Journal of Official Statistics, Vol. 15, No. 1, 1999, pp. 25–38

A New Bias-reducing Modification of the Finite Population Ratio Estimator and a Comparison Among Proposed Alternatives

Abeidi S. Mussa [1]

A modification of the standard ratio estimator is proposed which, like Chakrabarty's (1979) estimators, is in the family of Srivastava's (1980) class of generalized estimators of a finite-population mean or total. The proposed estimator is contrasted with the sample mean, the standard ratio estimator, and Chakrabarty's modified estimators in terms of bias and variance properties across a variety of finite-population characteristics. The new estimator performed comparably to Chakrabarty's estimators overall, and performed better than Chakrabarty's estimators for several sets of finite-population characteristics.

Key words: Unbiased estimator; ratio estimator; mean squared error; efficiency.

1. Introduction

The use of auxiliary information in estimating the finite population total or mean is common. It is widely used at the design stage, either for stratification or for selecting units for inclusion in the sample with probability proportional to the size of the auxiliary variable. Auxiliary information is also frequently employed at the stage of estimation in the form of ratio, regression, product and difference estimators, because of their simplicity and efficiency. Such estimators take advantage of the correlation between the characteristics of interest and the auxiliary variable. These estimators, under certain conditions, give more reliable estimates of the population value under study than those based on simple averages (Sukhatme and Sukhatme 1974). These estimators use statistics expressed in arithmetic means of several sample pairs (the variable of interest and its auxiliary) drawn from a population. The population mean of the auxiliary variable is generally assumed to be known.

1.1. Bias of ratio estimator

Ratio estimators are often used to estimate the population total, Y, or the population mean, $\overline{Y}$. Given a random sample of size n, the ratio estimator y_x is given by $\overline{y}_r = r\overline{X}$ where r is $\overline{y}/\overline{x}$, the ratio of the sample means of the two variables, and $\overline{X}$ is the population mean of the auxiliary variable. A well-known defect of the estimator is the fact that it is usually biased. However, by taking advantage of the correlation between y and x, $\rho_{yx'}$, the ratio estimator can provide a more reliable estimate of the population value than that based on a simple

[1] Department of Statistics, University of Dar es Salaam, PO Box 35047, Dar es Salaam, Tanzania. E-mail Fass@udsm.ac.tz.

arithmetic mean. The ratio r is a biased estimate of $R = \overline{Y}/\overline{X}$, but Cochran (1977) has shown that if the coefficient of variation of $\bar{x}$ is less than 0.1, then the bias is small relative to the standard error.

1.2. Efficiency of the standard ratio estimator

Following Cochran (1977), the estimated variance for $\bar{y}_r$ can be written as

$$v(\bar{y}_r) = \frac{1-f}{n}\left(s_y^2 + r^2 s_x^2 - 2rs_{yx}\right)$$

where $f = n/N$, s_y is the sample variance of $y_1, y_2, \ldots, y_n$, and s_{yx} is the sample covariance between y and x, which is given as

$$s_{yx} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

The relative efficiency of $\bar{y}_r$ compared to $\bar{y}$ is given as

$$\omega = \frac{s_y^2}{s_y^2 + r^2 s_x^2 - 2rs_{yx}} = \frac{1}{1 + \dfrac{C_x^2}{C_y^2} - 2\rho_{yx}\dfrac{C_x}{C_y}}$$

where $C_x = \dfrac{s_x}{\bar{x}}$ and $C_y = \dfrac{s_y}{\bar{y}}$

It follows that in relatively large samples the ratio estimate will be more efficient than the corresponding sample estimate based on the simple arithmetic mean if

$$\rho_{yx}\frac{C_x}{C_y} > \frac{1}{2}$$

1.3. Some ratio estimators

It has been observed in Section 1.1. that often in practical surveys, the bias is a small fraction of the standard deviation of the estimate and can be neglected. There is, however, an important class of survey designs in which this bias may become considerable. With small samples drawn from each of a large number, L, of strata, then, if the bias in each stratum has the same sign, which according to Goodman and Hartley (1958) does often happen, the bias in the estimate of the population total will be approximately L times that of an individual stratum, while the standard deviation is only multiplied by $\sqrt{L}$ (Cochran 1977). Therefore, the mean squared error will be of the order of magnitude L^2. Had it been that the estimate in each stratum total was unbiased, the order of magnitude would have been L. It is in these situations that the use of an alternative estimator with lower bias than the standard ratio estimator can be advantageous.

2. Proposed Modifications of the Standard Ratio Estimator

This section reviews several proposed methods in the literature for estimating the mean of a finite population with a modification of the standard ratio estimator and considers the motivation for each. In addition, a new estimator is proposed which is related to some of the earlier methods.

2.1. *Unbiased estimation*

As shown by Lahiri (1951), the customary ratio estimate r is unbiased if the sample is drawn with probability proportional to $\sum_{i=1}^{n} x$. Along the same lines Midzuno (1951) suggested a method of drawing the first member of the sample with probability proportional to the value of x for the first individual and drawing remaining $(n-1)$ members of the sample with equal probability.

Hartley and Ross (1954) proposed an alternative estimator that also considered the sampling fraction. Mickey (1959) derived an estimator by a process of averaging the ratios $r_{(j)}$ obtained when the j^{th} pair (x_j, y_j) is omitted, which we now recognize as the jackknife method of estimation. Several other estimators have been developed that are approximately unbiased and that offer potential efficiency gains; we now turn our attention to a few of these.

2.2. *Quenouille's bias reduction*

Quenouille (1956) managed to reduce bias to order n^{-2}, that is $0(n^{-2})$, though the use of half-sample ratio estimators. Specifically, we let $r = \overline{y}/\overline{x}$ as before and consider analogous ratios from two random halves of the data, which we denote $r_j = \overline{y}_j/\overline{x}_j$ for $j = 1, 2$. Quenouille's estimator is then $t_Q = 2r - 1/2(r_1 + r_2)$. It might be thought that any reduction in bias would be achieved at the expense of a corresponding increase of variance, but Quenouille managed to show that any such increase in variance is of small order compared with the variance itself. The argument has been followed by Durbin (1959), who demonstrated that for the class of estimators considered by Quenouille, the variance is smaller than the variance of the standard ratio estimator, in spite of the fact that for sufficiently large sample sizes, Quenouille's estimator also has a smaller bias than the standard ratio estimator.

2.3. *Tin estimator*

Tin (1965) developed a modified estimator $t_{MOD} = r\{1 + \eta(C_{xy} - C_x^2)\}$, where $\eta = (1/n - 1/N)$, $C_{xy} = s_{xy}/\overline{X}$, and as before $r = \overline{y}/\overline{x}$ and $C_x = s_x/\overline{X}$. He compared this estimator with several others in terms of bias, efficiency, and convergence to normality. Specifically, Tin considered the standard ratio estimator, Quenouille's (1956) estimator, and a modification proposed by Beale (1962), namely $t_Q = r(1 + \eta C_{xy})/1 + \eta C_x^2$. Tin concluded that t_{MOD} was the most efficient in certain cases such as when X and Y had a bivariate normal distribution.

2.4. *Chakrabarty's estimators*

Chakrabarty (1979) proposed two ratio estimators of the form $\overline{y}_{c1} = (1 - W)\overline{y} + W\overline{y}_r$ and $\overline{y}_{c2} = (1 - W)\overline{y} + Wt_Q\overline{X}$ for $W \geq 0$ which are in the family of Srivastava's (1980) class of generalized estimators of a finite-population mean or total. He compared the two estimators with $\overline{y}_r$ and with Srivastava's (1967) estimator of the form

$\bar{y}_s = \bar{y}\{\bar{x}/\bar{X}\}^{\alpha}$, which is discussed briefly in the next section. He concluded that the three alternative estimators were preferable to both $\bar{y}$ and $\bar{y}_r$ and that their efficiencies were the same in large samples and were practically of the same order in small samples. Also, he noted that computationally $\bar{y}_{c1}$ was simplest and that the bias of $\bar{y}_{c2}$ was least.

2.5. Other work

Srivastava (1967) considered an estimator of the form

$$\bar{y}_s = \bar{y}\left\{\frac{\bar{x}}{\bar{X}}\right\}^{\alpha}$$

while Srivastava, Jhajj and Sharma (1986) proposed estimators of the following kinds

$$\bar{y}_c = (1 - W)\bar{y} + W\left\{\frac{\bar{X}}{\bar{x}}\right\}, \qquad \bar{y}_v = (1 - a)\bar{y} + a\left\{\frac{\bar{x}}{\bar{X}}\right\}$$

and $\bar{y}_w = \bar{y}\dfrac{\bar{X}}{A\bar{x} + (1 - A)\bar{x}}$

The respective values of α, W, a and A that optimize these expressions are $-\alpha = W = -a = A = (\rho_{yx}C_y)/C_x$. Reddy (1978) has shown that the value is fairly stable in repeated surveys. In a similar approach, Sisodia and Dwivedi (1981) proposed estimators of the form

$$\bar{y}_{SD} = (1 - b)\bar{y} + b\bar{y}\left\{\frac{\bar{X}}{\bar{x}}\right\}^{p}$$

which reduces to $\bar{y}_{c1}$ if $p = 1$ and to $\bar{y}_s$ if $b = 1$. Also, if both $b = 1$ and $p = 1$ then $\bar{y}_{SD}$ reduces to the standard ratio estimator $\bar{y}_r$.

Royall and Cumberland (1981) considered a weighted least squares unbiased predictor model. However, they pointed out that in a number of examples of real data where a ratio estimate might be used, the estimate and its estimated variance can be badly biased unless the sample is balanced with respect to the X variable, in particular when $\bar{x}$ and $\bar{X}$ are not close.

Robinson (1987) obtained a conditional bias adjusted ratio estimate. The proposed method of adjustment for the bias of the estimate and its variance was based only on the assumption of simple random sampling, which uses an approximate conditional distribution of the estimate given the mean of the auxiliary variable. The condition is on the observed number in each (random) stratum and simply uses the usual estimate based on stratification or on this observed number and the observed mean in each stratum. The choice of the number of strata is arbitrary. Large number implies reduction in bias but also might reduce the efficiency of the estimator.

2.6. A new estimator

It has been observed that Quenouille's estimator performed relatively well in Tin's (1965) comparison and Durbin (1959) showed that it is more efficient than the standard ratio estimator for an auxiliary variable that has either a normal or gamma distribution. The Chakrabarty (1979) estimator made use of Quenouille's t_Q; here we consider a modification of this approach, which is easier to compute, that uses Tin's t_{MOD} in place of t_Q. That is, the new modified estimator of the form $\bar{y}_{MOD} = (1 - W)\bar{y} + W\, t_{MOD}\bar{X}$. In the next section, we discuss bias and variance properties of $\bar{y}_{MOD}$ along with Chakrabarty's $\bar{y}_{c1}$ and $\bar{y}_{c2}$ as well as the ratio estimator $\bar{y}_r$.

3. Asymptotic Bias and Variance Properties of Estimators

In this section we summarize key results of derivations that are postponed to Appendix 1. It is obvious that $\bar{y}_{c1}$, $\bar{y}_{c2}$ and $\bar{y}_{MOD}$ are consistent, but in general biased like the estimator $\bar{y}_r$. As noted in David and Sukhatme (1974), we can assume the sampled values are positive without loss of generality. The bias for $\bar{y}_r$ is given as

$$Bias\,(\bar{y}_r) = \frac{\bar{Y}}{n}(C_x^2 - C_{xy}) + O(n^{-2}), \text{ and the bias for } \bar{y}_{c1} \text{ is}$$

$$Bias\,(\bar{y}_{c1}) = \frac{W\bar{Y}}{n}(C_x^2 - C_{xy}) + O(n^{-2})$$

From Appendix 1 the bias of $\bar{y}_{MOD}$ is given as

$$Bias\,(\bar{y}_{MOD}) = \frac{W\bar{Y}}{n}\frac{n}{N}(C2_x - C_{xy}) + O(n^{-2}) = \frac{n}{N}\,Bias\,(\bar{y}_{c1})$$

The asymptotic bias of $\bar{y}_{c2}$ has been shown to be of a lower magnitude than that of $\bar{y}_r$, $\bar{y}_{c1}$ and $\bar{y}_{MOD}$. However the bias of $\bar{y}_{MOD}$ is smaller than those of $\bar{y}_r$ and $\bar{y}_{c1}$ and will be closer to that of $\bar{y}_{c2}$ when $N \gg n$. Also when $(C_x^2 - C_{xy}) = 0$, the regression of y on x passes though the origin, and all the estimators become unbiased to the term of order n^{-2}. Also worth noting is that the bias of $\bar{y}_r$ from Hartley and Ross (1954), the bias of $\bar{y}_{c1}$ becomes negligible in relation to the standard error of $\bar{y}_r$. This also applies to the bias of $\bar{y}_{MOD}$, which is a fraction n/N that of $\bar{y}_{c1}$.

In defining the variance of estimators $\bar{y}_{c1}$ and $\bar{y}_{c2}$, Chakrabarty (1979) showed that the variances are identical and, omitting the finite correction factor, are given by

$$V(\bar{y}_{c1}) = V(\bar{y}_{c2}) = \frac{S_y}{n}(1 + WK(WK - 2\rho_{yx}))$$

where $K = C_x/C_y$. The value of W which minimizes this variance is $W_{opt} = \rho_{yx}/K$, with the minimum variance given by $V_{min} = (S_y/n)(1 - \rho_{yx}^2)$ which is equal to the variance of the linear regression estimator up to terms of $o(n^{-2})$. Using a similar method suggested by Chakrabarty (1979) the variance is

$$V(\bar{y}_{MOD}) = \frac{S_y}{n}\left(1 + WK(WK - 2\rho_{yx} + \frac{2n}{N}(K - \rho_{yx})\right)$$

which is slightly larger than that of $\bar{y}_{c1}$ and $\bar{y}_{c2}$. However for large n it has been seen that $\rho_{yx}(C_y/C_x) > \frac{1}{2}$ which implies that $K < 2\rho_{yx}$, and hence the factor $2n/N(K - \rho_{yx})$ is always less than $(2n/N)\rho_{yx}$ and hence relatively small. Again for $N \gg n$ the term disappears. Hence, the variance of $\bar{y}_{MOD}$ is roughly equivalent to those of $\bar{y}_{c1}$ and $\bar{y}_{c2}$, and thus for $N \gg n$ their efficiencies will be roughly equivalent as well. Ordinarily, one would prefer a method that has advantages both in terms of bias and variance, and as shown here, $\bar{y}_{c2}$ has both smaller bias and variance than $\bar{y}_{MOD}$.

However, the advantages are slight, and the simpler computation involved in $\bar{y}_{MOD}$ might be viewed as an advantage in some situations.

Table 1 shows the efficiencies of $\bar{y}_{c1}$, $\bar{y}_{c2}$, and $\bar{y}_{MOD}$ relative to $\bar{y}$, and $\bar{y}_r$ for various values of ρ, K, and W. In the table E_1 reflects (in percentage terms) $V(\bar{y})/V(\bar{y}_j)$ and E_2 reflects $V(\bar{y}_r)/V(\bar{y}_j)$ for $j = c1$, $c2$ and MOD; thus higher values favour $\bar{y}_{c1}$, $\bar{y}_{c2}$, and $\bar{y}_{MOD}$. When X and Y have comparable coefficients of variation, i.e., when $K = 1$, the Chakrabarty and modified estimators offer better efficiency when correlations are low. There is a tradeoff in efficiency between $W = 0.25$ and $W = 0.5$ when $K = 1$, with the standard ratio estimator's advantage for large correlations more pronounced when $W = 0.25$. The modified ratio estimators with $W = 0.25$ generally offer efficiency improvements over the sample mean. When $K \neq 1$, efficiency gains are more prominent when the coefficient of variation X is larger than that of Y, i.e., when $K > 1$. The fact that K is not known in practice means that it has to be estimated in the formula for the variance of $\bar{y}_{c1}, \bar{y}_{c2}$, and $\bar{y}_{MOD}$. Thus, the efficiencies seen in practice may differ from those shown in Table 1. The asymptotic variances of the estimators $\bar{y}_{c1}$, $\bar{y}_{c2}$, and $\bar{y}_{MOD}$ with optimum value of $W = \rho_{yx}/K$ are equal to the asymptotic variance of the linear regression estimator, $\bar{y}_{1r}$. Therefore these estimators are asymptotically no more efficient than $\bar{y}_{1r}$ with constant weights ($W = 1/4$ or $1/2$). However, the estimator $\bar{y}_{1r}$ suffers appreciably in the ratio of bias to standard error when the relationship is not linear. Cochran (1977) has shown that the bias in $\bar{y}_{1r}$ is of the order n^{-1}, as compared to biases of order n^{-2} for $\bar{y}_{c2}$ and $\bar{y}_{MOD}$. Thus in situations where freedom from bias is important, $\bar{y}_{c2}$ and $\bar{y}_{MOD}$ may be preferable to $\bar{y}_{1r}$.

4. Bias and Variance Under a Linear Model

We also investigated the properties of the estimators under the following linear model:

$$y_i = \alpha + \beta x_i + u_i; \quad \beta > 0$$

for a sample of pairs (x_i, y_i) for $i = 1, 2, \ldots, n$.

The expectation of the error component u_i conditional on x_i, $E(u_i/x_i)$, is zero, and also $E(u_i, u_j/x_i, x_j) = 0$, the covariance of u_i and u_j given the values (x_i, x_j). The variance $v(u_i/x_i) = n\,\delta$ (δ is a constant of order n^{-1}), the variantes x_i/n have a gamma distribution with parameter $m = nh$. This model or slight variation on it has been used by Durbin (1959) and Rao and Webster (1966) to investigate bias in the setting where a ratio estimator or another alternative estimator is used. Chakrabarty (1979) showed that for any sample size n, the biases of $\bar{y}_{c1}$ and $\bar{y}_{c2}$ are, respectively,

$$Bias\,(\bar{y}_{c1}) = \alpha W/(m-1) \text{ and } Bias\,(\bar{y}_{c2}) = -2\alpha W/\{(m-1)\,(m-2)\}$$

The bias of $\bar{y}_r$ can be found by substituting $W = 1$ in the bias of $\bar{y}_{c1}$. In a similar way the bias of estimator $\bar{y}_{MOD}$ is derived as $Bias\,(\bar{y}_{MOD}) = \alpha W/\{(m-1)(m+1)\}$.

It can be seen that the biases of $\bar{y}_{c2}$ and $\bar{y}_{MOD}$ are of the order n^{-2}, while those of $\bar{y}_r$ and $\bar{y}_{c1}$ are of the order n^{-1}. The bias of $\bar{y}_{c1}$ is less than the bias of $\bar{y}_r$ if $W < 1$. For the special

Table 1a. Efficiencies for selected values of ρ and K with W = 0.25

	$K = 0.5$		$K = 1.0$		$K = 1.5$		$K = 2.0$	
ρ	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2
0.1	101	116	99	178	94	277	89	400
0.2	104	109	104	166	101	268	95	400
0.3	106	101	110	153	109	257	105	400
0.4	109	93	116	139	119	244	118	400
0.5	112	84	123	123	131	229	133	400
0.6	116	75	131	105	145	210	154	400
0.7	119	65	140	84	162	187	182	400
0.8	123	55	150	63	185	157	222	400
0.9	126	44	163	33	215	118	285	400

Table 1b. Efficiencies for selected values of ρ and K with W = 0.50

	$K = 0.5$		$K = 1.0$		$K = 1.5$		$K = 2.0$	
ρ	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2
0.1	99	157	71	114	87	209	56	256
0.2	104	152	79	109	95	210	62	262
0.3	110	147	90	104	105	211	71	271
0.4	116	141	104	99	117	213	83	283
0.5	123	133	123	92	133	215	100	300
0.6	131	123	151	85	153	219	125	325
0.7	140	109	195	77	182	224	167	367
0.8	150	89	276	68	222	234	250	450
0.9	163	57	471	57	286	259	500	700

E_1 = Efficiency of $\overline{y}_{c1}$, $\overline{y}_{c2}$, and $\overline{y}_{MOD}$ relative to $\overline{y}$ for $N \gg n$.
E_2 = Efficiency of $\overline{y}_{c1}$, $\overline{y}_{c2}$, and $\overline{y}_{MOD}$ relative to $\overline{y}_r$ for $N \gg n$.
$K = C_x/C_y$ where $C_x = \frac{S_x}{\overline{x}}$ and $C_y = \frac{S_y}{\overline{y}}$.

case of linear regression through the origin, that is when $\alpha = 0$, the estimators $\overline{y}_r, \overline{y}_{c1}, \overline{y}_{c2}$ and $\overline{y}_{MOD}$ are unbiased.

Applying the computations of Rao and Webster (1966), the variance of $\overline{y}_{c1}$ was found to be

$$v(\overline{y}_{c1}) = \frac{W^2 m}{(m-1)^2(m-2)}\alpha^2 + (1-W)^2 m\beta + \left[\frac{W^2m^2}{(m-1)(m-2)} + \frac{W(1-W)(m+1)}{(m-1)} + (1-W)\right]\delta + \frac{2W(1-W)m}{(m-1)}\alpha\beta$$

Putting $W = 1$ and $W = 0$ to the values of $v(\overline{y}_{c1})$ will give $v(\overline{y}_r)$ and $v(\overline{y})$, respectively.

Similarly the variance of $\overline{y}_{c2}$ was given as

$$v(\overline{y}_{c2}) = \frac{W^2m^2(m^2-6m+17)\alpha^2}{(m-1)^2(m-2)^2(m-4)} - \frac{2W(1-W)m(m-3)\alpha\beta}{(m-1)(m-2)} + (1-W)^2m\beta^2 + \left[(1-W)^2 + \frac{W^2(m^2-7m+18)m^2}{(m-1)(m-2)^2(m-4)} + \frac{2W(1-W)m(m-3)}{(m-1)(m-2)}\right]\delta$$

The variance of $\overline{y}_{MOD}$ can be derived in a similar manner. Hence

$$v(\overline{y}_{MOD}) = (1-W)^2\beta^2 m \frac{W^2m^3}{(m-1)(m-2)(m+1)(m+2)(m+3)} + \left[\frac{m^3+4m^2+8m-1)}{(m-1)(m+1)} + \frac{(n+1)}{(n-1)}\right]\alpha^2 + \left[(1-W)^2 + \frac{2(1-W)Wm^2}{(m-1)(m+1)} + \frac{W^2m^3}{(m-1)^2(m+1)(m+2)(m+3)}\left((m^2+4m+1) + \frac{(n+1)}{(n-1)}\right)\right]\delta - \frac{2(1-W)W\alpha\beta m^2}{(m-1)(m+1)}$$

It is noted (Chakrabarty 1979) that in terms of the model, $\alpha = \overline{Y}[(K-\rho_{yx})/K]$, $\beta = \overline{Y}[\rho_{yx}/Km]$, $\delta = \overline{Y}^2[(1[\rho_{yx}^2)/Km^2)]$ and $K = C_x/C_y$. The efficiencies of these estimators relative to that of $\overline{y}$ are given by

$$E(estimator) = \frac{v(\overline{y})}{[MSE(estimator)]^{1/2}}$$

Table 2 reports efficiencies for various values of W, ρ, m, and K. The efficiency of the ratio estimator to the sample mean depends strongly on ρ. The alternative estimators exhibit similar efficiencies across various values of ρ, with the impact on efficiency being somewhat less dramatic when $W = 0.25$ as compared to when $W = 0.50$.

Table 3 reports the percentage of the square root of mean squared error accounted for by absolute bias. With the standard ratio estimator, a substantial proportion of overall error is accounted for by bias in a number of scenarios. With the alternative estimators, especially $\overline{y}_{c2}$ and $\overline{y}_{MOD}$, a much more modest proportion of overall error is accounted for by bias.

5. Discussion

The estimators $\overline{y}_{c1}$, $\overline{y}_{c2}$ and $\overline{y}_{MOD}$ have comparatively smaller bias than the customary ratio

Table 2a. Efficiencies with $W = 0.25$ for selected values of ρ, m and K

		$\rho = 0.2$				$\rho = 0.8$			
m	*K*	E(R)	E(C1)	E(C2)	E(MOD)	E(R)	E(C1)	E(C2)	E(MOD)
8	0.25	68	94	100	101	89	102	107	109
	0.50	61	95	100	103	139	116	121	122
	1.00	37	93	96	101	160	142	148	150
	1.50	21	87	88	96	67	176	166	181
16	0.25	85	98	101	102	117	107	110	110
	0.50	77	100	103	103	178	119	122	122
	1.00	49	99	102	103	203	149	150	151
	1.50	29	94	96	99	90	181	180	183
20	0.25	89	99	102	102	123	108	110	110
	0.50	81	100	103	103	186	120	122	122
	1.00	51	100	102	103	212	149	150	151
	1.50	30	96	98	100	95	182	181	184
32	0.25	94	100	102	102	133	109	110	110
	0.50	86	102	103	103	199	121	122	123
	1.00	55	102	103	104	226	150	151	151
	1.50	33	98	99	100	103	183	183	184

Table 2b. Efficiencies with $W = 0.5$ for selected values of ρ, m and K

		$\rho = 0.2$				$\rho = 0.8$			
m	*K*	E(R)	E(C1)	E(C2)	E(MOD)	E(R)	E(C1)	E(C2)	E(MOD)
8	0.25	68	86	91	102	89	101	104	117
	0.50	61	85	89	101	139	129	135	147
	1.00	37	73	71	89	160	192	193	218
	1.50	21	55	50	69	67	198	167	242
16	0.25	85	95	100	103	117	112	117	120
	0.50	77	95	99	103	178	141	146	150
	1.00	49	84	87	93	203	208	213	221
	1.50	29	67	69	75	90	238	235	261
20	0.25	89	97	101	103	123	114	118	121
	0.50	81	97	100	103	186	143	147	150
	1.00	51	86	89	93	212	211	216	221
	1.50	30	69	71	76	95	245	245	265
32	0.25	94	99	102	103	133	118	120	122
	0.50	86	99	102	104	199	146	149	150
	1.00	55	90	92	94	226	216	218	222
	1.50	33	73	75	77	103	257	258	269

$$E(R) = \frac{v(\bar{y})}{MSE(\bar{y}_r)}, \quad E(C1) = \frac{v(\bar{y})}{MSE(\bar{y}_{c1})}, \quad E(C2) = \frac{v(\bar{y})}{MSE(\bar{y}_{c2})} \quad \text{and} \quad E(MOD) = \frac{v(\bar{y})}{MSE(\bar{y}_{MOD})}$$

Table 3a. $|Bias|/(MSE)^{1/2}$, *with* $W = 0.25$ *for selected values of* ρ*, m and K (%)*

		$\rho = 0.2$				$\rho = 0.8$			
m	*K*	B(R)	B(C1)	B(C2)	B(MOD)	B(R)	B(C1)	B(C2)	B(MOD)
8	0.25	1.66	0.49	0.16	0.05	20.95	5.62	1.91	0.64
	0.50	9.48	2.96	1.01	0.34	14.28	3.25	1.11	0.37
	1.00	19.74	7.80	2.64	0.90	10.22	2.43	0.81	0.27
	1.50	25.02	12.31	4.06	1.43	23.23	9.37	3.03	1.05
16	0.25	1.23	0.33	0.40	0.01	15.87	3.79	0.54	0.22
	0.50	7.03	2.00	0.29	0.11	10.66	2.18	0.31	0.13
	1.00	14.90	5.31	0.77	0.31	7.59	1.62	0.23	0.09
	1.50	18.56	8.42	1.22	0.50	17.71	6.28	0.89	0.37
20	0.25	1.10	0.29	0.03	0.01	14.38	3.35	0.37	0.16
	0.50	6.34	1.77	0.20	0.08	9.63	1.93	0.21	0.09
	1.00	13.50	4.71	0.53	0.22	6.84	1.43	0.16	0.06
	1.50	16.85	7.48	0.84	0.36	16.06	5.55	0.61	0.26
32	0.25	0.88	0.22	0.01	0.00	11.58	2.61	0.17	0.07
	0.50	5.08	1.38	0.09	0.04	7.72	1.50	0.10	0.04
	1.00	10.86	3.68	0.25	0.11	5.48	1.11	0.07	0.03
	1.50	13.62	5.86	0.39	0.17	12.97	4.32	0.28	0.13

Table 3b. $|Bias|/(MSE)^{1/2}$, *with* $W = 0.50$ *for selected values of* ρ*, m and K (%)*

		$\rho = 0.2$				$\rho = 0.8$			
m	*K*	B(R)	B(C1)	B(C2)	B(MOD)	B(R)	B(C1)	B(C2)	B(MOD)
8	0.25	1.66	0.93	0.32	0.11	20.95	11.17	3.77	1.33
	0.50	9.48	5.59	1.90	0.67	14.28	6.89	2.35	0.81
	1.00	19.73	13.77	4.53	1.68	10.22	5.60	1.87	0.66
	1.50	24.27	19.46	6.19	2.42	23.23	19.87	6.09	2.44
16	0.25	1.23	0.65	0.09	0.03	15.87	7.77	1.13	0.47
	0.50	7.03	3.89	0.56	0.25	10.66	4.74	0.69	0.28
	1.00	14.90	9.78	1.42	0.60	7.59	3.84	0.55	0.23
	1.50	18.56	14.17	2.05	0.88	17.71	14.38	2.04	0.88
20	0.25	1.10	0.57	0.06	0.02	14.38	6.92	0.78	0.33
	0.50	6.34	3.47	0.39	0.17	9.63	4.21	0.47	0.20
	1.00	13.49	8.75	0.98	0.43	6.84	3.42	0.38	0.16
	1.50	16.85	12.73	1.43	0.63	16.06	12.90	1.43	0.63
32	0.25	0.88	0.45	0.03	0.01	11.58	5.44	0.36	0.16
	0.50	5.07	2.73	0.18	0.08	7.72	3.30	0.22	0.10
	1.00	10.86	6.91	0.46	0.21	5.48	2.67	0.17	0.08
	1.50	13.61	10.13	0.68	0.31	12.97	10.23	0.68	0.31

$$B(R) = \frac{|Bias\,(\bar{y}_r)|}{[MSE(y_r)]^{1/2}}, \quad B(C1) = \frac{|Bias\,(\bar{y}_{c1})|}{[MSE(y_{c1})]^{1/2}},$$

$$B(C2) = \frac{|Bias\,(\bar{y}_{c2})|}{[MSE(y_{c2})]^{1/2}} \quad \text{and} \quad B(MOD) = \frac{|Bias\,(\bar{y}_{MOD})|}{[MSE(y_{MOD})]^{1/2}}$$

estimator, $\overline{y}_r$, and also are more efficient for a wide range of values of ρ, K and m, and with a ratio of absolute value of bias to standard error less than 10%. For the estimator $\overline{y}_{MOD}$ the ratio is less than 1% for most values of ρ, K and m compared to estimators $\overline{y}_{c1}$ and $\overline{y}_{c2}$.

The criterion proposed in the earlier section may be somewhat arbitrary, however, depending on the features of application, this appears to be the case. For example when ρ, K and m are unknown in a realistic situation, a researcher should first obtain preliminary data to determine the strength of ρ and K for considerable values of m depending on financial and manpower resources. The decision, on the estimator to be used, is then based on information about ρ, K and m obtained thereof.

Appendix

For sufficiently large n we have

$$|\delta_{\overline{x}}| = \left|\frac{\overline{x} - \overline{X}}{\overline{X}}\right| < 1 \tag{A1.1}$$

which is not unreasonable in an actual sample survey where x's are usually positive or can be adjusted to become positive. Also in order to facilitate the asymptotic expansions the following result is needed (see Kendall and Stuart 1963, Tin 1965 and Wu 1982)

$$E(\overline{x} - \overline{X})^r(\overline{y} - \overline{Y})^2 = \begin{cases} O(n^{-1/2(r+s)} & , \quad r+s \text{ even} \\ O(n^{-1/2(r+s+1)}, & r+s \text{ odd} \end{cases}$$

Since all terms of order n^{-2} in the calculations will be ignored, then all expectations with combinations of $(r+s) \geq 4$ when $(r+s)$ is even or $(r+s+1) \geq 4$ when $(r+1)$ is odd will be included in the term of $O(n^{-2})$. Using (A1.1), it can be shown that

$$r = R\left(\frac{1+\delta_{\overline{y}}}{1+\delta_{\overline{x}}}\right) = R(1+\delta_{\overline{y}})(1+\delta_{\overline{x}})^{-1} \tag{A1.2}$$

where $R = \overline{Y}/\overline{X}$, the population ratio. Expanding the above by Taylor's series, and taking expectations on both sides, gives $E(r) = R + R/n(C_x^2 - C_{xy}) + O(n^{-2})$.

The bias of r is found as

$$\begin{aligned} Bias\,(r) &= E(r) - R \\ &= \frac{R}{n}(C_x^2 - C_{xy}) + O(n^{-2}) \end{aligned}$$

Therefore

$$\begin{aligned} Bias\,(\overline{y}_r) &= \overline{x}\,Bias\,(r) \\ &= \frac{R\overline{X}}{n}(C_x^2 - C_{xy}) + O(n^{-2}) \\ &= \frac{\overline{y}}{n}(C_x^2 - C_{xy}) + O(n^{-2}) \end{aligned}$$

Since $t_Q = 2\,r - 1/2(r_1 + r_2)$, r_1 and r_2 are independent, then $E(t_Q) = R + O(n^{-2})$ and

$Bias\,(t_Q) = E(t_Q) - R = 0 + O(n^{-2})$. Consequently the biases of $\overline{y}_{c1}$ and $\overline{y}_{c2}$ are, respectively

$$Bias\,(\overline{y}_{c1}) = W\,Bias\,(\overline{y}_r) = \frac{W\overline{Y}}{n}(C_x^2 - C_{xy}) + O(n^{-2}) \text{ and}$$

$$Bias\,(\overline{y}_{c2} = W\,\overline{X}\,Bias\,(t_Q) = 0 + O(n^{-2})$$

Likewise

$$t_{MOD} = r\big(1 + \eta\,(C_{xy} - C_x^2)\big)$$

$$= r\left(1 + \eta\,\frac{S_{xy}}{\overline{XY}} - \eta\,\frac{S_x^2}{\overline{X}^2}\right)$$

Substituting and expanding r from equation (A1.2) the above can be rewritten as

$$t_{MOD} = R\big\{1 + \delta_{\overline{y}} - \delta_{\overline{x}} + \delta_{\overline{x}}^2 - \delta_{\overline{x}}\delta_{\overline{y}} + \ldots\big\}\left\{1 + \eta\,\frac{S_{xy}}{\overline{X}\,\overline{Y}} - \eta\,\frac{S_x^2}{\overline{X}^2}\right\}$$

Expanding the above and taking expectations on both sides gives

$$E(t_{MOD}) = R\left(1 + \frac{S_x^2}{n\overline{x}^2} - \frac{S_{xy}}{n\overline{X}\,\overline{Y}} + \frac{\eta S_{xy}}{\overline{X}\,\overline{Y}} - \frac{\eta S_x^2}{\overline{x}^2} + O(n{-}^{-2})\right)$$

$$= R\left(1 + \frac{C_x^2}{n} - \frac{C_{xy}}{n} + \frac{C_{xy}}{n} - \frac{C_{xy}}{N} + \frac{C_x^2}{n} + \frac{C_x^2}{N} + O(n^{-2})\right)$$

$$= R + \frac{R}{N}\left(C_x^2 - C_{xy}\right) + O(n^{-2})$$

Therefore the bias of t_{MOD} is expressed as

$$Bias\,(t_{MOD}) = \frac{1}{N}\left(C_x^2 - C_{xy}\right) + O(n^{-2})$$

and that of $\overline{y}_{MOD}$ as

$$Bias\,(\overline{y}_{MOD}) = \frac{W\overline{X}R}{N}(C_x^2 - C_{xy}) + O(n^{-2})$$

$$= \frac{W\overline{Y}}{n}\frac{n}{N}\left(C_x^2 - C_{xy}\right) + O(n^{-2})$$

that is

$$Bias\,(\overline{y}_{MOD}) = \frac{n}{N}\;Bias\,(\overline{y}_{c1})$$

In determining the variance of estimators $\overline{y}_{c1}$ and $\overline{y}_{c2}$, Chakrabarty (1979) considered up to terms of $O(n^{-2})$ only. He therefore showed that the variances are identical and, omitting the finite correction factor, are given by

$$v(\overline{y}_{c1}) = v(\overline{y}_{c2}) = \frac{S_y}{n}\left(1 + W\,K(W\,K - 2\rho_{yx})\right)$$

where $K = C_x/C_y$ and that the value of W which minimizes this variance is $W_{opt} = \rho_{yx}/K$. The minimum variance was given as

$$v_{min} = \frac{S_y}{n}\left(1 - \rho_{yx}^2\right)$$

which is equal to the variance of the linear regression estimator up to terms of $O(n^{-2})$. The variance of $\overline{y}_{MOD}$ using a similar method suggested by Chakrabarty (1979) is therefore

$$v(\overline{y}_{MOD}) = \frac{S_y}{n}\left(1 + WK(WK - 2\rho_{yx} + \frac{2n}{N}(K - \rho_{yx}))\right)$$

which is slightly larger than that of $\overline{y}_{c1}$ and $\overline{y}_{c2}$.

Substituting $W = 1$ in (1.6) the variance of $\overline{y}_r$ is expressed as

$$v(\overline{y}_r) = \frac{S_y}{n}\left(1 + K(K - 2\rho_{yx})\right)$$

Thus the asymptotic efficiencies of $\overline{y}_{c1}$ ($\overline{y}_{c2}$ and $\overline{y}_{MOD}$) over $\overline{y}$ and $\overline{y}_r$ are respectively given by

$$E_1 = \frac{v(\overline{y})}{v(\overline{y}_{c1})} = \frac{1}{(1 + WK(WK - 2\rho_{yx}))} \text{ and}$$

$$E_2 = \frac{v(\overline{y}_r)}{v(\overline{y}_{c1})} = \frac{v(\overline{y}_r)}{v(\overline{y}_{c2})} = \frac{v(\overline{y}_r)}{v(\overline{y}_{MOD})} = \frac{\left(1 + K(K - 2\rho_{yx}\right)}{\left(1 + WK(WK - 2\rho_{yx})\right)}$$

The following is achieved as a result of the above:

$$E_1 \geq 1 \text{ if } W \leq 2\frac{\rho_{yx}}{K} \text{ and}$$

$$E_2 \geq 1 \text{ if } \frac{2(\rho_{yx} - K)}{K} \leq W \leq 1$$

6. References

Beale, E.M.L. (1962). Some Use of Computers in Operational Research. Industrielle Organization, 31, 27–28.

Chakrabarty, R.P. (1973). A Note on the Small Sample Theory of the Ratio Estimator in Certain Specified Populations. Journal of the Indian Society of Agricultural Statistics, XXV, 2, 49–57.

Chakrabarty, R.P. (1979). Some Ratio-type Estimators. Journal of the Indian Society of Agricultural Statistics, XXXI, 1, 49–62.

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

David, I.P. and Sukhatme, B.V. (1974). On the Bias and Mean Square Error of the Ratio Estimator. Journal of the American Statistical Association, 69, 464–466.

Durbin, J. (1959). A Note on the Application of Quenouille's Method of Bias Reduction to Estimation of Ratios. Biometrika, 46, 477–480.

Goodman, L.A. and Hartley, H.O. (1958). The Precision of Unbiased Ratio-type Estimators. Journal of the American Statistical Association, 491–508.

Hartley, H.O. and Ross, A. (1954). Unbiased ratio estimators. Nature, 174, 270–271.

Kendall, M.G. and Stuart, A. (1963). The Advanced Theory of Statistics, 1, (2nd edition). Charles Griffin: London.

Lahiri, D.BB. (1951). A Method for Sample Selection Providing Unbiased Ratio Estimates. Bulletin of the International Statistical Institute, 33, 2, 133–140.

Mickey, M.R. (1959). Some Finite Population Unbiased Ratio and Regression Estimators. Journal of the American Statistical Association, 54, 594–612.

Midzuno, H. (1951). On the Sampling System with Probability Proportionate to Sum of Sizes. Ann. Inst. Stat. Math., 2, 99–108.

Quenouille, M.H. (1956). Notes on Bias Estimation. Biometrika, 43, 353–360.

Rao, J.N.K. and Webster, J.T. (1966). On Two Methods of Bias Reduction in Estimation of Ratios. Biometrika, 53, 571–577.

Reddy, V.N. (1978). A Study on the Use of Prior Knowledge on Certain Population Parameters in Estimation. Sankhyā, 4, Series C, 29–37.

Robinson, J. (1987). Conditioning Ratio Estimates under Simple Random Sampling. Journal of the Americal Statistical Association, 82, 826–831.

Royall, R.M. and Cumberland, W.G. (1981). An Empirical Study of the Ratio Estimator and Estimators of its Variance. Journal of the American Statistical Association, 76, 66–88.

Sisodia, B.V.S. and Dwivedi, V.K. (1981). A Modified Ratio Estimator Using Coefficient of Variation of Auxiliary Variable. Journal of the Indian Society of Agricultural Statistics, 33, 13–18.

Srivastava, S.K. (1967). An Estimator Using Auxiliary Information in Sample Surveys. Calcutta Statistical Association Bulletin, 16, 121–132.

Srivastava, S.K. (1980) A Class of Estimators Using Information in Sample Survey. Canadian Journal of Statistics, 8, 252–254.

Srivastava, S.K., Jhajj, H.S., and Sharma, M.K. (1986). Comparison of Some Estimators of K and B in Finite Populations. Journal of the Indian Society of Agricultural Statistics, Vol. XXXVIII, No. 2, 230–236.

Tin, M. (1965). Comparison of Some Ratio Estimates. Journal of the American Statistical Association, 60, 294–307.

Wu, C.F. (1982). Estimation of Variance of the Ratio Estimator. Biometrika, 69, 183–189.

Received February 1995
Revised September 1998