References

[1] Hall, P. (1985). *Stoch. Processes Appl.*, **20**, 231–246. (Suggests both fixed- and moving-block bootstrap methods in the setting of spatial data.)

[2] Carlstein, E. (1986). *Ann. Statist.*, **14**, 1171–1179. (Suggests a form of the fixed-block bootstrap for time series data.)

[3] Künsch, H. R. (1989). *Ann. Statist.*, **17**, 1217–1241. (Suggests the moving-block bootstrap for time series data.)

[4] Freedman, D. (1984). *Ann. Statist.*, **12**, 827–842. (An early contribution to bootstrap methods for dependent data in the setting of structural models.)

[5] Politis, D. N. and Romano, J. P. (1994). *J. Amer. Statist. Assoc.*, **89**, 1303–1313. (Suggests the stationary bootstrap for time series data.)

[6] Hall, P. and Jing, B. (1996). *J. R. Statist. Soc. B*, **58**. (Suggests moving-window methods for the bootstrap with dependent data.)

(BOOTSTRAPPING—II
SPATIAL DATA ANALYSIS
SPATIAL PROCESSES
TIME SERIES)

PETER HALL

BREAKDOWN POINT

In many statistical estimation situations, the only data available are possibly contaminated by recording errors, miscoding, or some other factors. The *breakdown point* of an estimator is, roughly speaking, the minimum proportion of the data for which contamination can lead to a completely noninformative estimation result. It is a measure of the resistance* of the estimator to data contamination; the higher the breakdown point is, the more resistant the estimator is to data contamination.

HODGES' TOLERANCE

If a location estimate is attracted to a small number of extreme values, the estimate fails to convey location information about the other observations. A natural way of defining the breakdown point in location estimation is thus to consider what happens to the estimate if some of the observations are replaced with positively or negatively extreme values. The *tolerance* proposed by Hodges [4] formally implements this idea.

Let $\hat{\theta}$ be a location estimate of interest, and $(y_1, y_2, \ldots, y_n)$ the data set of arbitrary real numbers sorted in ascending order. Suppose that there exist integers $m_l \geq 0$, $m_r \geq 0$ such that: (1) $y_{m_l+1} \leq \hat{\theta} \leq y_{n-m_r}$ for any $(y_1, y_2, \ldots, y_n)$; (2) when $(y_{m_l+2}, y_{m_l+3}, \ldots, y_n)$ are fixed, for any sequence of the replacements of $(y_1, y_2, \ldots, y_{m_l+1})$ along which $y_{m_l+1} \to -\infty$ with the order of $(y_1, y_2, \ldots, y_{m_l+1})$ being preserved, $\hat{\theta} \to -\infty$; and (3) when $(y_1, y_2, \ldots, y_{n-m_r-1})$ are fixed, for any sequence of the replacements of $(y_{n-m_r}, y_{n-m_r+1}, \ldots, y_n)$ along which $y_{n-m_r} \to \infty$ with the order of $(y_{n-m_r}, y_{n-m_r+1}, \ldots, y_n)$ being preserved, $\hat{\theta} \to \infty$. Then it is said that $\hat{\theta}$ can tolerate m_l extreme values on the left and m_r extreme values on the right.

Example 1. Both the sample average and the sample median are location estimators. Let n be the sample size. Then the sample average can tolerate no extreme values either on the left or on the right, while the sample median can tolerate $\lceil n/2 \rceil - 1$ extreme values both on the left and on the right, where $\lceil a \rceil$ is the minimum integer greater than or equal to a for each real number a.

HAMPEL'S BREAKDOWN POINT

Unlike Hodges, Hampel [3] considers the effect of data contamination strictly in the probabilistic framework. Suppose that the "true" distribution is F in a class $\mathcal{G}$ of probability distributions. The data of interest would be generated by random draws from probability distribution F, if there were no data contamination. Actual draws are from the probability distribution $G \in \mathcal{G}$ in some neighborhood of F (in terms of Prokhorov distance), due to data contamination. Let $B_\delta(F)$ be the set of all probability distributions in $\mathcal{G}$ within distance δ of F. Also let $\hat{\theta}_n$ be the estimator of interest with sample size n.

Suppose that G can be any member of $B_\delta(F)$. If the probability limit of $\hat{\theta}_n$ as $n \to \infty$ can be any point in the parameter space Θ depending on G in $B_\delta(F)$, the estimator fails to convey any information about F. On the other hand, if there is a proper subset K of Θ such that $G \in B_\delta(F)$ implies that the probability of the estimator $\hat{\theta}_n$ belonging to K converges to one as the sample size grows to ∞, then $\hat{\theta}_n$ is considered to convey some information about F when the observations are drawn from any G in $B_\delta(F)$. The supremum of δ for which the estimator $\hat{\theta}_n$ is informative about F when the observations are drawn from any probability distribution in $B_\delta(F)$ is the *Hampel breakdown point.*

Example 2. Let $\mathcal{G}$ be the set of all probability distributions on the entire real line with finite first absolute moment. Suppose that $F \in \mathcal{G}$. Pick an arbitrary $\delta \in (0, 1)$. Consider a probability distribution G created by reducing by 100δ percent the probability of each event specified by F, and putting probability δ on some positive point on the real line. By moving the point to which probability δ is added, we can set the expectation of the probability distribution G to any value we like. Because the sample average converges to the expectation of G by the Kolmogorov law of large numbers, this implies that the Hampel breakdown point is no higher than δ. Because δ is an arbitrary positive number, this implies that the Hampel breakdown point is zero for the sample average. In the same situation, the Hampel breakdown point of the sample median is 50%.

DONOHO–HUBER FINITE-SAMPLE BREAKDOWN POINT

Although Hampel's breakdown point can be applied to general parametric estimation problems, it only gives us insight into the large-sample behavior of the estimators. On the other hand, Hodges' tolerance, which is defined without recourse to a probability model, is a sensible measure for the resistance of the estimator with each fixed sample size, though it is only applicable to location estimators. While preserving its useful properties for finite sample sizes, Donoho and Huber [1] extend Hodges' tolerance to cover a wider range of problems. This notion is called the *finite-sample breakdown point.* Among the three versions of the finite-sample breakdown point proposed in ref. [1], we first explain the one with "ϵ-replacement."

Let $Q \subset \mathbb{R}^v$ be a sample space, where v is a natural number. Let n denote the sample size. For given $m \in \{0, \ldots, n\}$ and a given data set $\mathbf{z}^n \in Q^n \equiv \times_{i=1}^n Q$, consider choosing m arbitrary observations in $\mathbf{z}^n$ and replacing each of them with an arbitrary point in Q. Let $D_m^n(\mathbf{z}^n)$ denote the set of all data sets with m replaced observations and $n - m$ original observations.

Let a given parameter space Θ be a subset of a metric space with metric d (e.g., a Euclidean space, as assumed in Donoho and Huber [1]). The finite-sample breakdown point of an estimator $\hat{\theta}_n : Q^n \to \Theta$ at a data set $\mathbf{z}^n \in Q^n$ is defined as the minimum element of the following set of fractions:

$$\left\{ \frac{m}{n} \quad b(m, n; \mathbf{z}^n, \hat{\theta}_n \qquad \infty,\ m \in \qquad n\} \right| \qquad (1)$$

where $b(m, n; \mathbf{z}^n, \hat{\theta}_n) \equiv \sup_{z^n \in D_m^n(\mathbf{z}^n)} d(\hat{\theta}_n(z^n), \hat{\theta}_n(\mathbf{z}^n))$. We call this finite-sample breakdown point the *Donoho–Huber breakdown point with ϵ-replacement* (or simply the Donoho–Huber breakdown point) for convenience and clarity in what follows.

The value $b(m, n; \mathbf{z}^n, \hat{\theta}_n)$ can be considered to be the maximum bias (more accurately, the supremum bias) caused by m observation replacements. By checking this maximum bias for $m = 1$, $m = 2$, and so on, one may find that the maximum bias becomes ∞ for some m. Then m/n is the Donoho–Huber breakdown point with this m. When there is no $m \leq n$ for which the maximum bias is ∞, the breakdowh point is one. An example of an estimator whose Donoho–Huber breakdown point is one is the single-parameter estimator whose value is always a fixed value, say zero, regardless of the observations.

The Donoho–Huber breakdown point of an estimator is dependent on the noncontaminated data set $\mathbf{z}^n$ in general. Nevertheless, it takes the same value at all or at most possible data sets in many interesting applications. For this reason, the Donoho–Huber breakdown point of an estimator is often referred to without specifying what the noncontaminated data set z^n is.

Example 3. In the location estimation problem, the Donoho–Huber breakdown point is essentially the same as Hodges' tolerance if the metric of the location parameter space is Euclidean distance. Thus, the Donoho–Huber breakdown points of the sample average and the sample median are $1/n$ and $[n/2]/n$, respectively, where n is the sample size.

The other two versions of the breakdown points proposed by Donoho and Huber [1] are derived from the ϵ-replacement version by changing the way that the data are corrupted. In some situations, data corruption is caused by m erroneous observations adjoined to the original data set. Donoho and Huber call this type of data corruption *ϵ-contamination*. The breakdown point with ϵ-contamination is the minimum fraction of the contaminated observations, $m/(n + m)$, for which the distance between the estimate based on the contaminated data set can be arbitrarily far from that based on the original data set.

The third version of the Donoho–Huber breakdown point, unlike the previous two versions, does not limit the way in which the data are corrupted. All possible data corruption, including replacement and contamination, is considered. The degree of data corruption is measured by the distance between the empirical distribution of the modified data and that of the original data. The breakdown point is defined as the minimum data corruption level for which the estimate based on the modified data set can be arbitrarily far from that based on the original data set. Donoho and Huber call this breakdown point the finite-sample breakdown point with *ε-modification*.

As Donoho and Huber [1] mention, it is necessary to modify the definition of the three Donoho–Huber breakdown points to cover certain problems. Scale estimators* are an example of this. Suppose that a small number of contaminated observations can cause a scale estimate to be arbitrarily close to zero. We can call this behavior of the estimator *implosion*. The implosion of a scale estimator is as problematic as its explosion. The Donoho–Huber breakdown point ignores implosion, because the maximum bias in scale estimation caused in this way is never infinity. To take account of the possibility of implosion, the breakdown point with ε-replacement, for instance, needs to be redefined as m/n, where m is the minimum number of replaced observations for which the maximum bias is ∞ or for which the estimate can be arbitrarily close to zero.

EXTENSION OF THE DONOHO–HUBER BREAKDOWN POINT

Although the location estimation problem can be considered as a special case of the regression estimation problem, the Donoho–Huber breakdown point of a regression estimator may be very different from that of the corresponding location estimator. The L_1-estimator (the least-absolute-deviation estimator) is a good example. In the simple location setting L_1-estimation delivers the sample median. Nevertheless, as Ellis and Morgenthaler [2] precisely describe, the L_1-estimator for linear regression can be greatly affected by an observation which is far from the majority of observations in the regressor space. (Such an observation is a *leverage* point*.) It thus has the Donoho–Huber breakdown point $1/n$. Many other M-estimators* share the same property. On the other hand, the Donoho–Huber breakdown points of certain other estimators are known to have positive lower bound independent of n. For an excellent treatment of the properties of estimators in the linear regression setting, *see* ROBUST REGRESSION, POSITIVE BREAKDOWN.

A natural question that arises from studying the breakdown properties of linear regression estimators is what breakdown properties hold for estimators for other types of models. The Donoho–Huber breakdown point is inconvenient for answering this question, in that it is not invariant with respect to reparametrizations (possibly involving change of the metric of the parameter space). To see this point, consider the regression model

$$y = X\alpha + \epsilon,$$

where the set of all real numbers is denoted $\mathbb{R}$. $Y \in \mathbb{R}$ is the dependent variable, $X \in \mathbb{R}$ is the explanatory variable, α is the model parameter belonging to $\mathbb{R}$, and ε is an error term. The Donoho–Huber breakdown point of the least squares (LS) estimator for this model is known to be $1/n$, where n is the sample size. Nevertheless, we can rewrite this model with an alternative parametrization as

$$Y = X \tan^{-1}\theta + \epsilon, \qquad (2)$$

where $\theta \in (-\pi/2, \pi/2)$. If we employ Euclidean distance in this new parameter space $(-\pi/2, \pi/2)$, the Donoho–Huber breakdown point for the LS estimator is one with the new parametrization. Note that the LS estimator picks exactly the same regression function in the two estimation problems. This means that essentially the same estimator may have very different breakdown points depending on the parameterization.

The dependence of the breakdown point on the parametrization is not such a serious problem in this simple linear regression example. As long as we restrict the model to be a simple linear regression model, we would always use (1), because (2) is not linear in the parameter. Nevertheless, there are models for which parameterization may be arbitrary. Nonlinear regression models are an example of this. The dependence of the breakdown point on the parametrization makes the analysis unnecessarily complicated.

Note also that the bias considered by the Donoho–Huber breakdown point may be an inappropriate measure for the effects of data contamination in certain situations. Stromberg and Ruppert [6] give the following example. Consider the nonlinear regression* model known as the Michaelis–Menten model,

$$Y = \frac{VX}{K + X} + \epsilon, \qquad (3)$$

where $Y > 0$ is the dependent variable, $X > 0$ is the explanatory variable, ε is the error term, and V and K are parameters of this model, both of which can be any positive real numbers. Suppose that replacing m data points can drive both an estimate $\hat{V}$ for V and an estimate $\hat{K}$ for K to ∞, keeping their ratio $\alpha = \hat{V}/\hat{K}$ constant. This behavior of the parameter estimates does not completely invalidate the estimated regression function, which still retains information on the conditional location of Y given X. This point is made clear by comparing this case with that in which the ratio of the estimated parameters, α, can be arbitrarily close to zero or arbitrarily large. We thus need to modify the definition of the breakdown point in such estimation problems as (3).

Taking into account this limitation of the Donoho–Huber breakdown point, Stromberg and Ruppert [6] propose an alternative finite-sample breakdown point for nonlinear regression estimators. In Stromberg and Ruppert's version of the finite-sample breakdown point, the behavior of the estimated regression function is of concern instead of the estimated parameters. At each point in the explanatory-variable space, consider the possible range of the regression function obtained by letting the parameter range over the parameter space. If the estimated regression function can be arbitrarily close to the upper or lower boundary of its range at a certain contamination level, the estimator is considered to break down at the point in the explanatory variable space at the given contamination level. The breakdown point at that point in the explanatory-variable space is thus defined. The minimum value of the breakdown point over the points in the explanatory variable space is the *Stromberg–Ruppert breakdown point.*

This idea resolves the problems related to the dependence of the finite-sample breakdown point on the model parameterization, because the definition of the Stromberg–Ruppert breakdown point does not depend on the parameter space. Nevertheless, there are some cases in which it is inappropriate to judge the breakdown of the estimators in the way proposed by Stromberg and Ruppert, as the next example shows.

Example 4. A regression estimator is said to have the *exact fit property* if the estimate of the regression coefficient vector is $\bar{\theta}$ whenever more than the half of the observations are perfectly fitted by the regression function with $\bar{\theta}$. Some estimators (e.g., the least-median-of-squares estimator) are known to have the exact fit property.

Let $\Theta = [0, \infty)$, and consider the linear model regressing $Y \in \mathbb{R}$ on $X \in \mathbb{R}$:

$$Y = \theta X + \epsilon$$

where $\theta \in \Theta$, and ε is an error term. Suppose that we have a data set that consists of the following pairs of observations on Y and X: (0.0, −4.0), (0.0, −3.0), (0.0, −2.0), (0.0, −1.0), (0.001, 0.1), (0.01, 1.0), (0.02, 2.0), (0.03, 3.0), (0.04, 4.0). Then any estimator that has the exact fit property gives 0.01 as the estimate for θ with the given data set. The estimate, however, becomes 0 if we replace the fifth observation with (0.0, −0.1), i.e., the estimated regression function attains its lower bound at each $x \in (0, \infty)$ and its upper bound at each $x \in (-\infty, 0)$. Thus the Stromberg–Ruppert breakdown point of the estimator at the above dataset is $\frac{1}{9}$. For the dependent variable whose sample space is $\mathbb{R}$, would everyone agree that $\theta = 0$ is a crucially bad choice?

Although Stromberg and Ruppert mention the possibility that their breakdown point is inappropriate and one may instead need to use either the upper or the lower breakdown point similarly defined, both upper and lower breakdown points agree with the Stromberg–Ruppert breakdown point in the above example. This suggests that we may want to reconsider the criterion for crucially bad behavior of estimators. Also, the Stromberg–Ruppert breakdown point is not easily extended to parametric estimation problems other than nonlinear regression.

To overcome these limitations, Sakata and White [5] propose another version of the finite-sample breakdown point based on the fundamental insights of Donoho and Huber [1] and Stromberg and Ruppert [6]. This version of the breakdown point is like the Stromberg and Ruppert's in that it is based on the behavior of the estimated object of interest, e.g., a regression function, instead of the estimated parameters. For the Sakata–White breakdown point the model user specifies the criterion against which breakdown is to be judged. This criterion (e.g., the negative of goodness of fit) is the *badness measure*; the breakdown point is the *badness-measure-based breakdown point*. It constitutes the minimum proportion of observations for which date contamination leads to the worst value of the badness measure. Because of its flexibility, the Sakata–White breakdown point is generic. It can generate the Donoho–Huber breakdown point and the Stromberg–Ruppert breakdown point for appropriate badness measures. It can also be applied in nonregression contexts.

Table 1 summarizes the five breakdown points explained here.

Table 1 Comparison of Breakdown-Point Concepts

Breakdown Point	Applicable	Notion of Breakdown
Tolerance	Location	The estimate is arbitrarily (negatively or positively) large.
Hampel	Parametric in general	The stochastic limit of the estimator can be anywhere in the parameter space.
Donoho–Huber	Univariate and multivariate location, etc.	The estimate under data corruption can be arbitrarily far from that without data corruption.
Stromberg–Ruppert	Nonlinear regression	The estimated regression function can be arbitrarily close to its possible highest or lowest value at some point in the regressor space.
Sakata–White	Parametric in general	The estimated model can be arbitrarily bad in terms of the model user's criterion.

References

[1] Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann in Honor of His 65th Birthday*, P. J. Bickel, K. A. Doksum, and J. L. Hodges, Jr., eds. Wadsworth International Group, Belmont, Calif., pp. 157–184.

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

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

[4] Hodges, J. L., Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proc. Fifth Berkeley Symp. Math. Statist. and Probab.*, vol. 1, pp. 163–168.

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

[6] Stromberg, A. J. and Ruppert, D. (1992). Breakdown in nonlinear regression. *J. Amer. Statist. Ass.*, **87**, 991–997.

(ROBUST REGRESSION, POSITIVE BREAKDOWN)

SHINICHI SAKATA
HALBERT WHITE

BRIER SCORE

In Brier [2] this score is introduced as a means of evaluating (weather) forecasts expressed in terms of probability. Brier gave an example where forecasts of rain or no rain are available and a forecast consists of a probability of rain. The Brier score is used to evaluate these forecasts after the occurrence of rain or no rain has been observed. Hitherto the score has been applied frequently in meteorology* and fields such as medical diagnosis.*

Suppose that on each of n occasions exactly one out of r events can occur and that the forecast probabilities are $f_{i1}, f_{i2}, \ldots, f_{ir}$ on the ith occasion ($\sum_j f_{ij} = 1$). The Brier score is defined as

$$\sum_{i=1}^{n}\sum_{j=1}^{r}(f_{ij} - E_{ij})^2,$$

where E_{ij} takes the value 1 or 0 according to whether or not event j occurred on the ith occasion. The score is the average squared distance between the forecast probability distribution and the probability distribution of the perfect forecast, i.e., having probability one in the realized event. The score has a minimum value of zero (perfect forecasting) and a maximum value of 2 (worst possible forecasting).

To illustrate the formula, suppose that on ten occasions probability forecasts have been given of rain or no rain (i.e. $n = 10$ and $r = 2$). The forecasts are 0.7, 0.9, 0.8, 0.4, 0.2, 0, 0, 0, 0, 0.1, respectively, and it turned out to rain only on the second, third, and fourth occasions. The score for these forecasts is $\frac{1}{10} \times 2 \times (0.7^2 + 0.1^2 + 0.2^2 + 0.6^2 + 0.2^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0.1^2) = 0.19$ [2].

In some papers the Brier score is called "probability score" or "quadratic scoring rule." The latter term is somewhat confusing, as in Staël von Holstein and Murphy [6] a family of quadratic scoring rules is introduced, of which the Brier score (as well as the Epstein scoring rule*) is only a special case.

An attractive property of the Brier score is that it is *strictly proper*, i.e., a forecaster minimizes his expected score only by honestly expressing his personal probability assessments. Brier noted this and wrote that using this score "cannot influence the forecaster in any undesirable way." For a more detailed description of properness *see* DISTRIBUTIONAL INFERENCE.

The Brier score also can be written explicitly as the sum of *calibration* ("validity," "reliability") and *refinement* ("sharpness," "resolution") components (Murphy [3], Sanders [5]). The calibration component measures the extent to which forecast probabilities and observed frequencies correspond. The refinement component measures the extent to which, in a sequence of events receiving the same forecast probability, the occurrence of the event is uniquely determined (i.e. always or never). A graphical exposition of the relationship between calibration, refinement, and Brier score can be found in Blattenberger and Lad [1]. A different decomposition of the Brier score is given in Murphy [4].