A reinvestigation of robust scale estimation in finite samples

John A. Randal [1,*]

[a] *School of Economics and Finance, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand, phone +64 4 463 5558, email john.randal@vuw.ac.nz.*

Abstract

This paper reworks and expands on the results of existing simulation studies, investigating the performance of various robust estimators of scale for Tukey's three corner distributions. We focus attention on the popular biweight A-estimator, but also propose a new estimator based on the Student's t-distribution, which attains efficiency close to that of the A-estimator. We investigate use of more efficient auxiliary location and scale estimators in the two-pass estimators such as the A- and t-estimators, and find overall efficiency can be improved. Using much larger simulation sizes than previous studies, significant departures from existing efficiencies are obtained, and these lead to different recommendations for estimation.

Key words: Robustness; Scale estimation; EM algorithm; M-estimation.

* Corresponding author.
Email address: john.randal@vuw.ac.nz (John A. Randal).

[1] Thanks to Peter Thomson for advice and support, Robert Davies, and conference participants at the New Zealand Statistical Association Conference 2004, the 22nd International Biometrics Conference, and the 2005 International Conference on Robust Statistics for suggestions.

In this paper, we consider identification of promising robust estimators of scale in finite samples. This problem has already been addressed in the literature (Lax 1975, Rousseeuw & Croux 1993, Iglewicz 2000) but primarily by Lax (1985). The Lax study is a fundamental one, being the only study to survey the efficiency properties of robust scale estimators, following from Andrews, Bickel, Hampel, Huber, Rogers & Tukey (1972) and their survey of robust location estimators. Since the publication of Lax's results, new estimators have been proposed, primarily Rousseeuw & Croux's (1993) S_n and Q_n, and these need to be subjected to the same analysis. Unfortunately, Lax conducted his simulation in an era when computing power was much more limited than what is available today, and as a consequence, his efficiency estimates were very imprecise, and need to be reassessed in light of modern computing developments.

The contribution of this study is threefold: the simulation study we perform dwarfs that of Lax, with the simulation size several orders of magnitude greater; unlike Lax, we benchmark all estimators' efficiencies against the optimal (maximum likelihood) estimators; and we consider new estimators of scale in addition to those examined by Lax.

While much is understood of the asymptotic properties of scale estimators, finite sample performance is still an important consideration. Robust estimators are particularly applicable for financial data, which often feature the three situations we are protecting against through use of a robust estimator: occasional rogue values, many small errors (induced by properties of financial markets such as discrete price intervals and discontinuous trading) and underlying non-normality (Rydén, Teräsvirta & Åsbrink 1998, Cont 2001), but also in other areas, e.g. GPS navigation (Mertikas 1994). In the financial setting, sample sizes are by necessity small, to allow estimation of evolving scale.

1 Assessing robust scale estimators

A scale estimator satisfies the condition $S(a+b\mathbf{X}) = |b|S(\mathbf{X})$, and consequently scale estimators are not unique, i.e. if $S_1(\mathbf{X})$ is a scale estimator, then so is $kS_1(\mathbf{X})$ for any positive constant k. Due to this constant of proportionality, evaluating scale estimators requires a non-standard specification for efficiency, and we discuss two alternative approaches.

The first method is used by Lax (1985). Based on m independent samples, he estimates efficiency using

$$\widehat{\text{eff}}(S) = \frac{\text{sample variance of } \ln\hat{\sigma}_1, \ldots, \ln\hat{\sigma}_m}{\text{sample variance of } \ln S(\mathbf{X}_1), \ldots, \ln S(\mathbf{X}_m)} \tag{1}$$

where each sample of n observations is $\mathbf{X} = (X_1, \ldots, X_n)$, $\hat{\sigma}_i$ is the optimum scale estimate for sample i and $S(\mathbf{X}_i)$ is the scale estimate for sample i. The log transformation reduces the constant of proportionality to an additive constant which is then eliminated by the variance operation.

The second measure we consider was employed by Rousseeuw & Croux (1993), and is based on the squared coefficient of variation. Multiplied by sample size n, this quantity is called the standardised variance, and is given by

$$n\,\text{cv}(S)^2 = \frac{n(\text{sample variance of } S(\mathbf{X}_1), \ldots, S(\mathbf{X}_m))}{(\text{sample mean of } S(\mathbf{X}_1), \ldots, S(\mathbf{X}_m))^2}. \tag{2}$$

Again, scaling constants are removed through cancellation in the numerator and denominator. The efficiency based on standardised variances is given by

$$\widehat{\text{eff}}'(S) = \left(\frac{\text{cv}(\hat{\sigma})}{\text{cv}(S)}\right)^2 \tag{3}$$

where again $\hat{\sigma}$ is the optimal estimator. Unlike (1), in which a stabilising log transformation is used, the variance used in (3) penalises heavily the inefficient

scale estimators for heavy-tailed data.

The scale estimators considered in Lax's study and here, are computed for independent samples from Tukey's three corners. The first of these is the normal corner, consisting of a random sample of size n from the standard normal distribution. The second corner is the one-wild (also known as 1-wider), where $n-1$ of the observations are independent and identically distributed standard normal random variables and the remaining observation is also normal but with standard deviation 10. The final corner is the slash, where the independent observations are formed by dividing a standard normal random variable by an independent uniform random variable on the interval $[0, 1]$. As our interest is in finite sample performance, rather than asymptotic results (see Huber (1981) and Hoaglin, Mosteller & Tukey (2000) for discussion of asymptotic theory), we conduct simulations for samples of 20 observations, as in Lax (1985).

The normal, one-wild and slash sampling situations used here were considered by Tukey to reflect the three extreme cases of importance to robust statistics. The normal distribution is symmetric and has rapidly decaying tails; the one-wild allows the presence of a single, potentially outlying, but otherwise well behaved value (in the upper or lower tail with equal probability); and the slash has very slowly decaying tails, and even an undefined mean. In practice, most one-wild samples will be highly asymmetric, with the presence of the single outlier. The scaling factor of 10 most likely was motivated by the possibility of transcription error, where the decimal point is incorrectly moved one place to the right.

An estimator's overall quality was assessed by Lax (1985) using the trieffi-ciency criterion promoted by Tukey (see also Goodall (2000) and Iglewicz (2000)). An estimator's triefficiency is simply the minimum efficiency of the estimator over the three corners, and the "best" estimator will have the largest

possible triefficiency. We would expect the triefficiency of this estimator to be less than 100% since no single estimator can be optimal at all three corners. Should we have data whose distribution is unknown, but not as "extreme" as any of the three corners, the triefficient estimator will give us a reliable estimate of scale for this data, regardless of the actual distribution.

To further enhance the importance of the triefficiency criterion, Yatrakos (1991) shows that for any convex combination of the three corner distributions, any estimator will have efficiency at this distribution at least as great as its triefficiency. If we believe that the corners are indeed the extremes, then we can be confident in using the estimators that perform well in the simulations that are presented in Section 4. We note that the estimators which perform well in these simulations will be good general-purpose estimators, but are unlikely to be optimal in any single sampling situation. Their most likely use is when a reliable measure of scale is required for some data whose distribution remains unspecified.

2 The Lax study and triefficiency estimation

In Section 4 we compare our results with those of Lax (1985). The estimators examined by Lax include those that are well known, and also others which are not well known. Prominent in the Lax study, and here, is the biweight A-estimator. The A-estimator for the sample $\mathbf{X} = (X_1, \ldots, X_n)$, with ψ-function $\psi(u) = u\, w(u)$ is

$$s_\psi(\mathbf{X}; M_0, S_0, c) = \left[\frac{1}{n-1} \frac{\sum_{i=1}^n w(U_i)^2 (X_i - M_0)^2}{\left[\frac{1}{n}\sum_{i=1}^n \psi'(U_i)\right]^2} \right]^{\frac{1}{2}} \tag{4}$$

where $U_i = (X_i - M_0)/(cS_0)$, $M_0 \equiv M_0(\mathbf{X})$ is an auxiliary estimator of location, $S_0 \equiv S_0(\mathbf{X})$ is an auxiliary estimator of scale, $c > 0$ is a scaling constant,

and $\psi'(u) = \frac{d}{du}\psi(u)$. The A-estimator is a weighted average of the squared deviations of the data about an auxiliary location estimator, and the most successful weighting function identified by Lax is based on the biweight function, with

$$w(u) = \begin{cases} (1-u^2)^2 & |u| \leq 1 \\ 0 & |u| > 1. \end{cases} \tag{5}$$

This weight function is used in the non-parametric robust smoothing algorithm loess, developed by Cleveland, Grosse & Shyu (1992). This symmetric weight function decays relatively slowly about $u = 0$, but as $|u|$ increases it declines smoothly to zero, which it attains at $u = \pm 1$. Since the biweight function has zero weight for $|u| > 1$, it cannot be the maximum likelihood estimator for any target distribution (Goodall 2000), and consequently it cannot be 100% efficient for some combination of the three corners.

In Lax's study, the biweight A-estimator with M_0 equal to the sample median, S_0 equal to the median absolute deviation (MAD), and scaling constant $c = 9$ had the highest triefficiency. Its triefficiency was 85.8% indicating a minimum efficiency of 85.8% for the three corners and any convex combination of them. The actual efficiencies reported were 86.7%, 85.8% and 86.1% for the normal, one-wild and slash corners respectively, and these figures were based on 1000 samples for the normal corner, and 640 samples for the one-wild and slash corners, each of size $n = 20$.

Replicating Lax's efficiencies 100 times for this estimator (based on 1000 samples) indicates a standard deviation of 3.0%, 2.9% and 2.2% for the normal, one-wild and slash efficiencies respectively. In an attempt to increase the effective sample size, Lax used the variance reduction technique of Simon (1976). While in some cases, an increase in precision was attained (up to that of roughly four times the number of samples), in many other situations the variance reduction technique made no gain. Overall, the lack of precision in indi-

vidual efficiency estimates may have a large influence on the triefficiency.

While the normal efficiency was calculated by Lax relative to the (optimal) sample standard deviation, both the one-wild and slash efficiencies were simply calculated relative to the best performing estimator in the study, and hence, regardless of any precision issues, the triefficiencies are too high.

For the one-wild corner, efficiencies were calculated relative to a trimmed sample standard deviation with 20% trimming for both the mean and variance calculations. In the simulations reported in Section 4, we show that this estimator has an actual efficiency of 71.2%, and hence, each of Lax's one-wild efficiencies will be too large by a factor of $1.40 \approx 1/0.712$. Similarly, Lax's slash efficiencies were calculated relative to an M-estimator, which is also not optimal. We find the M-estimator in question has an efficiency of 92.4%, and conclude that Lax's slash efficiencies are overstated by a factor of $1.08 \approx 1/0.924$. On this basis alone, we might revise Lax's efficiencies for the best biweight A-estimator to be 86.7%, 61.1% and 79.6% indicating a much lower triefficiency of 61.1%.

It follows that every estimator's one-wild and slash efficiencies are overstated by the same factors, and consequently that Lax's triefficiencies have limited interpretation. Rescaling Lax's efficiencies indicate that the one-wild corner will be the smallest for many of the estimators, and hence the one-wild efficiency will often correspond to the triefficiency. Differences between Lax's results and those we obtain cannot be attribued solely to this effect though.

It is clear that while relative efficiencies are interesting when comparing estimators for a single distribution, in order to correctly calculate triefficiencies, efficiencies at each corner *must* be correctly benchmarked. In order to achieve this, maximum likelihood estimators of scale for the one-wild and slash corners are needed. In the latter case, the maximum likelihood recursions are

well known (see Kafadar 1982), however these were not used by Lax. Kafadar (1982) stated that the one-wild sample is not a sample from any particular distribution, and that no maximum likelihood method is helpful. However the parameters of a one-wild sample do indeed have maximum likelihood estimates, and the EM algorithm yields these. Details of the recursion equations can be found in Randal & Thomson (2004).

In the simulation study reported in the following sections, we use maximum likelihood estimates for each corner in order to correctly calculate the trief-ficiency statistics. In addition, larger simulation sizes ensure our estimated efficiencies are much more precise than those given by Lax. As a result of these two facts, we find considerable differences to the point estimates given by Lax.

3 Simulation methodology

In this simulation study, we consider independent samples of size $n = 20$. Each efficiency estimate is based on $m = 20{,}000$ samples from one of the three corners. In addition, we obtain 100 independent realisations of the efficiency estimate for each combination of distribution and estimator and report average efficiencies.

Recall that Lax (1985) used only $m = 1000$ samples for the normal and $m = 640$ for the one-wild and slash corners, with a variance reduction technique to improve precision. Since the variance reduction technique significantly increases simulation time, but has no effect on precision for many estimators, it was not employed in this study. We attain high precision in the reported efficiencies by our larger choice of m, but also by replicating these point estimates 100 times, and reporting the average. The individual efficiencies have standard deviation approximately equal to 0.55% on average, so the averages we report

have standard errors equal to approximately 0.055% on average. Application of the central limit theorem to the average efficiency then facilitates interval estimates, of width approximately 0.22%.

All simulations were conducted using the statistical software R (R Development Core Team 2004), and code is available from the author. Maximum likelihood estimates for each one-wild and slash sample were computed using the EM algorithm (Dempster, Laird & Rubin 1977) and the results of Randal & Thomson (2004).

As in Lax (1985), we evaluate the performance of: the sample standard deviation; trimmed standard deviation with 20% trimming for both the mean and variance calculations; median absolute deviation about the median (MAD); M-estimators using the Huber ψ-function, fully iterated with $b = 1.4$ and 1.7, and one-step with $b = 1.4$, 1.7 and 2.0; biweight A-estimators with $S_0 = \text{MAD}$ and scaling constants $c \in \{6, 7, 8, 9, 10\}$; the modified biweight A-estimator with $c = 6$; the Andrew's sine A-estimator with $c = 2.1$ and the modified sine with $c = 2.1$. We also include the interquartile range, and S_n and Q_n (Rousseeuw & Croux 1993).

In addition, we investigate the effect of alternative auxiliary estimators for the biweight A-estimators. Namely, we analyse the effect of replacing the sample median by the more efficient one-step biweight M-estimator (see Iglewicz 2000), and the MAD by either S_n or Q_n. In the latter cases, scaling constants were selected based on maximum triefficiency for samples of size $n = 20$, subject to a grid search over increments of 0.5 for A-estimators with $S_0 = S_n$ and Q_n.

Finally, we introduce an additional estimator based on the t-distribution. The t-distribution is chosen as a target family of distributions which, when appropriately tuned, may describe a compromise position between the normal

and slash distributions in particular. An EM algorithm (Dempster et al. 1977) can be developed to calculate maximum likelihood estimates of location and scale for a random sample from the t-distribution with ν degrees of freedom. A t_ν random variable can be written in the Gaussian compound scale model framework of Randal & Thomson (2004) as

$$X_i = \mu + \sigma \frac{Z_i}{\sqrt{S_i}}$$

where Z_i is standard normal, and νS_i is an independent χ^2_ν random variable. In this context, we observe only the X_i, and the EM algorithm leads us to maximise the conditional expected value of a complete likelihood function which assumes both the X_i and S_i are observed. Noting that given S_i, the X_i are independent normal random variables with mean μ and variance $\sigma^2 S_i^{-1}$, the complete likelihood function can be formed by evaluating $L_c(\mu, \sigma^2) = \prod_{i=1}^n f_{X_i|S_i}(x_i|s_i) f_{S_i}(s_i)$. Gathering terms involving μ and σ^2, the complete log-likelihood is given by

$$\ln L_c(\mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n S_i (X_i - \mu)^2 + \text{constant}.$$

Since the X_i are independent, $E_0\{\ln L_c(\mu, \sigma^2)|X_1, \ldots, X_n\}$ is

$$E_0\{\ln L_c(\mu, \sigma^2)|X_1, \ldots, X_n\} = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n E_0(S_i|X_i)(X_i - \mu)^2 + \text{constant}$$

and maximising this with respect to μ and σ^2 yields the iterative equations for the maximum likelihood estimates

$$\begin{aligned} \hat{\mu} &= \frac{\sum_{i=1}^n E_0(S_i|X_i) X_i}{\sum_{i=1}^n E_0(S_i|X_i)} \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n E_0(S_i|X_i)(X_i - \hat{\mu})^2 \end{aligned} \tag{6}$$

where the conditional expectations are evaluated using the previous parameter estimates.

Randal & Thomson (2004) show that $E_0(S_i|X_i)$ can be conveniently evaluated in the case of independent data using

$$E_0(S_i|X_i) = -\frac{d}{dt} \ln M_i(t)\Big|_{t=\frac{1}{2}(X_i-\hat{\mu}_0)^2/\hat{\sigma}_0^2}$$

where $M_i(t) = \int_0^\infty e^{-ts}\sqrt{s} f_{S_i}(s)ds$. In the special case where the X_i are independent t_ν random variables, $f_{S_i}(s) = \nu f_{\chi^2_\nu}(\nu s)$ and $\sqrt{s} f_{S_i}(s) \propto s^{\frac{1}{2}(\nu+1)-1} e^{-\frac{1}{2}\nu s}$ and so $M_i(t)$ is the moment generating function of a gamma random variable with parameters $\frac{1}{2}(\nu+1)$ and $\frac{1}{2}\nu$, i.e. $M_i(t) = (\nu/(\nu+2t))^{\frac{1}{2}(\nu+1)}$. This yields

$$E_0(S_i|X_i) = \frac{\nu+1}{\nu}\left(1 + \frac{(X_i-\hat{\mu}_0)^2}{\nu\hat{\sigma}_0^2}\right)^{-1}$$

which can be inserted into (6) to give a scale estimator based on the shape of the t_ν distribution.

Reparameterising this weight function and ignoring multiplicative constants, we obtain the t-estimator

$$s_t(\mathbf{X}; M_0, S_0, c) = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{1+U_i^2}\right)(X_i - M_0)^2\right]^{\frac{1}{2}} \tag{7}$$

where $U_i = (X_i - M_0)/(cS_0)$, $M_0 \equiv M_0(\mathbf{X})$ is an auxiliary location estimator, $S_0 \equiv S_0(\mathbf{X})$ is an auxiliary scale estimator, and $c > 0$ is a scaling constant. Like the A-estimator, the t-estimator calculates an updated estimate of scale using an auxiliary estimate, and it has a positive tuning constant.

Since the fully iterated t-estimator is a maximum likelihood estimator, we can evaluate its asymptotic breakdown point (Huber 1981, pp 109-110). In

particular, the χ-function is

$$\chi(x) = -x\frac{f'(x)}{f(x)} - 1 = \frac{\frac{1}{2}(\nu+1)}{1+\frac{\nu}{x^2}}$$

where $f(x)$ is the density function of the t_ν random variable . The asymptotic breakdown point is given by

$$\epsilon^* = \frac{-\chi(0)}{\chi(\infty)-\chi(0)} = \frac{2}{\nu+3} \leq \frac{1}{2}$$

and this quantity attains its upper bound when $\nu = 1$. As $\nu \to \infty$, and the t_ν approaches the Gaussian distribution, the asymptotic breakdown point converges to zero as expected. Consequently, as the degrees of freedom parameter increases, the t-estimator can be expected to have greater exposure to contamination, and be less robust.

The factor cS_0 in the t-estimator implicitly determines the value of the degrees of freedom parameter corresponding to the optimal scaling constant. As is clear from the derivation above, the actual form of the EM algorithm uses $\sqrt{\nu}\hat{\sigma}$ in the denominator of U_i, and so if S_0/k is a consistent estimator of the scale parameter σ, then $\nu = k^2c^2$. Fixing c to be the same for all three corners, it is impossible to infer a unique choice of ν however, since the factor k depends on the interaction between the distribution of the data and behaviour of the estimator S_0, and so ν will be different for each corner.

4 Simulation results

4.1 Replicated results

The first results we report are a direct comparison with Lax (1985) and other published studies. Differences in methodology include much larger simulation

sizes, correct benchmarking of the efficiencies, and possible beneficial influence of modern random number generators and numerical routines. The (point) efficiency estimates are shown in Table 1 along with the historical results. The majority of these are from Lax (1985), however Iglewicz's (2000) results for the fourth spread are compared to ours for the sample interquartile range, and Rousseeuw & Croux's (1993) figures for S_n and Q_n at the normal distribution are also included.

Consistent with Lax's normal efficiencies being correctly benchmarked against the sample standard deviation, we see that the results for the normal distribution are in close agreement, except for the trimmed standard deviation, the M-estimators, and the modified sine A-estimator. In contrast, efficiencies for the one-wild corner have been typically overstated in previous studies, however, the inflation factors are not constant across estimators. This indicates incorrect benchmarking is not the only factor, and that the efficiencies calculated in previous studies are possibly erroneous. Similar comments apply to the slash efficiencies, however the differences are less systematic with both under and overstatement common.

The estimators in Table 1 can be divided into four groups: the standard deviations, the MAD and its alternatives IQR, S_n and Q_n, the M-estimators, and the A-estimators.

The standard deviations are poor estimators under a triefficiency criterion. With no trimming, the standard deviation is terribly inefficient for slash data (the log estimates have infinite variance in theory), and while trimming adequately combats this effect, it results in decreased efficiency for normal data. As a consequence, both estimators have low triefficiency. No attempt has been made to optimise the trimmed standard deviation. We do note that decreasing the amount of trimming will improve efficiency for normal data, but worsen efficiency for slash data. Consequently, we might expect trimming parameters

estimator	normal		one-wild		slash	
sample standard deviation	100.0	(100.0)	11.4	(10.9)		
trimmed sd with $p = r = 0.2$	65.0	(89.9)	70.8	(100.0)	76.1	(28.1)
median absolute deviation	37.8	(35.3)	40.5	(41.5)	87.3	(91.8)
interquartile range	39.4	(41)	42.4	(47)	84.0	(94)
S_n	56.3	(54.1)	55.9		95.8	
Q_n	68.3	(68.8)	68.4		94.9	
M-Estimators (Huber ψ-function)						
$b = 1.4$ (iterated)	66.4	(48.1)	70.6	(56.8)	92.4	(100.0)
$b = 1.7$ (iterated)	78.8	(72.3)	82.8	(83.8)	77.8	(83.8)
$b = 1.4$ (one-step)	47.0	(55.2)	50.2	(68.1)	95.2	(86.8)
$b = 1.7$ (one-step)	57.3	(60.5)	60.3	(71.8)	88.8	(83.1)
$b = 2.0$ (one-step)	63.5	(69.8)	65.9	(76.1)	82.2	(75.9)
A-Estimators						
biweight with $c = 6$	67.3	(65.2)	67.8	(77.1)	88.3	(90.1)
biweight with $c = 7$	75.9	(74.8)	74.1	(82.9)	89.1	(89.3)
biweight with $c = 8$	81.9	(81.8)	77.5	(85.4)	88.8	(87.6)
biweight with $c = 9$	86.2	(86.7)	79.1	(85.8)	88.0	(86.1)
biweight with $c = 10$	89.4	(90.0)	79.2	(84.8)	86.8	(84.6)
modified biweight with $c = 6$	50.0	(47.5)	53.3	(56.8)	92.5	(96.8)
sine with $c = 2.1$	78.5	(77.5)	75.5	(83.7)	88.6	(88.4)
modified sine with $c = 2.1$	78.1	(82.1)	75.3	(89.6)	89.0	(94.5)

Table 1
Comparison of efficiencies with those from published studies, shown in parentheses. Results for the interquartile range are compared to results given in Iglewicz (2000) for the fourth spread, the results for S_n and Q_n based on standardised variances to those given in Rousseeuw & Croux (1993), and all others to those given in Lax (1985). The efficiencies stated are averages based on 100 individual values, each from 20,000 samples of size 20.

less than 0.2 to give the best performance for this class of estimators, with a triefficiency approximately equal to 70% (roughly the average of the current normal and slash efficiencies).

With the exception of the iterated Huber M-estimator with $b = 1.7$, the

M-estimators considered by Lax perform poorly here. The iterated estimator with $b = 1.7$ attains the best efficiency for one-wild data of those estimators reported in Table 1, but does not do as well for either of the other two corners. Due to computational difficulties, i.e. convergence concerns discussed by Lax, the M-estimators were not considered further in this study.

4.2 Evaluating alternatives to the MAD

The MAD is often used as an auxiliary scale estimator (see Cleveland et al. (1992) for example) but we see in Table 1 that while highly efficient for slash data, the MAD is a relatively inefficient estimator for normal and one-wild data. The interquartile range (IQR) does not do as well as MAD for slash samples, but is more efficient than the MAD for both normal and one-wild data, and consequently has higher triefficiency. Thus, under the triefficiency criterion, the IQR should be preferred to the MAD, which is a surprising conclusion given the IQR's relative simplicity, and the MAD's popularity.

Table 1 also indicates that Rousseeuw & Croux's (1993) "alternatives to the MAD" S_n and Q_n perform well, with both of them dominating each of the MAD and the IQR, due to higher efficiencies at every corner. S_n is marginally better than Q_n at the slash corner, but Q_n has a higher efficiency at both normal and one-wild corners, and consequently a higher triefficiency. We conclude that Q_n is the best of these four estimators, with a triefficiency almost double that of the MAD. As indicated by Rousseeuw & Croux, S_n and Q_n are both superior scale estimators to the MAD.

These conclusions are unchanged by using the alternative measure of efficiency (3). The rankings for the estimators are unchanged for each corner, and both Q_n and S_n still dominate each of the MAD and IQR. Actual efficiencies increase in each instance for normal and one-wild data, while slash efficiencies

decrease. Since the triefficiencies are always based on the normal corner, these also are slightly higher under this measure.

Due to the result of Yatrakos (1991), use of Q_n generally is likely to guarantee the user an efficiency of at least 66.9%, based on the efficiency measure (1). We note that the best trimmed standard deviation is likely to have a triefficiency higher than this, but it will not offer the same high efficiency for slash data.

4.3 Tuning the biweight A-estimator

A simultaneous virtue and vice of A-estimation is the degree of flexibility these estimators have. The user can not only freely specify the form of the ψ-function, but also the auxiliary location estimator M_0, the auxiliary scale estimator S_0, and the scaling constant c. This can be regarded a virtue due to the fact that there are so many ways in which to potentially improve the performance of this class of estimator, however, with no theoretical basis on which to appraise our choices, it can be difficult to optimise performance.

In this study, we have eliminated the choice of ψ-function by focusing attention on the biweight ψ-function, possibly at the expense of failing to identify better estimators. We select the biweight on the basis of popularity and also the success of this estimator in the Lax study. The median is chosen as the auxiliary location estimator in each case due to its simplicity and its reasonable efficiency at each of the three corners. Randal & Thomson (2004) calculate efficiencies for the sample median to be 67.5%, 67.3% and 83.7% for the normal, one-wild and slash respectively, using the same methodology as this simulation. We investigate the effect of this choice in two cases, and the results of this particular experiment are reported later in this section.

Having fixed the ψ-function and M_0 for the A-estimator, we investigate in detail the choice of S_0 and c. These two "parameters" are inextricably linked,

since the A-estimator depends on these only through the product cS_0. We consider the MAD, and S_n and Q_n of Rousseeuw & Croux (1993) for S_0, and use a grid search for the best choice of c for each of these estimators. For the normal corner, MAD is known to (asymptotically) estimate 0.6745σ and Rousseeuw & Croux report (asymptotic) biases for S_n and Q_n, however these are inappropriate for both the one-wild and slash samples and we lose any intuition on the product cS_0 when we shift focus from the normal corner. In the smoothing algorithm loess (Cleveland et al. 1992), $c = 6$ is the default, and observations approximately four standard deviations from the centre of the data are given zero weight, provided the normal distribution is assumed, i.e. the multiple four is no longer appropriate for non-Gaussian data generating processes.

Table 2 lists results for various A-estimators for both choices of efficiency. Unlike the one-wild efficiencies for the MAD, IQR, S_n and Q_n, the one-wild efficiencies for the A-estimators typically decrease when shifting from efficiency based on (1) to (3); however, efficiencies increase for normal samples, and decrease for slash samples. The three blocks in the table correspond to three alternative choices for S_0: the MAD, S_n and Q_n respectively.

For each auxiliary scale estimator, the choice of scaling constant depends on the evaluation criterion, and in each case, the best scaling constant under Lax's efficiency measure (1) is greater than that using the coefficient of variation-based measure (3). A systematic feature of the results is that the estimators perform best for the normal samples, with average efficiencies close to 90%. In most instances, the triefficiencies are attained at the one-wild corner, and typically, the scaling constant c has been chosen to maximise the one-wild efficiency. The one-wild efficiencies average approximately 80%, indicating the estimators will have at least 80% efficiency for a range of sampling situations. This is an improvement on the triefficiency of Q_n which was approximately

	efficiency using (1)				efficiency using (3)			
biweight-A with:	normal	one-wild	slash	trieff	normal	one-wild	slash	trieff
MAD and $c = 9$	86.2	79.1	88.0	79.1	87.9	79.2	84.2	79.2
MAD and $c = 10$	89.4	79.2	86.8	79.2	90.5	78.6	82.6	78.6
MAD and $c = 11$	91.7	78.2	85.5	78.2	92.5	77.1	81.0	77.1
S_n and $c = 6.5$	86.8	80.8	86.9	80.8	88.1	80.7	82.2	80.7
S_n and $c = 7$	89.0	81.1	85.8	81.1	90.1	80.6	80.8	80.4
S_n and $c = 7.5$	90.8	80.8	84.6	80.8	91.6	79.9	79.5	79.3
Q_n and $c = 10.5$	88.0	82.1	83.9	82.1	88.8	81.7	77.9	77.9
Q_n and $c = 11$	89.4	82.2	82.9	82.1	90.1	81.6	76.9	76.9
Q_n and $c = 11.5$	90.6	82.1	82.0	81.7	91.2	81.2	75.9	75.9

Table 2
Average efficiencies for selected A-estimators, based on 100 realisations of the efficiencies, each estimated from 20,000 samples of size 20. Efficiency is computed using (1) and (3) as indicated.

67%.

Using efficiency based on (1), the best choice of auxiliary estimator is Q_n with a scaling constant of $c = 11$ and a triefficiency of 82.1%, while using efficiency based on (3), the best choice of auxiliary estimator is S_n with a scaling constant of $c = 6.5$ and an efficiency of 80.7%. We note that while these are the best of these estimators under the competing efficiency measures, any of the three auxiliary scale estimators used in the A-estimator with an appropriate scaling constant will give almost the same triefficiency. While statistically significant gains are made using Q_n and $c = 11$ over MAD and $c = 10$ (or indeed MAD and $c = 9$ promoted by Lax), the actual benefits in practice may be limited.

There is conflicting evidence whether using a more triefficient auxiliary estimator of location improves overall triefficiency. Using efficiency based on (1), and replacing the sample median by a one-step M-estimator, and using Q_n and $c = 11$, triefficiency remains constant, although efficiencies for the normal and one-wild corners increase. This behaviour is not completely consistent

with the relative performances of the median and M-estimator documented by Randal & Thomson (2004), where the M-estimator was more efficient for all three corners (although its absolute improvement was smallest for slash). When MAD and $c = 9$ are chosen, a higher slash efficiency allows the one-wild corner to dominate, and triefficiency to increase. Clearly, there are complex interactions between the tuning choices we make, with simulation and trial and error the only way to draw firm conclusions. Optimisation of scaling constants and other choices of M_0 are not pursued here.

4.4 Estimators based on the t-distribution

The t-distribution is one which may be used to motivate a general-purpose scale estimator by regarding the t-distribution as a compromise distribution for the three corners. It is also a choice which, by using the form of the EM algorithm, yields a weighting function which can be used as the basis for a simple one-step estimator defined in (7).

Efficiencies for the t-estimator are shown in Table 3. In each case, the sample median is used as the auxiliary location estimate, and we consider MAD, S_n and Q_n for the auxiliary scale estimator. The results indicate that while the t-estimator has not led to significant gains in efficiency, the simplicity of this estimator implies it may be a useful choice in practice. Unlike the A-estimators using the biweight ψ-function, the t-estimator has a differentiable weight function $\frac{1}{n}(1 + U_i^2)^{-1}$, which will be non-zero (albeit very small for extreme data) for all observations in practice.

Comparison of Tables 2 and 3 indicates that unlike the A-estimators, the t-estimators are very sensitive to the choice of auxiliary scale estimator. In particular, the one-wild efficiencies are much higher when Q_n is used than when either MAD or S_n are used, and again the one-wild corner typically

	efficiency using (1)				efficiency using (3)			
t-estimator with:	normal	one-wild	slash	trieff	normal	one-wild	slash	trieff
MAD and $c = 4$	78.9	69.1	90.8	69.1	79.9	69.4	87.4	69.4
MAD and $c = 4.25$	80.8	69.3	89.7	69.3	81.8	69.3	86.2	69.3
MAD and $c = 4.5$	82.6	69.1	88.5	69.1	83.5	69.0	85.0	69.0
S_n and $c = 2.75$	83.1	76.6	89.8	76.6	83.8	76.7	85.8	76.7
S_n and $c = 3$	85.3	76.6	87.9	76.6	85.9	76.6	83.9	76.6
S_n and $c = 3.25$	87.3	76.2	86.0	76.2	87.7	76.1	82.0	76.1
Q_n and $c = 4$	85.7	81.7	86.2	81.7	86.2	82.0	80.9	80.9
Q_n and $c = 4.25$	86.9	81.8	85.0	81.8	87.4	82.0	79.6	79.6
Q_n and $c = 4.5$	88.1	81.7	83.7	81.7	88.5	81.8	78.4	78.4

Table 3
Average efficiencies for selected t-estimators, defined in (7), based on 100 realisations of the efficiencies, each estimated from 20,000 samples of size 20. Efficiency is computed using (1) and (3) as indicated.

determines the triefficiency. Using Lax's measure of efficiency, the t-estimator with Q_n and $c = 4.25$ is the best estimator of this class, with triefficiency 81.8%, which is marginally smaller than the best biweight A-estimator. Using the efficiency measure favoured by Rousseeuw & Croux (1993), we retain Q_n but reduce the scaling constant to $c = 4$, yielding a triefficiency of 80.9%. This is higher than the best A-estimator, and is the estimator with the highest triefficiency using (3).

5 Summary and conclusions

Results from an extensive simulation study of robust scale estimators have been summarised for selected estimators in Section 4. The analysis focused on two primary issues: the triefficiency of simple estimators, MAD, IQR, S_n and Q_n; and the triefficiency of weighted average-based estimators relying on auxiliary estimates, the A- and t-estimators. Simulation sizes reflect modern computing power, and new theoretical results for Tukey's corners (Randal &

Thomson 2004), and results indicate flaws in the most comprehensive published study of this type, Lax (1985).

We are able to recommend use of the IQR in place of the MAD, but note that further improvement can be attained through use of either S_n or Q_n. These estimators are useful as scale estimators in their own right, but triefficiency gains may be made by using them as auxiliary estimators in a biweight A-estimator, or an estimator motivated by the t-distribution.

Under one of the two efficiency criteria used in this study, the t-estimator with appropriate tuning provides the best triefficiency of the estimators considered, however the more complicated A-estimators achieve triefficiency only slightly less.

We do not claim to have identified the best general purpose scale estimator, however this is the first study to correctly benchmark efficiencies for Tukey's triefficiency. The biweight A-estimators remain strong contenders, however the t-estimators provide a simpler alternative which is competitive under the triefficiency criterion.

References

Andrews, D., Bickel, P., Hampel, F., Huber, P., Rogers, W. & Tukey, J. (1972), *Robust estimates of location: survey and advances*, Princeton University Press.

Cleveland, W. S., Grosse, E. & Shyu, W. M. (1992), *Statistical Models in S*, Wadsworth and Brooks/Cole, chapter 8: Local regression models, pp. 309–376.

Cont, R. (2001), 'Empirical properties of asset returns: stylized facts and statistical issues', *Quantitative Finance* **1**, 223–236.

Dempster, A., Laird, N. & Rubin, D. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society, Series B* **39**, 1–38.

Goodall, C. (2000), *Understanding Robust and Exploratory Data Analysis*, Wiley Classics Library Edition, chapter 11: M-estimators of location: an outline of the theory, pp. 339–403.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, Wiley Series in Probability and Statistics.

Hoaglin, D., Mosteller, F. & Tukey, J., eds (2000), *Understanding Robust and Exploratory Data Analysis*, Wiley Classics Library Edition.

Huber, P. J. (1981), *Robust Statistics*, John Wiley and Sons.

Iglewicz, B. (2000), *Understanding Robust and Exploratory Data Analysis*, Wiley Classics Library Edition, chapter 12: Robust scale estimators and confidence intervals for location, pp. 404–431.

Kafadar, K. (1982), 'A biweight approach to the one-sample problem', *Journal of the American Statistical Association* **77**, 416–424.

Lax, D. A. (1975), An interim report of a Monte Carlo study of robust estimators of width, Technical report 93, series 2, Department of Statistics, Princeton University.

Lax, D. A. (1985), 'Robust estimators of scale: finite-sample performance in long-tailed symmetric distributions', *Journal of the American Statistical Association* **80**, 736–741.

Mertikas, S. P. (1994), 'Description of accuracy using conventional and robust estimates of scale', *Journal of Marine Geodesy* **17**, 251–269.

R Development Core Team (2004), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

Randal, J. A. & Thomson, P. J. (2004), 'Maximum likelihood estimation for Tukey's three corners', *Computational Statistics and Data Analysis* **46**(4), 677–687.

Rousseeuw, P. J. & Croux, C. (1993), 'Alternatives to the median absolute deviation', *Journal of the American Statistical Association* **88**, 1273–1283.

Rydén, T., Teräsvirta, T. & Åsbrink, S. (1998), 'Stylized facts of daily return series and the hidden Markov model', *Journal of Applied Econometrics* **13**, 217–244.

Simon, G. (1976), 'Computer simulation swindles, with applications to estimates of location and dispersion', *Applied Statistics* **25**(3), 266–274.

Yatrakos, Y. G. (1991), 'A note on Tukey's polyefficiency', *Biometrika* **78**(3), 702–703.