1995

Applications of Some Robust Statistics in Forestry.

Kunjin Shi
Louisiana State University and Agricultural & Mechanical College

APPLICATIONS OF SOME ROBUST STATISTICS IN FORESTRY

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The School of Forestry, Wildlife, and Fisheries

by
Kunjin Shi
M.S., Beijing Forestry University, 1989
Master of Applied Statistics, Louisiana State University, 1993
May 1995

UMI Number: 9530496

ACKNOWLEDGMENTS

I would like to thank the School of Forestry, Wildlife, and Fisheries for providing me the opportunity to pursue a graduate education in forest biometrics. The research assistantship I received from the School provided the economic support I needed during my study. I appreciate all the help and friendship from the faculty, staff and students of the School.

This study was conducted under the direction of my major professor, Dr. Quang V. Cao. Without his invaluable guidance and support, it would have been impossible for me to finish this research. I am grateful for all his guidance and help in both academic and non-academic areas.

I would also like to thank the members of my graduate advisory committee, Drs. Jim Chambers, Thomas J. Dean, James P. Geaghan, Richard Pace, and Mark Williams for their guidance, encouragement and generous help. I am especially thankful for their patience and the extra guidance they have given me as a foreign student.

I would especially like to thank my wife Xuejian Guo for her continuous understanding. Her encouragement and support have been a great source of my strength. I would also like to share this pleasure with my lovely daughter Debbie Dingyi Shi.

TABLE OF CONTENTS

ABSTRACT

All statistical procedures are based on a set of assumptions, such as normality, independence and linearity. In practical applications, these assumptions can rarely be satisfied completely. Deviation from classical parametric assumptions may result in loss of efficiency or even lead to misleading conclusions. Robust statistics that have been extensively studied in the past two decades provide alternatives to deal with such problems.

This research was designed to investigate the applicability of robust estimation of population means and robust linear regression in forestry. Five robust estimators, Huber's minimax estimator, Hampel's three parameter redescending estimator, Andrew's wave estimator, Tukey's biweight estimator, and sample median were examined for their performance in estimating population means. Simulations on four families of distributions, beta, gamma, lognormal, and Weibull, suggested that the five robust estimators have a bias problem in estimating means of skewed populations. Analyses of simulated data revealed that magnitudes of robust estimator bias were closely related to a proposed robust sample skewness measure, $Skew_{\alpha}$. Regression models were developed to predict bias of the five robust estimators from $Skew_{\alpha}$. The predicted bias was then used in constructing a bias corrected robust estimator from each of the five estimators. The modified estimators were evaluated against corresponding original estimators and the sample mean on simulated data from four families of distributions and also on a forestry data set. The bias-corrected robust

estimators were better than the original estimators in terms of bias and mean square error.

Two robust linear regression procedures, least median of squares and least trimmed squares, were used to fit two individual tree volume equations on nine data sets and two yield models on one data set. The two robust regressions were evaluated against ordinary least squares based on prediction capabilities. For most of the data sets the robust procedures and least squares method produced similar prediction error values. For data sets that contained extreme outliers, the two robust procedures yielded smaller prediction error values than least-squares estimation.

INTRODUCTION

Statistical inference requires two bases; one is a sample of observations, and the other is assumptions about the underlying data generating mechanism, often called model assumptions. Without either of the two bases, no statistical inference could be properly made (Huber 1981).

According to Tukey (1970), there have been three important developments in statistics: classical parametric statistics, nonparametric statistics, and robust statistics. Most classical statistical procedures are based on two assumptions: the sample observations are independently and identically distributed, and the underlying distribution is normal. Given these properties, some standard, well-understood paradigms, such as maximum likelihood procedure, can be applied to make required inferences. In practice, one rarely knows exactly how data are generated, and only has some vague ideas about data distribution, e. g. symmetric, bell-shaped instead of exactly Gaussian. If classical parametric procedures are insensitive to deviations from the assumptions, they still can be used with confidence. Unfortunately, if the assumptions are not completely satisfied, the classical procedures may lose efficiency or may even provide unreasonable conclusions (Tukey 1960, Huber 1981, Hampel et al. 1986).

Loss of efficiency refers to increased variance of an estimator compared with another estimator. When testing a hypothesis, efficiency loss means decreased power of a test while keeping the significance level fixed. The sample mean is commonly

used to estimate the population location parameter. Problems may arise if there is even a single outlier in the sample. If the outliers are not very large, the sample mean may still give a reasonable answer, but the variance of the estimate is expanded by outliers which causes efficiency loss. If the outliers are large, the arithmetic mean might be far away from the true population location parameter.

To deal with this problem, the nonparametric approach emerged. The goal of nonparametric procedures is to make as few assumptions about the data as possible and still get the answers to some specific questions. The classical parametric methods are at one extreme of relying on strict assumptions, while the nonparametric approaches move to the other extreme by not using available information such as our vague knowledge of data distributions. As a result, in cases where classical assumptions hold entirely or even approximately, the analogous classical parametric procedures are generally more efficient than their nonparametric counterparts (Singhal and Sheather 1989).

The first two approaches represent two extremes with respect to model assumptions. Robust statistics are somewhere between these two extremes, combining the virtues of both approaches. When model assumptions are entirely met, robust estimators are nearly as efficient as classical parametric. If the assumptions are not met exactly, robust procedures are considerably more efficient overall. In the cases of large outliers, robust procedures can prevent misleading results that could be given by classical parameter estimators (Huber 1981, Hampel et al. 1986).

The importance of preventing misleading results is obvious. Efficiency gains are also highly desired. One would like to use the data at hand effectively to get precise estimates of the parameters of interest. Efficiency gains usually mean more precise estimates with the same data set, or estimates of the same precision with less data, thus saving labor, time and money.

In current statistical practice, robust procedures are playing an increasingly important role. They have been used successfully in many fields, such as biology (Marazzi et al. 1988), economics (Koenker 1982, Kassab 1990), agriculture (Swinton and King 1991), and astronomy (Freeman et al. 1992). However, two major problems have slowed the pace of their application. First, robust procedures are not available in some major statistical software packages, such as SAS, BMDP and SPSS. Second, there is usually more than one robust procedure applicable to the same task, forces users to choose among competitive procedures (Hampel et al. 1986).

Forestry data are usually very expensive and time-consuming to collect, especially growth and yield data. Therefore, the data should be used as effectively as possible. Since forest growth is affected by many factors and the measuring conditions are usually rough, forestry data tend to vary excessively, and gross errors are likely to occur. The observations suggest that robust procedures should be very valuable in forestry. However, a search of forestry literature revealed that we have still been too comfortable with the classical assumptions of "identically and independently distributed as normal" and their close relatives, using classical parametric methods

almost exclusively. It would be advantageous to realize that the strict assumptions of classical parametric procedures might not be quite true and to take proper actions to deal with the resulting problems.

The objective of this study was to explore the applicability of robust estimation of population means and robust linear regression in forestry. Efforts were made to correct bias of five robust estimators of population mean when data were not symmetrical. Two robust linear regression procedures were employed to estimate parameters of two individual tree volume equations and two forest yield models. The performance of the robust procedures was compared with commonly-used least squares procedure.

The objectives of this study were to

(1) correct the bias problem of five existing robust estimators of population means when data were not symmetrical,

(2) evaluate original and bias-corrected robust estimators of population means, and

(3) evaluate two robust linear regression procedures in estimating parameters of two individual tree volume equations and two yield models.

REVIEW OF LITERATURE

The term "robustness" was first introduced into statistics by Box (1953). A statistical procedure is called "robust" if it is insensitive to departures from the model assumptions. In the broad sense, robust procedures probably date back to the prehistory of statistics. Looking at data and checking conspicuous observations is a step towards robustness; excluding highly deviant values is an informal robust procedure. The median is a robust estimator of location. Using sample median instead of the sample mean when data are distributed with long tails is a robust method.

A systematic search for robust procedures did not start until quite recently. Tukey (1960), in summarizing earlier work of his group in the 1940s and 1950s, demonstrated the drastic nonrobustness of the arithmetic mean and also investigated some useful robust alternatives. His work made robust estimation a general research area. The first attempts at a manageable, rather realistic, and comprehensive theory of robustness were made by Huber (1964) and Hampel (1968).

Since then much work has been done in many areas, such as location and scale estimation (Andrews et al. 1972, Bickel 1976, Huber 1981, Hampel et al. 1986), hypothesis testing (Ringland 1983, Singhal and Sheather 1989, Marianthi et al. 1991), linear and nonlinear regression (Rousseeuw and Leroy 1987, Birkes and Dodge 1993), and analysis of variance (Schrader and Hettmansperger 1980, Tan and Tabatabai 1985). The following review will mainly focus on robust estimation and regression.

Deviations from Classical Parametric Models

Most classical parametric models define the joint distribution of the observations as independently, identically, and normally distributed. Thus deviations from the distribution form (usually normal), identity, and independence should be considered. The majority of robustness studies have focused on deviations from the assumed distribution, especially the normal distribution.

Experience with real life data reveals that some populations from many fields do not behave in a normal fashion. This was known by Bessel (cf. Hampel et al. 1986) shortly after the invention of least squares. A famous citation about normality is "Every one believes in the normal law, the experimenters because they imagine it a mathematical theorem, and the mathematicians because they think it an experimental fact" (cf. Stigler 1977). An extreme statement was made by Geary (1947): "Normality is a myth; there never was, and never will be a normal distribution." This might be an overstatement, but the fact is that nonnormal distributions are more prevalent in practice, and to assume normality might lead to erroneous statistical inferences.

There are mainly three reasons why a parametric model does not hold exactly.

1. The occurrence of gross errors. Gross errors are errors due to a source of deviations that act only occasionally but are quite powerful. They are the most frequent causes for outliers. Some reasons for gross errors are copying errors, interchange of two values or groups of values in a structured design, inadvertent

observation of a member of a different population, equipment failure, and also transient effects. A single unnoticed gross error can ruin a statistical analysis (for example least squares). Some sources for gross errors, such as keypunch errors or wrong decimal points, may easily change values by orders of magnitude. With the modern trend of putting masses of data into the computer, outliers can easily escape attention if no precautions are taken. Previous research showed that several percentages of gross errors are rather common. Paul Olmstead (cited by Tukey 1962) maintained that engineering data typically involves about 10% "wild shots." After reviewing the frequency of gross errors in the literature, Hampel et al (1986) concluded that 1-10% gross errors in routine data seem to be more the rule rather than the exception.

2. Rounding and grouping. All data are measured with only limited accuracy and are thus basically discrete; moreover, the data are often rounded, grouped, or classified even more coarsely. There can also be small systematic but localized inaccuracies in the measuring scale.

3. Distribution approximation. All too often, distribution assumptions are based on experience, asymptotic properties, central limit theorem, or even only for convenience's sake. They have been conceived as an approximation in the first place. Even large sets of accurately measured data still tend to show small but noticeable deviations from the normal model (Jefrey 1932, Bickel 1976). Jefrey (1961) analyzed nine long series of careful observations made under uniform circumstance, and

concluded that the errors might well be described by the t distribution with about 5 to 9 degrees of freedom.

Nonrobustness of Classical Parametric Procedures

Because actual data may deviate from usual model assumptions, the statistical tools used to analyze them should be robust to the deviations. Unfortunately, most classical procedures do not have the desired property. Commonly used sample mean and standard deviation are examples. If data are normally distributed and no gross errors are present in the sample, then the sample mean and standard deviation are the optimal estimators of the population location and scale parameter. For data from a t distribution with 9 degrees of freedom, the asymptotic efficiencies of sample mean and standard deviation reduce to 93% and 83%, respectively. If the degree of freedom reduces to 5, the corresponding asymptotic efficiencies become 80% and 40% (Hampel et al. 1986).

Tukey (1960) considered a situation where the majority of the sample was drawn from a standard normal distribution, while a small proportion of the sample was from a normal distribution with the mean zero and standard deviation 3. Let x be the standard normal random variable, and ε a small constant ($0 \leq \varepsilon \leq 1$), then the mixed distribution can be written as $F(x) = (1-\varepsilon)\ \Phi(x) + \varepsilon\ \Phi(x/3)$, where Φ is the standard normal cumulative distribution function. This model can be thought as a distribution contaminated with gross errors of a realistic amount. Tukey showed that the

asymptotic efficiency of the mean decreases quickly from 100% (for $\varepsilon = 0$) to about 70% (for $\varepsilon = 0.10$). Two outliers in a thousand observations can reduce the asymptotic efficiency of the standard deviation from 100% to 88%. Tukey also showed that it is virtually impossible to distinguish between $\varepsilon = 0\%$ and $\varepsilon = 1\%$ even with a sample size of 1000, unless one or a few points arising from the contaminating distribution $\Phi(x/3)$ are rather extreme. It has also been shown that the chi-square test and F-test for variance, as well as analysis of variance, are highly susceptible to slight nonnormality, in the sense of inaccurate level or low power (Hampel et al. 1986, Singhal and Sheather 1989).

As classical parametric procedures are not robust to deviations from assumptions, and deviations from parametric models do exist, robust procedures are needed. One might suggest a two-step approach: (1) reject outliers using some subjective or objective criteria, and (2) go on using classical parametric procedures. It may not be very difficult to reject extreme outliers, but there are some problems rejecting mild outliers. Sometimes, it is hard to determine whether an observation is an outlier or not. In the case of high-dimension problems, detecting and rejecting outliers become more difficult or even impossible. For heavy-tailed distributions, robust procedures can prevent avoidable efficiency loss compared to the two-step approach.

Robust Estimation of Population Means

A robust estimator should have two properties: resistance and robustness of efficiency. An estimator is resistant if it is affected to only a limited extent either by a small number of gross errors or by any number of small rounding and grouping errors. An estimator has robustness of efficiency over a range of distributions if its variance (or, for biased estimators, its mean squared error) is close to the minimum for each distribution. Robustness of efficiency guarantees that the estimator is good when repeated samples are drawn from a distribution that is not known precisely.

There are two main approaches to robust estimation: Huber's minimax approach (Huber 1964, 1981) and Hampel's infinitesimal approach (Hampel 1968, Hampel et al. 1986). Both approaches assume a parametric model for the observations and then try to construct estimators that do well over the neighborhood of the assumed model. The optimality problem of parametric statistics is modified by introducing, in addition to the classical consistency and efficiency requirements, a robustness condition that refers to the behavior of an estimator in the neighborhood of the assumed model. The optimal solution entails a trade-off between efficiency and robustness (Peracchi 1990).

Hampel's infinitesimal approach focuses on the asymptotic behavior of an estimator in an infinitesimal neighborhood of a given model (Hampel 1968, Hampel et al. 1986). He introduced the concept of an influence function that measures the effect, on the asymptotic bias of an estimator of an arbitrarily small contamination of the

assumed statistical model (Hampel 1971, 1974). He also defined the "breakdown point" as an important global measure of robustness. Breakdown point is the smallest percentage of contamination in the data that may cause the estimator to take on arbitrarily large values. A high breakdown point indicates that an estimator can tolerate high percentage of gross errors. The maximum breakdown point an estimator can reach is 50%. These concepts are very helpful for developing and evaluating new robust procedures.

Huber's aim was to optimize the worst that can happen over the neighborhood of the model, as measured by the asymptotic variance of the estimator. He used the logic of a game between Nature and a statistician: Nature chooses a distribution from the neighborhood of a model, the statistician choose an estimator, and the gain for Nature and loss for the statisticians is the asymptotic variance. Instead of believing in a strict parametric model of the form $G(x, \theta)$ for known G, where θ is the location parameter to be estimated, Huber assumed that a fraction ε ($0 \leq \varepsilon \leq 1$) of the data may consist of gross errors with an arbitrary (unknown) distribution $H(x, \theta)$. The distribution underlying the observation is thus $F(x, \theta) = (1-\varepsilon)\, G + \varepsilon\, H$. This is often called the gross error model (Huber 1964, 1981).

On the basis of the gross error model, Huber introduced a class of estimators, called "M-estimators". These estimators are a generalization of maximum likelihood estimators. Given that x_i is distributed independently with density $f(x_i, \theta)$, the maximum likelihood estimator T of a parameter θ is obtained by minimizing

$$-\sum \ln f(x_i, T)$$

where ln denotes natural logarithm. This is equivalent to solving

$$\sum f'(x_i, T) / f(x_i, T) = 0 \qquad \text{where } f'(x_i, T) \text{ exists}$$

Let $\rho(x_i, T)$ be a function of x and let $\psi(x_i, T)$ be the derivative of $\rho(x_i, T)$ with respect to T. Huber defined the M-estimator as the value that minimizes

$$\sum_{i=1}^{n} \rho(x_i, t)$$

or solving

$$\sum_{i=1}^{n} \psi(x_i, t) = 0$$

without assuming that ρ and ψ are of the form $-\ln f(x_i, T)$ and $-f'(x_i, \theta) / f(x_i, \theta)$, respectively, for any probability density function $f(x_i, \theta)$. Different forms of $\rho(x_i, T)$ or $\psi(x_i, T)$ define different M-estimators. Sample mean and sample median are special cases of M-estimators.

An M-estimator of location should be location-and-scale equivalent. Location-and-scale equivalent means that if one multiplies the whole sample by a nonzero constant and then shift the result by another constant, the estimator of location follows the same change. An auxiliary estimator of scale S_n that is a function of $x_1, \ldots, x_n$, can be used to rescale the x_i to get the centered and rescaled observations:

$$\mu_i = \frac{x_i - t}{S_n}$$

Then, the M-estimate is the value of t that minimizes $\sum \rho(\mu_i)$ or satisfies $\sum \psi(\mu_i)=0$. Following Huber's approach, many similar procedures have been proposed. Among them, three methods have received broad attention. They are Hampel's linear redescending estimator, Andrew's wave estimator, and Tukey's biweight estimator (Andrews et al. 1972, Stigler 1977, Rocker et al. 1982, Hoaglin et al. 1983).

Robust Estimation of Linear Models

The linear regression model is one of the most widely used tools in statistical analysis, and the ordinary least square (OLS) method is a popular estimation technique for this kind of model. However, in spite of its mathematical elegance and computational simplicity, the least-square estimator suffers a dramatic lack of robustness (Huber 1973, Hampel et al. 1986, Birkes and Dodge 1993). In regression analysis, outliers may occur in the response variable and predictor variables (often called leverage points). A single outlier of either type may totally spoil an ordinary least squares analysis. A good robust regression procedure should be insensitive to both types of outliers.

Huber (1973) extended his results on robust estimation of a location parameter to the case of linear regression. His method is robust with respect to regression residual distribution, but is sensitive to leverage points in the predictors. Its breakdown point is never greater than 25%, far less than the maximum value of 50%. Breakdown point is the smallest percentage of contamination in the data that may cause the

estimated model to fail to follow the trend formed by the majority of data. A high breakdown point indicates that an estimator can tolerate high percentage of gross errors. Maronna and Yohai (1981) defined the generalized M-estimator for linear models, and the breakdown point cannot exceed 1/p where p is the dimension of the parameter space, thus the breakdown point becomes very low when the number of parameters is even moderately large (Rousseeuw and Leroy 1987).

Recent results on robust regression have centered on estimates with a high breakdown point. Siegel (1982) proposed the first robust regression estimator with a 50 % breakdown point, the repeated median. However, the repeated median estimator is not affine equivariant. Affine equivariant is also called equivariant with respect to linear transformation of the independent variables. It means that a linear transformation of independent variables should transform the estimator accordingly, and will not affect the estimated value of the dependent variable.

Rousseeuw (1984) introduced the least median of squares (LMS). Rather than minimizing the sum of the squared residuals as least squares regression does, least median of squares minimizes the median of the squared residuals. Least median of squares regression is affine equivariant and has a very high break down point of almost 50%. That is almost half of the data can be corrupted in an arbitrary fashion and the least median of squares estimates still follow the trend of the majority of the data.

There are, however, two disadvantages of least median of squares. The first disadvantage is that least median of squares has low asymptotic efficiency. The second disadvantage is that there is no feasible algorithm to compute the actual least median of squares for models with large number of many predictors. For a model with p predictors (including the intercept term, if present), subsamples of size p are taken from the data set. Each of these subsamples is used to get a trial set of coefficients. The set of coefficients that yields the smallest median of squares forms the estimates of the regression. For models with many predictors the number of all subsamples of size p can grow very large, which increases computing time rapidly. In practice, only a certain number of subsamples are randomly drawn from the data set. A trial set of coefficients is estimated for each of the subsamples. The set of coefficients that gives the smallest median of squares is considered the estimates of the model parameters. Usually the number of subsamples is limited to 3,000. For models with nine or fewer independent variables, the probability of getting estimates with 50% breakdown point is larger than 99%. This probability drops sharply as the number of independent variables grows beyond ten.

Two more efficient variants of the LMS have been proposed. One is least trimmed squares (LTS). The objective that least trimmed squares minimizes is the sum of the h smallest squared residuals, where h is the largest integer that satisfies

$$h < (n/2) + (p+1)/2,$$

where n is the number of observation in the data set, and p is the number of parameters to be estimated (Rousseeuw 1984). Statistically least trimmed squares is a more efficient objective than least median of squares. Least trimmed squares shares the same computing time problem with least median squares (Birkes and Dodge 1993).

The other variants of the LMS is called S-estimator. The objective that S-estimator minimizes is a certain measure of the scale or variation of the residuals (Rousseeuw and Yohai 1984, Yohai and Zamar 1988, Birkes and Dodge 1993). van Zomeren (1987) showed that S-estimator requires tremendous computing time.

BIAS-CORRECTED ROBUST ESTIMATION OF POPULATION MEANS

Four common robust M-estimators of population means were used in this study. Their $\rho(\mu)$ and $\psi(\mu)$ functions are listed as follows.

1. Huber (k)

$$\rho(\mu)=\begin{cases}\frac{1}{2}\mu^2 & |\mu|\le k\\ k|\mu|-\frac{1}{2}k^2 & |\mu|>k\end{cases}$$

$$\psi(\mu)=\begin{cases}\mu & |\mu|\le k\\ k\,\mathrm{sgn}(\mu) & |\mu|>k\end{cases}$$

2. Hampel (a, b, c)

$$\rho(\mu)=\begin{cases}\frac{1}{2}\mu^2 & |\mu|\le a\\ a|\mu|-\frac{1}{2}a^2 & a<|\mu|\le b\\ (c-b)\frac{a}{2}\left[1-\left(\frac{c-|\mu|}{c-b}\right)^2\right] & b<|\mu|\le c\\ ab-\frac{1}{2}a^2+(c-b)\frac{a}{2} & |\mu|>c\end{cases}$$

$$\psi(\mu)=\begin{cases}\mu & |\mu|\le a\\ a\,\mathrm{sgn}(\mu) & a<|\mu|\le b\\ a\frac{c-|\mu|}{c-b}\mathrm{sgn}(\mu) & b<|\mu|\le c\\ 0 & |\mu|>c\end{cases}$$

3. Tukey's biweight (c)

$$\rho(\mu)=\begin{cases}\frac{1}{2}\left[1-\left(1-\mu^2\right)^3\right] & |\mu|\le c\\ \frac{1}{6} & |\mu|>c\end{cases}$$

$$\psi(\mu)=\begin{cases}\mu(1-\mu^{2)^2} & |\mu|\le c\\ 0 & |\mu|>c\end{cases}$$

4. Andrews' wave (c)

$$\rho(\mu)=\begin{cases}\frac{1}{\pi^2}\left(1-\cos\pi\mu\right) & |\mu|\le c\\ \frac{2}{\pi^2} & |\mu|>c\end{cases}$$

$$\psi(\mu)=\begin{cases}\frac{1}{\pi}\sin\pi\mu & |\mu|\le c\\ 0 & |\mu|>c\end{cases}$$

In addition to the above four estimators, sample median was also included as a robust estimator of population means. The performances of the five robust estimators of population means were compared with the traditional sample mean, the least squares estimator of population means. The $\psi(\mu)$ functions of the six estimators are shown on Figure 1.

Bias of M-Estimators

The performance of M-estimators have been evaluated for many symmetric distributions, such as normal, student - t and Cauchy. They have smaller variance than

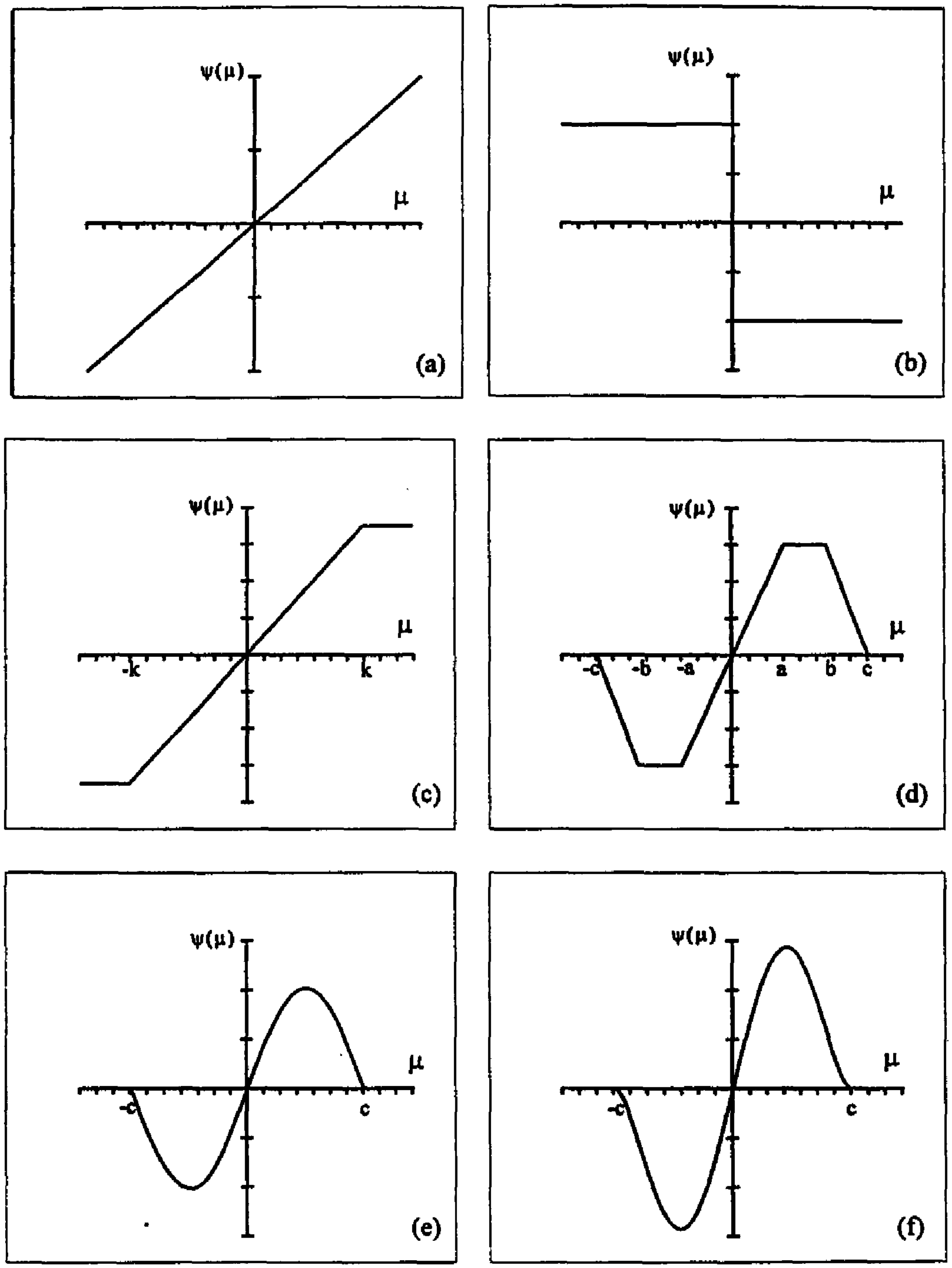


Figure 1. Plots of $\psi(\mu)$ functions for (a) sample arithmetic mean, (b) sample median, (c) Huber, (d) Hampel, (e) Andrews, and (f) Tukey estimators.

sample mean for heavy tailed distributions, such as student - t with small degree of freedom and Cauchy (Andrews et al. 1972, Huber 1981, and Hampel et al. 1986). However, for skewed distributions, M-estimators exhibit large bias in estimating population means. Simulations can be used to demonstrate the biases.

For simplicity's sake, a commonly used member of the Huber estimator family, the Huber estimator with k = 1.5, noted as $Huber_{15}$, was studied first. To demonstrate the performance of $Huber_{15}$ on skewed populations, a family of distributions with flexible shape is needed. The beta distribution is a good candidate for this purpose because it can have a wide range of skewness, negative or positive, depending on different combinations of the parameters. Table 1 lists the nine parameter settings simulated to demonstrate the bias problem of $Huber_{15}$.

The skewness of the simulated distributions ranges from -4 to 4. The simulated sample size was 50 and replication size was 500. The sample mean and $Huber_{15}$ estimate were calculated for each simulated sample. Biases were calculated as follows:

$$Huber_{15}\ \text{Bias} = \text{Population Mean} - Huber_{15}$$

$$\text{Sample Mean Bias} = \text{Population Mean} - \text{Sample Mean}$$

Figure 2 shows plots of sample mean bias and $Huber_{15}$ bias vs. population skewness. Simulation results on nine beta distributions were used to construct the plots. The first plot shows that the average bias of the arithmetic mean is very close to

Table 1. Population parameters of the simulated beta distributions [1/]

Parameter α	Parameter β	Population Mean	Population SD	Population Skewness[2/]	Population Kurtosis[3/]
10.00	0.17	0.9834	0.0383	-4	26.4592
10.00	0.29	0.9715	0.0495	-3	15.8833
10.00	0.62	0.9418	0.0687	-2	8.3375
10.00	1.86	0.8432	0.1014	-1	3.9308
10.00	10.00	0.5000	0.1091	0	2.7391
1.86	10.00	0.8432	0.1014	1	3.9308
0.62	10.00	0.9418	0.0687	2	8.3375
0.29	10.00	0.9715	0.0495	3	15.8833
0.17	10.00	0.9834	0.0383	4	26.4592

[1/] The probability density function of beta distribution is

$$p_X(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad (0 \le x \le 1)$$

[2/] Skewness = $E[(X - \mu)^3] / \{E[(X - \mu)^2]\}^{3/2}$
where X is a random variable, $\mu = E(X)$.

[3/] Kurtosis = $E[(X - \mu)^4] / \{E[(X - \mu)^2]\}^2$
where X is beta random variable, $\mu = E(X)$.

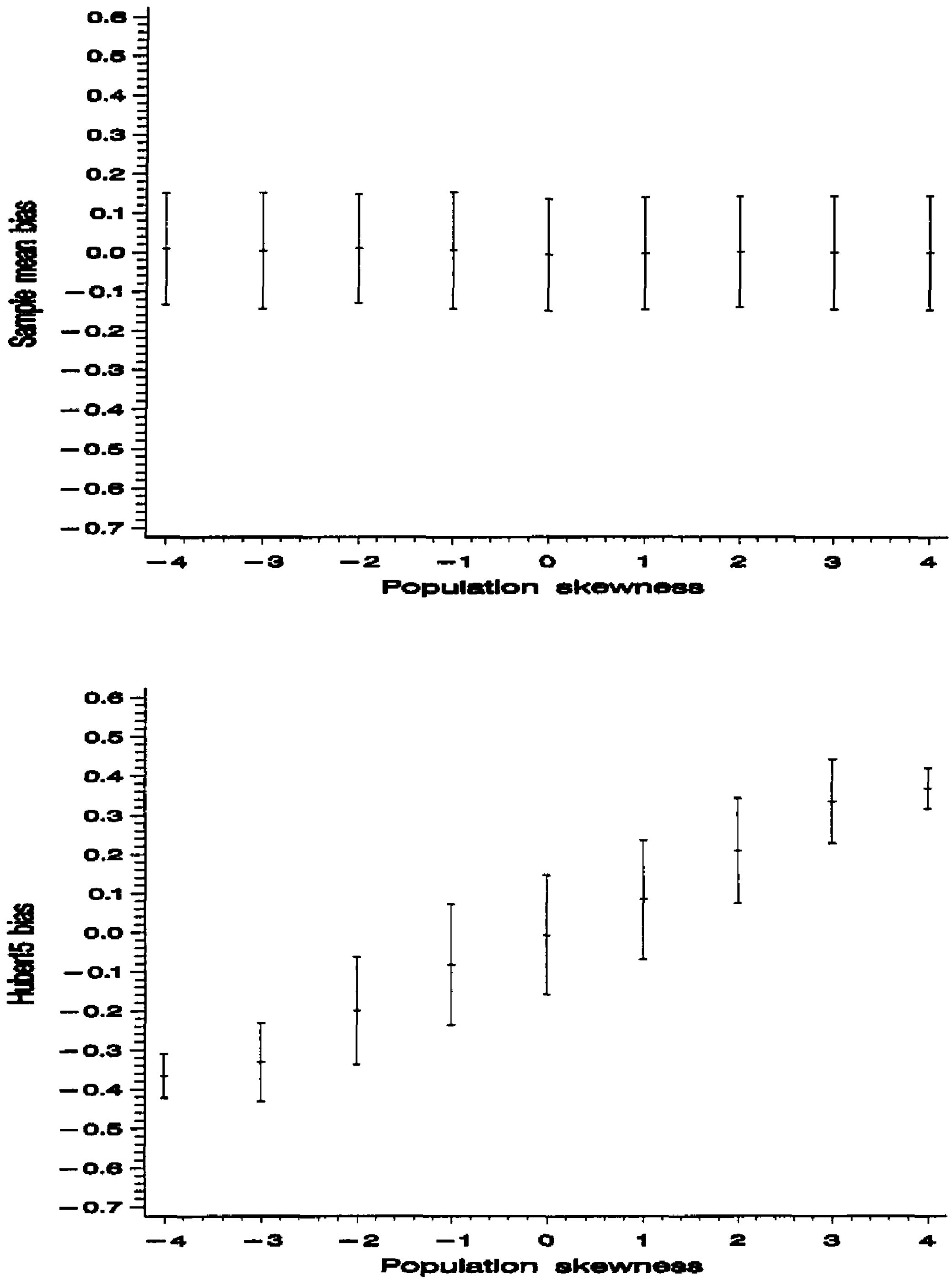


Figure 2. Bias (mean ± standard deviation) of sample mean and Huber_{15} vs. population skewness for beta distributions with sample size 50

zero for all nine populations with various skewness properties. On the other hand, the absolute value of average Huber_{15} bias increases steadily with the absolute value of population skewness.

Modified Robust Estimators

It has been demonstrated in the above simulations that Huber_{15} is biased for estimating means of one family of asymmetric distributions. The magnitude of the bias is related to population skewness. Therefore, the relationship between estimation bias and a robust skewness measure needs be established. If estimation bias can be predicted from a skewness measure, the predicted bias can be removed from the original robust estimator, resulting in a modified robust estimator. For Huber_{15}, the modified robust estimator can be expressed as

$$\text{Modified Huber15} = \text{Huber15} + \text{Bias Correction}$$

The bias correction term in the above formula is a function of a population skewness measure. It should satisfy the following requirements.

1. To retain the robust property for the modified Huber_{15}, the bias correction term itself should be robust.

2. The bias correction term should be zero for symmetric distributions because Huber_{15} is unbiased for means of symmetrically distributed populations.

Development of a new robust skewness measure

Sample skewness is defined as

$$\text{Sample skewness} = \frac{n}{(n-1)(n-2)s^3}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

where n is the sample size, s is sample standard deviation and $\bar{x}$ is sample mean. Sample skewness measures the degree of asymmetry of a sample. Sample skewness should be zero for symmetric samples.

As deviations of all observations from sample mean are raised to the third power and then summed for in calculating sample skewness, extreme outliers affect the skewness measure dramatically. In other words sample skewness is not robust to outliers and does not satisfy the robust requirement.

Let p_α stand for the α percentile of a sample. Then the median of a sample can be denoted as p_{50}. Suppose $0 \leq \alpha < 50$. A new skewness measure, Skew_α, can be defined as:

$$\text{Skew}_\alpha = \frac{p_{100-\alpha} - 2\,(p_{50}) + p_\alpha}{p_{100-\alpha} - p_\alpha}$$

Skew_α is therefore a measure of skewness or asymmetry. It is robust to α percent of outliers at each tail of the sample, and is scale equivariant. For symmetric samples $\text{Skew}_\alpha = 0$. For positively skewed samples, $\text{Skew}_\alpha > 0$, and for negatively skewed samples, $\text{Skew}_\alpha < 0$. The range of Skew_α is from -1 to 1. The rationale for the definition of Skew_α is explained as follows.

For a symmetric distribution, the median, p_{50}, is at the center of the distribution. P_α and $p_{1-\alpha}$ should be symmetric to the median . The distance between p_α and p_{50}, that is p_{50}-p_α, should be the same as the distance between $p_{100-\alpha}$ and p_{50}, or $p_{100-\alpha} - p_{50}$.

$$p_{50} - p_\alpha = p_{100-\alpha} - p_{50}$$

$$(p_{100-\alpha} - p_{50}) - (p_{50} - p_\alpha) = 0, \text{ and}$$

$$p_{100-\alpha} - 2\,(p_{50}) + p_\alpha = 0$$

For skewed distribution, the distance from p_α to median is not the same as the distance from $p_{100-\alpha}$ to the median. So the difference between these two distances can be used as a skewness measure. For samples that are skewed to the right,

$$p_{50} - p_\alpha < p_{100-\alpha} - p_{50}$$

$$(p_{100-\alpha} - p_{50}) - (p_{50} - p_\alpha) > 0, \text{ and}$$

$$p_{100-\alpha} - 2\,(p_{50}) + p_\alpha > 0$$

The value of the right-hand-side term increases with sample skewness. On the other hand, for samples that are skewed to the left,

$$p_{50} - p_\alpha > p_{100-\alpha} - p_{50}$$

$$(p_{100-\alpha} - p_{50}) - (p_{50} - p_\alpha) < 0, \text{ and}$$

$$p_{100-\alpha} - 2\,(p_{50}) + p_\alpha < 0$$

The absolute value of the left-hand-side term of the above equation increases with sample skewness.

Note that when all observations in a sample are multiplied by a constant, the skewness property of the sample is not changed. Thus a skewness measure should be scale equivarient. However the value of $p_{100-\alpha}$ - 2 (p_{50}) + p_α does not satisfy this requirement. It needs to be divided by a robust measure of sample variation to force it to be scale equivarient. In this case, the sample inner range

$$Range_\alpha = p_{100-\alpha} - p_\alpha$$

is a good candidate.

Figure 3 shows plots of sample mean bias and $Huber_{15}$ bias vs. a robust sample skewness measure $Skew_{05}$. The plots were constructed from the simulation results of the nine beta distributions (Table 1). Deviations of sample mean from the true population mean were distributed about evenly around zero. The bias from $Huber_{15}$ estimates formed a band that was negative for negative $Skew_{05}$, near zero for $Skew_{05} = 0$, and positive for positive $Skew_{05}$.

Analysis of $Huber_{15}$ bias

Figure 3 shows that there was a correlation between $Huber_{15}$ bias and the robust sample skewness measure. However, the relationship was not very close. A

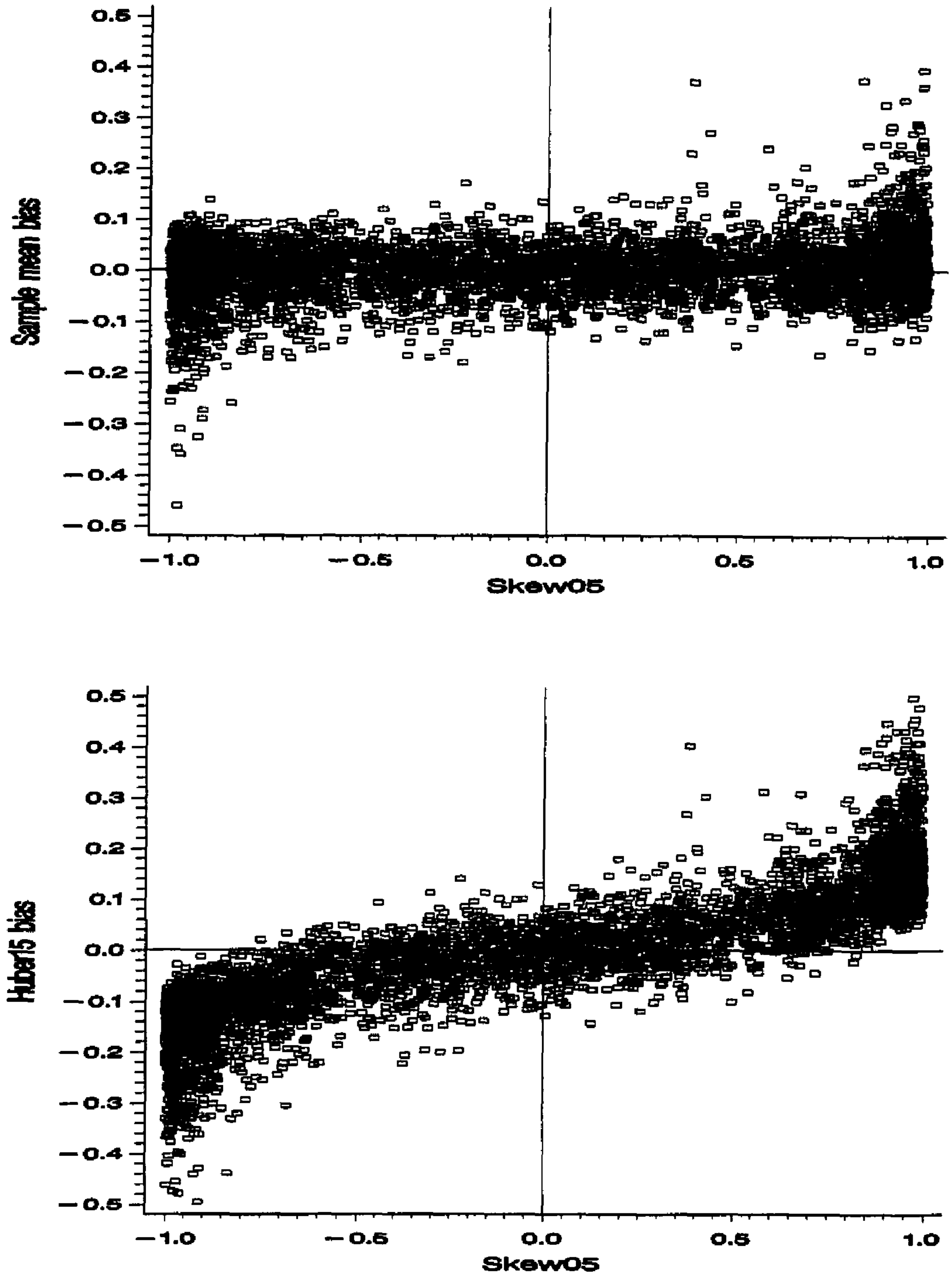


Figure 3. Bias of sample arithmetic mean and Huber_{15} vs. Skew_5 for beta distributions with sample size 50

careful analysis indicated that the variation shown on the $Huber_{15}$ bias vs. $Skew_{05}$ plot came from two sources.

Let μ be the true population mean, $Huber_{15}$ bias is defined as

$$Huber_{15} \text{ bias} = \mu - Huber_{15}$$

The right-hand-side term can be partitioned into two parts:

$$Huber_{15} \text{ bias} = (\mu - \bar{x}) + (\bar{x} - Huber_{15})$$

The first part, $(\mu - \bar{x})$, is the difference between population mean and sample mean. The mean of a sample is not likely to be exactly the same as the mean of the population from which the sample is drawn. The difference comes from sample variation and can be called sample bias. This part of the $Huber_{15}$ estimator is not caused by sample skewness.

$Huber_{15}$ and other robust estimators give very low weights or even zero weights (truncation) to extreme observations. This in turn causes bias for asymmetric samples and the bias can be named truncation bias. Truncation bias is the second component of the Huber15 partition. It is the difference between sample mean and $Huber_{15}$, or $(\bar{x}$ - $Huber_{15})$. The magnitude of the truncation bias is closely related to degree of sample asymmetry. For symmetric samples, the truncation bias should be zero.

Only one of the two components of $Huber_{15}$ bias, truncation bias, is related to sample skewness measure. Sample bias, which is the other component, is not affected

by the measure. If sample bias dominates in $Huber_{15}$ bias, the correlation between $Huber_{15}$ bias and sample skewness measure may be hidden.

To reveal how sample skewness measure affects $Huber_{15}$ bias, the truncation part needs be separated from the total $Huber_{15}$ bias. It can be calculated using the following formula.

$$Huber_{15} \text{ Truncation Bias} = \bar{x} - Huber_{15}$$

Note that when all observations in a sample are multiplied by a constant, the truncation bias changes by the same magnitude. To remove the scale effect, $Huber_{15}$ truncation bias can be divided by inner range $Range_\alpha = (p_{1-\alpha} - p_\alpha)$ to produce a standardized truncation bias measure, $HubTB_{15}$:

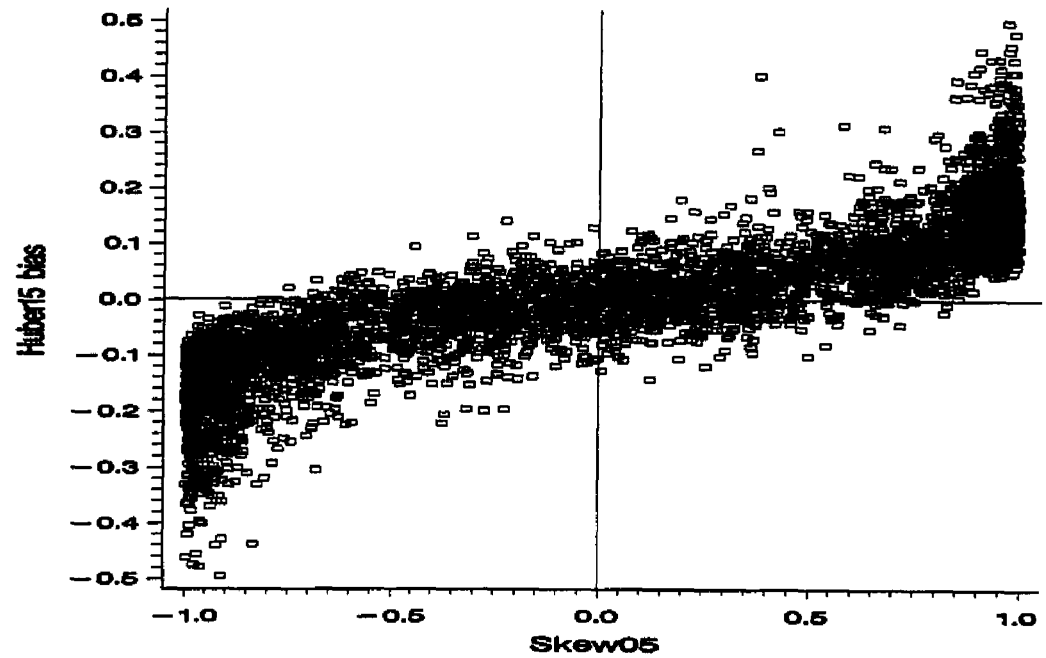
$$HubTB_{15} = \frac{\bar{x} - Huber_{15}}{p_{100-\alpha} - p_\alpha}$$

A plot of the $Huber_{15}$ truncation bias vs. $Skew_{05}$ based on the simulation on nine beta distributions (Table 1) is shown on Figure 4. For comparison, a plot of the unpartitioned $Huber_{15}$ bias vs. $Skew_{05}$ is also shown in the figure.

Relationship between truncation bias and $Skew_\alpha$

Which $Skew_\alpha$ to use

For $Skew_\alpha$, α ranges from 0 to 50. A natural question would be what value of α to use. For a certain α, the values of the observations that are outside α percentile



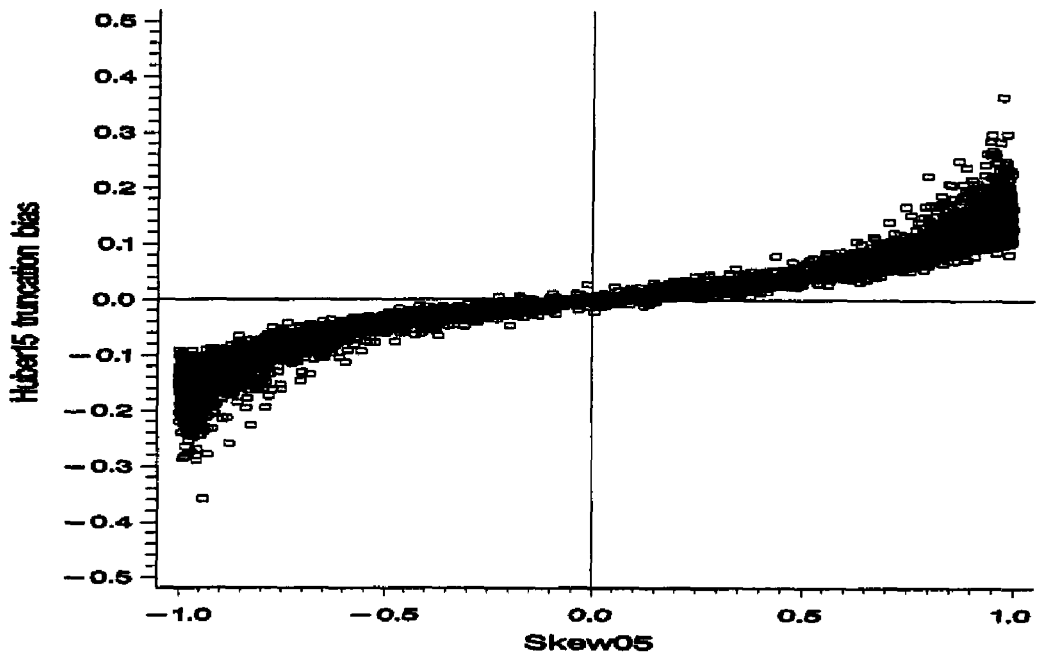


Figure 4. Bias and truncation bias of $Huber_{15}$ vs. $Skew_5$ for beta distributions with sample size 50

and (100-α) percentile do not affect the corresponding skewness measure $Skew_{\alpha}$ (i.e. $Skew_{\alpha}$ is robust up to some α percent of outliers at both extremes). From this point of view, a large α value makes the skewness measure robust to more outliers.

On the other hand, by choosing a large α value, more information in a sample is lost than when a small α value for skewness measure is chosen. As an extreme, if α = 50, the skewness measure $Skew_{\alpha}$ would be zero no matter how skewed the sample distribution was. From this point of view, a small α value is preferred.

The above two points of view contradict each other. As the skewness measure is used for developing a robust estimator, it should be robust to at least a minimum percentage of outliers in a sample, say 5% at each tail. Otherwise, the resulting robust estimators would not be robust at all.

$HubTB_{15}$ and $Skew_{\alpha}$ for α = 5, 10, 15, 20 and 25 were calculated for the nine beta distributions (Table 1). Figure 5 shows plots of $HubTB_{15}$ vs. $Skew_{\alpha}$ for α = 5, 10, 15, 20 and 25, respectively.

Figure 5 shows that for α = 5 and 10, there is a close relationship between $HubTB_{15}$ and $Skew_{\alpha}$. As α increases, the points on the plots scatter wider, indicating looser correlation between the two variables. In the following discussions, α is fixed to 5.

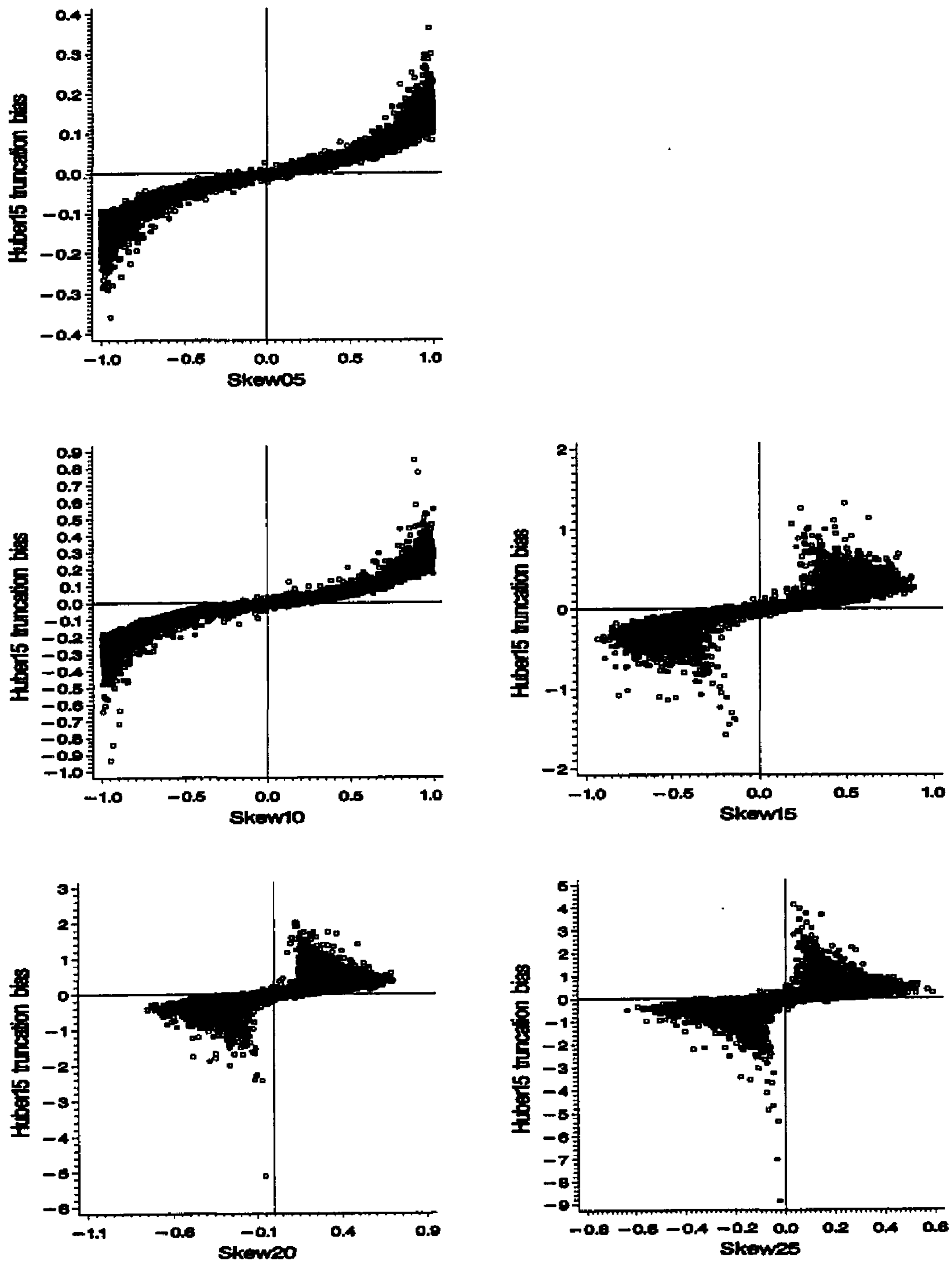


Figure 5. Truncation bias of $Huber_{15}$ vs. $Skew_{\alpha}$ for α=5, 10, 15, 20, and 25 for beta distributions with sample size 50

Simulations on other distributions

A close relationship between $HubTB_{15}$ and $Skew_5$ for beta distributions was shown in Figure 5. To reveal if the form of the relationship depends on population distributions, simulations on gamma, lognormal and Weibull distributions were also conducted (Table 2). The range of population skewness simulated were not the same because of the limitations of some distributions. The simulated sample size was 50 and number of replications was 500 for all the distributions.

Figure 6 shows plots of Huber15 truncation bias vs. Skew05 for the simulated gamma, lognormal and Weibull distributions. A graph of Huber15 truncation bias vs. Skew05 for beta distributions was also included in the figure for comparison. Data points from all four simulated distributions were also overlaid in a single graph to make the comparison easier. The plots indicate that the relationship between HubTB15 and. Skew5 is similar for the four families of distributions.

Sample size effect

A sample size of 50 was used in all previous simulations of the four distributions. To determine how sample sizes affect the relationship between $Huber_{15}$ truncation bias and $Skew_{05}$, simulations with sample sizes of 30 and 100 were also carried out. The two sample sizes were chosen to represent small and large samples. Figure 7 shows plots of $Huber_{15}$ truncation bias vs. $Skew_{05}$ for each of the three different sample sizes on all four distributions. A plot that combines all the data of the

Table 2. Parameters of the simulated gamma, lognormal and Weibull distributions

Distributions	Parameter	Mean	SD	Skewness[4/]	Kurtosis[5/]
Gamma[1/]	$\alpha = 4.00$	4.0000	4.0000	1	4.5000
	1.00	1.0000	1.0000	2	9.0000
	0.44	0.4400	0.4400	3	16.6364
	0.25	0.2500	0.2500	4	27.0000
Lognormal[2/]	$\sigma = 0.32$	1.0525	0.3456	1	4.9073
	0.55	1.1633	0.6914	2	10.8035
	0.72	1.2959	1.0681	3	22.8860
	0.83	1.4112	1.4052	4	40.4264
Weibull[3/]	$c = 41.00$	0.9865	0.0303	-1	4.7767
	3.60	0.9011	0.2780	0	2.7167
	1.56	0.8988	0.5886	1	4.1723
	1.00	1.0000	1.0000	2	9.0000
	0.76	1.1779	1.5703	3	18.2512
	0.64	1.3904	2.2588	4	31.9063

[1/] The probability density function of gamma distribution is

$$p_X(x) = \frac{(x-\gamma)^{\alpha-1}\exp(-\frac{x-\gamma}{\beta})}{\beta^{\alpha}\Gamma(\alpha)} \qquad (x \geq \gamma, \gamma = 0, \beta = 1)$$

[2/] The probability density function of lognormal distribution is

$$p_X(x) = \frac{\exp\left\{-\frac{1}{2}\frac{[\ln(x-\theta)-\zeta]^2}{\sigma^2}\right\}}{(x-\theta)\sqrt{2\pi}\sigma} \qquad (x > \theta, \theta = 0, \zeta = 0)$$

[3/] The probability density function of Weibull distribution is

$$p_X(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1}\exp\left[-\left(\frac{x-a}{b}\right)^{c}\right] \qquad (x \geq a, a = 0, b = 1)$$

[4/] Skewness = $E[(X - \mu)^3] / \{E[(X - \mu)^2]\}^{3/2}$
where X is a random variable, $\mu = E(X)$.

[5/] Kurtosis = $E[(X - \mu)^4] / \{E[(X - \mu)^2]\}^2$
where X is beta random variable, $\mu = E(X)$.

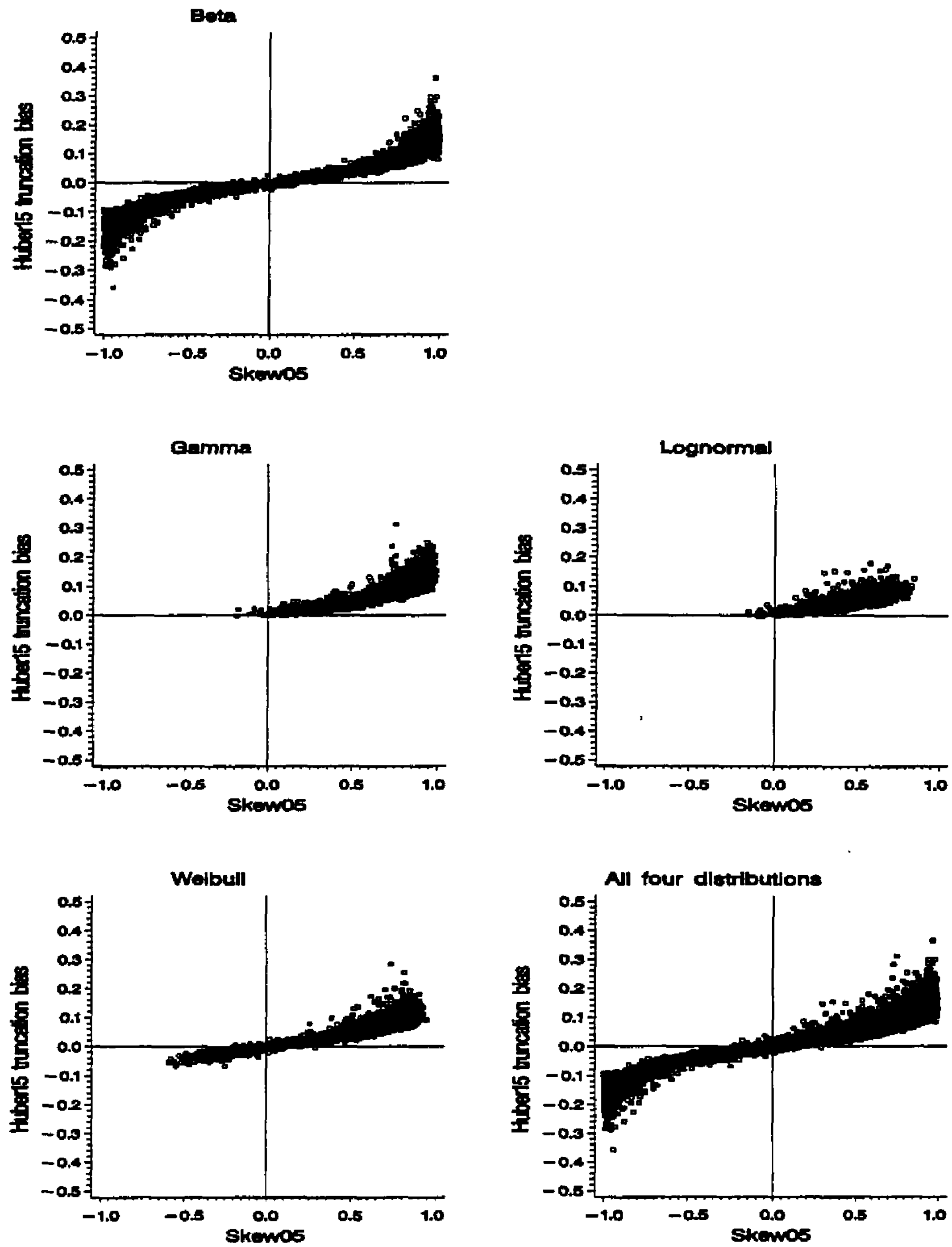


Figure 6. Truncation bias of $Huber_{15}$ vs. $Skew_5$ for beta, gamma, lognormal and Weibull distributions with sample size 50

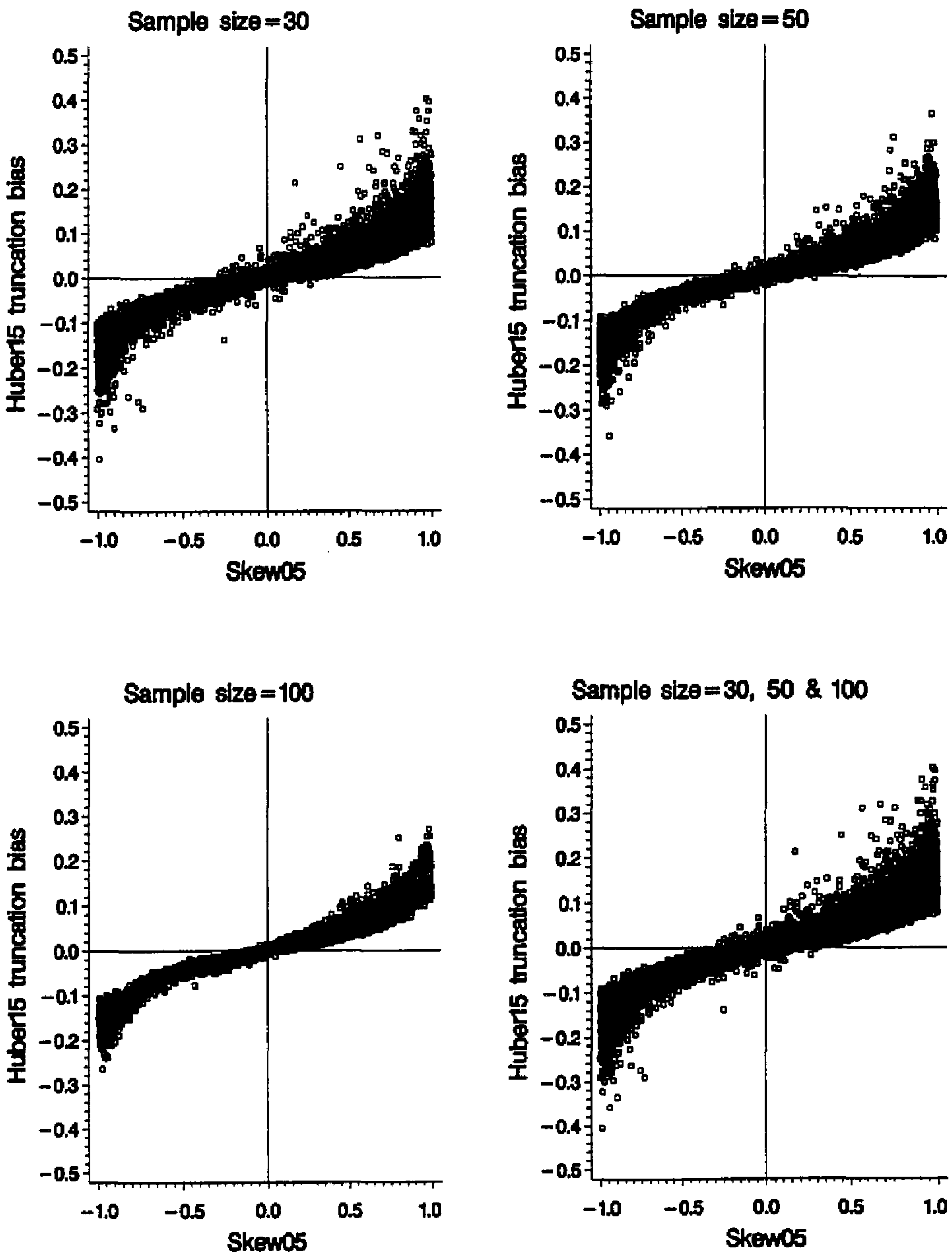


Figure 7. Truncation bias of $Huber_{15}$ vs. $Skew_5$ for beta, gamma, lognormal and Weibull distributions with different sample sizes

three different sample sizes is also show on the figure. Points representing different sample sizes exhibited a similar trend. However, points from sample size 100 were distributed in a narrower band than those of sample sizes 30 and 50, which indicated that large sample size produced smaller variations.

Members of the Huber estimator family

$Huber_{15}$ is one member of the Huber estimator family. The simulations for $Huber_{15}$ were also repeated for $Huber_{05}$, $Huber_{10}$, $Huber_{20}$, $Huber_{25}$ and $Huber_{30}$. Different members of Huber estimator family all exhibit close relationship between truncation bias and $Skew_{05}$ (Figure 8).

Different models were employed in an effort to describe the relationship between $HubTB_{\alpha}$ and $Skew_5$. The models were fit to the pooled data of all four distributions and three sample sizes. Among them, a simple third degree polynomial with only the linear and cubic terms, i.e.

$$HubTB_{\alpha} = a\,(Skew_5) + b\,(Skew_5)^3$$

performed well. Table 3 summarizes the results of fitting the cubic model to the simulation results for the five estimators.

Table 3 shows that parameter "a" decreased with the constant (k) used to calculate the Huber estimates. On the other hand, parameter "b" increased with k. Values of coefficient of determination (R^2) were high, ranging from 0.9076 to 0.9411.

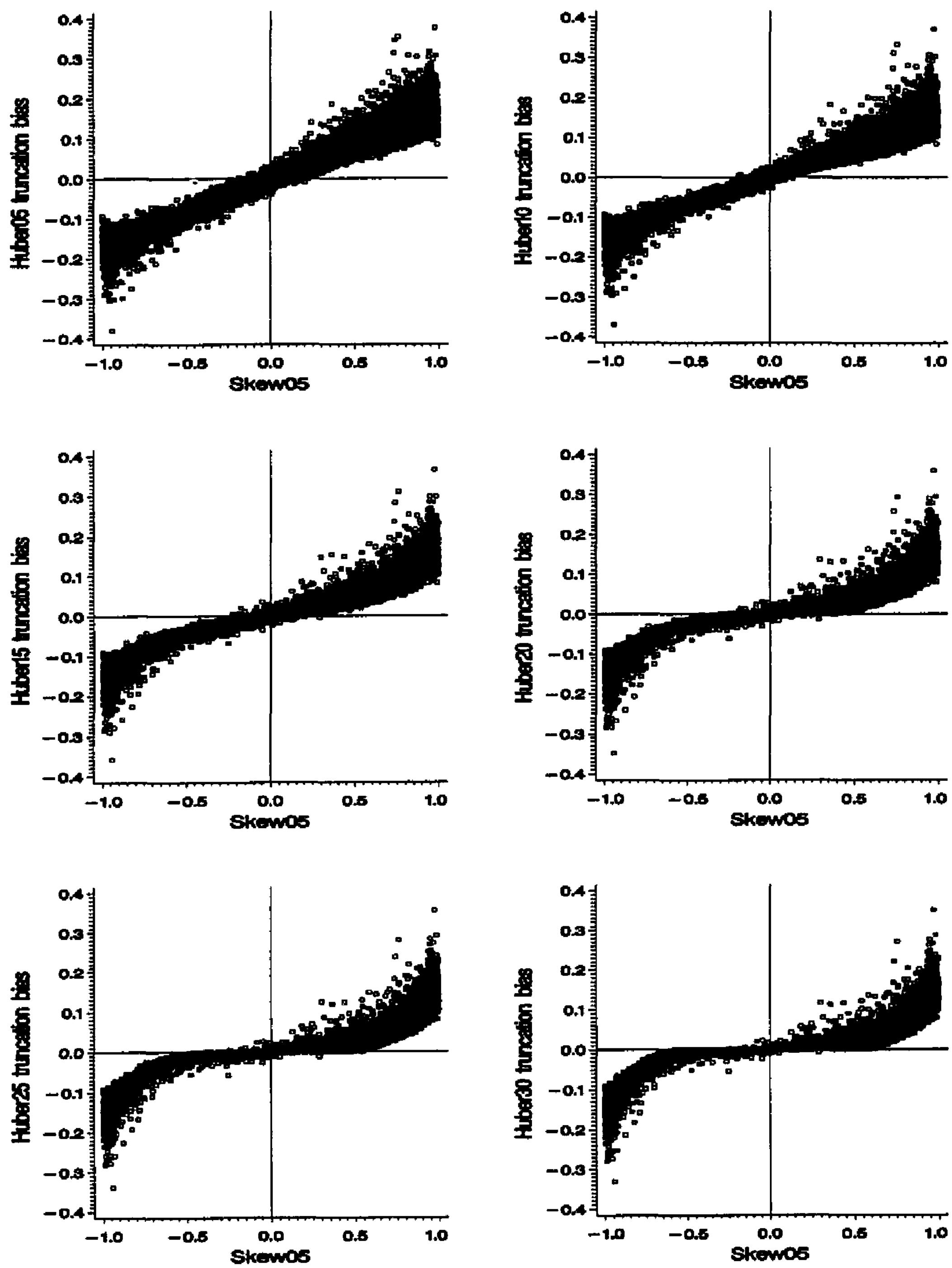


Figure 8. Truncation bias of Huber$_\alpha$ vs. Skew$_5$ for α=5, 10, 15, 20, 25 and 30 for beta distributions with sample size 50

Table 3. Estimated parameters and R^2 for fitting Huber truncation biases to $Skew_5$ [1/]

Estimates	a	b	R^2
$Huber_5$	0.1829	-0.0073	0.9407
$Huber_{10}$	0.1340	0.0355	0.9411
$Huber_{15}$	0.0718	0.0937	0.9315
$Huber_{20}$	0.0278	0.1333	0.9237
$Huber_{25}$	-0.0008	0.1561	0.9164
$Huber_{30}$	-0.0193	0.1683	0.9076

[1/] $HubTB_\alpha = a\,(Skew_5) + b\,(Skew_5)^3$

Sample median, Hampel, Andrews and Tukey estimators

Different estimators provide different weighting the of observations, and thus have different bias properties for asymmetric distributions. Four other robust estimators were also simulated. They were sample median, Hampel with a = 1.7, b = 3.4, and c = 8.5, Andrews with c = 6.6, and Tukey with c = 6.0. The parameters chosen for these estimators were recommended by Andrews et al. (1972).

Figure 9 shows plots of truncation bias vs. $Skew_{05}$ for the four estimators. A graph of $Huber_{15}$ truncation bias vs. $Skew_{05}$ were also include in the figure for comparison. Table 4 summarizes the results of fitting a similar cubic regression to simulation data for each of the four estimators.

Modified robust estimators

In the previous sections, it has been demonstrated for sample median, Huber, Hampel, Andrews and Tukey estimators that truncation bias of a robust estimator was closely related to a robust sample skewness measure $Skew_{05}$. The relationship did not appear to depend on distribution types and sample sizes simulated, and could be adequately described by a third degree polynomial with the linear and cubic terms.

Based on the above observations, a modified robust estimator can be constructed for each of the robust estimators. For $Huber_{15}$, the modified robust estimator, noted as $MHub_{15}$, is defined as

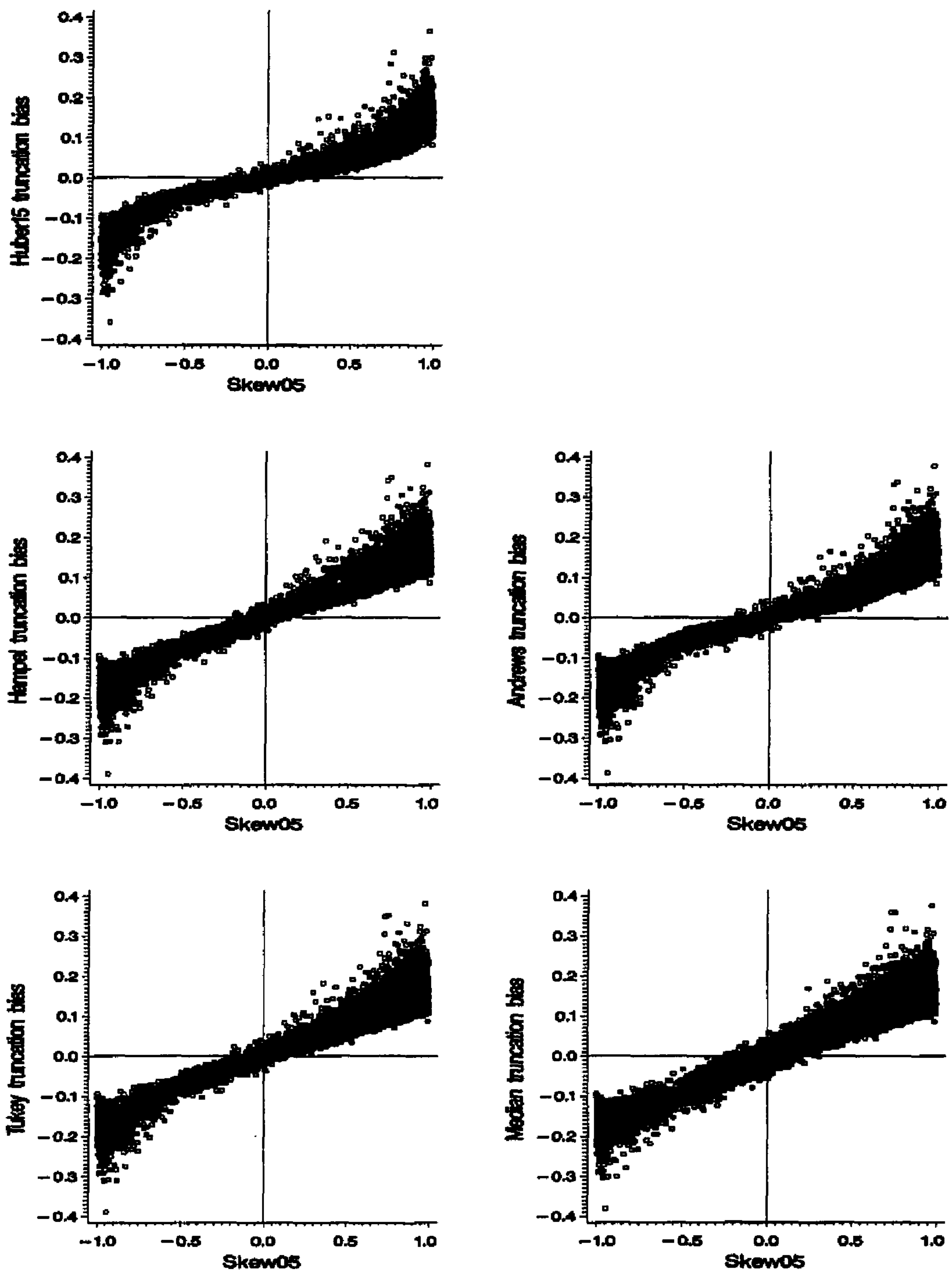


Figure 9. Truncation bias of Hampel, Andrews, Tukey and sample median vs. $Skew_5$ for beta, gamma, lognormal and Weibull distributions with sample size 50

Table 4. Estimated parameters and R^2 for fitting truncation biases of four estimators to $Skew_5$ [1/]

Estimators	a	b	R^2
Median	0.1924	-0.0169	0.9329
Hampel	0.1549	0.0314	0.9402
Andrews	0.1129	0.0768	0.9334
Tukey	0.1525	0.0363	0.9375

[1/] Truncation bias = a ($Skew_5$) + b $(Skew_5)^3$

$$MHub_{15} = Huber_{15} + (HubTB_5)\ (Range_5)$$

Substitute HubTB15 with the fitted model,

$$MHub_{15} = Huber_{15} + [0.0718\ (Skew_5) + 0.0937\ (Skew_5)^3]\ (Range_5)$$

Modification for the other robust estimators can be constructed in the same way as for $Mhub_{15}$ using the parameters provided in Table 4. The bias correction term in each modified robust estimator satisfies the two requirements stated earlier.

Validation of the Modified Robust Estimators

Validation for the modified sample median, Huber15, Hampel, Andrews and Tukey was carried out on simulated populations and also on a forestry data set. Each modified robust estimator was compared with corresponding original robust estimator and sample mean in terms of bias, standard deviation and square root of mean square error (MSE).

Bias of an estimate was calculated as

$$\text{Bias} = \text{True population mean - Estimate}$$

Mean square error (MSE) was calculated as

$$\text{MSE} = (\text{Bias})^2 + \text{Variance}$$

Evaluation on simulated populations

Four distributions, beta, gamma, lognormal and Weibull, were simulated to evaluate the modified robust estimators. Population skewness of the simulated distributions ranged from -8 to 8. For each distribution with a specific skewness, simulation was run for sample sizes 30, 50 and 100. For each sample size, the simulation was repeated 500 times. Sample mean, median, Huber15, Hampel, Andrews, Tukey and the corresponding modified robust estimates were calculated for each sample simulated.

The mean bias of the estimators for sample size = 30, 50, and 100 are listed in Appendix Tables 14 through 16. Appendix Tables 17 through 19 display standard deviation of the estimates for the three different sample sizes. Appendix Tables 20 through 22 show square root of MSE of the estimates for sample sizes 30, 50 and 100.

Visual comparisons can be made from Figures 10 through 12 that show bias, standard deviation, and square root of MSE for sample mean, $Huber_{15}$, and $MHub_{15}$ for Weibull distributions with different population skewness levels (sample size = 50). For populations with low skewness levels, the three estimates performed similarly in terms of bias (Figure 10). However, the modified $Huber_{15}$ estimator performed much better than $Huber_{15}$ for highly-skewed populations, and is comparable with sample mean.

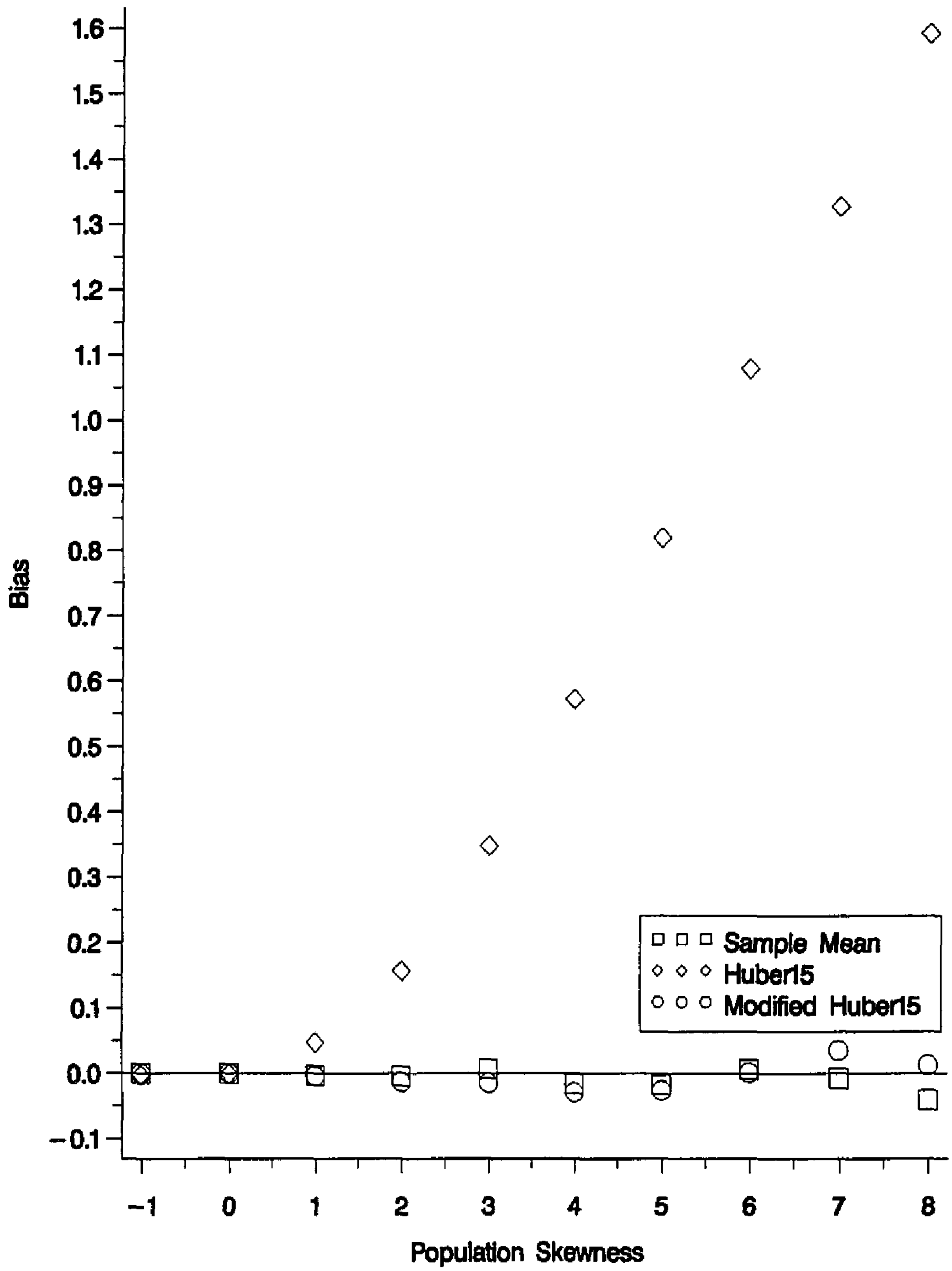


Figure 10. Bias of sample mean, $Huber_{15}$ and modified $Huber_{15}$ vs. population skewness (Weibull distributions with sample size 50)

Figure 11 shows that standard deviation of the modified Huber_{15} was very close to that of sample mean, while Huber_{15} yielded lower standard deviation for populations with high kurtosis.

In terms of square root of MSE, Figure 12 indicates that the performance of MHub_{15} and sample mean was similar, whereas Huber_{15} exhibited higher MSE for skewed populations.

Results for Hampel (Appendix Figures 20 - 22), Andrews (Appendix Figures 23 - 25) , Tukey (Appendix Figures 26 -28) and median (Appendix Figures 29 -31) were similar to those for Huber_{15}. The comparisons for other sample sizes and distributions produced similar results to those reported for Weibull distributions with sample size 50.

Figure 13 plots the biases of sample mean and all the five modified robust estimates for Weibull distributions with sample size 50. None of the estimators performed consistently better than the others. Figures 14 and 15 are similar graphs for standard deviation and square root of MSE. Points of the six estimates overlapped with one another for populations with low skewness levels. For highly-skewed populations, sample mean performed slightly better than the modified robust estimators, with the modified Huber_{15} in second place.

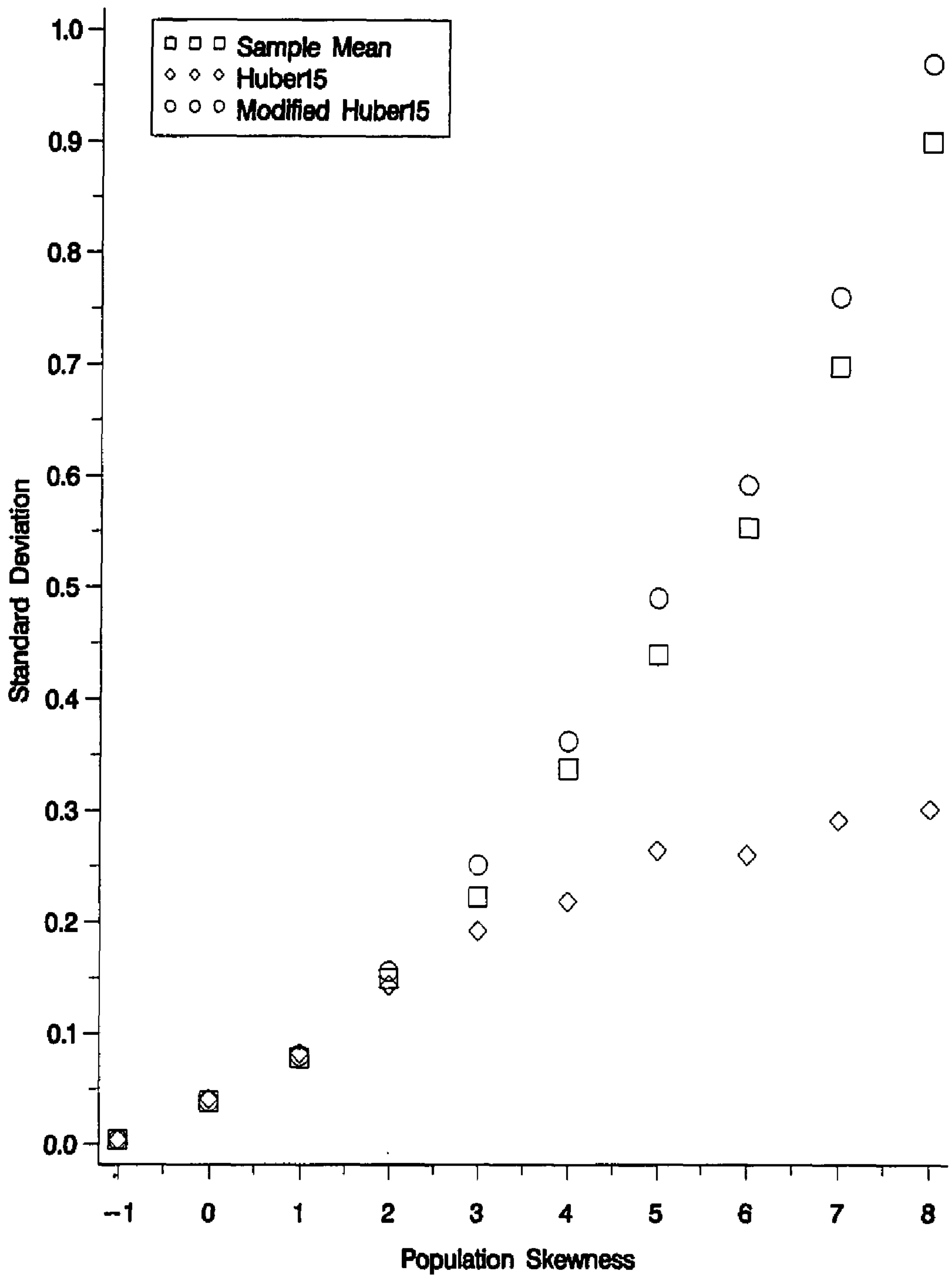


Figure 11. SD of sample mean, $Huber_{15}$ and modified $Huber_{15}$ vs. population skewness (Weibull distributions with sample size 50)

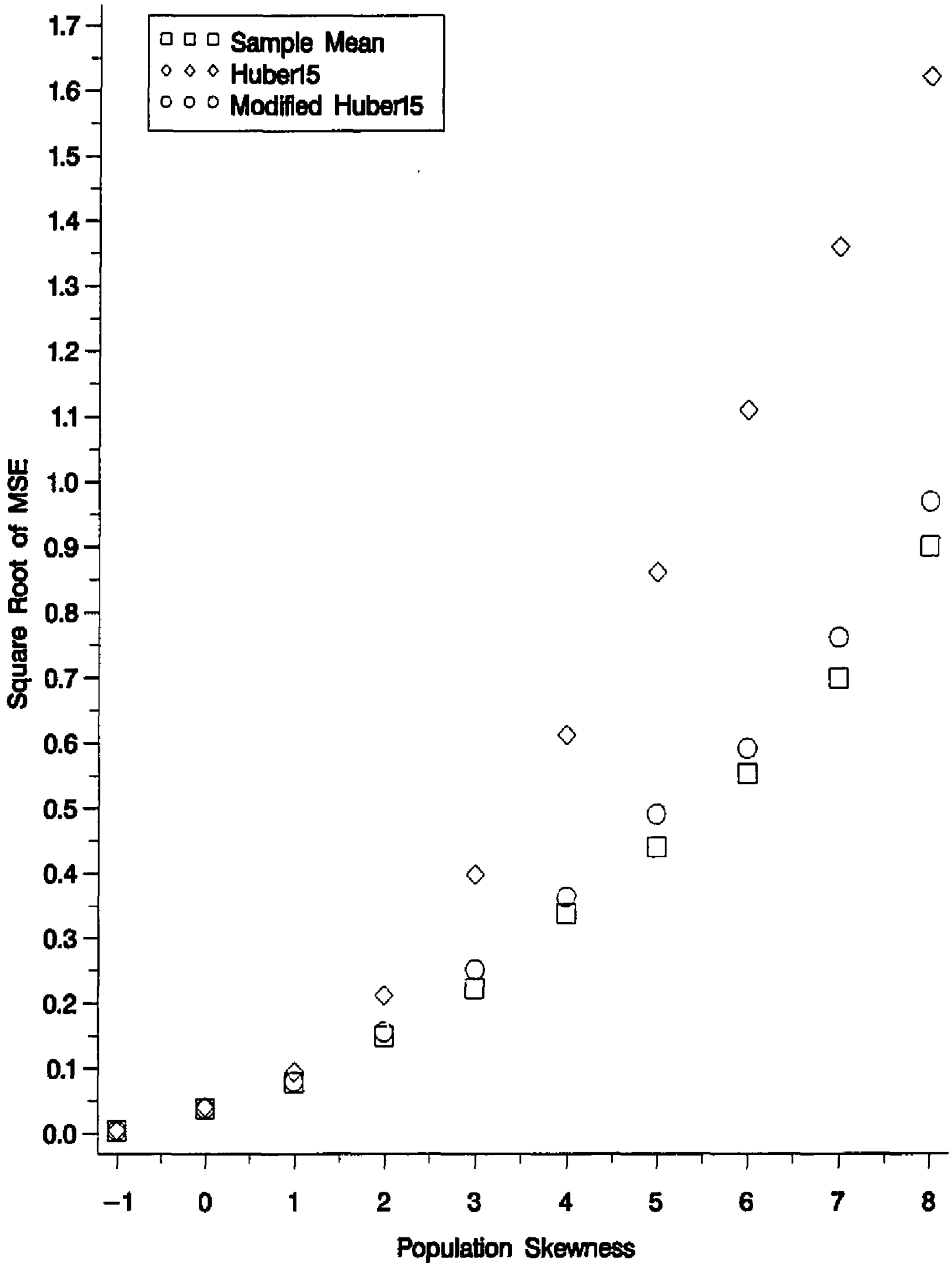


Figure 12. $\sqrt{\text{MSE}}$ of sample mean, Huber_{15} and modified Huber_{15} vs. population skewness (Weibull distributions with sample size 50)

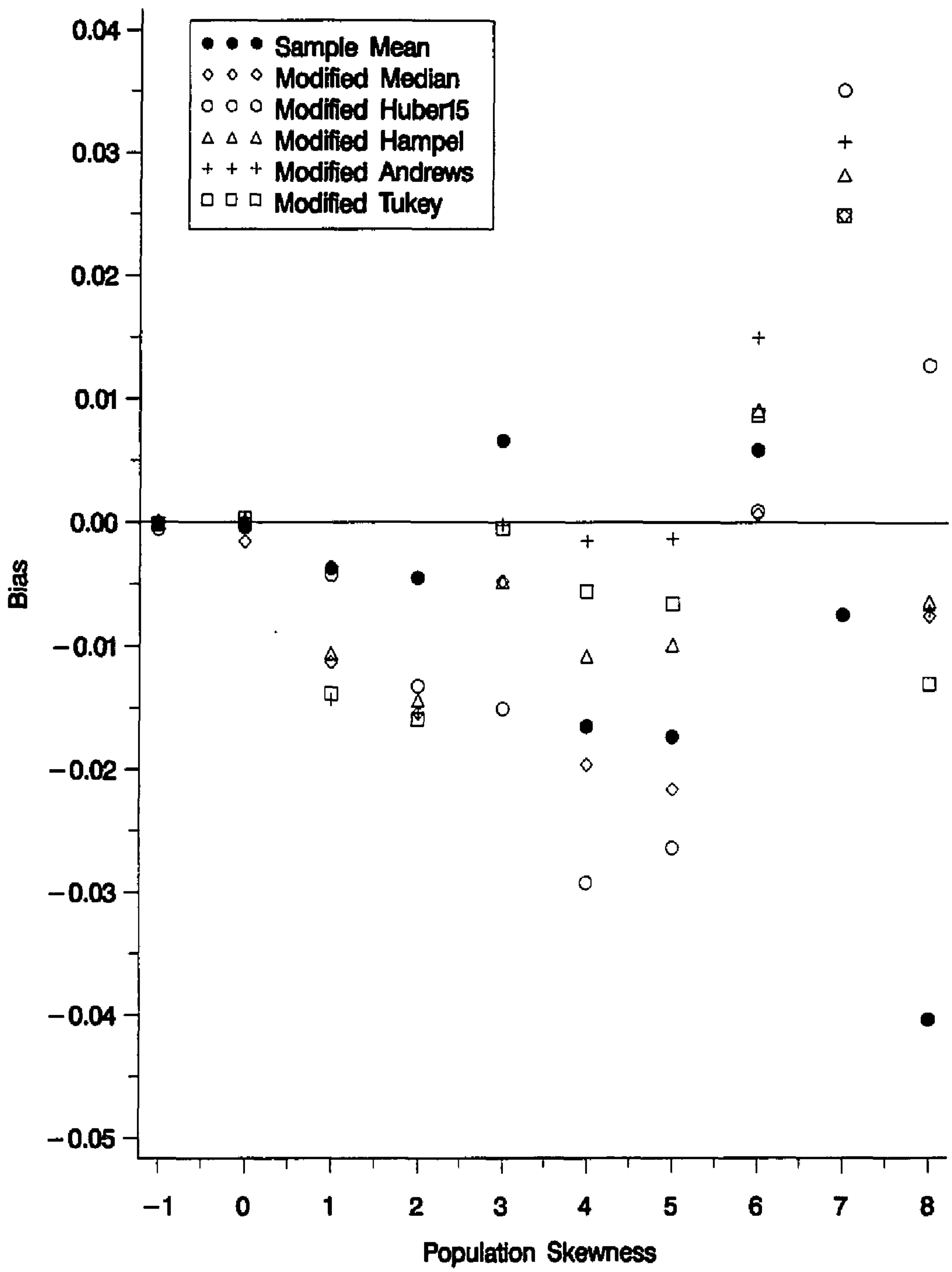


Figure 13. Bias of sample mean and the five modified estimators vs. population skewness (Weibull distributions with sample size 50)

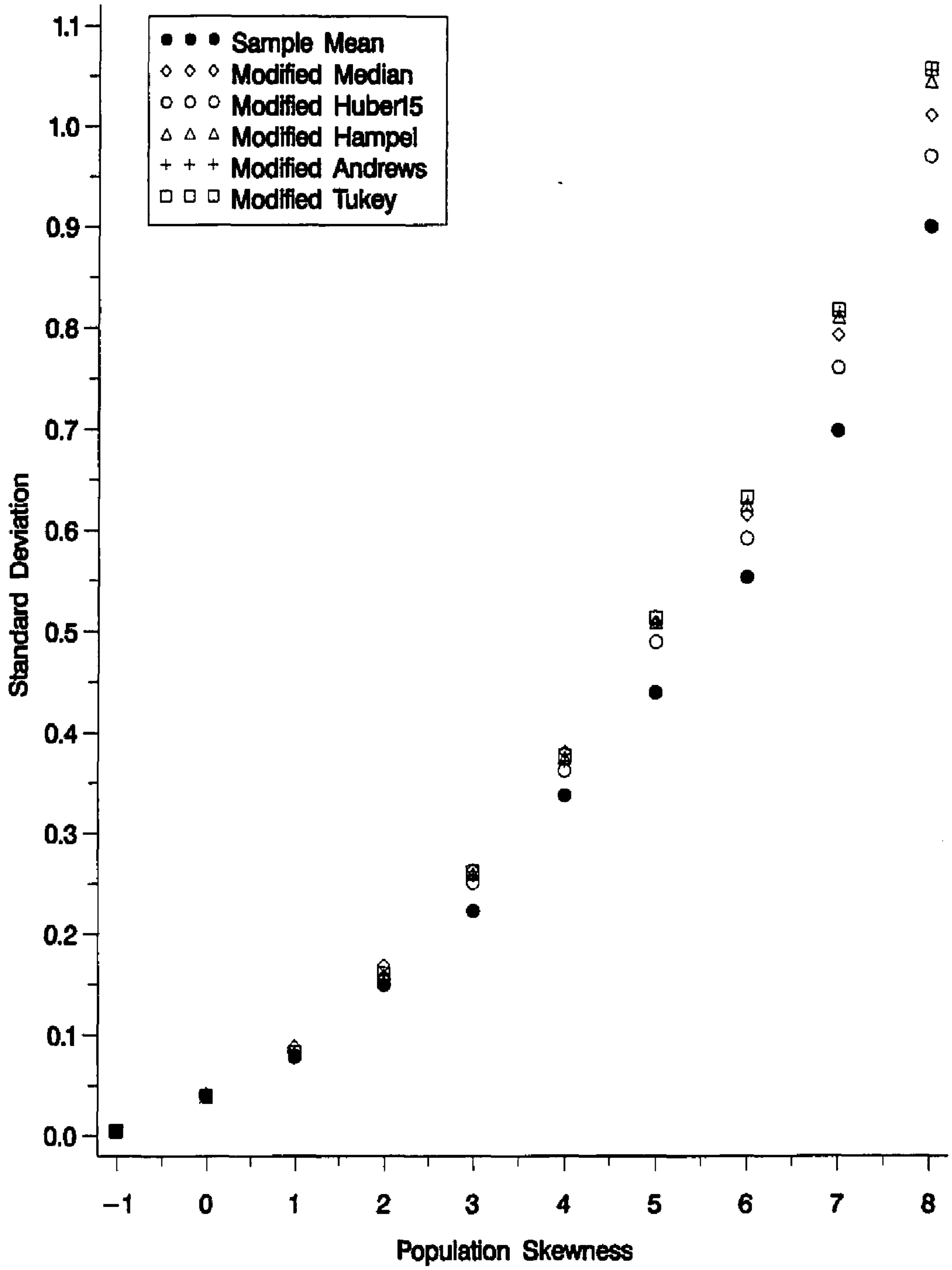


Figure 14. SD of sample mean and the five modified estimators vs. population skewness (Weibull distributions with sample size 50)

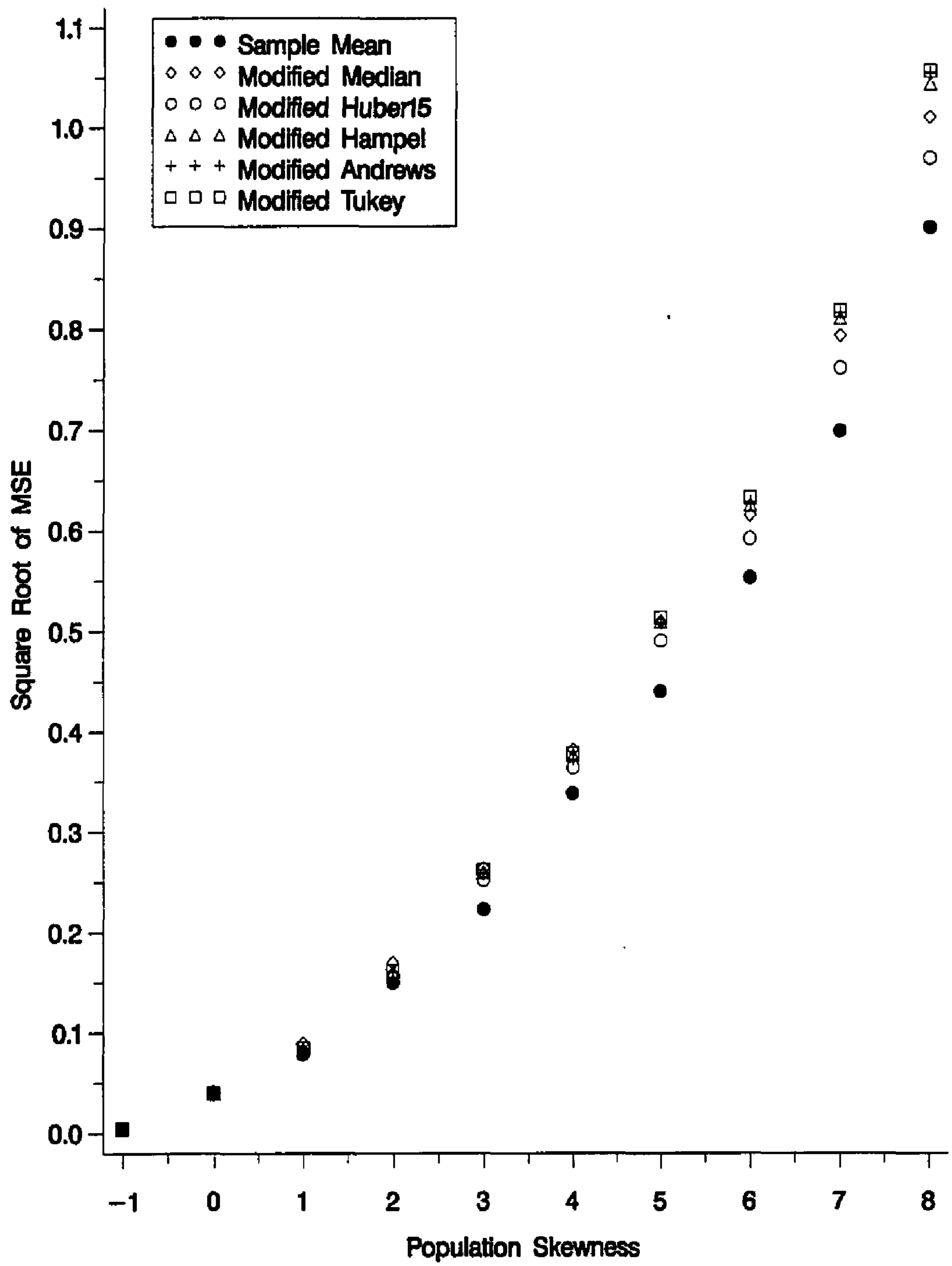


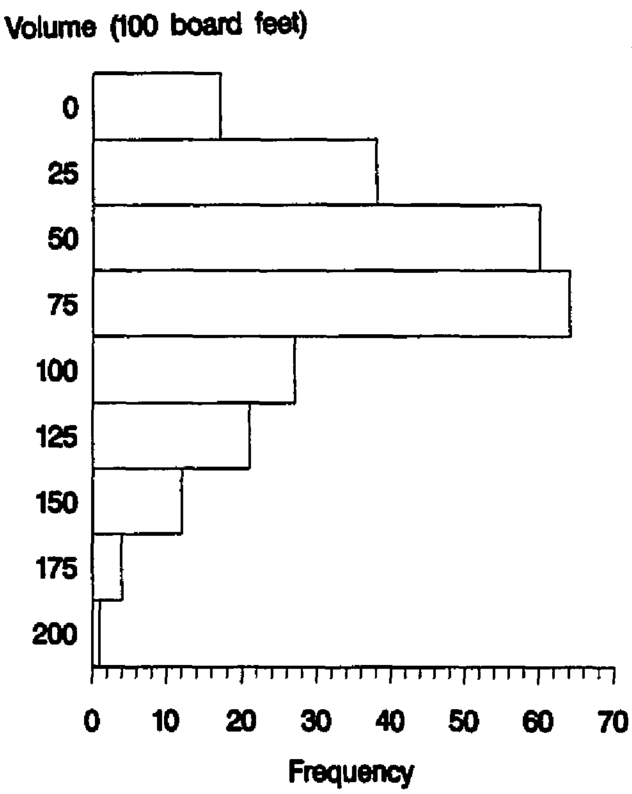
Figure 15. $\sqrt{\text{MSE}}$ of sample mean and the five modified estimators vs. population skewness (Weibull distributions with sample size 50)

Evaluation on a forestry data set

In the spring of 1994, 244 0.2 acre circular plots were measured in an area of 124 acres at Lee Memorial Forest, Washington Parish, Louisiana. Locations of the plots were determined by laying a 2 x 2 chain grid over the track. All pine trees with diameter at breast height (DBH) 10 inches and above and at least one log (16 feet) of merchantable timber were tallied as saw timber. For each tree tallied, diameter at breast height was measured in two perpendicular directions to the nearest 0.1 inch using a pair of calipers. Two measurements were averaged and rounded to the nearest 1 inch and recorded. Merchantable heights of two trees on each plot were measured to ensure the accuracy of height estimation for the rest of the trees in the plot. Number of 16-foot logs was estimated to the nearest 0.5 log to a top diameter of 8 inches. A Doyle log rule volume table for form 78 was used to calculate board feet volume from DBH and number of logs.

Some summary statistics of pine sawtimber volume (in Doyle board feet per acre) of the 244 plots are: sample mean 6935.16, standard deviation 4063.19, skewness 0.5205, and kurtosis 0.0028. A histogram of the data and frequency information was shown on Figure 16. Shapiro-Wilk test statistics (Shapiro and Wilk 1965) was calculated. The test rejected the null hypothesis that the sample was from a normal population at one percent level ($W = 0.9597$, p value = 0.0001).

Nonparametric bootstrapping (Efron 1979, 1982, and 1987, Léger and Politid 1992) was applied to evaluate the estimators. In doing so, the 244 plots were



FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
17	17	6.97	6.97
38	55	15.57	22.54
60	115	24.59	47.13
64	179	26.23	73.36
27	206	11.07	84.43
21	227	8.61	93.03
12	239	4.92	97.95
4	243	1.64	99.59
1	244	0.41	100.00

Figure 16. Histogram and frequency information of pine sawtimber volume from 244 plots.

considered to form a population. From this population, 500 subsamples of size 30 were drawn randomly with replacement. For each sample drawn, sample mean, the five robust estimates and corresponding modified robust estimates were calculated. Average bias, standard deviation and square root of MSE were obtained upon completion of the bootstrapping. The same process was carried out for sample sizes of 50, 100, and 244. Bootstrapping results are summarized in Table 5.

Results shown in Table 5 are consistent with the results obtained from simulations on theoretical populations. Notice that the skewness of the sample is 0.52, indicating that the sample is not very skewed. However, the five original robust estimators exhibited much higher bias (ranging from 180 to 325 board feet) than sample mean and the modified robust estimators (ranging from -74 to 14 board feet). On the other hand, the performances of the modified estimators and the sample mean were comparable in terms of bias. The standard deviation and square root of MSE of the five original robust estimators are consistently higher than those of their modified counterparts. All of the five bias-corrected robust estimators were comparable with sample mean, with no clear-cut winners. Sample size did not appear to affect the relative standings of the estimators.

Table 5. Simulation results from 500 subsamples of different sizes from 244 plots

Sample size		Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
	Bias	-23.39	324.59	-74.14	180.42	13.98	294.51	-34.44	217.84	-31.44	293.76	-31.05
30	SD	741.73	810.45	718.73	765.65	747.88	783.99	735.42	773.69	747.42	790.78	744.10
	$\sqrt{MSE}$	742.10	873.03	722.54	786.62	748.01	837.48	736.23	803.77	748.08	843.58	744.75
	Bias	-17.32	372.88	-51.00	205.88	33.089	326.89	-20.96	237.87	-23.54	327.67	-15.57
50	SD	581.88	640.80	553.46	602.52	586.77	612.73	571.67	604.35	583.48	618.62	581.48
	$\sqrt{MSE}$	582.14	741.39	555.80	636.73	587.71	694.47	572.05	649.48	583.95	700.04	581.69
	Bias	-6.47	412.96	-24.97	228.46	54.38	350.83	-6.54	252.57	-13.65	349.57	-2.83
100	SD	408.87	463.89	399.04	426.10	413.90	434.07	401.98	428.58	412.04	440.70	411.27
	$\sqrt{MSE}$	408.93	621.08	399.82	483.48	417.46	558.12	402.03	497.47	412.27	562.51	411.28
	Bias	-2.47	445.50	-0.47	244.03	68.86	366.17	3.19	261.10	-8.17	363.12	5.30
244	SD	254.78	296.87	253.03	267.83	259.78	274.83	253.98	271.22	261.11	280.34	261.95
	$\sqrt{MSE}$	254.79	535.35	253.03	362.33	268.75	457.83	254.00	376.48	261.24	458.74	262.00

APPLICATIONS OF ROBUST LINEAR REGRESSION

Two robust linear regression procedures, least median of squares (LMS) and least trimmed squares (LTS), were used in this study to compare with the ordinary least square (OLS) estimation of linear models.

Linear models can be written as

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i \qquad i = 1, \cdots, n$$

where p is number of parameters in the model, n is number of observations in a data set, y_i is the dependent variable value on the i^{th} observation, x_{ij} is the j^{th} dependent variable value on the i^{th} observation, β_j is the parameter of the j^{th} independent variable, and ε_i is error term.

Given a vector of parameter estimates, $(\hat{\beta}_1, \cdots, \hat{\beta}_p)$, the residuals can be calculated as

$$e_i = y_i - \sum_{j=1}^{p} x_{ij}\hat{\beta}_j \qquad i = 1, \cdots, n$$

Least squares estimates of a linear model are found by minimizing the sum of the squared residuals:

$$\min_{\beta} \sum_{i=1}^{n} e_i^{\,2}$$

On the other hand, least median of squares estimates of a linear model are found by minimizing the median of the squared residuals instead of the sum (Rousseeuw 1984).

$$\min_{\beta} \left[\underset{i}{median} \, (e_i)^2 \right]$$

Least trimmed squares estimates of a linear model are defined as a set of β values that minimizes a trimmed sum of the squared residuals (Rousseeuw and Yohai 1984, Yohai 1987, Yohai and Zamar 1988, Marazzi 1991).

$$\min_{\beta} \sum_{i=1}^{h} (e^2)_{i:n}$$

where $(e^2)_{i:n}$ are the squared and then ordered residuals, and h is the largest integer satisfying $h < (n/2) + (p+1)/2$.

The above three estimators of linear regression were used to estimate two individual tree volume equations on nine test data sets. A tenth data set was used to obtain coefficients of two forest yield models using the three procedures. The three estimators were then compared based on two evaluation criteria.

Robust Estimation of Two Volume Equations

Data

Nine test data sets (Table 6) that contain detailed measurement of 1231 loblolly pine (*Pinus taeda* L.) and slash pine (*Pinus elliottii* var. *elliottii* Engelm.) trees were

Table 6. Means (and standard deviations) of the variables in the nine data sets used to fit individual tree volume equations

Data set	Species	Number of trees	DBH (in.)	Total Height (ft.)	Total Volume (cu. ft.)
1	Loblolly pine	284	9.15(2.47)	56.85(8.86)	12.13(7.91)
2	Loblolly pine	476	11.05(2.43)	70.81(7.71)	22.46(11.01)
3	Slash pine	56	11.52(2.36)	72.97(7.01)	23.43(10.96)
4	Loblolly pine	31	11.96(2.24)	62.16(6.00)	20.95(8.57)
5	Slash pine	14	7.41(0.90)	46.30(4.73)	4.37(0.72)
6	Loblolly pine	14	8.09(1.24)	57.07(3.93)	7.58(1.30)
7	Loblolly pine	13	6.89(0.63)	45.41(3.09)	5.68(1.33)
8	Slash pine	147	9.94(2.08)	73.83(6.00)	21.73(9.90)
9	Slash pine	196	10.17(1.97)	74.57(5.82)	22.67(10.01)

used to estimate volume equations. These trees were from plantations at the Hill Farm Research Station located in Homer, Louisiana. Each tree was felled and diameters outside bark were measured (to the nearest 0.1 in.) at 25 inch intervals, starting from the stump. Total height and stump height were measured to the nearest 0.1 ft. The cubic-foot volume of each 2 ft. 1 in. section was calculated using Smalian's formula (Avery and Burkhart 1994). The top section was assumed a cone in the volume calculation. The sum of volumes of all sections of a tree provided the total outside-bark volume.

Methods

Volume equations have been widely used in forestry to estimate individual tree volume from either diameter or both diameter and height. Two commonly used model forms are the combined variable volume equation (Spurr 1952)

$$V = b_0 + b_1 D^2 H + \varepsilon$$

and Schumacher and Hall (1933) equation

$$\ln (V) = c_0 + c_1 \ln (D) + c_2 \ln (H) + \varepsilon$$

where V is total tree volume in cubic feet, D is diameter at breast height in inches, H is total tree height in feet, b_j's and c_j's are regression parameters, ln (x) is natural logarithm of x, and ε is random error.

Prediction sum of squares, often called PRESS (Allen et al. 1973), was used to compare the predictive abilities of the models estimated using OLS, LMS, and LTS

procedures. PRESS was calculated by removing one observation at a time from the data set, and calculating the regression coefficients for the model based on the remaining observations. The model is then used to predict for the withheld observation. Differences between the predicted and observed values were then squared and summed together:

$$\text{PRESS} = \sum_{i=1}^{n}(y_i - \hat{y}_{(i)})^2$$

where y_i is the i^{th} observation on the dependent variable, $\hat{y}_{(i)}$ is the predicted value of y_i from the model estimated without the i^{th} observation, and n is sample size.

Another statistics, prediction sum of absolute errors (PRESAE), was also calculated for evaluating the estimation procedures:

$$\text{PRESAE} = \sum_{i=1}^{n}|y_i - \hat{y}_{(i)}|$$

The average PRESAE and PRESS statistics, which are PRESAE and PRESS values divided by number of observations in a data set, were calculated for all three estimation procedures on each of the nine data sets. For each data set, the three estimation methods were ranked from 1 (best) to 3 (worst) according to their average PRESAE and PRESS respectively.

Results

Combined variable volume equation

Table 7 lists the OLS, LMS and LTS parameter estimates for the combined variable volume equation on the nine data sets. Figure 17 shows plots of observed and predicted individual tree volumes vs. D^2H for each of the nine data sets.

For all the nine data sets, the two regression lines estimated by LMS and LTS were similar. For data sets 2, 4, 7, 8, and 9, the observed points fell in a tight linear pattern, causing the three regression lines to be close to one another. However, for data set 6, the OLS line was quite different from the two robust lines. It was affected by an outlier (a single observation that fell outside of the linear pattern) and failed to follow the general trend. On the other hand, the LMS and LTS lines fit the main pattern better than the OLS line.

The calculated average PRESAE and PRESS statistics for all three estimation procedures and nine data sets were shown in Figure 18 and Table 8. Relative ranks of the estimation procedures were enclosed in parentheses. For data set 6, which had an obvious outlier, the OLS procedure yielded higher average PRESS and PRESAE than the LMS and LTS procedures. For data sets 1, 2, 5, 7, 8, and 9, the three different estimation procedures exhibited similar average PRESS and PRESAE. Overall, the LMS and LTS had smaller rank sum than OLS in terms of both average PRESS and PRESAE. The LMS also performed slightly better than the LTS regression.

Table 7. Estimated parameters of the combined variable volume equation[1/]

Data Set	OLS b_0	OLS b_1	LMS b_0	LMS b_1	LTS b_0	LTS b_1
1	0.7760	0.002136	0.3043	0.002286	0.4115	0.002224
2	0.7358	0.002333	0.2202	0.002393	0.2675	0.002381
3	-1.5077	0.002427	0.3182	0.002175	-0.1380	0.002215
4	0.4232	0.002205	0.5210	0.002153	0.9151	0.002127
5	2.3433	0.000776	2.1896	0.000873	2.1653	0.000859
6	4.5839	0.000775	3.8154	0.001031	3.6141	0.001093
7	0.3680	0.002423	0.3471	0.002424	0.2121	0.002487
8	0.5675	0.002718	0.7131	0.002637	0.3796	0.002729
9	-0.4945	0.002841	-0.7381	0.002831	-0.4477	0.002797

[1/] Equation $V = b_0 + b_1 D^2 H + \varepsilon$, where V, D, and H are the individual tree volume (cu. ft.), DBH (in.), and total height (ft.), respectively

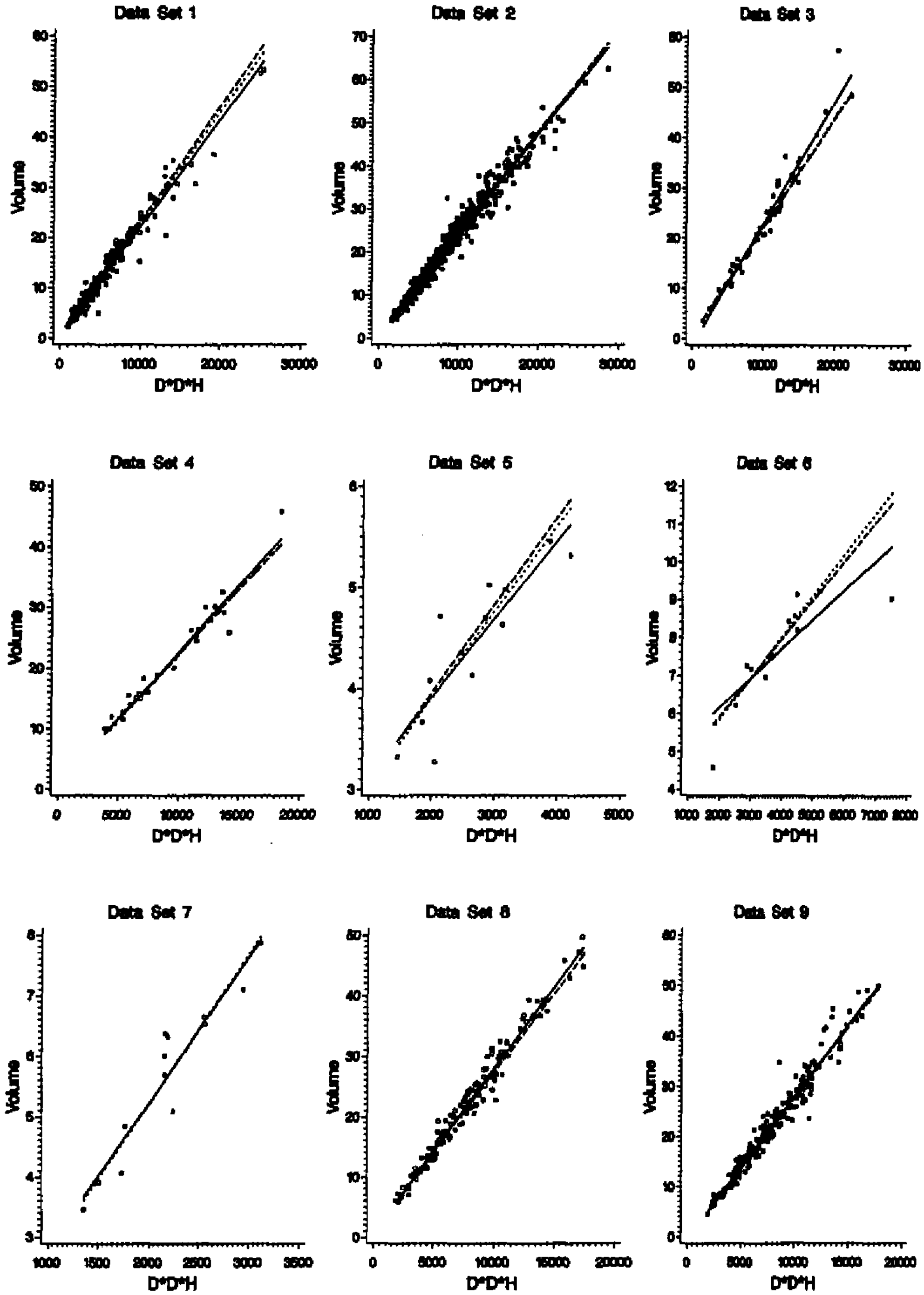
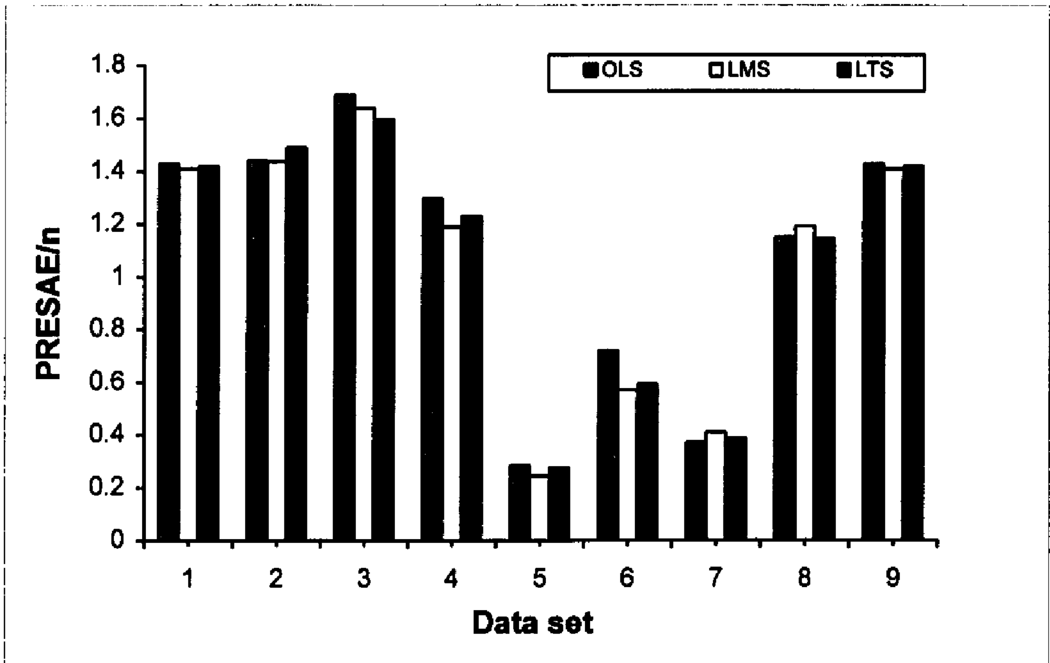


Figure 17. Observed and predicted individual tree volume vs. D^2H of the nine data sets (solid line - OLS, dashed line - LMS, and dotted line - LTS)



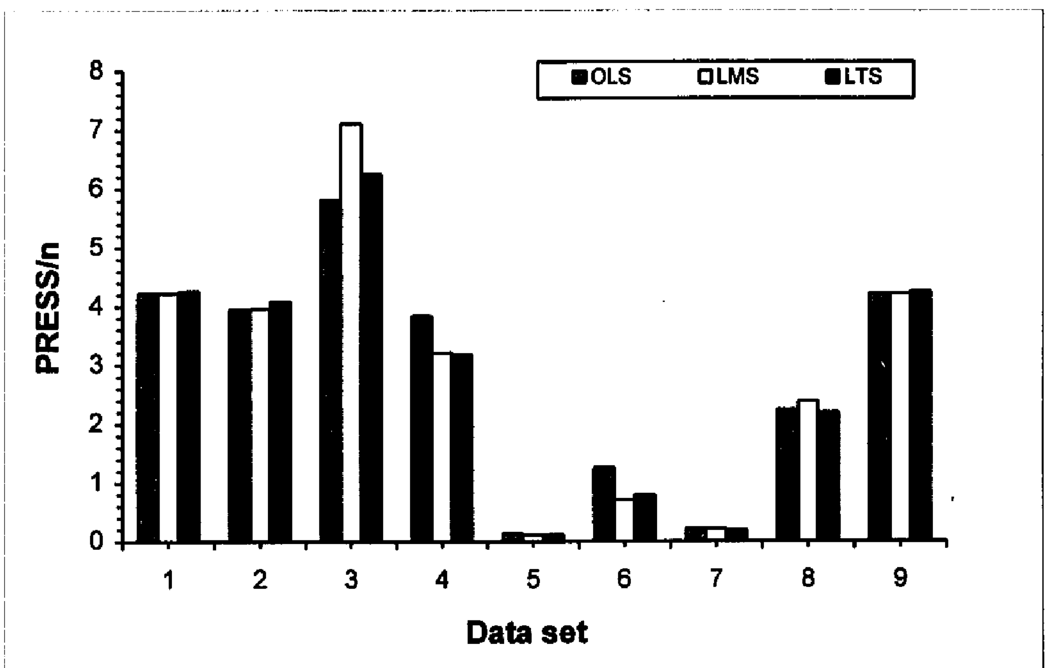


Figure 18. Histogram of PRESAE and PRESS for the combined variable volume equation fitted to the nine data sets.

Table 8. Average PRESAE and PRESS of the combined variable volume equation

Data	Average PRESAE (Rank)			Average PRESS (Rank)		
Set	OLS	LMS	LTS	OLS	LMS	LTS
1	1.4282 (3)	1.4101 (1)	1.4199 (2)	4.2193 (2)	4.2166 (1)	4.2530 (3)
2	1.4413 (2)	1.4387 (1)	1.4908 (3)	3.9496 (1)	3.9603 (2)	4.0799 (3)
3	1.6899 (3)	1.6392 (2)	1.5968 (1)	5.8188 (1)	7.1140 (3)	6.2588 (2)
4	1.2970 (3)	1.1906 (1)	1.2308 (2)	3.8413 (3)	3.2198 (2)	3.1837 (1)
5	0.2851 (3)	0.2445 (1)	0.2753 (2)	0.1349 (3)	0.1101 (1)	0.1146 (2)
6	0.7190 (3)	0.5708 (1)	0.5919 (2)	1.2682 (3)	0.7138 (1)	0.7941 (2)
7	0.3728 (1)	0.4098 (3)	0.3871 (2)	0.2128 (3)	0.2104 (2)	0.1960 (1)
8	1.1517 (2)	1.1937 (3)	1.1450 (1)	2.2536 (2)	2.3958 (3)	2.2093 (1)
9	1.4282 (3)	1.4101 (1)	1.4199 (2)	4.2193 (2)	4.2166 (1)	4.2530 (3)
Rank Sum	(23)	(14)	(17)	(20)	(16)	(18)

Schumacher and Hall volume equation

The Schumacher and Hall (1933) volume equation was also fitted to the nine data sets using OLS, LMS and LTS procedures. The estimated parameters are shown in Table 9.

In the Schumacher and Hall's equation, logarithm transformation is used to obtain a linear form. Because volume is really the variable of interest, the average PRESAE and PRESS were calculated in terms of volume V, not ln (V) (Figure 19 and Table 10).

Comparing with the combined variable volume equation, the Schumacher and Hall equation has one more independent variable. In terms of average PRESAE and PRESS, the Schumacher and Hall equation performed about the same as the combined volume equation on most of the data sets. However for data sets 1, 3 and 6, the OLS, LMS, and LTS coefficients of the Schumacher and Hall equation produced much smaller average PRESAE and PRESS (see tables 8 and 10). They indicated that the Schumacher and Hall equation fit the three data sets better than the combined variable equation.

Based on the average PRESAE and PRESS, the OLS and LTS regressions delivered similar overall performances, with the LMS a distant third.

Table 9. Estimated parameters of the Schumacher and Hall volume equation [1/]

Data	OLS			LMS			LTS		
Set	c_0	c_1	c_2	c_0	c_1	c_2	c_0	c_1	c_2
1	-5.3401	1.9790	0.8294	-5.8659	1.9033	1.0086	-5.3493	1.9000	0.8844
2	-5.2928	2.0570	0.7956	-4.7625	1.9747	0.7212	-4.8926	2.0297	0.7192
3	-5.4639	2.1625	0.7605	-5.6017	2.1635	0.7867	-4.1055	2.2064	0.4159
4	-6.1603	1.8449	1.1092	-5.5028	2.0255	0.8377	-5.6139	1.9169	0.9297
5	-2.9634	0.7623	0.7581	-3.4754	0.4578	1.0561	-2.4575	0.7178	0.6517
6	-6.1881	0.4296	1.8073	-6.0857	0.3457	1.8200	-5.9712	0.5205	1.7064
7	-8.1345	1.2861	1.9318	-8.3164	1.1516	2.0534	-8.2047	1.2328	1.9791
8	-5.4203	1.9389	0.9261	-5.6441	1.9978	0.9483	-6.2096	1.9397	1.1117
9	-7.4096	1.9548	1.3758	-7.9328	1.9248	1.5148	-7.9862	1.8778	1.5528

[1/] Equation $\ln(V) = c_0 + c_1 \ln(D) + c_2 \ln(H) + \varepsilon$, where V, D, and H are the individual tree volume (cu. ft.), DBH (in.), and total height (ft.), respectively

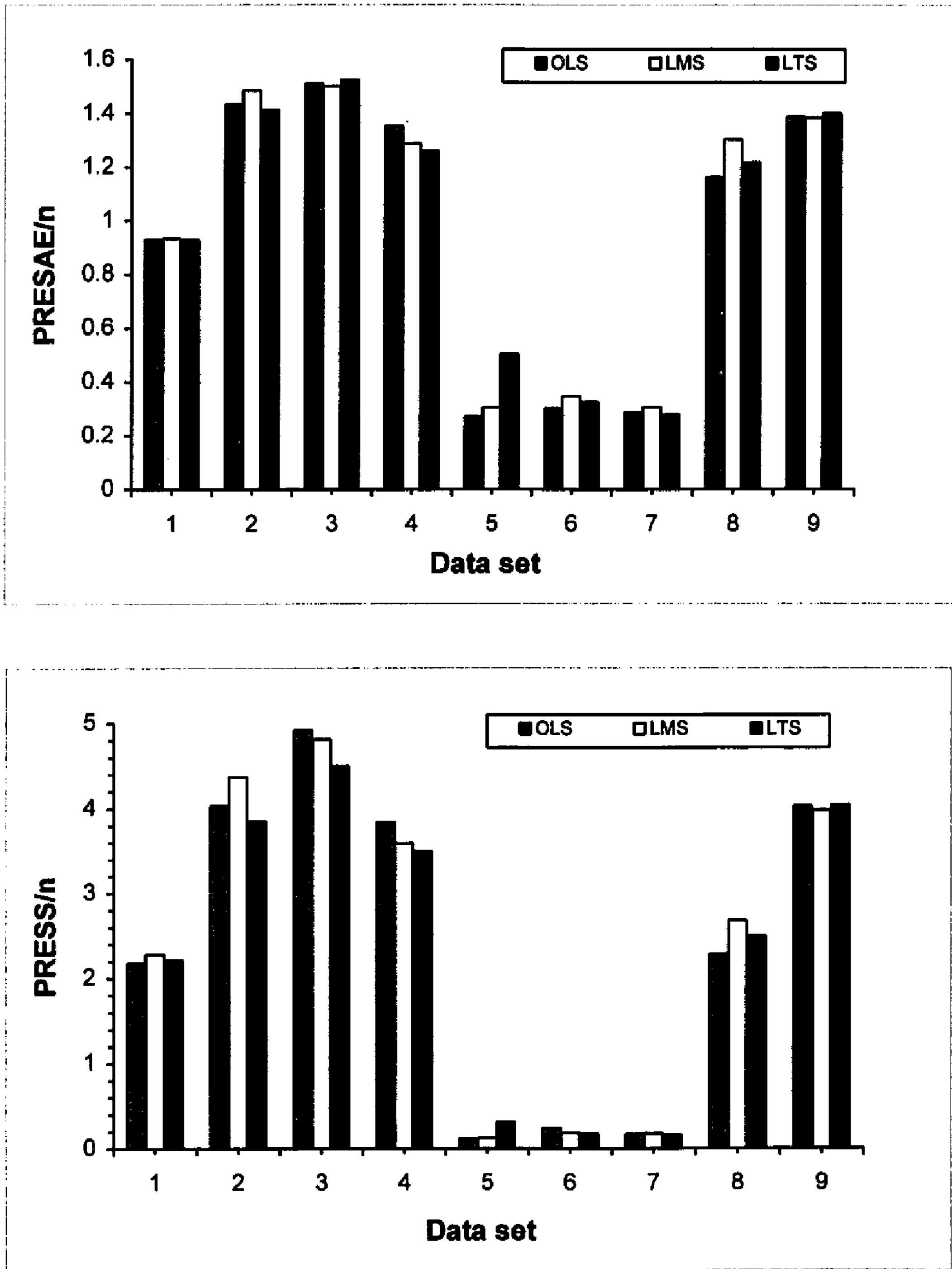


Figure 19. Histogram of PRESAE and PRESS for the Schumacher and Hall volume equation fitted to the nine data sets.

Table 10. Average PRESAE and PRESS of the Schumacher and Hall volume equation

Data	Average PRESAE (Rank)			Average PRESS (Rank)		
Set	OLS	LMS	LTS	OLS	LMS	LTS
1	0.9298 (2)	0.9323 (3)	0.9284 (1)	2.1788 (1)	2.2796 (3)	2.2153 (2)
2	1.4346 (2)	1.4867 (3)	1.4150 (1)	4.0392 (2)	4.3731 (3)	3.8590 (1)
3	1.5115 (2)	1.5019 (1)	1.5236 (3)	4.9214 (3)	4.8143 (2)	4.4986 (1)
4	1.3548 (3)	1.2887 (2)	1.2618 (1)	3.8408 (3)	3.5946 (2)	3.4948 (1)
5	0.2714 (1)	0.3045 (2)	0.5034 (3)	0.1186 (1)	0.1293 (2)	0.3133 (3)
6	0.3009 (1)	0.3464(3)	0.3253 (2)	0.2414 (3)	0.1891 (2)	0.1764 (1)
7	0.2867 (2)	0.3056 (3)	0.2775 (1)	0.1707 (2)	0.1767 (3)	0.1602 (1)
8	1.1629 (1)	1.3042 (3)	1.2175 (2)	2.2846 (1)	2.6797 (3)	2.4945 (2)
9	1.3893 (2)	1.3856 (1)	1.4015 (3)	4.0326 (2)	3.9844 (1)	4.0456 (3)
Rank Sum	(16)	(21)	(17)	(18)	(21)	(15)

Robust Estimation of Two Yield Equations

Yield information is essential for forest management planning. Most analyses of management strategies require growth and yield predictions. Yield tables for even-aged stands have been in use in the United States since 1900 (Burkhart 1985). A multiple regression approach to yield estimation was suggested by MacKinney et al. (1937), and subsequently used to construct a yield prediction equation (MacKinney and Chaiken 1939). Since that time, many studies have used multiple regression to predict yield. A review of literature showed that efforts have been mainly focused on developing model forms; robust procedures have not been investigated in estimating coefficients of yield models.

Data

The data set used to fit yield equations was obtained from the Southern Forest Experiment Station at Pineville, Louisiana, and include a total of 542 plots from unthinned loblolly pine plantations in the West Gulf region. Diameter at breast height was measured to the nearest 0.1 inch for each tree on the plots. The average height of dominant and codominant trees was obtained for each plot to determine site index. Total heights of some trees on each plot were measured to the nearest foot. Cubic-foot volume of each tree was predicted from an individual volume equation. Yield was calculated by summing volumes of all trees in each plot and converting to a per acre

basis. A more detailed description of the procedure can be found in Baldwin and Feducia (1987).

Methods

Two yield equations were fitted to the data set. The first one is from Burkhart et al. (1972):

$$\ln (V) = b_0 + b_1/A + b_2 (H/A) + b_3 N + b_4 A \ln (N) + \varepsilon$$

The second equation is from Goebel and Warner (1969):

$$\ln (V) = c_0 + c_1/A + c_2/H + c_3 N + c_4 A \ln (N) + c_5 A/\ln (N) + \varepsilon$$

In the above two models, V is cubic-foot volume per acre, H is average height in feet of dominant and codominant trees, A is stand age in years, N is number of trees per acre, b_j's and c_j's are regression coefficients, and ε is error term.

As with tree volume equations, coefficients of the yield models were estimated using OLS, LMS, and LTS procedures. These regression techniques were evaluated based on the average PRESAE and PRESS statistics.

Results

The estimated parameters for the Burkhart yield model are listed in table 11. Table 12 shows the estimated parameters for the Goebel and Warner model. Calculated average PRESAE and PRESS are shown in table 13 for both models.

Table 11. Estimated parameters of the Burkhart yield model [1/]

	b_0	b_1	b_2	b_3	b_4
OLS	4.124068	-17.112035	1.045114	0.000822	0.012336
LMS	6.077946	-16.814203	0.592183	0.000629	0.008140
LTS	6.216893	-18.526720	0.614207	0.000703	0.007058

[1/] Equation $\ln(V) = b_0 + b_1/A + b_2 (H/A) + b_3 N + b_4 A \ln(N) + \varepsilon$, where V is stand volume (cu. ft./acre), A is age (years), N is number of trees per acre, and H is average height (ft.) of dominants and codominants.

Table 12. Estimated parameters of the Goebel and Warner yield model [1/]

	c_0	c_1	c_2	c_3	c_4	c_5
OLS	8.574987	10.335761	-100.198637	0.000924	0.005477	-0.028450
LMS	9.874455	3.525613	-133.381047	0.001025	0.002614	-0.071104
LTS	9.163415	6.819897	-100.101032	0.000710	0.004768	-0.066364

[1/] Equation $\ln(V) = c_0 + c_1/A + c_2/H + c_3 N + c_4 A \ln(N) + c_5 A/\ln(N) + \varepsilon$, where V is stand volume (cu. ft./acre), A is age (years), N is number of trees per acre, and H is average height (ft.) of the dominants and codominants.

Table 13. Average PRESAE and PRESS of the two yield models

Model	Average PRESAE (Rank)			Average PRESS (Rank)		
	OLS	LMS	LTS	OLS	LMS	LTS
Burkhart	573 (3)	560 (1)	570 (2)	543158 (1)	652151 (2)	680922 (3)
Goebel	1061 (3)	753 (2)	522 (1)	1476132 (3)	975723 (2)	562603 (1)

For the Burkhart yield model, the OLS procedure ranked first in terms of average PRESS, bus last in terms of average PRESAE. On the other hand, the two robust regressions performed better than OLS based on both criteria. It was not clear, however, which robust procedure was more suitable for these data. The LMS procedure was better than LTS in fitting Burkhart et al. model, whereas the reverse was true for the Goebel and Warner model.

SUMMARY AND CONCLUSIONS

Robust Estimation of Population Means

The original objective of this study was to explore the applicability of robust estimation of population means and robust linear regression in forestry. At the early stage of the research, it was found that the robust estimators of population means possessed large bias for skewed distributions.

The bias problem was observed when a robust estimator ($Huber_{15}$) and the sample mean were used to estimate the population means of nine beta distributions with various combinations of parameters. For populations with skewness above 0.5, the bias component of the mean square error completely offset the claimed efficiency gain of the robust estimator.

Efforts were made to correct the bias problem of robust estimators. The resulting solution for the problem was to add a correction term to a robust estimator to remove its bias. The added term is basically an estimate of the bias caused by asymmetric property of a population distribution. As population skewness is rarely known, a new robust skewness measure was developed to describe sample skewness.

The analysis of bias from robust estimators revealed that the bias actually consisted of two parts, sample bias and truncation bias. After removing the sample bias component, a close relationship emerged between truncation bias and the robust

skewness measure. This relationship formed the basis for correcting robust estimation bias in case of asymmetric populations.

Simulations were conducted on four families of distributions and three different sample sizes. For each robust estimator, the relationship between truncation bias and the robust skewness measure remained almost constant for different families of distributions and sample sizes simulated. A set of parameters that determines the bias correction term for each of the five robust estimators were tabulated.

The five modified robust estimators were evaluated on both simulated data from theoretical distributions and an actual forestry data set. They were compared with their unmodified counterparts and the sample mean in terms of mean bias, standard deviation and square root of mean square error.

The modified robust estimators were comparable in terms of bias to the sample mean, which is a unbiased estimator of population mean regardless of population distributions. The modified robust estimates yielded much smaller bias than the original robust estimates for highly-skewed distributions.

The sample mean performed slightly better than the modified robust estimators in terms of standard deviation and square root of mean square error. The original robust estimators resulted in smaller standard deviation than their modified counterparts and the sample mean for some distributions. However, because of their high bias, the original robust estimators yielded much higher square root of mean

square error, a measure that takes into account both standard deviation and bias, than the modified robust estimators and the sample mean

The sample mean is an unbiased estimator for population mean, whether or not the population is symmetrically distributed. However, it is sensitive to outliers in samples. One single large outlier can "pull" the estimate far away from the true population mean. On the other hand, the five original robust estimators used in this research were robust to outliers, but produced biased estimates for skewed populations. The modified versions of these robust estimators were compromises between the above two types of estimators. They were robust to outliers, and comparable in terms of bias to the sample mean for both symmetric and skewed populations.

Among the five modified robust estimators, there were not many differences in performance in terms of bias, standard deviation and square root of mean square error. If one needs to select a robust estimator, the modified Huber_{15} can be recommended for its slightly better performance, and the modified median for its simplicity in calculation.

Applications of Robust Linear Regression

The second part of the research used ordinary least squares and two robust linear regression procedures, least median of squares and least trimmed squares, to fit two individual tree volume equations and two yield models. The two individual tree

volume equations were fitted to nine data sets and the yield models were fitted to one data set.

For each data set, the models were fitted by both of the robust procedures and the traditional ordinary least squares. Allen's PRESS statistics and a similar PRESAE statistics were calculated for all three estimation procedures to evaluate their prediction capabilities.

For most of the ten data sets used in this study, the model parameters estimated by the two robust procedures and the least squares method were very close. The calculated PRESS and PRESAE statistics were also similar.

However, among the ten data sets, at least one data set contained an extreme outlier (data set 6 on figure 17). When outliers are present, OLS estimation of model parameters could be affected severely and lead to wrong conclusions. In the case of the combined variable volume equation, the OLS estimation of model parameters for data set 6 was obviously pulled away from the main trend by a single outlier. As a result, the least-squares-estimated model exhibited higher prediction error than the two robust-estimated-models in terms of both PRESS and PRESAE. On the other hand, robust procedures LMS and LTS fitted the main pattern formed by majority of the data quite well.

Giving the fact that outliers may exist in a data set, and the OLS estimates may well be affected by the outliers, it is beneficial to scrutinize results from both OLS and

a robust procedure. Similar average PRESAE and PRESS values from OLS and a robust procedure provide evidence that the model fits the data well and no extreme outliers exist. Otherwise, measures should be taken to look at the model and data in more depth. The problem may be caused by outliers in the data set, or poor model specification. If the user is certain that the model is correctly specified, results from the robust procedure should be used.

Theoretically, LTS is more efficient than LMS. In this study, both of them performed equally well. From a practical point of view, it may not matter which one to use; the important point is to use a robust regression method to evaluate against the popular but unrobust OLS procedure.

LITERATURE CITED

Allen, D. M., J. O. Street, and A. Mazzoleni. 1973. PRESS used to study relationships between component cardiac weights and electrocardiographic patterns. p. 124-135 *in* Proc. of Univ. of Kentucky conf. on regression with a large number of predictor variables. W. O. Thompson and F. B. Cady (ed.)

Andrews, D. F., Bickel, P. J., Hampel, R. F., Huber, P. J., Rogers, W. H., and Tukey, J. H. 1972. Robust estimators of locations: Survey and advances. Princeton University Press. Princeton, N. J., 373p.

Avery, T. U., and H. E. Burkhart. 1994. Forest measurements. McGraw-Hill. New York. 408 p.

Baldwin, V. C. Jr., and D. P. Feducia. 1987. Loblolly pine growth and yield prediction for managed West Gulf plantations. U.S.D.A. For. Serv. Res. Pap. SO-236. 27 p.

Bickel, P. J. 1976. Another look at robustness: a review of review and some new development. Scand. J. Statist. 3:145-168.

Birkes, David, and Y. Dodge. 1993 Alternative methods of regression. Wiley, New York. 228 p.

Box, G. E. 1953. Non-normality and tests on variances. Biometrika. 40:318-335.

Burkhart, H. E. 1985. New developments in growth and yield prediction. Presented at Southern Forestry Symposium. Atlanta, Georgia. (unpublished).

Burkhart, H. E, R. C. Parker, M. R. Strub, and R. G. Oderwald. 1972. Yield of old-field loblolly pine plantations. Virginia Polytechnic Institute and State University Publication FWS-3-72. 49 p.

Efron, B. 1979. Bootstrap methods: another look at the jacknife. Ann. Statist. 7:1-26.

Efron, B. 1982. The jacknife, the bootstrap, and other resampling plans. CBMS 38, SIAM-NSF. 92 p.

Efron, B. 1987. Better bootstrap confidence intervals. J. Am. Statist. Assoc. 82:171-185.

Freeman, J. D., T. J. Henry, and D. W. McCarthy, Jr. 1992. Robust regression applied to estimation of object parameters from astronomical speckle interferometry. J. Opt. Soc. Am. A 9:2149-2158.

Geary, R. C. 1947. Testing for normality. Biometrika. 34:209-242.

Goebel, N. B., and J. R. Warner. 1969. Volume yield of loblolly pine plantations for a variety of sites in the South Carolina Piedmont. S. C. Agric. Exp. Sta. For. Res. Series No. 13. 15p.

Hampel, F. R. 1968. Contributions to the theory of robust estimation. Ph. D. thesis. University of California, Berkeley. Dissertation Abstracts International 30:432-B.

Hampel, F. R. 1971. A general qualitative definition of robustness. Ann. of Math. Statist. 42:1887-1096.

Hampel, F. R. 1973. Robust estimation: a condensed partial survey. Z. Wahr. verw. Gebiet. 27:87-104.

Hampel, F. R. 1974. The influence curve and its role in robust estimation. J. Am. Statist. Assoc. 69:383-393.

Hampel, F. R., Ronchetti, E. Rousseeuw, P. J., and Stahel, W. A., 1986. Robust Statistics: the approach Based on influence functions. Wiley, New York. 502 p.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey. 1983. Understanding robust and exploratory data analysis. Wiley. New York. 447 p.

Huber, P. J. 1964. Robust estimation of a location parameter. Ann. Math. Statist. 35:73-101.

Huber, P. J. 1973. Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Statist. 1:799:821.

Huber, P. J. 1981. Robust statistics. Wiley, New York. 308 p.

Jeffrey, S. H. 1932. An alternative to the rejection of outliers. Proc. Roy. Soc. London Ser. A137:78-87.

Jeffrey, S. H. 1961. Theory of probability. Clarendon Press, Oxford. 447 p.

Kassab, C. 1990. Studying economic change: are robust regression procedures needed? Rural. Soc. 55:357-375.

Koenker, R. W. 1982. Robust methods in econometrics. Econ. Rev. 1:213-255.

Léger C., and D. N. Politis. 1992. Bootstrap Technology and applications. Technometrics. 34:378-398.

MacKinney, A. L., F. X. Schumacher, and L. E. Chaiken. 1937. Construction of yield tables for non-normal loblolly pine stands. J. Agri. Res. 54:531-545.

MacKinney, A. L. and L. E. Chaiken. 1939. Volume, yield, and growth of loblolly pine in the Mid-Atlantic Coastal Region. U.S.D.A. For. Serv. Tech. Note No. 33. 30 p.

Marazzi, A., C. Ruffieux, and A. Randriamharisoa. 1988. Robust regression *in* biological assay: application to the evaluation of alternative experimental techniques. Experientia. 44:857-873.

Marazzi, A. 1991. Algorithms and programs for robust linear regression. p. 183-193 *in* W. Stahel and S. Weisberg (ed.). Directions in robust statistics and diagnostics. Springer-Verlag. New York.

Marianthi, M., W. A. Stahel, and E. Ronchett. 1991. Robust M-type testing procedures for linear models. p. 201-220. *in* W. Stahel and S. Weisberg (ed.). Directions in robust statistics and diagnostics. Springer-Verlag. New York.

Maronna, R. A., and V. J. Yohai. 1981. Asymptotic behavior of general M-estimates for regression and scale with random carries. Z. Wahrsch. verw. Geb. 58:7-20.

Peracchi, F. 1990. Robust M-estimators. Ecol. Rev. 9:1-30.

Ringland, J. T. 1983. Robust multiple comparisons. J. Am. Statist. Assoc. 78:145-151.

Rocker, D. M., Downs, G. W., and Rocke, A. J. 1982. Are robust estimators really necessary? Technometrics. 24:95-101.

Rousseeuw, P. J. 1984. Least median of squares regression. J. Am. Statist. Assoc. 79:871-880.

Rousseeuw, P. J., and A. Leroy. 1987. Robust regression and outlier detection. John Wiley. New York. 328 p.

Rousseeuw, P. J., and V. Yohai. 1984. Robust regression by means of S-estimators. p 256-272 *in* J. Franke, W. Härdle, and R. D. Martin (ed.) Robust and nonlinear time series analysis . Springer Verlag.

Schrader, R. M., and T. P. Hettmansperger. 1980. Robust analysis of variance based upon a likelihood ratio criterion. Biometrika. 67:93-101.

Schumacher, F. X., and F. S. Hall. 1933. Logarithmic expression of timber-tree volume. J. Agri. Res. 47:719-734.

Shapiro, S. S., and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). Biometrika. 52:591-611.

Siegel, A. F. 1982. Robust regression using repeated medians. Biometrika. 69:242:244.

Singhal, R. G., and S. J. Sheather. 1989. Robust estimation and testing. John Wiley & Sons. New York. 351 p.

Spurr, S. H. 1952. Forestry Inventory. The Ronald Press. New York. 476 p.

Stigler, R. S. 1977. Do robust estimators work with real data? Ana. Statist. 5:1055-1098.

Swinton, S. M., and R. P. King. 1991. Evaluating robust regression techniques for detrending crop yield data with nonnormal error. Amer. J. Agr. Econ. 446-451.

Tan, W. Y., and M. A. Tabatabai. 1985. Some robust ANOVA procedures under heteroscedasticity and nonnormality. Commun. Statist. -Simula. Computa. 14:1007-1026.

Tukey, M. L. 1960. A survey of sampling from contaminated distributions. p. 448-485 *in* Contributions to probability and statistics, I. Olkin (ed.). Stanford University Press. Calif.

Tukey, J. W. 1962. The future of data analysis. Ann. Statist. 3:145-168.

Tukey, J. W. 1970/71. Exploratory data analysis (Preliminary Edition). Addison-Wesley. Mass. 668 p.

van Zomeren, B. C. 1987. A comparison of some high-breakdown regression estimators. Comp. Statist. Quart. 4:207-218.

Yohai, V. J. 1987. High breakdown-point and high efficiency robust estimates for regression. Ann. of Statist. 15:642-656.

Yohai, V. J., and R. H. Zamar. 1988. High breakdown estimates of regression by means of the minimization of an efficient scale. J. Am. Statist. Assoc. 83:406-413.

APPENDIX A. TABLES

Table 14. Bias of sample mean, robust estimates and modified robust estimates for sample size = 30
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	0.0001	-0.0043	0.0009	-0.0043	0.0006	-0.0043	0.0013	-0.0043	0.0014	-0.0043	0.0013
B	-7	-0.0002	-0.0056	0.0009	-0.0056	0.0005	-0.0056	0.0012	-0.0056	0.0014	-0.0056	0.0013
B	-6	-0.0001	-0.0075	0.0014	-0.0075	0.0009	-0.0075	0.0019	-0.0075	0.0021	-0.0075	0.0021
B	-5	0.0001	-0.0105	0.0014	-0.0103	0.0008	-0.0106	0.0020	-0.0106	0.0022	-0.0106	0.0021
B	-4	0.0001	-0.0148	0.0010	-0.0137	0.0006	-0.0152	0.0014	-0.0150	0.0015	-0.0152	0.0015
B	-3	-0.0003	-0.0199	0.0002	-0.0158	0.0003	-0.0202	-0.0002	-0.0194	-0.0002	-0.0204	-0.0002
B	-2	0.0006	-0.0235	0.0013	-0.0140	0.0020	-0.0217	0.0013	-0.0191	0.0013	-0.0219	0.0011
B	-1	-0.0000	-0.0180	0.0020	-0.0088	0.0005	-0.0152	0.0018	-0.0111	0.0022	-0.0146	0.0022
B	0	0.0002	0.0011	0.0009	0.0002	0.0001	0.0003	0.0001	0.0003	0.0001	0.0004	0.0002
B	1	0.0012	0.0190	-0.0002	0.0098	0.0009	0.0162	0.0000	0.0121	-0.0006	0.0155	-0.0005
B	2	0.0006	0.0237	-0.0002	0.0142	-0.0011	0.0216	-0.0005	0.0190	-0.0006	0.0218	-0.0003
B	3	0.0012	0.0205	0.0008	0.0166	0.0007	0.0208	0.0012	0.0202	0.0012	0.0210	0.0012
B	4	-0.0001	0.0148	-0.0005	0.0137	-0.0001	0.0152	-0.0008	0.0151	-0.0009	0.0153	-0.0009
B	5	-0.0000	0.0105	-0.0014	0.0103	-0.0008	0.0106	-0.0020	0.0106	-0.0022	0.0106	-0.0022
B	6	-0.0003	0.0075	-0.0014	0.0075	-0.0009	0.0075	-0.0019	0.0075	-0.0020	0.0075	-0.0020
B	7	0.0003	0.0056	-0.0010	0.0056	-0.0006	0.0056	-0.0013	0.0056	-0.0015	0.0056	-0.0014
B	8	0.0003	0.0043	-0.0006	0.0043	-0.0003	0.0043	-0.0009	0.0043	-0.0010	0.0043	-0.0010
G	1	0.0122	0.3040	-0.0236	0.1727	0.0226	0.2729	-0.0034	0.2157	-0.0006	0.2695	-0.0039
G	2	0.0013	0.2770	-0.0156	0.1510	-0.0146	0.2448	-0.0162	0.2024	-0.0183	0.2418	-0.0182
G	3	-0.0048	0.2406	-0.0089	0.1652	-0.0154	0.2343	-0.0060	0.2171	-0.0051	0.2372	-0.0040
G	4	-0.0058	0.1971	-0.0129	0.1674	-0.0116	0.2022	-0.0121	0.1975	-0.0130	0.2036	-0.0128

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	-0.0030	0.1462	-0.0107	0.1372	-0.0053	0.1498	-0.0143	0.1488	-0.0161	0.1501	-0.0160
G	6	0.0024	0.1072	-0.0074	0.1052	-0.0018	0.1083	-0.0129	0.1082	-0.0148	0.1084	-0.0144
G	7	-0.0052	0.0795	-0.0186	0.0790	-0.0132	0.0797	-0.0243	0.0797	-0.0261	0.0797	-0.0256
G	8	0.0007	0.0599	-0.0068	0.0598	-0.0031	0.0599	-0.0109	0.0599	-0.0121	0.0599	-0.0118
L	1	-0.0010	0.0495	-0.0047	0.0268	0.0022	0.0436	-0.0019	0.0342	-0.0013	0.0432	-0.0019
L	2	-0.0025	0.1471	-0.0136	0.0847	0.0032	0.1340	-0.0051	0.1109	-0.0021	0.1340	-0.0041
L	3	0.0014	0.2786	-0.0206	0.1749	0.0050	0.2624	-0.0047	0.2279	0.0018	0.2618	-0.0042
L	4	-0.0052	0.3894	-0.0315	0.2496	-0.0038	0.3691	-0.0131	0.3277	-0.0028	0.3702	-0.0112
L	5	-0.0192	0.5238	-0.0201	0.3497	0.0066	0.5040	0.0031	0.4552	0.0146	0.5070	0.0063
L	6	-0.0193	0.5834	-0.0686	0.4018	-0.0248	0.5688	-0.0386	0.5189	-0.0227	0.5694	-0.0385
L	7	0.0179	0.6968	-0.0237	0.4932	0.0127	0.6787	0.0033	0.6258	0.0195	0.6820	0.0057
L	8	-0.0250	0.7930	-0.0387	0.5655	-0.0059	0.7757	-0.0113	0.7207	0.0064	0.7799	-0.0091
W	-1	-0.0003	-0.0046	-0.0001	-0.0027	-0.0006	-0.0040	-0.0003	-0.0033	-0.0003	-0.0041	-0.0003
W	0	-0.0004	-0.0014	-0.0002	-0.0003	0.0002	-0.0012	-0.0002	-0.0005	0.0003	-0.0007	0.0002
W	1	0.0004	0.1080	-0.0077	0.0525	-0.0016	0.0911	-0.0070	0.0669	-0.0105	0.0872	-0.0099
W	2	-0.0045	0.2909	-0.0109	0.1559	-0.0162	0.2545	-0.0152	0.2125	-0.0161	0.2534	-0.0153
W	3	0.0216	0.5346	-0.0019	0.3543	-0.0030	0.5071	0.0046	0.4613	0.0103	0.5115	0.0083
W	4	-0.0322	0.7779	-0.0533	0.5586	-0.0554	0.7673	-0.0386	0.7192	-0.0320	0.7749	-0.0348
W	5	0.0089	1.0540	-0.0270	0.8238	-0.0223	1.0537	-0.0158	1.0101	-0.0081	1.0623	-0.0145
W	6	-0.0318	1.3179	-0.0616	1.0653	-0.0549	1.3328	-0.0503	1.2888	-0.0456	1.3444	-0.0498
W	7	-0.0137	1.5277	-0.0764	1.2679	-0.0565	1.5521	-0.0659	1.5079	-0.0627	1.5624	-0.0697
W	8	0.0762	1.8562	-0.0102	1.6005	0.0203	1.8854	-0.0150	1.8435	-0.0179	1.8944	-0.0241

Table 15. Bias of sample mean, robust estimates and modified robust estimates for sample size = 50
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	-0.0000	-0.0043	0.0003	-0.0043	0.0001	-0.0043	0.0006	-0.0043	0.0007	-0.0043	0.0007
B	-7	0.0001	-0.0056	0.0008	-0.0056	0.0005	-0.0056	0.0012	-0.0056	0.0013	-0.0056	0.0013
B	-6	-0.0000	-0.0075	0.0009	-0.0075	0.0004	-0.0075	0.0014	-0.0075	0.0016	-0.0075	0.0015
B	-5	-0.0002	-0.0106	0.0006	-0.0104	0.0000	-0.0107	0.0011	-0.0107	0.0013	-0.0107	0.0013
B	-4	0.0003	-0.0151	0.0008	-0.0140	0.0004	-0.0155	0.0011	-0.0153	0.0013	-0.0155	0.0013
B	-3	0.0002	-0.0206	0.0004	-0.0164	0.0006	-0.0210	0.0000	-0.0203	0.0000	-0.0212	0.0000
B	-2	0.0006	-0.0241	0.0009	-0.0137	0.0021	-0.0217	0.0014	-0.0190	0.0014	-0.0219	0.0012
B	-1	0.0004	-0.0187	0.0013	-0.0085	0.0007	-0.0153	0.0016	-0.0108	0.0024	-0.0145	0.0022
B	0	-0.0008	-0.0008	-0.0009	-0.0008	-0.0009	-0.0008	-0.0009	-0.0008	-0.0009	-0.0008	-0.0009
B	1	-0.0004	0.0190	-0.0013	0.0083	-0.0008	0.0151	-0.0019	0.0107	-0.0026	0.0143	-0.0025
B	2	0.0000	0.0245	-0.0007	0.0142	-0.0018	0.0221	-0.0011	0.0194	-0.0011	0.0224	-0.0008
B	3	-0.0000	0.0207	-0.0001	0.0166	-0.0004	0.0211	0.0003	0.0205	0.0003	0.0214	0.0004
B	4	-0.0000	0.0151	-0.0009	0.0141	-0.0004	0.0155	-0.0012	0.0154	-0.0014	0.0155	-0.0014
B	5	-0.0003	0.0105	-0.0014	0.0103	-0.0008	0.0106	-0.0019	0.0106	-0.0022	0.0106	-0.0021
B	6	-0.0002	0.0075	-0.0014	0.0075	-0.0009	0.0075	-0.0019	0.0075	-0.0021	0.0075	-0.0020
B	7	0.0002	0.0056	-0.0004	0.0056	-0.0001	0.0056	-0.0008	0.0056	-0.0009	0.0056	-0.0009
B	8	0.0001	0.0043	-0.0004	0.0043	-0.0001	0.0043	-0.0007	0.0043	-0.0008	0.0043	-0.0007
G	1	0.0015	0.3098	-0.0213	0.1639	0.0176	0.2685	-0.0083	0.2024	-0.0115	0.2631	-0.0105
G	2	0.0007	0.2996	-0.0121	0.1636	-0.0125	0.2642	-0.0137	0.2209	-0.0138	0.2626	-0.0141
G	3	-0.0017	0.2548	-0.0014	0.1781	-0.0093	0.2487	0.0011	0.2323	0.0024	0.2520	0.0033
G	4	0.0035	0.2019	0.0034	0.1731	0.0038	0.2079	0.0052	0.2038	0.0048	0.2092	0.0046

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	0.0004	0.1485	-0.0061	0.1405	-0.0006	0.1518	-0.0102	0.1510	-0.0121	0.1521	-0.0120
G	6	-0.0001	0.1077	-0.0082	0.1060	-0.0024	0.1087	-0.0140	0.1085	-0.0160	0.1087	-0.0156
G	7	-0.0002	0.0797	-0.0082	0.0794	-0.0033	0.0799	-0.0134	0.0798	-0.0150	0.0799	-0.0146
G	8	0.0009	0.0600	-0.0043	0.0599	-0.0007	0.0600	-0.0083	0.0600	-0.0095	0.0600	-0.0092
L	1	0.0005	0.0492	-0.0015	0.0273	0.0053	0.0432	0.0010	0.0336	0.0012	0.0428	0.0011
L	2	-0.0032	0.1525	-0.0139	0.0876	0.0055	0.1392	-0.0038	0.1146	-0.0004	0.1382	-0.0036
L	3	-0.0107	0.2860	-0.0193	0.1710	0.0022	0.2648	-0.0057	0.2271	0.0003	0.2642	-0.0050
L	4	-0.0046	0.3919	-0.0223	0.2492	0.0052	0.3707	-0.0030	0.3275	0.0069	0.3707	-0.0020
L	5	-0.0172	0.5456	-0.0233	0.3613	0.0042	0.5200	-0.0032	0.4703	0.0110	0.5219	-0.0009
L	6	0.0005	0.6140	-0.0092	0.4219	0.0242	0.5930	0.0169	0.5411	0.0321	0.5943	0.0182
L	7	-0.0109	0.7094	-0.0118	0.4915	0.0173	0.6845	0.0115	0.6282	0.0272	0.6867	0.0131
L	8	0.0004	0.7816	-0.0016	0.5582	0.0344	0.7600	0.0252	0.7030	0.0427	0.7632	0.0274
W	-1	-0.0002	-0.0046	0.0001	-0.0026	-0.0005	-0.0040	-0.0001	-0.0031	-0.0001	-0.0040	-0.0001
W	0	-0.0004	-0.0042	-0.0015	-0.0011	-0.0001	-0.0021	0.0001	-0.0013	0.0003	-0.0018	0.0003
W	1	-0.0037	0.1025	-0.0113	0.0472	-0.0042	0.0851	-0.0106	0.0602	-0.0144	0.0808	-0.0138
W	2	-0.0045	0.2888	-0.0154	0.1568	-0.0132	0.2560	-0.0144	0.2121	-0.0155	0.2533	-0.0160
W	3	0.0066	0.5440	-0.0048	0.3483	-0.0151	0.5084	-0.0047	0.4593	-0.0002	0.5133	-0.0004
W	4	-0.0165	0.8005	-0.0196	0.5722	-0.0292	0.7824	-0.0108	0.7358	-0.0015	0.7912	-0.0056
W	5	-0.0174	1.0642	-0.0216	0.8204	-0.0264	1.0631	-0.0099	1.0188	-0.0013	1.0735	-0.0066
W	6	0.0059	1.3358	0.0007	1.0799	0.0009	1.3454	0.0091	1.3021	0.0150	1.3556	0.0087
W	7	-0.0074	1.5839	0.0250	1.3300	0.0352	1.6043	0.0283	1.5641	0.0310	1.6149	0.0250
W	8	-0.0404	1.8643	-0.0075	1.5961	0.0128	1.8988	-0.0064	1.8585	-0.0071	1.9103	-0.0130

Table 16. Bias of sample mean, robust estimates and modified robust estimates for sample size = 100
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	0.0000	-0.0043	0.0001	-0.0043	-0.0002	-0.0043	0.0003	-0.0043	0.0004	-0.0043	0.0004
B	-7	-0.0000	-0.0056	0.0003	-0.0056	-0.0000	-0.0056	0.0007	-0.0056	0.0008	-0.0056	0.0007
B	-6	-0.0001	-0.0075	0.0007	-0.0075	0.0003	-0.0075	0.0012	-0.0075	0.0014	-0.0075	0.0014
B	-5	-0.0001	-0.0106	0.0005	-0.0105	-0.0001	-0.0107	0.0011	-0.0107	0.0013	-0.0107	0.0013
B	-4	-0.0003	-0.0154	0.0002	-0.0145	-0.0003	-0.0158	0.0005	-0.0157	0.0007	-0.0158	0.0007
B	-3	0.0001	-0.0211	0.0001	-0.0170	0.0004	-0.0216	-0.0002	-0.0210	-0.0003	-0.0218	-0.0003
B	-2	0.0001	-0.0242	0.0007	-0.0137	0.0018	-0.0215	0.0013	-0.0187	0.0013	-0.0218	0.0010
B	-1	0.0003	-0.0189	0.0013	-0.0083	0.0007	-0.0152	0.0017	-0.0105	0.0026	-0.0142	0.0026
B	0	0.0003	0.0006	0.0004	0.0004	0.0003	0.0005	0.0003	0.0004	0.0003	0.0005	0.0003
B	1	-0.0008	0.0186	-0.0019	0.0081	-0.0009	0.0151	-0.0021	0.0103	-0.0029	0.0141	-0.0028
B	2	-0.0001	0.0244	-0.0004	0.0137	-0.0018	0.0215	-0.0012	0.0187	-0.0012	0.0218	-0.0009
B	3	-0.0001	0.0209	0.0001	0.0167	-0.0002	0.0214	0.0005	0.0207	0.0006	0.0216	0.0006
B	4	0.0002	0.0154	-0.0001	0.0144	0.0004	0.0157	-0.0004	0.0157	-0.0006	0.0158	-0.0005
B	5	-0.0001	0.0106	-0.0009	0.0105	-0.0003	0.0107	-0.0015	0.0107	-0.0017	0.0107	-0.0016
B	6	0.0001	0.0075	-0.0007	0.0075	-0.0003	0.0075	-0.0012	0.0075	-0.0014	0.0075	-0.0013
B	7	-0.0000	0.0056	-0.0005	0.0056	-0.0002	0.0056	-0.0009	0.0056	-0.0010	0.0056	-0.0010
B	8	-0.0000	0.0043	-0.0001	0.0043	0.0002	0.0043	-0.0004	0.0043	-0.0004	0.0043	-0.0004
G	1	-0.0068	0.3091	-0.0240	0.1506	0.0089	0.2567	-0.0193	0.1868	-0.0237	0.2474	-0.0252
G	2	0.0002	0.2986	-0.0082	0.1615	-0.0077	0.2626	-0.0091	0.2175	-0.0100	0.2594	-0.0110
G	3	0.0008	0.2601	0.0013	0.1800	-0.0091	0.2519	0.0020	0.2358	0.0038	0.2555	0.0044
G	4	0.0025	0.2031	0.0026	0.1736	0.0024	0.2090	0.0042	0.2048	0.0036	0.2103	0.0035

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	0.0006	0.1502	-0.0004	0.1428	0.0048	0.1536	-0.0045	0.1529	-0.0064	0.1538	-0.0062
G	6	-0.0003	0.1085	-0.0062	0.1071	-0.0004	0.1092	-0.0122	0.1091	-0.0143	0.1092	-0.0138
G	7	0.0007	0.0798	-0.0061	0.0796	-0.0013	0.0799	-0.0112	0.0799	-0.0129	0.0799	-0.0124
G	8	-0.0012	0.0600	-0.0056	0.0600	-0.0018	0.0600	-0.0096	0.0600	-0.0109	0.0600	-0.0105
L	1	-0.0006	0.0498	-0.0030	0.0265	0.0044	0.0431	-0.0006	0.0328	-0.0003	0.0425	-0.0006
L	2	-0.0047	0.1567	-0.0098	0.0865	0.0063	0.1405	-0.0017	0.1136	0.0002	0.1385	-0.0024
L	3	-0.0043	0.2926	-0.0133	0.1763	0.0109	0.2714	0.0020	0.2326	0.0087	0.2698	0.0018
L	4	-0.0049	0.3996	-0.0058	0.2500	0.0179	0.3734	0.0106	0.3275	0.0196	0.3720	0.0105
L	5	-0.0197	0.5508	-0.0109	0.3597	0.0132	0.5217	0.0079	0.4692	0.0209	0.5227	0.0094
L	6	0.0074	0.6388	0.0070	0.4405	0.0398	0.6165	0.0336	0.5634	0.0496	0.6173	0.0345
L	7	0.0028	0.7300	0.0107	0.5122	0.0425	0.7051	0.0353	0.6493	0.0526	0.7072	0.0370
L	8	-0.0181	0.7975	0.0021	0.5652	0.0371	0.7731	0.0287	0.7150	0.0478	0.7749	0.0296
W	-1	-0.0001	-0.0046	0.0002	-0.0025	-0.0005	-0.0040	-0.0000	-0.0031	-0.0000	-0.0039	-0.0000
W	0	0.0016	-0.0006	0.0018	0.0011	0.0020	0.0002	0.0021	0.0010	0.0024	0.0007	0.0025
W	1	-0.0028	0.1041	-0.0065	0.0456	-0.0026	0.0841	-0.0080	0.0577	-0.0132	0.0781	-0.0130
W	2	-0.0012	0.2971	-0.0104	0.1573	-0.0119	0.2578	-0.0143	0.2124	-0.0153	0.2549	-0.0159
W	3	0.0095	0.5564	0.0075	0.3601	-0.0011	0.5222	0.0098	0.4738	0.0160	0.5277	0.0148
W	4	-0.0095	0.8193	-0.0046	0.5847	-0.0196	0.7973	0.0004	0.7497	0.0089	0.8065	0.0060
W	5	-0.0068	1.0710	-0.0024	0.8243	-0.0087	1.0653	0.0065	1.0207	0.0158	1.0755	0.0098
W	6	-0.0274	1.3488	0.0005	1.0910	0.0012	1.3582	0.0087	1.3161	0.0162	1.3691	0.0090
W	7	-0.0044	1.5874	0.0626	1.3244	0.0648	1.6069	0.0685	1.5660	0.0725	1.6185	0.0668
W	8	-0.0100	1.8936	0.0484	1.6292	0.0654	1.9253	0.0459	1.8867	0.0451	1.9355	0.0380

Table 17. Standard deviation of sample mean, robust estimates and modified robust estimates for sample size = 30
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	0.0035	0.0000	0.0052	0.0000	0.0049	0.0000	0.0055	0.0000	0.0056	0.0000	0.0056
B	-7	0.0038	0.0000	0.0059	0.0001	0.0056	0.0000	0.0063	0.0000	0.0064	0.0000	0.0064
B	-6	0.0048	0.0003	0.0072	0.0004	0.0068	0.0002	0.0076	0.0002	0.0077	0.0002	0.0077
B	-5	0.0052	0.0005	0.0073	0.0008	0.0069	0.0004	0.0077	0.0005	0.0078	0.0004	0.0078
B	-4	0.0070	0.0020	0.0088	0.0031	0.0085	0.0020	0.0091	0.0022	0.0092	0.0020	0.0092
B	-3	0.0093	0.0056	0.0103	0.0072	0.0102	0.0061	0.0104	0.0066	0.0104	0.0061	0.0104
B	-2	0.0125	0.0127	0.0142	0.0129	0.0136	0.0133	0.0140	0.0140	0.0140	0.0137	0.0142
B	-1	0.0175	0.0215	0.0196	0.0183	0.0179	0.0193	0.0183	0.0191	0.0183	0.0196	0.0184
B	0	0.0207	0.0261	0.0222	0.0218	0.0209	0.0232	0.0210	0.0222	0.0209	0.0230	0.0210
B	1	0.0183	0.0227	0.0204	0.0194	0.0189	0.0203	0.0192	0.0202	0.0194	0.0208	0.0195
B	2	0.0120	0.0114	0.0134	0.0120	0.0128	0.0121	0.0132	0.0129	0.0132	0.0125	0.0133
B	3	0.0084	0.0050	0.0098	0.0066	0.0097	0.0054	0.0099	0.0059	0.0100	0.0054	0.0100
B	4	0.0070	0.0019	0.0084	0.0029	0.0081	0.0018	0.0086	0.0020	0.0087	0.0018	0.0087
B	5	0.0055	0.0005	0.0075	0.0008	0.0071	0.0004	0.0079	0.0004	0.0080	0.0004	0.0080
B	6	0.0046	0.0001	0.0061	0.0002	0.0057	0.0001	0.0064	0.0001	0.0066	0.0001	0.0065
B	7	0.0039	0.0000	0.0060	0.0001	0.0057	0.0000	0.0064	0.0000	0.0065	0.0000	0.0065
B	8	0.0033	0.0000	0.0050	0.0000	0.0047	0.0000	0.0053	0.0000	0.0054	0.0000	0.0053
G	1	0.3756	0.4542	0.4318	0.3832	0.3816	0.4028	0.3953	0.3962	0.3902	0.4061	0.3957
G	2	0.1775	0.1740	0.2025	0.1682	0.1875	0.1703	0.1929	0.1766	0.1903	0.1742	0.1925
G	3	0.1239	0.0905	0.1390	0.1064	0.1348	0.0990	0.1369	0.1065	0.1363	0.0996	0.1369
G	4	0.0881	0.0357	0.1061	0.0509	0.1032	0.0379	0.1080	0.0419	0.1088	0.0371	0.1088

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	0.0741	0.0161	0.0886	0.0234	0.0848	0.0150	0.0918	0.0162	0.0930	0.0145	0.0928
G	6	0.0625	0.0044	0.0799	0.0074	0.0758	0.0031	0.0842	0.0035	0.0857	0.0030	0.0853
G	7	0.0544	0.0011	0.0808	0.0020	0.0762	0.0007	0.0856	0.0007	0.0871	0.0007	0.0868
G	8	0.0395	0.0005	0.0564	0.0008	0.0532	0.0003	0.0599	0.0003	0.0610	0.0003	0.0607
L	1	0.0608	0.0729	0.0702	0.0620	0.0629	0.0654	0.0649	0.0643	0.0642	0.0659	0.0649
L	2	0.1277	0.1209	0.1404	0.1156	0.1326	0.1157	0.1358	0.1186	0.1349	0.1182	0.1363
L	3	0.1972	0.1604	0.2329	0.1619	0.2142	0.1587	0.2246	0.1648	0.2190	0.1619	0.2245
L	4	0.2625	0.1912	0.2927	0.1988	0.2697	0.1897	0.2782	0.2004	0.2728	0.1948	0.2802
L	5	0.3524	0.2338	0.3927	0.2491	0.3670	0.2365	0.3811	0.2506	0.3754	0.2411	0.3825
L	6	0.4096	0.2551	0.4830	0.2746	0.4414	0.2574	0.4639	0.2721	0.4536	0.2611	0.4664
L	7	0.4266	0.2464	0.5124	0.2683	0.4708	0.2469	0.4962	0.2612	0.4868	0.2494	0.4988
L	8	0.4851	0.2542	0.5397	0.2869	0.5035	0.2641	0.5288	0.2795	0.5216	0.2680	0.5330
W	-1	0.0055	0.0064	0.0059	0.0056	0.0056	0.0059	0.0057	0.0058	0.0058	0.0059	0.0058
W	0	0.0500	0.0642	0.0550	0.0527	0.0508	0.0567	0.0518	0.0537	0.0511	0.0559	0.0517
W	1	0.1106	0.1381	0.1230	0.1167	0.1126	0.1236	0.1152	0.1227	0.1160	0.1262	0.1167
W	2	0.1882	0.1832	0.2098	0.1831	0.1964	0.1828	0.2008	0.1916	0.2010	0.1885	0.2021
W	3	0.2693	0.2088	0.3172	0.2291	0.2970	0.2222	0.3096	0.2361	0.3057	0.2266	0.3115
W	4	0.4204	0.2495	0.4804	0.2995	0.4659	0.2723	0.4827	0.2931	0.4823	0.2741	0.4850
W	5	0.5633	0.2507	0.7038	0.3184	0.6749	0.2718	0.7137	0.2929	0.7141	0.2735	0.7204
W	6	0.7964	0.2613	0.9550	0.3375	0.9146	0.2795	0.9746	0.3017	0.9780	0.2789	0.9835
W	7	0.8904	0.3035	1.0524	0.3984	1.0015	0.3111	1.0695	0.3366	1.0734	0.3071	1.0802
W	8	1.0405	0.2673	1.3131	0.3747	1.2567	0.2851	1.3570	0.3138	1.3691	0.2829	1.3698

Table 18. Standard deviation of sample mean, robust estimates and modified robust estimates for sample size = 50
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	0.0028	0.0000	0.0041	0.0000	0.0038	0.0000	0.0043	0.0000	0.0044	0.0000	0.0044
B	-7	0.0031	0.0000	0.0046	0.0000	0.0043	0.0000	0.0049	0.0000	0.0050	0.0000	0.0049
B	-6	0.0037	0.0000	0.0048	0.0001	0.0045	0.0000	0.0051	0.0000	0.0052	0.0000	0.0052
B	-5	0.0041	0.0003	0.0057	0.0004	0.0054	0.0002	0.0060	0.0002	0.0061	0.0002	0.0061
B	-4	0.0054	0.0014	0.0068	0.0022	0.0066	0.0013	0.0069	0.0014	0.0070	0.0012	0.0070
B	-3	0.0073	0.0037	0.0085	0.0049	0.0084	0.0040	0.0085	0.0044	0.0086	0.0040	0.0086
B	-2	0.0096	0.0087	0.0106	0.0094	0.0103	0.0095	0.0105	0.0101	0.0105	0.0097	0.0106
B	-1	0.0151	0.0183	0.0169	0.0157	0.0155	0.0165	0.0160	0.0163	0.0159	0.0168	0.0161
B	0	0.0156	0.0206	0.0172	0.0166	0.0159	0.0177	0.0160	0.0168	0.0159	0.0175	0.0161
B	1	0.0145	0.0172	0.0160	0.0154	0.0153	0.0159	0.0155	0.0160	0.0157	0.0163	0.0158
B	2	0.0097	0.0085	0.0110	0.0092	0.0105	0.0091	0.0108	0.0098	0.0107	0.0094	0.0109
B	3	0.0072	0.0039	0.0083	0.0053	0.0082	0.0043	0.0084	0.0047	0.0084	0.0042	0.0084
B	4	0.0056	0.0014	0.0070	0.0020	0.0067	0.0012	0.0072	0.0013	0.0073	0.0011	0.0073
B	5	0.0044	0.0004	0.0057	0.0006	0.0054	0.0003	0.0060	0.0003	0.0061	0.0003	0.0061
B	6	0.0039	0.0000	0.0055	0.0001	0.0052	0.0000	0.0058	0.0000	0.0059	0.0000	0.0059
B	7	0.0030	0.0000	0.0044	0.0000	0.0042	0.0000	0.0047	0.0000	0.0048	0.0000	0.0047
B	8	0.0028	0.0000	0.0040	0.0000	0.0037	0.0000	0.0042	0.0000	0.0043	0.0000	0.0043
G	1	0.2810	0.3171	0.3039	0.2791	0.2857	0.2894	0.2938	0.2864	0.2931	0.2928	0.2943
G	2	0.1437	0.1372	0.1597	0.1359	0.1510	0.1378	0.1556	0.1442	0.1540	0.1418	0.1556
G	3	0.0927	0.0682	0.1110	0.0802	0.1065	0.0759	0.1106	0.0821	0.1102	0.0771	0.1117
G	4	0.0689	0.0250	0.0799	0.0364	0.0779	0.0257	0.0808	0.0282	0.0813	0.0250	0.0814

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	0.0524	0.0110	0.0666	0.0165	0.0639	0.0101	0.0692	0.0110	0.0702	0.0099	0.0701
G	6	0.0458	0.0032	0.0560	0.0053	0.0531	0.0024	0.0590	0.0027	0.0601	0.0023	0.0598
G	7	0.0403	0.0006	0.0537	0.0010	0.0507	0.0003	0.0570	0.0004	0.0580	0.0003	0.0577
G	8	0.0346	0.0001	0.0465	0.0002	0.0439	0.0000	0.0494	0.0001	0.0503	0.0000	0.0500
L	1	0.0463	0.0560	0.0521	0.0474	0.0471	0.0504	0.0487	0.0493	0.0483	0.0507	0.0488
L	2	0.1003	0.0960	0.1128	0.0933	0.1040	0.0929	0.1075	0.0957	0.1060	0.0950	0.1076
L	3	0.1531	0.1349	0.1749	0.1302	0.1597	0.1291	0.1652	0.1351	0.1624	0.1330	0.1660
L	4	0.1964	0.1533	0.2311	0.1559	0.2140	0.1527	0.2223	0.1605	0.2172	0.1563	0.2228
L	5	0.2583	0.1678	0.2895	0.1802	0.2635	0.1712	0.2770	0.1804	0.2696	0.1748	0.2789
L	6	0.2929	0.1899	0.3163	0.2029	0.2911	0.1892	0.2999	0.2014	0.2934	0.1929	0.3014
L	7	0.3475	0.1815	0.4013	0.2021	0.3699	0.1829	0.3908	0.1927	0.3815	0.1864	0.3937
L	8	0.3896	0.2128	0.4243	0.2407	0.3946	0.2197	0.4161	0.2336	0.4078	0.2217	0.4189
W	-1	0.0043	0.0049	0.0047	0.0043	0.0043	0.0044	0.0044	0.0044	0.0044	0.0045	0.0045
W	0	0.0387	0.0512	0.0415	0.0410	0.0389	0.0442	0.0390	0.0416	0.0386	0.0434	0.0390
W	1	0.0782	0.0992	0.0886	0.0824	0.0800	0.0876	0.0821	0.0864	0.0822	0.0894	0.0828
W	2	0.1496	0.1496	0.1688	0.1434	0.1557	0.1434	0.1593	0.1500	0.1588	0.1483	0.1606
W	3	0.2227	0.1706	0.2633	0.1923	0.2512	0.1838	0.2601	0.1966	0.2578	0.1869	0.2615
W	4	0.3376	0.1844	0.3809	0.2188	0.3623	0.1975	0.3756	0.2119	0.3720	0.1996	0.3773
W	5	0.4400	0.2064	0.5094	0.2642	0.4901	0.2198	0.5101	0.2382	0.5088	0.2187	0.5138
W	6	0.5539	0.2028	0.6157	0.2603	0.5924	0.2189	0.6258	0.2362	0.6283	0.2190	0.6331
W	7	0.6997	0.2186	0.7940	0.2915	0.7616	0.2272	0.8115	0.2458	0.8164	0.2246	0.8186
W	8	0.9016	0.2189	1.0130	0.3016	0.9716	0.2255	1.0467	0.2447	1.0573	0.2233	1.0591

Table 19. Standard deviation of sample mean, robust estimates and modified robust estimates for sample size = 100
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	0.0021	0.0000	0.0027	0.0000	0.0026	0.0000	0.0029	0.0000	0.0030	0.0000	0.0030
B	-7	0.0020	0.0000	0.0028	0.0000	0.0027	0.0000	0.0030	0.0000	0.0031	0.0000	0.0031
B	-6	0.0025	0.0000	0.0035	0.0000	0.0033	0.0000	0.0037	0.0000	0.0037	0.0000	0.0037
B	-5	0.0031	0.0001	0.0040	0.0003	0.0038	0.0001	0.0042	0.0001	0.0042	0.0001	0.0042
B	-4	0.0038	0.0007	0.0046	0.0011	0.0044	0.0006	0.0048	0.0007	0.0048	0.0006	0.0048
B	-3	0.0048	0.0027	0.0058	0.0037	0.0056	0.0030	0.0058	0.0032	0.0059	0.0029	0.0059
B	-2	0.0069	0.0068	0.0078	0.0069	0.0074	0.0070	0.0076	0.0074	0.0076	0.0072	0.0076
B	-1	0.0100	0.0128	0.0108	0.0108	0.0103	0.0115	0.0104	0.0113	0.0105	0.0118	0.0105
B	0	0.0107	0.0138	0.0120	0.0111	0.0109	0.0118	0.0110	0.0113	0.0110	0.0117	0.0111
B	1	0.0106	0.0128	0.0117	0.0108	0.0107	0.0114	0.0109	0.0112	0.0109	0.0115	0.0109
B	2	0.0069	0.0062	0.0079	0.0066	0.0076	0.0066	0.0078	0.0070	0.0077	0.0068	0.0078
B	3	0.0050	0.0026	0.0059	0.0036	0.0058	0.0029	0.0059	0.0032	0.0059	0.0029	0.0059
B	4	0.0038	0.0008	0.0046	0.0013	0.0045	0.0007	0.0048	0.0007	0.0048	0.0006	0.0048
B	5	0.0031	0.0001	0.0040	0.0002	0.0038	0.0001	0.0043	0.0001	0.0043	0.0001	0.0043
B	6	0.0025	0.0000	0.0034	0.0000	0.0032	0.0000	0.0036	0.0000	0.0037	0.0000	0.0037
B	7	0.0023	0.0000	0.0032	0.0000	0.0030	0.0000	0.0034	0.0000	0.0035	0.0000	0.0035
B	8	0.0021	0.0000	0.0029	0.0000	0.0027	0.0000	0.0031	0.0000	0.0031	0.0000	0.0031
G	1	0.2055	0.2422	0.2168	0.2135	0.2094	0.2230	0.2114	0.2205	0.2135	0.2262	0.2143
G	2	0.0975	0.0986	0.1116	0.0950	0.1022	0.0966	0.1053	0.1011	0.1041	0.0997	0.1050
G	3	0.0650	0.0459	0.0736	0.0556	0.0713	0.0520	0.0731	0.0564	0.0725	0.0525	0.0735
G	4	0.0466	0.0189	0.0540	0.0277	0.0530	0.0196	0.0550	0.0217	0.0555	0.0191	0.0554

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	0.0403	0.0065	0.0463	0.0105	0.0445	0.0052	0.0480	0.0057	0.0486	0.0050	0.0485
G	6	0.0313	0.0015	0.0407	0.0026	0.0385	0.0009	0.0430	0.0010	0.0437	0.0009	0.0435
G	7	0.0279	0.0002	0.0358	0.0004	0.0338	0.0001	0.0380	0.0001	0.0387	0.0001	0.0385
G	8	0.0252	0.0000	0.0329	0.0001	0.0311	0.0000	0.0350	0.0000	0.0356	0.0000	0.0354
L	1	0.0338	0.0384	0.0367	0.0339	0.0342	0.0352	0.0350	0.0348	0.0350	0.0355	0.0351
L	2	0.0688	0.0688	0.0756	0.0650	0.0720	0.0649	0.0733	0.0672	0.0732	0.0674	0.0739
L	3	0.1070	0.0937	0.1216	0.0914	0.1075	0.0905	0.1112	0.0946	0.1079	0.0930	0.1111
L	4	0.1386	0.1053	0.1583	0.1077	0.1447	0.1027	0.1498	0.1080	0.1445	0.1047	0.1490
L	5	0.1964	0.1194	0.2140	0.1299	0.1976	0.1189	0.2042	0.1262	0.1978	0.1208	0.2048
L	6	0.2131	0.1180	0.2334	0.1286	0.2148	0.1171	0.2224	0.1250	0.2167	0.1188	0.2229
L	7	0.2544	0.1300	0.2709	0.1426	0.2493	0.1313	0.2616	0.1397	0.2556	0.1335	0.2639
L	8	0.2790	0.1324	0.2935	0.1503	0.2726	0.1375	0.2853	0.1469	0.2788	0.1396	0.2874
W	-1	0.0030	0.0035	0.0034	0.0031	0.0031	0.0031	0.0032	0.0032	0.0032	0.0032	0.0033
W	0	0.0281	0.0352	0.0296	0.0292	0.0280	0.0310	0.0281	0.0294	0.0280	0.0304	0.0281
W	1	0.0602	0.0737	0.0667	0.0623	0.0611	0.0655	0.0623	0.0644	0.0624	0.0664	0.0627
W	2	0.1051	0.0981	0.1169	0.0982	0.1111	0.0975	0.1138	0.1020	0.1127	0.1001	0.1136
W	3	0.1600	0.1169	0.1776	0.1302	0.1715	0.1230	0.1739	0.1322	0.1721	0.1253	0.1749
W	4	0.2271	0.1290	0.2616	0.1572	0.2505	0.1417	0.2597	0.1537	0.2574	0.1430	0.2617
W	5	0.3078	0.1375	0.3535	0.1711	0.3386	0.1482	0.3547	0.1599	0.3538	0.1484	0.3582
W	6	0.3978	0.1457	0.4446	0.1936	0.4301	0.1558	0.4534	0.1692	0.4549	0.1546	0.4575
W	7	0.4654	0.1414	0.5144	0.1933	0.4932	0.1487	0.5232	0.1621	0.5247	0.1472	0.5281
W	8	0.5945	0.1469	0.6361	0.2052	0.6117	0.1519	0.6519	0.1657	0.6568	0.1506	0.6588

Table 20. Square root of MSE of sample mean, robust estimates and modified robust estimates for sample size = 30
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	0.0035	0.0043	0.0053	0.0043	0.0050	0.0043	0.0057	0.0043	0.0058	0.0043	0.0058
B	-7	0.0038	0.0056	0.0060	0.0056	0.0056	0.0056	0.0064	0.0056	0.0065	0.0056	0.0065
B	-6	0.0048	0.0075	0.0073	0.0075	0.0068	0.0075	0.0078	0.0075	0.0080	0.0075	0.0079
B	-5	0.0052	0.0105	0.0074	0.0103	0.0070	0.0106	0.0079	0.0106	0.0081	0.0106	0.0081
B	-4	0.0070	0.0150	0.0089	0.0140	0.0085	0.0153	0.0092	0.0152	0.0093	0.0154	0.0093
B	-3	0.0093	0.0206	0.0103	0.0174	0.0102	0.0211	0.0104	0.0205	0.0104	0.0213	0.0104
B	-2	0.0125	0.0267	0.0143	0.0190	0.0138	0.0255	0.0141	0.0237	0.0141	0.0259	0.0143
B	-1	0.0175	0.0280	0.0197	0.0203	0.0179	0.0246	0.0184	0.0221	0.0184	0.0244	0.0185
B	0	0.0207	0.0262	0.0222	0.0218	0.0209	0.0232	0.0210	0.0222	0.0209	0.0230	0.0210
B	1	0.0183	0.0296	0.0204	0.0218	0.0189	0.0260	0.0192	0.0236	0.0194	0.0259	0.0195
B	2	0.0121	0.0263	0.0134	0.0186	0.0129	0.0248	0.0132	0.0229	0.0132	0.0251	0.0133
B	3	0.0085	0.0211	0.0099	0.0178	0.0097	0.0215	0.0100	0.0210	0.0101	0.0217	0.0101
B	4	0.0070	0.0150	0.0084	0.0140	0.0081	0.0154	0.0086	0.0152	0.0087	0.0154	0.0087
B	5	0.0055	0.0105	0.0076	0.0103	0.0071	0.0106	0.0081	0.0106	0.0083	0.0106	0.0083
B	6	0.0046	0.0075	0.0062	0.0075	0.0058	0.0075	0.0067	0.0075	0.0069	0.0075	0.0068
B	7	0.0040	0.0056	0.0061	0.0056	0.0057	0.0056	0.0065	0.0056	0.0067	0.0056	0.0066
B	8	0.0033	0.0043	0.0050	0.0043	0.0047	0.0043	0.0054	0.0043	0.0055	0.0043	0.0054
G	1	0.3758	0.5466	0.4325	0.4203	0.3822	0.4865	0.3953	0.4511	0.3902	0.4874	0.3957
G	2	0.1775	0.3271	0.2031	0.2261	0.1880	0.2982	0.1936	0.2686	0.1912	0.2980	0.1933
G	3	0.1240	0.2571	0.1393	0.1965	0.1357	0.2543	0.1370	0.2419	0.1364	0.2573	0.1369
G	4	0.0883	0.2003	0.1069	0.1749	0.1039	0.2058	0.1087	0.2019	0.1096	0.2070	0.1095

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	0.0741	0.1471	0.0892	0.1392	0.0850	0.1506	0.0929	0.1497	0.0944	0.1508	0.0942
G	6	0.0625	0.1073	0.0803	0.1055	0.0758	0.1084	0.0852	0.1082	0.0869	0.1084	0.0865
G	7	0.0546	0.0795	0.0829	0.0791	0.0774	0.0797	0.0890	0.0797	0.0910	0.0797	0.0905
G	8	0.0396	0.0599	0.0569	0.0598	0.0533	0.0599	0.0609	0.0599	0.0622	0.0599	0.0618
L	1	0.0608	0.0881	0.0704	0.0675	0.0630	0.0787	0.0650	0.0728	0.0642	0.0788	0.0650
L	2	0.1277	0.1904	0.1410	0.1433	0.1326	0.1770	0.1359	0.1624	0.1349	0.1787	0.1363
L	3	0.1972	0.3215	0.2338	0.2384	0.2143	0.3066	0.2246	0.2812	0.2190	0.3078	0.2246
L	4	0.2626	0.4338	0.2944	0.3191	0.2698	0.4150	0.2785	0.3841	0.2728	0.4184	0.2804
L	5	0.3529	0.5737	0.3932	0.4293	0.3671	0.5567	0.3811	0.5196	0.3757	0.5614	0.3825
L	6	0.4100	0.6367	0.4879	0.4867	0.4421	0.6244	0.4655	0.5859	0.4542	0.6264	0.4680
L	7	0.4270	0.7391	0.5130	0.5615	0.4710	0.7222	0.4962	0.6781	0.4872	0.7262	0.4988
L	8	0.4857	0.8327	0.5411	0.6341	0.5036	0.8194	0.5290	0.7730	0.5216	0.8246	0.5331
W	-1	0.0055	0.0079	0.0059	0.0062	0.0057	0.0071	0.0058	0.0067	0.0058	0.0072	0.0058
W	0	0.0500	0.0642	0.0550	0.0527	0.0508	0.0567	0.0518	0.0537	0.0511	0.0559	0.0517
W	1	0.1106	0.1753	0.1233	0.1280	0.1126	0.1535	0.1154	0.1397	0.1165	0.1534	0.1171
W	2	0.1883	0.3438	0.2101	0.2405	0.1971	0.3133	0.2014	0.2861	0.2017	0.3158	0.2027
W	3	0.2702	0.5739	0.3172	0.4219	0.2970	0.5537	0.3097	0.5182	0.3059	0.5595	0.3116
W	4	0.4216	0.8169	0.4834	0.6338	0.4692	0.8142	0.4842	0.7766	0.4833	0.8219	0.4862
W	5	0.5634	1.0834	0.7043	0.8832	0.6753	1.0882	0.7139	1.0517	0.7141	1.0969	0.7205
W	6	0.7971	1.3435	0.9570	1.1175	0.9162	1.3618	0.9758	1.3237	0.9790	1.3731	0.9847
W	7	0.8905	1.5576	1.0552	1.3290	1.0031	1.5830	1.0715	1.5450	1.0752	1.5923	1.0824
W	8	1.0433	1.8754	1.3131	1.6438	1.2569	1.9068	1.3571	1.8701	1.3692	1.9154	1.3700

Table 21. Square root of MSE of sample mean, robust estimates and modified robust estimates for sample size = 50
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	0.0028	0.0043	0.0041	0.0043	0.0038	0.0043	0.0044	0.0043	0.0044	0.0043	0.0044
B	-7	0.0031	0.0056	0.0047	0.0056	0.0044	0.0056	0.0050	0.0056	0.0051	0.0056	0.0051
B	-6	0.0037	0.0075	0.0049	0.0075	0.0046	0.0075	0.0053	0.0075	0.0054	0.0075	0.0054
B	-5	0.0042	0.0106	0.0057	0.0104	0.0054	0.0107	0.0061	0.0107	0.0063	0.0107	0.0063
B	-4	0.0054	0.0151	0.0068	0.0142	0.0066	0.0155	0.0070	0.0154	0.0071	0.0156	0.0071
B	-3	0.0073	0.0209	0.0085	0.0171	0.0084	0.0214	0.0085	0.0208	0.0086	0.0216	0.0086
B	-2	0.0096	0.0257	0.0106	0.0167	0.0105	0.0237	0.0106	0.0215	0.0106	0.0240	0.0107
B	-1	0.0151	0.0262	0.0170	0.0178	0.0155	0.0225	0.0160	0.0196	0.0161	0.0222	0.0162
B	0	0.0157	0.0206	0.0172	0.0166	0.0159	0.0177	0.0160	0.0168	0.0159	0.0175	0.0161
B	1	0.0145	0.0256	0.0160	0.0175	0.0153	0.0220	0.0157	0.0192	0.0160	0.0217	0.0160
B	2	0.0097	0.0259	0.0111	0.0169	0.0107	0.0239	0.0108	0.0217	0.0108	0.0243	0.0109
B	3	0.0072	0.0211	0.0083	0.0174	0.0082	0.0216	0.0084	0.0210	0.0084	0.0218	0.0084
B	4	0.0056	0.0152	0.0070	0.0142	0.0067	0.0155	0.0073	0.0154	0.0074	0.0156	0.0074
B	5	0.0044	0.0105	0.0059	0.0104	0.0055	0.0106	0.0063	0.0106	0.0065	0.0106	0.0065
B	6	0.0039	0.0075	0.0057	0.0075	0.0053	0.0075	0.0061	0.0075	0.0063	0.0075	0.0062
B	7	0.0030	0.0056	0.0044	0.0056	0.0042	0.0056	0.0047	0.0056	0.0048	0.0056	0.0048
B	8	0.0028	0.0043	0.0040	0.0043	0.0037	0.0043	0.0043	0.0043	0.0043	0.0043	0.0043
G	1	0.2810	0.4433	0.3046	0.3236	0.2862	0.3947	0.2939	0.3507	0.2933	0.3937	0.2945
G	2	0.1437	0.3295	0.1602	0.2126	0.1515	0.2979	0.1562	0.2638	0.1546	0.2985	0.1562
G	3	0.0928	0.2638	0.1110	0.1954	0.1069	0.2600	0.1106	0.2464	0.1102	0.2635	0.1118
G	4	0.0690	0.2034	0.0799	0.1768	0.0780	0.2094	0.0810	0.2057	0.0814	0.2107	0.0815

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	0.0524	0.1489	0.0669	0.1415	0.0639	0.1521	0.0700	0.1514	0.0713	0.1524	0.0711
G	6	0.0458	0.1078	0.0566	0.1061	0.0532	0.1087	0.0607	0.1085	0.0621	0.1087	0.0618
G	7	0.0403	0.0797	0.0543	0.0794	0.0508	0.0799	0.0585	0.0798	0.0599	0.0799	0.0596
G	8	0.0346	0.0600	0.0467	0.0599	0.0439	0.0600	0.0500	0.0600	0.0511	0.0600	0.0509
L	1	0.0463	0.0745	0.0522	0.0547	0.0474	0.0664	0.0487	0.0596	0.0483	0.0663	0.0488
L	2	0.1003	0.1802	0.1137	0.1280	0.1042	0.1674	0.1076	0.1493	0.1060	0.1677	0.1077
L	3	0.1535	0.3162	0.1760	0.2149	0.1597	0.2946	0.1653	0.2642	0.1624	0.2958	0.1661
L	4	0.1965	0.4208	0.2322	0.2940	0.2141	0.4010	0.2224	0.3647	0.2173	0.4023	0.2229
L	5	0.2589	0.5708	0.2905	0.4037	0.2636	0.5474	0.2770	0.5038	0.2698	0.5504	0.2789
L	6	0.2929	0.6427	0.3165	0.4682	0.2921	0.6225	0.3004	0.5774	0.2952	0.6248	0.3020
L	7	0.3476	0.7323	0.4015	0.5315	0.3703	0.7085	0.3910	0.6571	0.3824	0.7115	0.3939
L	8	0.3896	0.8101	0.4243	0.6079	0.3961	0.7911	0.4169	0.7408	0.4100	0.7948	0.4197
W	-1	0.0043	0.0067	0.0047	0.0050	0.0044	0.0060	0.0044	0.0054	0.0044	0.0060	0.0045
W	0	0.0387	0.0514	0.0416	0.0411	0.0389	0.0442	0.0390	0.0416	0.0386	0.0435	0.0390
W	1	0.0783	0.1426	0.0893	0.0949	0.0801	0.1222	0.0827	0.1053	0.0835	0.1205	0.0840
W	2	0.1497	0.3253	0.1695	0.2125	0.1562	0.2934	0.1599	0.2597	0.1595	0.2935	0.1614
W	3	0.2227	0.5701	0.2634	0.3978	0.2516	0.5406	0.2602	0.4996	0.2578	0.5462	0.2615
W	4	0.3380	0.8214	0.3814	0.6126	0.3635	0.8069	0.3757	0.7658	0.3720	0.8160	0.3774
W	5	0.4403	1.0841	0.5099	0.8619	0.4909	1.0855	0.5102	1.0462	0.5088	1.0956	0.5138
W	6	0.5539	1.3511	0.6157	1.1109	0.5924	1.3631	0.6259	1.3233	0.6284	1.3732	0.6331
W	7	0.6997	1.5989	0.7944	1.3616	0.7624	1.6203	0.8120	1.5833	0.8170	1.6305	0.8190
W	8	0.9025	1.8771	1.0131	1.6244	0.9717	1.9122	1.0467	1.8745	1.0573	1.9233	1.0592

Table 22. Square root of MSE of sample mean, robust estimates and modified robust estimates for sample size = 100
(Four distributions: B - beta, G - gamma, L - lognormal, and W - Weibull)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
B	-8	0.0021	0.0043	0.0027	0.0043	0.0026	0.0043	0.0029	0.0043	0.0030	0.0043	0.0030
B	-7	0.0020	0.0056	0.0029	0.0056	0.0027	0.0056	0.0031	0.0056	0.0032	0.0056	0.0032
B	-6	0.0025	0.0075	0.0035	0.0075	0.0033	0.0075	0.0039	0.0075	0.0040	0.0075	0.0039
B	-5	0.0031	0.0106	0.0040	0.0105	0.0038	0.0107	0.0043	0.0107	0.0044	0.0107	0.0044
B	-4	0.0038	0.0154	0.0046	0.0145	0.0044	0.0158	0.0048	0.0157	0.0049	0.0158	0.0049
B	-3	0.0048	0.0213	0.0058	0.0174	0.0056	0.0218	0.0058	0.0212	0.0059	0.0220	0.0059
B	-2	0.0069	0.0251	0.0078	0.0154	0.0076	0.0226	0.0077	0.0202	0.0077	0.0230	0.0077
B	-1	0.0100	0.0228	0.0109	0.0136	0.0103	0.0191	0.0105	0.0154	0.0108	0.0184	0.0109
B	0	0.0107	0.0139	0.0120	0.0112	0.0109	0.0118	0.0110	0.0113	0.0110	0.0117	0.0111
B	1	0.0106	0.0226	0.0118	0.0136	0.0107	0.0189	0.0111	0.0152	0.0113	0.0182	0.0113
B	2	0.0069	0.0251	0.0079	0.0152	0.0078	0.0225	0.0079	0.0200	0.0078	0.0228	0.0079
B	3	0.0050	0.0211	0.0059	0.0170	0.0058	0.0216	0.0059	0.0209	0.0059	0.0218	0.0060
B	4	0.0038	0.0154	0.0046	0.0145	0.0045	0.0158	0.0048	0.0157	0.0049	0.0158	0.0049
B	5	0.0031	0.0106	0.0041	0.0105	0.0038	0.0107	0.0045	0.0107	0.0046	0.0107	0.0046
B	6	0.0025	0.0075	0.0035	0.0075	0.0032	0.0075	0.0038	0.0075	0.0039	0.0075	0.0039
B	7	0.0023	0.0056	0.0033	0.0056	0.0030	0.0056	0.0035	0.0056	0.0036	0.0056	0.0036
B	8	0.0021	0.0043	0.0029	0.0043	0.0027	0.0043	0.0031	0.0043	0.0032	0.0043	0.0031
G	1	0.2056	0.3927	0.2181	0.2612	0.2096	0.3401	0.2123	0.2890	0.2148	0.3353	0.2158
G	2	0.0975	0.3145	0.1119	0.1874	0.1025	0.2798	0.1056	0.2399	0.1045	0.2779	0.1056
G	3	0.0650	0.2641	0.0736	0.1884	0.0719	0.2572	0.0731	0.2424	0.0726	0.2608	0.0737
G	4	0.0466	0.2039	0.0541	0.1758	0.0530	0.2099	0.0552	0.2059	0.0556	0.2111	0.0555

(Table con'd.)

Dist.	Popul. Skew.	Sample mean	Sample median	Modif. median	$Huber_{15}$	Modif. $Huber_{15}$	Hampel	Modif. Hampel	Andrews	Modif. Andrews	Tukey	Modif. Tukey
G	5	0.0404	0.1504	0.0463	0.1432	0.0448	0.1537	0.0482	0.1530	0.0491	0.1539	0.0489
G	6	0.0313	0.1085	0.0411	0.1072	0.0385	0.1092	0.0447	0.1091	0.0460	0.1092	0.0457
G	7	0.0279	0.0798	0.0363	0.0797	0.0338	0.0799	0.0396	0.0799	0.0407	0.0799	0.0404
G	8	0.0253	0.0600	0.0334	0.0600	0.0311	0.0600	0.0362	0.0600	0.0372	0.0600	0.0370
L	1	0.0338	0.0629	0.0368	0.0430	0.0345	0.0556	0.0350	0.0478	0.0350	0.0554	0.0351
L	2	0.0690	0.1712	0.0763	0.1082	0.0723	0.1548	0.0734	0.1320	0.0732	0.1541	0.0740
L	3	0.1071	0.3072	0.1224	0.1986	0.1080	0.2861	0.1112	0.2511	0.1083	0.2853	0.1112
L	4	0.1387	0.4132	0.1584	0.2722	0.1458	0.3872	0.1501	0.3449	0.1458	0.3864	0.1493
L	5	0.1973	0.5636	0.2143	0.3824	0.1980	0.5351	0.2044	0.4859	0.1989	0.5364	0.2050
L	6	0.2132	0.6496	0.2335	0.4589	0.2185	0.6275	0.2249	0.5771	0.2223	0.6287	0.2256
L	7	0.2544	0.7415	0.2711	0.5317	0.2529	0.7172	0.2639	0.6641	0.2609	0.7197	0.2665
L	8	0.2796	0.8084	0.2935	0.5848	0.2751	0.7852	0.2868	0.7300	0.2829	0.7874	0.2889
W	-1	0.0030	0.0058	0.0034	0.0039	0.0032	0.0051	0.0032	0.0044	0.0032	0.0051	0.0033
W	0	0.0281	0.0352	0.0297	0.0292	0.0281	0.0310	0.0282	0.0294	0.0281	0.0304	0.0282
W	1	0.0603	0.1276	0.0670	0.0772	0.0611	0.1066	0.0628	0.0865	0.0637	0.1025	0.0640
W	2	0.1051	0.3129	0.1174	0.1855	0.1118	0.2756	0.1147	0.2356	0.1138	0.2739	0.1147
W	3	0.1603	0.5686	0.1778	0.3829	0.1715	0.5365	0.1742	0.4919	0.1729	0.5423	0.1755
W	4	0.2273	0.8294	0.2617	0.6054	0.2512	0.8098	0.2597	0.7653	0.2575	0.8191	0.2618
W	5	0.3079	1.0798	0.3535	0.8418	0.3387	1.0755	0.3548	1.0331	0.3542	1.0857	0.3583
W	6	0.3988	1.3566	0.4446	1.1080	0.4301	1.3671	0.4535	1.3269	0.4552	1.3778	0.4576
W	7	0.4654	1.5937	0.5182	1.3384	0.4974	1.6137	0.5276	1.5744	0.5296	1.6252	0.5323
W	8	0.5945	1.8993	0.6380	1.6421	0.6152	1.9313	0.6535	1.8940	0.6584	1.9413	0.6599

APPENDIX B. FIGURES

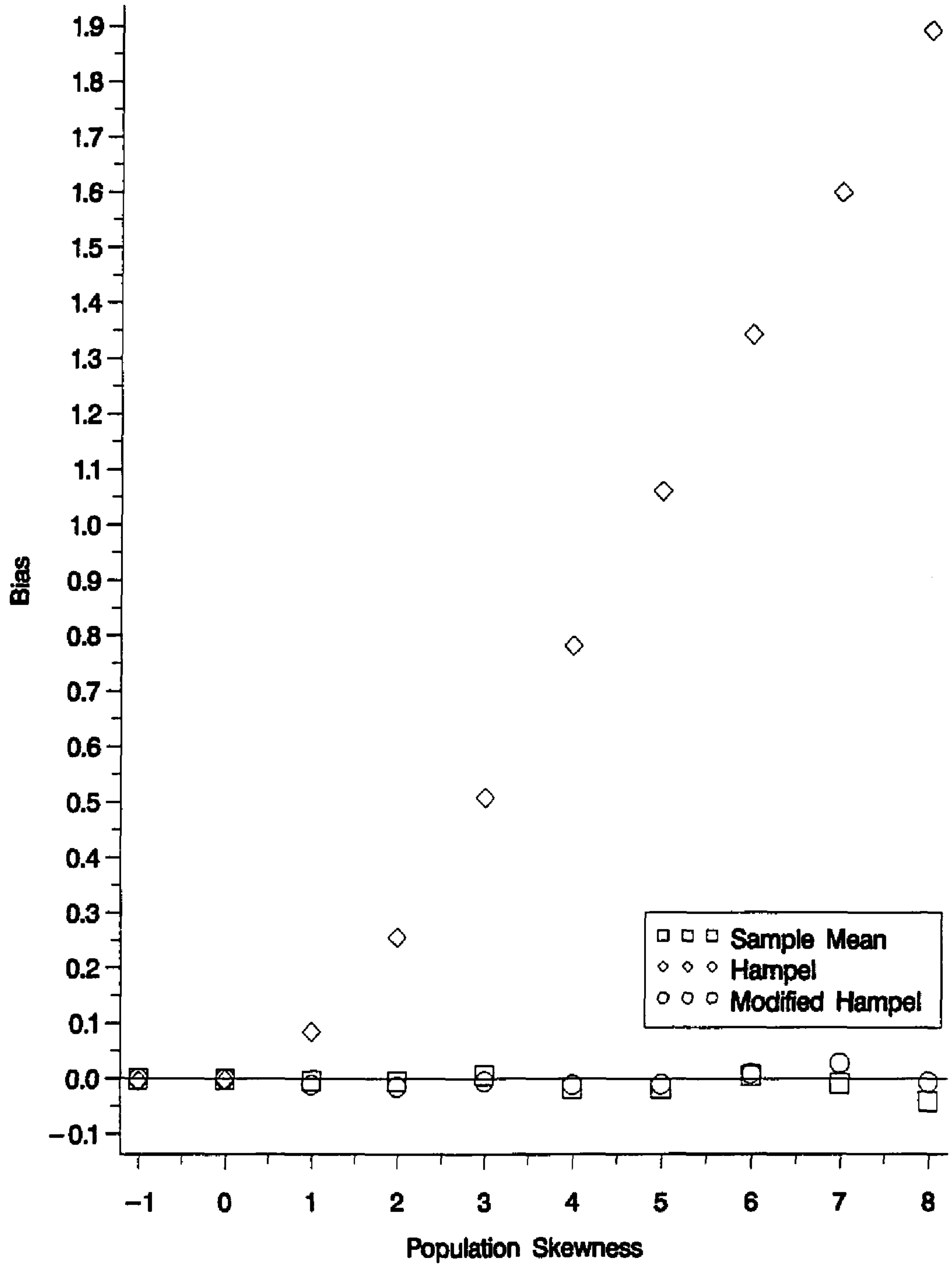


Figure 20. Bias of sample mean, Hampel and modified Hampel$_{15}$ vs. population skewness (Weibull distributions with sample size = 50)

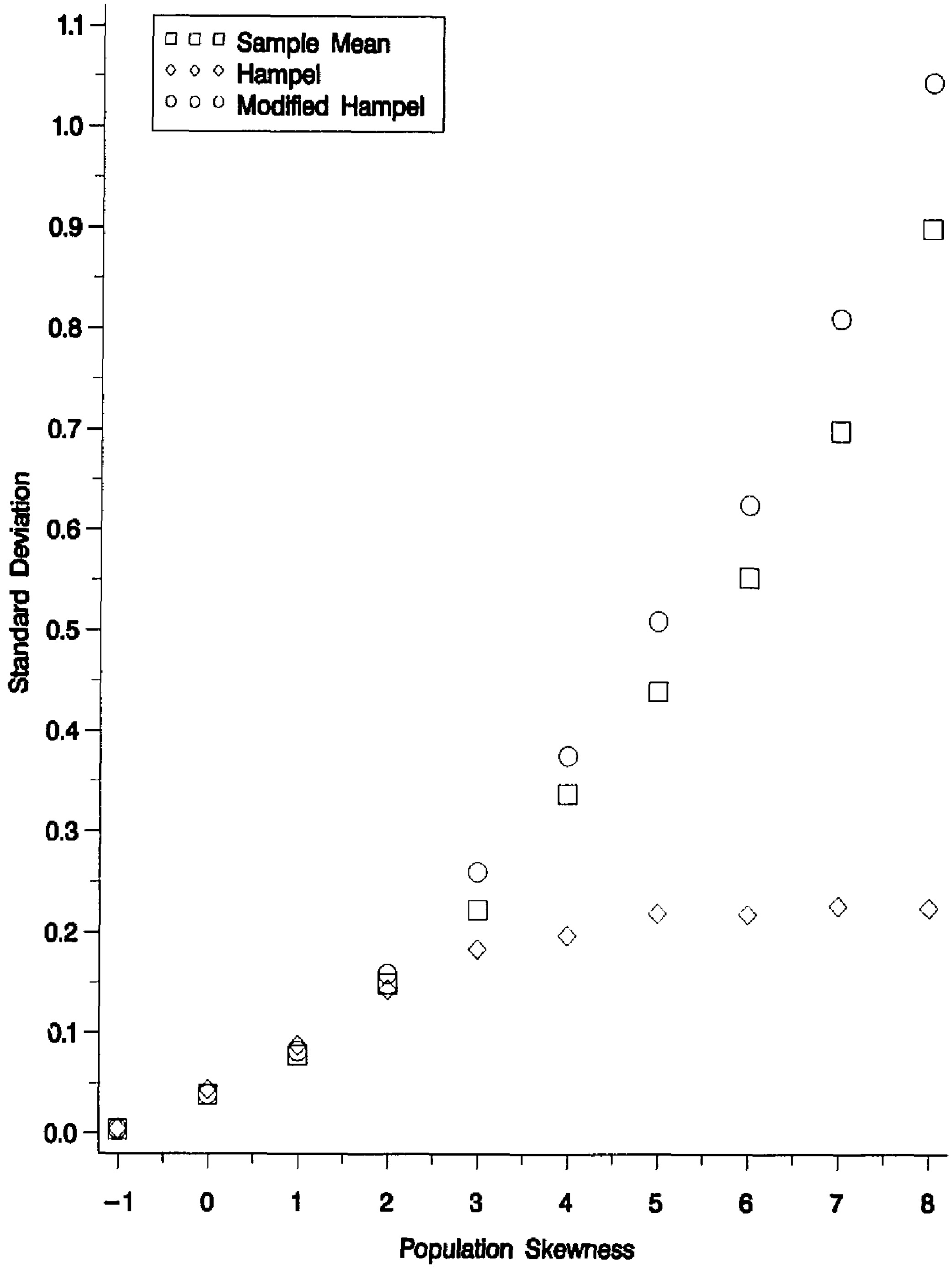


Figure 21. SD of sample mean, Hampel and modified Hampel vs. population skewness (Weibull distributions with sample size = 50)

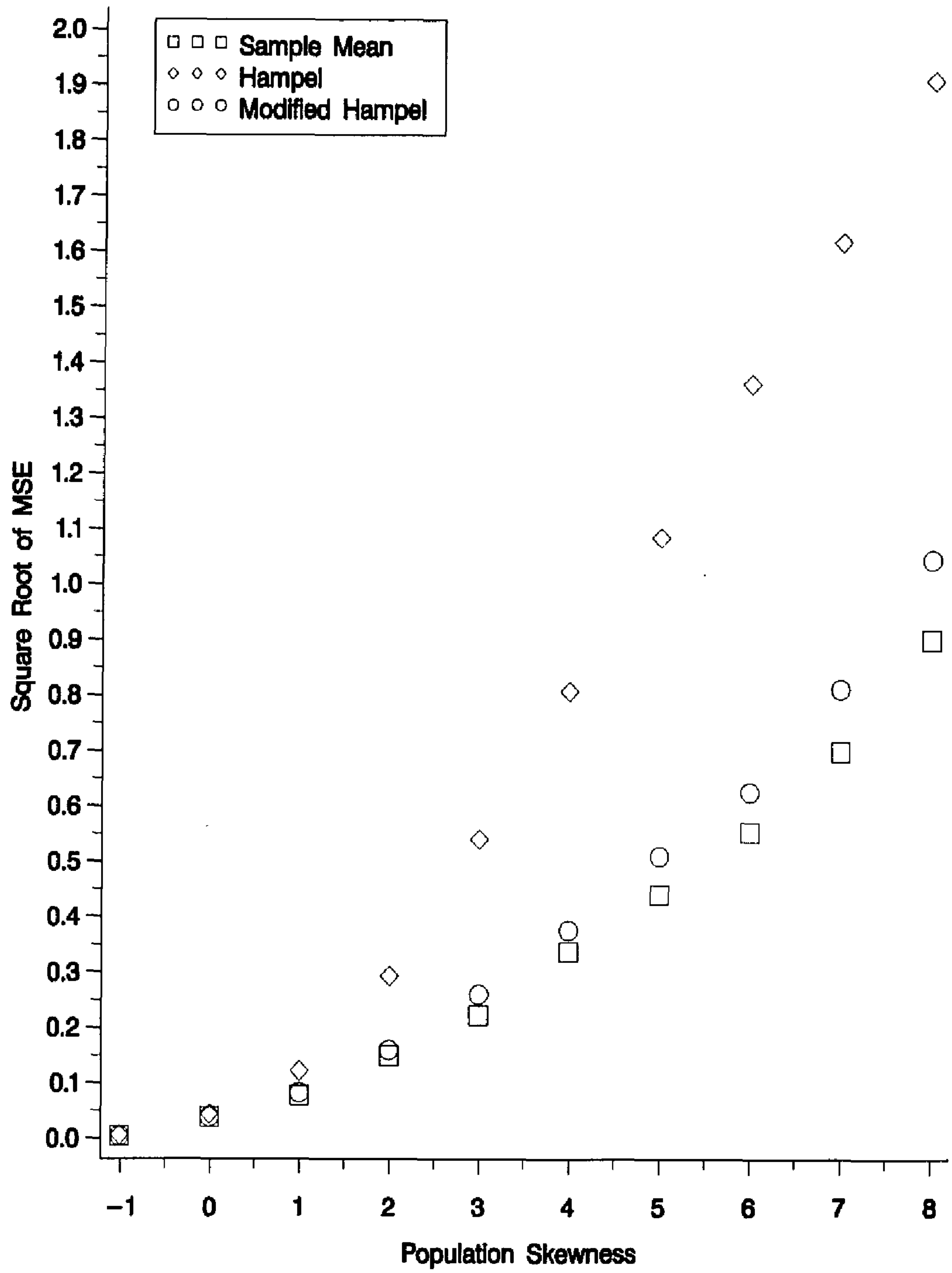


Figure 22. $\sqrt{\text{MSE}}$ of sample mean, Hampel and modified Hampel vs. population skewness (Weibull distributions with sample size = 50)

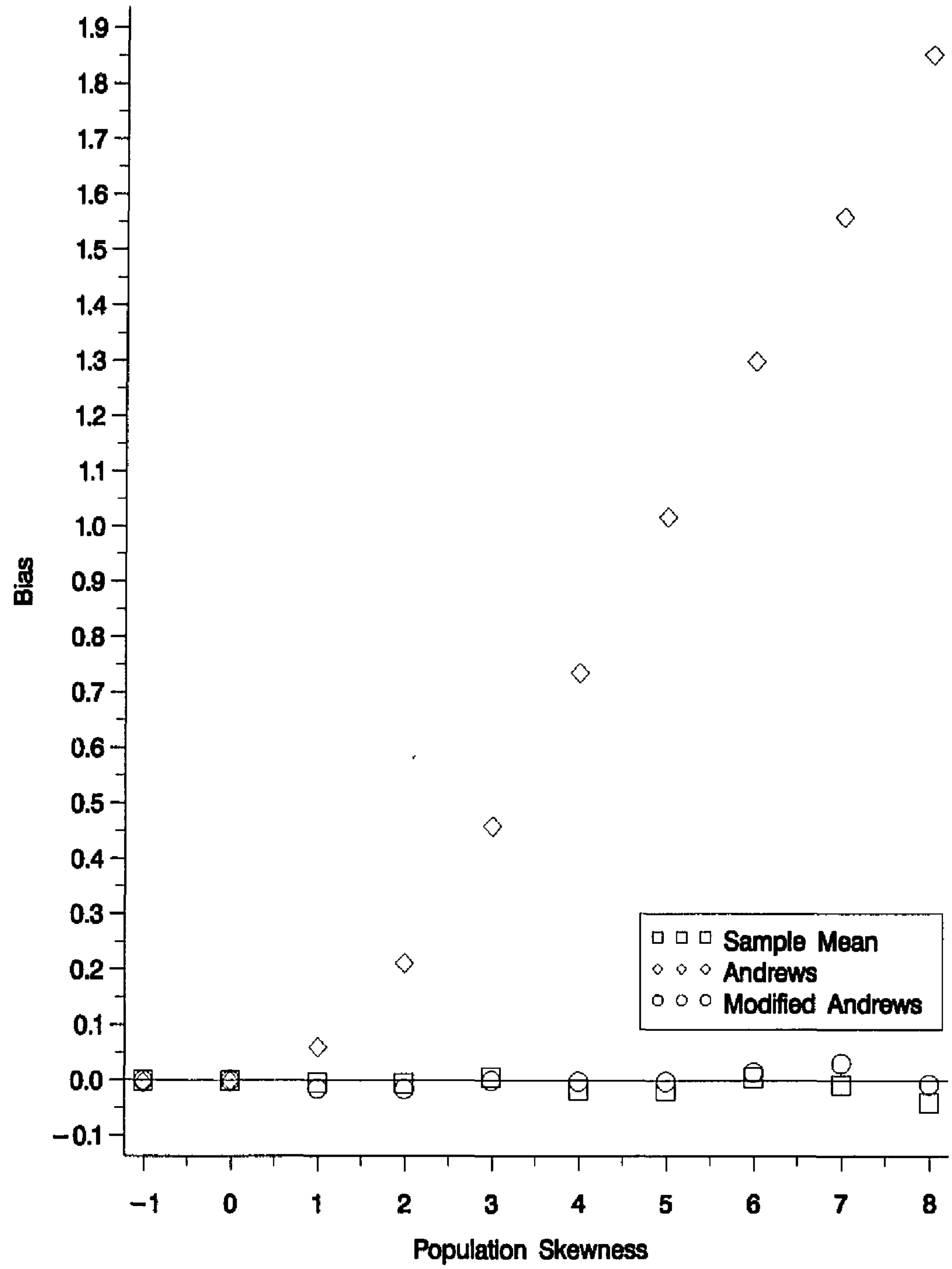


Figure 23. Bias of sample mean, Andrews and modified Andrews vs. population skewness (Weibull distributions with sample size = 50)

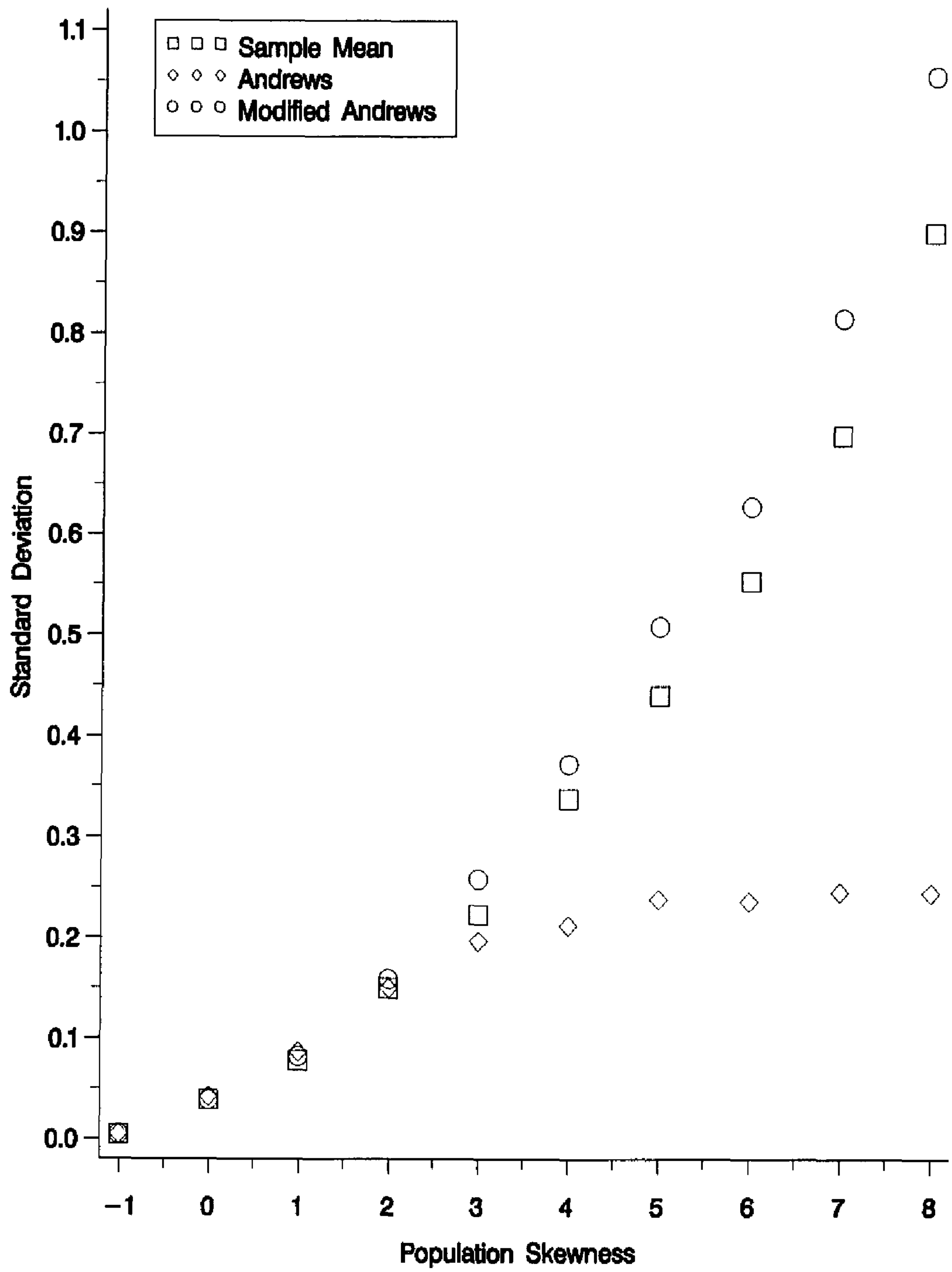


Figure 24. SD of sample mean, Andrews and modified Andrews vs. population skewness (Weibull distributions with sample size = 50)

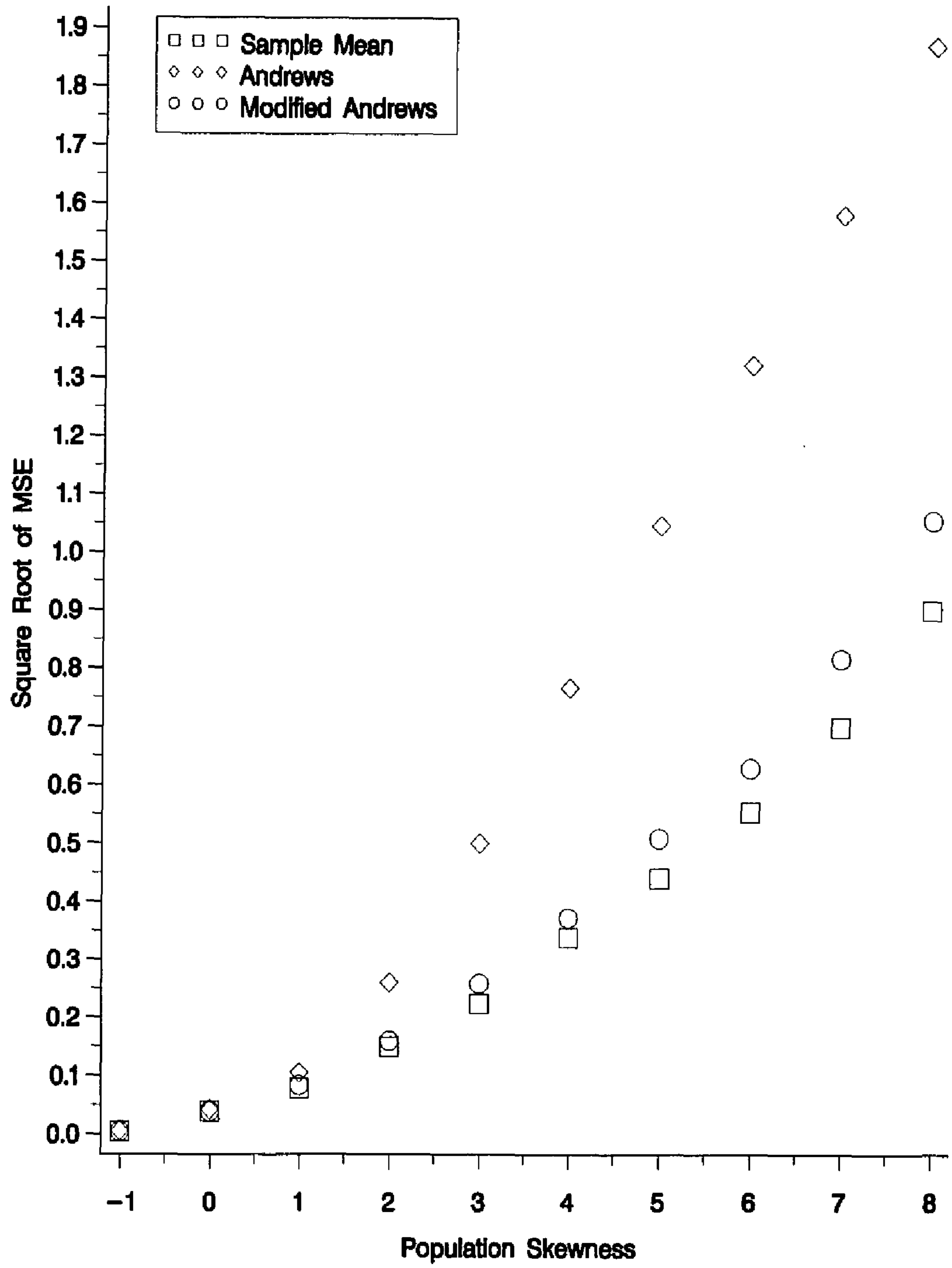


Figure 25. $\sqrt{\text{MSE}}$ of sample mean, Andrews and modified Andrews vs. population skewness (Weibull distributions with sample size = 50)

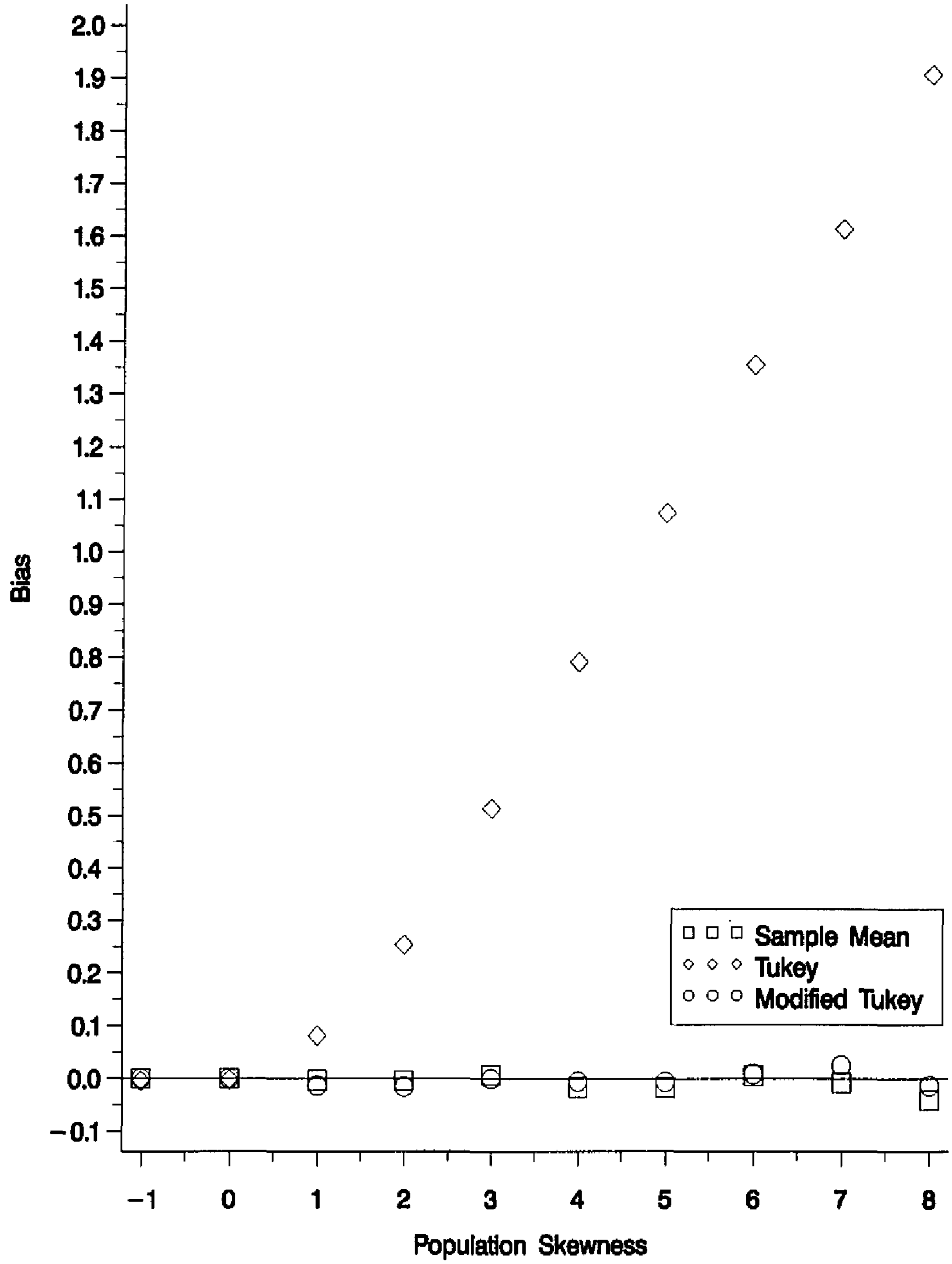


Figure 26. Bias of sample mean, Tukey and modified Tukey vs. population skewness (Weibull distributions with sample size = 50)

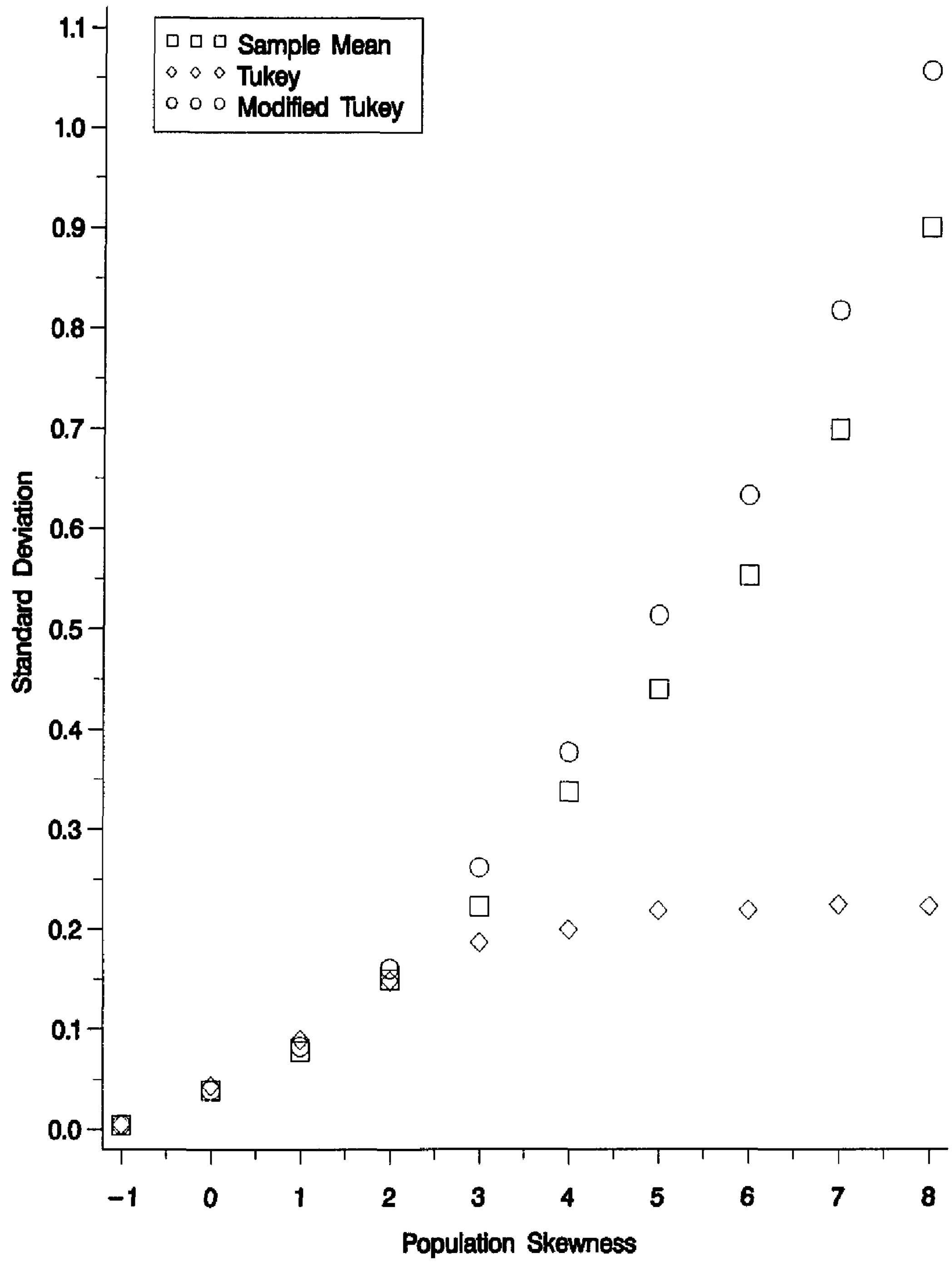


Figure 27. SD of sample mean, Tukey and modified Tukey vs. population skewness (Weibull distributions with sample size = 50)

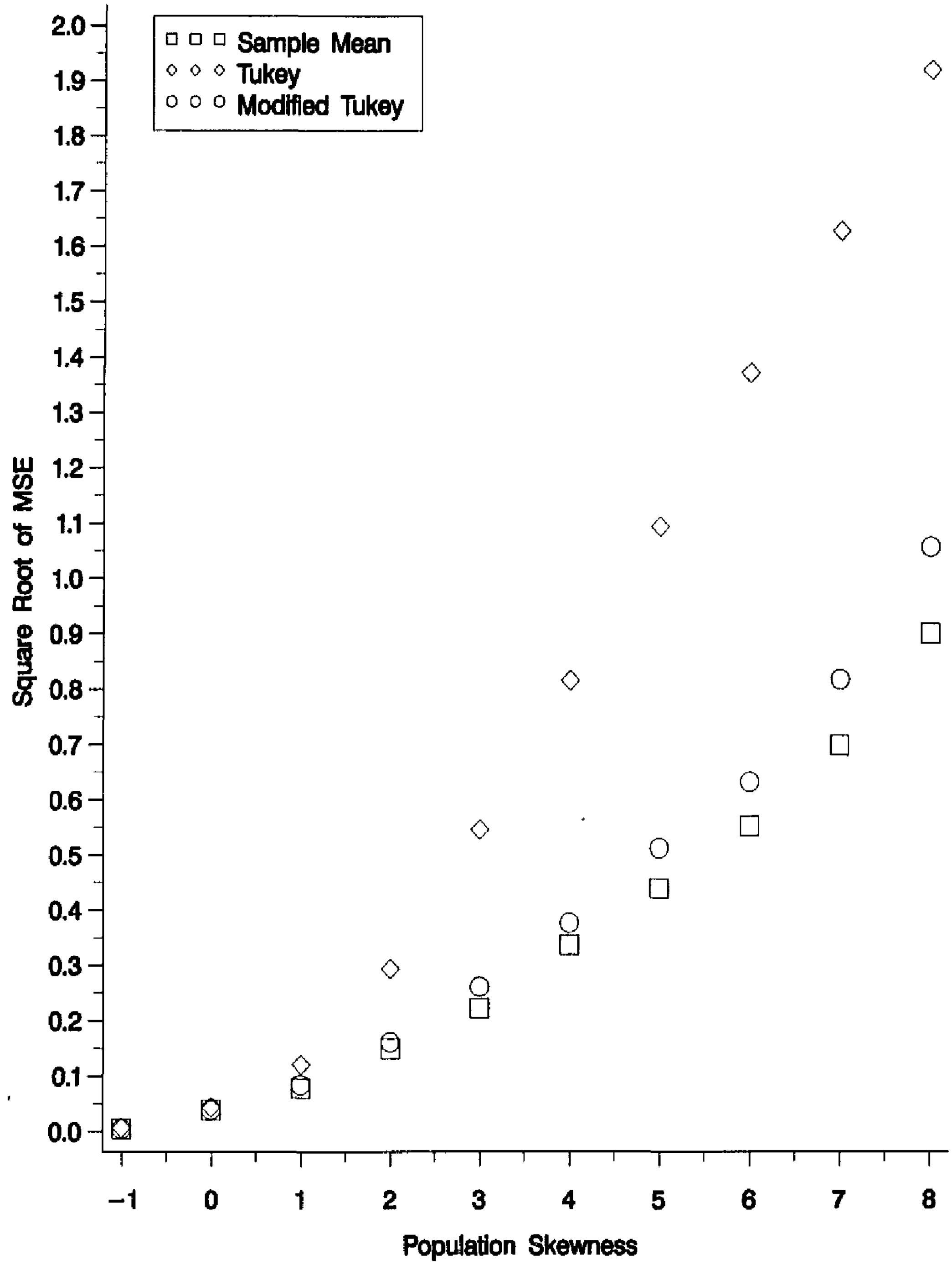


Figure 28. $\sqrt{\text{MSE}}$ of sample mean, Tukey and modified Tukey vs. population skewness (Weibull distributions with sample size = 50)

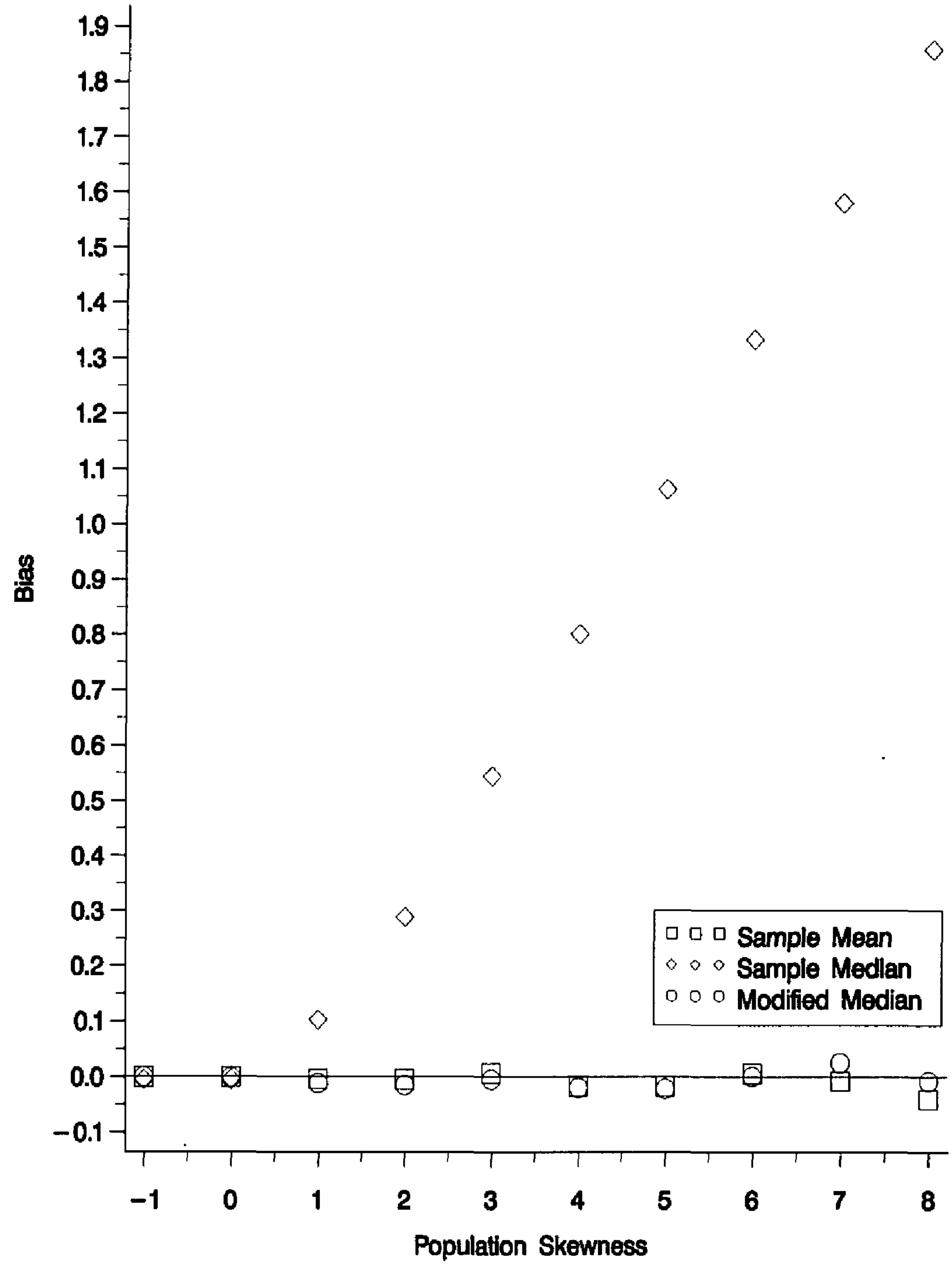


Figure 29. Bias of sample mean, median and modified median vs. population skewness (Weibull distributions with sample size = 50)

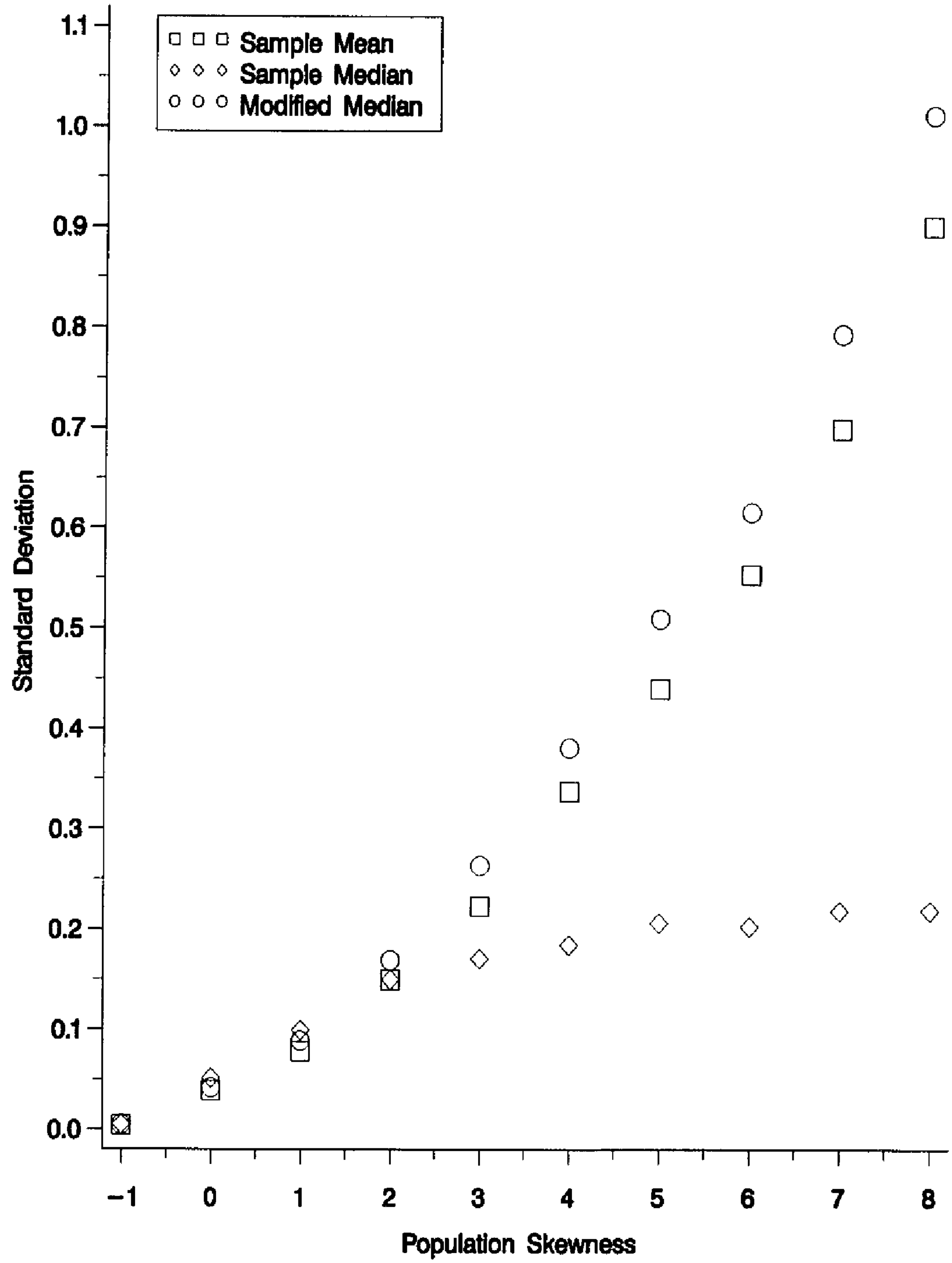


Figure 30. SD of sample mean, median and modified median vs. population skewness (Weibull distributions with sample size = 50)

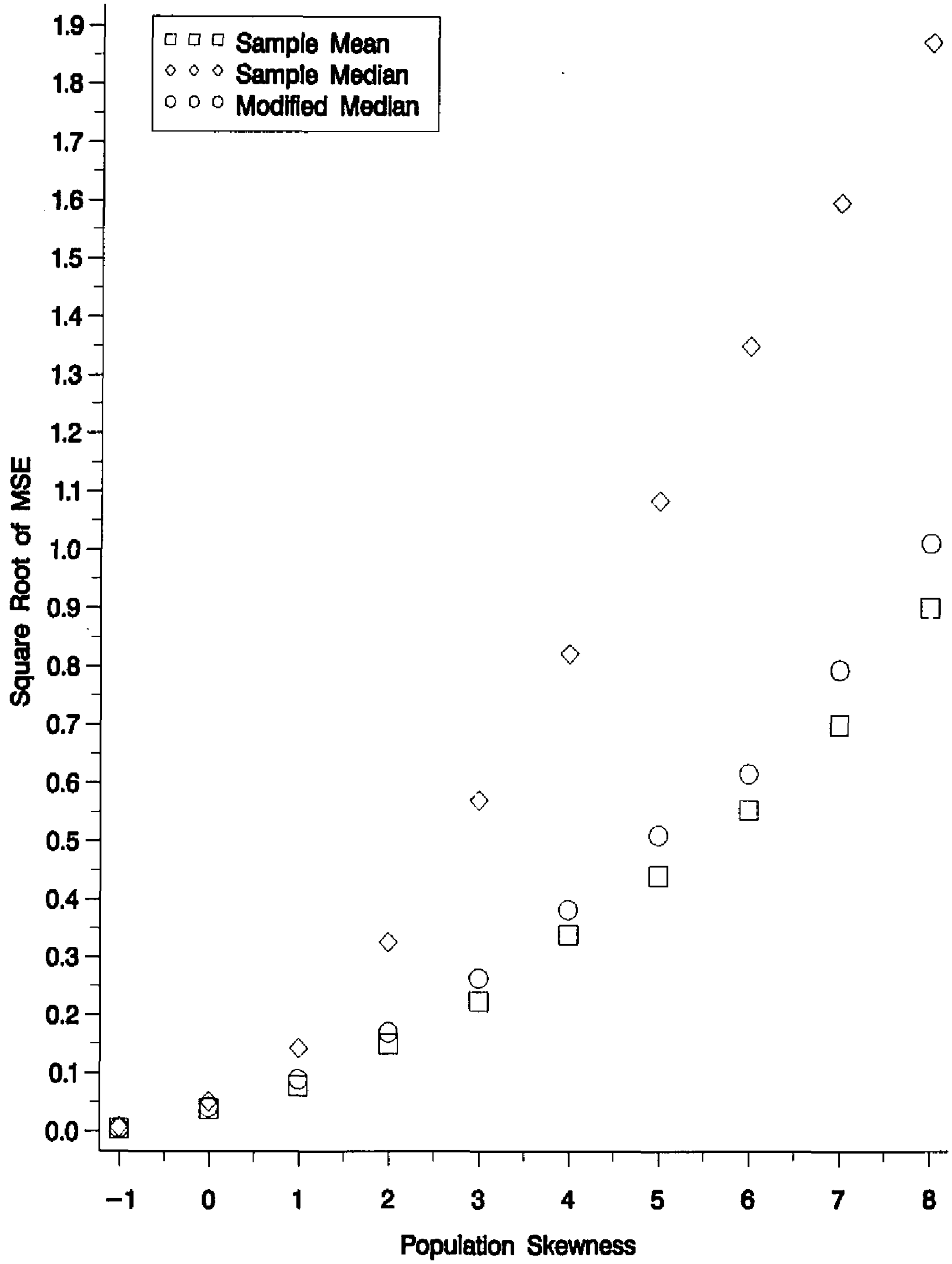


Figure 31. $\sqrt{\text{MSE}}$ of sample mean, median and modified median vs. population skewness (Weibull distributions with sample size = 50)

VITA

Kunjin Shi was born on July 9, 1962 in Anhui Province, People's Republic of China.

In 1981, he graduated from Huangshan Forestry School and worked as a forest technician for four years. He entered Beijing Forestry University in 1985 to pursue a graduate education, and graduated in 1987. After that he worked for four years as a lecturer and researcher at Beijing Forestry University.

Kunjin arrived in the Unite States in 1991 and became a graduate student at Louisiana State University. He earned a M.S. degree in Applied Statistics in May 1993, and is currently a candidate for a Ph.D. degree in Forest Biometrics.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Kunjin Shi

Major Field: Forestry

Title of Dissertation: Applications of Some Robust Statistics in Forestry

Approved:

Quang V Cao
Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Jim L. Chambers

Richard M Pace, III

M. L. Williams

James P. Geaghan

Thomas J. Dean

Date of Examination:

December 6, 1994