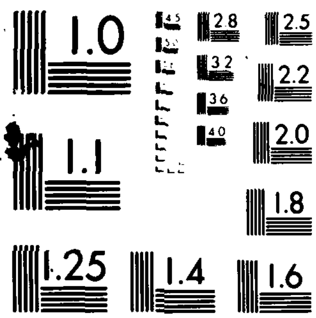
AD-A110 476 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
ESTIMATION OF VARIANCE OF THE RATIO ESTIMATOR.(U)
AUG 81 C WU DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2267 NL

END
DATE
FILMED
3 82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

LEVEL II

2

AD A110476

MRC Technical Summary Report # 2267

ESTIMATION OF VARIANCE OF THE RATIO ESTIMATOR

Chien-Fu Wu

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

17

August 1981

(Received July 20, 1981)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

and

National Science Foundation
Washington, DC 20550

82 02 03 056

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

ESTIMATION OF VARIANCE OF THE RATIO ESTIMATOR

Chien-Fu Wu*

Technical Summary Report #2267

August 1981

ABSTRACT

v sub ∅

v sub 2

A general class of estimators of the variance of the ratio estimator is considered, which includes two standard estimators ν_0 and ν_2 and approximates another estimator v sub H ν_H, suggested by Royall and Eberhardt (1975). Asymptotic expansions for the variances and biases of the proposed estimators are obtained. Based on this ~~we obtain~~ optimal variance estimator in the class, is obtained and compare the relative merits of three estimators v sub ∅, v sub 1 and v sub 2 ν_0, ν_1 and ν_2 without any model assumption. Under a simple regression model a more definite comparison of v sub ∅, v sub 1 and v sub 2 ν_0, ν_1 and ν_2 is made in terms of variance and bias.

AMS (MOS) Subject Classification: 62D05

Key Words: Asymptotic expansion, Bias, Optimal variance estimator, Ratio estimator, Superpopulation, Variance

Work Unit Number 4 - Statistics and Probability

---

* Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-7901846.

SIGNIFICANCE AND EXPLANATION

In estimating the population mean of a character y, we often make use of an auxiliary covariate x whose information is more readily available and is positively correlated with y. One commonly used estimator in survey sampling is the ratio estimator (y-sample mean)(x-population mean)/(x-sample mean). To assess the variability of the estimator, we need an estimator of its variance. Several variance estimators have been compared under the assumption that the finite population itself is a random sample from an infinite superpopulation that is described by a linear model. Such an assumption may not be realistic in practice and usually is hard to verify. We propose a class of variance estimators, which includes or approximates several existing variance estimators in the literature. We then find the asymptotic variance and bias of these estimators and determine the optimal estimators for minimizing variance or bias. No superpopulation model is assumed. If we do assume a regression model over the finite population, strong optimality results are obtained and more definite comparisons of estimators are made.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

ESTIMATION OF VARIANCE OF THE RATIO ESTIMATOR

Chien-Fu Wu*

1. Introduction

Suppose a population consists of N distinct units with values (y_i, x_i), $x_i > 0$, $i=1,\ldots,N$. Denote the population means of y_i and x_i by $\bar{Y}$ and $\bar{X}$. To estimate $\bar{Y}$, it is customary to take a simple random sample of size n and use the ratio estimator $\hat{Y}_R = \bar{y}\,\bar{X}/\bar{x}$, where $\bar{y}$ and $\bar{x}$ are respectively the sample means of y_i and x_i. The mean square error and variance of $\hat{Y}_R$ are each approximated by (Cochran, 1977, p. 155)

$$V = \frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \frac{\bar{Y}}{\bar{X}}x_i\right)^2 , \tag{1}$$

where $f = n/N$ is the sampling fraction. Two commonly used estimators of V are

$$v_0 = \frac{1-f}{n}\frac{1}{n-1}\sum_{i=1}^{n}(y_i - rx_i)^2 \tag{2}$$

and

$$v_2 = \frac{1-f}{n}\left(\frac{\bar{X}}{\bar{x}}\right)^2\frac{1}{n-1}\sum_{i=1}^{n}(y_i - rx_i)^2 , \tag{3}$$

where $r = \bar{y}/\bar{x}$. The asymptotic consistency of v_0 and v_2 and the asymptotic normality of $\hat{Y}_R$ were rigorously established in Scott and Wu (1981). Although the original motivation for $v_2' = v_2/\bar{X}^2$ as a variance

---

* Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706.

Sponsored by the United States Army under Contract No. DAAG29-80-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-7901846.

estimator of the ratio $R = \bar{Y}/\bar{X}$ is the unavailability of $\bar{X}$, it is not clear whether v_2 is worse than v_0 (Cochran, 1977, p. 155). Rao and Rao (1971) studied the small-sample properties of v_0 and v_2 by assuming that the sample is a random sample directly from an infinite superpopulation that can be described by a simple linear regression model of y on x and that x has a gamma distribution. The nature of random sampling from a finite population is not taken into full account in their formulation as distinguished from the usual superpopulation approach. Royall and Eberhardt (1975) noted the bias of v_0 when the finite population is a random sample from a superpopulation in which

$$y_i = \beta x_i + \varepsilon_i \quad , \tag{4}$$

where ε_i are independent with mean zero and variance $\sigma^2 x_i^t$. They suggested the simple modification

$$v_H = v_0 \frac{\bar{x}_c \bar{X}}{\bar{x}^2} \left(1 - \frac{C_x^2}{n}\right)^{-1} \quad , \tag{5}$$

where C_x is the x-sample coefficient of variation and $\bar{x}_c$ is the mean of the $N-n$ units not in the sample. They also noted that $v_H \simeq v_2$ for large n and $N >> n$, thus justifying the use of v_2 from a superpopulation viewpoint. The estimator v_2 was previously recommended by Hájek (1958). The estimator v_H is ξ-unbiased under model (4) with $t = 1$, and remains approximately ξ-unbiased when the variance of ε_i in model (4) is not proportional to x_i. An empirical study of v_H was reported in Royall and Cumberland (1978). All these authors assume that the actual population satisfies a hypothetical infinite population model. It is desirable to have a model-free comparison of these estimators. As Royall and Cumberland (1978, p. 334) pointed out, the conventional theory in sample surveys does not provide any comparison on the relative merits of v_0 and v_2. One of the purposes of this paper is to provide such a model-free comparison of v_0 and v_2.

We consider a more general class of variance estimators

$$v_g = \left(\frac{\bar{X}}{\bar{x}}\right)^g v_0 ,$$

which includes v_0, v_2 and a new estimator

$$v_1 = \frac{1-f}{n} \frac{\bar{X}}{\bar{x}} \frac{1}{n-1} \sum_{i=1}^{n} (y_i - rx_i)^2 ,$$

which is equal to $(v_0 v_2)^{1/2}$. Since the numerator of v_H is a linear combination of v_1 and v_2, it can be adequately approximated by some v_g. In §2 we obtain the leading terms of the mean square error and variance of v_g, which happens to be a quadratic function in g. The optimal variance estimator is then obtained by minimizing this quadratic function. The optimal g denoted g_{opt} is equal to the population regression coefficient of $z_i/\bar{Z}$ over $x_i/\bar{X}$, $i = 1,\ldots,N$, where z_i, defined in (12), depends on the "residual" $e_i = y_i - Rx_i$ and e_i^2. Therefore among v_0, v_1 and v_2, v_0 is the best if $g_{opt} < 0.5$; v_1 the best if $0.5 < g_{opt} < 1.5$; and v_2 the best if $g_{opt} > 1.5$. By further assuming the superpopulation model (4) we show the optimality of v_0 among v_g under $t = 0$ and the optimality of v_1 among v_g under $t = 1$. Note that the ratio estimator $\hat{\bar{Y}}_R$ is the best linear unbiased estimator of $\bar{Y}$ under model (4) with $t = 1$ (Brewer, 1963; Royall, 1970). If the ratio estimator is adopted with this optimality property in mind, then according to our result, one ought to use v_1 as the estimate of variance. Under model (4) with $t > 1$, it is also shown in §2 that $g_{opt} > 1$ with the implication that ν_1 and ν_2 are better than ν_0. Therefore our study justifies the use of ν_1 and ν_2 in practice when one believes that model (4) with $t > 1$ adequately describes the population. Under further distributional assumptions on x, g_{opt} is determined as a function of t. In §3 we obtain the leading terms of the bias of ν_g. We then compare ν_0, ν_1 and ν_2 in terms of their biases under model (4) with general variance pattern t. By further assuming that x has a gamma distribution, we show that, among ν_0, ν_1 and ν_2, ν_0 is the least biased for $t > 1.5$ or $t < 0.6$, ν_1 the least biased for $0.6 < t < 6/7$ and ν_2 the least biased for $6/7 < t < 1.5$.

2. Variance of ν_g

2.1. Asymptotic expansions

We need the following results for asymptotic expansions

$$r-R = \frac{\bar{e}}{\bar{X}} + O(\tfrac{1}{n}) = \frac{\bar{e}}{\bar{X}} - \frac{\bar{e}}{\bar{X}}\,\delta\bar{x} + O(\tfrac{1}{n^2}) \quad , \tag{6}$$

$$\left(\frac{\bar{X}}{\bar{x}}\right)^g = 1-g(\delta\bar{x}) + O(\tfrac{1}{n}) = 1-g(\delta\bar{x}) + \frac{g(g+1)}{2}(\delta\bar{x})^2 + O(\tfrac{1}{n^2}) \quad , \tag{7}$$

$$\begin{aligned}(\bar{x}-\bar{X})^r(\bar{y}-\bar{Y})^s &= O(n^{-\frac{r+s}{2}}) \quad \text{if } r+s \text{ is even} \quad ,\\ &= O(n^{-\frac{r+s+1}{2}}) \quad \text{if } r+s \text{ is odd} \quad ,\end{aligned} \tag{8}$$

$$(\bar{x} - \bar{X})(\bar{y} - \bar{Y})(\bar{z} - \bar{Z}) = O(n^{-2}) \quad , \tag{9}$$

where $\delta\bar{x} = (\bar{x}-\bar{X})/\bar{X}$, $\bar{z}$ is the sample mean of character z from the <u>same</u> simple random sample as $\bar{x}$ and $\bar{y}$, $\bar{Z}$ is the corresponding population mean, $O(n^{-1})$ is of stochastic order n^{-1}, $\bar{e} = n^{-1}(e_1 +\dots+ e_n)$ and $e_i = y_i - Rx_i$ is the residual of y_i to the line connecting $(\bar{X},\bar{Y})$ and the origin. Note $e_1 + \ldots + e_N = 0$. Formulas (6) and (7) follow easily from (8), and formulas (8), (9) can be rigorously justified as in David and Sukhatme (1974). Using (6), (8) and (9), we expand

$$\nu_0 = \frac{1-f}{n-1} \{\frac{1}{n} \sum_1^n e_i^2 - 2 \frac{\Sigma_1^N x_i e_i}{\Sigma_1^N x_i} \bar{e} - 2 \frac{\bar{e}}{\bar{X}} (\frac{1}{n} \sum_1^n x_i e_i - \frac{1}{N} \sum_1^N x_i e_i)$$

$$+ 2 \frac{\bar{e}(\delta\bar{x})}{\bar{X}} \frac{1}{N} \sum_1^N x_i e_i + \frac{\bar{e}^2}{\bar{X}^2} \frac{1}{N} \sum_1^N x_i^2\} + O(\frac{1}{n^3}) \tag{10}$$

$$= \frac{1-f}{n} \bar{z} + O(\frac{1}{n^2}) \quad , \tag{11}$$

where

$$\bar{z} = n^{-1} \sum_1^n z_i \quad , \quad z_i = e_i^2 - 2 \frac{\Sigma_1^N x_i e_i}{\Sigma_1^N x_i} e_i \quad . \tag{12}$$

Note that $\bar{Z} = N^{-1}(e_1^2 + \ldots + e_N^2)$ and $E\ \nu_0 = V + O(n^{-2})$. From (7), (8) and (11),

$$\nu_g = \frac{1-f}{n} \{\bar{z} - g(\delta\bar{x})\bar{Z}\} + O(\frac{1}{n^2}) \tag{13}$$

and the mean square error and variance of ν_g are

$$\mathrm{var}(\nu_g) = (\frac{1-f}{n})^3 \{S_z^2 - 2g \frac{\bar{Z}}{\bar{X}} S_{zx} + g^2 \frac{\bar{Z}^2}{\bar{X}^2} S_x^2\} + O(\frac{1}{n^4}) \quad , \tag{14}$$

where S_z^2 and S_{zx} are the population variance of z and the population covariance of z and x, respectively. The bias square of ν_g is of lower order than the variance and will be studied in §3.

2.2 Optimal choice of ν_g and comparison of ν_0, ν_1, ν_2

The optimal variance estimator is now obtained by minimizing expression (14) with respect to g, the optimal g denoted g_{opt} being

$$g_{opt} = \frac{S_{xz}/\bar{X}\,\bar{Z}}{S_x^2/\bar{X}^2} , \tag{15}$$

which is the population regression coefficient of $z_i/\bar{Z}$ over $x_i/\bar{X}$, $i = 1,\ldots,N$. Therefore, if n is large and computational cost is not a problem, we propose the optimal estimator $v_{\hat{g}}$, where $\hat{g}$ is a sample analogue of g_{opt}. For estimation of the population mean Srivastava (1967) suggested a similar estimator $\bar{y}(\bar{X}/\bar{x})^g$. Das and Tripathi (1978) considered estimators of a similar type for the finite population variance of y. In both papers the optimal g for minimizing the asymptotic mean square errors were found.

In practice we may not want to compute $\hat{g}$ and will choose the variance estimate among v_0, v_1 and v_2. Since the leading terms in $\mathrm{var}(v_g)$ are quadratic in g, we conclude that, among the three, v_0 is the best if $g_{opt} > 0.5$, v_1 the best if $0.5 < g_{opt} < 1.5$ and v_2 the best if $g_{opt} > 1.5$. To further relate our estimator v_g to the usual ratio estimator and the more general estimator proposed by Srivastava (1967), we approximate $y_i - rx_i$ by $y_i - RX_i = e_i$ and the variance estimation problem is now reduced to estimating the population mean $\bar{D} = N^{-1}(e_1^2 +\ldots+ e_N^2)$ by the sample mean $v_0 = n^{-1}(e_1^2 +\ldots+ e_n^2)$, or by the ratio estimator $v_1 = v_0\bar{X}/\bar{x}$, or by the unfamiliar $v_2 = v_0(\bar{X}/\bar{x})^2$. The usual comparison of the sample mean and the ratio estimator (Cochran, 1977, §6.6) and the more general comparison in Srivastava (1967) would suggest that v_0 is less efficient than v_1 and v_2 if the population regression coefficient of $e_i^2/\bar{D}$ over $x_i/\bar{X}$ is greater than $\frac{1}{2}$. Because of the error introduced in the approximation $y_i - rx_i \simeq e_i$, the exact condition (15) involves the less intuitive z_i rather than e_i^2. Some readers may prefer the preceding interpretation in terms of the regression of residual square

e_i^2 over x_i. Under model (4) $\mathcal{E}S_{xz} = \mathcal{E}S_{xe^2}$, where S_{xe^2} is the population covariance of x_i and e_i^2, so that z_i can indeed be replaced by e_i^2. Here $\mathcal{E}$ denotes expectation with respect to model.

To gain further insight, we now assume the superpopulation model (4) with variance proportional to x^t. For most of the computations involving model (4), it is important to note that, under (4),

$$R = \beta + O(N^{-1/2}),\ e_i = y_i - Rx_i = \varepsilon_i + O(N^{-1/2}) \quad ,$$

if $N^{-1}(x_1^t + \ldots + x_N^t)/\bar{X}^2$ is bounded. The g that minimizes $\mathcal{E}\{var(\nu_g)\}$ under (4) is, up to a term of order N^{-1},

$$g_* = \frac{(\Sigma_1^N x_i^{t+1} - \bar{X}\,\Sigma_1^N x_i^t)(\Sigma_1^N x_i)}{\Sigma_1^N (x_i - \bar{X})^2 (\Sigma_1^N x_i^t)} \quad . \tag{16}$$

We have $g_* = 0$ for $t = 0$ and $g_* = 1$ for $t = 1$, which is stated as a proposition.

Proposition 1. Under model (4) with $t = 0$ (or 1), ν_0 (or ν_1) is the optimal estimator of V among ν_g.

We shall point out that $\hat{\bar{Y}}_R$ is the best linear unbiased estimator of $\bar{Y}$ under model (4) with $t = 1$ (Brewer, 1963; Royall, 1970). Therefore the ratio-type estimator ν_1 for variance should be used in situations where the ratio estimator $\hat{\bar{Y}}_R$ is optimal for estimating the mean. On the other hand, $t = 0$ implies that e_i^2 and x_i are not correlated and the optimal variance estimator ν_0 does not incorporate information on $\bar{x}$. For $t > 1$ we have $\Sigma\, x_i^{t+1}\, \Sigma\, x_i > \Sigma\, x_i^t\, \Sigma\, x_i^2$, which implies $g_* > 1$ for $t > 1$. Its implication as to the choice of estimators is states as follows.

Proposition 2. Under model (4) with $t > 1$, the optimal $g_* > 1$ and ν_1, ν_2 are b th better han ν_0 for estimating V.

When $t \neq 0$ or 1, there seems to be no clear-cut comparison of ν_0, ν_1, ν_2. We now assume a distribution on x to facilitate such a comparison. The optimal g becomes

$$g_{**} = \frac{\mathrm{Cov}(x^t,x)E(x)}{E(x^t)\ \mathrm{var}(x)} , \tag{17}$$

where expectation is taken with respect to the distribution of x. When x has a gamma distribution with two parameters, $g_{**} = t$ irrespective of the values of the parameters. When x has a beta distribution on $[0,M]$, $M > 0$, with parameters p and q, $p,q > 0$, $g_{**} = t(p+q+1)/(p+q+t)$. Note that $g_{**} < t$ for $t > 1$ and $g_{**} > t$ for $t < 1$. In particular, when x has a uniform distribution, $g_{**} = 3t/(t+2)$. When x has a lognormal distribution with parameters δ and γ, i.e., $\gamma + \delta \log x$ is standard normal, then $g_{**} = (w^{2t}-1)/(w^2-1)$, where $w = exp(1/2\delta^2)$. Note that $g_{**} > t$ for $t > 1$ and $g_{**} < t$ for $t < 1$ in this case. Some selected g_{**} values as function of t are given in Table 1.

Table 1. Optimal g_{**} as function of t in model (4)

	values of t						
distribution of x	0	0.5	1.0	1.5	2.0	2.5	3.0
gamma	0	0.5	1.0	1.5	2.0	2.5	3.0
uniform $p+q=2$	0	0.6	1.0	1.29	1.5	1.67	1.8
beta $p+q=5$	0	0.55	1.0	1.38	1.71	2.0	2.25
lognormal $\delta=2$	0	0.47	1.0	1.60	2.28	3.06	3.93
lognormal $\delta=1$	0	0.38	1.0	2.03	3.72	6.51	11.11

Rao and Rao (1971) compared the stability of ν_0 and ν_2 under an infinite population regression model and gamma distribution on x. They reported that ν_0 is more stable than ν_2 for $t = 0$ or 1 and ν_2 is more stable than ν_0 for $t = 2$. Their results are consistent with ours.

3. Bias of ν_g

3.1. Asymptotic expansions

Multiplying (7) to (10) and using (8) and (9) to collect terms of order n^{-3}, we obtain

$$\nu_g = \frac{1-f}{n-1} \{\frac{1}{n} \sum_1^n e_i^2 - 2 \frac{\bar{e}}{\bar{X}} \frac{1}{n} \sum_1^n x_i e_i + (2+2g) \frac{\bar{e}(\delta\bar{x})}{\bar{X}} \frac{1}{N} \sum_1^N x_i e_i +$$

$$\frac{\bar{e}^2}{\bar{X}^2} \frac{1}{N} \sum_1^N x_i^2 - g(\delta\bar{x}) \frac{1}{n} \sum_1^n e_i^2 + \frac{g(g+1)}{2} (\delta\bar{x})^2 \frac{1}{N} \sum_1^N e_i^2\} + O(\frac{1}{n^3}) .$$

From Theorem 2.3 of Cochran (1977), the bias of ν_g is

$$\text{bias}(\nu_g) = \frac{(1-f)^2}{n(n-1)} \frac{1}{N-1} \frac{1}{N\bar{X}^2} \{\frac{g^2+g+2}{2} \sum_1^N e_i^2 \sum_1^N x_i^2 + 2(g+1)(\sum_1^N x_i e_i)^2$$

$$- \frac{(g+1)(g-2)}{2} N\bar{X}^2 \sum_1^N e_i^2 - (g+2) \sum_1^N x_i \sum_1^N x_i e_i^2\} + O(\frac{1}{n^3}) . \quad (18)$$

Since the bias square of ν_g is of $O(n^{-4})$, smaller than $\text{var}(\nu_g)$, one would not choose ν_g based on its bias for large samples. But in practice the variance estimators can be seriously biased in small samples (Rao, 1968). The bias can then be reduced by subtracting the sample analogue of expression (18) from the estimate.

3.2. Least biased ν_g and comparison of ν_0, ν_1, ν_2

Without further assumptions on the population, there is no clear-cut comparison on the biases of ν_0, ν_1 and ν_2. By assuming model (4), we have

$$\mathcal{E}\{\mathrm{bias}(\nu_g)\} = \frac{(1-f)^2}{n(n-1)} \frac{1}{N-1} \frac{\sigma^2}{N\bar{X}^2} \{\frac{g^2+g+2}{2} \sum_1^N x_i^t \sum_1^N x_i^2$$

$$- \frac{(g+1)(g-2)}{2} N\bar{X}^2 \sum_1^N x_i^t - (g+2) \sum_1^N x_i \sum_1^N x_i^{t+1}\} + O(\frac{1}{n^3}) \quad . \qquad (19)$$

Based on (19), the least biased ν_g for $t = 0$ or 1 are easily found.

<u>Proposition 3</u>. Under model (4) with $t = 0$, $\nu_{-1/2}$ is the least biased estimator of V among ν_g and the bias is of order $O(n^{-2})$. Under model (4) with $t = 1$, ν_2 and ν_{-1} are the least biased estimators of V among ν_g and are the only ones among ν_g with bias of order $O(n^{-3})$.

For $t \neq 0$ or 1 expression (19) does not have a very nice form. To gain further insight we assume that x has a gamma distribution with shape parameter $\alpha > 0$. Then, from (19), the bias of ν_g is $C\{(g^2+g+2)/2 - (g+2)t\}$, where C is a positive constant independent of g. From this we conclude that, among ν_0, ν_1, ν_2, (i) ν_0 is the least biased for $t > 1.5$ or $t < 0.6$, (ii) ν_1 the least biased for $0.6 < t < 6/7$ and (iii) ν_2 the least biased for $6/7 < t < 1.5$.

Rao and Rao (1971) compared the biases of ν_0 and ν_2 under an infinite population regression model and gamma distribution on x. They reported that ν_2 is less biased than ν_0 for $0 < t < 1.5$ and ν_0

is less biased than ν_2 for $t = 2$. Their results are again in good accord with ours.

Of course the definition of "bias" here is in terms of estimating the approximate variance V of the ratio estimator, not its true mean square error. From a Monte Carlo study by Rao (1968) on some natural populations, the percent underestimate in V of the true mean square error is between 10 and 15 percent for $n = 4,6,8,12$. For a summary of results see Cochran (1977, p. 164). Therefore, if V underestimates the true mean square error, we shall prefer a variance estimator with a small amount of positive "bias". The only impact of this observation on the previous comparison is for $0 < t < 1.0$, where the biases of ν_0, ν_1 can be either positive or negative. Since $\text{bias}(\nu_2) > \text{bias}(\nu_1)$ or bias (ν_0) for $0 < t < 1.0$, ν_2 may be preferred on this ground.

The research was supported by a grant from the National Science Foundation. The comments of the referee are appreciated.

REFERENCES

Brewer, K. W. R. (1963). Ratio estimation in finite populations: some results deducible from the assumption of an underlying stochastic process. Australian Journ. Statist. 5, 93-105.

Cochran, W. G. (1977). Sampling Techniques, 3rd edition, New York: Wiley.

Das, A. K. and Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance. Sankyā, Ser. C 40, 139-148.

David, I. P. and Sukhatme, B. V. (1974). On the bias and mean square error of the ratio estimator. J. Am. Statist. Assoc. 69, 464-466.

Hájek, J. (1958). On the theory of ratio estimates. Aplikace Mathematiky 3, 384-398.

Rao, J. N. K. (1968). Some small sample results in ratio and regression estimation. J. Indian Statist. Assoc. 6, 160-168.

Rao, P. S. R. S. and Rao, J. N. K. (1971). Small sample results for ratio estimators. Biometrika 58, 625-630.

Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. Biometrika 57, 377-387.

Royall, R. M. and Cumberland, W. G. (1978). An empirical study of prediction theory in finite population sampling: simple random sampling and the ratio estimator. In Survey Sampling and Measurements, ed. N. K. Namboodiri, pp. 293-309. New York: Academic Press.

Royall, R. M. and Eberhardt, K. R. (1975). Variance estimates for the ratio estimator, Sankyā, Ser. C 37, 43-52.

Scott, A. J. and Wu, C. F. (1981). On the asymptotic distribution of ratio and regression estimators. J. Am. Statist. Assoc. 76, 98-102.

Srivastava, S. K. (1967). An estimator using auxiliary information in sample surveys. Calcutta Statist. Assoc. Bulletin 16, 121-132.

CFW/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2267	2. GOVT ACCESSION NO. AD-A110 474	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ESTIMATION OF VARIANCE OF THE RATIO ESTIMATOR		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Chien-Fu Wu		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041 MCS-7901846
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics & Probability
11. CONTROLLING OFFICE NAME AND ADDRESS (see Item 18 below)		12. REPORT DATE August 1981 13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

and

National Science Foundation
Washington, DC 20550

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

Asymptotic expansion, Bias, Optimal variance estimator, Ratio estimator, Superpopulation, Variance

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

A general class of estimators of the variance of the ratio estimator is considered, which includes two standard estimators v_0 and v_2 and approximates another estimator v_H suggested by Royall and Eberhardt (1975). Asymptotic expansions for the variances and biases of the proposed estimators are obtained. Based on this we obtain optimal variance estimator in the class and compare the relative merits of three estimators v_0, v_1 and v_2 without any model assumption. Under a simple regression model a more definite comparison of v_0, v_1 and v_2 is made in terms of variance and bias.

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DAT
FILM