ON ESTIMATING VARIANCES OF ROBUST ESTIMATORS WHEN THE ERRORS ARE ASYMMETRIC

by

Raymond J. Carroll*
University of North Carolina at Chapel Hill

ABSTRACT

We investigate the effects of asymmetry on estimates of variance of robust estimator in location and regression problems, showing that heavy skewness of errors can seriously bias the common variance estimates for location and intercept, a problem that can be corrected by jackknifing for location but is more intractable for the intercept in regression. The scale parameters in regression seem not to be as seriously subject to this bias if the sample size is large compared to the number of parameters.

*Research supported by the Air Force Office of Scientific Research under Grant AFOSR-75-2796.

Introduction

The theoretical results and Monte-Carlo studies in the area of robustness have in the main focused on symmetric distributions (Andrews, et al (1972)) or procedures which are not scale equivariant(which effectively eliminates most problems due to asymmetry when the number of dimensions in the problem is fixed). Recently, Huber (1973) and Bickel (1978) have examined situations in which the asymmetry of errors can lead to quite complicated results. In this paper we study the effects of asymmetric errors in two very simple situations: (one and two dimensional) location problems and simple linear regression. We have focused our attention on estimating the variability of robust point estimators in these problems.

A major difficulty with considering asymmetric errors has been that location (intercept) is not uniquely defined. However, asymmetric data do occur and there are situations where data transformations to achieve symmetry either make no sense or are not possible. In regression, it might be conjectured that asymmetry has different effects on intercept and slope (see Section 3); if so, there will be situations where one might invest much effort in data transformations, when the parameters of interest are not influenced by the asymmetry. This alone seems a good reason for studying asymmetry, but as another example, in ranking and selection problems, one is often interested in the stochastically largest population; robust estimators have been proposed in place of the sample mean for this problem, and a data transformation which achieves symmetry for all populations may not be possible.

In Section 2, we study the effects of asymmetry of errors for the one-dimensional location problem, using M-estimates, trimmed means, and an adaptive trimmed mean due to Hogg (1974). In Section 3, we study M-estimates of regression. Our major qualitative conclusions are as follows:

(1) The published estimates of variance for M-estimates of location will be consistently small if the errors are heavily skewed. This problem appears amenable to solution if a variance estimate suggested by jackknifing is used.

(2) In linear regression, a similar conclusion holds true for the intercept. Theoretical and Monte-Carlo results show that slope parameters are influenced only negligibly by skewness. Jackknifing for the intercept does not work as well here, with the variance estimates now being slightly too large.

(3) In the two-sample problem with equal scales, a robust test statistic can be constructed with a consistent variance estimtate, even if the errors are asymmetric. This statistic, however, has unknown sensitivity to heteroscedasticity of variances.

Location Estimates

In this section we define the location estimates used in the study, state two Lemmas and give a discussion of expected results. An M-estimator is defined as a solution to the equation

$$\Sigma\, \psi((X_i - T_n)/s_n) = 0, \tag{2.1}$$

where s_n is an estimate of scale. In the Monte-Carlo study, we used two functions ψ defined by

$$\psi(x) = -\psi(-x) = \max(-2, \min(x,2)) \tag{2.2}$$

$$\begin{aligned} \psi(x) &= -\psi(-x) \\ &= x && 0 \le x \le 2.25 \\ &= 2.25 && 2.25 \le x \le 3.75 \\ &= 2.25\ (15-x)/11.25 && 3.75 \quad x \le 15 \\ &= 0 && x \ge 15\ . \end{aligned} \tag{2.3}$$

The function (2.2) is referred to as Huber's ψ, while (2.3) is Hampel's ψ. The solutions to (2.1) were generated iteratively (Gross (1977)), starting with the sample median and with two estimates of scale:

(2.4) $$s_{in} = (\text{median absolute residual})/.6745$$

(2.5) $$s_{2n}^2(\text{new}) = \{(n-1)^{-1} \sum_{i=1}^{n} \psi^2((X_i - T_n)/s_{2n}(\text{old}))\} s_{2n}^2(\text{old})/a ,$$

where $a = E_\Phi \psi^2(Z)$, the expectation taken under the standard normal distribution function. Two estimates of the variance of $n^{\frac{1}{2}} T_n$ were used, namely

(2.6) $$D_{1n}(i) = H^2 s_{in}^2 (n-1)^{-1} \sum_{j=1}^{n} \psi^2((X_j - T_n)/s_{in})/b_i^2 , \text{ where}$$

$$b_i = n^{-1} \sum_{j=1}^{n} \psi'((X_j - T_n)/s_{in}), \; H = 1 + n^{-1}(1-b_i)/b_i$$

(2.7) $$D_{2n}(i) = 1.08(n-1)^{-1} \sum_{j=1}^{n} w_{ij}(Y_j - T_n)^2, \text{ where}$$

$$w_{ij} = \psi((X_j - T_n)/s_{in})/((X_j - T_n)/s_{in}) .$$

We used the factor 1.08 in $D_{2n}(i)$ so that the two estimates have approximately the same value for normal samples in our Monte-Carlo experiment, and we used the factor H^2 in $D_{1n}(i)$ as suggested by Huber (1973).

The variance estimates $D_{1n}(i)$ and $D_{2n}(i)$ are essentially as suggested by Gross (1977), the first motivated by the asymptotic variance formula for $n^{\frac{1}{2}} T_n$

$$\int \psi^2(x)dF(x)/\{ \int \psi'(x)dF(x)\}^2 ,$$

appropriate when F is symmetric, the second suggested by the weighted least squares nature of the iterative estimation procedure. In the general case with possible asymmetry, the behavior of M-estimates T_n can be summarized by the following result, for which exact conditions can be generated from the method of proof of Theorem 1 of Carroll (1978a).

Lemma 1. Suppose that for some constants $T_i(F)$, $\sigma_i(F)$ we have that $T_n \overset{p}{\to} T_i(F)$, $n^{\frac{1}{2}}(s_{in} - \sigma_i(F)) = O_p(1)$, and $E\,\psi((X_1 - T_i(F))/\sigma_i(F)) = 0$. Taking $T_i(F) = 0$, $\sigma_i(F) = 1$ (without loss of generality), for ψ smooth,

$$(2.8) \quad (E_F\psi'(X_1))T_n = n^{-1} \sum_{i=1}^{n} \psi(X_i) + (1 - s_{in})E_F\, X_1\psi'(X_1) + O_p(n^{-1}). \qquad \square$$

Discussion of Lemma 1

The following qualitative considerations emerge from Lemma 1 and are verified in the Monte-Carlo experiment.

(1) If the distribution function F is not symmetric, then, in general, $E_F X_1\psi'(X_1) \neq 0$ so that $D_{1n}(i)$ (which is asymptotically correct in the symmetric case) will typically underestimate the true variance. The same should hold for $D_{2n}(i)$.

(2) The bias in $D_{1n}(i)$ and $D_{2n}(i)$ should be small for distributions which are nearly symmetric because T_n is a smooth function of the data. The bias will become large as the degree of asymmetry increases.

(3) When s_{2n} is used as scale, T_n should be a particularly smooth function of F. From Jaeckel (1972), this means that jackknifing will probably be effective in estimating the variance of T_n. As pointed out by Huber (1977, p.26), such a variance estimate will be more appropriate for T_n than for the jackknifed version of T_n.

We also consider trimmed means and two adaptive versions. Define $U_n(\alpha)(L_n(\alpha))$ as the mean of the largest (smallest) $n\alpha$ order statistics. As a measure of tail length, when $n\alpha$ is an integer (true in the situations considered here), Hogg (1974) proposed

$$Q_n = (U_n(.20) - L_n(.20))/(U_n(.50) - L_n(.50)).$$

Then, if m% refers to a m% symmetrically trimmed mean, the Hogg adaptive estimate is defined by

$$\begin{aligned} HG1 &= 5\% \quad \text{if} \quad Q_n \le 1.81 \\ &= 10\% \quad \text{if} \quad 1.81 < Q_n \le 1.87 \\ &= 25\% \quad \text{if} \quad Q_n > 1.87 \;. \end{aligned}$$

The estimate of an m% trimmed mean is given by Shorack (1974), Huber (1977, eq.(10.4)) has shown that this estimate essentially arises from jackknifing the m% trimmed mean. Calling this variance $\sigma^2(m\%)$, we define

$$D(\text{Hogg}) = \sigma^2(m\%) \text{ if } HG1 = m\%.$$

We also consider an estimate suggested by Switzer (1970), defined by

$$\text{Switzer} = m\% \text{ if } \sigma^2(m\%) \text{ is minimum among } \sigma^2(5\%),\ \sigma^2(10\%) \text{ and } \sigma^2(25\%).$$

If one assumes that $n^{\frac{1}{2}}(Q_n - Q(F))$ has a limit distribution, it is easy to obtain the following disturbing result, precise conditions for which could be given but are omitted.

Lemma 2. Suppose $Q(F) = 1.81$. If F is symmetric about $\theta(F)$, then $n^{\frac{1}{2}}(HG1 - \theta(F))$ has a non-normal limit distribution. If F is not symmetric, there typically exists no finite θ for which $n^{\frac{1}{2}}(HG1 - \theta)$ has a limit distribution.

Sketch of proof. Let $T_n(m)$ be the m% symmetrically trimmed mean with $n^{\frac{1}{2}}(T_n(m) - \theta(m))$ having a normal limit distribution. If F is symmetric about $\theta(F)$, since Q_n is even and $T_n(m)$ is odd, $n^{\frac{1}{2}}(T_n(m) - \theta(F))$ is asymptotically independent of $n^{\frac{1}{2}}(Q_n - Q(F))$. Define

$$A_{1n} = \{Q_n \le Q(F)\},\quad A_{2n} = \{Q_n > Q(F)\}\quad B_1 = \{n^{\frac{1}{2}}(HG1 - \theta(F)) \le z\}.$$

An $n \to \infty$,

$$P(B_1) = P(A_{1n} \cap B_1) + P(A_{2n} \cap B_1)$$
$$\approx P(A_{1n} \text{ and } n^{\frac{1}{2}}(T_n(5) - \theta(F)) \leq z)$$
$$+ P(A_{2n} \text{ and } n^{\frac{1}{2}}(T_n(10) - \theta(F)) \leq z)$$
$$\approx \tfrac{1}{2}\{P(n^{\frac{1}{2}}(T_n(5) - \theta(F)) \leq z) + P(n^{\frac{1}{2}}(T_n(10) - \theta(F)) \leq z)\},$$

verifying the first part of Lemma 2. If F is asymmetric and $\theta(5) \neq \theta(10)$ (the usual case), for any θ ,

$$(2.9) \qquad P(B_1) \approx P(A_{1n} \text{ and } n^{\frac{1}{2}}(T_n(5) - \theta(5)) \leq z + n^{\frac{1}{2}}(\theta - \theta(5)))$$
$$+ P(A_{2n} \text{ and } n^{\frac{1}{2}}(T_n(10) - \theta(10)) \leq z + n^{\frac{1}{2}}(\theta - \theta(10))).$$

Clearly, no θ can be chosen so that the right hand side of (2.9) is a probability distribution. □

Because of Lemma 2 we expect HG1 to do very poorly in terms of efficiency and estimating variance if F is asymmetric and $Q(F) \approx 1.81$. For the negative exponential distribution, $Q(F) \approx 1.805$, so particular problems might be expected here.

The location estimates used in the study we report here are given in Table 1. These represent a portion of the results in a larger study which leads to the same conclusions.

The Monte-Carlo study used a shuffled congruential random number generator to obtain the uniform random deviates. The Box-Muller algorithm was used to obtain the standard normal deviates. Due to time and financial considerations, various sample sizes ($500 \leq N \leq 2000$) were used, so that we also report standard errors.

TABLE 1

A description of the location estimates used in the study

Code	
M	Sample Mean
10%	10% symmetrically trimmed mean
HG1	See text
Switzer	See text
Hub 11	Huber with s_{n1} and $D_{n1}(1)$
Hub 22	Huber with s_{n2} and $D_{n2}(2)$
Hamp 12	Hampel with s_{n1} and $D_{n2}(1)$
Hamp 21	Hampel with s_{n2} and $D_{n1}(2)$
Hub 1J	Huber jackknife variance estimate with s_{1n}
Hamp 1J	Hampel jackknife variance estimate with s_{1n}
Hub 2J	Huber jackknife variance estimate with s_{2n}

Let Z be a standard normal random variable. While a wide range of sampling distributions were investigated (both symmetric and asymmetric), the five presented here are representative.

Type	Code
Z	$N(0,1)$
$Z + .10Z^2$	$.10N^2$
$Z + .50Z^2$	$.50N^2$
Negative Exponential, mean 1.25	NE
.50 Exp(Z)	EXP(Z)

The second ($.10N^2$) is only slightly skewed and was chosen from a larger set ($Z + .05Z^2$, $Z + .25Z^2$, EXP(.10Z), EXP(.25Z)) as a reasonable representative of the class of distributions close to, but not exactly, symmetric. Such data might arise for example from data transformations which only achieve approximate symmetry.

The sample size is $n = 20$. If N is the number of Monte-Carlo experiments and $Y_1, Y_2, \ldots, Y_n$ the realized value of a location estimator, the (standardized) Monte-Carlo variance is

$$\sigma^2 = \frac{n}{N} \sum_1^N (Y_i - \bar{Y}_N)^2 \ .$$

The average value of a variance estimate of $n^{\frac{1}{2}} T_n$ is denoted by $\hat{\sigma}_n^2$. Table 2 presents the values of σ^2 and $\hat{\sigma}_n^2$ and their standard errors.

TABLE 2

The first row for each estimator gives the Monte-Carlo variance σ^2 of the location estimate and the Monte-Carlo average $\hat{\sigma}_n^2$ of the appropriate variance estimate. The second row consists of standard errors.

	N(0,1)		$.10N^2$		$.50N^2$		NE		EXP(Z)	
Code	σ^2	$\hat{\sigma}^2$	σ^2	$\hat{\sigma}^2$	σ^2	$\hat{\sigma}^2$	σ^2	$\hat{\sigma}^2$	σ^2	$\hat{\sigma}^2$
M	1.03	.99	1.07	1.01	1.50	1.47	1.58	1.55	1.10	1.10
	.03	.01	.03	.01	.05	.02	.05	.02	.05	.04
10%	1.09	1.06	1.10	1.07	1.24	1.25	1.31	1.36	.51	.52
	.03	.01	.03	.01	.04	.02	.04	.02	.02	.01
Switzer	1.14	.90	1.15	.90	1.57	.93	1.57	1.04	.52	.34
	.04	.01	.04	.01	.05	.01	.05	.02	.01	.01
HG1	1.09	.96	1.12	.97	1.75	1.05	1.63	1.16	.58	.38
	.03	.01	.04	.01	.05	.02	.05	.02	.01	.01
Hub 11	1.05	1.01	1.06	.99	1.49	1.08	1.50	1.20	.68	.46
	.03	.01	.03	.01	.05	.02	.05	.02	.03	.01
Hub 22	1.05	1.02	1.07	1.02	1.54	1.21	1.52	1.31	.68	.56
	.03	.01	.03	.01	.05	.02	.05	.02	.03	.01
Hamp 12	1.04	1.06	1.06	1.07	1.64	1.24	1.54	1.38	.77	.53
	.03	.01	.03	.01	.05	.02	.05	.02	.03	.01
Hamp 21	1.05	1.01	1.07	1.00	1.60	1.14	1.54	1.28	.74	.45
	.03	.01	.03	.01	.05	.02	.05	.02	.03	.01
Hub 1J	1.05	1.01	1.06	1.02	1.49	1.49	1.50	1.62	.68	.70
	.03	.01	.04	.01	.05	.03	.05	.03	.03	.02
Hamp 1J	1.04	1.01	1.06	1.02	1.64	1.53	1.54	1.57	.77	.81
	.03	.01	.03	.01	.05	.02	.05	.02	.03	.02
Hub 2J	1.05	1.03	1.07	1.05	1.54	1.61	1.52	1.70	.68	.75
	.03	.01	.03	.01	.05	.05	.05	.05	.03	.03

Discussion of the results for location estimates

The tabled results confirm our earlier speculations and might be summarized as follows:

(1) It is difficult to estimate the variance of either HG1 or SWITZER. Jackknifing will probably not help here as the estimates are not smooth functionals of the data.

(2) The two commonly used scales (s_{1n}, s_{2n}) lead to similar results.

(3) The two variance estimates $D_{n1}(i)$ and $D_{n2}(i)$ do quite well for N(0,1) and $.10N^2$, which are symmetric or nearly so, but are not to be trusted for heavily skewed data. Either a jackknife or transformtions appear necessary.

(4) The jackknifed variance estimates are a dramatic improvement on $D_{n1}(i)$ and $D_{n2}(i)$. There is no simple pattern to these figures, but generally the jackknifed variance estimates appear to become more conservative as we pass from N(0,1) to EXP(Z).

The optimistic interpretation given here to the use of the jackknife for variance estimation in robustness of location contrasts with that of Braun (1975), who concluded that jackknifing and robust estimation do not get along very well, and that variance estimates obtained from jackknifing are not reliable. The closest he came to considering our M-estimates was a one-step M-estimate starting at the median and the median absolute deviations from the median (MAD, essentially our s_{1n}). We believe that our results are more favorable to the jackknife because the statistics we have considered are much smoother functionals of the data then are one-step M-estimates starting at (median, MAD), (a conjecture of this type has been made by Gross (1976)), because instead of comparing σ^2 and $\hat{\sigma}_n^2$ as we do here, he compared (in the main) $\log(\hat{\sigma}_n^2)$ to $\log \sigma^2$, which is more unstable when $\sigma^2 \approx 1$, and because asymmetry is unfavorable to (2.6) and (2.7). We used the one-step M-estimate with ψ given by (2.2) and starting at (median, MAD/.6745) and obtained the table

	Z	$.10Z^2$	$.50Z^2$	NE	EXP(Z)
σ^2	1.06	1.08	1.64	1.55	.58
$\hat{\sigma}_n^2$ from (2.6)	.99	.97	.85	1.04	.33
$\hat{\sigma}_n^2$ from jackknife	1.06	1.09	1.90	1.76	.65

Thus, for one-step M-estimates, the jackknife is more unstable than it is with smoother M-estimates.

Regression

We next investigate regression to see if similar phenomena to those found in the previous section continue. Our intuition says that, under asymmetry, the common estimates of variance will be inconsistent for intercept but acceptable for slope. In this section we sketch a proof confirming this conjecture and illustrate small sample results with a Monte-Carlo experiment.

The model we consider is (the use of σ_0 will become clear later)

$$Y_i = \underline{x}_i \, \underline{\beta}_0 + \varepsilon_i \, \sigma_0 \quad (i = 1,\ldots,n) \,, \tag{3.1}$$

where the ε_i are i.i.d. random variables with $E\psi(\varepsilon_1) = 0$ and $\underline{x}_i = (1 \;\; x_{i1} \; \ldots \; x_{ip})$. We consider a version of Huber's Proposal 2 (Huber (1977), p.37), which involves solving the equations

$$n^{-1} \sum_{i=1}^{n} \psi((Y_i - \underline{x}_i \, \underline{\beta}_n)/s_n) = 0 \tag{3.2}$$

$$(n - p)^{-1} \sum_{i=1}^{n} \psi^2((Y_i - \underline{x}_i \, \underline{\beta}_n)/s_n) = a. \tag{3.3}$$

While we will assume the $\underline{x}_i$ are constants, the first two conditions of Lemma 3 (to follow) are reasonable and may be justified by quoting results of Maronna and Yohai (1978). In a subset of their paper, they assume $(y_1,\underline{x}_1)$, $(y_2,\underline{x}_2),\ldots$ is a sample from a distribution function P with $\underline{\beta}_0$, σ_0 solving

$$E\psi((y - \underline{x}\,\beta_0)/\sigma_0) = 0$$

$$E\psi^2((y - \underline{x}\,\beta_0)/\sigma_0) = a \ .$$

Defining $\varepsilon_i = (y_i - \underline{x}_i\,\beta_0)/\sigma_0$, they show essentially that if the ε_i are independent of $\underline{x}_i$ and if $\underline{\beta}_n$, s_n satisfy (3.2) and (3.3) (the latter with (n - p) replaced by n), then $n^{\frac{1}{2}}(\beta_n - \underline{\beta}_0)$ and $n^{\frac{1}{2}}(s_n - \sigma_0)$ are asymptotically normally distributed.

The proof of our result involves Taylor expansions along the lines of Carroll (1978a, 1978b) and is omitted to avoid cluttering up the paper with messy calculations. Recall that our $\underline{x}_i$ are non-stochastic.

<u>Lemma 3</u>. Suppose that

$$n^{\frac{1}{2}}(\underline{\beta}_n - \underline{\beta}_0) = O_p(1)$$

$$n^{\frac{1}{2}}(s_n - \sigma_0) = O_p(1)$$

$$n^{-1}X'X \to V \text{ (positive definite)}$$

$$n^{-1}\sum_{i=1}^{n} \underline{x}_i \to (1,0,0,\ldots,0) = \underline{w}$$

Then, for ψ sufficiently smooth,

(3.4)
$$(a_3V - (a_1a_Y/a_2)\underline{w}'\underline{w})(\underline{\beta}_n - \underline{\beta}_0)/\sigma_0$$
$$= n^{-1}\sum_{i=1}^{n}\{\underline{x}_i'\,\psi(\varepsilon_i) - (a_1/a_2)\underline{w}'(\psi^2(\varepsilon_i) - a)\}$$
$$+ O_p(n^{-1}) \ ,$$

where

$$a_1 = E\,\varepsilon_1\psi'(\varepsilon_1) \qquad a_3 = E\psi'(\varepsilon_1)$$

$$a_2 = 2E\,\varepsilon_1\,\psi(\varepsilon_1)\psi'(\varepsilon_1) \qquad a_4 = 2E\,\psi(\varepsilon_1)\psi'(\varepsilon_1)\ . \qquad \square$$

In simple linear regression with

$$V = \begin{pmatrix} 1 & 0 \\ 0 & v_1 \end{pmatrix}$$

we obtain that if $\underline{\beta}_0' = (\beta_{int}, \beta_{slope})$ and $\underline{\beta}_n' = (\hat\beta_{int}, \hat\beta_{slope})$, then

Corollary. Under the conditions of Lemma 3, for simple linear regression

(3.5) $$(\hat\beta_{int} - \beta_{int})/\sigma_0 = (a_3 - (a_1a_4/a_2))^{-1}n^{-1}\sum_{i=1}^{n}\{\psi(\varepsilon_i) - (a_1/a_2)(\psi^2(\varepsilon_i) - a)\} + O_p(n^{-1}),$$

(3.6) $$(\hat\beta_{slope} - \beta_{slope})/\sigma_0 = (a_3v_1)^{-1}\ n^{-1}\sum_{i=1}^{n} x_{i1}\ \psi(\varepsilon_i) + O_p(n^{-1}). \qquad \square$$

Similar results hold for the general regression problem.

Discussion of Lemma 3. The following conclusions emerge from Lemma 3 (for large samples) and are partially confirmed for small samples in the Monte-Carlo experiment to follow.

(1) It can be shown that the representations (2.8) and (3.5) are equivalent when scale is estimated by s_{2n}.

(2) The representation (3.6) implies that in large samples, even with asymmetry

(3.7) $$\hat\beta_{slope} - \beta_{slope} \sim \text{Normal}\ (0,\ \sigma_0^2E\psi^2(r_1)/a_3^2v_1)\ ,$$

which is precisely the result obtained by Huber (1973) for known scale and p fixed.

(3) Thus, for fixed p as n becomes large, the asymptotic variance *formulas* for $\underline{\beta}_n$ obtained in the symmetric case with known scale are correct in general except when applied to the intercept. We expect to experience the same difficulty in estimating the variance of the intercept that we found in the previous section.

(4) For estimating the variance of the intercept, the equation (3.6) shows that $\hat{\beta}_{slope}$ is sufficiently smooth for jackknifing.

We constructed a Monte-Carlo experiment for n = 20 in the simple linear regression model

$$y_i = 1 + \tfrac{1}{2} x_i + \varepsilon_i,$$

where the values of x are -.95, -.90,...,.90,.95 and the error distributions were as in the previous section. There were N=2000 iterations for least squares and the Huber estimate, and N=1200 iterations for the jackknife. In Table 2 we present the values of $\hat{\sigma}_n^2/\sigma^2$ for $\hat{\beta}_{slope}$ and $\hat{\beta}_{int}$ using least squares, the Huber Proposal 2 with $\psi(x) = \max(-2, \min(x,2))$ and the jackknifed variance estimate for the Huber Proposal 2. The estimate of variance used is essentially $D_{1n}(2)$, and if $r_i = (y_i - \underline{x}_i\, \underline{\beta}_n)/s_{2n}$,

$$D_{1n}(2) = H^2(n-p)^{-1} \sum_{j=1}^{n} \psi^2(r_j)/b^2, \text{ where}$$

$$b^2 = n^{-1} \sum_{j=1}^{n} \psi'(r_j),\ H = 1 + p(1-b)/(bn).$$

In Table 3 we present the average values $\hat{\beta}_{slope}$ and $\hat{\beta}_{int}$ for both least squares and Proposal 2. We conclude

(1) The slope estimate $\hat{\beta}_{slope}$ is hardly influenced by even large dosages of asymmetry. It appears relatively unbiased and its variance can be assessed relatively accurately. In separate Monte-Carlo experiments, we have found that this phenomena extends to a quadratic regression with uniform design (n=30) and to the poison-treatment 3×4 design of Box and Cox (1964).

(2) The intercept estimate $\hat{\beta}_{int}$ is influenced by asymmetry and standard variance formulas will tend to underestimate the true variance (in asymmetric cases) even more than will happen in least squares.

(3) The jackknife appears to be conservative, more so than in the location case. Hinkley (1977) shows that the jackknife as employed here (the so-called balanced jackknife) leads in general to a biased variance estimate even in least squares, unless the design is balanced. In his example, the design is much less balanced than ours and the behavior of the balanced jackknife much worse. Further study of his "weighted" jackknife appears necessary, although for this particular design, our unreported simulation results for the weighted jackknife are virtually identical to those given here for the usual jackknife.

(4) Consider a two-sample location problem

$$Y_{ij} = \mu_j + \varepsilon_{ij} \quad (j = 1,2 \qquad i = 1,\ldots,n_j).$$

The results of the previous section show that a statistic formed by analogy with the t-statistic ($\overline{X}$ replaced by $T_n(X_1,\ldots,X_n)$, etc.) will tend to have a higher type I error than advertised. However, define $(n_1 + n_2)\beta_0 = n_1\mu_1 + n_2\mu_2$, $(n_1 + n_2)\beta_1 = n_2(\mu_1 - \mu_2)$, and

$$Z_i = Y_{i1} \qquad x_i = 1 \qquad i = 1,\ldots,n_1$$

$$Z_i = Y_{i-n_1,2} \qquad x_i = -n_1/n_2 \qquad i = n_1 + 1,\ldots,n_2 \ .$$

Then we have the linear model and, by estimating scale simultaneously through (3.2) and (3.3), Lemma 3 and Corollary 1 tell us that we can consistently estimate the variance of the estimate of $\mu_1 - \mu_2$ and hence obtain proper tests. This of course leads to the usual t-statistic if $\psi(x) = x$.

TABLE 2

The entries are the ratio $\hat{\sigma}_n^2/\sigma^2$ for the appropriate estimates and sampling situations.

Estimation Method	N(0,1)		$.10N^2$		$.50N^2$		NE		EXP(Z)	
	β_{int}	β_{slope}	β_{int}	β_{slope}	β_{int}	β_{slope}	β_{int}	β_{slope}	β_{int}	β_{slope}
Least Squares	.98	.98	.96	.98	.98	.99	.99	1.00	1.00	1.00
Huber	.95	.98	.93	.98	.75	.96	.83	.99	.69	.94
Jackknifed Huber	1.02	1.06	-	-	1.07	1.11	-	-	1.11	1.14

TABLE 3

The entries are the average values of the parameters over the sampling situations.

The true values are $\beta_{int} = 1.00$, $\beta_{slope} = .50$.

	N(0,1)	$.10N^2$	$.50N^2$	NE	EXP(Z)
β_{int}, Least Squares	1.00	1.05	1.00	1.00	1.00
β_{slope}, Least Squares	.49	.49	.48	.48	.48
β_{int}, Huber	1.00	.99	.92	.94	.90
β_{slope}, Huber	.49	.49	.49	.48	.50

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 University Press.

Bickel, P.J. (1978). Using residuals robustly I: tests for heteroscedasticity, nonlinearity. *Ann. Statist.* 6, 266-291.

Box, G.E.P., and Cox, D.R. (1964). An analysis of transformations. *J. R. Statist. Soc. B 26,* 211-252.

Braun, H. (1975). Robustness and the jackknife. Technical report 94, Series 2, Department of Statistics, Princeton University.

Carroll, R.J. (1978a). On almost sure expansions for M-estimates. *Ann. Statist.* 6, 314-318.

Carroll, R.J. (1978b). On almost sure expansions for multivariate M-estimates. To appear in *J. Mult. Analysis.*

Gross, A.M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *J. Am. Statist. Assoc.* 71, 409-416.

Gross, A.M. (1977). Confidence intervals for bisquare regression estimates. *J. Am. Statist. Assoc.* 72, 341-354.

Hinkley, D.V. (1976). On jackknifing in unbalanced situations. Technical Report No. 22, Division of Biostatistics, Stanford University.

Hogg, R.V. (1974). Adaptive robust procedures: a partial review and suggestions for future applications and theory. *J. Amer. Statist. Assoc. 69,* 909-923.

Huber, P.J. (1973). Robust regressions: asymptotics, conjectures, and Monte-Carlo. *Ann. Statist.* 1, 799-821.

Huber, P.J. (1977). *Robust Statistical Procedures.* SIAM, Philadelphia.

Jaeckel, L.A. (1972). The infinitesimal jackknife. *Bell Laboratories Memorandum* MM-72-1215-11.

Naronna, R.A. and Yohai, V.J. (1978). Robust M-estimators for regression with contaminated independent variables, Unpublished manuscript.

Shorack, G.R. (1974). Random means. *Ann. Statist.* *2*, 661-675.

Switzer, P. (1970). Efficiency robustness of estimators. Proc. Sixth *Berkeley Symp. Math. Statistic. Prob.* 1, 283-291.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) On Estimating Variance of Robust Estimators When the Errors are Asymmetric	5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
	6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1177
7. AUTHOR(s) Raymond J. Carroll	8. CONTRACT OR GRANT NUMBER(s) AFOSR-75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina Chapel Hill, North Carolina 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB, Washington, D.C.	12. REPORT DATE August, 1978
	13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Regression, Robustness, Asymmetry, Jackknife, M-estimates

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We investigate the effects of asymmetry on estimates of variance of robust estimator in location and regression problems, showing the heavy skewness of errors can seriously bias the common variance estimates for location and intercept, a problem that can be corrected by jackknifing for location but is more intractable for the intercept in regression. The scale parameters in regression seem not to be as seriously subject to this bias if the sample size is large compared to the number of parameters.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BIOMATHEMATICS TRAINING PROGRAM