A MONTE CARLO STUDY OF ROBUST ESTIMATORS OF LOCATION

by

Raymond J. Carroll

and

Edward J. Wegman¹

Department of Statistics

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #1040

November, 1975

¹ The work of this author was supported by the Air Force Office of Scientific Research under grant number AFOSR-75-2840. The development of many of the computer routines was also supported by AFOSR-75-2849.

ABSTRACT

Andrews et al (1972) carried out an extensive Monte Carlo study of robust estimators of location. Their conclusions were that the hamper and the skipped estimates, as classes, seemed to be preferable to some of the other currently fashionable estimators. The present study extends this work to include estimators not previously examined. The estimators are compared over short-tailed as well as long-tailed alternatives and also over some dependent data generated by first-order autoregressive schemes. The conclusions of the present study are threefold. First, from our limited study, none of the so-called robust estimators are very robust over short-tailed situations. More work seems to be necessary in this situation. Second, none of the estimators perform very well in dependent data situations, particularly when the correlation is large and positive. This seems to be a rather pressing problem. Finally, for long-tailed alternatives, the hamper estimators and Hogg-type adaptive versions of the hampers are the strongest classes. The adaptive hampers neither uniformly outperform nor are they outperformed by the hampers. However, the superiority in terms of maximum relative efficiency goes to the adaptive hampers. That is the adaptive hampers, under their worst performance, are superior to the usual hampers, under their worst performance.

Key words and phrases: M-estimators, hamper estimators, trimmed means, one-step M-estimators, adaptive estimation, rank estimators, robustness, Monte Carlo simulation, sensitivity curves, influence curves, relative efficiency, small sample, long-tails, short-tails, time series, first-order autoregressive, dependent data.

A.M.S. Classification Numbers: Primary 62G35; Secondary 62C05, 62F10, 62G05.

1. Introduction. In 1972, Andrews, Bickel, Hampel, Huber, Rogers and Tukey published a very extensive and informative Monte Carlo study of robust estimation of location in a symmetric probability density. This study involved some 68 estimators of location as well as 14 distinct sampling distributions. Sample sizes used were 5, 10, 20 and 40 although for most of the sampling situations only the sample size of 20 was investigated. For the estimators and sampling situations examined, Andrews et al (1972) provide a very complete and satisfactory picture.

We are motivated, however, to supplement this study for three basic reasons. In Andrews et al (1972), hereafter referred to as *PRS (Princeton Robustness Study)*, short-tailed sampling situations are ruled out of consideration with the statement, "Robustness for short-tailed distributions was thought to be a rather special case, arising in situations that are usually rather easily recognized in practice."¹ Since some of the location estimators were designed to protect against short-tailed alternatives as well as long-tailed possibilities, we feel that examination of long-tailed alternatives alone faults such estimators unfairly, rather akin to discovering a two-sided test is not most powerful against one-sided alternatives. Even ignoring this aspect, the question remains, "What are desirable estimators given short-tailed alternatives?"

A second motivation was provided by the rather dismal situation in the time series context. The sample mean is routinely shown to be a consistent estimator of the "center" of a stationary time series with the clear implication

¹ *PRS*, p. 67.

that a time series is centered by subtracting the sample mean. If the sample mean is unsatisfactory for independent, but non-normal data, how much worse must it be for correlated data?

Our third motivation arises from a personal conviction that adaptive estimators, properly formulated, ought to be very successful. In particular, we observe that because of the contributions of Professors Hampel and Huber to the *PRS*, the so-called Hampel and closely related M-estimators were extensively studied. At least 25 of the 68 estimators studied involve either a Hampel or a M-estimator. The M-estimators and particularly the Hampels perform very well both because they are innately good estimators and because they were fine-tuned. The adaptive estimators, however, were not similarly fine tuned and do not fare as well in the final analysis. To state our conviction succinctly, if Hampels are good, adaptive Hampels should be better.

The paper is divided into six parts. Section 2 lists the estimators studied in this paper and gives a short discussion of each. Section 3 discusses the details of the Monte Carlo aspects of this work while section 4 contains the tables of results. Section 5 presents stylized sensitivity curves for a selection of the estimators and finally section 6 contains our reactions to and conclusions about the results.

2. The Estimators. In choosing which estimators to examine in this study, we were guided by the results of the *PRS*. As standards of comparison, we chose M-estimators and the related Hampels and the trimmed means. Against these standards, we measured various forms of adaptive estimators, one-step rank

estimators and several miscellaneous estimators - including the Hodges-Lehmann, Johns', Gastwirth's, Andrews' and an estimator based on skipping. We follow the routine established in the *PRS* by listing the estimators together with their mnemonic codes in the following table. A short description of the estimators follows the table. We note here that the codes described below agree with those in the *PRS* when an estimator is common to both works.

<u>Number</u>	<u>Code</u>	<u>Short Description</u>
1	5%	5% symmetrically trimmed mean
2	10%	10% symmetrically trimmed mean
3	19%	18.75% symmetrically trimmed mean
4	25%	25% symmetrically trimmed mean (midmean)
5	38%	37.5% symmetrically trimmed mean
6	50%	50% symmetrically trimmed mean (median)
7	M	mean
8	OM	Outer mean, mean of trimmings after 25% symmetrical trimming
9	D10	One-step Huber, k=1.0, start = median
10	D15	One-step Huber, k=1.5, start = median
11	D20	One-step Huber, k=2.0, start = median
12	12A	M-estimate, ψ bends at 1.2, 3.5, 8.0
13	17A	M-estimate, ψ bends at 1.7, 3.4, 8.5
14	21A	M-estimate, ψ bends at 2.1, 4.0, 8.2
15	22A	M-estimate, ψ bends at 2.2, 3.7, 5.9
16	25A	M-estimate, ψ bends at 2.5, 4.5, 9.5
17	ADA	Adaptive M-estimate, ψ bends at ADA, 4.5, 8.0
18	HG1	Hogg-type adaptor using trimmed means 38%, 19%, 11, OM
19	HG2	Hogg-type adaptor using trimmed means 38%, 25%, 10%
20	HG3	Hogg-type adaptor using trimmed means 38%, 10%, 5%
21	HG4	Hogg-type adaptor using trimmed means 33%, 25%, 19%
22	HG5	Hogg-type adaptor using trimmed means 33%, 25%
23	HG6	Hogg-type adaptor using trimmed means 33%, 19%
24	1.81A	Hogg-type adaptor using hampels ADA, 25A, 17A
25	1.81*A	Hogg-type adaptor using hampels 25A, 21A, 12A
26	1.90A	Hogg-type adaptor using hampels ADA, 25A, 17A
27	1.95A	Hogg-type adaptor using hampels ADA, 25A, 17A
28	2.00A	Hogg-type adaptor using hampels 21A, 12A
29	1.81D	Hogg-type adaptor using M-estimators D10, D15, D20
30	1.90D	Hogg-type adaptor using M-estimators D10, D15, D20

(table con't. next page)

(con't.)

<u>Number</u>	<u>Code</u>	<u>Short Description</u>
31	1.95D	Hogg-type adaptor using M-estimators D10, D15, D20
32	1.60D	Hogg-type adaptor using M-estimators D10, D15, D20
33	1.75D	Hogg-type adaptor using M-estimators D10, D15, D20
34	GAS	Gistwirth's estimator
35	AMT	Andrew's estimator
36	H/L	Hodges-Lehmann estimator
37	RN	Normal-scores rank estimator
38	RN50	One-step, normal-scores, rank estimator, $\theta=50\%$
39	RNA	One-step, normal-scores, rank estimator, $\theta=\text{the adaptive estimator}$
40	RU50	One-step, uniform-scores rank estimator, $\theta=50\%$
41	RUA	One-step, uniform-scores rank estimator, $\theta=\text{the adaptive estimator}$
42	J0A	John's adaptive estimator
43	5T4	Multiply-skipped mean, $\max(5k, 2) \leq .6N$ deleted

TABLE 2.1.

A brief description of the 43 estimators of location together with a short mnemonic code for each.

Trimmed Means

It is well-known that the usual arithmetic mean is non-robust in the presence of extreme outliers. A simple scheme for robustifying the mean is to eliminate "extreme" observations. The $\alpha(100)\%$ symmetric trimmed mean discards the $[(N+1)\alpha]$ largest and $[(N+1)\alpha]$ smallest observations from the sample and computes the arithmetic average of the remaining observations. Here $[\cdot]$ is the greatest integer function. For short-tailed situations, the outer mean is sometimes calculated. The outer mean is the mean of the trimmings with $\alpha=.25$. More precisely, if $U(\alpha)$ is the mean of the $[(N+1)\alpha]$ largest observations and $L(\alpha)$ is the mean of the $[(N+1)\alpha]$ smallest observations, then $OM = (U(.25) + L(.25))/2$. In this study, the OM is used as a robust estimator in its own right as well as in HGL. The trimmed-mean estimators are 5%, 10%, 19%, 25%, 38%, 50% and, of course, with 0% trimming, M. Adaptive

estimators based on trimmed means include HG1, HG2, HG3, HG4, HG5 and HG6.

Huber's M-estimators

M-estimators of location are solutions, T , of an equation of the form

$$\sum_{j=1}^n \psi\left(\frac{x_j - T}{s}\right) = 0,$$

where ψ is an odd function and s is a scale estimate which is either estimated independently or from an equation of the form

$$\sum_{j=1}^n \chi\left(\frac{x_j - T}{s}\right) = 0.$$

Here, of course, $x_1, \dots, x_n$ are the observations. Huber (1964) proposed ψ of the form

$$\psi(x; k) = \begin{cases} -k & x < -k \\ x & -k \leq x \leq k \\ k & x > k \end{cases}.$$

Hampel (1974) proposed using ψ as above but with s estimated by

$$\text{med}|x_i - 50\%| / .6754.$$

We refer to this latter quantity as the hamper scale estimator.

The Huber M-estimators examined in this work are one-step estimators (cf. Bickel (1975)). That is, rather than selecting T as the exact root of the equation

$$\sum_{i=1}^n \psi\left(\frac{x_i - T}{s}\right) = 0,$$

we perform one iteration of the Newton-Raphson method using 50% (the median) as the starting value. Thus the estimator T is given by

$$T = (50\%) + s \cdot \frac{n^{-1} \sum_{i=1}^n \psi \left(\frac{x_i - (50\%)}{s} \right)}{n^{-1} \sum_{i=1}^n \psi' \left(\frac{x_i - (50\%)}{s} \right)} .$$

The scale s was chosen according to two algorithms. We investigated the Hampel scale and also the scale $s = 1.0Q/1.1805$. Results based on Hampel scale were similar and generally somewhat superior to those based on the inter-quartile range. We will report only the results based on Hampel scale in section 4. The Huber estimators discussed include D10, D15, and D20 which have $k=1.0, 1.5$ and 2.0 respectively. Closely related are the Hampel-type estimators 12A, 17A, 21A, 22A, 25A and ADA. Also 1.81D, 1.90D, 1.95D, 1.60D and 1.75D are adaptive M-estimators and 2.00A, 1.81A, 1.81*A, 1.90A, 1.95A and 2.20A are adaptive hampels. Finally AMT, the Andrews estimator, is an M-estimate with ψ chosen by

$$\psi(x) = \begin{cases} \sin(x/2.1) & |x| < 2.1\pi \\ 0 & \text{otherwise} \end{cases} .$$

Hampels

The hampel estimators are Huber-type M-estimators with ψ given by

$$\psi(x) = \text{sgn } x \cdot \begin{cases} |x| & 0 \leq |x| < a \\ a & a \leq |x| < b \\ \frac{c-|x|}{c-b} \cdot a & b \leq |x| < c \\ 0 & |x| \geq c \end{cases} .$$

The parameters, a , b and c are given below together with the code symbol.

code	a	b	c
12A	1.2	3.5	8.0
17A	1.7	3.4	8.5
21A	2.1	4.0	8.2
22A	2.2	3.7	5.9
25A	2.5	4.5	9.5
ADA	ADA	4.5	8.0

The parameter, ADA, is chosen as a function of the estimated tail length of the distribution which is estimated by

$$L = \frac{\text{ave}}{|x_i - (50\%)| > s} \left\{ \left\{ \frac{s}{|x_i - (50\%)|} \right\} \right\}.$$

Here s is the Hampel scale estimator. The parameter ADA is defined by

$$ADA = \begin{cases} 1.0 & L \leq .44 \\ (75L - 25)/8 & .44 < L \leq .6 \\ 2.5 & .6 \leq L \end{cases}.$$

All of the hampel estimators are one-step M-estimators computed by the algorithm discussed above. Hampel estimators include 12A, 17A, 21A, 22A, 25A and ADA and adaptive Hampel-type estimators 1.81A, 1.81*A, 1.90A, 1.95A and 2.00A. Related are the Huber M-estimators and the Andrew's estimator already mentioned.

Hogg-Type Adaptors

Hogg (1974) has suggested an adaptation procedure which chooses among 2 or more estimators of the center depending on the value of some statistic which measures the length of the tail. Hogg first suggested the Kurtosis as a measure of tail length. More recent suggestions include

$$Q_1 = \frac{U(.05) - L(.05)}{U(.50) - L(.50)}$$

and

$$Q_2 = \frac{U(.20) - L(.20)}{U(.50) - L(.50)}$$

where U and L are defined as in the section on trimmed means. The exact formulation of the various adaptors is given below.

<u>Code</u>	<u>Formulation</u>
HG1	$T = \begin{cases} 38\% & Q_1 > 3.2 \\ 19\% & 2.6 < Q_1 \leq 3.2 \\ M & 2.6 < Q_1 \leq 2.6 \\ OM & Q_1 \leq 2.0 \end{cases}$
HG2	$T = \begin{cases} 38\% & Q_2 > 1.87 \\ 25\% & 1.81 < Q_2 \leq 1.87 \\ 10\% & Q_2 \leq 1.81 \end{cases}$
HG3	$T = \begin{cases} 38\% & Q_2 > 1.87 \\ 10\% & 1.81 < Q_2 \leq 1.87 \\ 5\% & Q_2 \leq 1.81 \end{cases}$
HG4	$T = \begin{cases} 38\% & Q_2 > 1.80 \\ 25\% & 1.55 < Q_2 \leq 1.80 \\ 19\% & Q_2 \leq 1.55 \end{cases}$
HG5	$T = \begin{cases} 38\% & Q_2 > 2.20 \\ 25\% & Q_2 \leq 2.20 \end{cases}$
HG6	$T = \begin{cases} 38\% & Q_2 > 2.20 \\ 19\% & Q_2 \leq 2.20 \end{cases}$

Using Q_2 as a measure of tail length, several adaptive Hampel estimators were also constructed.

$$1.81A \quad T = \begin{cases} ADA & 1.81 < Q_2 \leq 1.87 \\ 25A & Q_2 \leq 1.81 \\ 17A & Q_2 > 1.87 \end{cases}$$

$$\begin{aligned}
1.81^*A \quad T &= \begin{cases} 21A & 1.81 < Q_2 \leq 1.87 \\ 25A & Q_2 \leq 1.81 \\ 12A & Q_2 > 1.87 \end{cases} \\
1.90A \quad T &= \begin{cases} ADA & 1.90 < Q_2 \leq 2.05 \\ 25A & Q_2 \leq 1.90 \\ 17A & Q_2 > 2.05 \end{cases} \\
1.95A \quad T &= \begin{cases} ADA & 1.95 < Q_2 \leq 2.10 \\ 25A & Q_2 \leq 1.95 \\ 17A & Q_2 > 2.10 \end{cases} \\
2.00A \quad T &= \begin{cases} 21A & Q_2 \leq 2.00 \\ 12A & Q_2 > 2.00 \end{cases} .
\end{aligned}$$

Also included in this study are Hogg-type adaptors using Huber M-estimators.

$$\begin{aligned}
1.81D \quad T &= \begin{cases} D10 & Q_2 > 1.87 \\ D15 & 1.81 < Q_2 \leq 1.87 \\ D20 & Q_2 \leq 1.81 \end{cases} \\
1.90D \quad T &= \begin{cases} D10 & Q_2 > 2.05 \\ D15 & 1.90 < Q_2 \leq 2.05 \\ D20 & Q_2 \leq 1.90 \end{cases} \\
1.95D \quad T &= \begin{cases} D10 & Q_2 > 2.10 \\ D15 & 1.95 < Q_2 \leq 2.10 \\ D20 & Q_2 \leq 1.95 \end{cases} \\
1.60D \quad T &= \begin{cases} D10 & Q_2 > 1.70 \\ D15 & 1.60 < Q_2 \leq 1.70 \\ D20 & Q_2 \leq 1.60 \end{cases} \\
1.75D \quad T &= \begin{cases} D10 & Q_2 > 1.75 \\ D15 & 1.70 < Q_2 \leq 1.75 \\ D20 & Q_2 \leq 1.70 \end{cases} .
\end{aligned}$$

Rank Estimators

In what follows, let $R_i(\theta)$ denote the signed ranks of $x_i - \theta$. That is, we rank $x_1 - \theta, x_2 - \theta, \dots, x_n - \theta$ according to magnitude (but not sign) and $R_i(\theta)$ is the product of $\text{sgn}(x_i - \theta)$ and rank of $(x_i - \theta)$. Further let us define $J(u) = \phi^{-1}\left(\frac{1+u}{2}\right)$, where ϕ is some distribution function. Typically, ϕ is taken to be normal or uniform. The rank statistic with normal scores is taken as the root, T , of the equation

$$J^*(T) - \bar{J} = 0$$

where

$$J^*(T) = n^{-1} \sum^* J(R_i(T)/n+1)$$

and

$$\bar{J} = n^{-1} \sum_{i=1}^n J(i/n+1).$$

The symbol, $\sum^*$, indicates the summation over all positive ranks. Roots were found by the method of bisection using a maximum of twelve iterations. If ϕ is taken as uniform, the resulting rank statistic is asymptotically equivalent to the Hodges-Lehmann estimator - namely the median of all pairwise means of the sample (c.f. Hodges-Lehmann (1963)). This proved to be a very close approximation so that in this study we report only the Hodges-Lehmann estimator and not the uniform-scores rank estimator.

Just as with the Huber-type M-estimators, one may calculate one-step rank estimators based on one iteration of a Newton-Raphson method (Kraft and Van Eeden(1969)). In our previous notation, the one-step rank estimator, T , becomes

$$T = \theta + \frac{J^*(\theta) - \frac{1}{2}J}{B^*}.$$

θ is chosen as an initial robust estimator of location and B^* is an estimator of $B = \int J'(F(x))f^2(x)dx$, where, of course F and f are respectively the distribution and density of the sample.

In this study, we employed a nonparametric estimate of B proposed by Schweder (1974)

$$B^* = \sum_{i,j=1}^n J'(i/n) \epsilon_n^{-1} K\left(\frac{x(i) - x(j)}{\epsilon_n}\right)$$

where $x_{(1)} \leq \dots \leq x_{(n)}$ are the ordered observations. This estimate, B^* , is based on the standard kernel estimator of a probability density function (c.f. Parzen, (1962)) with uniform kernel, K .

We chose $\epsilon_{20} = 1.60$, although T seems relatively insensitive to the choice of ϵ . In the study of one-step rank estimators, we chose ϕ to be uniform on $(-1,1)$, normal, logistic, exponential and triangular. The logistic-scores, exponential-scores and the triangular-scores estimators are not reported since they were typically bettered by one of the others. The initial estimator θ was chosen according to two schemes. The simpler choice was $\theta=50\%$, the median. The more complex choice was an adaptive estimator given by

$$\theta = \begin{cases} 21A & Q_2 \leq 2.25 \\ 50\% & Q_2 > 2.25. \end{cases}$$

Estimators based on ranks include H/L , RN , $RN50$, RNA , $RU50$ and RUA .

Other estimators

Johns (1974) provides an adaptive estimator which can be written as a linear combination of two trimmed means

$$T = c_1 T_1 + c_2 T_2$$

where

$$T_1 = \frac{1}{2s} \sum_{j=r+1}^{r+s} [x_{(j)} + x_{(n-j+1)}]$$

and

$$T_2 = \frac{1}{n-2r-2s} \sum_{j=r+s+1}^{n/2} [x_{(j)} + x_{(n-j+1)}]$$

with c_1 and c_2 determined from the sample by

$$c_1 + c_2 = 1$$

and

$$c_1 \propto \frac{1}{D_1^2} \left(\frac{2s(n+r+2s)}{(2r+s)(n-2r)} \right) - \frac{2}{D_1 D_2} \frac{n-2r-2s}{n-2r}$$

$$c_2 \propto \frac{1}{D_2^2} \left(\frac{2(n-2r-2s)}{n-2r} \right) - \frac{4}{D_1 D_2} \left(\frac{s}{n-2r} \right).$$

The D_i are proportional to the lengths of the segments of the sample entering in the T_i . Values of r and s for $n=20$ were $r=1$ and $s=4$. The program to compute this estimator was adapted directly from *PRS*. For further details see Johns (1974). Johns' estimator is coded as *JØH*.

A skipping procedure is used to identify and delete extreme points. For $n=20$, let

$$h_1 = \frac{1}{2}(x_{(5)} + x_{(6)}) \quad \text{and} \quad h_2 = \frac{1}{2}(x_{(16)} + x_{(15)}) \quad \text{and let}$$

$t_1 = h_1 - 1.5(h_2 - h_1)$ and $t_2 = h_2 + 1.5(h_2 - h_1)$. A singly skipped estimate deletes all observations smaller than t_1 and larger than t_2 . If a single skipping deletes a total of k points, the estimator *ST4* deletes a further ℓ points

from each end where

$$k = \min(\max(1, 2k), \frac{.6n-k}{2}),$$

and the mean of the remaining sample is used as the estimator. This estimator was formulated in the *PRS*.

Finally, we included an estimator proposed by Gastwirth (1966), defined by

$$T = .3x_{([\frac{n}{3}+1])} + .4(50\%) + .3x_{(n-[\frac{n}{3}])}.$$

3. Monte Carlo Details. In addition to the long-tailed sampling situations investigated in the *PRS*, we also investigated short-tailed situations and situations involving dependent data. Pseudo-random numbers were generated by the familiar multiplicative congruential generator. (See the *PRS* for description and cautions in the use of these generators). The output, scaled between 0 and 1, form the basis for all subsequent calculations since they approximate a uniformly distributed sample. Normal pseudo-random deviates were generated according to the well-known Box-Muller transform² and Cauchy pseudo-random deviates were generated by the arctangent transform. These basic three distributions, or mixtures thereof, form the basis of all results reported in this work.

The basic sampling situations are summarized in Table 3.1.

² This transform is widely attributed to Box and Muller who published their work in 1959. The result is, however, pointed out by Paley and Wiener in their 1934 book (pages 146-147).

Table 3.1.

Sampling situations for independent observations. All had sample size 20.

N/U represents a variate formed by dividing a Normal (0,1) variate by a variate uniform on the interval (0,1).

<u>Distributions</u>	<u>Tag</u>
Uniform, mean 0, variance 1	U
97% Uniform (0,1) + 3% Normal (0,1) contamination	U + .03N(0,1)
90% Uniform (0,1) + 10% Normal (0,1) contamination	U + .10N(0,1)
50% Uniform (0,1) + 50% Normal (0,1) contamination	U + .50N(0,1)
Normal, mean 0, variance 1	N(0,1) = N
99% Normal (0,1) + 1% Normal (0,9) contamination	N + .01N(0,9)
97.5% Normal (0,1) + 2.5% Normal (0,9) contamination	N + .025N(0,9)
95% Normal (0,1) + 5% Normal (0,9) contamination	N + .05N(0,9)
90% Normal (0,1) + 10% Normal (0,9) contamination	N + .10N(0,9)
85% Normal (0,1) + 15% Normal (0,9) contamination	N + .15N(0,9)
75% Normal (0,1) + 25% Normal (0,9) contamination	N + .25N(0,9)
95% Normal (0,1) + 5% Normal (0,100) contamination	N + .05N(0,100)
90% Normal (0,1) + 10% Normal (0,100) contamination	N + .10N(0,100)
75% Normal (0,1) + 25% Normal (0,100) contamination	N + .25N(0,100)
97% Normal (0,1) + 3% Cauchy contamination	N + .03C
90% Normal (0,1) + 10% Cauchy contamination	N + .10C
50% Normal (0,1) + 50% Cauchy contamination	N + .50C
Cauchy, median = 0	C
90% Normal (0,1) + 10% N/U	N + .10N/U
75% Normal (0,1) + 25% N/U	N + .25N/U

In addition to the sets of independent observations discussed above, we generated observations according to a first order autoregressive scheme, $x_j = \rho x_{j-1} + \varepsilon_j$, $j = 1, 2, \dots, 20$. The shock, ε_j , was chosen according to either N, U or C and the correlation coefficient, ρ , as either .2, .5 or .9. The initial value, x_0 , was chosen as 0. All of the 43 estimators discussed in section 2 were investigated in the 20 independent sampling situations. Only the non-adaptive estimators were studied in the 9 time-series sampling schemes.

The results we report are based on 1000 Monte Carlo replications. The sample variances of the estimators were calculated and then scaled by a factor of 20 (the sample size) to make the results comparable to those in the *PRS*. Assuming approximate normality for the estimators, the variance of $20 \times$ sample variance would be about $.7992\sigma_T^4$ where σ_T^2 is the true variance of the estimator, T . For example if $T = \bar{X}$, then σ_T^2 is $\frac{1}{20}$ and the variance of 20 times the sample variance is .001998. Thus one may expect about one decimal place of accuracy with correspondingly less significance as the value σ_T^2 climbs. We report two decimal places, in most cases, because all estimators were calculated over the same samples. Without attempting inferences outside these samples, the extra decimal place(s) are meaningful.

We also report pseudo-variances as discussed in the *PRS*. For long-tailed distributions, the variance is essentially determined by a few observations in the extreme tails. Thus if the distribution of the estimator, T , is itself long-tailed, the sample variance of T would itself not be a robust estimate of the scale of T . As in the *PRS*, we compute the 2.5%-ile point, square this quantity and divide by the normalizing constant $(1.96)^2$. This pseudo-variance is further rescaled by multiplying by 20 (the sample size) to yield quantities comparable to those in the *PRS*.

4. The Results: The main results of the Monte Carlo study are summarized in 3 tables. Table 4.1 is a table of sample variances (over the 1000 Monte Carlo replications) for the 20 independent sampling schemes and 43 estimators of location. Table 4.2 is a table of pseudo-variances (based on the 2.5%-ile

point) for the same 20 independent sampling schemes and 43 estimators.

Table 4.3 is a table of biases and variances for the 9 sampling schemes involving first order autoregressive variates. None of the Hogg-type adaptive estimators are listed for the two-fold reason that the choice of Q_1 and Q_2 is predicated on independent observations and that the basic estimators - trimmed means, hampels, etc. - perform poorly. Some preliminary calculations indicate that Hogg-type adaptors perform similarly under the time series alternatives.

We withhold discussion of these results to section 6.

NUMBER	CODE	U	U+ .03N(0,1)	U+ .10N(0,1)	U+ .50N(0,1)	N(0,1)	N+ .01N(0,9)	N+ .025N(0,9)	N+ .05N(0,9)	N+ .10N(0,9)	N+ .15N(0,9)
1	5%	1.18	1.15	1.21	1.02	1.08	1.03	1.11	1.25	1.42	1.63
2	10%	1.34	1.32	1.37	1.13	1.11	1.05	1.13	1.24	1.32	1.51
3	19%	1.50	1.49	1.53	1.22	1.14	1.09	1.16	1.26	1.31	1.49
4	25%	1.82	1.80	1.83	1.41	1.23	1.18	1.23	1.35	1.43	1.56
5	38%	2.30	2.08	2.41	1.61	1.36	1.32	1.36	1.50	1.56	1.66
6	50%	2.48	2.43	2.51	1.82	1.55	1.49	1.54	1.67	1.72	1.84
7	M	1.01	.98	1.05	.94	1.05	1.04	1.24	1.45	1.82	2.09
8	QM	.61	.60	.67	.84	1.25	1.31	1.87	2.27	3.37	4.08
9	D10	1.95	2.10	1.97	1.47	1.16	1.15	1.25	1.38	1.46	1.60
10	D15	1.55	1.66	1.58	1.27	1.07	1.06	1.19	1.30	1.40	1.51
11	D20	1.23	1.32	1.29	1.11	1.04	1.01	1.15	1.25	1.38	1.49
12	12A	1.82	1.97	1.85	1.42	1.14	1.13	1.26	1.36	1.46	1.59
13	17A	1.41	1.52	1.46	1.22	1.06	1.05	1.18	1.27	1.39	1.50
14	21A	1.20	1.30	1.26	1.11	1.05	1.01	1.16	1.24	1.38	1.49
15	22A	1.17	1.29	1.25	1.13	1.08	1.03	1.20	1.26	1.39	1.52
16	25A	1.08	1.17	1.14	1.04	1.03	.98	1.13	1.23	1.38	1.51
17	ADA	1.19	1.31	1.28	1.11	1.06	1.02	1.17	1.28	1.43	1.55
18	HG1	.80	.80	.88	.96	1.11	1.06	1.19	1.28	1.41	1.60
19	HG2	1.39	1.37	1.41	1.19	1.16	1.12	1.21	1.31	1.37	1.58
20	HG3	1.21	1.20	1.25	1.08	1.12	1.09	1.17	1.28	1.35	1.57
21	HG4	1.77	1.72	1.78	1.40	1.25	1.22	1.28	1.40	1.47	1.60
22	HG5	1.82	1.80	1.83	1.41	1.23	1.18	1.23	1.35	1.43	1.56
23	HG6	1.50	1.49	1.53	1.22	1.14	1.09	1.16	1.26	1.31	1.49
24	1.81A	1.10	1.20	1.17	1.07	1.05	1.00	1.16	1.26	1.40	1.52
25	1.81*A	1.11	1.21	1.17	1.08	1.05	1.01	1.18	1.26	1.41	1.54
26	1.90A	1.09	1.18	1.15	1.05	1.03	.99	1.15	1.25	1.40	1.51
27	1.95A	1.08	1.18	1.14	1.04	1.03	.98	1.14	1.24	1.39	1.52
28	2.00A	1.20	1.30	1.26	1.11	1.05	1.02	1.16	1.24	1.39	1.50
29	1.81D	1.25	1.34	1.31	1.14	1.04	1.04	1.18	1.27	1.40	1.53
30	1.90D	1.24	1.32	1.29	1.12	1.04	1.02	1.15	1.25	1.38	1.49
31	1.95D	1.23	1.32	1.29	1.11	1.04	1.02	1.15	1.25	1.39	1.50
32	1.60D	1.49	1.58	1.54	1.30	1.10	1.12	1.24	1.34	1.46	1.56
33	1.75D	1.34	1.44	1.41	1.22	1.08	1.08	1.22	1.32	1.45	1.56
34	GAS	1.90	1.96	1.93	1.44	1.28	1.19	1.25	1.37	1.40	1.59
35	AMT	1.22	1.31	1.27	1.12	1.12	1.04	1.16	1.25	1.32	1.50
36	H/L	1.21	1.28	1.25	1.09	1.03	1.01	1.14	1.26	1.38	1.51
37	RN	.89	.94	.97	.94	1.01	.98	1.14	1.26	1.45	1.62
38	RNS0	1.41	1.49	1.45	1.18	1.07	1.15	1.22	1.34	1.46	1.58
39	RNA	1.01	1.09	1.08	1.01	1.02	.98	1.13	1.23	1.38	1.50
40	RUS0	2.47	2.60	2.50	1.82	1.40	1.49	1.54	1.67	1.70	1.84
41	RUA	1.19	1.30	1.26	1.11	1.05	1.01	1.16	1.24	1.38	1.49
42	JØH	.88	.88	.93	.95	1.06	1.02	1.20	1.25	1.47	1.59
43	ST4	1.22	1.28	1.27	1.10	1.08	1.05	1.20	1.29	1.47	1.57

VARIANCES

Table 4.1: Variances of 43 estimators of location based on 1000 Monte Carlo replications.

NUMBER	CODE	N+ .25N(0.9)	N+ .05N(0.100)	N+ .10N(0.100)	N+ .25N(0.100)	N+ .03C	N+ .10C	N+ .50C	C	N+ .10N/U	N+ .25N/U
1	5%	2.34	1.68	4.35	16.0	1.10	1.40	8.53	27.2	1.60	3.67
2	10%	2.02	1.22	2.16	9.12	1.13	1.22	2.70	8.34	1.23	1.80
3	19%	1.89	1.21	1.61	5.32	1.17	1.23	1.96	4.99	1.24	1.64
4	25%	1.84	1.29	1.57	2.78	1.25	1.30	1.74	3.24	1.31	1.67
5	38%	1.90	1.41	1.72	2.49	1.37	1.43	1.84	2.77	1.43	1.79
6	50%	2.08	1.58	1.94	2.56	1.55	1.59	2.06	2.83	1.61	2.00
7	M	2.96	5.45	11.5	26.0	72.9	381.	***	***	121.	***
8	OM	5.80	17.3	38.1	85.3	289.	***	***	***	475.	***
9	D10	1.86	1.32	1.59	2.68	1.27	1.33	1.76	3.10	1.34	1.70
10	D15	1.87	1.23	1.54	3.12	1.19	1.25	1.82	3.63	1.26	1.64
11	D20	1.95	1.20	1.59	3.80	1.14	1.23	1.96	4.37	1.23	1.67
12	12A	1.87	1.27	1.44	2.48	1.26	1.30	1.68	2.76	1.29	1.58
13	17A	1.89	1.18	1.37	2.82	1.17	1.22	1.72	3.21	1.21	1.51
14	21A	1.96	1.16	1.36	3.11	1.14	1.21	1.73	3.58	1.19	1.50
15	22A	2.01	1.18	1.36	2.97	1.17	1.24	1.78	3.46	1.21	1.51
16	25A	2.07	1.15	1.39	3.60	1.11	1.18	1.89	3.93	1.18	1.53
17	ADA	1.98	1.19	1.42	2.67	1.15	1.24	1.74	3.01	1.23	1.57
18	HG1	2.14	1.28	1.74	4.26	1.16	1.30	1.88	3.30	1.31	1.69
19	HG2	1.93	1.29	1.66	2.80	1.20	1.30	1.84	2.83	1.30	1.66
20	HG3	2.02	1.28	1.66	3.00	1.16	1.28	1.87	2.91	1.30	1.66
21	HG4	1.88	1.34	1.69	2.60	1.28	1.37	1.83	2.78	1.35	1.72
22	HG5	1.84	1.29	1.59	2.78	1.26	1.31	1.79	3.01	1.31	1.69
23	HG6	1.89	1.21	1.60	4.73	1.18	1.24	1.89	3.68	1.25	1.65
24	1.81A	1.99	1.16	1.37	2.90	1.13	1.21	1.73	3.26	1.19	1.52
25	1.81*A	1.97	1.19	1.42	2.59	1.15	1.24	1.69	2.84	1.22	1.54
26	1.90A	2.01	1.17	1.38	2.86	1.12	1.21	1.75	3.30	1.20	1.52
27	1.95A	2.01	1.16	1.38	3.01	1.12	1.20	1.74	3.31	1.20	1.52
28	2.00A	1.94	1.18	1.42	2.77	1.15	1.23	1.74	2.98	1.21	1.49
29	1.81D	1.91	1.24	1.56	2.88	1.17	1.26	1.78	3.16	1.26	1.63
30	1.90D	1.94	1.23	1.56	3.06	1.15	1.24	1.80	3.31	1.24	1.62
31	1.95D	1.95	1.21	1.56	3.21	1.14	1.24	1.82	3.44	1.24	1.62
32	1.60D	1.87	1.30	1.58	2.77	1.24	1.31	1.76	3.11	1.31	1.67
33	1.75D	1.89	1.28	1.57	2.83	1.21	1.29	1.77	3.11	1.28	1.65
34	GAS	1.87	1.34	1.60	2.67	1.28	1.34	1.78	3.19	1.35	1.72
35	AMT	1.97	1.16	1.36	2.99	1.14	1.21	1.73	3.44	1.19	1.52
36	H/L	1.97	1.22	1.65	4.04	1.14	1.24	2.01	4.57	1.25	1.73
37	RN	2.24	1.28	1.95	5.93	1.12	1.28	2.46	6.76	1.31	1.93
38	RN50	1.97	1.31	1.76	2.90	1.21	1.32	2.05	3.41	1.32	1.81
39	RVA	2.01	1.17	1.48	3.33	1.13	1.23	1.96	3.84	1.20	1.67
40	RU50	2.08	1.57	1.94	2.56	1.55	1.59	2.06	2.83	1.61	2.00
41	RUA	1.95	1.17	1.39	3.13	1.16	1.23	1.91	1.41	1.20	1.58
42	JPH	2.19	1.24	1.52	4.38	1.16	1.27	1.73	3.25	1.28	1.62
43	ST4	1.98	1.24	1.42	3.73	1.20	1.25	1.66	3.00	1.26	1.54

VARIANCES

Table 4.1. (continued): Variances of 43 estimators of location based on 1000 Monte Carlo replications.

NUMBER	CODE	U	U+ .03N(0,1)	U+ .10N(0,1)	U+ .50N(0,1)	N(0,1)	N+ .01N(0,9)	N+ .025N(0,9)	N+ .05N(0,9)	N+ .10N(0,9)	N+ .15N(0,9)
1	5%	1.13	1.09	1.19	1.18	.99	1.06	1.20	1.19	1.51	1.60
2	10%	1.33	1.27	1.31	1.25	1.02	1.07	1.20	1.12	1.46	1.43
3	19%	1.55	1.38	1.48	1.36	1.00	1.13	1.31	1.16	1.40	1.39
4	25%	1.83	1.69	1.89	1.40	1.09	1.25	1.33	1.32	1.45	1.49
5	38%	2.23	2.05	2.25	1.54	1.19	1.41	1.53	1.30	1.42	1.75
6	50%	2.40	2.36	2.67	1.86	1.52	1.58	1.51	1.63	1.73	2.04
7	M	.94	.96	1.02	1.06	.98	1.07	1.24	1.31	2.13	2.26
8	OM	.63	.65	.73	.89	1.22	1.23	2.09	2.19	4.25	3.85
9	D10	1.94	2.10	2.07	1.43	1.12	.92	1.34	1.27	1.43	1.60
10	D15	1.56	1.63	1.68	1.27	1.09	.96	1.27	1.17	1.45	1.46
11	D20	1.26	1.36	1.28	1.20	1.07	.94	1.20	1.17	1.35	1.48
12	12A	1.82	1.82	1.96	1.42	1.17	.94	1.33	1.24	1.50	1.66
13	17A	1.41	1.51	1.55	1.28	1.07	1.01	1.27	1.13	1.42	1.43
14	21A	1.26	1.28	1.25	1.21	1.06	.94	1.24	1.18	1.37	1.38
15	22A	1.22	1.28	1.22	1.21	1.07	.98	1.32	1.17	1.37	1.41
16	25A	1.15	1.18	1.12	1.20	1.04	.95	1.17	1.19	1.42	1.44
17	ADA	1.20	1.30	1.46	1.19	1.05	.97	1.25	1.28	1.46	1.55
18	HG1	.83	.84	.90	1.06	1.00	1.05	1.26	1.25	1.49	1.81
19	HG2	1.33	1.35	1.47	1.36	1.04	1.15	1.32	1.24	1.40	1.67
20	HG3	1.23	1.12	1.24	1.20	.98	1.18	1.25	1.18	1.41	1.67
21	HG4	1.82	1.67	1.84	1.40	1.14	1.35	1.36	1.35	1.30	1.73
22	HG5	1.83	1.68	1.89	1.40	1.09	1.25	1.33	1.32	1.45	1.49
23	HG6	1.55	1.38	1.48	1.36	1.00	1.13	1.31	1.16	1.39	1.38
24	1.81A	1.18	1.20	1.12	1.19	1.06	.96	1.25	1.18	1.44	1.51
25	1.81*A	1.15	1.20	1.12	1.20	1.07	.96	1.25	1.19	1.48	1.53
26	1.90A	1.15	1.19	1.12	1.20	1.04	.95	1.25	1.22	1.44	1.51
27	1.95A	1.15	1.18	1.12	1.20	1.04	.96	1.20	1.19	1.43	1.47
28	2.00A	1.26	1.28	1.25	1.21	1.06	.94	1.24	1.18	1.37	1.41
29	1.81D	1.52	1.57	1.61	1.39	1.11	.97	1.35	1.22	1.41	1.55
30	1.90D	1.26	1.36	1.28	1.20	1.07	.94	1.20	1.18	1.40	1.48
31	1.95D	1.26	1.36	1.28	1.20	1.07	.94	1.20	1.17	1.43	1.49
32	1.60D	1.52	1.57	1.61	1.39	1.11	.97	1.35	1.22	1.41	1.55
33	1.75D	1.41	1.42	1.46	1.26	1.10	.92	1.29	1.19	1.40	1.59
34	GAS	1.97	1.98	2.03	1.40	1.02	1.13	1.33	1.35	1.48	1.62
35	AMT	1.25	1.35	1.27	1.21	.93	.99	1.24	1.20	1.41	1.35
36	H/L	1.31	1.36	1.24	1.19	1.00	.90	1.18	1.11	1.37	1.53
37	RN	1.00	1.09	1.02	1.04	1.00	.92	1.20	1.20	1.49	1.59
38	RN50	1.41	1.62	1.57	1.23	1.05	.91	1.35	1.30	1.47	1.72
39	RNA	1.14	1.14	1.13	1.13	1.05	.93	1.22	1.17	1.36	1.48
40	RU50	2.39	2.90	2.67	1.86	1.20	1.16	1.51	1.63	1.75	2.04
41	RUA	1.25	1.28	1.25	1.21	1.06	.94	1.24	1.18	1.37	1.38
42	JØH	.91	.89	1.02	1.08	.99	1.03	1.34	1.14	1.55	1.78
43	ST4	1.21	1.40	1.24	1.20	1.10	.98	1.23	1.19	1.47	1.62

PSEUDO-VARIANCES

Table 4.2.: Pseudo-variances of 43 estimators of location based on the 2.5%-ile and 1000 Monte Carlo replications.

NUMBER	CODE	N+ .25N(0,9)	N+ .05N(0,100)	N+ .10N(0,100)	N+ .25N(0,100)	N+ .03C	N+ .10C	N+ .50C	C	N+ .10N/U	N+ .25N/U
1	5%	2.55	1.82	4.62	16.2	1.24	1.59	7.28	27.8	1.28	3.56
2	10%	2.28	1.35	2.34	10.5	1.33	1.25	2.97	9.14	1.18	1.95
3	19%	1.98	1.31	1.91	5.60	1.41	1.31	2.28	5.02	1.22	1.72
4	25%	1.88	1.36	1.59	2.35	1.36	1.31	1.92	3.49	1.28	1.83
5	38%	1.89	1.48	1.82	2.09	1.45	1.39	1.94	2.66	1.33	1.94
6	50%	2.12	1.61	2.08	2.29	1.56	1.60	2.16	2.59	1.42	1.72
7	M	3.15	6.85	12.5	29.0	1.85	9.01	159.	680.	22.0	65.0
8	OM	6.06	24.7	41.6	93.9	3.06	29.5	628.	***	81.0	264.
9	D10	1.86	1.38	1.63	2.12	1.42	1.34	1.98	3.28	1.27	1.90
10	D15	1.92	1.29	1.73	2.63	1.44	1.35	2.28	3.74	1.21	1.84
11	D20	1.96	1.31	1.66	3.43	1.30	1.30	2.48	4.42	1.21	1.78
12	12A	1.96	1.31	1.48	1.82	1.38	1.36	1.59	2.54	1.32	1.85
13	17A	1.96	1.30	1.46	2.22	1.33	1.26	1.72	3.11	1.19	1.80
14	21A	1.92	1.27	1.39	2.47	1.21	1.25	1.77	3.36	1.12	1.89
15	22A	2.11	1.30	1.44	2.32	1.17	1.26	1.68	3.06	1.16	1.80
16	25A	2.09	1.31	1.47	3.03	1.26	1.20	1.85	3.53	1.09	1.86
17	ADA	1.98	1.26	1.43	2.00	1.26	1.23	1.77	2.57	1.17	1.93
18	HG1	2.12	1.52	1.86	2.98	1.32	1.27	1.99	3.02	1.25	1.81
19	HG2	1.93	1.43	1.83	2.12	1.31	1.25	1.94	2.59	1.33	1.87
20	HG3	2.25	1.48	1.84	2.12	1.31	1.26	1.98	2.77	1.33	1.81
21	HG4	1.87	1.48	1.82	2.09	1.42	1.35	1.94	2.62	1.32	1.90
22	HG5	1.88	1.36	1.66	2.33	1.36	1.28	2.07	3.08	1.29	1.83
23	HG6	1.98	1.51	1.82	4.86	1.41	1.27	2.32	3.66	1.22	1.76
24	1.81A	1.95	1.27	1.50	2.21	1.20	1.25	1.72	3.11	1.12	1.76
25	1.81*A	1.96	1.31	1.49	1.83	1.22	1.34	1.59	2.69	1.24	1.82
26	1.90A	2.01	1.31	1.46	2.07	1.28	1.22	1.75	3.27	1.17	1.77
27	1.95A	1.98	1.31	1.41	2.14	1.28	1.22	1.71	3.27	1.14	1.75
28	2.00A	1.94	1.22	1.48	2.00	1.21	1.25	1.83	2.82	1.18	1.82
29	1.81D	1.82	1.38	1.72	2.13	1.36	1.35	1.98	3.28	1.22	1.86
30	1.90D	1.97	1.34	1.72	2.48	1.37	1.30	2.03	3.38	1.21	1.78
31	1.95D	1.96	1.31	1.72	2.56	1.37	1.31	2.28	3.51	1.21	1.78
32	1.60D	1.82	1.38	1.72	2.13	1.36	1.35	1.98	3.28	1.22	1.86
33	1.75D	1.82	1.36	1.72	2.13	1.35	1.30	1.98	3.28	1.22	1.86
34	GAS	2.01	1.42	1.69	2.23	1.42	1.32	1.95	3.31	1.25	1.94
35	AMT	2.01	1.27	1.41	2.43	1.22	1.24	1.88	2.96	1.12	1.91
36	H/L	2.05	1.31	1.66	3.51	1.28	1.28	2.41	4.64	1.19	1.89
37	RN	2.40	1.38	2.12	5.50	1.24	1.28	3.06	7.07	1.27	2.01
38	RN50	2.17	1.45	1.67	2.53	1.36	1.25	2.21	3.38	1.23	1.96
39	RNA	2.20	1.30	1.55	2.84	1.18	1.24	2.31	3.79	1.15	1.78
40	RU50	2.12	1.61	2.07	2.29	1.56	1.60	2.16	2.59	1.42	2.27
41	RUA	1.99	1.29	1.38	2.47	1.21	1.28	1.99	3.36	1.14	1.93
42	JPH	2.27	1.43	1.57	2.46	1.31	1.24	2.03	3.15	1.21	1.80
43	ST4	1.92	1.37	1.48	2.59	1.22	1.25	1.92	2.70	1.28	1.82

PSEUDO-VARIANCES

Table 4.2. (continued): Pseudo-variances of 43 estimators of location based on the 2.5%-ile and 1000 Monte Carlo replications.

NUMBER	CODE	DISTRIBUTION		NORMAL		NORMAL		NORMAL		CAUCHY		CAUCHY		CAUCHY		UNIFORM		UNIFORM		UNIFORM	
		p		.2		.5		.9		.2		.5		.9		.2		.5		.9	
1	5%		BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	
2	10%																				
3	19%																				
4	25%																				
5	38%																				
6	50%																				
7	M																				
8	CM																				
9	D10																				
10	D15																				
11	D20																				
12	12A																				
13	17A																				
14	21A																				
15	22A																				
16	25A																				
17	ADA																				
34	CAS																				
35	AMT																				
36	H/L																				
39	RNA																				
41	RUA																				
42	JOM																				
45	ST4																				

Table 4.3.: Variances for 24 estimators of location based on 1000 Monte Carlo replications with first order autoregressive data. Note that *** means a value larger than 9999.99.

5. Sensitivity Curves. Tukey (1970) proposes the use of sensitivity curves to study the finite sample behavior of estimators. These sensitivity curves are similar to the influence curves proposed by Hampel (1968). In the *PRS*, stylized sensitivity curves are computed by the following algorithm:

- i. Generate the 19 expected order statistics from a normal sample of size 19
- ii. To these 19 values add a moving point x and compute the estimator, $T(x)$.
- iii. Plot $20T(x)$ as a function of x .

Stylized sensitivity curves are produced in the *PRS* for a variety of estimators including trimmed means, M -estimators, hampels, skipped estimators and so on. We reproduce some additional sensitivity curves.

Again we defer discussion until section 6.

FIGURE 5.1. Sensitivity curve for 10% symmetrically trimmed mean, 10%.

FIGURE 5.2. Sensitivity curve for mean, M .

FIGURE 5.3. Sensitivity curve for M estimate, $k=1.5$, $D15$.

FIGURE 5.4. Sensitivity curve for 1-step hampel estimate, 25A.

FIGURE 5.5. Sensitivity curve for mildly adaptive hamper estimator, ADA.

FIGURE 5.6. Sensitivity curve for the outer-mean, OM.

FIGURE 5.7. Sensitivity curve for Hogg adapter using 0M, M and trimmed means, Hg1. Notice that because of the "regular" spacing of the quantiles, this stylized curve does not reflect the adaption to 0M.

FIGURE 5.8. Sensitivity curve for Hogg-type adaptor based on trimmed means, HG3.

FIGURE 5.10. Sensitivity curve for a Hogg-type adaptor using hampels, 1.81^*A .

FIGURE 5.9. Sensitivity curve for Hogg-type adaptor using hampels, 1.95A.

FIGURE 5.11. Sensitivity for 1-step rank estimator with uniform scores,
start = adaptor, RUA.

FIGURE 5.12. Sensitivity curve for 1-step, rank estimator with uniform scores,
start = 50%, RU50%.

FIGURE 5.13. Sensitivity curve for normal scores rank estimator, R_N . The estimator was computed by a 12 step method of bisection using the median as initial estimate. Use of other starting estimators will give slightly different sensitivity curve.

6. Discussion and Conclusions. Tables 4.1 and 4.2 contain many estimators with fairly comparable variance or pseudo-variance. In order to make the better estimators more apparent, we have constructed an additional table, Table 6.1. Let T_i , $i = 1, \dots, 1000$ represent the 1000 observations of an estimator of location, and let

$$s_T^2 = \frac{1}{1000} \sum_{i=1}^{1000} (T_i - \bar{T})^2$$

be the sample variance. Assuming the T_i 's are approximately normal, then $\frac{1000s_T^2}{\sigma_T^2}$ is distributed approximately as a χ^2 with 999 degrees of freedom.

Accordingly, the variance of s_T^2 is $.001998\sigma_T^4$ and the variance of $20s_T^2$ is $.7992\sigma_T^4$ as pointed out earlier. An estimate of the standard deviation of s_T^2 can be found by calculating $\sqrt{.001998} s_T^2 = .0447s_T^2$. Table 6.1 was constructed by calculating the estimated standard deviation for the estimator with the smallest variance. The variance of any estimator, whose estimated variance was no more than one standard deviation from the minimal variance, was replaced with a plus sign while the variance of any estimator, whose estimated variance was more than 5 standard deviations from the minimal variance, was replaced with a minus sign. All others were left blank. The resultant picture, Table 6.1, leaves a clear impression of the consistently stronger estimators.

In terms of the 4 light-tailed alternatives the estimator of choice appears to be the outer mean. From Tables 4.1 and 4.2 we observe that HG1, RN and JØH also perform creditably but significantly more poorly than OM. The pseudo-variance also suggests that the mean, M , is not too bad.

We feel that the uniform is a highly artificial (notional in *PRS*) situation. The contamination of a normal by a distribution whose support is a bounded interval seems less so. While this very small selection of short-tailed alternatives is inadequate for sweeping commitments to certain types of estimators, it is very suggestive of what is reasonable.

NUMBER	CODE	U	U+ .03N(0,1)	U+ .10N(0,1)	U+ .50N(0,1)	N(0,1)	N+ .01N(0,9)	N+ .025N(0,9)	N+ .05N(0,9)	N+ .10N(0,9)	N+ .15N(0,9)	N+ .25N(0,9)	N+ .05N(0,100)	N+ .10N(0,100)	N+ .25N(0,100)	N+ .03C	N+ .10C	N+ .50C	C	N+ .10N/U	N+ .25N/U
1	5%	-	-	-	-	-	-	+	+	+	-	-	-	-	-	+	-	-	-	+	-
2	10%	-	-	-	-	-	-	+	+	+	+	+	-	-	-	+	+	-	-	+	-
3	19%	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	+	-	-	-	-
4	25%	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-
5	38%	-	-	-	-	-	-	-	-	-	+	+	-	-	+	-	-	-	-	-	-
6	50%	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	-
7	M	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	OM	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	D10	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
10	D15	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	+	+	-	+	-
11	D20	-	-	-	-	+	+	+	+	+	+	+	+	-	-	+	+	+	-	+	-
12	12A	-	-	-	-	-	-	-	+	+	+	+	-	+	+	-	+	+	+	+	+
13	17A	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
14	21A	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
15	22A	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
16	25A	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
17	ADA	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
18	HG1	-	-	-	-	-	-	-	+	+	+	+	-	-	-	+	-	-	+	+	+
19	HG2	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	+	+	+
20	HG3	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	+	+
21	HG4	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	+	+	+
22	HG5	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	+	-	+	+
23	HG6	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	+	+
24	1.81A	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
25	1.81*A	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
26	1.90A	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
27	1.95A	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
28	2.00A	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
29	1.81D	-	-	-	-	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+
30	1.90D	-	-	-	-	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+
31	1.95D	-	-	-	-	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+
32	1.60D	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	+	+	+	+	+
33	1.75D	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	+	+	+	+	+
34	GAS	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
35	AMT	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
36	H/L	-	-	-	-	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+
37	RN	-	-	-	-	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+
38	RN50	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-
39	RNA	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
40	RU50	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+
41	RUA	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
42	JØH	-	-	-	-	-	+	-	+	+	+	+	-	-	-	-	+	+	+	+	+
43	ST4	-	-	-	-	-	-	-	-	-	+	+	-	+	-	-	+	+	+	+	+

VARIANCES

Table 6.1.: Deviations of the variances of 43 estimators of location. A plus, +, means the variance of the estimator is within one standard error of the minimum variance while a minus, -, means the variance of the estimator is more than 5 standard errors away from that of the smallest variance.

	D20	17A	21A	22A	25A	ADA	1.81A	1.81*A	1.90A	2.00A	AMT	H/L	5T4
N(0,1)	-.010	-.029	-.019	-.047	.000	-.029	-.019	-.019	.000	.010	-.084	.000	-.047
N+.01N(0,9)	-.030	-.069	-.030	-.050	.000	-.040	-.020	-.030	-.010	-.040	-.059	-.030	-.069
N+.025N(0,9)	-.009	-.034	-.017	-.051	.009	-.026	-.017	-.034	-.009	-.017	-.017	.000	-.051
N+.05N(0,9)	-.008	-.024	.000	-.016	.008	-.031	-.016	-.016	-.008	.000	-.008	-.016	-.062
N+.10N(0,9)	.007	.000	.007	.000	.007	-.028	-.007	-.014	-.007	.000	.052	.007	-.056
N+.15N(0,9)	.020	.013	.020	.000	.007	-.020	.000	-.013	.007	.013	.013	.007	-.032
N+.25N(0,9)	.030	.062	.025	.000	-.029	.015	.010	.020	.000	.035	.020	.020	.015
N+.05N(0,100)	-.034	-.017	.000	-.017	.009	-.026	.000	-.026	-.008	-.017	.000	.050	-.067
N+.10N(0,100)	-.142	.007	.015	.015	-.007	-.029	.007	-.029	.000	-.029	.015	-.179	-.029
N+.25N(0,100)	-.233	.065	-.033	.013	-.179	.120	.037	.150	.051	.083	.007	-.294	-.214
N+.03C	-.018	-.044	-.018	-.044	.009	-.026	-.009	-.026	.000	-.026	-.018	-.018	-.069
N+.10C	-.025	-.016	-.008	-.033	.017	-.033	-.008	-.033	-.008	-.025	-.008	-.033	-.041
N+.50C	-.119	.012	.006	-.023	-.083	.000	.006	.029	-.006	.000	.006	-.144	.047
C	-.278	.031	-.078	-.044	-.172	.095	.015	.153	.003	.105	-.039	-.322	.098
N+.10N/U	-.025	-.008	.008	-.008	.017	-.025	.008	-.017	.000	-.008	.008	-.041	-.049
N+.25N/U	-.094	.007	.013	.007	-.007	-.032	.000	-.013	.000	.020	.000	-.129	-.013

VARIANCES

Table 6.2.: Logarithm of the relative efficiency. Efficiency in this table is computed as the ratio of the variance of 1.95A to the variance for various alternatives.

	D20	17A	21A	22A	25A	ADA	1.81A	1.81*A	1.90A	2.00A	AMT	H/L	5T4
N(0,1)	-.028	-.028	-.019	-.028	.000	-.009	-.019	-.028	.000	-.019	.112	.029	-.056
N+.01N(0,9)	.021	-.051	.021	-.021	.010	-.010	.000	.000	.010	.021	-.031	.065	-.021
N+.025N(0,9)	.000	-.057	-.033	-.095	.025	-.041	-.041	-.041	-.041	-.033	-.033	.017	-.025
N+.05N(0,9)	.017	.052	.008	.017	.000	-.073	.008	.000	.025	.008	-.008	.070	.000
N+.10N(0,9)	.058	.007	.043	.043	.007	-.021	-.007	-.034	-.007	.043	.014	.043	-.028
N+.15N(0,9)	-.007	.028	.063	.042	.021	-.053	-.027	-.040	-.027	.042	.085	-.040	-.097
N+.25N(0,9)	.010	.010	.031	-.064	-.054	.000	.015	.010	-.015	.020	-.015	-.035	.031
N+.05N(0,100)	.000	.008	.031	.008	.000	.039	.031	.000	.000	.071	.031	.000	-.045
N+.10N(0,100)	-.163	-.035	.014	-.021	-.042	-.014	-.062	-.055	-.035	-.048	.000	-.163	-.048
N+.25N(0,100)	-.472	-.037	-.143	-.081	.348	.068	-.032	.700	.033	.068	-.127	-.495	-.191
N+.03C	-.016	-.038	.056	.090	.016	.016	.064	.048	.000	.056	.048	.000	.048
N+.10C	-.064	-.032	-.024	-.032	.016	-.008	-.024	-.094	.000	-.024	-.016	-.048	-.024
N+.50C	-.372	-.006	-.034	.018	-.079	-.034	-.006	.073	-.023	-.068	-.095	-.343	-.081
C	-.301	.050	-.027	.066	-.076	.241	.050	.195	.000	.148	.100	-.350	.192
N+.10N/U	-.060	-.043	.018	-.017	.045	-.026	.018	-.084	-.026	-.034	.018	-.043	-.116
N+.25N/U	-.017	-.028	-.077	-.028	-.061	-.098	-.006	-.039	-.011	-.039	-.087	-.077	-.039

PSEUDO-VARIANCES

Table 6.3.: Logarithm of relative efficiency. Efficiency in this table is measured as the ratio of pseudo-variance of 1.95A to the pseudo-variances for various alternatives.

Just as dramatic as the OM's good performance in the short-tailed situations is its bad performance in every long-tailed situation. Table 6.1 also distinguishes several classes of strong performers in heavy-tailed situations. As a class, the hampels, numbers 12 through 17 appear strong in spite of the appearance of several minuses. The adaptive hampels, numbers 24 through 28 also appear very strong with the added bonus of no minus signs. Single estimators worthy of mention here include D20, AMT, H/L, RNA and RUA.

The estimators, RNA and RUA, are one-step rank estimators with an adaptive initial estimator. There seems to be evidence that the relatively good performance in these cases is due to the relatively good initial estimator rather than the one-step rank estimator itself. For example, the corresponding estimators, RN50% and RU50%, based on the median as an initial estimator do not perform particularly well. Also notice in Figures 5.11 and 5.12, the sensitivity curves portrayed there are essentially those for the initial estimator. For this reason, we excluded RNA and RUA from the following comparisons.

In the *PRS*, the concept of *deficiency* is introduced in order to provide a comparison of two estimators. The efficiency of an estimator under test relative to a standard estimator is defined by

$$\text{efficiency} = \frac{\text{variance of standard}}{\text{variance}}$$

and the deficiency by

$$\text{deficiency} = 1 - \text{efficiency}.$$

Notice that a negative deficiency means the estimator is more efficient than the standard. Thus deficiency is centered at 0 with negative meaning the

standard is less efficient and positive meaning more efficient.

We feel the advantage of the zero reference point is outweighed by the following anomaly. In a sense, efficiencies of 2 and $\frac{1}{2}$ mean the same thing (interchanging the roles of the standard estimator with the test estimator). The corresponding deficiencies of $\frac{1}{2}$ and -1 are not symmetrically located about 0 and, on the intuitive level, not apparently related.

Rather than compute deficiency, we have computed the natural logarithm of the likelihood ratio which also has a 0 reference (for efficiency 1) and is symmetrical in the sense that $\ln 2 = -\ln \frac{1}{2}$. The logarithm of the efficiency also has the advantage that in order to shift standards, only a simple subtraction is necessary. Tables 6.2 and 6.3 give logarithms of efficiencies for a variety of the better performing estimators relative to 1.95A. To illustrate the change of standard, the log efficiency of D20 relative to 1.95A under Cauchy alternative is -.278 while the log efficiency of 1.81*A relative to 1.95A is .153. Thus the log efficiency of D20 relative to 1.81*A is just $-.278 - .153 = -.431$. Note that a negative log efficiency means the standard is more efficient while a positive log efficiency means the standard is less efficient than the test estimator.

In Tables 6.2 and 6.3, 1.95A was used as a standard although it was not uniformly the best. However, we feel it adequately represents the class of adaptive hampels. For the moment, let us consider Table 6.2 based on variances

When compared to D20, ADA, H/L and 5T4, 1.95A dominates in the sense that there are more negatives than positives. In these cases, the magnitude of the negative log efficiency is generally much greater than that of positive log efficiency. For example, the largest negative log efficiency for D20 is

-.278 while the largest positive log efficiency is only .030. The only clear exception to this rule in these cases is the positive .120 for ADA. We can point out that relative to 1.81^*A , ADA does not exhibit this behavior.

When compared to the hampels, the adaptive hampels do not usually dominate in terms of sign. However, the worst case behavior still applies. For example, 25A has a positive log efficiency in 10 of the 16 cases. However, the maximum positive log efficiency is only .017 whereas the smallest negative log efficiency is -.179. For 22A the maximum positive log efficiency is .015 while the worst negative case is -.051, and so on.

Table 6.3 is based on pseudo-variances. Again the typical pattern is that the adaptive hampels do not uniformly dominate the other estimators, but their worst case behavior is usually considerably better than that of the other estimators.

To summarize, with adaptive hampels, one may lose a slight bit of efficiency in some cases, but gain a rather large amount in others. In this sense we say that if hampels are good, adaptive hampels are better.

There is also some question in our minds whether Q_1 and Q_2 are the most suitable measures of heavy-tailedness. Other more precise measures may yield higher efficiencies. Also we feel work needs to be done in adapting to light tails. A more sensitive measure of light-tailedness may improve the efficiencies there also. Finally, we mention the possibility of adaptive estimators in a time series context. In regard to adaptive estimation, we must take exception to the *PRS*. Properly formulated, blatantly adaptive estimators perform at least as well as non-adaptors for sample sizes less than 40.

Sensitivity curves for a variety of estimators were produced in the *PRS*. Figures 5.1 through 5.4 reproduce for completeness curves found in the *PRS*. Our figures 5.3 and 5.4 contain very slight irregularities in their central portions which we attribute to the use of the median as a starting estimator in the one-step procedures. Figure 5.5 is the sensitivity curve for ADA which was not given in the *PRS*. Figure 5.6 is the outer mean's curve. Notice that it is "insensitive" to middle observations. Figure 5.7 is the sensitivity curve for the Hogg 4 way adaptor, HGI. The point of interest here is that there is no "insensitivity" to middle values. Thus, in this stylized plot, HGI never has an opportunity to adapt to OM - the tails of the normal quantiles being too long to allow this adaptation. In general one may suspect that the concept of stylized sensitivity curves is not appropriate to adaptors. This is true of those adaptors appearing in the *PRS* also.

With these comments we remark that figures 5.8 through 5.10 represent other adaptive estimators. Figures 5.11 and 5.12 are of interest as mentioned before because they are essentially the sensitivity curves for a hampel and for the median. It appears to us as though the convergence of rank estimators is slow and, hence, that one-step rank estimators are unsatisfactory for this reason.

Figure 5.13 portrays the sensitivity curve for a normal scores rank estimator. The general raggedness again suggests slow convergence. Here a maximum of 12 steps was used. Possibly, more steps would have improved the performance of this estimator and the appearance of the curve.

Finally we may address ourselves to Table 4.3 and the time series alternative. A word on the design of the sampling situations is in order. For the

first order autoregressive model

$$x_t = \rho x_{t-1} + \epsilon_t, \quad t = 1, 2, \dots, n$$

a negative value of ρ guarantees a spectrum dominated by high frequencies. In this circumstance observations will occur on both "sides" of the center. For positive values of ρ low frequencies dominate. Hence for positive ρ , particularly for ρ close to 1 the time series could make long excursions on one "side" of the center. Thus the difficulty in location estimation in a time series occurs when ρ is positive and close to one or more generally, when the time series is close to nonstationarity. Gastwirth and Rubin (1975) discuss robust estimation in precisely this situation. In particular, they derive efficiencies of several estimators relative to the mean M . Unfortunately as Table 4.3 clearly demonstrates M is on the whole unsatisfactory. It is our assessment that no estimator presently fashionable is very satisfactory in the presence of positively correlated data.

A more complicated time series situation might have been to examine a time series which is contaminated by (possibly) uncorrelated observations. The general performance as seen in Table 4.3 did not seem to warrant this investigation, however.

List of References

- Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972), *Robust Estimates of Location: Survey and Advances*, Princeton U. Press, Princeton, N.J.
- Bickel, P.J. (1975), One-step Huber estimates in the linear model, *J. Am. Statist. Assoc.*, 70, pp. 428-434.
- Box, G.E.P. and Muller, M.E. (1959), A note on the generation of random normal deviates, *Ann. Math. Statist.*, 29, pp. 610-611.
- Gastwirth, J. (1966), On robust procedures, *J. Am. Statist. Assoc.*, 61, pp. 929-948.
- Gastwirth, J. and Rubin, H. (1975), The behavior of robust estimators on dependent data, *Ann. Statist.*, 3, pp. 1070-1100.
- Hampel, F. (1968), *Contributions to the Theory of Robust Estimation*, Ph.D. Dissertation, Berkeley.
- Hampel, F. (1974), The influence curve and its role in robust estimation, *J. Am. Statist. Assoc.*, 69, pp. 383-393.
- Hodges, J.L. and Lehmann, E.L. (1963), Estimates of location based on rank tests, *Ann. Math. Statist.*, 34, pp. 598-611.
- Hogg, R.V. (1974), Adaptive robust procedures: a partial review and some suggestions for future applications and theory, *J. Am. Statist. Assoc.*, 69, pp. 909-923.
- Huber, P.J. (1964), Robust estimation of a location parameter, *Ann. Math. Statist.*, 35, pp. 73-101.
- Johns, M.V. Jr. (1974), Nonparametric estimators of location, *J. Am. Statist. Assoc.*, 69, pp. 453-460.
- Kraft, C. and Van Eeden, C. (1969), Efficient linearized estimates based on ranks, *Proceedings of the First International Symposium on Nonparametric Statistics*, Cambridge U. Press, pp. 267-273.
- Paley, R.E.A.C. and Wiener, N. (1934), *Fourier Transforms in the Complex Domain*, American Mathematical Society, Providence, R.I.
- Parzen, E. (1962), On the estimation of a probability density function and mode, *Ann. Math. Statist.*, 33, pp. 1065-1076.
- Schweder, T. (1974), Window estimation of the asymptotic variance of rank estimators of location, unpublished manuscript.