The Annals of Statistics
1994, Vol. 22, No. 2, 1024–1044

FINITE SAMPLE BREAKDOWN POINTS OF PROJECTION BASED MULTIVARIATE LOCATION AND SCATTER STATISTICS[1]

BY DAVID E. TYLER

Rutgers University

Finite sample breakdown points are obtained for two classes of projection based multivariate location and scatter statistics: the Stahel-Donoho statistics and the Maronna-Yohai statistics. The definition of these multivariate statistics are dependent on the value of location and scale statistics for all univariate projections of the data, and consequently their properties depend on the nature of the corresponding univariate location and scale statistics used on the projected data. The finite sample breakdown points of the multivariate statistics, though, are not directly related to those of the corresponding univariate location and scale statistics. A uniform finite sample breakdown point concept is needed.

The median and the median absolute deviation about the median (M.A.D.) are one possible choice for the univariate location and scale statistics, respectively. For sparse data sets in high dimensions, though, it is not recommended that the M.A.D. be used as the univariate scale statistic for the projected data since its uniform finite sample breakdown point is shown to be much less than optimum. A simple modification to the M.A.D., however, is shown to alleviate this problem.

For various reasons, one may wish to consider univariate location and scale statistics other than the median and the M.A.D., respectively. A very broad and natural class of univariate location and scale statistics to consider for the projected data is the simultaneous M-estimates of location and scale. New results on their breakdown properties are given in this paper. Implicit formulas for the breakdown points of monotonic simultaneous M-estimates of location and scale are known, and they tend to imply rather low breakdown points for smooth choices of the defining weight functions. It is shown here that this phenomenon does not occur for a large class of nonmonotonic simultaneous M-estimates of location and scale. Furthermore, explicit rather than implicit expressions for the uniform finite sample breakdown points are given for these nonmonotonic M-estimates.

1. Introduction and summary. In the past decade, there has been considerable interest in the study of affine equivariant multivariate location and scatter statistics which have high breakdown points regardless of the dimensionality of the data. The first multivariate location and scatter statistics which were shown to be both affine equivariant and to have breakdown points near $\frac{1}{2}$ were independently introduced by Stahel (1981) and Donoho (1982). Subsequently, other high breakdown point affine equivariant multivariate location and scatter statistics have been introduced, with perhaps the most well known

Received May 1991; revised April 1993.
[1]Research supported by NSF Grants DMS-87-01198 and DMS-91-06706.
AMS 1991 *subject classifications.* 62F35, 62H12.
Key words and phrases. Breakdown, empirical processes, M-estimates, multivariate location and scatter, projection based statistics, redescending influence functions, robustness, S-estimates.

being Rousseeuw's (1985) minimum volume ellipsoid estimator (MVE) and its generalization, the multivariate S-estimator [see Davies (1987) and Lopuhaä (1989)]. More recently, a new class of high breakdown point affine equivariant scatter statistics has been introduced by Maronna and Yohai (1989). Both the Stahel–Donoho and the Maronna-Yohai statistics are projection based statistics. Their definitions depend on location and scale statistics for univariate projections of the data.

The breakdown point of an estimator or a statistic can be defined in a number of different ways. As originally defined by Hodges (1967) and Hampel (1968), the breakdown point was essentially an asymptotic rather than a finite sample size robustness measure. An increasingly popular and intuitively appealing concept, the finite sample size breakdown point, was introduced by Donoho (1982) and Donoho and Huber (1983). For the MVE-estimators and the S-estimators of multivariate location and scatter, both the asymptotic and the finite sample size breakdown points have been well studied [see Davies (1987), Lopuhaä (1989) and Lopuhaä and Rousseeuw (1991)]. However, for the projection based statistics only the asymptotic breakdown points have been previously studied, with the exception of Donoho's (1982) results on the finite sample breakdown point of the Stahel–Donoho statistics when the univariate location and scale statistic is taken to be, respectively, the median and the median absolute deviation about the median (M.A.D.).

In this paper, the finite sample breakdown points of the Stahel–Donoho statistics are obtained in general (Section 3.1), along with the finite sample breakdown points of the Maronna–Yohai scatter statistics and some extensions of their approach to multivariate location statistics (Section 3.2). The behavior of projection based statistics depends on the nature of the corresponding univariate location and scale statistics used on the projected data. Unlike the asymptotic breakdown points, the finite sample breakdown points of the multivariate statistics are not directly related to those of the corresponding univariate location and scale statistics. A uniform finite sample breakdown point concept for the location and scale statistics is needed [see (2.8)].

The median and M.A.D. are one possible choice for the univariate location and scale statistics. For sparse data sets in high dimensions, however, the M.A.D. is not recommended to be the univariate scale statistic used on the projected data since the resulting finite sample breakdown point of the corresponding multivariate location and scatter statistics is shown to be much less than optimum. A simple modification to the M.A.D., though, can be made to alleviate this problem [see (3.5)].

For various reasons, one may wish to consider univariate location and scale statistics other than the median and variants of the M.A.D., respectively. A very broad and natural class of univariate location and scale statistics to consider for the projected data is the simultaneous M-estimates of location and scale. A study of their breakdown properties is a major part of this paper (Section 4). Although the breakdown points of the M-estimates of location-only have been studied extensively, there is relatively little known for the breakdown points of the simultaneous M-estimates. This is partially due to the perception that scale is often just a nuisance parameter in the univariate setting [e.g., see Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 105]. However, here, the scale statistic is instrumental in defining projection based multivariate shape statistics, that is, functions of the scatter statistic which are invariant under scalar multiplication of the scatter statistic, and it is often the shape statistics which are of primary interest in multivariate analysis. There is also a perception that the simultaneous M-estimates of location and scale may be less robust than M-estimates of location-only in conjunction with an auxiliary scale statistic such as the M.A.D. In particular, Huber [(1981), Section 6.6] gives an implicit formula for the asymptotic breakdown points of simultaneous M-estimates for which the influence functions of both the location and scale components are monotone. His results imply that these statistics tend to have low breakdown points whenever the influence functions of the location and scale statistics are considerably smoother than the influence functions for the median and the M.A.D. As shown in Section 4, though, this does not necessarily occur for certain classes of nonmonotonic simultaneous M-estimates. Moreover, explicit rather than implicit expressions for the uniform finite sample breakdown points can be given for a large class of nonmonotonic M-estimates.

Section 2 establishes some concepts and notation. The proof for the main theorem of Section 4 is given in the Appendix.

2. Equivariance and breakdown. Let $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ represent a point cloud of n data points in $\mathbb{R}^p$. Location and scatter statistics for the point cloud $\mathbf{X}$ are denoted, respectively, $\mathbf{t}(\mathbf{X}) \in \mathbb{R}^p$ and $V(\mathbf{X}) \in$ PDS (p), where PDS (p) represents the set of all $p \times p$ positive definite symmetric matrices. The statistics $[\mathbf{t}(\mathbf{X}), V(\mathbf{X})]$ are said to be affine equivariant if, for any $\mathbf{b} \in \mathbb{R}^p$ and any $p \times p$ nonsingular matrix A,

$$\mathbf{t}(A\mathbf{X} + \mathbf{b}) = A\mathbf{t}(\mathbf{X}) + \mathbf{b} \tag{2.1}$$

and

$$V(A\mathbf{X} + \mathbf{b}) = AV(\mathbf{X})A', \tag{2.2}$$

where $A\mathbf{X} + \mathbf{b} = \{A\mathbf{x}_1 + \mathbf{b}, \dots, A\mathbf{x}_n + \mathbf{b}\}$.

Donoho and Huber (1983) discuss two types of finite sample breakdown points, replacement and contamination. The concept employed here is the finite sample replacement breakdown point. Suppose m arbitrary data points $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_m\}$ replace m data points from the original data $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, producing a corrupted sample $\mathbf{Z}$ which consists of a fraction of $\varepsilon_m = m/n$ bad values. For a given ε_m, a statistic is said to break down under ε_m-corruption if the difference between the statistic on the original sample $\mathbf{X}$ and on the corrupted sample $\mathbf{Z}$ can be made arbitrarily large in some sense for varying choices of $\mathbf{Y}$ and of the m replaced data points from $\mathbf{X}$. The finite sample breakdown point of the statistic at the sample $\mathbf{X}$ is defined to be $\varepsilon^*(\mathbf{X})$, the infimum of all ε_m causing breakdown.

More formally, let $[\mathbf{t}(\mathbf{X}), V(\mathbf{X})] \in \mathbb{R}^p \times \text{PDS}\,(p)$ and $[\mathbf{t}(\mathbf{Z}), V(\mathbf{Z})] \in \mathbb{R}^p \times \text{PDS}\,(p)$ represent location and scatter statistics for the original data $\mathbf{X}$ and the contaminated data $\mathbf{Z}$, respectively. Assuming $[\mathbf{t}(\mathbf{X}), V(\mathbf{X})]$ exists, define the maximum bias caused by ε_m-corruption to be

$$b(\varepsilon_m, \mathbf{X}; \mathbf{t}, V) = \sup_{\mathbf{Z}} \left[\max \left\{ \left\| V(\mathbf{X})^{-1/2} \{\mathbf{t}(\mathbf{X}) - \mathbf{t}(\mathbf{Z})\} \right\|, \operatorname{tr}\{V(\mathbf{X})V(\mathbf{Z})^{-1} + V(\mathbf{X})^{-1}V(\mathbf{Z})\} \right\} \right] \tag{2.3}$$

provided $[\mathbf{t}(\mathbf{Z}), V(\mathbf{Z})]$ exists for all possible $\mathbf{Z}$, and $b(\varepsilon_m, \mathbf{X}; \mathbf{t}, V) = \infty$ otherwise, where the norm $\|\mathbf{t}\| = (\mathbf{t}'\mathbf{t})^{1/2}$. Note that the maximum bias $b(\varepsilon_m, \mathbf{X}; \mathbf{t}, V)$ is invariant under affine transformations, that is,

$$b(\varepsilon_m, A\mathbf{X} + \mathbf{b}; \mathbf{t}, V) = b(\varepsilon_m, \mathbf{X}; \mathbf{t}, V), \tag{2.4}$$

for any nonsingular A and any $\mathbf{b} \in \mathbb{R}^p$, whenever $[\mathbf{t}(\cdot), V(\cdot)]$ is affine equivariant. Breakdown occurs under ε_m-corruption whenever $b(\varepsilon_m, \mathbf{X}; \mathbf{t}, V) = \infty$. This implies either $[\mathbf{t}(\mathbf{Z}), V(\mathbf{Z})]$ does not exist for some form of ε_m-corruption, $\|\mathbf{t}(\mathbf{Z})\|$ can be made arbitrarily large, the largest root of $V(\mathbf{Z})$ can be made arbitrarily large or the smallest root of $V(\mathbf{Z})$ can be made arbitrarily close to zero. The finite sample replacement breakdown point of $[\mathbf{t}(\mathbf{X}), V(\mathbf{X})]$ at $\mathbf{X}$ is then defined to be

$$\varepsilon^*(\mathbf{X}; \mathbf{t}, V) = \min_m \{\varepsilon_m = m/n \mid b(\varepsilon_m, \mathbf{X}; \mathbf{t}, V) = \infty\}. \tag{2.5}$$

The finite sample breakdown point can be dependent not only on the particular location and scatter statistics used but also on the nature of the good data $\mathbf{X}$. As with the S-estimators studied by Davies (1987), the breakdown points of the projection based statistics turn out to be the same for all point clouds $\mathbf{X}$ in *general position*. This means that the space generated by any $p + 1$ vectors in $\mathbf{X}$ equals $\mathbb{R}^p$. This occurs with probability 1 when $\mathbf{X}$ represents a realization from an absolutely continuous distribution in $\mathbb{R}^{np}$. For $\mathbf{X}$ in general position, Davies (1987) gives a strict upper bound for the finite sample replacement breakdown point for affine equivariant location and scatter statistics, namely,

$$\varepsilon^*(\mathbf{X}; \mathbf{t}, V) \leq \lfloor (n - p + 1)/2 \rfloor / n, \tag{2.6}$$

where $\lfloor k \rfloor$ represents the maximum of zero and the greatest integer less than or equal to k.

In this paper, general point clouds are considered. In practice, due to possible rounding or discreteness of the data, the good data may not be in general position. For general $\mathbf{X}$, the strict upper bound in (2.6) becomes

$$\varepsilon^*(\mathbf{X}; \mathbf{t}, V) \leq \left\lfloor \{n - c(\mathbf{X}) + 1\}/2 \right\rfloor / n, \tag{2.7}$$

where $c(\mathbf{X})$ represents the maximum number of data points in any $(p-1)$-dimensional hyperplane. When $\mathbf{X}$ is in general position, $c(\mathbf{X}) = p$ and so (2.6) and (2.7) coincide. The proof of (2.7) is identical to Davies' (1987) proof of (2.6), so it is omitted.

A univariate projection of the point cloud $\mathbf{X}$ can be represented by

$$\mathbf{a}'\mathbf{X} = \{\mathbf{a}'\mathbf{x}_1, \dots, \mathbf{a}'\mathbf{x}_n\}.$$

Let $[\mu(\cdot), \sigma(\cdot)]$ represent translation and scale equivariant univariate location and scale statistics, that is, $[\mu(\cdot), \sigma^2(\cdot)]$ are affine equivariant. For example, $\mu(\mathbf{a}'\mathbf{X})$ may be the sample median of $\mathbf{a}'\mathbf{X}$ and $\sigma(\mathbf{a}'\mathbf{X})$ may be the sample M.A.D. of $\mathbf{a}'\mathbf{X}$. The definition of the finite sample replacement breakdown point $\varepsilon^*(\mathbf{x}; \mu, \sigma)$ of $[\mu(\cdot), \sigma^2(\cdot)]$ at a univariate data set $\mathbf{x} = \{x_1, \dots, x_n\}$ is simply the univariate version of (2.5). A stronger concept, however, is needed in later sections when considering all univariate projections of the p-dimensional data cloud $\mathbf{X}$. For $[\mu(\cdot), \sigma(\cdot)]$, define the uniform finite sample replacement breakdown point at $\mathbf{X}$ to be

$$\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = \min_m \left\{ \varepsilon_m = \frac{m}{n} \;\middle|\; \sup_{\mathbf{a}} b(\varepsilon_m, \mathbf{a}'\mathbf{X}; \mu, \sigma^2) = \infty \right\}, \tag{2.8}$$

where the supremum is taken over $\mathbf{a} \in \mathbb{R}^p \backslash \{\mathbf{0}\}$. The definition of the maximum bias function $b(\varepsilon_m, \mathbf{x}; \mu, \sigma^2)$ is the univariate version of (2.3), applied to $[\mu(\cdot), \sigma^2(\cdot)]$. By the invariance property (2.4), it is sufficient to let the supremum in (2.8) be taken over $\mathbf{a} \in \mathcal{S}_{p-1} = \{\mathbf{a} \in \mathbb{R}^p \mid \mathbf{a}'\mathbf{a} = 1\}$, the p-dimensional unit sphere. If $\mathbf{x}$ is a univariate data set, then $\varepsilon^{**}(\mathbf{x}; \mu, \sigma)$ and $\varepsilon^*(\mathbf{x}; \mu, \sigma)$ are equivalent.

A relationship between the uniform breakdown point and the breakdown points themselves for the univariate location and scale statistics can be obtained immediately from (2.8), namely,

$$\varepsilon^{**}(\mathbf{X}; \mu, \sigma) \le \inf_{\mathbf{a}} \varepsilon^*(\mathbf{a}'\mathbf{X}; \mu, \sigma), \tag{2.9}$$

where the infimum is over $\mathbf{a} \in \mathbb{R}^p \backslash \{\mathbf{0}\}$ or, equivalently, over $\mathbf{a} \in \mathcal{S}_{p-1}$. Equality can be shown to hold in (2.9) if for any data set $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ the statistics $\mu(\mathbf{a}'\mathbf{Z})$ and $\sigma(\mathbf{a}'\mathbf{Z})$ are continuous functions of $\mathbf{a}$. Otherwise, equality in (2.9) does not necessarily hold. A counterexample to equality in (2.9) is given at the end of Section 3.1.

The same strict upper bound given by (2.7) also holds for $\varepsilon^{**}(\mathbf{X}; \mu, \sigma)$, that is,

$$\varepsilon^{**}(\mathbf{X}; \mu, \sigma) \le \left\lfloor \{n - c(\mathbf{X}) + 1\}/2 \right\rfloor \Big/ n. \tag{2.10}$$

Although this upper bound can be shown directly, it is viewed here as a corollary to (2.7) and Theorem 3.2 of the next section.

3. Projection based statistics.

3.1. *The Stahel–Donoho statistics.* The Stahel-Donoho class of multivariate location and scatter statistics for the data cloud $\mathbf{X}$ are defined as follows. Given translation and scale equivariant univariate location and scale statistics $[\mu(\cdot), \sigma(\cdot)]$, define, for any $\mathbf{v} \in \mathbb{R}^p$,

$$r(\mathbf{v}, \mathbf{X}) = \sup_{\mathbf{a}} \left\{ \frac{|\mathbf{a}'\mathbf{v} - \mu(\mathbf{a}'\mathbf{X})|}{\sigma(\mathbf{a}'\mathbf{X})} \right\}, \tag{3.1}$$

where the supremum is over $\mathbf{a} \in \mathbb{R}^p$ or, equivalently, over $\mathbf{a} \in \mathcal{S}_{p-1}$. The value of $r(\mathbf{v}, \mathbf{X})$ is affine invariant, that is, $r(\mathbf{v}, \mathbf{X}) = r(A\mathbf{v} + \mathbf{b}, A\mathbf{X} + \mathbf{b})$ for any nonsingular A and any $\mathbf{b} \in \mathbb{R}^p$, and represents a measure of how far the point $\mathbf{v}$ lies from the data cloud $\mathbf{X}$. The Stahel–Donoho statistics are then obtained by taking a weighted mean vector and covariance matrix which downweights outlying data points. Specifically, define

$$\mathbf{t}(\mathbf{X}) = \frac{\sum_{i=1}^n w_i \mathbf{x}_i}{\sum_{i=1}^n w_i} \tag{3.2}$$

and

$$V(\mathbf{X}) = \frac{\sum_{i=1}^n w_i \{\mathbf{x}_i - \mathbf{t}(\mathbf{X})\}\{\mathbf{x}_i - \mathbf{t}(\mathbf{X})\}'}{\sum_{i=1}^n w_i}, \tag{3.3}$$

where $w_i = w[r(\mathbf{x}_i; \mathbf{X})]$ for $i = 1, \dots, n$ with w: $\mathbb{R} \to \mathbb{R}^+$ being bounded and continuous, and $r^2 w(r)$ being bounded. Here, $\mathbb{R}^+$ is the set $\{w > 0\}$.

As noted in the introduction, the Stahel–Donoho statistics were the first affine equivariant statistics to be shown to have breakdown point near $\frac{1}{2}$ regardless of the dimension p. Stahel (1981) shows that the statistics (3.2) and (3.3) have asymptotic breakdown point $\frac{1}{2}$ at continuous multivariate models whenever the corresponding univariate location and scale statistics have asymptotic breakdown point $\frac{1}{2}$. For a more readily accessible treatment of Stahel's proof, see Hampel, Ronchetti, Rousseeuw and Stahel [(1986), Theorem 5.5.3]. When $\mathbf{X}$ is in general position and the corresponding univariate location and scale statistics are the sample median and the M.A.D., respectively, Donoho (1982) derives the finite sample breakdown point of the statistics (3.2) and (3.3). His results are in terms of the finite sample contamination breakdown point, which when interpreted in terms of the finite sample replacement breakdown point [e.g., see Davies (1987), page 1288] is

$$\varepsilon^*(\mathbf{X}; \mathbf{t}, V) = \lfloor (n - 2p + 2)/2 \rfloor / n. \tag{3.4}$$

Expression (3.4) does not correspond to the finite sample replacement breakdown point of the median and M.A.D., which is $\lfloor n/2 \rfloor / n$ whenever the "good" univariate data has no duplications. From Donoho's (1982) derivation of (3.4), it can be verified that (3.4) corresponds to the uniform finite sample replacement breakdown point of the median and the M.A.D. The following theorem establishes a general relationship.

THEOREM 3.1. *For the statistics defined by* (3.2) *and* (3.3), $\varepsilon^*(\mathbf{X}; \mathbf{t}, V) \geq \varepsilon^{**}(\mathbf{X}; \mu, \sigma)$ *with equality holding whenever* $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = \lfloor\{n - c(\mathbf{X}) + 1\}/2\rfloor/n$.

PROOF. Let $\varepsilon_m < \varepsilon^{**}(\mathbf{X}; \mu, \sigma)$ and so there exists a $b_0 < \infty$ such that $b(\varepsilon_m, \mathbf{a}'\mathbf{X}; \mu, \sigma) < b_0$ for all $\mathbf{a} \in \mathcal{S}_{p-1}$. This implies the existence of scalars μ_0, σ_0 and σ_1 such that $|\mu(\mathbf{a}'\mathbf{Z})| < \mu_0 < \infty$ and $0 < \sigma_0 < \sigma(\mathbf{a}'\mathbf{Z}) < \sigma_1 < \infty$ for all $\mathbf{a} \in \mathcal{S}_{p-1}$ and all ε_m-corrupted data sets $\mathbf{Z}$. For any $\mathbf{x}_i \in \mathbf{X}$, $r(\mathbf{x}_i; \mathbf{Z})$ must therefore be uniformly bounded above over all possible $\mathbf{Z}$. By the conditions on w, the denominator in (3.2) and (3.3) is uniformly bounded away from zero over all possible $\mathbf{Z}$.

Next, consider the numerators in (3.2) and (3.3). For any $\mathbf{a} \in \mathcal{S}_{p-1}$,

$$|\mathbf{a}'\mathbf{z}| \leq \sigma(\mathbf{a}'\mathbf{Z})\left|\frac{\{\mathbf{a}'\mathbf{z} - \mu(\mathbf{a}'\mathbf{Z})\}}{\sigma(\mathbf{a}'\mathbf{Z})}\right| + |\mu(\mathbf{a}'\mathbf{Z})| \leq \sigma_1 r(\mathbf{z}; \mathbf{Z}) + \mu_0.$$

This implies $w[r(\mathbf{z}_i; \mathbf{Z})]|\mathbf{a}'\mathbf{z}_i| \leq \sigma_1 w_i r(\mathbf{z}_i; \mathbf{Z}) + \mu_0 w_i$, which by the conditions on w implies the numerator in (3.2) is uniformly bounded above over all possible $\mathbf{Z}$. Similarly, the numerator in (3.3) is uniformly bounded above over all possible $\mathbf{Z}$ since $r^2 w(r)$ is bounded above.

The proof is completed by showing that, for any $\mathbf{a} \in \mathcal{S}_{p-1}$, $\mathbf{a}'V(\mathbf{Z})\mathbf{a}$ is uniformly bounded away from zero for all possible $\mathbf{Z}$. This holds, since otherwise there must exist $(n - m)$ vectors from $\mathbf{X}$ (say, $\mathbf{X}_0$) such that, for some $\mathbf{a} \in \mathcal{S}_{p-1}$, $\mathbf{t} \in \mathbb{R}^p$ and all $\mathbf{x} \in \mathbf{X}_0$, $\mathbf{a}'(\mathbf{x} - \mathbf{t}) = 0$. This implies $(n - m) \leq c(\mathbf{X})$ and so $\varepsilon_m \geq 1 - c(\mathbf{X})/n$, which contradicts (2.9). The statement concerning equality follows from (2.7). □

The proof of Theorem 3.1 is similar to Stahel's proof for the asymptotic breakdown point. A brief technical remark concerning $\varepsilon^{**}(\mathbf{X}; \mu, \sigma)$ may be instructive here. As noted previously, $\varepsilon^{**}(\mathbf{X}; \mu, \sigma)$ is not equivalent to the finite sample replacement breakdown point of the location and scale statistics. This is in contrast to the asymptotic breakdown points considered by Stahel (1981), which are defined on location and scale functionals and employ the Prohorov distance between two distributions in measuring bias. At continuous multivariate models, a breakdown point of ε^* for the univariate location and scale functional implies the maximum bias of the location and scale functionals is uniformly bounded above over all marginal univariate distributions when the proportion of contamination is less than ε^*. Empirical distributions though are not continuous, and replacement neighborhoods are not equivalent to Prohorov neighborhoods. For discrete distributions, problems arise mostly because different point masses will coincide for certain projections. The introduction of $\varepsilon^{**}(\mathbf{X}; \mu, \sigma)$ is needed to account for this effect.

It is not clear whether equality holds in general in Theorem 3.1 without further conditions on the univariate location and scale statistics. When using the sample median and the M.A.D. in (3.2) and (3.3), Donoho (1982) proves only that (3.4) is a lower bound for $\varepsilon^*(\mathbf{X}; \mathbf{t}, V)$, but notes that it can also be shown to be an upper bound. A readily accessible treatment of Donoho's derivation of the lower bound is given in Huber (1985). Equality in (3.4) follows by noting that if

$m = \lfloor (n - 2p + 2)/2 \rfloor$ replacements lie in the same plane as p of the remaining data points, then the M.A.D. $= 0$ for the univariate projection orthogonal to that plane, and so (3.2) and (3.3) are not defined.

A sample of size $n \geq 10p - 5$, when n is odd, and of size $n \geq 10(p - 1)$, when n is even, is needed for (3.4) to be as large as 0.4, whereas a sample of size $n \geq 5p$, when $(n - p)$ is even, and of size $n \geq 5(p - 1)$, when n is odd, is needed for the strict upper bound (2.6) to be as large as 0.4. Breaking the previous statement into different cases is needed since neither (2.6) nor (3.4) is monotonic in n, although both are monotonic for even n and for odd n. For data clouds as sparse as $n = 2p$, (2.6) goes to $\frac{1}{4}$ whereas (3.4) goes to zero as $p \to \infty$. The drawback in using the M.A.D. in sparse data sets is due to the proportion of inliers arising from certain projections of the data. A slight modification to the M.A.D. can be made to alleviate this problem. Replace the M.A.D. by the $\{k + (n + 1)/2\}$th smallest absolute deviation about the median when n is odd, and by the average of the $(k + n/2)$th and $(k + 1 + n/2)$th smallest absolute deviation about the median when n is even. Together with the median, for $\mathbf{X}$ in general position,

$$(3.5)\quad \varepsilon^{**}(\mathbf{X}; \mu, \sigma) = \lfloor (n - 2p + k + 2)/2 \rfloor / n \quad \text{and} \quad \varepsilon^{*}(\mathbf{X}; \mathbf{t}, V) = \varepsilon^{**}(\mathbf{X}; \mu, \sigma),$$

for $k = 0, 1, \dots, p - 1$. The proof of (3.5) is analogous to that of (3.4). If $k = p - 1$, then the corresponding Stahel–Donoho statistics achieve the maximal possible breakdown point (2.6). Other location and scale statistics are considered in Section 4.

To construct a simple counterexample to equality in (2.9), suppose $\mathbf{X}$ is in general position and n is odd. Let $\mu(\cdot)$ be the sample median and define the scale statistic by choosing $\sigma(\cdot)$ to be the M.A.D. if no data points are duplicated and to be the $\{p - 1 + (n + 1)/2\}$th smallest absolute deviation about the median if the data set has duplicate points. For this example, it can be shown that $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = \lfloor (n - 2p + 2)/2 \rfloor / n$, that is, the same as if using the median and the M.A.D., whereas $\sup_{\mathbf{a}} \varepsilon^{*}(\mathbf{a}'\mathbf{X}; \mu, \sigma) = \lfloor (n - p + 1)/2 \rfloor / n$.

3.2. *The Maronna–Yohai statistics.* Maronna and Yohai (1989) recently introduced the following class of scatter statistics. Given a location and scale equivariant scale statistic $\sigma(\cdot)$, define

$$(3.6)\qquad V(\mathbf{X}) = \arg\inf \left[\sup_{\mathbf{a}} \left\{ \left| 1 - \frac{\sigma(\mathbf{a}'\mathbf{X})}{(\mathbf{a}'V\mathbf{a})^{1/2}} \right| \right\} \right],$$

where the infimum is over $V \in \mathrm{PDS}(p)$ and the supremum is over $\mathbf{a} \in \mathbb{R}^p$ or, equivalently, $\mathbf{a} \in \mathcal{S}_{p-1}$. Note that if $\sigma(\cdot)$ is taken to be the sample standard deviation, then $V(\mathbf{X})$ is the sample covariance matrix and $\mathbf{a}'V(\mathbf{X})\mathbf{a} = \sigma^2(\mathbf{a}'\mathbf{X})$ for all $a \in \mathbb{R}^p$. For general $\sigma(\cdot)$, as a function of $\mathbf{a}$ the quadratic form $\mathbf{a}'V(\mathbf{X})\mathbf{a}$ represents an approximation of $\sigma^2(\mathbf{a}'\mathbf{X})$.

For the scatter statistic (3.6), Maronna and Yohai (1989) show that its asymptotic breakdown point is at least ε^* whenever the corresponding asymptotic breakdown point of the univariate scale statistic is ε^*. The asymptotic breakdown points here refer to the breakdown points of the scatter and scale functionals over ε-contaminated neighborhoods of the model distribution rather than Prohorov neighborhoods. The results of Maronna and Yohai (1989) are obtained at elliptically symmetric continuous models, of which the multivariate normal distribution is one. Their results on the asymptotic breakdown points have been extended to general distributions by Maronna, Stahel and Yohai (1992).

Obtaining straightforward results on the finite sample breakdown behavior of the Maronna–Yohai statistics by using (3.6) directly is somewhat problematic. The problem lies in the asymmetric manner in which departures of $\sigma^2(\mathbf{a}'\mathbf{X})$ from $\mathbf{a}'V\mathbf{a}$ are treated. For a sequence $V_k \in \mathrm{PDS}(p)$, if $\mathbf{a}'V_k\mathbf{a} \to 0$, then $|1 - \sigma(\mathbf{a}'\mathbf{X})/(\mathbf{a}_k'V\mathbf{a})^{1/2}| \to \infty$ when $\sigma(\mathbf{a}'\mathbf{X}) > 0$, whereas if $\mathbf{a}'V_k\mathbf{a} \to \infty$, then $|1 - \sigma(\mathbf{a}'\mathbf{X})/(\mathbf{a}'V_k\mathbf{a})^{1/2}| \to 1$. The following modification to (3.6) is easier to analyze. For a given location and scale statistic $\sigma(\cdot)$, define

$$V(\mathbf{X}) = \arg\inf\left[\sup_{\mathbf{a}} d\left\{\frac{\sigma^2(\mathbf{a}'\mathbf{X})}{\mathbf{a}'V\mathbf{a}}\right\}\right], \tag{3.7}$$

where $d\colon \mathbb{R}^+ \to \mathbb{R}$ is nonnegative, continuous with $d(0^+) = \infty$ and $d(\infty) = \infty$. For example, $d(\cdot)$ could be taken as $|\log(\cdot)|$. The infimum is again over $V \in \mathrm{PDS}(p)$. Curiously, it can be shown that the solutions to (3.6) and (3.7) differ only up to a scalar multiple, that is, $V(\mathbf{X}) = V$ solves (3.7) if and only if $V(\mathbf{X}) = cV$ solves (3.6) for some scalar $c > 0$, if for some $x_0 > 0$ the function $d(x)$ decreases for $x < x_0$ and increases for $x > x_0$. This relationship was established by Maronna (1991).

The basic concept behind the Maronna–Yohai approach can be extended to multivariate location. Given translation and scale equivariant statistics $[\mu(\cdot), \sigma(\cdot)]$, define

$$\mathbf{t}(\mathbf{X}) = \arg\inf\{r(\mathbf{t}; \mathbf{X})\}, \tag{3.8}$$

where $r(\mathbf{t}; \mathbf{X})$ is defined in (3.1) and the infimum is over $\mathbf{t} \in \mathbb{R}^p$. If the univariate location statistic is taken to be the sample mean, then $\mathbf{t}(\mathbf{X})$ is the sample mean vector and $\mathbf{a}'\mathbf{t}(\mathbf{X}) = \mu(\mathbf{a}'\mathbf{X})$. For general $\mu(\cdot)$, the linear form $\mathbf{a}'\mathbf{t}(\mathbf{X})$, as a function of $\mathbf{a}$, can be viewed as an approximation of $\mu(\mathbf{a}'\mathbf{X})$. The uniform finite sample breakdown points of the corresponding univariate location and scale statistics again serve as a lower bound for the finite sample breakdown points of (3.7) and (3.8).

THEOREM 3.2. *For the statistics defined by* (3.7) *and* (3.8), $\varepsilon^*(\mathbf{X}; \mathbf{t}, V) \geq \varepsilon^{**}(\mathbf{X}; \mu, \sigma)$, *with equality whenever* $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = \lfloor\{n - c(\mathbf{X}) + 1\}/2\rfloor/n$.

PROOF. Let $\varepsilon_m < \varepsilon^{**}(\mathbf{X}; \mu, \sigma)$. As in the proof of Theorem 3.1, $|\mu(\mathbf{a}'\mathbf{Z})| < \mu_0 < \infty$ and $0 < \sigma_0 < \sigma(\mathbf{a}'\mathbf{Z}) < \sigma_1 < \infty$ for all $\mathbf{a} \in \mathcal{S}_{p-1}$ and all ε_m-corrupted data sets $\mathbf{Z}$. This implies that, for any fixed $\mathbf{t} \in \mathbb{R}^p$, $r(\mathbf{t}; \mathbf{Z})$ is uniformly bounded above over all possible $\mathbf{Z}$. However, the equivariance and robustness properties of $\mu(\cdot)$

and $\sigma(\cdot)$ imply that $\inf_{\mathbf{Z}} r(\mathbf{t};\mathbf{Z}) \to \infty$ as $\|\mathbf{t}\| \to \infty$, and so $\|\mathbf{t}(\mathbf{Z})\|$ is uniformly bounded above over all possible $\mathbf{Z}$.

Consider now the scatter component $V(\mathbf{Z})$. For any fixed $V \in \mathrm{PDS}(p)$ and $\mathbf{a} \in \mathcal{S}_{p-1}$,

$$\frac{\sigma_0^2}{\mathbf{a}'V\mathbf{a}} \le \frac{\sigma^2(\mathbf{a}'\mathbf{Z})}{\mathbf{a}'V\mathbf{a}} \le \frac{\sigma_1^2}{\mathbf{a}'V\mathbf{a}},$$

for all possible $\mathbf{Z}$. By the definition of $d(\cdot)$, this implies for any fixed $V \in \mathrm{PDS}(p)$ that $\sup_{\mathbf{a}} d\{\sigma^2(\mathbf{a}'\mathbf{Z})/\mathbf{a}'V\mathbf{a}\}$ is uniformly bounded above over all possible $\mathbf{Z}$ since the supremum can be restricted to $\mathbf{a} \in \mathcal{S}_{p-1}$. Now by definition of $V(\mathbf{Z})$, for any fixed $V \in \mathrm{PDS}(p)$,

$$\sup_{\mathbf{a}} d\left\{\frac{\sigma^2(\mathbf{a}'\mathbf{Z})}{\mathbf{a}'V(\mathbf{Z})\mathbf{a}}\right\} < \sup_{\mathbf{a}} d\left\{\frac{\sigma^2(\mathbf{a}'\mathbf{Z})}{\mathbf{a}'V\mathbf{a}}\right\},$$

and so the left-hand side is uniformly bounded over all possible $\mathbf{Z}$. However, $\sigma^2(\mathbf{a}'\mathbf{Z})$ is uniformly bounded away from zero and uniformly bounded above over all $\mathbf{a} \in \mathcal{S}_{p-1}$ and over all possible $\mathbf{Z}$ and so, by the definition of $d(\cdot)$, the same statement must also hold for $\mathbf{a}'V(\mathbf{Z})\mathbf{a}$. □

If one uses the median as the univariate location statistic and a variant of the M.A.D. discussed prior to (3.5) as the univariate scale statistics, then the uniform finite sample breakdown point of these statistics is given by the first part of (3.5) when $\mathbf{X}$ is in general position. The second part of (3.5) also holds for the Maronna–Yohai statistics, that is, $\varepsilon^*(\mathbf{X};\mathbf{t},V) = \varepsilon^{**}(\mathbf{X};\mu,\sigma)$. This again follows by noting that if $m = \lfloor (n-2p+k+2)/2 \rfloor$ and the replacements all lie in the same plane as p of the remaining data points, then the modified M.A.D. equals 0 for the univariate projection orthogonal to that plane, and so (3.7) and (3.8) are not defined.

An interesting observation arises when one considers the special case where $\mu(\cdot)$ is the sample median and $\sigma(\cdot)$ is the sample M.A.D. For this case, a heuristic argument suggests that under random sampling from continuous spherically symmetric models centered at $\mathbf{0}$, the asymptotic distribution of $\mathbf{t}(\mathbf{X})$ is

$$n^{1/2}\mathbf{t}(X) \to_d \arg\inf \sup_{\mathbf{a}}\{|\mathbf{a}'\mathbf{t} - \mathbf{G}(\mathbf{a})|\}, \tag{3.9}$$

where the supremum is over $\mathbf{a} \in \mathcal{S}_{p-1}$ and the infimum is over $\mathbf{t} \in \mathbb{R}^p$. The process $\mathbf{G}(\mathbf{a})$ corresponds to the limiting process of $n^{1/2}$ median$(\mathbf{a}'\mathbf{X})$, which is a Gaussian process on the unit sphere with zero mean and covariance structure $\mathrm{cov}\{\mathbf{G}(\mathbf{a}_1),\mathbf{G}(\mathbf{a}_2)\} = \{\frac{1}{4} - (2\pi)^{-1}\arccos(\mathbf{a}_1'\mathbf{a}_2)\}/f_0^2$, with f_0 being the value of any univariate marginal density at the origin. The limiting distribution (3.9) is the same as the limiting distribution of another multivariate location statistic proposed by Donoho (1982), namely, Donoho's deepest depth statistic. The finite sample breakdown point of Donoho's (1982) statistic depends on the structure of $\mathbf{X}$ even when it is in general position and is at best $\frac{1}{3}$ in large samples, whereas for (3.8) the breakdown point goes to $\frac{1}{2}$ for large samples. The asymptotic distribution of Donoho's deepest depth statistic has been obtained by Nolan (1992) using the theory of empirical processes. A rigorous justification of (3.9) is not given here.

4. M-estimators of location and scale. A natural class of univariate location and scale statistics which can be used in constructing projection based multivariate location and scatter statistics are the simultaneous M-estimates of location and scale. This section considers the finite sample breakdown problem for these M-estimates. The main theorem of the section, Theorem 4.1, gives an explicit formula for the uniform finite sample breakdown points for two broad classes of simultaneous M-estimates of location and scale, namely, the MLE-type M-estimates and the S-type M-estimates, or S-estimates. Theorem 4.1 is novel even when applied to the special case $p = 1$. For this case, it yields explicit formulas for the finite sample breakdown points of the MLE-type and S-type M-estimates of location and scale.

For a univariate data set $\mathbf{x} = \{x_i;\ i = 1, \dots, n\}$ the simultaneous M-estimates are usually defined as solutions (μ, σ) of a system of simultaneous equations of the form

$$\text{ave}\{\psi(s_i)\} = 0, \tag{4.1}$$

$$\text{ave}\{\chi(s_i)\} = 0, \tag{4.2}$$

where $s_i = (x_i - \mu)/\sigma$ for $i = 1, \dots, n$ and ψ and χ are real-valued functions with ψ commonly taken to be odd and χ commonly taken to be even. The notation "ave" refers to the average over $i = 1, \dots, n$. By choosing $\psi(s) = \text{sign}(s)$ and $\chi(s) = \text{sign}(|s| - 1)$, the M-estimates correspond to the median and the M.A.D.

For various reasons, one may wish to consider location and scale statistics other than the median and variants of the M.A.D., respectively. Whenever both ψ and χ are monotone, though, Huber's (1981) implicit formula for the asymptotic breakdown point of such M-estimates implies that smoother choices of ψ and χ tend to produce M-estimates with considerably lower breakdown points. This problem, however, does not necessarily arise for MLE-type M-estimates and for the S-estimates whenever the corresponding ψ function is not monotone.

The MLE-type M-estimates are defined as a solution $(\widehat{\mu}, \widehat{\sigma}) \in \mathbb{R} \times \mathbb{R}^+$ which minimizes

$$\ell(\mu, \sigma; \mathbf{x}) = \text{ave}\left\{\rho\left(s_i^2\right)\right\} + \log \sigma \tag{4.3}$$

over all $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$, where ρ is some real-valued continuous function and as before $s_i = (x_i - \mu)/\sigma$. Note that, when $\exp\{-\rho(s^2)\}$ is integrable over $\mathbb{R}$, $\ell(\mu, \sigma; \mathbf{x})$ can be regarded as a negative log-likelihood function of a univariate symmetric location–scale model and so the MLE-type M-estimates include the maximum likelihood estimates of such models. If ρ is differentiable, then any critical point of (4.3) satisfies (4.1)–(4.2) with $\psi_{\rho, M}(s) = s\rho'(s^2)$ and $\chi_\rho(s) = 2s^2\rho'(s^2) - 1$.

The S-type M-estimates, or S-estimates, are defined as follows: for $\mu \in \mathbb{R}$, consider the scale $\widehat{\sigma}(\mu)$ which minimizes (4.3), that is, define $\widehat{\sigma}(\mu) \in \mathbb{R}^+$ such that

$$\ell\left(\mu, \widehat{\sigma}(\mu); \mathbf{x}\right) \le \ell(\mu, \sigma; \mathbf{x}), \tag{4.4}$$

for all $\sigma \in \mathbb{R}^+$. An intuitively appealing estimate of location and scale is to choose μ so that the scale about μ, that is, $\widehat{\sigma}(\mu)$, is minimized. Accordingly, for a given ρ function, the S-estimates of location and scale are taken to be $(\widetilde{\mu}, \widetilde{\sigma}) \in \mathbb{R} \times \mathbb{R}^+$ such that

$$\widetilde{\sigma} = \widehat{\sigma}(\widetilde{\mu}) \leq \widehat{\sigma}(\mu) \tag{4.5}$$

over $\mu \in \mathbb{R}$. If ρ is twice differentiable, then the S-type M-estimates satisfy (4.1)–(4.2) with $\chi_\rho(s) = 2s^2\rho'(s^2) - 1$ and $\psi_{\rho, S}(s) = \chi_\rho'(s)$.

Whenever ψ or χ is not monotone, (4.1)–(4.2) may admit multiple solutions. For an MLE-type or S-type M-estimate, not every solution to the corresponding equations (4.1)–(4.2) represents the MLE-type or S-type M-estimate, only those which satisfy the full definition do. The need to define the right solution rather than considering any solution to (4.1)–(4.2) can be seen from the simpler location-only case. If ψ is strongly redescending, that is, $\psi(s) = 0$ for large s, then the latter definition would yield a statistic with breakdown point zero since there are always arbitrarily large solutions to the equation. Such tail solutions, though, are commonly viewed as wrong solutions. In studying redescending M-estimates of location-only, Huber (1984) defines the M-estimate of location to be a solution μ which minimizes $\text{ave}\{\rho_0(x - \mu)\}$ rather than a solution to $\text{ave}\{\psi_0(x - \mu)\} = 0$, where $\psi_0(s) = \rho_0'(s)$.

Throughout, the function $\rho(s)$ is taken to be continuous for $s \geq 0$, but not necessarily differentiable. Thus the definition of the MLE-type and S-type M-estimates includes cases which cannot be viewed as solutions to M-estimating equations of the form (4.1)–(4.2). For the S-estimates of location and scale, finite sample breakdown points have been previously obtained within the context of the more general regression model by Rousseeuw and Yohai (1984) and in the more general multivariate model by Davies (1987), Lopuhaä (1989) and Lopuhaä and Rousseeuw (1991). All previous results assume ρ to be differentiable and the corresponding χ function to be nondecreasing on $\mathbb{R}^+$, left-continuous and constant for large s. None of these additional conditions are needed here. The definition of the S-estimates given by (4.4) and (4.5) is thus more general than earlier definitions of S-estimates of location and scale. The generality of the conditions on ρ considered here helps to emphasize the crucial role of the tail of ρ on the breakdown point.

For a given ρ function, the breakdown points of the MLE-type and S-type M-estimates are the same and depend on ρ only through its tail behavior, and in particular its "sill" a_ρ, $-\infty \leq a_\rho \leq \infty$, which is defined by

$$a_\rho = \sup\Big\{a \in \mathbb{R} \mid t^{a/2} \exp\left[-\rho(t)\right] \to 0 \text{ as } t \to \infty\Big\}. \tag{4.6}$$

If $\rho(t)$ is differentiable for large t and $\chi_\rho(s)$ converges to a limit as $s \to \infty$, then the sill can be characterized by

$$\lim_{s \to \infty} \chi_\rho(s) = a_\rho - 1; \tag{4.7}$$

see Kent and Tyler [(1991), (2.3)].

Besides continuity, the following mild tail conditions on ρ are needed in deriving the breakdown points.

CONDITION T. Let $\rho(t)$ be continuous for $t \geq 0$ with sill a_ρ, and define, for $a \in \mathbb{R}$, $h_a(t) = t^{a/2}\exp\{-\rho(t)\}$.

(i) For $a < a_\rho$ and for some r_0, $h_a(t)/h_a(r)$ is bounded above over all $t > r > r_0$.

(ii) For $a > a_\rho$ and for some r_0, $h_a(t)/h_a(r)$ is bounded above for all $t \leq r$, $r > r_0$.

Note that from the definition of the sill, $h_a(t) \to 0$ as $r \to \infty$ for $a < a_\rho$. On the other hand, the definition of the sill and Condition T(ii) together imply $h_a(r) \to \infty$ as $r \to \infty$ for $a > a_\rho$. This second statement follows since if $a > a_\rho$, then from the definition of the sill there must exist a sequence $t_k \to \infty$ such that $h_a(t_k) \to \infty$. However, if $h_a(r) \not\to \infty$ as $r \to \infty$, then there exists a subsequence $r_k \to \infty$ with $r_k > t_k$ such that $h_a(r_k)$ is bounded and hence $h_a(t_k)/h_a(r_k) \to \infty$. This contradicts Condition T(ii).

In general, Condition T simply places a restriction on the variability of $\rho(t)$ in the tail. This is particularly clear when $\rho(t)$ is differentiable, in which case the following lemma provides a simple verification of Condition T in terms of the tail of $\chi_\rho(s)$.

LEMMA 4.1. *If $\rho(t)$ is continuous for $t \geq 0$, differentiable for large t, and $\chi_\rho(s)$ has a limit as $s \to \infty$, then Condition* T *holds.*

PROOF. By (4.7), $\chi_\rho(s) = a_\rho - 1$ as $s \to \infty$. The derivative of $\log h_a(t)$ is $\{\frac{1}{2}a - 1 - \chi_\rho(t^{1/2})\}/t$, which for large t is negative for $a < a_\rho$ and positive for $a > a_\rho$. This implies for large r that $h_a(r)$ is decreasing for $a < a_\rho$ and increasing for $a > a_\rho$. Condition T(i) is immediate since the ratio is bounded above by 1. Condition T(ii) also follows since the ratio is again bounded above by 1 for large t. Furthermore, if t is bounded above, then $h_a(t)$ is bounded above, whereas $h_a(r)$ is bounded away from zero for larger r. □

The main theorem of this section can now be stated.

THEOREM 4.1. *Suppose Condition* T *holds with $a_\rho > n/(n - c(\mathbf{X}))$, and let $\delta^*(\mathbf{X}) = \min[1 - 1/a_\rho - c(\mathbf{X})/n, 1/a_\rho]$. For both the MLE-type M-estimates $[\hat{\mu}(\cdot), \hat{\sigma}(\cdot)]$ and the S-estimates $[\tilde{\mu}(\cdot), \tilde{\sigma}(\cdot)]$, the following holds*:

(a) *If $\delta^*(\mathbf{X}) = m/n$ for some integer m, then either $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = m/n$ or $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = (m + 1)/n$.*

(b) *If $\delta^*(\mathbf{X})$ is not of the form m/n for some integer m, then $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = m^*/n$, where $m^* = \lfloor(n\delta^*(\mathbf{X}) + 1)\rfloor$.*

The proof of Theorem 4.1 is given in the Appendix. From the proof it can be noted that the term $(1 - 1/a_\rho - c(\mathbf{X})/n)$ in $\delta^*(\mathbf{X})$ is due to inlier contamination, that is, by letting all the contaminating data points lie within a hyperplane containing $c(\mathbf{X})$ of the good data. The second term $1/a_\rho$ in $\delta^*(\mathbf{X})$ is due to outlier contamination, that is, by letting all the contaminating data points go to infinity in one direction. At one extreme, if $a_\rho = \infty$, then $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = 1/n$ and breakdown can be caused by one outlier. At the other extreme, if $a_\rho < n/(n - c(\mathbf{X}))$, then the location and scale statistics do not exist for some projection of the good data itself.

If $\mathbf{X}$ is in general position, that is, $c(\mathbf{X}) = p$, then the value of $\delta^*(\mathbf{X})$ is maximized when $a_\rho = 2n/(n - p)$, giving $\delta^*(\mathbf{X}) = (n - p)/(2n)$. If the first part of Theorem 4.1 applies whenever $a_\rho = 2n/(n - p)$, then $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = (n - p)/(2n)$ since the other possibility $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = (n - p + 2)/(2n)$ exceeds the universal upper bound (2.10). If the second part of Theorem 4.1 applies, then $\varepsilon^{**}(\mathbf{X}; \mu, \sigma)$ also achieves the upper bound (2.10) whenever $a_\rho = 2n/(n - p)$. This is summarized in the following corollary.

COROLLARY 4.1. *If* $\mathbf{X}$ *is in general position*, $n > p$, *and Condition* T *holds with* $a_\rho = 2n/(n - p)$, *then*, *for both the MLE-type M-estimates and the S-estimates*, $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = \lfloor (n - p + 1)/2 \rfloor / n$.

There do exist MLE-type and S-type M-estimates which correspond to solutions of equations of the form (4.1)–(4.2) for which both ψ and χ are monotonic. Such cases, however, are of little interest here since if the ψ function exists but does not redescend to zero, then the corresponding χ_p function is unbounded and $a_\rho = \infty$. At the other extreme, for the MLE-type M-estimates, if the $\psi_{\rho, M}$ function exists and redescends strongly to zero, then $a_\rho = 0$. However, no restrictions on a_ρ are made if the ψ function exists and is strongly redescending for an S-estimate since $\psi_{\rho, S}(s) = \chi_\rho'(s) = 0$ for large s simply implies $\chi_\rho(s)$ is constant for large s.

Any ρ function for which $0 < a_\rho < \infty$ can be adjusted so that the corresponding MLE-type and S-type M-estimates of location and scale obtains the maximal value of $\varepsilon^{**}(\mathbf{X}; \mu, \sigma)$ for $\mathbf{X}$ in general position. This is done by multiplying the function ρ by $c_\rho = a_\rho(n - p)/(2n)$. The new ρ function $\rho^*(t) = c_\rho \rho(t)$ has sill $2n/(n - p)$ and consequently Corollary 4.1 applies. Note that multiplying ρ by a constant does not affect the ψ function in (4.1) for either an MLE-type M-estimate or an S-estimate. Consequently, given an M-estimating equation for location (4.1) for which ψ redescends to zero, one can choose a simultaneous M-estimate of scale of the MLE-type or of the S-type so that for the simultaneous location and scale estimate $\varepsilon^{**}(\mathbf{X}; \mu, \sigma) = \lfloor (n - p + 1)/2 \rfloor / n$ for $\mathbf{X}$ in general position. Thus, although a preference for an auxiliary estimate of scale over a simultaneous M-estimate of scale may be appropriate for monotonic ψ functions, this is not necessarily the case for redescending ψ functions. For strongly redescending M-estimates of location only, the finite sample replacement breakdown point depends on the configuration of $\mathbf{X}$, even when $\mathbf{X}$ is in general position, and can be considerably less than $\lfloor (n - p + 1)/2 \rfloor / n$ [see Huber (1984)].

As a concluding example, consider the maximum likelihood estimates for location and scale arising from a random sample from a location–scale family of distributions based on the Student's t-distribution with $\nu > 0$ degrees of freedom. This corresponds to an MLE-type M-estimate with $\rho_\nu(s^2) = (1/2)(\nu+1)\log(\nu+s^2)$, for which $a_\rho = \nu + 1$. For $\mathbf{X}$ in general position, the value of $\delta^*(\mathbf{X})$ in Theorem 4.1 is thus $\delta^*(\mathbf{X}) = 1/(\nu+1)$ for $\nu \geq (n+p)/(n-p)$, and $\delta^*(\mathbf{X}) = \{\nu/(\nu+1) - p/n\}$ for $p/(n-p) \leq \nu \leq (n+p)/(n-p)$. For $\nu = (n+p)/(n-p)$, Corollary 2.1 applies. The breakdown point decreases as $\nu \to \infty$ due to the influence of outliers and decreases as $\nu \to 0$ due to the influence of inliers. One practical drawback to nonmonotone M-estimates in general is the possibility of multiple solutions to the corresponding M-estimating equations (4.1)–(4.2). In this example, though, it is known that when $\nu \geq 1$ the likelihood function is unimodal and the maximum likelihood estimates of location and scale are the unique solution of the likelihood equations [see Copas (1975) and Märkeläinen, Schmidt and Styan (1981)]. A study of other nonmonotone MLE-type M-estimates for which the corresponding M-estimating equations have unique solutions can be found in Kent and Tyler (1991).

APPENDIX

Proof of Theorem 4.1. Some lemmas on the existence of the simultaneous M-estimates are first given. The conditions of Theorem 4.1 are understood to hold in all statements made here. A sufficient condition on $\rho(s)$ and on the univariate data set $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ for insuring that $\ell(\mu,\sigma;\mathbf{x})$ has a minimum for some $(\widehat{\mu}, \widehat{\sigma}) \in \mathbb{R} \times \mathbb{R}^+$ is that $P_n(\{x\}) < 1 - 1/a_\rho$ for all $x \in \mathbb{R}$, where P_n represents the empirical distribution measure for the data set $\mathbf{x}$. This follows since, as a special case of Lemma 2.2 in Kent and Tyler (1991),

$$\text{(A.1)} \qquad \ell(\mu,\sigma;\mathbf{x}) \to \infty \quad \text{if } |\mu| \to \infty,\ \sigma \to 0 \text{ or } \sigma \to \infty,$$

under the stated condition. Statement (A.1) implies more than the existence of MLE-type M-estimates $(\widehat{\mu}, \widehat{\sigma})$ for $\mathbf{x}$, as the first part of the following lemma shows.

LEMMA A.1. *Let $M(\mathbf{x})$ represent the set of all $(\widehat{\mu},\widehat{\sigma}) \in \mathbb{R} \times \mathbb{R}^+$ such that $\ell(\widehat{\sigma},\widehat{\mu};\mathbf{x}) \leq \ell(\mu,\sigma;\mathbf{x})$ for all $(\mu,\sigma) \in \mathbb{R}\times\mathbb{R}^+$.*

(a) *If $P_n(\{x\}) < 1 - 1/a_\rho$ for all $x \in \mathbb{R}$, then $M(\mathbf{x})$ is not empty and $M(X)$ is bounded away from $\partial(\mathbb{R}\times\mathbb{R}^+)$, the boundary of $\mathbb{R}\times\mathbb{R}^+$.*

(b) *If $P_n(\{x\}) > 1 - 1/a_\rho$ for some $x \in \mathbb{R}$, then $M(\mathbf{x})$ is empty.*

The second part of Lemma A.1 indicates that the sufficient condition for the existence of an MLE-type M-estimate is almost necessary. It is a restatement of Lemma 3.3 in Kent and Tyler (1991).

Analogous results hold for the existence of the S-estimates $(\widetilde{\mu},\widetilde{\sigma})$ for the data set $\mathbf{x}$, as the next lemma shows.

LEMMA A.2. *Let $S(\mathbf{x})$ represent the set of all $(\widetilde{\mu},\widetilde{\sigma}) \in \mathbb{R}\times\mathbb{R}^+$ such that $\widetilde{\sigma} =$*

$\widehat{\sigma}(\widetilde{\mu}) \le \widehat{\sigma}(\mu)$ for all $\mu \in \mathbb{R}$.

(a) *If $P_n(\{x\}) < 1 - 1/a_\rho$ for all $x \in \mathbb{R}$, then $S(\mathbf{x})$ is not empty and $S(\mathbf{x})$ is bounded away from $\partial(\mathbb{R} \times \mathbb{R}^+)$.*

(b) *If $P_n(\{x\}) > 1 - 1/a_\rho$ for some $x \in \mathbb{R}$, then $S(\mathbf{x})$ is empty.*

PROOF.

(a) By (A.1), $\widehat{\sigma}(\mu) \in \mathbb{R}^+$ exists for all $\mu \in \mathbb{R}$ and the set $\{\widehat{\sigma}(\mu) \mid a \le \mu \le b\}$ is bounded away from 0 and ∞ for any $-\infty < a < b < \infty$. Part (a) then follows by showing $\widehat{\sigma}(\mu) \to \infty$ as $|\mu| \to \infty$. This can be shown by contradiction. Suppose there exists a sequence $|\mu_k| \to \infty$ with $\sigma_k = \widehat{\sigma}(\mu_k) \to \sigma_\infty < \infty$, and define $L_k(\sigma) = \exp\{\ell(\mu_k, \sigma; \mathbf{x}) - \ell(\mu_k, \sigma_k; \mathbf{x})\}$. A contradiction is obtained by noting that $L_k(|\mu_k|) \to 0$ and hence $\sigma_k > |\mu_k| \to \infty$. This follows since $L_k(|\mu_k|) = \Pi_{i=1}^n b_k^{1/2} \exp\{-\rho(b_k, v_{i,k}^2)\}/\exp\{-\rho(v_{i,k}^2)\}$, where $b_k = |\mu_k|^2/\sigma_k^2 \to \infty$, $v_{i,k}^2 = (x_i - \mu_k)^2/\mu_k^2 \to 1$, and so $b_k^{1/2}\exp\{-\rho(b_k, v_{i,k}^2)\} \to 0$ since $a_\rho > 1$, while $\exp\{-\rho(v_{i,k}^2)\} \to \exp\{-\rho(1)\} > 0$.

(b) Suppose for some $x \in \mathbb{R}$, $\{i \mid x_i = x\} = q$ with $P_n(\{x\}) = q/n > 1 - 1/a_\rho$. This implies $n > (n-q)a_\rho$ and so, for some $a > a_\rho$, $n > (n-q)a$. For such a, express

$$\exp\{-\ell(x, \sigma; \mathbf{x}\} = \sigma^{-\{n-(n-q)a\}}\exp\{-q\rho(0)\} \prod_{x_i \ne x} (\sigma^2)^{-a/2}\exp\Big[-\rho\{(x_i-\mu)^2/\sigma^2\}\Big],$$

which is seen to go to infinity as $\sigma \to 0$. Thus, $\ell(x, \sigma; \mathbf{X}) \to -\infty$ as $\sigma \to 0$, whereas $\ell(\mu, \sigma; \mathbf{X}) > -\infty$ for $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$. □

Nonexistence of an MLE-type M-estimates or an S-estimate can essentially be interpreted as $\widehat{\sigma} = 0$ or $\widetilde{\sigma} = 0$, respectively. Breakdown for a location and scale statistic $[\mu(\cdot), \sigma(\cdot)]$ can occur in different ways. It is convenient here to divide the possible ways in which breakdown can occur into three disjoint categories: $\sigma(\cdot) \to \infty$ while $|\mu(\cdot)|/\sigma(\cdot)$ stays bounded above; $\sigma(\cdot) \to 0$ while $|\mu(\cdot)|$ stays bounded above; and $|\mu(\cdot)| \to \infty$ and $|\mu(\cdot)|/\sigma(\cdot) \to \infty$. The following lemma, which is a consequence of Condition T, is central to the derivation of the breakdown points of the MLE and S-type M-estimates. For any $x \in \mathbb{R}$ and $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$, let

$$g(x; \mu, \sigma) = \exp\Big[-\rho\{(x-\mu)^2/\sigma^2\}\Big]\Big/\exp\big[-\rho(x^2)\big]. \tag{A.2}$$

LEMMA A.3.

(a) *If $|x|$ is bounded above, then $g(x; \mu, \sigma)$ is bounded above.*

(b) *Suppose $\sigma \to \infty$ and $|\mu|/\sigma$ is bounded above. If $|x| \to \infty$, then $\sigma^{-a}g(x; \mu, \sigma) \to 0$ for $a > a_\rho$.*

(c) *Suppose $\sigma \to 0$ and $|\mu|$ is bounded above. If $|x - \mu|$ is bounded away from zero, then $\sigma^{-a}g(x; \mu, \sigma) \to 0$ for $a < a_\rho$.*

(d) *Suppose $|\mu| \to \infty$ and $|\mu|/\sigma \to \infty$.*

(i) *If* $|x|$ *is bounded above, then* $\{|\mu|/\sigma\}^a g(x;\mu,\sigma) \to 0$ *for* $a < a_\rho$.
(ii) *If* $|x| \to \infty$, *then* $|\mu|^{-a} g(x;\mu,\sigma) \to 0$ *for* $a > a_\rho$.

PROOF. Let $s^2 = (x-\mu)^2/\sigma^2$ throughout the proof.

(a) Since x^2 is bounded above, $\exp\{-\rho(x^2)\}$ is bounded away from zero and so $g(x;\mu,\sigma)$ is bounded above.

(b) Let $r = (s^2+1)/(x^2+1)$. Since $s^2/x^2 = \{\sigma^{-2} - (\mu/\sigma)x^{-1}\}^2 \to 0$, Condition T(ii) implies $h_a(s^2)/h_a(x^2)$ is bounded above for $a > a_\rho$. If $s^2 \to \infty$, then $r^{a/2} g(x;\mu,\sigma) = \{(1+s^{-2})/(1+x^{-2})\}^{a/2} h_a(s^2)/h_a(x^2)$ is bounded above. Moreover, it must go to zero since it is bounded above for any $a > a_\rho$ and $r \to 0$. If s^2 is bounded above, then

$$r^{a/2} g(x;\mu,\sigma) = (1+x^{-2})^{-1}\left\{\frac{h_a(s^2+1)}{h_a(x^2)}\right\}\exp\{-\rho(s^2)+\rho(s^2+1)\}$$

is still bounded above and so again must go to zero. The statement $r^{a/2} g(x;\mu,\sigma) \to 0$ holds in general since by the above results any sequence has a subsequence which goes to zero. The proof is completed by noting that $r\sigma^2 = \{(1-\mu/x)^2 + \sigma^2/x^2\}/(1+x^{-2})$ is bounded away from zero. Otherwise $\mu/x \to 1$, $\sigma^2/x^2 \to 0$ and so $\mu^2/\sigma^2 \to \infty$, a contradiction. Thus, $\sigma^{-a} g(x;\mu,\sigma) \to 0$ for any $a > a_\rho$.

(c) The proof of part (c) is divided into two cases, of which any sequence has a subsequence satisfying one of the two cases.

Case 1 ($|x|$ bounded above). For this case $s^2 \to \infty$, and so, for $a < a_\rho$, $h_a(s^2) \to 0$ or, equivalently, $s^a g(x;\mu,\sigma) \to 0$ since $\exp\{-\rho(x^2)\}$ is bounded away from zero. Furthermore, $\sigma^{-2}s^{-2} = 1/(x-\mu)^2$ is bounded above and so $\sigma^{-a} g(x;\mu,\sigma) \to 0$.

Case 2 ($x^2 \to \infty$). For this case $s^2/x^2 \to \infty$. By Condition T(i), for $a < a_\rho$, $(s^2x^2)^{a/2} g(x;\mu,\sigma) = h_a(s^2/h_a(x^2))$ is bounded above and hence goes to zero since it is bounded for any $a < a_\rho$ and $s^2/x^2 \to \infty$. Also, $\sigma^{-2}(x^2/s^2) = (1-\mu/x)^{-2}$ goes to 1. Thus, $\sigma^{-a} g(x;\mu,\sigma)$ for $a < a_\rho$.

(d) (i) As in the proof of Case 1 of part (c), $s^a g(x;\mu,\sigma) \to 0$ for any $a < a_\rho$. Furthermore $(\mu^2/\sigma^2)s^{-2} = \{1 - x/\mu\}^{-2}$ is bounded above and so $(|\mu|/\sigma)^a g(x;\mu,\sigma) \to 0$ for $a < a_\rho$.

(ii) Let $R = s^2/x^2$. The proof is divided into three cases. Any sequence of (μ,σ,x) has a subsequence satisfying at least one of the three cases.

Case 1 ($R \geq 1$). By Condition T(i), $R^{b/2} g(x;\mu,\sigma) = h_b(s^2)/h_b(x^2)$ is bounded above for $b < a_\rho$ and so $g(x;\mu,\sigma)$ is bounded above. Thus $(\mu^2)^{-a/2} g(x;\mu,\sigma) \to 0$ for any $a > 0$.

Case 2 ($R < 1$ and $R\mu^2 > \varepsilon > 0$). By condition T(ii), $R^{a/2} g(x;\mu,\sigma) = h_a(s^2)/h_a(x^2)$ is bounded above for $a > a_\rho$. Thus, $(\mu^2)^{-a/2} g(x;\mu,\sigma)$ is bounded above for any $a > a_\rho$ and consequently goes to zero since $\mu^2 \to \infty$.

Case 3 ($R\mu^2 \to 0$). Recall that Condition T(ii) implies $h_a(v) \to \infty$ as $v \to \infty$ for $a > a_\rho$. Thus, $x^{-a} g(x;\mu,\sigma) = \{h_a(1)/h_a(x)\}\exp\{-\rho(s^2)+\rho(1)\} \to 0$. The proof is completed by noting that $R\mu^2 \to 0$ implies $x/\mu \to 1$. □

To prove Theorem 4.1, it suffices to show that breakdown does not occur for $\varepsilon_m = m/n < \delta^*(\mathbf{X})$ and does occur for $\varepsilon_m > \delta^*(\mathbf{X})$. That is, the proof is completed by establishing the following lemma.

LEMMA A.4. *For either the MLE-type M-estimates* $[\widehat{\mu}(\cdot), \widehat{\sigma}(\cdot)]$ *or the S-estimates* $[\widetilde{\mu}(\cdot), \widetilde{\sigma}(\cdot)]$, *the following hold*:

(a) *If* $\varepsilon_m < \delta^*(\mathbf{X})$, *then* $\sup_{\mathbf{a}} b(\varepsilon_m, \mathbf{a}'\mathbf{X}; \mu, \sigma) < \infty$.
(b) *If* $\varepsilon_m > \delta^*(\mathbf{X})$, *then* $\sup_{\mathbf{a}} b(\varepsilon_m, \mathbf{a}'\mathbf{X}; \mu, \sigma) = \infty$.

PROOF. (a) The proof for both the MLE-type and S-type M-estimates is by contradiction. If $\sup_{\mathbf{a}} b(\varepsilon_m, \mathbf{a}'\mathbf{X}; \mu, \sigma) = \infty$, then there must exist a sequence $\mathbf{a}_k \in \mathcal{S}_{p-1}$ and a sequence of corrupted data $\mathbf{Z}_k$ such that $|\mu_k| \to \infty$, $\sigma_k \to 0$ or $\sigma_k \to \infty$, where $(\mu_k, \sigma_k) = [\widehat{\mu}(\mathbf{a}_k'\mathbf{Z}_k), \widehat{\sigma}(\mathbf{a}_k'\mathbf{Z}_k)]$ or $(\mu_k, \sigma_k) = [\widetilde{\mu}(\mathbf{a}_k'\mathbf{Z}_k), \widetilde{\sigma}(\mathbf{a}_k'\mathbf{Z}_k)]$ depending on whether the MLE-type or S-type M-estimate is being considered. Lemmas A.1(a) and A.2(a) assure that (μ_k, σ_k) exist. Without loss of generality, it can be assumed that $\mathbf{Z}_k = \mathbf{X}_0 \cup \mathbf{Y}_k$, where $\mathbf{X}_0$ consists of any $(n - m)$ fixed elements of $\mathbf{X}$ (say, $\mathbf{X}_0 = \{\mathbf{x}_1, \dots, \mathbf{x}_{n-m}\}$) and $\mathbf{Y}_k = \{\mathbf{y}_{1,k}, \dots, \mathbf{y}_{m,k}\}$). Furthermore, it can be assumed that $\mathbf{a}_k \to \mathbf{a} \in \mathcal{S}_{p-1}$ and $\mathbf{a}_k'\mathbf{y}_{i,k} \to y_i$ with $-\infty \le y_i \le \infty$ for $i = 1, \dots, m$. Define the index set $J = \{j \le m \mid |y_j| = \infty\}$ and let $m_\infty = \#J$. Consider separately now the MLE case and the S case.

(*MLE-type.*) Let $G(\mu, \sigma; \mathbf{x}) = \exp[-\{\ell(\mu, \sigma; \mathbf{x}) - \ell(0, 1; \mathbf{x})\}]$, and so

$$G(\mu, \sigma; \mathbf{a}_k'\mathbf{Z}_k) = \sigma^{-n}\left\{\prod_{i=1}^{n-m} g(\mathbf{a}_k'\mathbf{x}_i; \mu, \sigma)\right\}\left\{\prod_{j=1}^{m} g(\mathbf{a}_k'\mathbf{y}_{j,k}; \mu, \sigma)\right\}. \tag{A.3}$$

The proof consists of showing that $G(\mu_k, \sigma_k; \mathbf{a}_k'\mathbf{Z}) = 0$. This implies $\{\ell(\mu_k, \sigma_k; \mathbf{a}_k'\mathbf{Z}_k) - \ell(0, 1; \mathbf{a}_k'\mathbf{Z}_k)\} \to \infty$, which contradicts the assumption that (μ_k, σ_k) minimizes $\ell(\mu, \sigma; \mathbf{Z}_k)$. The assumption $|\mu_k| \to \infty$, $\sigma_k \to 0$ or $\sigma_k \to \infty$ is divided into the three cases corresponding to Lemma A.3(b)–(d).

Case 1. Suppose $\sigma_k \to \infty$ and $|\mu_k|/\sigma_k$ is bounded above. For this case factor $G(\mu_k, \sigma_k; \mathbf{a}_k'\mathbf{Z}_k) = H_{1,k}H_{2,k}$, where

$$H_{1,k} = \left\{\prod_{i=1}^{n-m} g(\mathbf{a}_k'\mathbf{x}_i; \mu_k, \sigma_k)\right\}\left\{\prod_{j \notin J} g(\mathbf{a}_k'\mathbf{y}_{j,k}; \mu_k, \sigma_k)\right\}$$

and

$$H_{2,k} = \sigma_k^{-n}\left\{\prod_{j \in J} g(\mathbf{a}_k'\mathbf{y}_{j,k}; \mu_k, \sigma_k)\right\}.$$

By Lemma A.3, (a) and (b), $H_{1,k}$ is bounded above and $H_{2,k} \to 0$ provided $n/m_\infty > a_\rho$, which holds since $\varepsilon_m < 1/a_\rho$.

Case 2. Suppose $\sigma_k \to 0$ and $|\mu_k|$ is bounded above. Without loss of generality, assume $\mu_k \to \mu$. Define $\mathcal{I} = \{i \le n - m \mid \mathbf{a}'\mathbf{x}_i \ne \mu\}$ and note that $\#\mathcal{I} \ge n - m - c(\mathbf{X})$. Factor $G(\mu_k, \sigma_k; \mathbf{a}'\mathbf{Z}_k) = H_{3,k}H_{4,k}$, where

$$H_{3,k} = \left\{\prod_{i \notin \mathcal{I}} g(\mathbf{a}_k'\mathbf{x}_i; \mu_k, \sigma_k)\right\}\left\{\prod_{j \notin J} g(\mathbf{a}_k'\mathbf{y}_{j,k}; \mu_k, \sigma_k)\right\}$$

and

$$H_{4,k} = \sigma_k^{-n}\left\{\prod_{i\in \mathfrak{I}} g(\mathbf{a}_k'\mathbf{x}_i;\mu_k,\sigma_k)\right\}\left\{\prod_{j\in J} g(\mathbf{a}_k'\mathbf{y}_{j,k};\mu_k,\sigma_k)\right\}.$$

By Lemma A.3(a) and (c), $H_{3,k}$ is bounded above and $H_{4,k} \to 0$ provided $n/(n-m-c(\mathbf{X})+m_\infty) < a_\rho$, which holds since $\varepsilon_m < 1 - 1/a_\rho - c(\mathbf{X})/n$.

Case 3. Suppose $|\mu_k| \to \infty$ and $|\mu_k|/\sigma_k \to \infty$. For this case, partition $G(\mu_k,\sigma_k;\mathbf{a}_k'\mathbf{Z}_k) = H_{5,k}H_{6,k}$, where

$$H_{5,k} = \left(\frac{|\mu_k|}{\sigma_k}\right)^n \prod_{i=1}^{n-m} g(\mathbf{a}_k'\mathbf{x}_i;\mu_k,\sigma_k)\prod_{j\notin J} g(\mathbf{a}_k'\mathbf{y}_i;\mu_k,\sigma_k)$$

and

$$H_{6,k} = |\mu_k|^{-k}\prod_{j\in J} g(\mathbf{a}_k'\mathbf{y}_{i,k};\mu_k,\sigma_k).$$

By Lemma A.3(d), $H_{5,k} \to 0$ and $H_{6,k} \to 0$ provided $n/(n-m_\infty) < a_\rho < n/m_\infty$, which holds since $\varepsilon_m < 1 - 1/a_\rho - c(\mathbf{X})/n$ and $\varepsilon_m < 1/a_\rho$.

(*S-type*.) From the definition of the S-estimates, for the same ρ function, $\tilde{\sigma}(\cdot) \le \hat{\sigma}(\cdot)$, and so the case $\sigma_k \to \infty$ can be discounted since the MLE-type M-estimator does not break down.The proof is divided into three remaining cases.

Case 1 ($\sigma_k \to 0$ with $|\mu_k|$ bounded above). It can be shown that $G(0,\sigma_k;\mathbf{a}_k'\mathbf{Z}_k - \mu_k) \to 0$, where G is defined by (A.3). This implies $\{\ell(\mu_k,\sigma_k;\mathbf{a}_k'\mathbf{Z}_k) - \ell(\mu_k,1;\mathbf{a}_k'\mathbf{Z}_k)\} \to \infty$, which contradicts the assumption that σ_k minimizes $\ell(\mu_k,\sigma;\mathbf{a}_k'\mathbf{Z}_k)$. The proof that $G(0,\sigma_k;\mathbf{a}_k'\mathbf{Z}_k - \mu_k) \to 0$ is identical to the proof for Case 2 for the MLE-type M-estimate after changing $(\mu_k,\mathbf{a}_k'\mathbf{Z}_k)$ in the former proof to $(0,\mathbf{a}_k'\mathbf{Z}_k - \mu_k)$.

Case 2 ($\sigma_k \to 0$ with $|\mu_k| \to \infty$). This proof is also obtained by arguing that $G(0,\sigma_k;\mathbf{a}_k',\mathbf{Z}_k - \mu_k) \to 0$. To show this, assume without loss of generality $\mathbf{a}_k'\mathbf{y}_{j,k} - \mu_k \to y_j^*$, with $-\infty \le y_j^* \le \infty$ for $j = 1,\dots,m$, and let $J^* = \{j = 1,\dots,m \mid y_j^* \ne 0\}$ and $m^* = \#J^*$. Factor $G(0,\sigma_k;\mathbf{a}_k'\mathbf{Z}_k - \mu_k) = M_{1,k}M_{2,k}$, where

$$M_{1,k} = \prod_{j\notin J^*} g(\mathbf{a}_k'\mathbf{y}_{j,k} - \mu_k;0,\sigma_k)$$

and

$$M_{2,k} = \sigma^{-n}\left\{\prod_{i=1}^{n-m} g(\mathbf{a}_k'\mathbf{x}_i - \mu_k;0,\sigma_k)\right\}\left\{\prod_{j\notin J^*} g(\mathbf{a}_k'\mathbf{y}_{j,k} - \mu_k;0,\sigma_k)\right\}.$$

By Lemma A.3(a), $M_{1,k}$ is bounded above. By Lemma A.3(c), $M_{2,k} \to 0$ provided $n/(n-m+m^*) < a_\rho$, which holds since $n/(n-m+m^*) \le 1/(1-\varepsilon_m) < a_\rho$.

Case 3 ($|\mu_k| \to \infty$ with σ_k bounded away from zero and infinity). It is to be shown that, for $\gamma_k = |\mu_k|$, $G[0,\sigma_k/\gamma_k;\gamma_k^{-1}(\mathbf{a}_k'\mathbf{Z}_k - \mu_k)] \to 0$. This implies $\{\ell(\mu_k,\sigma_k;\mathbf{a}_k'\mathbf{Z}_k) - \ell(\mu_k,\gamma_k;\mathbf{a}_k'\mathbf{Z}_k)\} \to \infty$, which contradicts the definition of σ_k. Without loss of generality, assume $(\mathbf{a}_k'\mathbf{y}_{i,k} - \mu_k)/\gamma_k \to y_j^0$, with $-\infty \le y_j^0 \le$

∞, $j = 1, \ldots, m$, and let $J^0 = \{j = 1, \ldots, m \mid |y_j^0| = \infty\}$ and $m^0 = \#J^0$. Partition $G[0, \sigma_k/\gamma_k; \gamma_k^{-1}(\mathbf{a}_k'\mathbf{Z}_k - \mu_k)] = M_{3,k} M_{4,k}$, where

$$M_{3,k} = \prod_{j \notin J^0} g\left\{\frac{(\mathbf{a}_k'\mathbf{y}_{j,k} - \mu_k)}{\gamma_k}; 0, \frac{\sigma_k}{\gamma_k}\right\}$$

and

$$M_{4,k} = (\sigma_k/\gamma_k)^{-n}\left\{\prod_{i=1}^{n-m} g\left\{\frac{(\mathbf{a}_k'\mathbf{x}_i - \mu_k)}{\gamma_k}\right\}; 0, \frac{\sigma_k}{\gamma_k}\right\}\left\{\prod_{j \in J^0} g\left\{\frac{(\mathbf{a}_k'\mathbf{y}_{j,k} - \mu_k)}{\gamma_k}\right\}; 0, \frac{\sigma_k}{\gamma_k}\right\}.$$

By Lemma A.2(a), $M_{3,k}$ is bounded above. By Lemma A.2(c), $M_{4,k} \to 0$ provided $n/(n - m + m^0) < a_\rho$, which holds since $n/(n - m + m^0) \le 1/(1 - \varepsilon_m) < a_\rho$.

(b) (*MLE-type.*) If $\varepsilon_m > 1 - 1/a_\rho - c(\mathbf{X})/n$, consider $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ such that $\mathbf{y}_j = \mathbf{x}_1$ for $j = 1, 2, \ldots, m$. Without loss of generality, suppose for some $\mathbf{a} \in \mathcal{S}_{p-1}$ that the size of the set $D = \{i \mid \mathbf{a}'\mathbf{x}_i = \mathbf{a}'\mathbf{x}_1\}$ is $c(\mathbf{X})$. If $\mathbf{Y}$ replaces m values in $\mathbf{X}$ with none of these m values $x_i \in D$, then Lemma A.1(b) implies $[\widehat{\mu}(\mathbf{a}'\mathbf{Z}), \widehat{\sigma}(\mathbf{a}'\mathbf{Z})]$ does not exist. Thus, $\sup_{\mathbf{a}} b(\varepsilon_m, \mathbf{a}'\mathbf{X}; \mu, \sigma) = \infty$.

If $\varepsilon_m > 1/a_\rho$, consider the sequence $\mathbf{Z}_k = \mathbf{X}_0 \cup \mathbf{Y}_k$, where $\mathbf{X}_0 = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n-m}\}$ and $\mathbf{Y}_k = \{\mathbf{y}_{1,k}, \ldots, \mathbf{y}_{m,k}\}$, with $\mathbf{y}_{j,k} = \mathbf{y}_k^0$ for $j = 1, \ldots, \infty$, and $\lambda_k = |\mathbf{a}'\mathbf{y}_k^0| \to \infty$ for some $\mathbf{a} \in \mathcal{S}_{p-1}$. If breakdown does not occur under this sequence, a contradiction arises by showing $G_k = \exp\{\ell(\mu_k, \lambda_k; \mathbf{a}'\mathbf{Z}_k) - \ell(\mu_k, \sigma_k; \mathbf{a}'\mathbf{Z}_k)\} \to 0$, where $\mu_k = \widehat{\mu}(\mathbf{a}'\mathbf{Z}_k)$ and $\sigma_k = \widehat{\sigma}(\mathbf{a}'\mathbf{Z}_k)$. To show this, express $G_k = \gamma_k^{-n} g^m(v_k; 0, \gamma_k) H_k$, where $\gamma_k = \sigma_k/\lambda_k \to 0$, $v_k = (\lambda_k - \mu_k)/\lambda_k \to 1$, and $H_k = \prod_{i=1}^{n-m} g(v_{i,k}; 0, \gamma_k)$ with $v_{i,k} = (\mathbf{a}'\mathbf{x}_i - \mu_k)/\lambda_k \to 0$. By Lemma A.3(a), H_k is bounded above and so, by Lemma A.3(c), $G_k \to 0$ since $n/m = 1/\varepsilon_m < a_\rho$.

(*S-type.*) The proof is analogous to the proof for the MLE-type M-estimates. □

Acknowledgments. I am extremely grateful to the two referees who made extensive comments on the original manuscript. Their comments were particularly helpful in clarifying many ambiguities in the presentation and proofs of the paper. My gratitude is also extended to the Ricardo Maronna, who afforded me access to his unpublished work and who made valuable comments on the original manuscript.

REFERENCES

COPAS, J. B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62** 701–704.

DAVIES, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Dept. Statistics, Harvard Univ.

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

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

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

HODGES, J. L., Jr. (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. (1984). Finite sample breakdown point of M- and P-estimators. *Ann. Statist.* **12** 119–126.

HUBER, P. J. (1985). Projection pursuit. *Ann. Statist.* **13** 435–474.

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

LOPUHAÄ, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17** 1662–1683.

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.

MÄRKELÄINEN, T., SCHMIDT, K. and STYAN, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter and in fixed sample sizes. *Ann. Statist.* **9** 758–767.

MARONNA, R. A. (1991). Personal correspondence.

MARONNA, R. A., STAHEL, W. A. and YOHAI, V. J. (1992). Bias-robust estimators of multivariate scatter based on projections. *J. Multivariate Anal.* **42** 141–161.

MARONNA, R. A. and YOHAI, V. J. (1989). Bias-robust estimates of multivariate scatter based on projections. Unpublished manuscript. (Paper presented at 1989 IMA summer workshop on robustness in statistics, Minneapolis, MN).

NOLAN, D. (1992). Asymptotics for multivariate trimming. *Stochastic Process. Appl.* **42** 157–169.

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

ROUSSEEUW, P. J. and YOHAI, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

STAHEL, W. A. (1981). Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, ETH, Zürich.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903