ARO-D-8049.9-M

AD 742678

ARO-D PROPOSAL

No. CRDARD-M-8049

"New Developments in Sample Survey Theory"

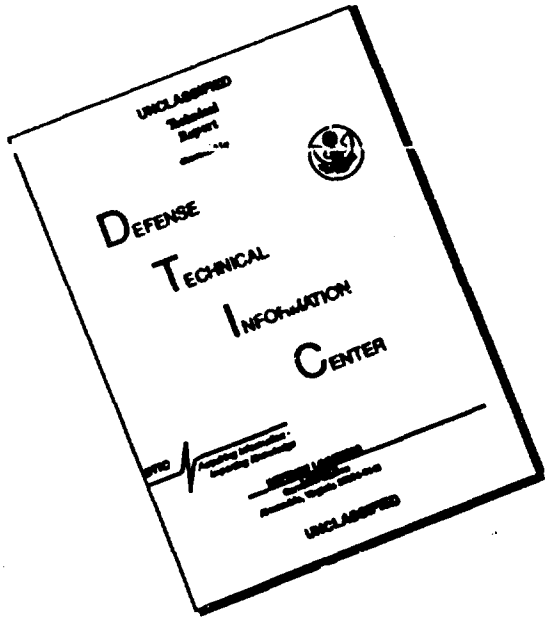
Technical Report No. 10

"SQUARE-ROOT AND CUBE-ROOT ESTIMATORS"

by

O. C. Jenkins, L. J. Ringer, and H. O. Hartley

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

1. INTRODUCTION

1.1 Preliminaries

For the estimation of the mean of a finite population from samples drawn with equal probability and without replacement the sample mean is almost universally recommended. Indeed, it has been shown by Hartley and Rao [1968] and independently in a somewhat different context by Royall [1968] that among the class of estimators described as "scale load estimators" it is the only one which is unbiased uniformly in the parameters of the population. Accordingly, among the class of uniformly unbiased estimators the sample mean is the only admissible competitor and is, therefore, "best" in any competition including that of minimum variance estimators.

Because the arithmetic mean of a random sample is always unbiased and because it has a variance that is a function of only the population variance and the sample size, it is a safe estimator. That is, even when there is no prior knowledge of the population distribution one still can be sure of a predictably "good" estimator. However, there are many occasions when sufficient prior knowledge of the population is available to limit the class of populations to those for which the population mean may be more adequately estimated by some statistic other than the sample mean.

The citations on the following pages will follow the style of Biometrika.

1.2 Objectives

When the situation arises that the population being sampled is known, a priori, to have characteristics that place it in a more restricted category, the question rightly may be asked as to whether the condition of uniform unbiasedness might not be dropped and the bias and variance be combined into a single measure of mean square error. The purpose of this research is to investigate a class of estimators, which shall be called "root estimators", that will usually have smaller mean square error than the arithmetic mean for certain classes of populations.

Root estimators, in general, are of the form

$$y^* = \sum_{j=1}^{m} C_j \left[\frac{1}{n} \sum_{i=1}^{n} (y_i)^{\frac{1}{j}} \right]^j .$$

The particular "root estimators" which will be investigated are:

(1) The square-root estimator, (Section II), $\hat{y} = C_1\bar{y} + C_2\bar{u}^2$ where $u_i = y_i^{1/2}$, which may be useful in the estimation of the means of populations which consist of positive quantities only;

(2) The cube-root estimator, (Section III), $\tilde{y} = (1-C)\bar{y} + C\bar{v}^3$ where $v_i = y_i^{1/3}$, which may be useful in the estimation of the means of populations which consist of both positive and negative quantities.

Each estimator is to be a weighted sum of the mean of the observations and the respective j^{th} power of the mean of the j^{th} roots of the observations. It is the values of the C_i, the weighting constants, that are to be optimized.

1.3 Procedure

For a specific known population of values the distribution of the appropriate roots of these values can be determined mathematically through well known and established procedures. It is then possible to express $\bar{y}$ in terms of the k statistics of this root distribution and through it to determine the bias and error mean square of the estimator in terms of the κ parameters of the root distribution. It is then possible to investigate the properties of the root estimator for various sample sizes.

In many practical sampling applications the sampler does not know the exact form of the population distribution but does know certain facts about it. In particular, he may know that all values are positive and have a large positive skewness. In another case, he may know that most values are zero with only an occasional deviation from zero which may be either positive or negative. In order to make a "root estimator" useful in such applications it must be determined:

(1) If there is a broad class of population distributions for which a particular value of C will substantially

reduce the mean square error;

(2) How much loss of efficiency the estimator will suffer if the population sampled is not in this class.

These two goals will be investigated through mathematical models, deriving the appropriate relationships and then applying them to various standard probability distributions that range across a broad class of population distributions. Graphs of twelve of these distributions are shown in Figure 1. The results will then be tested on real population data for verification of practicality.

2. THE SQUARE-ROOT ESTIMATOR

2.1 Introduction

One of the most common types of populations encountered is made up of values that are all positive. Such distributions are apt to be skewed positively because they are bounded at the lower end but not at the upper end, or for other reasons. It is in such a class of distributions that the square-root estimator will be tested.

The square-root transformation has a greater effect the further a number is from 1. It, therefore, has the effect of reducing the amount of positive skewness while reducing the variance. Negatively skewed distributions, on the other hand, will have the skewness emphasized by the square-root transformation.

The square-root estimator will be defined, in general, by $\tilde{y} = C_1\bar{y} + C_2\bar{u}^2$, or if $C_1 + C_2 = 1$, by $\hat{y} = (1-C)\bar{y} + C\bar{u}^2$. It would be more appropriate to start with a discussion of the general case, but for reasons of clarity the general case will be discussed in Section 2.3.

2.2 The Square-Root Estimator of the Form $\hat{y} = (1-C)\bar{y} + C\bar{u}^2$

2.2.1 Definitions

a. y_i, $i = 1, \ldots, n$; a set of observed values picked with

equal probability and without replacement from a population of all positive quantities.

b. $\bar{y} = \frac{1}{n}\sum_{1}^{n} y_i$; the sample mean.

c. $\bar{Y} = \frac{1}{N}\sum_{1}^{N} y_i$; the population mean.

d. $V(y) = E(y-\bar{Y})^2$; the variance of the y distribution.

e. $u_i = y_i^{1/2}$

f. $\bar{u} = \frac{1}{n}\sum_{1}^{n} u_i$.

g. $\hat{y} = (1-C)\bar{y} + C\bar{u}^2$; the square root estimator.

h. C; the weighting factor which is to be determined.

i. $V(\hat{y}) = E[y - E(\hat{y})]^2$; variance of $\hat{y}$.

j. $B(\hat{y}) = E(\hat{y}-\bar{Y})$; bias of $\hat{y}$ as an estimator of $\bar{Y}$.

k. $EMS(\hat{y}) = E(\hat{y}-\bar{Y})^2 = V(\hat{y}) + [B(\hat{y})]^2$.

ℓ. $R = \frac{EMS(\hat{y})}{V(\bar{y})}$; efficiency ratio of $EMS(\hat{y})$ over $V(\bar{y})$.

m. k statistics [Kendall and Stuart, Vol. 1, p. 280] .

(1) $k_1 = \frac{1}{n}\sum_{1}^{n} u_i = \bar{u}$

$$(2)\quad k_2 = \frac{1}{n-1}\sum_1^n (u_i - \bar{u})^2$$

$$(3)\quad k_3 = \frac{n}{(n-1)(n-2)}\sum_1^n (u_i - \bar{u})^3$$

$$(4)\quad k_4 = \frac{1}{n^{(4)}}\{(n^3+n^2)S_4 - 4(n^2+n)S_3S_1 - 3(n^2-n)S_2^2 + 12nS_2S_1^2 - 6S_1^4\}$$

$$S_j = \sum_{i=1}^n u_i^j \; ; \; n^{(4)} = n(n-1)(n-2)(n-3) \quad .$$

The use of k statistics is desirable because in each case $E(k_i) = \kappa_i$; that is, the value one would attain if n = N. The desirability is further enhanced by the availability of the various relationships which have been worked out by Wishart [1952].

2.2.2 Optimizing the square-root estimator for finite populations

The great advantage of using $\hat{\bar{y}} = (1-C)\bar{y} + C\bar{u}^2$ is that $\bar{y}$ can be expressed in terms of k statistics of the square roots (u_i),

$$\bar{y} = \frac{1}{n}\sum_1^n y_i = \frac{1}{n}\sum_1^n u_i^2 = \frac{1}{n}(\Sigma u_i^2 - n\bar{u}^2) + \bar{u}^2 = (\frac{n-1}{n})k_2 + k_1^2 \quad . \qquad (2.2.1)$$

Hence,

$$\hat{\bar{y}} = (1-C)[(\frac{n-1}{n})k_2 + k_1^2] + Ck_1^2 = (1-C)(\frac{n-1}{n})k_2 + k_1^2 \quad . \qquad (2.2.2)$$

Now, since we are to be estimating $\bar{Y}$, we shall define $B(\hat{y}) = E(\hat{y}) - \bar{Y}$ as the bias and determine

$$E(\hat{y}) = (1-C)\left(\frac{n-1}{n}\right) E(k_2) + E(k_1^2)$$

by utilizing the relationships from Wishart [1952]. They make

$$E(\hat{y}) = (1-C)\left(\frac{n-1}{n}\right)\kappa_2 + E\left(\frac{1}{n}k_2 + k_{11}\right)$$

$$= (1-C)\left(\frac{n-1}{n}\right)\kappa_2 + \frac{1}{n}\kappa_2 + \kappa_{11}$$

$$= \kappa_2 + \kappa_{11} - C\kappa_2\left(1 - \frac{1}{n}\right)$$

$$= \frac{1}{N-1}\sum_{1}^{N}(u_i-\bar{U})^2 + \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i} u_i u_j - C\kappa_2\left(1 - \frac{1}{n}\right)$$

$$= \frac{1}{N-1}\sum_{1}^{N} u_i^2 - \frac{1}{N(N-1)}\left(\sum_{1}^{N} u_i\right)^2 + \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i} u_i u_j - C\kappa_2\left(1-\frac{1}{n}\right)$$

$$= \frac{1}{N-1}\sum_{1}^{N} u_i^2 - \frac{1}{N(N-1)}\left[\sum_{1}^{N} u_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} u_i u_j\right]$$

$$+ \frac{1}{N(N-1)}\sum_{i}\sum_{j\neq i} u_i u_j - C\kappa_2\left(1 - \frac{1}{n}\right)$$

$$= \frac{1}{N}\sum_{1}^{N} u_i^2 - C\kappa_2\left(1 - \frac{1}{n}\right)$$

$$= \bar{Y} - C\kappa_2(1 - \frac{1}{n}) \qquad [\text{i.e., } \bar{Y} = \kappa_2 + \kappa_{11}] \ .$$

Therefore,

$$B(\hat{y}) = E(\hat{y}) - \bar{Y} = -C\kappa_2(1 - \frac{1}{n}) \ . \qquad (2.2.3)$$

Similarly the variances and covariances of the k statistics can be determined.

$$V(k_2) = E[k_2 - E(k_2)]^2 = E(k_2^2) - [E(k_2)]^2$$

$$= E[\frac{1}{n} k_4 + \frac{n+1}{n-1} k_{22}] - \kappa_2^2$$

$$= \frac{1}{n} \kappa_4 + \frac{n+1}{n-1} \kappa_{22} - [\frac{1}{N} \kappa_4 + \frac{N+1}{N-1} \kappa_{22}]$$

$$V(k_2) = [\frac{1}{n} - \frac{1}{N}]\kappa_4 + 2[\frac{1}{n-1} - \frac{1}{N-1}]\kappa_{22} \qquad (2.2.4)$$

$$V(k_1^2) = E(k_1^4) - [E(k_1^2)]^2 = E[\frac{1}{n} k_2 + k_{11}]^2 - [E(\frac{1}{n} k_2 + k_{11})]^2$$

$$= E[\frac{1}{n^2} k_2^2 + k_{11}^2 + \frac{1}{n} k_2 k_{11}] - [\frac{1}{n^2} \kappa_2^2 + \kappa_{11}^2 + \frac{2}{n} \kappa_2 \kappa_{11}]$$

$$= \frac{1}{n^2} [\frac{1}{n} \kappa_4 + \kappa_{22} + \frac{2}{n-1} \kappa_{22}] + [\frac{2}{n(n-1)} \kappa_{22} + \frac{4}{n} \kappa_{211} + \kappa_{1111}]$$

$$+ \frac{2}{n}[\frac{2}{n}\kappa_{31} - \frac{2}{n(n-1)}\kappa_{22} + \kappa_{211}] - \frac{1}{n^2}[\frac{1}{N}\kappa_4 + \kappa_{22} + \frac{2}{N-1}\kappa_{22}]$$

$$- [\frac{2}{N(N-1)}\kappa_{22} + \frac{4}{N}\kappa_{211} + \kappa_{1111}] - \frac{2}{n}[\frac{2}{N}\kappa_{31} - \frac{2}{N(N-1)}\kappa_{22} + \kappa_{211}]$$

$$V(k_1^2) = \frac{1}{n^2}[\frac{1}{n} - \frac{1}{N}]\kappa_4 + \frac{4}{n}(\frac{1}{n} - \frac{1}{N})\kappa_{31} + 4(\frac{1}{n} - \frac{1}{N})\kappa_{211}$$

$$+ 2\{(\frac{n-1}{n})[\frac{1}{n(n-1)} - \frac{1}{N(N-1)}] - \frac{1}{n(N-1)}(\frac{1}{n} - \frac{1}{N})\}\kappa_{22} \quad . \quad (2.2.5)$$

$$Cov(k_1^2, k_2) = E(k_1^2 k_2) - E(k_1^2)\,E(k_2)$$

$$= E[\frac{1}{n^2}k_4 + \frac{2}{n}k_{31} + \frac{1}{n}k_{22} + k_{211}] - [\frac{1}{n}\kappa_2 + \kappa_{11}]\kappa_2$$

$$= \frac{1}{n^2}\kappa_4 + \frac{2}{n}\kappa_{31} + \frac{1}{n}\kappa_{22} + \kappa_{211} - \frac{1}{n}\kappa_2^2 - \kappa_{11}\kappa_2$$

$$= \frac{1}{n^2}\kappa_4 + \frac{2}{n}\kappa_{31} + \frac{1}{n}\kappa_{22} + \kappa_{211} - \frac{1}{n}[\kappa_4 + \kappa_{22} + \frac{2}{N-1}\kappa_{22}]$$

$$- [\frac{2}{N}\kappa_{31} - \frac{2}{N(N-1)}\kappa_{22} + \kappa_{211}]$$

$$Cov(k_1^2, k_2) = [\frac{1}{n} - \frac{1}{N}][\frac{1}{n}\kappa_4 - \frac{2}{N-1}\kappa_{22} + 2\kappa_3\kappa_1] \quad . \quad (2.2.6)$$

Now $V(\bar{y})$ and $V(\hat{y})$ can be derived in terms of the kappa parameters of the distribution of u.

$$V(\bar{y}) = V[(\tfrac{n-1}{n})k_2 + k_1^2] = (\tfrac{n-1}{n})^2 V(k_2) + V(k_1^2) + 2(\tfrac{n-1}{n})\,\mathrm{Cov}(k_2, k_1^2)$$

$$= (\tfrac{n-1}{n})^2 \left[(\tfrac{1}{n} - \tfrac{1}{N})\kappa_4 + 2(\tfrac{1}{n-1} - \tfrac{1}{N-1})\kappa_{22}\right]$$

$$+ \tfrac{1}{n^2}(\tfrac{1}{n} - \tfrac{1}{N})\kappa_4 + \tfrac{4}{n}(\tfrac{1}{n} - \tfrac{1}{N})\kappa_{31} + 4(\tfrac{1}{n} - \tfrac{1}{N})\kappa_{211}$$

$$+ 2\left\{(\tfrac{n-1}{n})(\tfrac{1}{n(n-1)} - \tfrac{1}{N(N-1)}) - \tfrac{1}{n(N-1)}(\tfrac{1}{n} - \tfrac{1}{N})\right\}\kappa_{22}$$

$$+ 2(\tfrac{n-1}{n})\left\{\tfrac{1}{n}(\tfrac{1}{n} - \tfrac{1}{N})\kappa_4 - (\tfrac{2}{N-1})(\tfrac{1}{n} - \tfrac{1}{N})\kappa_{22} + 2(\tfrac{1}{n} - \tfrac{1}{N})\kappa_{31}\right\} .$$

$$V(\bar{y}) = (\tfrac{1}{n} - \tfrac{1}{N})\left[\kappa_4 + 4\kappa_{31} + 4\kappa_{211}\right] + 2\left\{(\tfrac{n-1}{n})^2(\tfrac{1}{n-1} - \tfrac{1}{N-1})\right.$$

$$\left. + (\tfrac{n-1}{n})(\tfrac{1}{n(n-1)} - \tfrac{1}{N(N-1)}) - (\tfrac{2n-1}{n(N-1)})(\tfrac{1}{n} - \tfrac{1}{N})\right\}\kappa_{22} . \quad (2.2.7)$$

$$V(\hat{y}) = (1-C)^2(\tfrac{n-1}{n})^2 V(k_2) + V(k_1^2) + 2(1-C)(\tfrac{n-1}{n})\,\mathrm{Cov}(k_2, k_1^2)$$

$$= V(\bar{y}) + C(C-2)(\tfrac{n-1}{n})^2 V(k_2) - 2C(\tfrac{n-1}{n})\,\mathrm{Cov}(k_2, k_1^2) \quad (2.2.8)$$

$$EMS(\hat{\bar{y}}) = V(\hat{\bar{y}}) + [B(\hat{\bar{y}})]^2 = C(1 - \tfrac{1}{n})\{C(1 - \tfrac{1}{n})\kappa_2^2 + (\tfrac{1}{n} - \tfrac{1}{N})[C(1-\tfrac{1}{n}) - 2]\kappa_4$$

$$+ [2(C-2)(1 - \tfrac{1}{n})(\tfrac{1}{n-1} - \tfrac{1}{N-1}) + \tfrac{4}{N-1}(\tfrac{1}{n} - \tfrac{1}{N})\kappa_{22}$$

$$- 4(\tfrac{1}{n} - \tfrac{1}{N})\kappa_{31}\} + V(\bar{\bar{y}}) \qquad (2.2.9)$$

It can be seen in (2.2.9) that for $EMS(\hat{\bar{y}})$ to be less than $V(\bar{\bar{y}})$ the bracketed term must be negative. It is also evident that the bracketed term is a quadratic in C. This makes minimizing quite simple by the usual process of equating the first derivative with respect to C to zero.

$$\frac{\partial EMS(\bar{y})}{\partial C} = 2(C-1)(\tfrac{n-1}{n})^2 V(k_2) - 2(\tfrac{n-1}{n})\ Cov(k_2, k_1^2) + 2C(\tfrac{n-1}{n})^2\ \kappa_2^2$$

which, when equated to zero and solved for C, yields

$$C_0 = \frac{(\frac{1}{n-1})(\frac{N-n}{N})\ [\kappa_4 + 2\kappa_{22} + 2\kappa_{31}]}{\kappa_2^2 + (\frac{1}{n} - \frac{1}{N})\kappa_4 + 2(\frac{1}{n-1} - \frac{1}{N-1})\kappa_{22}} \quad ; \quad n > 1 \qquad (2.2.10)$$

2.2.3 The square-root estimator for infinite populations

Analysis of these relationships is facilitated by examining the limiting equations as $N \to \infty$. Indeed, if N is moderately large there is little loss of accuracy by doing so.

$$\lim_{N\to\infty} V(\bar{y}) = \frac{1}{n}\{\kappa_4 + 4\kappa_3\kappa_1 + 4\kappa_2\kappa_1^2 + 2\kappa_2^2\} \qquad (2.2.11)$$

$$\lim_{N\to\infty} EMS(\hat{y}) = \frac{C}{n}\left(1 - \frac{1}{n}\right)\{[C(1 - \tfrac{1}{n}) - 2]\kappa_4$$

$$+ [C(n+1) - 4]\kappa_2^2 - 4\kappa_3\kappa_1\} + V(\bar{y}) \qquad (2.2.12)$$

$$\lim_{N\to\infty} C_0 = \frac{(\frac{1}{n-1})\ [\kappa_4 + 2\kappa_2^2 + 2\kappa_3\kappa_1]}{\frac{1}{n}\kappa_4 + \frac{n+1}{n-1}\kappa_2^2} \quad . \qquad (2.2.13)$$

The disappearing of such terms as κ_{31} is due to the fact that $\lim_{N\to\infty} \kappa_{31} = \kappa_3\kappa_1$, etc.

Examination is further facilitated by the substitution of the equivalent central moments of u

$$\kappa_1 = \bar{U}$$

$$\kappa_2 = \mu_2 = E(u-\bar{U})^2$$

$$\kappa_3 = \mu_3 = E(u-\bar{U})^3$$

$$\kappa_4 = \mu_4 - 3\mu_2^2; \quad \mu_4 = E(u-\bar{U})^4 \quad .$$

The optimum weight becomes, thereby,

$$C_0 = \frac{(\frac{1}{n-1})[\mu_4 - 3\mu_2^2 + 2\mu_2^2 + 2\mu_3\mu_1]}{\frac{1}{n}[\mu_4 - 3\mu_2^2] + \frac{n+1}{n-1}\mu_2^2}$$

$$= \frac{n[\mu_4 - \mu_2^2 + 2\bar{U}\mu_3]}{(n-1)\mu_4 + (n^2-2n+3)\mu_2^2} \quad ; \quad n > 1 \qquad (2.2.14)$$

Noting at this point that C_0 is positive if $\mu_4 + 2\bar{U}\mu_3 > \mu_2^2$, consider the inequality

$$0 < E[(x-\mu)^2 - E(x-\mu)^2]^2 = E(x-\mu)^4 - [E(x-\mu)^2]^2 = \mu_4 - \mu_2^2$$

which shows $\mu_4 > \mu_2^2$. It is evident, then, that $\mu_4 + 2\bar{U}\mu_2 > \mu_2^2$ if $\mu_3 > 0$; that is, if the distribution of the square-root transformed distribution has a positive third moment.

Now, converting $EMS(\hat{y})$ to moments about $\bar{U}$;

$$EMS(\hat{y}) = \frac{C}{n}(1-\frac{1}{n})\{[C(1-\frac{1}{n}) - 2][\mu_4-3\mu_2^2] + [C(n+1) - 4]\mu_2^2 - 4\bar{U}\mu_3\} + V(\bar{y})$$

$$= \frac{C}{n}(1-\frac{1}{n})\{[C(\frac{n-1}{n}) - 2]\mu_4 + [C(n+\frac{1}{n}-2) + 2]\mu_2^2 - 4\bar{U}\mu_3\} + V(\bar{y}) \; . \qquad (2.2.15)$$

It has already been shown that $C > 0$ when $\mu_3 > 0$. Inspection of (2.2.15) further indicates that a large μ_3 causes a smaller $EMS(\hat{y})$; more evidence that a population which is highly skewed to the right

is best benefitted by the square-root estimator.

2.2.4 The bias of the square-root estimator

The bias of $\hat{\bar{y}}$ at C_0 is

$$B(\hat{\bar{y}}) = -C_0(1 - \frac{1}{n})\kappa_2 = \frac{-[\kappa_4 + 2\kappa_2^2 + \kappa_3\kappa_1]}{\frac{\kappa_4}{\kappa_2} + \frac{n(n+1)}{n-1}\kappa_2}$$

$$= \frac{-\mu_2\left[\frac{\mu_4}{\mu_2^2} + 2\frac{\bar{U}\mu_3}{\mu_2^2} - 1\right]}{\frac{\mu_4}{\mu_2^2} - 3 + n(\frac{n+1}{n-1})}$$

which decreases with increasing n, but at a very slow rate when n is small. For example, when $n = 2$, $\frac{n(n+1)}{n-1} = 6$. In order to double this value it is necessary to make $n = 10$. Unless $\frac{\mu_4}{\mu_2^2}$ is small, even doubling $\frac{n(n+1)}{n-1}$ does not halve the bias.

2.2.5 Types of distributions for which $EMS(\hat{\bar{y}})$ can be made substantially less than $V(\bar{y})$

The investigations of the types of distribution functions for which the error-mean-square can be substantially reduced will be facilitated by the following two theorems.

Theorem 1. The value of C_0 is invariant to the scale parameter (a multiplicative constant).

Proof: Let y_i be distributed as f(y), and

$$u_i = y_i^{\frac{1}{2}} \quad .$$

Let

$$y_i^* = by_i$$

so that

$$u_i^* = \sqrt{y_i^*} = \sqrt{by_i} = bu_i \quad .$$

Then

$$\bar{y}^* = \frac{1}{n} \Sigma y_i^* = \frac{b}{n} \Sigma y_i = b\bar{y}$$

and similarly

$$\bar{Y}^* = b\bar{Y}$$

$$\bar{u}^* = \frac{1}{n} \Sigma u_i^* = \frac{1}{n} \Sigma \sqrt{b}\, u_i = \sqrt{b}\, \bar{u}$$

$$\hat{y}^* = (1-C)\bar{y}^* + C\bar{u}^{*2} = b(1-C)\bar{y} + bC\bar{u}^2 = b\hat{y}$$

$$B(\hat{y}^*) = E[\hat{y}^* - \bar{Y}^*] = bE[\hat{y} - \bar{Y}] = bB(\hat{y}) \quad .$$

Then

$$V(\hat{y}^*) = V(b\hat{y}) = b^2V(\hat{y})$$

and

$$EMS(\hat{y}^*) = b^2V(\hat{y}) + b^2[B(\hat{y})]^2 = b^2[EMS(\hat{y})] \quad .$$

Therefore, the value of $EMS(\hat{y}^*)$ will be minimized by minizing $EMS(\hat{y})$ which is accomplished by C_0.

Theorem 2. The efficiency of $EMS(\hat{y})$, (R), is invariant to the scale parameter.

Proof: Let y_i be distributed as f(y) and $V(\bar{y})$ be the variance of $\bar{y}$ for a sample of size n.

Then

$$R = \frac{EMS(\hat{y})}{V(\bar{y})} = \frac{V(\hat{y})}{V(\bar{y})} + \frac{[B(\hat{y})]^2}{V(\bar{y})} \quad .$$

Now, letting $y_i^* = by_i$ as in Theorem 1,

$$V(\bar{y}^*) = V(b\bar{y}) = b^2V(\bar{y})$$

$$V(\hat{y}^*) = V(b\hat{y}) = b^2V(\hat{y})$$

$$B(\hat{y}^*) = bB(\hat{y}) \quad .$$

Then

$$R^* = \frac{EMS(\hat{y}^*)}{V(\bar{y}^*)} = \frac{V(\hat{y}^*)}{V(\bar{y}^*)} + \frac{[B(\hat{y}^*)]^2}{V(\bar{y}^*)} = \frac{b^2V(\hat{y})}{b^2V(\bar{y})} + \frac{b^2[B(\hat{y})]^2}{b^2V(\bar{y})} = R \quad .$$

Theorem 3. The values of C_0 and R are not invariant to the position parameter. That is, a constant added to every element of the population will change the value of C_0 and R.

Proof: Let $y \sim f(y)$ such that $E(y) = \bar{Y}$ and $V(y) = \sigma^2$. Then, for a simple random sample of size n, $E(\bar{y}) = \bar{Y}$ and $V(\bar{y}) = \sigma^2/n$. Letting $y_i^* = y_i + b$, then $E(y^*) = \bar{Y} + b$ and $V(y^*) = \sigma^2$. Again, for a simple random sample of size n $E(\bar{y}^*) = \bar{Y} + b$ and $V(\bar{y}^*) = \sigma^2/n$, that is, there is no change in the variance of the unbiased estimator.

But, letting $u_i^* = \sqrt{y_i^*}$ and $\hat{y}^* = (1-C)\bar{y}^* + C\bar{u}^{*2}$, we see that

$$E(\hat{y}^*) = (1-C)[\bar{Y} + b] + CE(k_1^{*2})$$

$$= (1-C)(\bar{Y}+b) + C[\tfrac{1}{n}\kappa_2^* + \kappa_1^{*2}]$$

$$= [\bar{Y} + b] - C[\bar{Y} + b - \tfrac{1}{n}\kappa_2^* - \kappa_1^{*2}]$$

$$= [\bar{Y} + b] - C[\kappa_2 + \kappa_1^2 + b - \tfrac{1}{n}\kappa_2^* - \kappa_1^{*2}]$$

$$= [\bar{Y} + b] - C[(\kappa_2 - \frac{1}{n}\kappa_2^*) + (\kappa_1^2 - \kappa_1^{*2}) + b]$$

$$= [\bar{Y} + b] - C(\frac{n-1}{n})\kappa_2 - C[\frac{1}{n}(\kappa_2 - \kappa_2^*) + (\kappa_1^2 - \kappa_1^{*2}) + b] \quad .$$

Therefore, the bias of $\hat{y}^*$ as an estimate of $\bar{Y} + b$ is

$$B(\hat{y}^*) = B(\hat{y}) - C[\frac{1}{n}(\kappa_2 - \kappa_2^*) + (\kappa_1^2 - \kappa_1^{*2}) + b] \quad .$$

Now, since

$$\hat{y}^* = (1-C)\bar{y}^* + C\bar{u}^{*2}$$

$$= (1-C)(\bar{y}+b) + C\bar{u}^{*2}$$

$$= (1-C)\bar{y} + C\bar{u}^{*2} + (1-C)b$$

$$= (1-C)[(\frac{n-1}{n})k_2 + k_1^2] + Ck_1^{*2} + (1-C)b \quad ,$$

we have

$$V(\hat{y}^*) = (1-C)^2(\frac{n-1}{n})^2 V(k_2) + (1-C)^2 V(k_1^2) + C^2 V(k_1^{*2})$$

$$+ 2(1-C)^2(\frac{n-1}{n})\,\mathrm{Cov}(k_2, k_1^2) + 2C(1-C)(\frac{n-1}{n})\,\mathrm{Cov}(k_2, k_1^{*2})$$

$$+ 2C(1-C)\,\mathrm{Cov}(k_1^2, k_1^{*2})$$

$$= (1-C)^2\left(\frac{n-1}{n}\right)^2 V(k_2) + V(k_1^2) + C(C-2)\ V(k_1^2)$$

$$+ 2(1-C)\left(\frac{n-1}{n}\right)\ \mathrm{Cov}(k_2,\ k_1^2) - 2C(1-C)\left(\frac{n-1}{n}\right)\ \mathrm{Cov}(k_2,\ k_1^2)$$

$$+ 2C(1-C)\left(\frac{n-1}{n}\right)\ \mathrm{Cov}(k_2,\ k_1^{*2}) + 2C(1-C)\ \mathrm{Cov}(k_1^2,\ k_1^{*2})$$

$$= V(\hat{y}) + C(C-2)\ V(k_1^2) + C^2 V(k_1^{*2}) + 2C(1-C)\ \mathrm{Cov}(k_1^2,\ k_1^{*2})$$

$$- 2C(1-C)\left(\frac{n-1}{n}\right)\ \{\mathrm{Cov}(k_2,\ k_1^2) - \mathrm{Cov}(k_2,\ k_1^{*2})\} \quad .$$

And then

$$\mathrm{EMS}(\hat{y}^*) = V(\hat{y}^*) + [B(\hat{y}^*)]^2$$

$$= V(\hat{y}) + C(C-2)V(k_1^2) + C^2 V(k_1^{*2}) + 2C(1-C)\ \mathrm{Cov}(k_1^2,\ k_1^{*2})$$

$$- 2C(1-C)\left(\frac{n-1}{n}\right)\ \{\mathrm{Cov}(k_2,\ k_1^2) - \mathrm{Cov}(k_2,\ k_1^{*2})\}$$

$$+ [B(\hat{y})]^2 + C^2\left[\frac{1}{n}(\kappa_2 - \kappa_2^*) + (\kappa_1^2 - \kappa_1^{*2}) + b\right]^2$$

$$- 2B(\hat{y})\ C\left[\frac{1}{n}(\kappa_2 - \kappa_2^*) + (\kappa_1^2 - \kappa_1^{*2}) + b\right] \quad .$$

So,

$$EMS(\hat{y}^*) = EMS(\hat{y}) + C(C-2)V(k_1^2) + C^2V(k_1^{*2}) + 2C(1-C)\ Cov(k_1^2, k_1^{*2})$$

$$- 2C(1-C)\left(\frac{n-1}{n}\right)\{Cov(k_2, k_1^2) - Cov(k_2, k_1^{*2})\}$$

$$+ C^2\left[\frac{1}{n}(\kappa_2 - \kappa_2^*) + (\kappa_1^2 - \kappa_1^{*2}) + b\right]^2$$

$$- 2B(\hat{y})C\left[\frac{1}{n}(\kappa_2 - \kappa_2^*) + (\kappa_1^2 - \kappa_1^{*2}) + b\right] \quad .$$

It can be seen from this equation that the $EMS(\hat{y}^*)$ is not equal to $EMS(\hat{y})$, and that the value of C which will minimize it will be a function of b. These facts coupled with the fact that $V(\bar{y}^*) = V(\bar{y})$ are sufficient to show that

$$R^* = \frac{EMS(\hat{y}^*)}{V(\bar{y}^*)} = \frac{EMS(\hat{y}^*)}{V(\bar{y})} \neq R \quad .$$

An example later in this section will further illustrate this point.

Theorems 1 and 2 will allow investigations of such distributions as

$$f(y) = \frac{1}{\alpha!\ \beta^{\alpha+1}}\ y^{\alpha}\ e^{-y/\beta}$$

by letting $\beta = 1$, and then apply the results with equal effect to the same distribution with any other value of β. On the other hand, distributions that differ by an additive constant will not have the same optimum value of C nor will the square-root estimator have equal effficiencies on these distributions.

In order to investigate the efficiency and utility of the square-root estimator, three specific families of distributions were examined. These families were chosen because they represent a wide spectrum of population forms.

(1) The gamma distributions; $f(y) = \frac{1}{\alpha!} y^{\alpha} e^{-y}$, $0 \leq y < \infty$.

Due to Theorems 1 and 2 any results applicable to this distributions will also be applicable to

$$f(y) = \frac{1}{\alpha!\ \beta^{\alpha+1}} y^{\alpha} e^{-y/\beta}$$

for any value of β.

Making the transformation $u = y^{1/2}$ yields

$$f(u) = \frac{2}{\alpha!} u^{2\alpha+1} e^{-u^3}$$

from which the first four central moments can be calculated for various values of α. We shall consider the distributions generated by $\alpha = 0, 1, 2,$ and 3. These values are convenient because

$$f(y) = e^{-y}$$

is fairly skewed, while

$$f(y) = \frac{1}{6} y^3 e^{-y}$$

is rather symmetrical with only a slight positive skewness.

Figures 2(a), 2(b), 2(c) and 2(d) show the relative efficiencies (R) of a number of different distributions for n = 2, 3, 5 and 10, respectively. The top four of these efficiency curves (number 1, 2, 3, 4) are of the gamma distribution with α = 3, 2, 1 and 0, respectively. The accuracy to which these graphs can be read is sufficient for practical purposes. Exact values of R and C_0 for the various distributions are given in Table 1.

It can be seen that the more symmetrical the parent distribution the less gain attainable. However, it should also be noticed that for values of C between 0 and 2.5 there is, in every case, some improvement over $V(\bar{y})$. This is an important fact as it indicates that the square-root estimator will give an improvement in mean square error for any value of C between 0 and 2.5 as long as the parent distribution is at least as skewed as

$$f(y) = \frac{1}{6} y^3 e^{-y} \quad .$$

To illustrate Theorem 3, consider the distribution

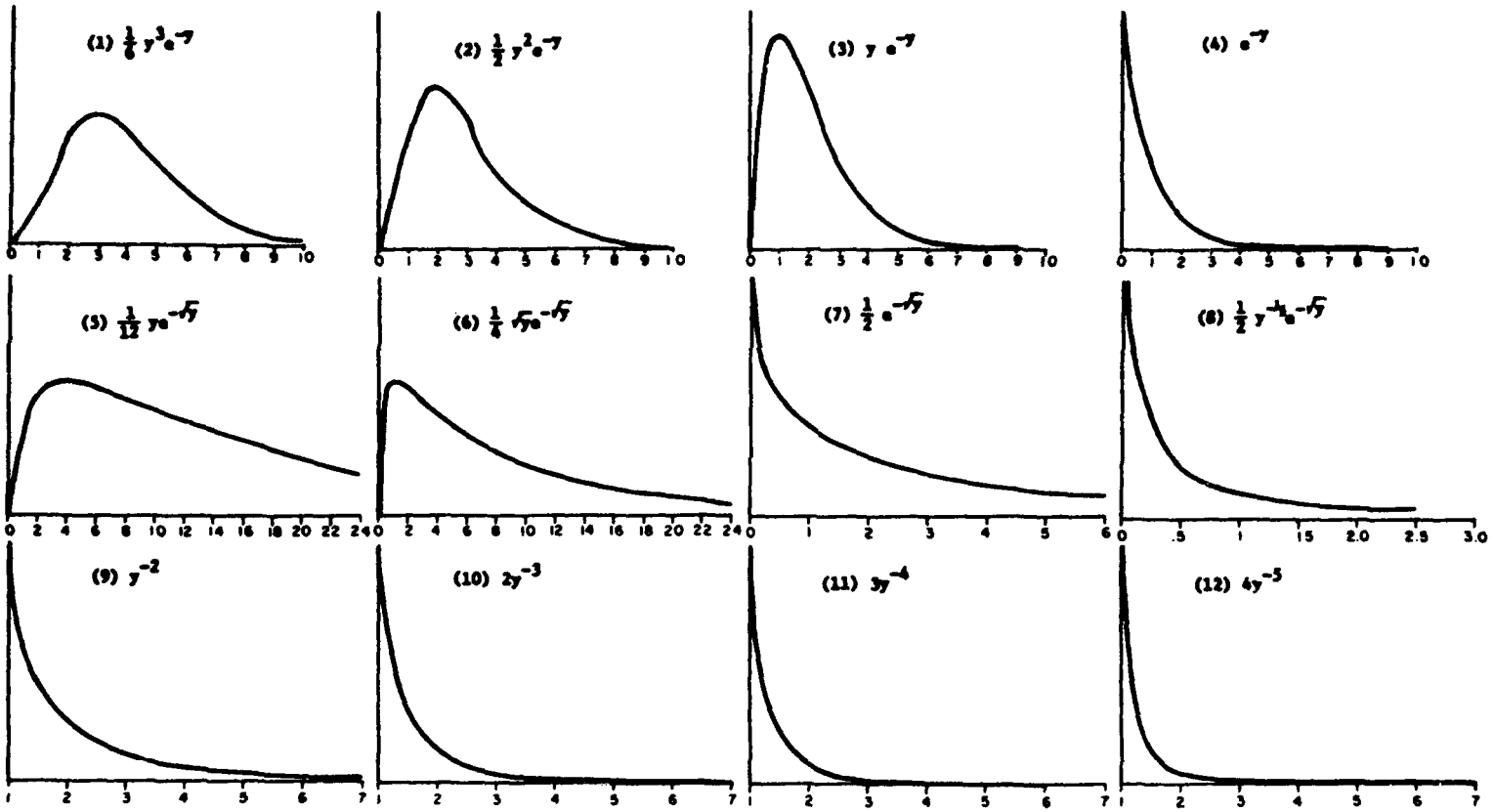


Figure 1. Graphs of twelve distributions used as standards.

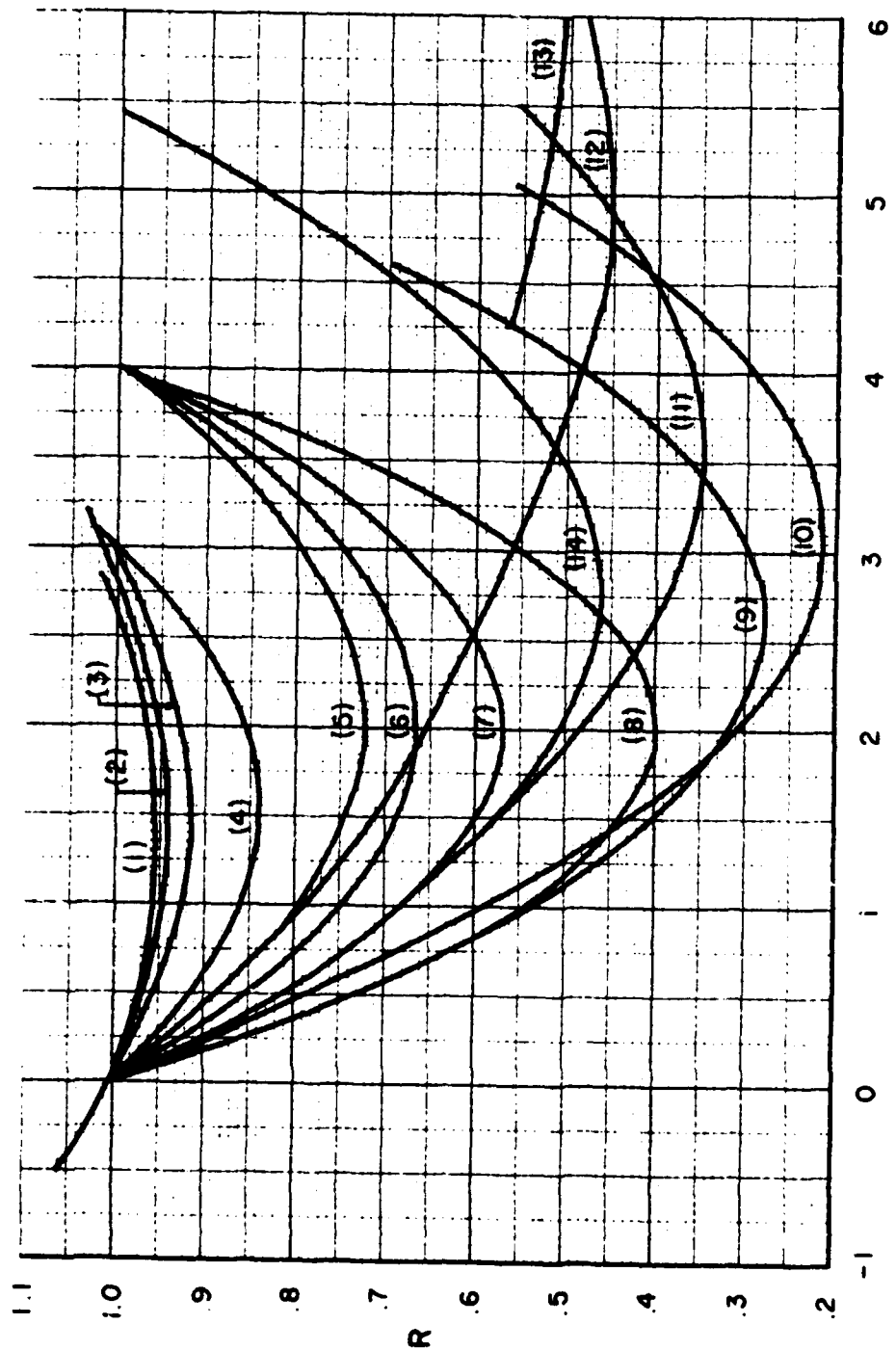


Figure 2(a). Relative efficiency (R) as a function of the weighting constant (C) for the square-root estimator, n = 2

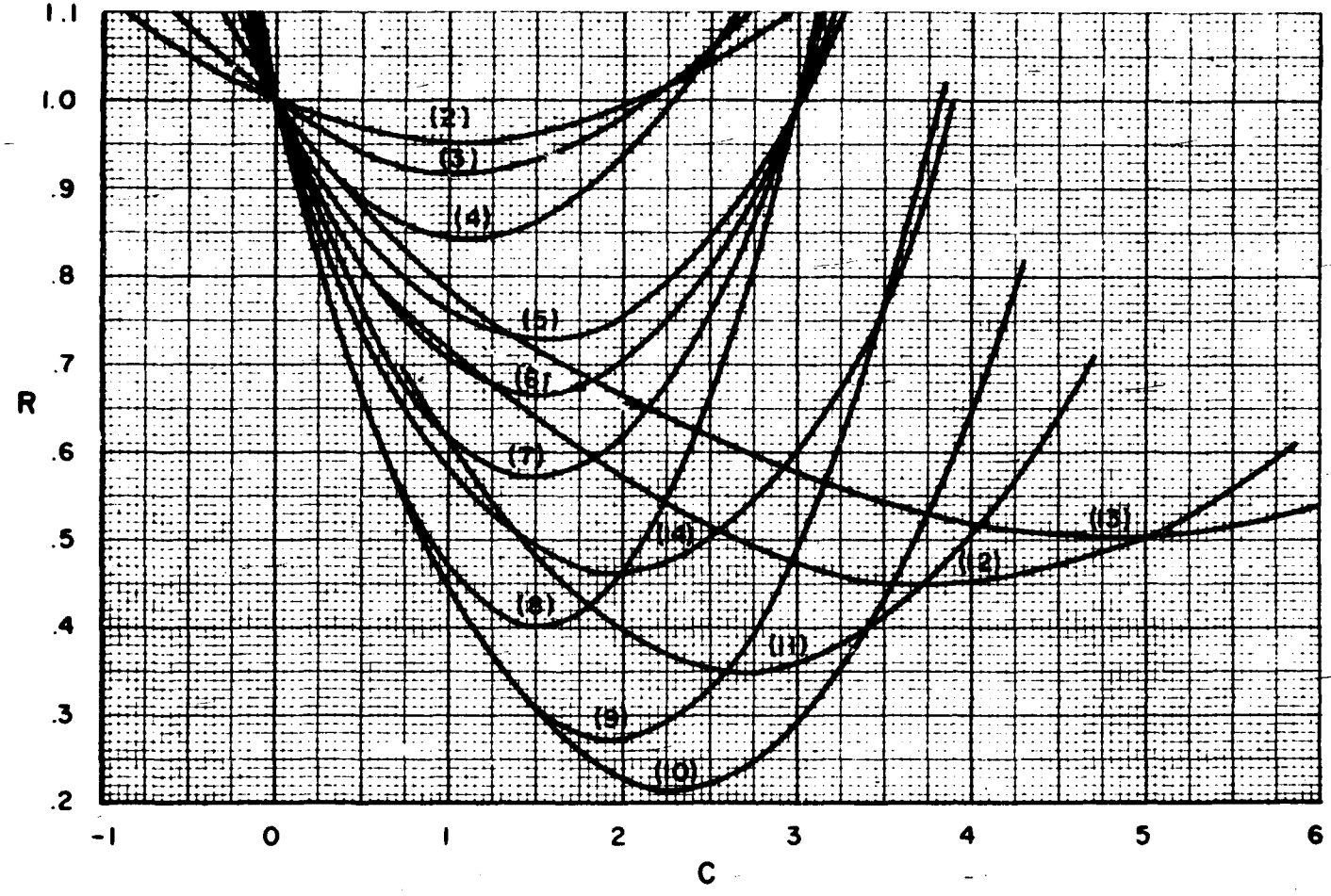


Figure 2(b). Relative efficiency (R) as a function of the weighting constant (C) for the square-root estimator, n = 3

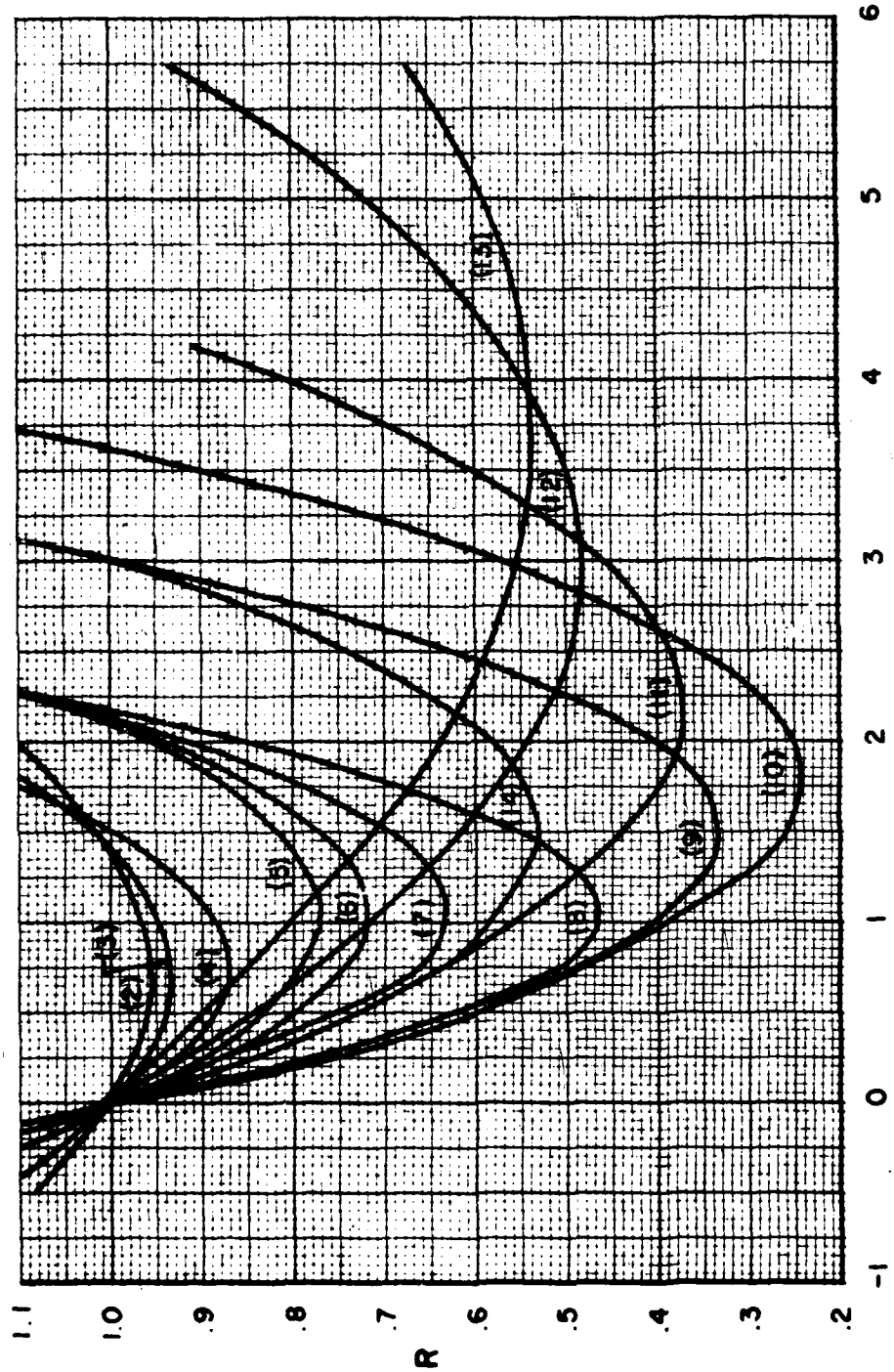


Figure 2(c). Relative efficiency (R) as a function of the weighting constant (C) for the square-root estimator, n = 5

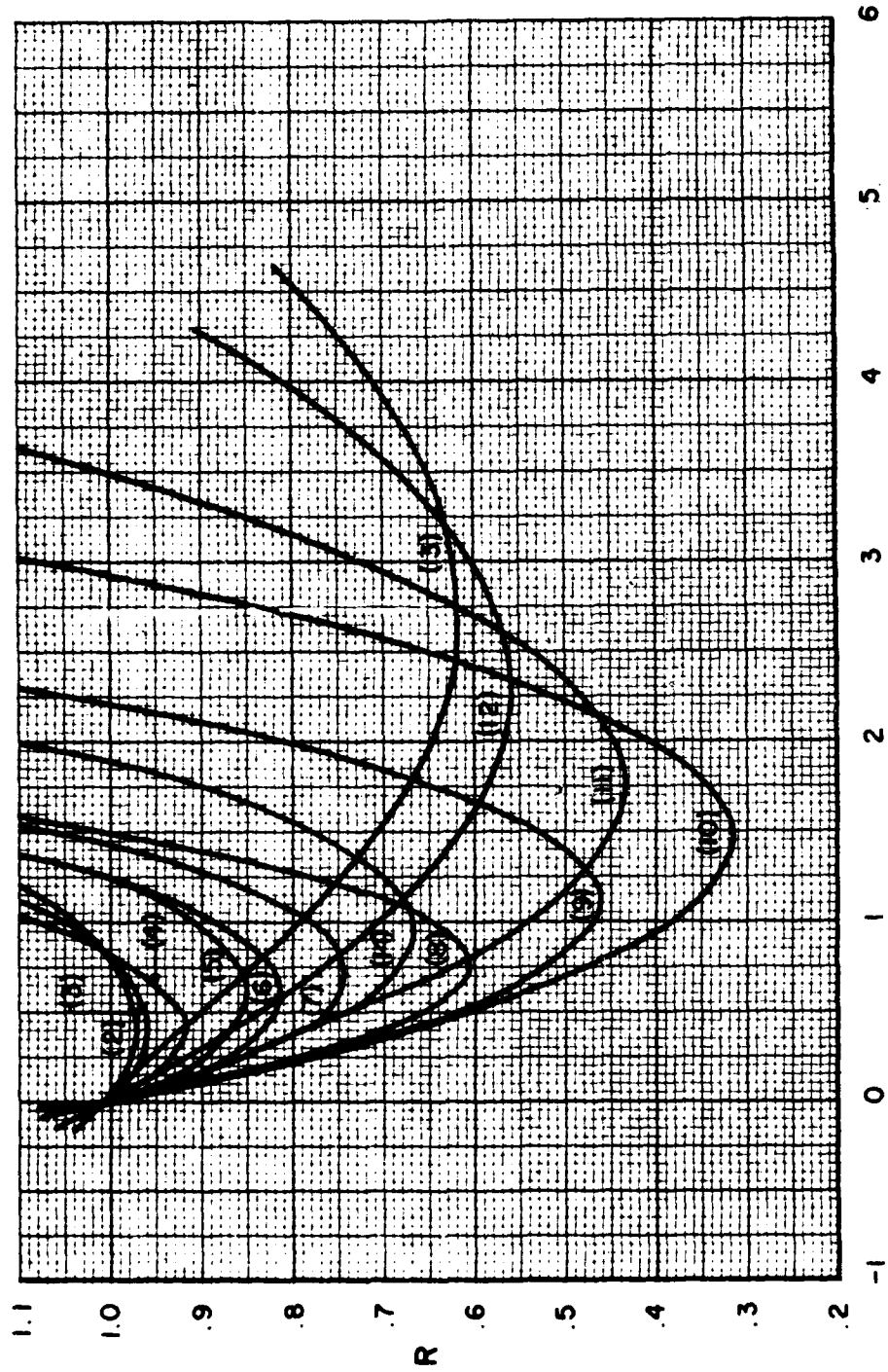


Figure 2(d). Relative efficiency (R) as a function of the weighting constant (C) for the square-root estimator, n = 10

TABLE 1. VALUES OF C FOR WHICH $EMS(\hat{\bar{y}}) < V(\bar{y})$ FOR THIRTEEN DISTRIBUTIONS; n = 2, 20

Distribution	n= 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
(1) $\frac{1}{6} y^3 e^{-y}$	**.96	.96	.96	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.99	.99	.99	.99	.99	.99
	*2.74	2.06	1.66	1.38	1.18	1.04	.92	.82	.76	.70	.64	.60	.56	.52	.49	.46	.43	.41	.39
(2) $\frac{1}{2} y^2 e^{-y}$	.94	.94	.95	.96	.96	.96	.97	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.98	.98
	2.77	2.08	1.67	1.39	1.19	1.04	.93	.83	.76	.70	.64	.60	.56	.52	.50	.46	.44	.42	.40
(3) $y e^{-y}$	.92	.92	.93	.93	.94	.95	.95	.96	.96	.96	.96	.97	.97	.97	.97	.97	.98	.99	.99
	2.83	2.13	1.70	1.42	1.22	1.07	.95	.85	.78	.71	.65	.61	.57	.53	.50	.47	.45	.43	.41
(4) e^{-y}	.84	.84	.86	.87	.88	.90	.91	.91	.92	.93	.93	.94	.94	.94	.95	.95	.95	.95	.96
	2.98	2.23	1.79	1.50	1.29	1.13	1.01	.91	.83	.76	.70	.65	.61	.57	.54	.51	.48	.46	.43
(5) $\frac{1}{12} ye^{-\sqrt{y}}$	.73	.73	.75	.77	.79	.81	.83	.84	.85	.86	.87	.88	.88	.89	.90	.90	.91	.91	.91
	4.00	3.00	2.45	2.08	1.81	1.61	1.45	1.32	1.21	1.12	1.04	.97	.91	.86	.81	.77	.73	.70	.67
(6) $\frac{1}{4} \sqrt{y} e^{-\sqrt{y}}$	.67	.67	.69	.72	.74	.76	.78	.80	.81	.82	.84	.84	.85	.86	.87	.87	.88	.89	.89
	4.00	3.00	2.46	2.10	1.85	1.65	1.49	1.35	1.25	1.16	1.08	1.01	.95	.90	.86	.82	.77	.74	.71
(7) $\frac{1}{2} e^{-\sqrt{y}}$	.57	.57	.60	.63	.66	.68	.71	.73	.75	.76	.78	.79	.80	.81	.82	.83	.83	.84	.85
	4.00	3.00	2.48	2.41	1.89	1.70	1.55	1.42	1.31	1.22	1.14	1.07	1.01	.95	.91	.86	.82	.79	.75
(8) $\frac{1}{2} y^{-\frac{1}{2}} e^{-\sqrt{y}}$	.40	.40	.43	.47	.50	.53	.56	.58	.60	.62	.64	.66	.68	.69	.70	.71	.72	.73	.74
	4.00	3.00	2.53	2.22	2.00	1.82	1.68	1.57	1.46	1.37	1.29	1.23	1.17	1.11	1.06	1.01	.97	.93	.89
(9) y^{-2} $(y \geq 1)$	.28	.28	.33	.36	.39	.42	.44	.46	.48	.50	.52	.54	.54	.55	.57	.58	.59	.60	.62
	5.26	3.85	3.38	3.03	2.79	2.59	2.43	2.29	2.17	2.07	1.97	1.89	1.81	1.74	1.67	1.62	1.56	1.51	1.46
(10) $2y^{-3}(y \geq 1)$	.22	.22	.23	.24	.26	.27	.29	.30	.32	.33	.34	.36	.37	.38	.39	.40	.41	.42	.43
	6.08	4.56	3.99	3.67	3.45	3.29	3.15	3.04	2.95	2.85	2.77	2.71	2.64	2.57	2.52	2.47	2.41	2.36	2.31
(11) $3y^{-4}(y \geq 1)$	.35	.35	.36	.37	.39	.40	.41	.42	.44	.45	.46	.47	.48	.49	.50	.51	.52	.52	.53
	7.20	5.40	4.72	4.34	4.08	3.89	3.73	3.59	3.48	3.37	3.28	3.19	3.12	3.04	2.97	2.91	2.85	2.79	2.73
(12) $4y^{-5}(y \geq 1)$	.45	.45	.47	.48	.50	.52	.53	.55	.56	.57	.59	.60	.61	.62	.63	.64	.65	.66	.66
	10.09	7.57	6.56	5.96	5.54	5.21	4.94	4.76	4.51	4.33	4.17	4.03	3.89	3.77	3.65	3.55	3.45	3.35	3.27
(13) $5y^{-6}(y \geq 1)$	.51	.51	.52	.54	.56	.57	.59	.61	.62	.63	.65	.66	.67	.68	.69	.70	.71	.71	.72
	12.89	9.66	8.32	7.51	6.93	6.48	3.11	5.79	5.51	5.27	5.05	4.84	4.66	4.49	4.34	4.20	4.07	3.94	3.82

*Maximum values of C for which $EMS(\hat{\bar{y}}) < V(\bar{y})$.
Optimum value is ½ the listed value.

**Value of $R = \frac{EMS(\hat{\bar{y}})}{V(\bar{y})}$ at C_0.
Minimum value of C for which $EMS(y) < V(\bar{y})$ is zero in each case.

$$f(x) = e^{-(x-1)} \quad , \quad x \geq 1 \quad .$$

This is exactly the same distribution as

$$f(y) = e^{-y}$$

except that 1 has been added to each population value. The values shown below illustrate the differences caused by this shift.

$f(y) = e^{-y}$; $y \geq 0$	$f(x) = e^{-(x-1)}$; $x \geq 1$	
$\bar{U} = .8862$	$\bar{U} = 1.3792$	
$\mu_2 = .2146$	$\mu_2 = 2$	
$\mu_3 = .0627$	$\mu_3 = 3.0688$	
$\mu_4 = .1495$	$\mu_4 = .0413$	
$C_0 = 1.49$	$C_0 = 4.11$	$n = 2$
$R = .84$	$R = .70$	$n = 2$

One needs not fear dire consequences because of such differences, however. If, in each case, a value of $C = 2$ had been used, the efficiencies attained would have been .86 and .78, respectively.

(2) The Wishart distributions; $f(y) = \frac{1}{2\alpha!}\, y^{\left(\frac{\alpha-1}{2}\right)}\, e^{-\sqrt{y}}$.

This group was chosen in order to investigate distributions with a bit more skewness than the gamma. In fact, the square root transformation of these distributions generate gamma distributions,

$$g(u) = \frac{1}{\alpha!}\, u^{\alpha}\, e^{-u} \quad ,$$

for which the moments are readily calculated.

The effectiveness of the square-root estimator for these distributions (numbers 5, 6 and 7) for n = 2, 3, 5 and 10 are also shown in Figures 2(a), 2(b), 2(c) and 2(d). Again it can be seen that the more skewed parent distributions offer greater gains through the square-root estimator. Equally important is the fact that any value of C from 0 to 4 will produce a gain in efficiency for these distributions with $C_0 \doteq 2$ being the optimum value.

(3) The Pareto distributions; $f(y) = \frac{\alpha\beta^{\alpha}}{y^{\alpha+1}}$; $0 < \alpha,\ \alpha < y$,

for which β is a scale parameter.

The Pareto distributions are reputed to be approximate for income distributions and similar cases. These were included to show what happens to the square-root estimator in such extraordinary cases. The square-root transformation, $u = y^{1/2}$, produces

$$g(u) = 2\alpha u^{-(2\alpha+1)} \quad , \qquad u \geq 1 \quad .$$

Moments higher than $2\alpha-1$ do not converge so the distribution was truncated to force convergence.

In the cases of $f(y) = y^{-2}$ and $f(y) = 2y^{-3}$ the square-root estimator makes even greater gains of efficiency over $\bar{y}$. It is, however, at an increasing value of C with the maximum gain for $f(y) = 2y^{-3}$ at $C \doteq 3$. If such a value of C were being utilized and the distribution was, in reality, a gamma with $\alpha = 3$ (number 1), there would be a loss of information of approximately 4%.

2.2.6 General Comments

Inspection of Figures 2(c) and 2(d) readily illustrate that for larger sample sample sizes some gains are realized, but two important facts should be noted. As the size increases the efficiency of the square-root estimator over $\bar{y}$ becomes less and the value of C_0 approaches zero for all distributions, indicating that the primary uses of the square-root estimator are cases where small sample sizes are necessary.

The following three properties are quite important to the usefulness of the square-root estimator.

(1) $EMS(\hat{\bar{y}})$ is quadratic in C of the form

$$h(C) = aC^2 + bC + d;\ a > 0 \ .$$

(2) $EMS(\hat{\bar{y}}) = V(\bar{y})$ when $C = 0$.

(3) $EMS(\hat{\bar{y}}) < V(\bar{y})$ only when $C > 0$ for positive populations that yield a $\mu_3 > 0$.

The implications of these properties are:

(1) For positive populations for which $\mu_3 > 0$ only values of C greater than 0 and less than $2C_0$ will produce a gain in efficiency and any value of C in this range will produce a gain.

(2) If it is known that the population being sampled is at least as skew as one of the standard distributions a value of C can be established which will guarantee that the square-root estimator will be more efficient than $\bar{y}$.

2.2.7 A simulation to verify the efficiency of the square-root estimator

The efficiency curves identified by (14) in Figures 2(a), 2(b), 2(c) and 2(d) are for a set of data from Cochran [1953]. These data are, actually, a sample of 200 sizes of cities in the United States in 1920 and are reproduced as Table 2(a). The cities sizes are grouped into categories of an interval width of 100,000 and the mid-points of the categories were used as representation of the entire category. To facilitate calculation the sizes have been coded by dividing by 50,000.

For this demonstration the 200 cities are taken to be a population of 200 which has a mean value of 2.66 (2.66 × 50,000) and a variance of 12.564. Letting u_i be the square root of the i^{th} category size the following central moments of the square-root distribution were calculated:

$$\bar{U} = 1.437$$

$$\mu_2 = .595$$

$$\mu_3 = .953$$

$$\mu_4 = 2.523$$

It was through the use of these values substituted into equation (2.2.15) and varying the value of C that the efficiency curve was generated.

With such a population as this it is not difficult to determine every possible sample of size n = 2 and to calculate $\hat{y}$. For instance, if C = 2 the equation for the square-root estimator is

$$y = (1-2)\bar{y} + 2\bar{u}^2$$

$$= -\frac{1}{2}(y_1 + y_2) + 2\left[\frac{\sqrt{y}_1 + \sqrt{y}_2}{2}\right]^2$$

$$= \sqrt{y_1 \cdot y_2} \quad .$$

The use of C = 2 was chosen because, by referring to Figure 2(a), it can be seen that it is a very safe value to use when n = 2. The frequency distribution of the various values of $\hat{y}$ are shown in Table 2(b).

Calculation of

$$EMS(\hat{y}) = \frac{1}{n} \Sigma f_i (\hat{y}_i - 2.66)^2$$

shows that $EMS(\hat{y}) = 3.3$. That makes the efficiency factor

$$\frac{EMS(\hat{y})}{V(\bar{y})} = \frac{3.3}{(12.564/2)} = .525$$

which agrees with the theoretical value within rounding error.

The expectation of $\hat{y}$ calculated from Table 2(b) is

$$E(\hat{y}) = 2.073$$

which makes the bias

$$B(\hat{y}) = 2.073 - 2.660 = -.593 \quad .$$

According to equation (2.3.3)

$$B(\hat{y}) = -C(1 - \frac{1}{n})\kappa_2 = -2(\frac{1}{2})(.595) = -.595$$

TABLE 2(a). DISTRIBUTION OF SIZES

OF 200 U.S. CITIES IN 1920

Population Size X50,000	u	f
1	1	133
3	1.732	36
5	2.236	11
7	2.645	5
9	3.000	4
11	3.317	4
13	3.606	0
15	3.873	4
17	4.123	0
19	4.358	1
21	4.583	2

$\bar{Y} = 2.66$
$\sigma^2 = 12.564$
} Original population parameters

$\bar{U} = 1.437$
$\mu_2 = .595$
$\mu_3 = .953$
$\mu_4 = 2.523$
} Square-root transformed population parameters

[Source: Cochran (1953) p.39].

TABLE 2(b). DISTRIBUTION OF SQUARE-ROOT ESTIMATOR FOR ALL SAMPLES OF SIZE n = 2 FROM SIZES OF 200 U.S. CITIES IN 1920

Sample	$\hat{y}$	f	Sample	$\hat{y}$	f	Sample	$\hat{y}$	f
1,1	1.000	8,778	5,7	5.916	55	9,21	13.748	8
1,3	1.732	4,778	5,9	6.708	44	11,11	11.000	6
1,5	2.236	1,463	5,11	7.416	44	11,13	11.958	0
1,7	2.646	665	5,13	8.062	0	11,15	12.845	16
1,9	3.000	532	5,15	8.660	44	11,17	13.675	0
1,11	3.317	532	5,17	9.220	0	11,19	14.475	4
1,13	3.606	0	5,19	9.747	11	11,21	15.199	8
1,15	3.873	532	5,21	10.247	22	13,13	13.000	0
1,17	4.123	0	7,7	7.000	10	13,15	13.964	0
1,19	4.359	133	7,9	7.937	20	13,17	14.866	0
1,21	4.583	266	7,11	8.775	20	13,19	15.716	0
3,3	3.000	630	7,13	9.539	0	13,21	16.523	0
3,5	3.873	396	7,15	10.247	20	15,15	15.000	6
3,7	4.583	180	7,17	10.909	0	15,17	15.969	0
3,9	5.196	144	7,19	11.533	5	15,19	16.882	4
3,11	5.745	144	7,21	12.124	10	15,21	17.748	8
3,13	6.245	0	9,9	9.000	6	17,17	17.000	0
3,15	6.708	144	9,11	9.950	16	17,19	17,972	0
3,17	7.141	0	9,13	10.817	0	17,21	18.894	0
3,19	7.550	36	9,15	11.619	16	19,19	19.000	0
3,21	7.937	72	9,17	12.369	0	19,21	19.975	2
5,5	5.000	55	9,19	13.077	4	21,21	21.000	1

$$E(\hat{y}) = \frac{41051.357}{19,800} = 2.073; \quad B(\hat{y}) = 2.073 - 2.660 = -.593$$

$$MSE(\hat{y}) = 3.3$$

which again shows agreement within rounding error.

2.3 Square-Root Estimators of the Form

$$\tilde{y} = C_1\bar{y} + C_2\bar{u}^{-2},\ C_1 + C_2 \neq 1$$

If the restriction that the weights sum to one be relaxed it is, of course, possible to attain even greater gains of efficiency. Using the definitions in Section 2.2;

$$\tilde{y} = C_1\bar{y} + C_2\bar{u}^{-2} = C_1[(\tfrac{n-1}{n})\ k_2 + k_1^2] + C_2k_1^2$$

$$= C_1(\tfrac{n-1}{n})\ k_2 + (C_1 + C_2)\ k_1^2 \qquad (2.3.1)$$

$$E(\tilde{y}) = C_1(\tfrac{n-1}{n})\ \kappa_2 + (C_1 + C_2)\ E(k_1^2)$$

$$= C_1(\tfrac{n-1}{n})\ \kappa_2 + (C_1 + C_2)[\tfrac{1}{n}\ \kappa_2 + \kappa_{11}]$$

$$= C_1\kappa_2 + \tfrac{1}{n}\ C_2\kappa_2 + (C_1 + C_2)\ \kappa_{11}$$

$$= C_1(\kappa_2 + \kappa_{11}) + \tfrac{1}{n}\ C_2\kappa_2 + C_1\kappa_{11}$$

$$= C_1\bar{Y} + (1-C)(\kappa_2 + \kappa_{11}) + \tfrac{1}{n}\ C_2\kappa_2 + C_1\kappa_{11}$$

$$= \bar{Y} + (1-C)(\kappa_2 + \kappa_{11}) + \tfrac{1}{n}\ C_2\kappa_2 + C_1\kappa_{11}$$

$$= \bar{Y} + (C_1 + \frac{C_2}{n} - 1)\kappa_2 + (C_1 + C_2 - 1)\kappa_{11}$$

$$B(\tilde{y}) = -[(1 - C_1 - \frac{C_2}{n})\kappa_2 + (1 - C_1 - C_2)\kappa_{11}] \qquad (2.3.2)$$

$$V(\tilde{y}) = C_1^2(\frac{n-1}{n})^2 V(k_2) + (C_1 + C_2)^2 V(k_1^2)$$

$$+ 2C_1(C_1 + C_2)(\frac{n-1}{n}) \text{Cov}(k_2, k_1^2) \qquad (2.3.3)$$

$$\text{EMS}(\tilde{y}) = V(\tilde{y}) + [B(\tilde{y})]^2$$

$$\frac{\partial \text{EMS}(\tilde{y})}{\partial C_1} = 2C_1(\frac{n-1}{n})^2 V(k_2) + 2(C_1 + C_2) V(k_1^2)$$

$$+ 2(2C_1 + C_2)(\frac{n-1}{n}) \text{Cov}(k_2, k_1^2)$$

$$- 2[(1 - C_1 - \frac{C_2}{n})\kappa_2 + (1 - C_1 - C_2)\kappa_{11}][\kappa_2 + \kappa_{11}] \quad .$$

Equating to zero and isolating C_1 and C_2:

$$C_1[(\frac{n-1}{n})^2 V(k_2) + V(k_1^2) + 2(\frac{n-1}{n}) \text{Cov}(k_2, k_1^2) + (\kappa_2 + \kappa_{11})^2]$$

$$+ C_2[V(k_1^2) + (\frac{n-1}{n}) \text{Cov}(k_2, k_1^2)$$

$$+ (\frac{\kappa_2}{n} + \kappa_{11})(\kappa_2 + \kappa_{11})] = (\kappa_2 + \kappa_{11})^2 \quad .$$

$$\frac{\partial EMS(\tilde{y})}{\partial C_2} = 2(C_1 + C_2)\, V(k_1^2) + 2C_1(\tfrac{n-1}{n})\, Cov(k_2, k_1^2)$$

$$- 2[(1 - C_1 - \tfrac{C_2}{n})\kappa_2 + (1 - C_1 - C_2)\kappa_{11}][\tfrac{\kappa_2}{n} + \kappa_{11}] \quad .$$

Again, equating to zero and isolating C_1 and C_2

$$C_1[V(k_1^2) + (\tfrac{n-1}{n})\, Cov(k_2, k_1^2) + (\kappa_2 + \kappa_{11})(\tfrac{\kappa_2}{n} + \kappa_{11})]$$

$$+ C_2[V(k_1^2) + (\tfrac{\kappa_2}{n} + \kappa_{11})^2] = (\kappa_2 + \kappa_{11})(\tfrac{\kappa_2}{n} + \kappa_{11}) \quad .$$

Letting

$$E_1 = (\kappa_2 + \kappa_{11})^2$$

$$E_2 = (\kappa_2 + \kappa_{11})(\tfrac{\kappa_2}{n} + \kappa_{11})$$

$$A = (\tfrac{n-1}{n})^2\, V(k_2) + V(k_1^2) + 2(\tfrac{n-1}{n})\, Cov(k_2, k_1^2) + E_1$$

$$B = V(k_1^2) + (\tfrac{n-1}{n})\, Cov(k_2, k_1^2) + E_2$$

$$D = V(k_1^2) + (\tfrac{\kappa_2}{n} + \kappa_{11})^2$$

then

$$AC_1 + BC_2 = E_1$$

and

$$BC_1 + DC_2 = E_2 \quad .$$

Solving simultaneously yields

$$C_1 = \frac{DE_1 - BE_2}{AD - B^2} \quad ; \qquad (2.3.4)$$

$$C_2 = \frac{AE_2 - BE_1}{AD - B^2} \quad . \qquad (2.3.5)$$

All of the equations pertinent to the general square-root estimator are complex and extremely difficult to analyze critically. However, the calculations for specific distributions are quite easy with the aid of a computer, so tables have been prepared showing the results of applying the general square-root estimator to the thirteen standard distributions.

Evaluation of the optimum values of C_1 and C_2 appear in Table 3. It is immediately obvious that the values are quite dependent upon the form of the distribution being sampled. For example, $C_2 = 0$ for all of the gamma distributions, while $C_1 = 0$ for all of the Wishart distributions. It is also interesting

to note that $C_1 + C_2 \doteq 1$ for all forms and sample sizes of the Pareto distribution.

Table 4 is a comparison of the optimum efficiency ratios of the two forms of the square-root estimator. The efficiency ratio of the general square-root estimator is, of course, better in all cases if the optimum values of C_1 and C_2 for the specific distribution are being used. If the specific type of distribution is unknown it would not be possible to incorporate "workable" values of C_1 and C_2 that would be safe for all distributions.

The use of the general square-root estimator should, therefore, be restricted to those cases where there is a priori knowledge of the form of the parent distribution.

TABLE 3. VALUES OF C_1 AND C_2 FOR OPTIMUM EFFICIENCY FOR THIRTEEN DISTRIBUTIONS; n = 2, 20

Distribution	n = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
(1) $\frac{1}{6} y^3 e^{-y}$	.89	.92	.94	.95	.96	.96	.97	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.98	.98	C_1
	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	C_2
(2) $\frac{1}{2} y^2 e^{-y}$	.86	.90	.92	.94	.95	.95	.96	.96	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.98	C_1
	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	C_2
(3) $y e^{-y}$	.80	.86	.89	.91	.92	.93	.94	.95	.95	.96	.96	.96	.96	.97	.97	.97	.98	.98	.98	C_1
	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	C_2
(4) e^{-4}	.67	.75	.80	.83	.86	.88	.89	.90	.91	.92	.92	.93	.93	.94	.94	.94	.95	.95	.95	C_1
	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	C_2
(5) $\frac{1}{12} y e^{-\sqrt{y}}$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	C_1
	.73	.85	.94	.99	1.03	1.05	1.08	1.09	1.11	1.12	1.13	1.14	1.15	1.15	1.16	1.16	1.17	1.17	1.18	C_2
(6) $\frac{1}{4} \sqrt{y} e^{-\sqrt{y}}$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	C_1
	.67	.82	.91	.98	1.03	1.06	1.09	1.12	1.14	1.15	1.17	1.18	1.19	1.20	1.20	1.21	1.22	1.22	1.23	C_2
(7) $\frac{1}{2} e^{-\sqrt{y}}$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	C_1
	.57	.75	.87	.96	1.03	1.08	1.12	1.16	1.19	1.21	1.23	1.25	1.26	1.28	1.29	1.30	1.31	1.32	1.33	C_2
(8) $\frac{1}{2} y^{-\frac{1}{2}} e^{-\sqrt{y}}$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	C_1
	.40	.60	.80	.90	1.00	1.09	1.16	1.23	1.28	1.33	1.37	1.41	1.44	1.47	1.50	1.52	1.54	1.56	1.58	C_2
(9) y^{-2} $(y \geq 1)$	-1.17	-1.17	-1.18	-1.16	-1.15	-1.14	-1.13	-1.12	-1.11	-1.10	-1.10	-1.09	-1.09	-1.09	-1.08	-1.08	-1.07	-1.07	-1.07	C_1
	1.97	2.29	2.49	2.54	2.60	2.64	2.67	2.69	2.71	2.73	2.74	2.75	2.76	2.76	2.77	2.78	2.78	2.79	2.80	C_2
(10) $2y^{-3}(y \geq 1)$	-2.09	-1.48	-1.25	-1.13	-1.06	-1.01	- .98	- .95	- .93	- .91	- .90	- .88	- .87	- .87	- .86	- .85	- .85	- .84	- .84	C_1
	3.10	2.56	2.37	2.27	2.21	2.17	2.14	2.11	2.10	2.08	2.07	2.06	2.05	2.05	2.04	2.04	2.03	2.03	2.02	C_2
(11) $3y^{-4}(y \geq 1)$	-2.68	-1.88	-1.59	-1.45	-1.36	-1.30	-1.26	-1.23	-1.20	-1.18	-1.17	-1.15	-1.14	-1.13	-1.12	-1.11	-1.11	-1.10	-1.10	C_1
	3.69	2.92	2.64	2.51	2.47	2.37	2.33	2.30	2.27	2.25	2.24	2.23	2.22	2.21	2.20	2.19	2.18	2.18	2.17	C_2
(12) $4y^{-4}(y \geq 1)$	-4.37	-3.25	-2.85	-2.64	-2.52	-2.43	-2.37	-2.33	-2.29	-2.20	-2.24	-2.22	-2.20	-2.19	-2.18	-2.16	-2.15	-2.15	-2.14	C_1
	5.39	4.29	3.90	3.69	3.59	3.49	3.43	3.38	3.35	3.32	3.30	3.28	3.26	3.25	3.23	3.22	3.21	3.21	3.20	C_2
(13) $5y^{-6}(y \geq 1)$	-6.01	-4.61	-4.09	-3.83	-3.67	-3.56	-3.48	-3.42	-3.38	-3.34	-3.31	-3.28	-3.26	-3.24	-3.23	-3.21	-3.20	-3.19	-3.18	C_1
	7.04	5.64	5.13	4.87	4.71	4.60	4.53	4.47	4.42	4.39	4.36	4.33	4.31	4.29	4.27	4.26	4.25	4.24	4.23	C_2

TABLE 4. COMPARISON OF OPTIMUM EFFICIENCY RATIOS OF $\tilde{y} = C_1\bar{y} + C_2\bar{u}^2$ (TOP)

AND $\hat{y} = (1-C)\bar{y} + C\bar{u}^2$ (BOTTOM) FOR THIRTEEN DISTRIBUTIONS; n = 2, 20

Distribution	n = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
(1) $\frac{1}{6}y^3e^{-y}$	.89	.92	.94	.95	.96	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.99	.99	.99	.99
	.96	.96	.96	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.99	.99	.99	.99	.99	.99
(2) $\frac{1}{2}y^2e^{-y}$	.86	.90	.92	.94	.95	.95	.96	.96	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.98
	.94	.94	.95	.96	.96	.96	.97	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.98	.98
(3) ye^{-y}	.80	.86	.89	.91	.92	.93	.94	.95	.95	.96	.96	.96	.96	.97	.97	.97	.97	.97	.98
	.92	.92	.93	.93	.94	.95	.95	.96	.96	.96	.96	.97	.97	.97	.97	.97	.98	.99	.99
(4) e^{-y}	.67	.75	.80	.83	.86	.88	.89	.90	.91	.92	.93	.93	.94	.94	.94	.94	.95	.95	.95
	.84	.84	.86	.87	.88	.90	.91	.91	.92	.93	.93	.94	.94	.94	.95	.95	.95	.95	.96
(5) $\frac{1}{n}ye^{-\sqrt{y}}$	.63	.70	.74	.77	.79	.81	.82	.83	.84	.84	.85	.85	.86	.86	.86	.86	.87	.87	.87
	.73	.73	.75	.77	.79	.81	.83	.84	.85	.86	.87	.88	.88	.89	.90	.90	.91	.91	.91
(6) $\frac{1}{4}\sqrt{y}\,e^{-\sqrt{y}}$	.56	.64	.69	.72	.74	.76	.77	.79	.80	.80	.81	.82	.82	.82	.83	.83	.83	.84	.84
	.67	.67	.69	.72	.74	.76	.78	.80	.81	.82	.84	.84	.85	.86	.87	.87	.88	.88	.89
(7) $\frac{1}{2}e^{-\sqrt{y}}$	.45	.54	.59	.63	.66	.68	.70	.72	.73	.74	.75	.75	.76	.77	.77	.78	.78	.79	.79
	.57	.57	.60	.63	.66	.68	.71	.73	.75	.76	.78	.79	.80	.81	.82	.83	.83	.84	.85
(8) $\frac{1}{2}y^{\frac{1}{2}}e^{-\sqrt{y}}$	.28	.36	.42	.46	.50	.53	.55	.57	.59	.60	.62	.63	.64	.65	.65	.66	.67	.67	.68
	.40	.40	.43	.47	.50	.53	.56	.58	.60	.62	.64	.66	.68	.69	.70	.71	.72	.73	.74
(9) y^{-2} $(y\geq 1)$	.26	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.26	.26
	.28	.28	.33.	.36	.39	.42	.44	.46	.48	.50	.52	.54	.54	.55	.57	.58	.59	.60	.62
(10) $2y^{-3}(y\geq 1)$	.22	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20
	.22	.22	.23	.24	.26	.27	.29	.30	.32	.33	.34	.36	.37	.38	.39	.40	.41	.42	.43
(11) $3y^{-4}(y\geq 1)$	.35	.34	.34	.34.	.34	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33
	.35	.35	.36	.37	.39	.40	.41	.42	.44	.45	.46	.47	.48	.49	.50	.51	.52	.52	.53
(12) $4y^{-5}(y\geq 1)$	.44	.42	.41	.41	.41	.41	.41	.41	.40	.40	.40	.40	.40	.40	.40	.40	.40	.40	.40
	.45	.45	.47	.48	.50	.52	.53	.55	.56	.57	.59	.60	.61	.62	.63	.64	.65	.66	.66
(13) $5y^{-6}(y\geq 1)$	.49	.46	.45	.44	.44	.44	.44	.44	.44	.43	.43	.43	.43	.43	.43	.43	.43	.43	.43
	.51	.51	.52	.54	.56	.57	.59	.61	.62	.63	.65	.66	.67	.68	.69	.70	.71	.71	.72

3. THE CUBE-ROOT ESTIMATOR

3.1 Introduction

When a population consists of both positive and negative numbers, the square root estimator is, of course, impossible to use. The cube root of a negative number is defined so an estimator of the form $\hat{y} = (1 - C)\,\bar{y} + C\bar{v}^3$ is suggested.

3.2 Definitions

(a) $v_i = y_i^{\frac{1}{3}}$

(b) $\bar{v} = \frac{1}{n} \sum_1^n v_i$

(c) $k_1 = \bar{v}$

(d) $k_2 = \frac{1}{n-1} \sum_1^n (v_i - \bar{v})^2$

(e) $k_3 = \frac{n}{(n-1)(n-2)} \Sigma (v_i - \bar{v})^3$

(f) $k_4 = \frac{1}{n^{(4)}} \{(n^3 + n^2)S_4 - 4(n^2 + n)S_3S_1 - 3(n^2 - n)S_2^2$

$+ 12n\, S_2S_1^2 - 6S_1^4\}$

(g) $k_5 = \frac{1}{n^{(5)}} \{(n^4 + 5n^3)S_5 - 5(n^3 + 5n^2)S_4S_1 - 10(n^3 - n^2)S_3S_2$

$+ 20(n^2 + 2n)S_3S_1^2 + 30(n^2 - n)S_2^2S_1 - 60S_2S_1^3 + 24S_1^5\}$

(h) $k_6 = \frac{1}{n^{(6)}} \{(n^5 + 16n^4 + 11n^3 - 4n^2)S_6$

$- 6(n^4 + 16n^3 + 11n^2 - 4n)S_5S_1 - 15n(n - 1)^2(n + 4)S_4S_2$

$- 10(n^4 - 2n^3 + 5n^2 - 4n)S_3^2 + 30(n^3 + 9n^2 + 2n)S_4S_1^2$

$+ 120(n^3 - n)S_3S_2S_1 + 30(n^3 - 3n^2 + 2n)S_2^3$

$- 120(n^2 + 3n)S_3S_1^3 - 270(n^2 - n)S_2^2S_1^2 + 360S_2S_1^4 - 120S_1^6\}$

(i) $S_j = \sum_{i=1}^{n} (v_i)^j$

3.3 Derivations of Bias and Error Mean Square

The cube root estimator will be used in the form $\hat{y} = (1 - C)\,\bar{y} + C\bar{v}^3$ with $\bar{y}$ being expressed in terms of k statistics of the u_i's.

Since $k_3 = \frac{n}{(n-1)(n-2)} \Sigma(v_i - \bar{v})^3$ it is possible to expand the last term and solve for $\Sigma v_i^3 = \frac{(n-1)(n-2)}{n} k_3 + 3(n-1)k_1k_2 + nk_1^3$.

Now,

$$\bar{y} = \frac{1}{n}\sum_{1}^{n} y_i = \frac{1}{n}\Sigma v_i^3 = \left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)k_3 + 3\left(\frac{n-1}{n}\right)k_1k_2 + k_1^3$$

so,

$$\hat{y} = (1 - C)\{\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)k_3 + 3\left(\frac{n-1}{n}\right)k_1k_2 + k_1^3\} + C\,k_1^3$$

$$= \bar{y} - C\left(\frac{n-1}{n}\right)\left[\left(\frac{n-2}{n}\right)k_3 + 3k_1k_2\right] \qquad (3.3.1)$$

or

$$\hat{y} = (1 - C)\left[\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)k_3 + 3\left(\frac{n-1}{n}\right)k_2k_1\right] + k_1^3 \qquad (3.3.1a)$$

$$E(\hat{y}) = \bar{Y} - C\left(\frac{n-1}{n}\right)\left[\left(\frac{n-2}{n}\right)E(k_3) + 3E(k_1k_2)\right]$$

$$= \bar{Y} - C\left(\frac{n-1}{n}\right)\left[\left(\frac{n-2}{n}\right)\kappa_3 + 3\left(\frac{1}{n}\kappa_3 + \kappa_2\kappa_1\right)\right]$$

$$= \bar{Y} - C\left(\frac{n-1}{n}\right)\left[\left(\frac{n+1}{n}\right)\kappa_3 + 3\kappa_2\kappa_1\right]$$

$$B(\hat{y}) = -C\left(\frac{n-1}{n}\right)\left[\left(\frac{n+1}{n}\right)\kappa_3 + 3\kappa_2\kappa_1\right] \qquad (3.3.2)$$

$$V(\bar{y}) = \left(\frac{n-1}{n}\right)^2\left(\frac{n-2}{n}\right)^2 V(k_3) + 9\left(\frac{n-1}{n}\right)^2 V(k_2k_1) + V(k_1^3)$$

$$+ 6\left(\frac{n-1}{n}\right)^2\left(\frac{n-2}{n}\right)\mathrm{Cov}(k_3, k_2k_1) + 2\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\mathrm{Cov}(k_3, k_1^3)$$

$$+ 6\left(\frac{n-1}{n}\right)\mathrm{Cov}(k_1^3, k_2k_1) \qquad (3.3.3)$$

$$V(\hat{y}) = (1 - C)^2 \left(\frac{n-1}{n}\right)^2 \left(\frac{n-2}{n}\right)^2 V(k_3) + 9(1 - C)^2 \left(\frac{n-1}{n}\right)^2 V(k_1 k_2)$$

$$+ V(k_1^3) + 6(1 - C)^2 \left(\frac{n-1}{n}\right)^2 \left(\frac{n-2}{n}\right) \mathrm{Cov}(k_3, k_2 k_1)$$

$$+ 2(1 - C)\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right) \mathrm{Cov}(k_3, k_1^3)$$

$$+ 6(1 - C)\left(\frac{n-1}{n}\right) \mathrm{Cov}(k_1^3, k_2 k_1) \qquad (3.3.4)$$

$$\mathrm{EMS}(\hat{y}) = V(\hat{y}) + [B(\hat{y})]^2$$

$$\frac{\partial \mathrm{EMS}(\hat{y})}{\partial C} = 2(C - 1)\left(\frac{n-1}{n}\right)^2 V(k_3) + 18(C - 1)\left(\frac{n-1}{n}\right)^2 V(k_1 k_2)$$

$$+ 12(C - 1)\left(\frac{n-1}{n}\right)^2 \left(\frac{n-2}{n}\right) \mathrm{Cov}(k_3, k_2 k_1)$$

$$- 2\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right) \mathrm{Cov}(k_3, k_1^3) - 6\left(\frac{n-1}{n}\right) \mathrm{Cov}(k_1^3, k_2 k_1)$$

$$+ 2C\left(\frac{n-1}{n}\right)^2 \left[\left(\frac{n+1}{n}\right)\kappa_3 + 3\kappa_2 \kappa_1\right]^2 .$$

Equating the derivative of $\mathrm{EMS}(\hat{y})$ to zero and solving for C yields

$$C_0 = \frac{A + \left(\frac{n-2}{n}\right) \mathrm{Cov}(k_3, k_1^3) + 3\mathrm{Cov}(k_1^3, k_2 k_1)}{A + \left(\frac{n-1}{n}\right)\left[\left(\frac{n+1}{n}\right)\kappa_3 + 3\kappa_1 \kappa_2\right]^2} \qquad (3.3.5)$$

where

$$A = \left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)^2 V(k_3) + 9\left(\frac{n-1}{n}\right) V(k_1 k_2) + 6\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right) \mathrm{Cov}(k_3, k_2 k_1) \ .$$

In order to evaluate $\mathrm{EMS}(\hat{y})$ it is necessary to determine those variance and covariance terms appearing in (3.3.3), (3.3.4), and (3.3.5). The derivations will be for infinite populations only.

$$V(k_3) = E(k_3^2) - [E(k_3)]^2 = E(k_3^2) - \kappa_3^2$$

$$= E\left[\frac{1}{n} k_6 + \frac{9}{n-1} k_{42} + \frac{n+8}{n-1} k_{33} + \frac{6n}{(n-1)(n-2)} k_{222}\right] - \kappa_3^2$$

$$= \frac{1}{n} \kappa_6 + \frac{9}{n-1} \kappa_4 \kappa_2 + \frac{n+8}{n-1} \kappa_3^2 + \frac{6n}{(n-1)(n-2)} \kappa_2^3 \ . \qquad (3.3.6)$$

$$V(k_1 k_2) = E(k_1^2 k_2^2) - [E(k_1 k_2)]^2$$

$$= E\left[\frac{1}{n^3} k_6 + \frac{2}{n^2} k_{51} + \frac{3n+1}{n^2(n-1)} k_{42} + \frac{2(n+1)}{n^2(n-1)} k_{33} + \frac{1}{n} k_{411}\right.$$

$$\left. + \frac{4(n+1)}{n(n-1)} k_{321} + \frac{n+1}{n(n-1)} k_{222} + \frac{n+1}{n-1} k_{2211}\right]$$

$$- \left[E\left(\frac{1}{n} k_3 + k_{21}\right)\right]^2$$

$$= \frac{1}{n^3} \kappa_6 + \frac{2}{n^2} \kappa_5 \kappa_1 + \frac{3n+1}{n^2(n-1)} \kappa_4 \kappa_2 + \frac{2(n+1)}{n^2(n-1)} \kappa_3^2 + \frac{1}{n} \kappa_4 \kappa_1^2$$

$$+ \frac{4(n+1)}{n(n-1)} \kappa_3 \kappa_2 \kappa_1 + \frac{n+1}{n(n-1)} \kappa_2^3 + \frac{n+1}{n-1} \kappa_2^2 \kappa_1^2$$

$$- \left[\frac{1}{n}\kappa_3 + \kappa_{21}\right]^2$$

$$= \frac{1}{n^3}\kappa_6 + \frac{2}{n^2}\kappa_5\kappa_1 + \frac{3n+1}{n^2(n-1)}\kappa_4\kappa_2 + \frac{n+3}{n^2(n-1)}\kappa_3^2 + \frac{1}{n}\kappa_4\kappa_1^2$$

$$+ \frac{2(n+3)}{n(n-1)}\kappa_3\kappa_2\kappa_1 + \frac{n+1}{n(n-1)}\kappa_2^3 + \frac{2}{n-1}\kappa_2^2\kappa_1^2 \ . \qquad (3.3.7)$$

$$V(k_1^3) = E(k_1^6) - [E(k_1^3)]^2 \ .$$

Using techniques similar to those used in (22) and (23), this yields;

$$V(k_1^3) = \frac{1}{n}\left\{\frac{1}{n^4}\kappa_6 + \frac{6}{n^3}\kappa_5\kappa_1 + \frac{15}{n^3}\kappa_4\kappa_2 + \frac{9}{n^3}\kappa_3^2 + \frac{15}{n^2}\kappa_4\kappa_1^2\right.$$

$$\left. + \frac{54}{n^2}\kappa_3\kappa_2\kappa_1 + \frac{15}{n^2}\kappa_2^3 + \frac{18}{n}\kappa_3\kappa_1^3 + \frac{36}{n}\kappa_2^2\kappa_1^2 + 9\kappa_2\kappa_1^4\right\} \qquad (3.3.8)$$

$$\mathrm{cov}(k_3, k_1k_2) = E(k_3, k_1k_2) - E(k_3)E(k_1k_2)$$

$$= \frac{1}{n^2}\kappa_6 + \frac{1}{n}\kappa_5\kappa_1 + \frac{n+5}{n(n-1)}\kappa_4\kappa_2 + \frac{6}{n(n-1)}\kappa_3^2$$

$$+ \frac{6}{n-1}\kappa_3\kappa_2\kappa_1 \ . \qquad (3.3.9)$$

$$\mathrm{Cov}(k_3,\ k_1^3) = E(k_3k_1^3) - E(k_3)E(k_1^3)$$

$$= \frac{1}{n^3}\kappa_6 + \frac{3}{n^2}\kappa_5\kappa_1 + \frac{3}{n^2}\kappa_4\kappa_2 + \frac{3}{n}\kappa_4\kappa_1^2 \ . \qquad (3.3.10)$$

$$\mathrm{Cov}(k_1k_2\ ,\ k_1^3) = E(k_1^4k_2) - E(k_1k_2)E(k_1^3)$$

$$= \frac{1}{n^4}\kappa_6 + \frac{4}{n^3}\kappa_5\kappa_1 + \frac{7}{n^3}\kappa_4\kappa_2 + \frac{3}{n^3}\kappa_3^2 + \frac{6}{n^2}\kappa_4\kappa_1^2$$

$$+ \frac{12}{n^2}\kappa_3\kappa_2\kappa_1 + \frac{3}{n^2}\kappa_2^3 + \frac{3}{n}\kappa_3\kappa_1^3 + \frac{3}{n}\kappa_2^2\kappa_1^2 \ . \qquad (3.3.11)$$

3.4 The Bias of the Cube-Root Estimator

$$B(y) = -C\left(\frac{n-1}{n}\right)\left[\left(\frac{n+1}{n}\right)\kappa_3 + 3\kappa_2\kappa_1\right]$$

is a function of the first three moments of the cube root distribution and is, therefore, sensitive to large values of these moments. The third moment of the cube-root distribution is apt to be small and cause little problem. The effect of $3\kappa_2\kappa_1$ is not so easily dismissed. The effect of a non-zero mean on the bias could be severe, especially since the mean square error is increased by the square of the bias.

3.5 Types of Distributions for which EMS($\hat{y}$) Can be Made Substantially Less than V($\bar{y}$)

Consider the distribution of errors that would be encountered in a corporate account audit. Most, by far, of the account entries would be correct, i.e., with zero error. A small portion would have errors and these errors would be both positive and negative. One of the more onerous duties of an auditor is to determine the average amount of error in such accounts in order to detect if the total is substantially in error. Since he is sampling for a rare attribute (error), his sample size usually must be quite large in order to be effective. An estimator which would contain the same amount of information with a smaller sample size would be valuable.

In order to evaluate the effectiveness of the cube-root estimator in such a situation, three types of error distributions will be considered using the following definitions:

P = proportion of population containing error.

S = proportion of the errors which are negative.

(1) The rectangular distribution of errors,

$$f(y) = PS \qquad -1 \leq y < 0$$

$$= 1 - P \qquad y = 0$$

$$= P(1 - S) \qquad 0 < y \leq 1 .$$

In this case we are assuming that the positive errors of various magnitudes are equally likely and the negative errors of various magnitudes are, likewise, equally likely.

If

$$v_i = y_i^{\frac{1}{3}}$$

then

$$g(v) = 3PS\, v^3 \qquad -1 \leq v < 0$$

$$= 1 - P \qquad v = 0$$

$$= 3P(1 - S) \qquad 0 < v \leq 1 \quad .$$

Figures 3(a), 3(b), and 3(c) illustrate the gains in efficiency which can be attained for P = .02, .10, and .20, and for S = .95 and .45 at each of these levels of P. The efficiency factor (R) is shown as a function of C.

Extremely large gains are possible in the populations which are only slightly unbalanced. For small sample sizes large gains are attainable even when 19% of the population is in error to one side of zero while only 1% are in error on the other side. For larger sample sizes, however, the possible gains in efficiency become much smaller and the range of values of C which will allow gain becomes much more critical.

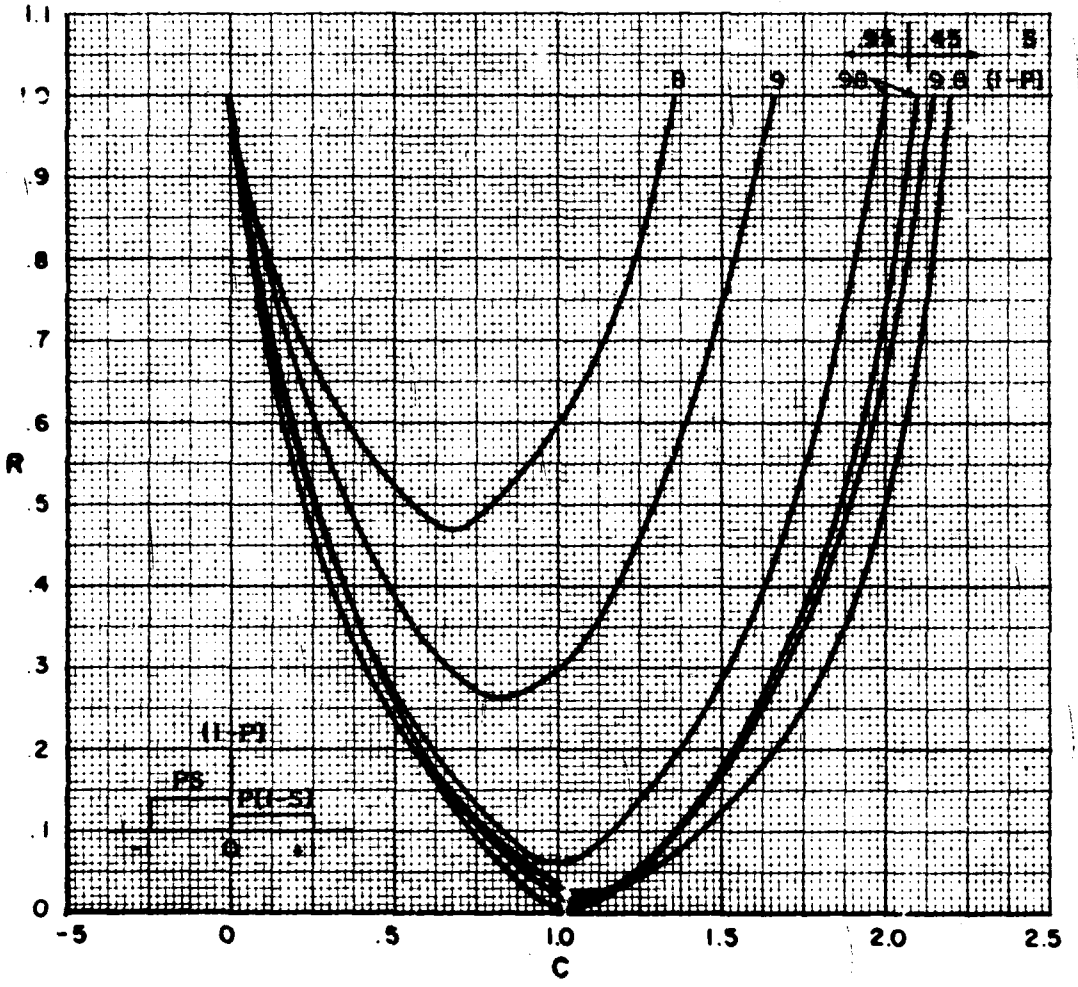


Figure 3(a). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Rectangular distribution of errors, n = 5

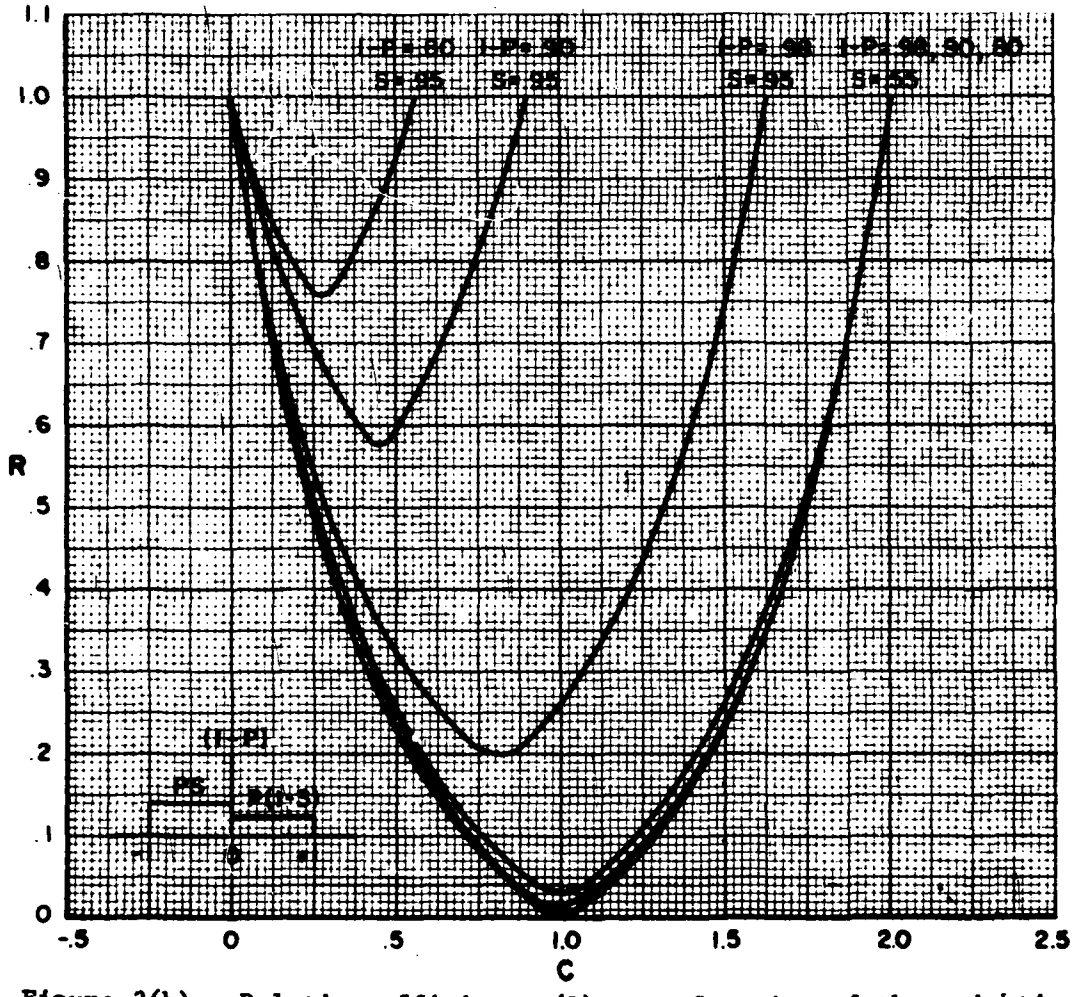


Figure 3(b). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Rectangular distribution of errors, n = 20

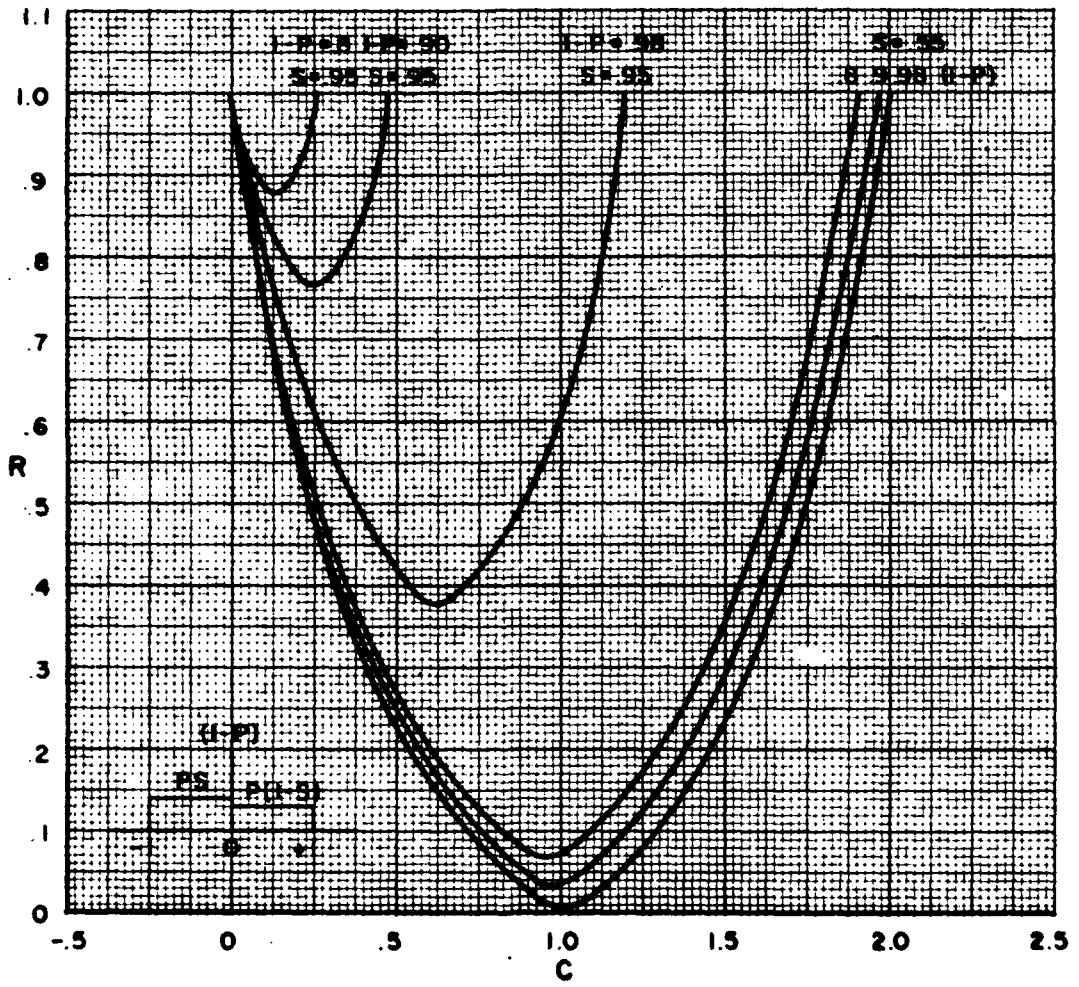


Figure 3(c). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Rectangular distribution of errors, n = 50

(2) The uniformly decreasing distribution

$$f(y) = 2PS(1 + y) \qquad -1 \leq y < 0$$

$$= (1 - P) \qquad y = 0$$

$$= 2P(1 - S)(1 - y) \qquad 0 < y \leq 1$$

This distribution is similar to the rectangular distribution except that the probability of error decreases with distance from 0.

Letting

$$v_i = y_i^{\frac{1}{3}}$$

$$g(v) = 6PSv^2(1 + v^3) \qquad -1 \leq v < 0$$

$$= 1 - P \qquad v = 0$$

$$= 6P(1 - S)v^2(1 - v^3) \qquad 0 < v \leq 1$$

Comparison of Figures 3(d), 3(e), and 3(f) shows the same results as occurs in the rectangular distribution except that the imbalance does not have as great an effect on the efficiency ratio. This, of course, is because there is a smaller effect on the first three central moments.

(3) The parabolic distribution of errors

$$f(y) = 3PS(1 + y)^2 \qquad -1 \leq y < 0$$

$$= 1 - P \qquad y = 0$$

$$= 3P(1 - S)(1 - y)^2 \qquad 0 < y \leq 1$$

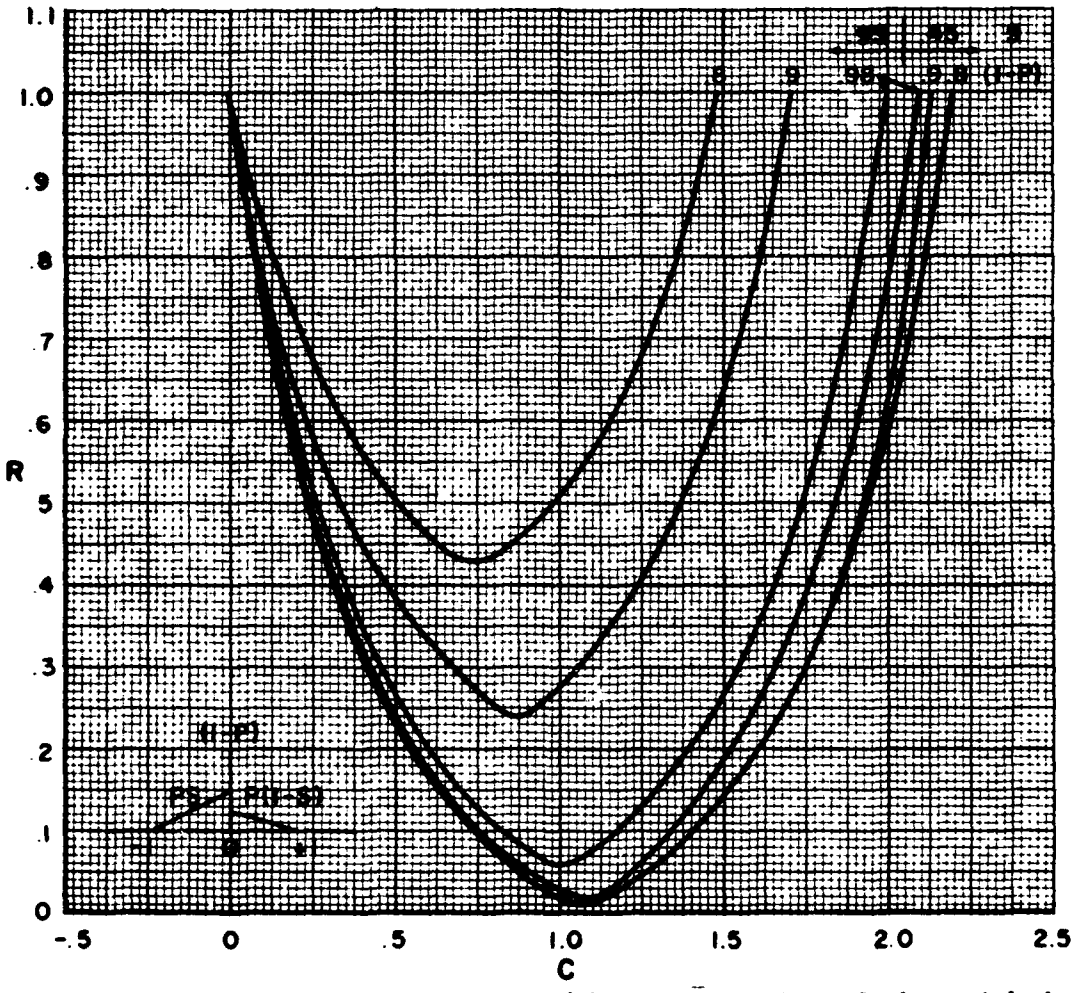


Figure 3(d). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Uniformly decreasing distribution of errors, n = 5

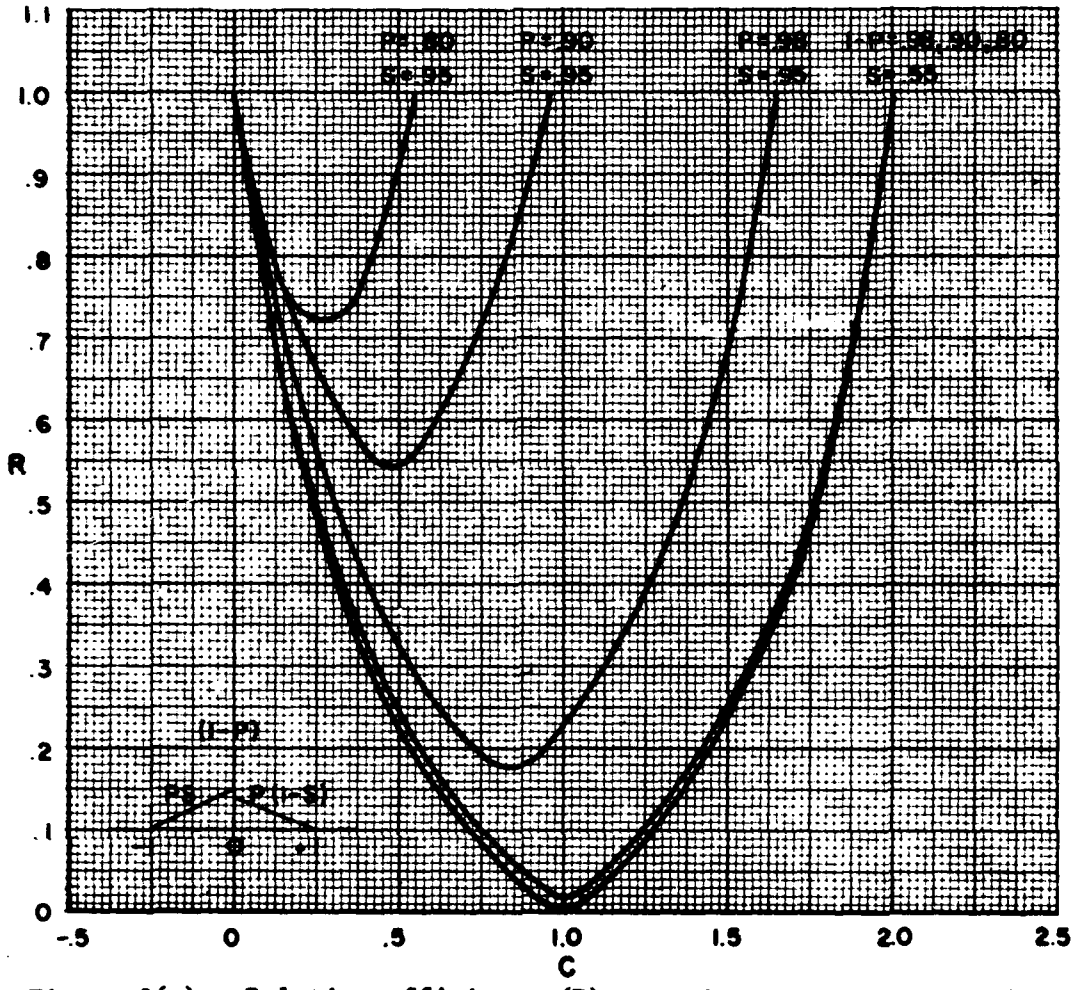


Figure 3(e). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Uniformly decreasing distribution of errors, n = 20

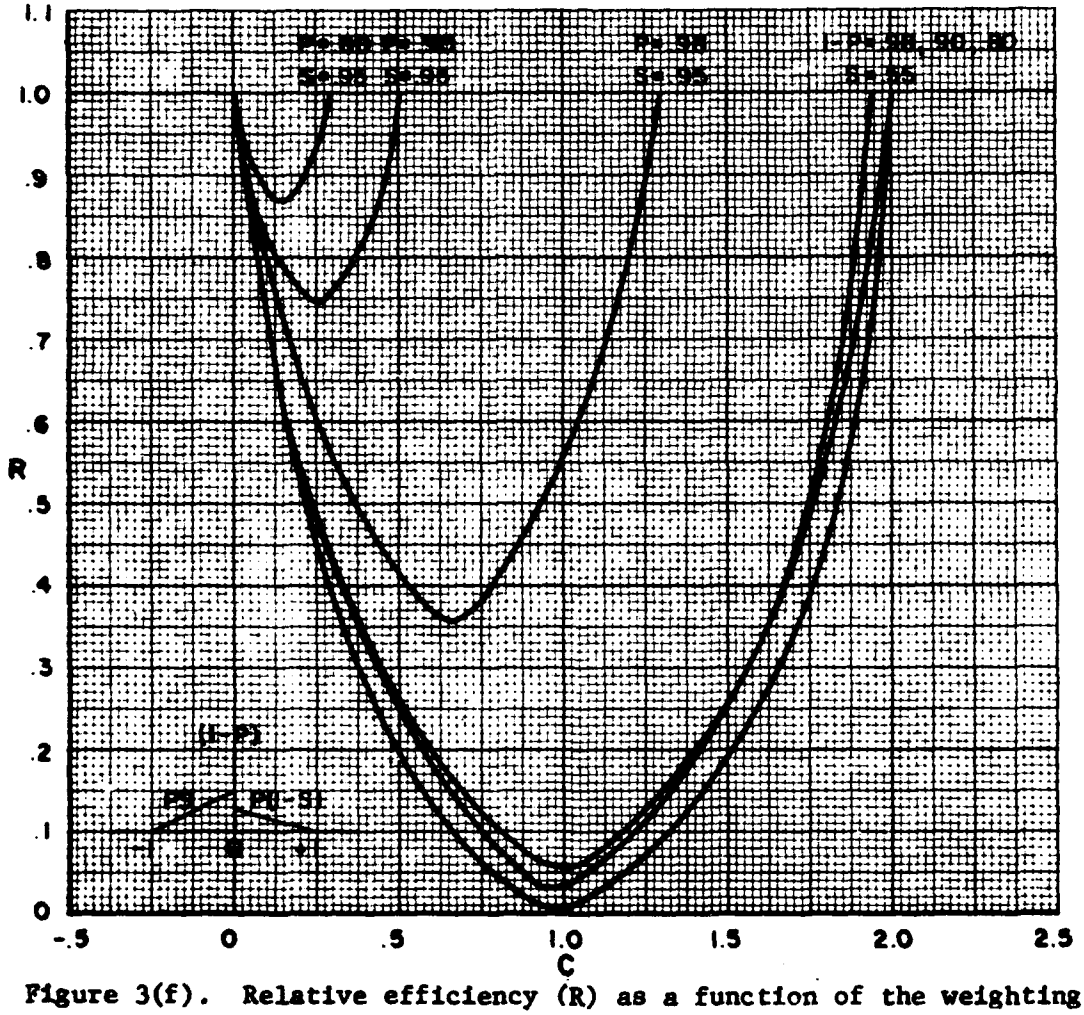


Figure 3(f). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Uniformly decreasing distribution of errors, n = 50

The parabolic distribution is similar to the rectangular and uniformly decreasing distributions except that it reduced the probabilities of larger errors by the square of the distance from zero.

Letting

$$v_i = y_i^{\frac{1}{3}}$$

$$g(v) = 9PS(v + v^4)^2 \qquad -1 \leq v < 0$$

$$= 1 - P \qquad v = 0$$

$$= 9P(1 - S)(v - v^4)^2 \qquad 0 < v \leq 1$$

Figures 3(g), 3(h), and 3(i) illustrate once again the same basic results shown by the rectangular and uniformly decreasing distributions. The better the balance, the greater the gain.

The similarity in the results of these distributions serves to indicate that for small sample sizes there is a value of C which will allow the cube-root estimator to be used on a class of distributions. For larger sample sizes it is important to have some idea of the amount of imbalance before choosing C. It is possible to estimate P and S from the sample, a posteriori, and to choose C from the results. This will change the mean square error of the estimator but it will allow some hedging against a loss of efficiency.

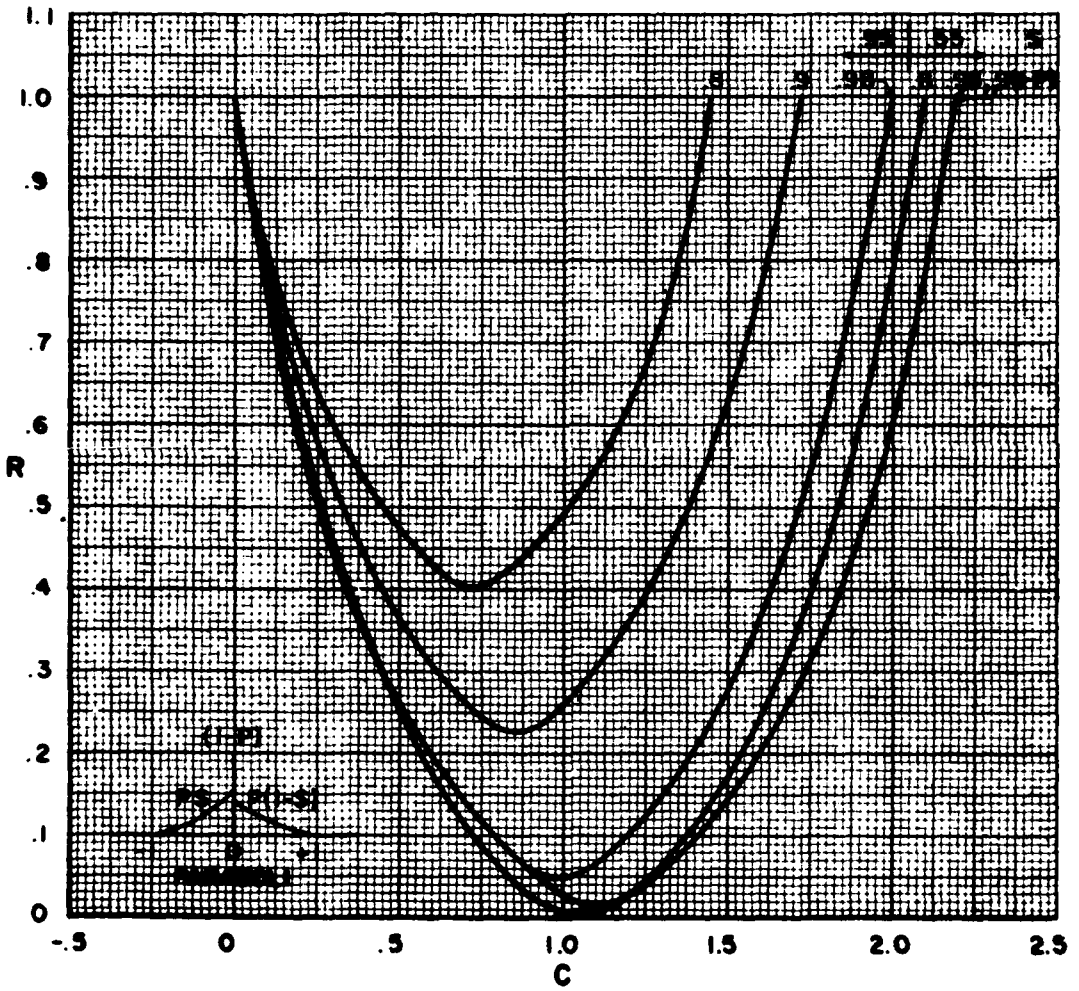


Figure 3(g). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Parabolic distribution of errors, n = 5

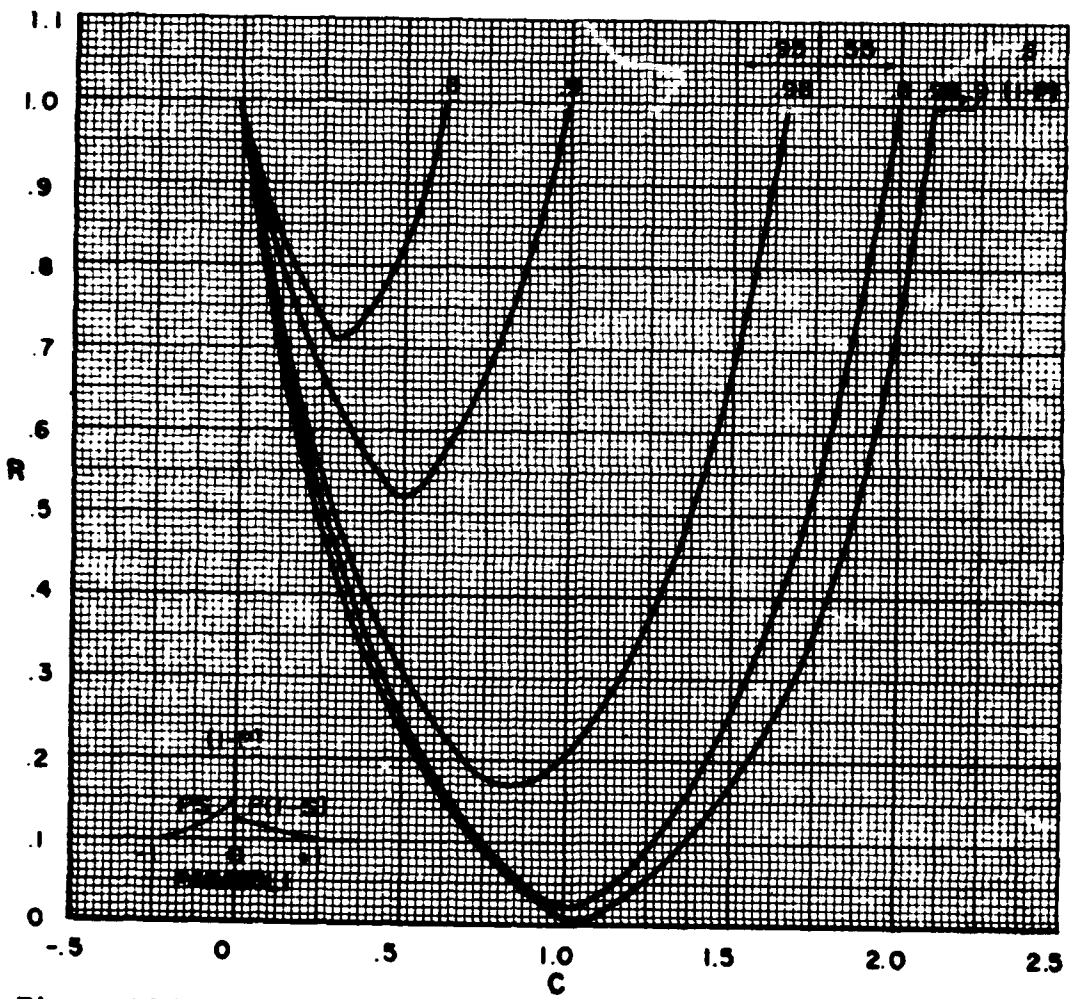


Figure 3(h). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Parabolic distribution of errors, n = 20

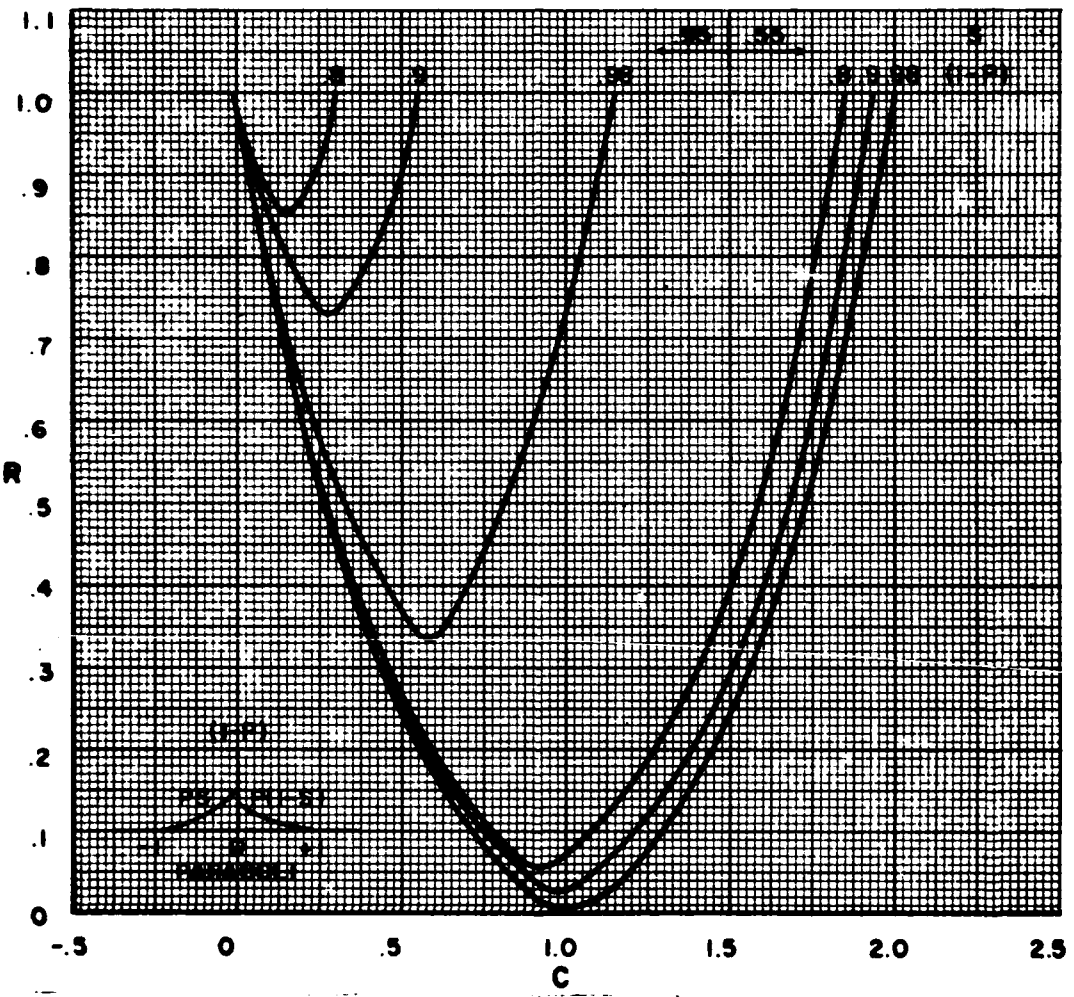


Figure 3(i). Relative efficiency (R) as a function of the weighting constant (C) for the cube root estimator. Parabolic distribution of errors, n = 50

3.6 The Use of the Cube-Root Estimator to Estimate Changes in the Mean

Small sample surveys are often used to detect if the mean of a population has changed over time or after some treatment. Consider, for example, a population which has previously been surveyed in total or, at least, by a very large sample. Then at a later date a small sample is taken to detect if the mean has changed. Letting μ_0 be the prior mean of the population, X_i be an observation made in the later survey, and $X_i - \mu_0 = y_i$. Then y_i is distributed identically with X with the exception that $\overline{Y} = \overline{X} - \mu_0$.

Utilizing the cube-root estimator on the sample values of y would produce an estimator $\tilde{y} = (1 - C)\bar{y} + c\bar{v}^3$ with the properties that have previously been described. If the values of the first six moments of the cube-root distribution can be estimated by using the information from the original survey it is possible to predetermine a value of C which is apt to give good results. Further, if the distribution is fairly symmetrical the bias and mean square error of the cube-root estimator may be quite small compared to $\bar{y}$.

3.7 A Simulation to Verify the Efficiency of the Cube-Root Estimator

In order to verify that the cube-root estimator does, indeed,

TABLE 5. DISTRIBUTION OF ERRORS IN 100 AUDITED ACCOUNTS OF A WHOLESALE FIRM

Error Size ($) y	v	f
0	0.00000	90
-.52	-.804145	1
-.80	-.928318	1
-1.00	-1.000000	2
-2.00	-1.259921	1
-3.00	-1.442250	1
+0.10	+0.464159	1
+0.40	+0.736806	1
+1.00	+1.000000	1
+10.00	+2.154435	1

$\bar{Y} = .0318$
$\sigma^2 = 1.1698$
} Original population parameters

$\bar{V} = .0016$
$\mu_2 = .1553$
$\mu_3 = .0403$
$\mu_4 = .3319$
$\mu_5 = .3865$
$\mu_6 = 1.2169$
} Cube-root transformed population parameters

attain the claimed efficiencies, a set of account errors was obtained from the auditing records of a wholesaling firm. The data shown in Table 5 are a sample of 100 account errors, but they will be used as if they constitute an entire population.

Referring to Figure 3(a) it can be seen that C = 1 is both safe and likely to produce a small mean square error when n = 3. There is the added advantage that the cube-root estimator is quite simple to calculate when the weighting constant is equal to one.

Equations (3.3.2) and (3.3.4) reveal that when n = 3 and C = 1

$$B(\hat{y}) = .0302 \quad ,$$

$$V(\hat{y}) = .0101$$

and

$$EMS(\hat{y}) = .0110 \quad .$$

Through the use of an electronic computer every combination of the one hundred values taken three at a time were picked and $\hat{y}$ was calculated for each combination. Calculating the moments of these actual sample values yielded;

$$E(\hat{y}) = .00144$$

$$B(\hat{y}) = .0318 - .0014 = -.0304$$

$$V(\hat{y}) = .0072$$

$$EMS(\hat{y}) = .0081 .$$

The error mean square calculated from the actual sample values is smaller than equations (3.3.2) and (3.3.4) predicted indicating that the finite population correction, which was ignored in the development of the dube-root estimator, has more influence in the case of the cube-root estimator than it has in the case of the sample mean.

4. USES OF ROOT-ESTIMATORS IN STRATIFIED SAMPLING

4.1 Introduction

Stratified random sampling is a well known and commonly used sampling technique wherein a random sample is drawn from each of the mutually exclusive strata (or subpopulations) of the population. This technique is particularly useful on populations which are naturally subdivided into subpopulations, each of which are more homogeneous than the whole. In such a case each stratum mean is estimated by the mean of individual observations drawn from that stratum. If the stratum sizes are known then the stratum totals are estimable, and through them the population total and mean are estimable. Estimates of a population mean or total obtained in this fashion have a variance that is smaller than the variance of estimates obtained from a simple random sample of the whole population. When each of the stratum means can be estimated with smaller mean square error through use of a biased estimator, it would seem that these biased estimators could be used to good advantage in the estimation of the population mean. As is shown below, this is not the case.

4.2 Definitions

(a) L = number of strata in population

(b) N_i = number of elements in the i^{th} stratum

(c) $\sum_{1}^{L} N_i = N$ = total number of elements in population

(d) n_i = size of sample from the i^{th} stratum

(e) $\Sigma n_i = n$ = total size of sample

(f) $\bar{Y}_i = \frac{1}{N_i} \sum_{i}^{N_i} y_i$

(g) $\hat{y}_i = f(y_{i,1}, y_{i,2}, \ldots, y_{i,n_i})$

(h) $\bar{Y} = \sum_{1}^{L} \frac{N_i}{N} \bar{Y}_i = \frac{1}{N} \sum_{i=1}^{L} \sum_{j=1}^{N_i} y_{ij}$

(i) $\hat{Y} = \sum_{1}^{L} w_i y_i$

(j) $B(\hat{Y}) = E(\hat{Y}) - \bar{Y}$ Bias of Y as estimator of $\bar{Y}$

(k) $B_i = E(\hat{y}_i) - \bar{Y}_i$ Bias of $\hat{y}_i$ as estimator of $\bar{Y}_i$

4.3 The Bias and Error Mean Square of $\hat{Y}$ as Estimator of $\bar{Y}$

The bias of $\hat{Y}$ as an estimator of $\bar{Y}$ is

$$B(\hat{Y}) = E(\hat{Y}) - \bar{Y} = E[\sum_{1}^{L} w_i \hat{y}_i] - \bar{Y} = \sum_{1}^{L} w_i (\bar{Y}_i + B_i) - \sum_{1}^{L} \frac{N_i}{N} \bar{Y}_i$$

$$= \Sigma (w_i - \frac{N_i}{N}) \bar{Y}_i + \sum_{1}^{L} w_i B_i \, .$$

The variance of $\hat{y}$ is, then

$$V(\hat{y}) = V(\Sigma w_i \hat{y}_i) = \sum_{1}^{L} w_1^2 V(\hat{y}_i)$$

and

$$EMS(\hat{Y}) = V(\hat{Y}) + [B(\hat{Y})]^2 = \Sigma w_i^2 V(\hat{y}_i) + [\Sigma(w_i - \frac{N_i}{N})\bar{Y}_i + \Sigma w_i B_i]^2 .$$

It is apparent that letting $w_i = \frac{N_i}{N}$ will reduce the error mean square to a minimum under the given conditions, so that

$$EMS(\hat{Y}) = \Sigma \frac{N_i^2}{N^2} V(\hat{y}_i) + [\sum_1^L \frac{N_i^2}{N^2} B_i]^2 .$$

If we now use the relationship $V(\hat{y}_i) = EMS(\hat{y}_i) - (B_i)^2$, the $EMS(\hat{Y})$ becomes

$$EMS(\hat{Y}) = \Sigma \frac{N_i^2}{N^2} [EMS(\hat{y}_i) - B_i^2] + [\Sigma \frac{N_i^2}{N^2} B_i]^2$$

$$= \frac{1}{N^2} \Sigma N_i^2 [EMS(\hat{y}_i)] + \frac{1}{N^2} \sum_{i=1}^{L} \sum_{i \neq 1}^{L} N_i N_{i'} B_i B_{i'}$$

$$= \frac{1}{N^2} [\sum_1^L N_i^2 [EMS(\hat{y}_i)] + \sum_{i=1}^{L} \sum_{i \neq 1} N_i N_{i'} B_i B_{i'}] .$$

4.4 Investigation of $EMS(\hat{Y})$ Compared to $V(\bar{y}_{st})$

If $\bar{y}_{st} = \frac{1}{N} \Sigma N_i \bar{y}_i$, and each $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_i$, which is unbiased for $\bar{Y}_i$, then

$$EMS(\bar{y}_{st}) = \frac{1}{N^2} \sum_1^L \Sigma N_i^2 V(\bar{y}_i) = V(\bar{y}_{st}) , \quad \text{since all } B_i = 0 .$$

$$\text{This makes } R = \frac{EMS(Y)}{V(\bar{y}_{st})} = \frac{\sum_1^L \Sigma N_i^2 [EMS(\hat{y}_i)]}{\Sigma N_i^2 V(\bar{y}_i)} + \frac{\Sigma_i \Sigma_{i' \neq i} N_i N_{i'} B_i B_{i'}}{\Sigma N_i^2 V(\bar{y}_i)}$$

which is a most unfortunate result. The second term is the sum of all the L(L-1) cross-products of the $N_i B_i$'s divided by $\Sigma N_i^2 V(\bar{y}_{st})$ which contains only L terms. In order to clear away some of the confusing factors, assume that each stratum has the same number of elements and equal variance. Then each $N_i = N/L$ and

$$EMS(\hat{Y}) = \frac{1}{N^2}\{L\frac{N^2}{L^2}[EMS(\bar{y})] + L(L-1)\frac{N^2}{L^2}B^2\}$$

$$= \frac{1}{L}\{EMS(\hat{y}) + (L-1)B^2\}$$

and

$$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{1}^{L}\frac{N^2}{L^2}V(\bar{y}_i) = \frac{1}{L}V(\bar{y}) \ .$$

This, then, makes

$$R = \frac{EMS(\hat{y})}{V(\bar{y})} + \frac{(L-1)B^2}{V(\bar{y})} \ .$$

Therefore, if the biased estimator $\hat{y}$ is sufficiently efficient compared to $\bar{y}$ and L is small, the resultant overall efficiency may be comparable. However, as L increases $\frac{(L-1)B^2}{V(\bar{y})}$ is certain to become large enough to cause a reduction in efficiency below that of $\bar{y}$.

It is easily seen, therefore, that estimators which are biased are rather dangerous to use for estimating the population mean or total when utilizing stratified random sampling. This is especially true when all B_i's are in the same direction as is the case with the square root estimator.

In the case of the cube-root estimator it is possible for the bias terms to be positive for some strata and negative for others. However, the purpose of such a survey, usually, is to detect a consistant bias in error, the very condition which will cause the cube-root error to be inefficient.

Root-estimators, therefore, are not recommended for use in estimating population means or totals in stratified sampling. They are recommended for use in the estimation of strata means and totals when the n_i are small and the distributions within strata are apt to be positively skewed.

5. CONCLUSION

5.1 Summary

The square-root estimator of the form $\hat{y} = (1 - C)\bar{y} + C\bar{u}^2$, $u_i = \sqrt{y_i}$, has been developed for populations consisting of all positive numbers. It was found that for small sample sizes from populations with a large positive skewness there is an optimum value of C (C_0) which will make the mean square error of $\hat{y}$ smaller than that of the sample mean. Indeed, it was found that any value of C between zero and $2C_0$ will have this effect to some extent and that, for a particular sample size, values of C can be determined which will produce smaller mean square errors for a wide class of positively skewed distributions.

The general square-root estimator of the form $\tilde{y} = C_1\bar{y} + C_2\bar{u}^2$ was also investigated. This form was found to produce improvement in the mean square error at the optimum values of C_1 and C_2. The wide variability in the values of C_1 and C_2 between types of distributions made the use of the general square-root estimator a bit unsafe if the form of the distribution is not known a priori.

The cube-root estimator of the form $\hat{y} = (1 - C)\bar{y} + C\bar{v}^3$, $v_i = y_i^{1/3}$, was developed for use on populations consisting of both positive and negative values. It was found that the mean square error of this estimator is quite sensitive to asymmetry in the population. However, for populations of errors which are

predominantly zero the cube-root estimator performed quite well in comparison to $\bar{y}$. It also showed promise for the estimation of small changes in the means of populations for which a previous large sample survey has established good estimators of the higher moments.

An investigation of the use of biased estimators in stratified sampling indicated that root-estimators can be used for the estimation of within stratum means and totals with good results. They should not be used for estimating the population mean or total in stratified sampling, however, because the bias accumulates in the total and overcomes the gains made within the individual strata.

5.2 Future Research

Although the square-root and the cube-root estimators show promise for practical application in small sample estimation, there are several aspects of the problem which need further investigation. Of primary importance is a better description of the classes of distributions for which the root estimators are advantageous. The sampling practioner could make use of a more complete set of model distributions than those exhibited in this paper. It also appears likely that the square-root estimator could be improved by the addition of a constant to populations which have values between zero and one.

This paper is, admittedly, but a beginning in the investigation of biased estimators which exhibit a reduced mean square error. However, we are hopeful that the properties that have been determined for these estimators will encourage further investigation.

REFERENCES

Cochran, W. G., Sampling Techniques, 2nd Ed., 1953, John Wiley and Sons, Inc., New York.

Hartley, H. O. and Rao, J. N. K., [1968]. "A new estimation theory for sample surveys", Biometrika Vol. 55, 547-557.

Kendall, M.S. and Stuart, A., The Advanced Theory of Statistics Vol. 1, 2nd Ed., 1963, Hafner Publishing Company, New York.

Royall, R., [1968]. "An old approach to finite population sampling theory", J. Am. Statist. Ass. Vol. 63, 1269-1279.

Wishart, J., [1952], "Moment coefficients of the k-statistics in samples from a finite population", Biometrika Vol. 39, 1-13.