| d\lambda \tag{4.22}$$

for any F in Θ and $g \in G$. Finally from (4.22) we have

$$D(F, h_{g^n}(F)) = 2n \int_{-\pi}^{\pi} |\log(g(\exp(i\lambda)))| d\lambda.$$

The continuity properties of $|g(\exp(i\lambda)|$ imply

$$\lim_{n\to\infty} n \int_{-\pi}^{\pi} |\log(g(\exp(i\lambda)))| d\lambda = \infty$$

unless $|g(\exp(i\lambda))| = 1, -\pi < \lambda \leq \pi$. This however would imply $g(z) = z$ and so we see that (3.6) holds for any g which is not the identity. Hence D also satisfies all the assumptions of Theorem 3.1. G_0 is seen to consist only of the identity. The theorem then implies

$$\varepsilon^*(T, P, d, D) \leq (1 - \Delta(P))/2.$$

5. Attaining the bound.

5.1. *Location functionals.* In Section 4.1 we proved that the maximum breakdown point for translation equivariant location functionals is 1/2. This bound is sharp as is shown by the location equivariant L_1-functional

$$T(P) = \mathrm{argmin}_\mu \int (\|x - \mu\| - \|x\|)dP(x). \tag{5.1}$$

In general the L_1-functional is not regarded as a satisfactory location functional as these are often required to be affinely equivariant. The attempt to prove Theorem 3.1 for affine equivariant functionals fails as there are affine equivariant location functionals with a higher breakdown point than would be suggested by Theorem 3.1. This is most clearly seen in one dimension where the breakdown point of the median is 1/2 at all distributions. The cause of the failure is (3.6) which no longer holds. The upper bound of 1/2 remains valid but it is not clear whether this can be attained for dimensions greater than one. Work has been done in this direction but it is not conclusive (Rousseeuw (1983), Niinimaa, Oja and Tableman (1990), Lopuhaä and Rousseeuw (1991), Gordaliza (1991), Lopuhaä (1992), Donoho and Gasko (1992)).

We first point out that the bound 1/2 is not globally sharp. Take a discrete measure in $\mathbb{R}^2$ with point mass 1/3 on the points $x_1 = (0, 1), x_2 = (0, -1), x_3 = (\sqrt{3}, 0)$. The points form a regular simplex. For symmetry reasons every affinely equivariant location functional must yield the value $(1/\sqrt{3}, 0)$. Replacing now $(\sqrt{3}, 0)$ by $(\eta\sqrt{3}, 0)$, it is easily shown that each affinely equivariant location functional must result in $(\eta/\sqrt{3}, 0)$. On letting $\eta \to \infty$ it follows that the breakdown point of every affinely equivariant location functional cannot exceed 1/3. In k dimensions one can prove in a similar manner that $1/(k+1)$ is the maximal breakdown point for points on a regular simplex with $k+1$ sides.

In spite of the above example we now show that there are probability distributions at which the finite sample replacement breakdown point is 1/2 even if this cannot be obtained globally. We consider a sample $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ of size n in $\mathbb{R}^k$ and form the empirical measure P_n given by $P_n = 1/n \sum_{i=1}^n \delta_{x_i}$. To obtain our goal we define an appropriate affinely equivariant location functional T at P_n^A for all affine transformations A and also at all measures of the form P_n^{*A}. Here P_n^* is any empirical measure obtained from $\boldsymbol{x}_n$ by altering the values of at most $\lfloor (n-1)/2 \rfloor$ of the x_i. The new sample will be denoted by $\boldsymbol{x}_n^* = (x_1^*, \ldots, x_n^*)$. We have to show that the values of $T(P_n^{*A})$ can be defined in such a way that

$$T(P_n^A) = A(T(P_n)) \tag{5.2}$$

$$T(P_n^{*A}) = A(T(P_n^*)) \tag{5.3}$$

and

$$\sup_{P_n^*} |T(P_n) - T(P_n^*)| < \infty. \tag{5.4}$$

This is done in Appendix A.

We note that the Sample conditions 1 and 2 in Appendix A are satisfied for an i.i.d. Gaussian sample of size n if n is sufficiently large. We indicate how this may be shown for Sample condition 2 in Appendix B.

5.2. *Scatter functionals.* The results given in this section are new even for the one dimensional case. In example (1.1) the median absolute deviation MAD has a finite sample breakdown point of 1/11 compared with the upper bound of 3/11 given by Theorem 3.1. We propose a modification of the median absolute deviation which does attain the upper bound.

For a probability measure P we define the interval $I(P, \lambda)$ by

$$I(P, \lambda) = [\mathrm{med}(P) - \lambda, \mathrm{med}(P) + \lambda]$$

and write

$$\Delta(P, \lambda) = \max\{P(\{x\}) : x \in I(P, \lambda)\}.$$

The new scale functional MAD* is defined by

$$\mathrm{MAD}^*(P) = \min\{\lambda : P(I(P, \lambda)) \geq (1 + \Delta(P, \lambda))/2\}.$$

Suppose we use the Kuiper metric d_1. Then as in Davies (1993) we find that

$$\varepsilon^*(\mathrm{MAD}^*, P, d_1, D) = (1 - \Delta(P))/3. \tag{5.5}$$

We now show that for the metric d_3 the upper bound for the breakdown point is attained. Given $\eta > 0$ we define $\lambda^*(P, Q, \eta)$ by

$$\lambda^*(P, Q, \eta) = \inf\{\lambda : P(I(Q, \lambda)) \geq 1 - \eta\}. \tag{5.6}$$

To ease the notation we write λ^* for $\lambda^*(P, Q, \eta)$ of (5.6). Suppose now that Q is such that

$$d_3(P, Q) \leq (1 - \Delta(P))/2 - \delta \tag{5.7}$$

with $\delta > 0$. As the breakdown point of the median is 1/2 it follows that

$$\sup\{\lambda^* : d_3(P, Q) \leq (1 - \Delta(P))/2 - \delta\} < \infty. \tag{5.8}$$

Let $\Delta(P, \lambda^*)$ and $\Delta(Q, \lambda^*)$ respectively denote measures of the largest G-invariant sets of P and Q in $I(Q, \lambda^*)$.

(1) Case 1: $\Delta(Q, \lambda^*) \leq \Delta(P, \lambda^*)$

From (5.7) it follows that

$$\begin{aligned} Q(I(Q,\lambda^*)) &\geq P(I(Q,\lambda^*)) - (1-\Delta(P))/2 + \delta \\ &\geq 1 - \eta - (1-\Delta(P))/2 + \delta \\ &\geq (1+\Delta(P))/2 + \delta - \eta \\ &\geq (1+\Delta(P,\lambda^*))/2 + \delta - \eta \\ &\geq (1+\Delta(Q,\lambda^*))/2 + \delta - \eta. \end{aligned} \tag{5.9}$$

As η may be chosen to be less than δ it follows from (5.8) that

$$\sup\{\mathrm{MAD}^*(Q) : d_3(P,Q) \leq (1-\Delta(P))/2 - \delta\} < \infty. \tag{5.10}$$

(2) Case 2: $\Delta(Q, \lambda^*) = \Delta(P, \lambda^*) + \gamma, \gamma > 0$

Let x be a point in $I(Q, \lambda^*)$ with $Q(\{x\}) = \Delta(Q, \lambda^*)$ and set $I \setminus \{x\} = I_1 \cup I_2$ for intervals I_1 and I_2. From the definition of d_3 it follows that

$$|Q(I_1) - P(I_1)| + |Q(\{x\}) - P(\{x\})| + |Q(I_2) - P(I_2)| \leq (1-\Delta(P))/2 - \delta.$$

As

$$|Q(\{x\}) - P(\{x\})| = \Delta(Q,\lambda^*) - P(\{x\}) \geq \Delta(Q,\lambda^*) - \Delta(P,\lambda^*) = \gamma$$

we may deduce

$$Q(I_1 \cup I_2) \geq P(I_1 \cup I_2) - (1-\Delta(P))/2 + \delta + \gamma.$$

It follows that

$$
\begin{aligned}
Q(I(Q,\lambda^*)) &\geq P(I_1 \cup I_2) - (1-\Delta(P))/2 + \delta + \gamma + \Delta(Q,\lambda^*) \\
&= P(I(Q,\lambda^*)) - P(\{x\}) - (1-\Delta(P))/2 + \delta + \gamma + \Delta(Q,\lambda^*) \\
&\geq 1 - \eta - \Delta(P,\lambda^*) - (1-\Delta(P))/2 + \delta + \gamma + \Delta(Q,\lambda^*) \\
&= 1 - \eta - (1-\Delta(P))/2 + \delta + 2\gamma \\
&= (1+\Delta(P))/2 - \eta + \delta + 2\gamma \\
&= (1+\Delta(Q)-\gamma)/2 - \eta + \delta + 2\gamma \\
&= (1+\Delta(Q))/2 - \eta + \delta + 3\gamma/2 \\
&\geq (1+\Delta(Q))/2. \qquad (5.11)
\end{aligned}
$$

for η sufficiently small. Again we have

$$\sup\{\mathrm{MAD}^*(Q) : d_3(P,Q) \leq (1-\Delta(P))/2 - \delta\} < \infty. \tag{5.12}$$

From (5.10) and (5.12) it follows that

$$\sup\{\mathrm{MAD}^*(Q) : d_3(P,Q) \leq (1-\Delta(P))/2 - \delta\} < \infty \tag{5.13}$$

which shows that $\mathrm{MAD}^*(Q)$ does not explode for Q satisfying (5.7) for any fixed $\delta > 0$. It is not difficult to show that $\mathrm{MAD}^*(Q)$ does not implode so that we have

$$\varepsilon^*(\mathrm{MAD}^*, P, d_3, D) = (1-\Delta(P))/2 \tag{5.14}$$

and the upper bound is attained globally.

In higher dimensions it seems plausible that by an appropriate choice of the metric and by a similar modification of known functionals (minimum volume ellipsoid, minimum covariance determinant) the upper bound given by Theorem 3.1 can be attained globally. We do not consider this any further.

5.3. *Regression functionals.* The remarks at the end of the last section also apply here. We suspect that by an appropriate choice of the metrics and by a modification of known functionals (least median of squares) the upper bound given by Theorem 3.1 can be attained globally. Again we do not consider this any further.

5.4. *Time series.* Here we have no results to offer for the simple reason that we are not aware of any functional T which is well defined on $\mathcal{P}$. It may be a topic worthy of further research.

6. Two further examples.

6.1. *Logistic regression.* We now consider the logistic regression model defined by

$$P(Y=1|x) = \exp(\theta_0 + x^t\tilde{\theta})/(1+\exp(\theta_0 + x^t\tilde{\theta})),\ \theta = (\theta_0, \tilde{\theta}^t)^t \in \mathbb{R}^{k+1}, \tag{6.1}$$

where $x^t = (x_1, \ldots, x_k)$ are the covariates associated with the random variable Y. The sample space is given by

$$\mathcal{X} = \{0,1\} \times \mathbb{R}^k \tag{6.2}$$

and $\mathcal{P}$ is the set of all nondegenerate probability measures on the Borel sets of $\mathcal{X}$. The metric d on $\mathcal{P}$ is defined by

(6.3)

$$d(P,Q) = \sup\{|P(B) - Q(B)| : B = \{(y, x^t)^t : u_0 y + x^t u \le v\}, u \in \mathbb{R}^k, u_0, v \in \mathbb{R}\}.$$

The parameter space Θ is $\mathbb{R}^{k+1}$. The group G on $\mathcal{X}$ is generated by the compositions of transformations of the form

$$(y, x^t)^t \to (1 - y, x^t)^t \tag{6.4}$$

$$(y, x^t)^t \to (y, \mathcal{A}(x)^t)^t, \tag{6.5}$$

where $\mathcal{A}$ is a nonsingular affine transformation $\mathcal{A}(x) = Ax + a$. The group H_G of transformations of Θ induced by G is given by

$$h_g(\theta) = -\theta, \quad g \text{ as in } (6.4) \tag{6.6}$$

$$h_g((\theta_o, \tilde{\theta}^t)^t) = (\theta_0 - a^t(A^t)^{-1}\tilde{\theta}, ((A^t)^{-1}(\tilde{\theta}))^t)^t, \quad g \text{ as in } (6.5) \tag{6.7}$$

The metric D on Θ will be defined as in Section 3.3 for a suitable choice of P. We define P as follows. Under P the random variable $(Y, X^t)^t$ is such that X is $N(0, I_k)$ and, given $X = x$, the random variable Y satisfies (6.1) with $\theta = (1, \ldots, 1)^t$. With this P and g such that $\mathcal{A}(x) = Ax + a$, $\det A > 1$, under P^g, the random variable X is $\mathcal{N}(0, A^tA)$. Hence under P^{g^n} it is $\mathcal{N}(0, (A^tA)^n)$. From this it follows $\lim_{n\to\infty} d(P^{g^n}, P) = 1$ so that (2.1) and (2.2) are satisfied. All the conditions for Theorem 3.1 are satisfied apart from the condition (3.6) which is not satisfied. Indeed if we define the functional $T : \mathcal{P} \to \Theta$ by $T(P) = 0$ for all P then it is seen that $T(P^g) = h_g(T(P)) = h_g(0) = 0$ and T is equivariant with respect to the group structure. As a constant functional its breakdown point is one.

The parameter value $\theta = 0$ corresponds to Y being $Bin(1, 0.5)$ independent of X. Some authors have tried to alter the definition of breakdown point in such a way that $\theta = 0$ is regarded as a breakdown. We are not convinced by these arguments. Firstly, if observed y-values have little dependence on the observed x-values (toss a coin and take the x-values to be the share prices of all companies listed on the London Stock Exchange at the time of tossing the coin) then a value of $\theta = 0$ would seem perfectly reasonable. It is not clear why it should be judged as a breakdown. Secondly, even if all the y values are one the probability under the model with $\theta = 0$ is 2^{-n} which is not zero. To say this is a breakdown seems rather odd for $n = 3$ say, irrespective of everything else. Of course one could define breakdown if $2^{-n} < 10^{-10}$ but this is rather arbitrary. It is similar to saying a location functional has broken down if its value exceeds a certain specified finite bound. If we accepted such a definition for location functionals then the theory of breakdown, affine equivariance etc. would itself be broken down. Breakdown is a limiting behaviour and its elegance derives from this and an appropriate equivariance structure. It should not be confused with exceeding a finite but arbitrary bound.

6.2. *Nonlinear regression* We consider the model

$$Y = h(x, \theta) + \varepsilon \tag{6.8}$$

where $h(\cdot, \theta), \theta \in \Theta$, is a parametric family of functions. Stromberg and Ruppert (1992) proposed the following definition of finite sample breakdown point for the above model. Given a data set

$$\chi = ((x_1, y_1), \ldots, (x_n, y_n))$$

they define the upper breakdown point at x by

$$\varepsilon_+(x, h, \hat{\theta}, \chi) = \min_{0 \le m \le n} \left\{ \frac{m}{n} \sup_{\chi^m} h(x, \hat{\theta}(\chi^m)) = \sup_{\theta} h(x, \theta) \right\} \tag{6.9}$$

where χ^m denotes a sample obtained from χ by altering m points. The lower breakdown point $\varepsilon_-(x, h, \hat{\theta}, \chi)$ is defined similarly and the breakdown point at x is then given by

$$\varepsilon(x, h, \hat{\theta}, \chi) = \min\{\varepsilon_+(x, h, \hat{\theta}, \chi), \varepsilon_-(x, h, \hat{\theta}, \chi)\}.$$

Finally the finite sample breakdown point is defined by

$$\varepsilon(h,\hat{\theta},\chi) = \inf_{x}\ \varepsilon(x,h,\hat{\theta},\chi).$$

As Stromberg and Ruppert impose no restrictions on the functionals we may consider the constant functional $T(P) = \theta_0$. It is clear that this functional has a finite sample breakdown point of one.

APPENDIX

A. We consider the constraints imposed upon us when defining $T(P_n^*)$. We start with the internal constraints which apply to each P_n^* without reference to the other measures.

- **Case 1** $P_n^{*A_1} \neq P_n^{*A_2}$ for any two different affine transformations A_1 and A_2. This is seen to reduce to $P_n^{*A} \neq P_n^*$ for any affine transformation A which is not the identity. If this is the case then there are no restrictions on the choice of $T(P_n^*)$. Having chosen it we extend the definition of T to all the measures P_n^{*A} by $T(P_n^{*A}) = A(T(P_n^*))$.

- **Case 2** $P_n^{*A} = P_n^*$ for some affine transformation A which is not the identity. If this is the case then A is unique and there exists a permutation π of $\{1,\ldots,n\}$ such that $A(x_i) = x_{\pi(i)}$. This implies that for each i we can form cycles

$$(x_i, A(x_i), \ldots, A^{m_i-1}(x_i))$$

with $A^{m_i}(x_i) = x_i$. From this we see that for some sufficiently large m $A^m(x_i) = x_i$ for all i. On writing $A(x) = \alpha(x) + a$ we see that if the $x_i, i = 1,\ldots,n$, span $\mathbb{R}^k$ then $\alpha^m = I$ where I denotes the identity transformation on $\mathbb{R}^k$. This implies that α must be an orthogonal transformation and that

$$\text{(A.1)} \qquad \sum_{j=0}^{m-1} \alpha^j(a) = 0.$$

It follows that if we set $T(P_n^*) = \mu$, we must have $A(\mu) = \mu$ for any affine transformation for which $P_n^{*A} = P_n^*$. The choice of μ is arbitrary subject only to these constraints. Having chosen such a μ the values of $T(P_n^{*B})$ are defined to be $B(\mu)$ for all other affine transformations B.

The above argument shows the internal consistency relationships which must be placed on T so that $T(P_n^{*A}) = A(T(P_n))$ for any affine transformation A. We now consider what one may call the external restrictions.

- **Case 3** Suppose that P_n^* is such that there does not exist a $P_n'^*$ and an affine transformation A such that $P_n^{*A} = P_n'^*$. In this case the choice of $T(P_n^*)$ is only restricted by the considerations of Case 2 above if that case applies and otherwise not at all.

- **Case 4** Suppose that P_n^* is such that there exists a $P_n'^*$ and an affine transformation A such that $P_n^* = P_n'^{*A}$. In this case we require $T(P_n^*) = A(T(P_n'^*))$.

We now place the following conditions on the sample $\boldsymbol{x}_n$:

Sample condition 1: There do not exist two distinct subsets of $\boldsymbol{x}_n$ each of size at least $k+2$ and an affine transformation A which transforms one subset into the other.

Sample condition 2: If

$$|A(\boldsymbol{x}_n) \cap B(\boldsymbol{x}_n)| \geq \lfloor (n+1)/2 \rfloor - 2k$$

for two affine transformations A and B then $A = B$.

Sample condition 3: $k < \lfloor (n-1)/2 \rfloor$.

We now construct a functional T which satisfies (5.2), (5.3) and (5.4). If the sample conditions hold then for any affine transformation $A \neq I$ we have $P_n^A \neq P_n^*$ where P_n^* derives from a subset $\boldsymbol{x}_n^*$ which differs from $\boldsymbol{x}_n$ by at least one and at most $\lfloor (n-1)/2 \rfloor$ points. This follows on noting that at most $k+1$ of the $A(x_i)$ belong to $\boldsymbol{x}_n$ by Sample condition 1. Because of this we can define the $T(P_n)$ without reference to the values of $T(P_n^*)$. We set

$$T(P_n) = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

If P_n^* satisfies the conditions of Case 3 above we set

$$T(P_n^*) = \frac{1}{n^*} \sum_{i=1}^{n^*} x_{\pi(i)}$$

where the $x_{\pi(i)}$ are those $n^* \geq \lceil (n+1)/2 \rceil$ points of the sample $\boldsymbol{x}_n$ which also belong to the sample $\boldsymbol{x}_n^*$. Finally we consider Case 4 above. We show that the sample assumptions and the condition $P_n^* = P_n'^{*A}$ uniquely determine the affine transformation A. To see this we suppose that there exists a second affine transformation B and a distribution $P_n''^*$ such that $P_n^* = P_n''^{*B}$. Let $x^*_{\pi(1)}, \ldots, x^*_{\pi(N')}$ denote those points of $\boldsymbol{x}_n^*$ not contained in the sample $\boldsymbol{x}_n$. Because of Sample condition 1 this set contains at least $\lceil (n+1)/2 \rceil - k - 2$ points of the form $A(x_i)$. Similarly it also contains at least $\lceil (n+1)/2 \rceil - k - 2$ points of the form $B(x_i)$. The intersection of these two sets is of size at least $\lfloor (n+1)/2 \rfloor - 2k$ and we may conclude from Sample condition 2 that $A = B$. The representation is therefore unique. Let $x_{\pi(1)}, \ldots x_{\pi(m)}$ be those points of $\boldsymbol{x}_n$ which belong to the sample $\boldsymbol{x}_n'^*$ and for which $A(x_{\pi(1)}), \ldots, A(x_{\pi(m)})$

belong to the sample $\boldsymbol{x}_n$. It is clear that $m \geq 1$. We define

$$T(P_n'^*) = \frac{1}{m}\sum_{i=1}^{m} x_{\pi(i)}$$

and by equivariance

$$T(P_n^*) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{A}(x_{\pi(i)}).$$

It follows that $T(P_n^*)$ is well defined and in both cases the sums involved come from the sample $\boldsymbol{x}_n$. The functional T is extended to all $P_n{*}^B$ and $P_n'{*}^B$ by affine equivariance. In all cases the definition of $T(P_n^*)$ is as the mean of a subset of $\boldsymbol{x}_n$. From this it is clear that (5.4) is satisfied.

B. Let $\mathcal{A} = A + a$ and $\mathcal{B} = B + b$ with A and B nonsingular matrices and a and b points in $\mathbb{R}^k$. We suppose that $A \neq B$. On taking differences we see that there exist sample points $X_{i_1}, \ldots, X_{i_{k+1}}$ and $X_{j_1}, \ldots, X_{j_{k+1}}$ such that

$$A(X_{i_l} - X_{i_{k+1}}) = B(X_{j_l} - X_{j_{k+1}}), j = 1, \ldots, k.$$

This implies that $B^{-1}A$ and $B^{-1}(b-a)$ are functions of the chosen sample points

$$B^{-1}A = C(X_{i_1}, \ldots, X_{i_{k+1}}, X_{j_1}, \ldots, X_{j_{k+1}}) \tag{B.1}$$

$$B^{-1}(b-a) = c(X_{i_1}, \ldots, X_{i_{k+1}}, X_{j_1}, \ldots, X_{j_{k+1}}).$$

For n sufficiently large there exist four further sample points $X_i, i = 1, \ldots, 4$ which are not contained in $\{X_{i_1}, \ldots, X_{i_{k+1}}, X_{j_1}, \ldots, X_{j_{k+1}}\}$ and for which

$$A(X_1) + a = B(X_2) + b, \quad A(X_3) + a = B(X_4) + b.$$

This implies

$$B^{-1}A(X_3 - X_1) = X_4 - X_2. \tag{B.2}$$

However as the $X_i, i = 1, \ldots, 4$, are independent of $X_{i_1}, \ldots, X_{i_{k+1}}, X_{j_1}, \ldots, X_{j_{k+1}}$ it follows from (B.1) that (B.2) holds with probability zero. From this we conclude that $A = B$. Similarly we can show that $a = b$ and hence $\mathcal{A} = \mathcal{B}$.

REFERENCES

Croux, C. and Rousseeuw, P. J. (1992). A class of high-breakdown estimators based on subranges. *Commun. Statist. - Theory Meth.* **21**, 1935–1951.

de Luna, X. and Genton, M. G. (2001). Robust simulation-based estimation of ARMA models. *Journal of Computational and Graphical Statistics* **10**, 370–387.

Davies, P. L. (987) Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics*,**15**, 1269-1292.

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

Davies, P. L. and Gather, U. (1993). The identification of multiple outliers. *J. Amer.Statist. Assoc.* **88**, 782– 792.

Davies, P. L. and Kovac, A. (2002). Densities, spectral densities and modality. *Preprint. Department of Mathematics and Informatics, University of Essen.*

Dietel, G. (1993). Global location and dispersion functionals. *Ph. D. qualifying paper, Department of Mathematics and Informatics, University of Essen.*

Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. *Ph. D. qualifying paper, Dept. Statistics, Harvard University.*

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

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

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

Gordaliza, A. (1991). On the breakdown point of multivariate location estimators based on trimming procedures. *Statist. Probab. Letters* **11**, 387–394.

Hampel, F. R. (1968). Contributions to the theory of robust estimation. Ph. D. thesis, Dept. Statistics, Univ. California, Berkeley.

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

He, X. and Fung, W. K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *J. Multivariate Anal.* **72**, 151–162.

He, X. and Simpson, D. G. (1992). Robust direction estimation. *Ann. Statist.* **20**, 351–369.

He, X. and Simpson, D. G. (1993). Lower bounds for contamination bias: Globally minimax versus locally linear estimation. *Ann. Statist.* **21**, 314–337.

Hodges, J. L. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* 163–186. Univ. California Press, Berkeley.

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

Huber, P. J. (1995). Robustness: Where are we now? *Student* **1**, 75–86.

Hubert, M. (1997). The breakdown value of the L_1 estimator in contingency tables. *Statist. Probab. Letters* **33**, 419–425.

Kent, J. T. and Tyler, D. E. (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.* **19**, 2102–2119.

Lopuhaä, H. P. (1992). Highly efficient estimators of multivariate location with high breakdown point. *Ann. Statist.* **20**, 398–413.

Lopuhaä, H. P. and Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19**, 229–248.

Lucas, A. (1997). Asymptotic robustness of least median of squares for autoregressions with additive outliers. *Commun. Statist.–Theory Meth.* **26**, 2363–2380.

Ma, Y. and Genton, M. G. (2000). Highly robust estimation of the autocovariance function. *J. Time Series Analysis* **21**, 663–684.

Martin, R. D. and Jong, J. (1977). Asymptotic properties of robust generalized M-estimates for the first order autoregressive parameter. *Bell Laboratories Technical Memo.* Murray Hill, New Jersey.

MENDES, B. V. M. (2000). Asessing the bias of maximum likelihood estimates of contaminated GARCH models. *J. Statist. Comput. Simul.* **67**, 359–376.

MÜLLER, C. H. (1995). Breakdown points for designed experiments. *Journal of Statistical Planning and Inference* **45**, 413–427.

NIINIMAA, A., OJA, H. and TABLEMAN, M. (1990). The finite-sample breakdown point of the Oja bivariate median and of the corresponding half-samples version. *Statist. Probab. Letters* **10**, 325–328.

PAPANTONI-KAZAKOS, P. (1984). Some aspects of qualitative robustness in time series. In *Robust and Nonlinear Time Series* (J. Franke, W. Härdle, and R. D. Martin, eds.) 218–230, Springer, New York.

ROUSSEEUW, P. (1983). Multivariate estimation with high breakdown point. *Proc. of the Fourth Pannonian Symp. on Math. Stat.* (W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, eds.) 283–297. Bad Tatzmannsdorf, Austria.

ROUSSEEUW, P, (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**, 871-880.

RUCKSTUHL, A. F. and WELSH, A. H. (2001). Robust fitting of the binomial model. *Ann. Statist.* **29**, 1117–1136.

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.

STROMBERG, A. J. and RUPPERT, D. (1992). Breakdown in nonlinear regression. *J. Amer. Statist. Assoc.* **87**, 991–997.

TATUM, L. G. and HURVICH, C. M. (1993). High breakdown methods in time series analysis. *J. R. Statist. Soc. B* **55**, 881–896.

TERBECK, W. and DAVIES, P. L. (1998). Interactions and outliers in the two-way analysis of variance. *Ann. Statist.* **26**, 1279–1305.

VON MISES, R. (1937). Sur les fonctions statistiques. In: *Conf. de la Réunion Internationale des Math, Gauthier-Villars, Paris*, 1–8.

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18**, 309–348.