VARIANCES FOR ADAPTIVE TRIMMED MEANS

by

John E. Boyer, Jr.

Technical Report No. 145
Department of Statistics ONR Contract

August, 1981

Research sponsored by the by the Office of Naval Research
Contract N00014-77-C-0699/N00014-76-C-0613

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

Variances for Adaptive Trimmed Means

John E. Boyer, Jr.
Southern Methodist University

ABSTRACT

Variances for adaptive estimators of the location parameter in a family of symmetric distributions including the uniform, normal, and double exponential are examined at small to moderate sample sizes. The estimators are all trimmed means or means of trimmings where the proportion of trimming is determined by an easily computed measure of nonnormality. Comparisons are made to the asymptotic variances.

Variances for Adaptive Trimmed Means

John E. Boyer, Jr.
Southern Methodist University

1. INTRODUCTION

In a recent article Prescott (1978) discussed the use of adaptive trimmed means and means of trimmings for estimating a location parameter from a symmetric family of distributions. The proportion of the sample trimmed or retained is determined by the value of the quantity $\hat{Q}$, a measure of the length of the tails of the distribution based on the means of groups of observations from the extremes of the ordered sample.

Asymptotic properties based on the corresponding population quantity, $\hat{Q}$, were derived for several different such estimates under the assumption that the underlying distribution belongs to the exponential power family of distributions. Since the population quantity will not, in practice, be available, the corresponding properties are examined in the study below and compared with the values found in Prescott's computations.

2. Asymptotic Variances for Trimmed Means

Let $x_1 < x_2 < \cdots < x_n$ be an ordered sample of size n from a population distribution function F(x) and density function f(x). The α-trimmed mean is defined as

$$m(\alpha) = \frac{1}{n(1-2\alpha)} \left\{ \sum_{i=[n\alpha]+2}^{n-[n\alpha]-1} x_i + (1 + [n\alpha] - n\alpha)(x_{[n\alpha]+1} + x_{n-[n\alpha]}) \right\} . \quad (2.1)$$

The mean of the observations discarded in m(α) is the α-mean of the trimmings, denoted $m^c(\alpha)$ and is given by

$$m^c(\alpha) = \frac{1}{2n\alpha}\left\{\sum_{i=1}^{[n\alpha]}(x_i + x_{n-i+1}) + (n\alpha-[n\alpha])(x_{[n\alpha]+1} + x_{n-[n\alpha]})\right\} . \qquad (2.2)$$

It should be noted that the limiting forms for these estimators are commonly encountered estimators, i.e., $m(0.5)$ is the median, $m(0) = m^c(0.5)$ is the mean and $m^c(0)$ is the midrange.

Prescott (1978) considers these estimators for the location parameter θ in the exponential power family of symmetric distributions defined by the density function

$$f(x) = \frac{1}{2\Gamma(\frac{\tau+1}{\tau})} e^{-|x-\theta|^\tau} \qquad -\infty < x < \infty \ , \qquad \tau \geq 1. \qquad (2.3)$$

These distributions are symmetric about θ with variance $\sigma_\tau^2 = \Gamma(3/\tau)/\Gamma(1/\tau)$. If we regard $\gamma = \frac{1}{\tau}$ as a continuous parameter in the interval $[0,1]$, this family may be thought of as containing distributions which change gradually from the uniform $(\gamma=0)$, through short-tailed symmetric distributions to the normal $(\gamma = \frac{1}{2})$, then through long-tailed symmetric distributions to the double exponential $(\gamma=1)$.

Prescott discusses the robustness properties and derives the asymptotic variances for $m(\alpha)$ and $m^c(\alpha)$ for distributions belonging to this family by using influence curve techniques. As all of the above estimators are unbiased for θ in all of the distributions belonging to the exponential power family, the asymptotic variance of the particular estimator, when compared to the Cramér-Rao lower bound, provides a measure of the efficiency of the estimation.

As different estimators from the family of trimmed means and means-of-trimmings are more efficient depending on which member of the exponential family is being considered, adaptive estimation techniques enter in a natural way. In particular, several statistics which choose a trimming proportion α

based on the measure of nonnormality (or tailweight)

$$\hat{Q} = \left(\bar{U}_{(.05)} - \bar{L}_{(.05)}\right) / \left(\bar{U}_{(.05)} - \bar{L}_{(.50)}\right) \tag{2.4}$$

proposed by Hogg (1974), where $\bar{U}_{(\beta)}$ ($\bar{L}_{(\beta)}$) is the average of the largest (smallest) $n\beta$ order statistics, with fractional items used if $n\beta$ is not an integer, are presented as possible adaptive estimators for the exponential power family. The choice of $\hat{Q}$ over other measures of nonnormality or tailweight such as sample kurtosis is discussed in detail in Hogg (1972, 1974) and Davenport (1971) and the choice of the particular 5% and 50% proportions and some asymptotic properties for $\hat{Q}$ are discussed there.

Prescott's computations of the variances of the above suggested trimmed means are, however, based on knowledge of Q, the population quantity which corresponds to $\hat{Q}$. Nevertheless, it is clear that in this situation, Q would not be known to the statistician or there would be no need to adapt in the first place.

Parr (1979) points out that the variance of an adaptive estimator will be dependent on the relative frequency with which the adapting statistic, $\hat{Q}$ picks different trimming proportions. That is, the appropriate variances to consider for the various adaptive estimators are the weighted sums of the conditional variances, where the weights are the proportion of observations that yield the corresponding values of $\hat{Q}$, and the conditioning is on the observed $\hat{Q}$. Since Prescott's estimates of the variances use only the trimmed mean or mean-of-the trimmings which appears to do best or "nearly" best for a given member of the exponential power family, they do not truly reflect the adaptive nature of the statistics. The results of taking that aspect of the statistic into account will be seen in the simulation discussed below.

Analytically, if T is the trimmed mean chosen by the adaptive procedure, then $\text{Var } T = E\{\text{Var}[T|\hat{Q}]\} + \text{Var}\{E[T|\hat{Q}]\}$. However, since all the trimmed means are symmetric estimators and all of the members of the exponential family are symmetric distributions, $E[T|\hat{Q}] \equiv \theta$ and thus the second term will always be zero. If T_Q is chosen to (nearly) minimize $\mathrm{Var}[T_Q]$, then $\mathrm{Var}[T|\hat{Q}] \geq \mathrm{Var}[T_Q]$, so $\mathrm{Var}\ T = E[\mathrm{Var}\ [T|Q]] \geq \mathrm{Var}[T_Q]$. Thus the asymptotic variances given by the influence curve calculations will in general be smaller than the variances which can be achieved in practice, and the difference can be attributed to the problem of estimating Q.

3. Adaptive Trimmed Means

The adaptive estimators investigated in the simulation study discussed below are the same as those presented in Prescott. The first is an estimator suggested by Hogg (1974) given by

$$T_1 = \begin{cases} m^c(\frac{1}{4}) & \hat{Q} < 2.0 \\ m(0) & 2.0 \leq \hat{Q} < 2.6 \\ m(3/16) & 2.6 \leq \hat{Q} \leq 3.2 \\ m(3/8) & 3.2 < \hat{Q} \,. \end{cases} \tag{3.1}$$

The second is an estimator suggested by Prescott (and denoted by T* in that article)

$$T_2 = \begin{cases} m^c(.2) & \hat{Q} < 2.2 \\ m^c(.3) & 2.2 \leq \hat{Q} < 2.4 \\ m(0) & 2.4 \leq \hat{Q} \leq 2.8 \\ m(.2) & 2.8 < \hat{Q} \leq 3.0 \\ m(.3) & 3.0 < \hat{Q} \,. \end{cases} \tag{3.2}$$

The third, also suggested by Prescott, and denoted by T** there, is an extension of the notion that more intervals for $\hat{Q}$ would produce asymptotic variances closer to the minima suggested by the Cramér-Rao bounds. Thus the statistic below adapts or adjusts continuously rather than in a step-wise fashion.

$$T_3 = \begin{cases} m^c(0) \equiv \text{midrange} & \hat{Q} < 1.9 \\ m^c[(Q-1.9)\cdot 0.7)] & 1.9 \le \hat{Q} \le 2.6 \\ m[(Q-2.6)\cdot 0.7] & 2.6 < \hat{Q} \le 3.3 \\ m(.5) \equiv \text{median} & 3.3 < \hat{Q}\ . \end{cases} \tag{3.3}$$

The simulation consisted of 2000 repetitions for each of the three statistics suggested above, at each of the nine parameter values $\gamma = 0(\frac{1}{8})1$ and at each of the three sample sizes $n = 10$, 20, and 40. The results of the simulation as well as the asymptotic variances provided by the influence curve calculations and the values of the Cramér-Rao lower bound appear in Table 1. In addition figures A, B and C are graphs of the corresponding data for the estimators T_1, T_2, and T_3 respectively, using diagrams as in Prescott's work.

4. Properties of the Variances of $m(\alpha)$ and $m^c(\alpha)$

Since the adaptive estimator chosen is dependent on the value $\hat{Q}$, estimation of Q is of interest in its own right. Davenport has shown that for the uniform, normal and double exponential distributions, $\hat{Q}$ is asymptotically distributed as a normal random variable with mean Q and finite variance. In the appendix to this paper, the condition on the distribution required for that result is shown to hold for any member of the exponential power family.

As seen in Figure D, the tendency is for $\hat{Q}$ to underestimate Q at small to moderate sample sizes. The results graphed there are the true values of Q minus the averages of $\hat{Q}$ obtained from 2000 samples of the given sample size. Additionally, it is clear that the magnitude of the underestimation increases as Q or $(1/\tau)$ increases, for any of the sample sizes studied. This underestimation is a substantial factor in the variance calculations.

As an example of the point illustrated in Section 3, consider the estimator T_2 at $1/\tau = 5/8$ with samples of size 20. For this parameter configuration Q is 2.77 so the asymptotic results are based on the sample mean, which has (standardized) variance 1.0. In 2000 samples, however $\hat{Q}$ was less than 2.4 715 times, between 2.4 and 2.8 699 times, and larger than 2.8 586 times. Thus the sample mean is chosen only about 35% of the time, in practice. The observed variance of 1.1222 reflects the fact that the 65% of the time when estimators other than m(0) are chosen the variance will be larger than that of m(0), which is nearly best among trimmed means and means of trimmings.

On the other hand when the adaptive estimator does not choose a nearly optimal trimming proportion, the finite sample results may give variances which are smaller than those suggested by the asymptotic variances.

For example, for the 2000 samples of size 10 at $1/\tau = 1/8$, the value of $\hat{Q}$ was less than 2.0 1358 times, between 2.0 and 2.6 572 times, between 2.6 and 3.2 66 times and larger than 3.2 4 times. Thus, if T_1 is the estimator being considered, even though the population value of Q is 2.05 and the sample mean, m(0), is the estimator that should be chosen, more than 67% of the sample runs choose $m^c(1/4)$. The variance for this estimator could therefore also be expected to be quite different from that of the sample mean. In fact, $m^c(1/4)$ with asymptotic variance .5869, is a good deal closer to the best trimmed mean or mean of trimmings available for $1/\tau = 1/8$ than the sample mean. Thus the estimator T_1, for samples of size 10 actually performs considerably better than the asymptotic variance would suggest. This is reflected in Figure A, and Table 1 where the variance for samples of size 10 is seen to be .8216, while the asymptotic value is 1.0. Note that the best available estimator at $\gamma = 1/8$ is a mean of trimmings with α between .05 and .10, with an asymptotic variance of approximately .40. It is seen that in this particular case, the fact that $\hat{Q}$ underestimates Q is advantageous.

The estimator T_3 exhibits the properties claimed by Prescott in his concluding section. Particularly for small samples ($n < 20$) and long-tail ($\gamma \geq 1/2$) distributions, $\hat{Q}$ severely underestimates the parameter Q. Consequently, the trimming proportion chosen is considerably smaller than the optimal one, and the resulting variance is considerably larger than that of the optimal trimmed mean. The performance is somewhat better for shorter-tailed distributions, but still not particularly impressive.

As the sample size increases however, the continuous adaptation begins to fare very well. By $n = 40$, the performance of T_3 is as good as any of the estimators studied. This coincides with the assertions made by Prescott.

TABLE 1

VARIANCES OF ESTIMATORS T_1, T_2, T_3, BASED ON 2000 SAMPLES

	$1/\tau$	.000	.125	.250	.375	.500	.675	.750	.875	.1000
	Q	1.90	2.05	2.20	2.40	2.58	2.77	2.95	3.13	3.30
T_1	n = 10	.7692	.8216	1.0124	1.0805	1.1140	1.0608	1.0344	1.0242	.9729
	n = 20	.7564	.8670	1.0163	1.0973	1.0592	1.0253	.9388	.8597	.7338
	n = 40	.6710	.8079	1.0428	1.0580	1.0301	.9660	.9302	.7895	.6607
	asympt	.5000	1.0000	1.0000	1.0000	1.0000	.9389	.8655	.7552	.5478
T_2	n = 10	.6327	.7140	.9471	1.0970	1.2435	1.2459	1.2915	1.3711	1.3280
	n = 20	.5524	.6763	.8891	1.0750	1.1254	1.1222	1.0657	1.0014	.8285
	n = 40	.4570	.5773	.8632	1.0123	1.0719	1.0126	.9871	.8506	.6809
	asympt	.4000	.5186	.7966	1.0000	1.0000	1.0000	.8632	.7071	.5844
T_3	n = 10	.5682	.7014	1.0168	1.1937	1.3902	1.3593	1.4454	1.5070	1.4302
	n = 20	.4049	.6605	.9474	1.1645	1.2328	1.1912	1.1152	1.0206	.8595
	n = 40	.2374	.5850	.9262	1.0734	1.1206	1.0497	1.0060	.8530	.6902
	asympt	.0000	.4259	.7566	.9468	1.0013	.9397	.8575	.6883	.5000
Cramér-Rao Lower Bound		.0000	.3924	.7295	.9358	1.0000	.9485	.8225	.6626	.5000

Figure A. ASYMPTOTIC VARIANCE OF T_1 AND VARIANCE USING SAMPLES OF SIZE 10,20 AND 40

$$T_1 = \begin{cases} m^c(1/4) & Q < 2.0 \\ m\,(0) & 2.0 \le Q \le 2.6 \\ m\,(3/16) & 2.6 < Q \le 3.2 \\ m\,(3/8) & 3.2 < Q \end{cases}$$

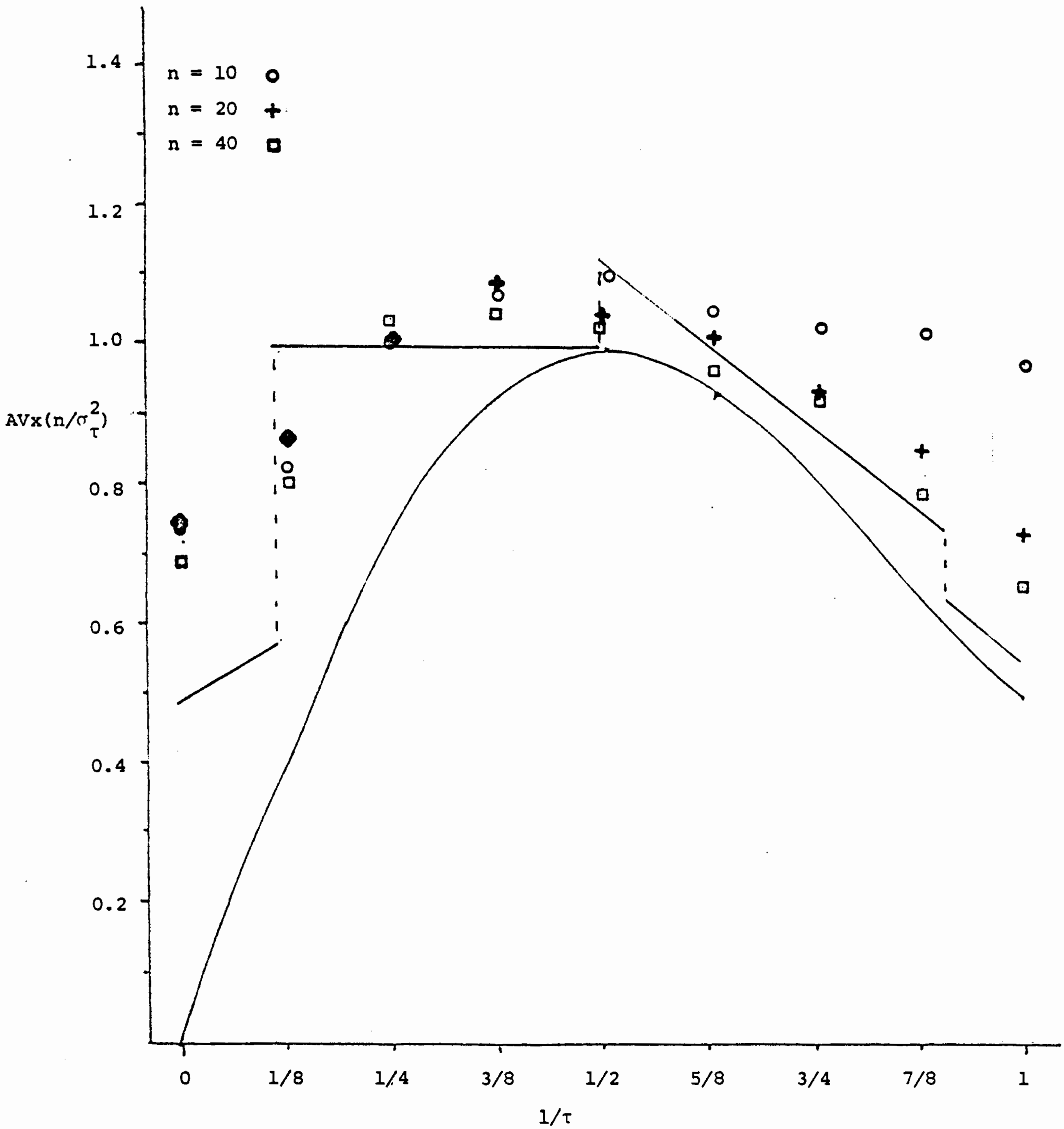


Figure B. ASYMPTOTIC VARIANCE OF T_2 AND VARIANCE USING SAMPLES OF SIZE 10,20 AND 40

$$T_2 = \begin{cases} m^c(.2) & Q < 2.2 \\ m^c(.3) & 2.2 \le Q < 2.4 \\ m\ (0) & 2.4 \le Q \le 2.8 \\ m\ (.2) & 2.8 < Q \le 3.0 \\ m\ (.3) & 3.0 < Q \end{cases}$$

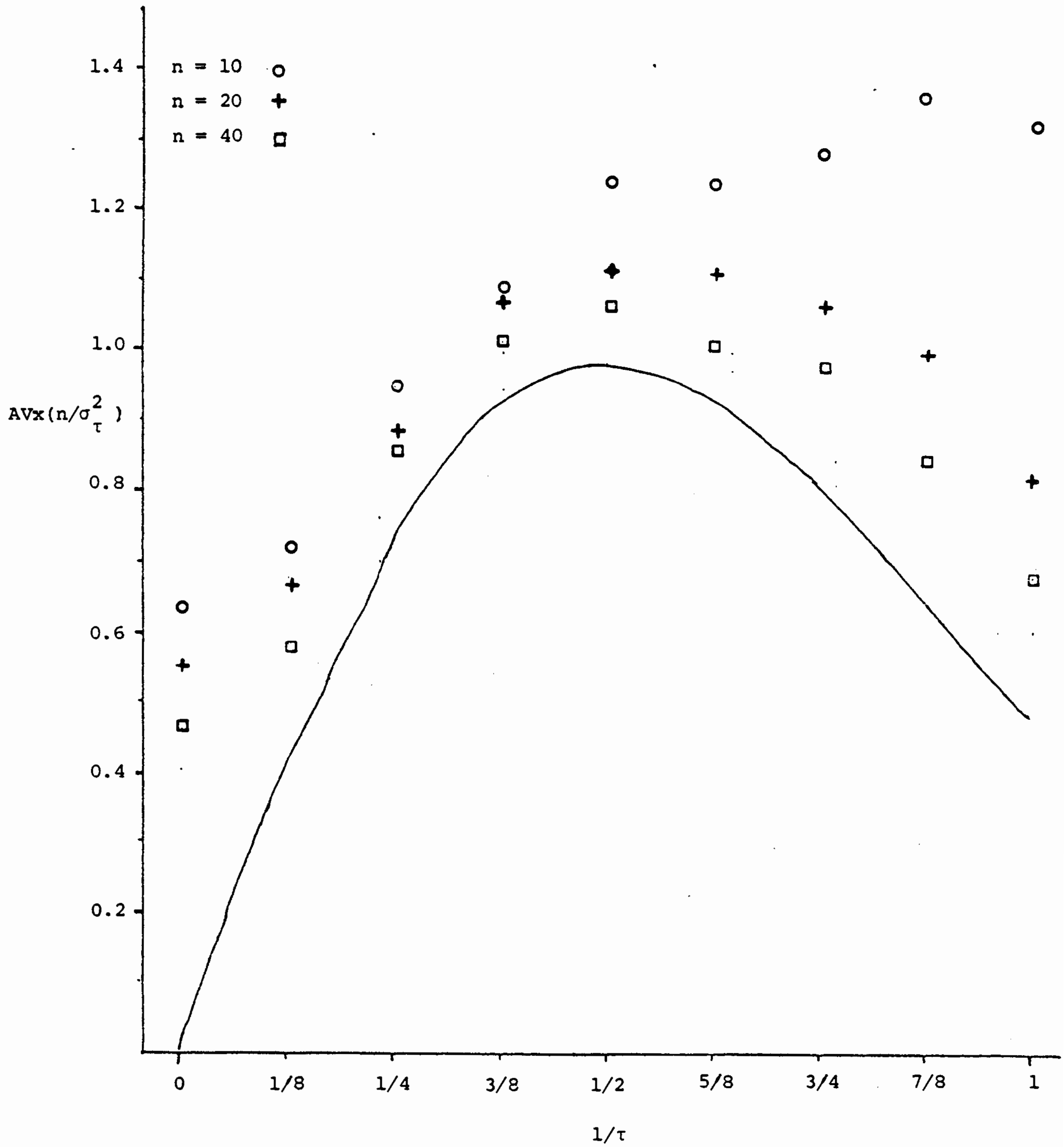


Figure C. ASYMPTOTIC VARIANCE OF T_3 AND VARIANCE USING SAMPLES OF SIZE 10,20 AND 40

$$T_3 = \begin{cases} m^c(0) \equiv \text{midrange} & Q < 1.9 \\ m^c[(Q-1.9)0.7] & 1.9 \leq Q < 2.6 \\ m[(Q-2.6)0.7] & 2.6 \leq Q \leq 3.3 \\ m\left(\frac{1}{2}\right) \equiv \text{median} & 3.3 < Q \end{cases}$$

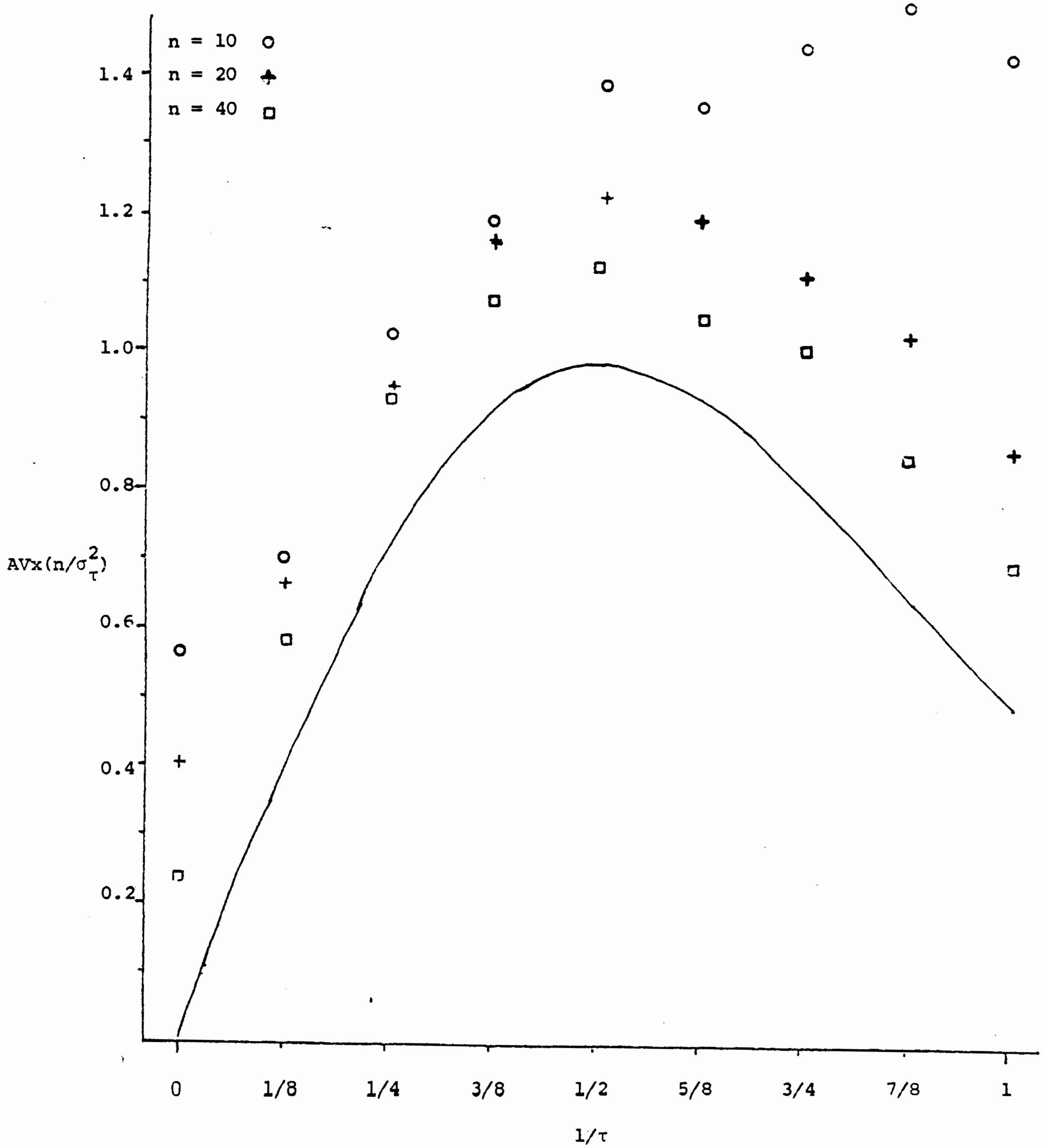
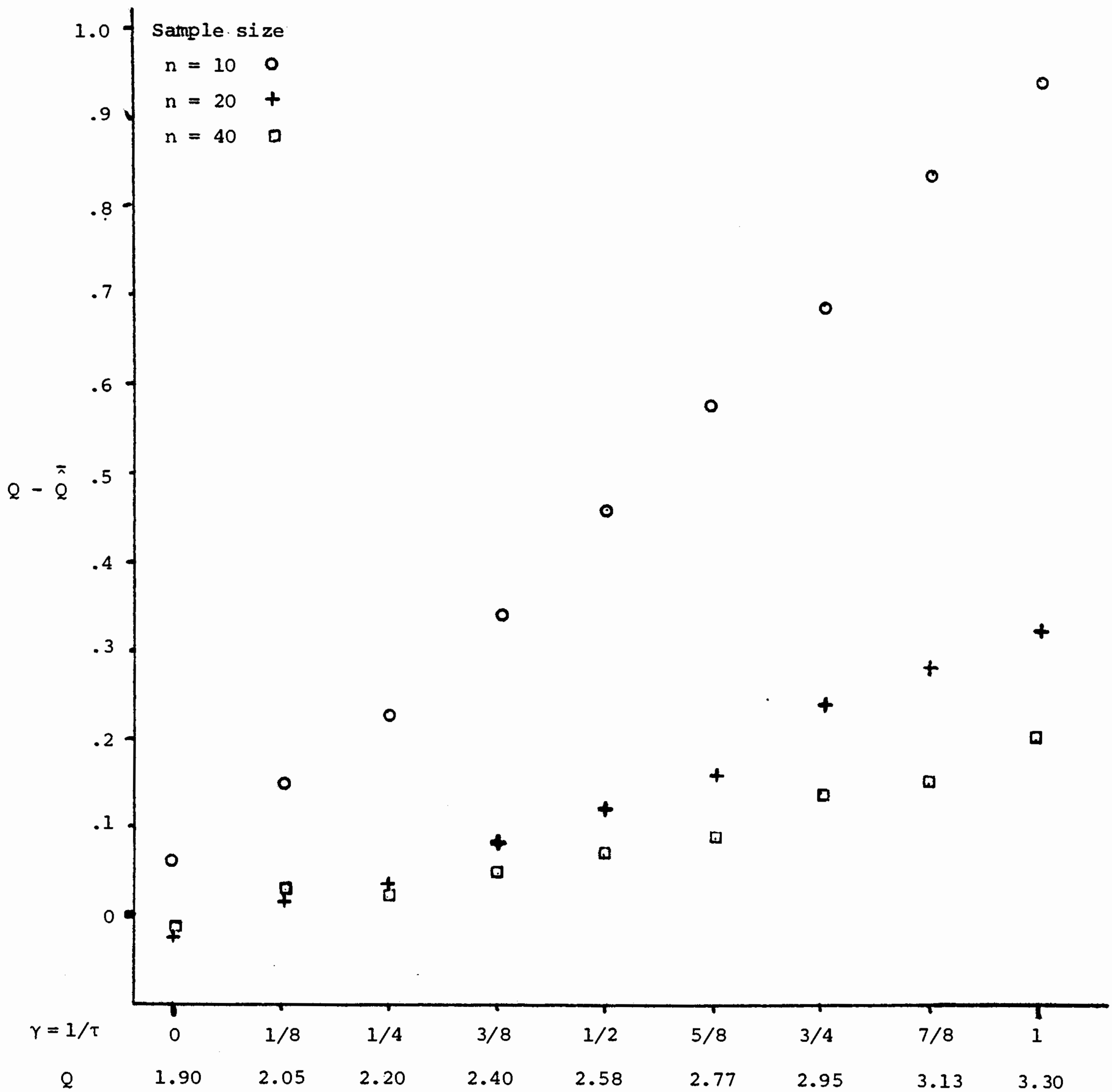


Figure D. True Value of Q Minus Average Value of $\hat{Q}$ for 2000 Samples from Exponential Power Family of Distributions



5. Concluding Remarks

Investigations using small to moderate sample size indicate that adapting too closely may not be worth the effort. $\hat{Q}$ will give only a general idea of the value of Q, and at small sample sizes gives a value that is too small. In fact, in a slightly different setting (a two-sample problem) Randles and Hogg (1973) propose deciding that the underlying distributions are light, medium, or heavy tailed as

$$\hat{Q} < 2.08 - 2/N,$$

$$2.08 - 2/N \leq \hat{Q} \leq 2.96 - 5.5/N,$$

$$2.96 - 5.5/N < \hat{Q},$$

respectively, where the samples are of size m and n and $N = (m^2 + n^2)/m+n$. In a different discussion, Hogg (1974) recommends an adapting scheme that uses only one trimmed mean if the sample size is less than or equal to 10, one of two adaptively chosen trimmed means if $10 < n \leq 20$, one of three adaptively chosen trimmed means if $20 < n \leq 30$, etc., thus adapting more closely as the sample size increases. Both of these suggestions and the current study support Prescott's conclusion that if the sample size is fairly large ($n \geq 50$) the continuously adapting T_3 should be a useful robust estimator, but if $n < 50$, and particularly if one suspects long-tailed non-normality, then T_2 might be preferable.

6. APPENDIX

In section 4, reference was made to the requirement on a distribution in order that $\hat{Q}$ might be asymptotically normal. In Davenport (1971, p. 8), four conditions on a distribution are described as sufficient to ensure that asymptotic property. The fourth condition, essentially a requirement on smoothness of the tails of the distribution, comes directly from Chernoff (1967, p. 61, Assumption 5). The Chernoff condition can be easily shown to be satisfied under the assumption that the distribution is a member of the exponential power family of distributions (with the exception of $\tau = \infty$, the uniform distribution, for which the result can be shown directly) by way of the following theorem which establishes the relationship between the density and the cumulative distribution function for all members of the family.

Theorem. Let $f(x) = \frac{1}{2\Gamma(1+1/\tau)} \exp(-|x|^{\tau})$, $1 \le \tau < \infty$, the density for a member of the exponential power family, and let $F(\cdot)$ be the corresponding cumulative distribution function. For any $x > 0$

$$\tau x^{\tau-1} (1-F(x)) \le f(x) \le (\tau x^{\tau-1} + (\tau-1)1/x)(1-F(x)).$$

Proof: The derivative of $f(x)$ is $-\tau x^{\tau-1} f(x)$ and the derivative of $1 - F(x)$ is $-f(x)$ so that

$$f(x) = \int_x^{\infty} \tau y^{\tau-1} f(y)\,dy$$

$$= -\tau y^{\tau-1} (1-F(y)) \Big|_x^{\infty} + \tau(\tau-1) \int_x^{\infty} y^{\tau-2} (1-F(y))\,dy$$

where the second equality follows from integration by parts. The second term in the latter expression is positive and the first term in that expression is $\tau x^{\tau-1} (1-F(x))$. Thus $f(x) \ge \tau x^{\tau-1} (1-F(x))$ as desired.

Now replace $1 - F(y)$ in the second term by its upper bound $\frac{1}{\tau} y^{1-\tau} f(y)$ just obtained and the result is

$$f(x) \le \tau x^{\tau-1} (1-F(x)) + \tau(\tau-1) \int_x^{\infty} y^{\tau-2} \frac{1}{\tau} y^{1-\tau} f(y)\,dy$$

$$= \tau x^{\tau-1} (1-F(x)) + (\tau-1) \int_x^{\infty} \frac{1}{y} f(y)\,dy$$

$$\le \tau x^{\tau-1} (1-F(x)) + (\tau-1) \frac{1}{x} \int_x^{\infty} f(y)\,dy$$

$$= (1-F(x))(\tau x^{\tau-1} + (\tau-1)\frac{1}{x}).$$

Combining gives

$$\tau x^{\tau-1}(1-F(x)) \leq f(x) \leq (\tau x^{\tau-1} + (\tau-1)\frac{1}{x})(1-F(x)) \qquad \text{Q.E.D.}$$

Note that the inequalities are strict for $\tau > 1$ and when $\tau = 1$, the equality $1 - F(x) = f(x)$ holds for all $x > 0$. An equivalent result holds for $x \leq 0$ and the two conditions together provide all the necessary machinery to satisfy the Chernoff requirements.

REFERENCES

Chernoff, Herman, Gastwirth, Joseph L., and Johns, M.V., Jr. (1967) "Asymptotic Distributions of Linear Combinations of Functions of Order Statistics with Applications to Estimation," Annals of Mathematical Statistics, 38, 52-72.

Davenport, Alan S. (1971), "Classification of Distributions Based on Measures of Spread," Ph.D. Dissertation, University of Iowa.

Hogg, Robert V. (1972), "More Light on Kurtosis and Related Statistics," Journal of the American Statistical Association, 67, 422-424.

Hogg, Robert V. (1974), "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory," Journal of the American Statistical Association, 69, 909-923.

Parr, William C. (1980), "A Conditional Property of Adaptive Estimators," Journal of Statistical Computation and Simulation, 10, 289-295.

Prescott, P. (1978), "Selection of Trimming Propprtions for Robust Adaptive Trimmed Means," Journal of American Statistical Association, 73, 133-140.

Randles, Ronald H. and Hogg, Robert B. (1973) "Adaptive Distribution-Free Tests," Communications in Statistics, 2, 337-356.