Statistical Papers 42, 505–515 (2001)

Statistical Papers

On constructing almost unbiased estimators of finite population mean using transformed auxiliary variable

Horng-Jinh Chang and Kuo-Chung Huang

Department of Management Sciences, Tamkang University, Tamsui, Taipei, Taiwan 25137, R.O.C.

Received: October 12, 1999; revised version: April 25, 2000

This paper is intended as an investigation of constructing almost unbiased estimators of finite population mean by suitably combining a set of transformed estimators. A generalization of Tracy, Singh and Singh (1999) estimator is suggested, and, to the first degree of approximation, each member of the proposed class is as efficient as the usual regression estimator. Further, it is proved that Reddy (1973) estimator is also a particular case of the proposed class. To the second degree of approximation, a new almost unbiased estimator is established. Moreover, an empirical study is carried out in order to understand better the performance of the new estimator compared to the usual unbiased $\bar{y}$, ratio $\bar{y}_R$ and Tracy et al. (1999) estimators.

Key words: bias; mean squared error; simple random sampling without replacement (SRSWOR); transformed auxiliary variable.

1. INTRODUCTION

Consider a finite population $U = (U_1, U_2, \ldots, U_N)$ of size N from which a sample of size n is drawn through simple random sampling without replacement (SRSWOR). Let x_i and y_i be the values of the auxiliary variable x and the study variable y for the unit U_i respectively. Denote by $\bar{X}$ and $\bar{Y}$ the population

† E-MAIL: chj@mail.tku.edu.tw

means of x and y. The problem of interest is to estimate the population mean $\bar{Y}$ on the basis of units in a sample from U with the knowledge that $\bar{X}$ is known. Let $\bar{x}$ and $\bar{y}$ be the unbiased estimators of the population means $\bar{X}$ and $\bar{Y}$ respectively, and $\xi = \bar{X}/\bar{x}$ which is a pivotal quantity.

Generally, positively or negatively correlated between the study variable and the auxiliary variable would lead to different types of estimators. For estimating the population mean $\bar{Y}$, the following estimators are commonly used.

When the correlation between x and y is positive, Cochran (1940) suggests the ratio estimator

$$\bar{y}_R = \bar{y}\xi \tag{1.1}$$

and Bandyopadhyay (1980) and Srivankataramana (1980) suggest the ratio-type estimator, with $g = n/(N-n)$,

$$\bar{y}_R^* = \bar{y}[\xi + g(\xi - 1)]/\xi . \tag{1.2}$$

And, when x is negatively correlated with y, Robson (1957) and Murthy (1964) suggest the product estimator

$$\bar{y}_P = \bar{y}/\xi \tag{1.3}$$

and Bandyopadhyay (1980) and Srivankataramana (1980) suggest

$$\bar{y}_P^* = \bar{y}\xi/[\xi + g(\xi - 1)]. \tag{1.4}$$

In addition to these, several researchers have attempted to formulate modified ratio estimators in order to provide better alternatives, such as Srivastava (1967), Reddy (1973), Gupta (1978), Sahai (1979), Adhvaryu and Gupta (1983) etc. Kothwala and Gupta (1988) investigated the behavior of these mentioned estimators, when second degree approximation is considered, and, it is concluded that one should go for the Reddy (1973) estimator.

Moreover, up to terms of order $O(n^{-1})$, Tracy, Singh and Singh (1999) adopt a procedure, suggested by Singh and Singh (1993), to construct an almost unbiased estimator, with $K = \rho C_y/C_x$, given by

$$\hat{Y}_{\delta 0}^{(u)} = \bar{y}\left\{1 - \frac{K[2K-(\gamma-1)g]}{\alpha[\alpha+1-(\gamma-1)g]} - \frac{K[\alpha+1-2K]}{\gamma g[\alpha+1-(\gamma-1)g]} + \frac{K[2K-(\gamma-1)g]}{\alpha[\alpha+1-(\gamma-1)g]}\xi^{\alpha} + \frac{K[\alpha+1-2K]}{\gamma g[\alpha+1-(\gamma-1)g]}[\{\xi+g(\xi-1)\}/\xi]^{\gamma}\right\} \tag{1.5}$$

It is well known that the usual ratio (product) estimator of the population mean $\bar{Y}$ using auxiliary variable x is optimum if the regression of y on x is linear and goes through the origin. Often, however, the regression of y on x is linear but does not go through the neighborhood of the origin. In that case the usual ratio (product) estimator of $\bar{Y}$ is inappropriate. In such a case, it is more appropriate to use the transformed auxiliary variable to estimate the population mean $\bar{Y}$, see Mohanty and Das (1971). In this paper, a number of almost unbiased estimators based on a transformed auxiliary variable are proposed. In section 2, a generalization of Tracy et al. (1999) estimator is suggested with its properties. And section 3 provides evidence that the ratio-cum-product estimator $\bar{y}\xi[\xi - K(\xi-1)]^{-1}$ suggested by Reddy (1973) belongs to the proposed class. Up to the second degree of approximation, a new almost unbiased estimator is developed in section 4, assuming that N is large and the population follows a bivariate normal distribution. In addition, an empirical study is reported in section 5 with the help of live data.

2. THE PROPOSED CLASS OF ESTIMATORS

In this section we consider a generalization of Tracy et al. (1999) estimator and its specific cases will be provided in the next two sections.

Suppose that the values of x and y are positive for all units in the population. Define $e_0 = (\bar{y}/\bar{Y}) - 1$ and $e_1 = (\bar{x}/\bar{X}) - 1$. Since the sample is drawn by the method of SRSWOR, we have, with $\theta = (N-n)/(nN)$,

$$E(e_0) = E(e_1) = 0, \quad E(e_0 e_1) = \theta K C_x^2,$$
$$E(e_0^2) = \theta C_y^2, \quad E(e_1^2) = \theta C_x^2$$

where

$$C_y^2 = \sum_{i=1}^{N} (y_i - \bar{Y})^2 / (N-1)\bar{Y}^2, \quad C_x^2 = \sum_{i=1}^{N} (x_i - \bar{X})^2 / (N-1)\bar{X}^2$$

and ρ denotes the correlation coefficient between x and y.

According to the transformed ratio method, the ratio $\bar{X}/\bar{x}$ is transformed to be

$$\frac{\bar{X}+L}{\bar{x}+L} = \frac{(1+L')\xi}{1+L'\xi} \tag{2.1}$$

where $L' = L/\bar{X}$, L is a nonnegative constant. Consider the transformed estimators $\bar{y}_1 = \bar{y}$, $\bar{y}_2 = \bar{y}[(1+L')\xi/(1+L'\xi)]^{\alpha}$ and $\bar{y}_3 = \bar{y}[\{(1+L')\xi + g(\xi-1)\}/(1+L')\xi]^{\gamma}$ where (α, γ) are constants which may take values $(1, 1)$ or $(-1, -1)$ according as x and y are positively or negatively correlated. By taking the three estimators mentioned above and considering their convex combination, we can define a new class of estimators, which takes the form

$$\hat{\bar{Y}} = \sum_{i=1}^{3} \delta_i \bar{y}_i \tag{2.2}$$

such that

$$\sum_{i=1}^{3} \delta_i = 1 \tag{2.3}$$

where the weights δ_1, δ_2 and δ_3 are suitably chosen constants so that the bias in the estimator $\hat{\bar{Y}}$ to the first degree of approximation is zero.

Expressing $\hat{\bar{Y}}$ in terms of e_0 and e_1, it follows that

$$\hat{\bar{Y}} = \bar{Y}\{1 + e_0 - (1+L')^{-1}(\alpha\delta_2 + \gamma g \delta_3)e_1\} + O_p(e^2), \tag{2.4}$$

and thus,

$$\hat{\bar{Y}} - \bar{Y} = \bar{Y}\{e_0 - (1+L')^{-1}(\alpha\delta_2 + \gamma g\delta_3)e_1\} + O_p(e^2). \tag{2.5}$$

Squaring (2.5), neglecting terms involving powers in e_0 and e_1 higher than the second, and taking expectation, we have, up to terms of order $O(n^{-1})$,

$$MSE\left(\hat{\bar{Y}}\right) = \theta\bar{Y}^2\left[C_y^2 - 2\left(\frac{\alpha\delta_2 + \gamma g\delta_3}{1+L'}\right)KC_x^2 + \left(\frac{\alpha\delta_2 + \gamma g\delta_3}{1+L'}\right)^2 C_x^2\right] \tag{2.6}$$

which is minimized when

$$(1+L')^{-1}(\alpha\delta_2 + \gamma g\delta_3) = K. \tag{2.7}$$

Consequently, we get the minimum mean square error as

$$MSE_{opt}\left(\hat{\bar{Y}}\right) = \theta\bar{Y}^2 C_y^2(1-\rho^2). \tag{2.8}$$

Further, in order to eliminate the bias of $\hat{\bar{Y}}$, we have a linear restriction

$$\sum_{i=1}^{3}\delta_i B(\bar{y}_i) = 0 \tag{2.9}$$

where $B(\bar{y}_i)$ denotes the bias in the i th estimator.

It is evident to see that the biases of $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}_3$, up to terms of order $O(n^{-1})$, are respectively given by

$$B(\bar{y}_1) = 0, \tag{2.10}$$

$$B(\bar{y}_2) = \theta(\bar{Y}/2)C_x^2(1+L')^{-2}\alpha[\alpha + 1 - 2(1+L')K], \tag{2.11}$$

$$B(\bar{y}_3) = \theta(\bar{Y}/2)C_x^2(1+L')^{-2}\gamma g[(\gamma-1)g - 2(1+L')K]. \tag{2.12}$$

In the light of the expressions of (2.3), (2.7) and (2.9), we have

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & \alpha(1+L')^{-1} & \gamma g(1+L')^{-1} \\ 0 & \alpha[\alpha+1-2(1+L')K] & \gamma g[(\gamma-1)g - 2(1+L')K] \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} = \begin{bmatrix} 1 \\ K \\ 0 \end{bmatrix}. \tag{2.13}$$

Solving (2.13) for δ_i $(i = 1, 2, 3)$, it follows that

$$\delta_1 = 1 - \frac{K(1+L')[2(1+L')K - (\gamma-1)g]}{\alpha[\alpha+1-(\gamma-1)g]} - \frac{K(1+L')[\alpha+1-2(1+L')K]}{\gamma g[\alpha+1-(\gamma-1)g]}, \tag{2.14}$$

$$\delta_2 = \frac{K(1+L')[2(1+L')K - (\gamma-1)g]}{\alpha[\alpha+1-(\gamma-1)g]}, \tag{2.15}$$

$$\delta_3 = \frac{K(1+L')[\alpha+1-2(1+L')K]}{\gamma g[\alpha+1-(\gamma-1)g]}. \tag{2.16}$$

Substituting from the optimum weights (2.14), (2.15) and (2.16) in (2.2) and after some simplification, we obtain a general class of almost optimum unbiased estimators (AOUEs) t_1 (say) as

$$t_1 = \bar{y}\left\{1 - \frac{K(1+L')[2(1+L')K - (\gamma-1)g]}{\alpha[\alpha+1-(\gamma-1)g]} - \frac{K(1+L')[\alpha+1-2(1+L')K]}{\gamma g[\alpha+1-(\gamma-1)g]}\right.$$

$$+\frac{K(1+L')[2(1+L')K-(\gamma-1)g]}{\alpha[\alpha+1-(\gamma-1)g]}\left[\frac{(1+L')\xi}{1+L'\xi}\right]^{\alpha}$$

$$\left.+\frac{K(1+L')[\alpha+1-2(1+L')K]}{\gamma g[\alpha+1-(\gamma-1)g]}[\{(1+L')\xi+g(\xi-1)\}/(1+L')\xi]^{\gamma}\right\} \tag{2.17}$$

with the variance

$$V(t_1)=\theta\bar{Y}^2C_y^2(1-\rho^2) \tag{2.18}$$

which is equivalent to that of the usual linear regression estimator.

Remark 1: Corresponding to the various suitable choices of δ_i $(i=1,2,3)$ and L, the expression (2.2) generates a class of estimators. For example, if $\delta_1=1$ and $\delta_2=\delta_3=0$, then the estimator becomes $\bar{y}$, the simple unbiased estimator of $\bar{Y}$. In a similar way, it is obvious to see that $\bar{y}_R$, $\bar{y}_P$, $\bar{y}_R^*$ and $\bar{y}_P^*$ belong to this class.

Remark 2: In case of $L=0$ in (2.17), the almost optimum unbiased estimator t_1 reduces to Tracy et al. (1999) estimator $\hat{Y}_{\delta 0}^{(u)}$.

Remark 3: It is to be noted that the AOUE t_1 could be used in practice if the parameter K is known. A prior knowledge of K sometimes can be obtained from a most recent survey taken in the past or by conducting a survey using double sampling technique. This problem has been investigated by many researchers such as Murthy (1967, p.96-99), Reddy (1978) and Srivenkataramana and Tracy (1984). Thus, it is not entirely unrealistic to assume a prior knowledge of K.

When the exact value of K or its guessed value is not known in advance, it is advisable to replace it with its sample analogue $\hat{K}$ and then we get the following estimator, given by

$$t_1^*=\bar{y}\left\{1-\frac{\hat{K}(1+L')[2(1+L')\hat{K}-(\gamma-1)g]}{\alpha[\alpha+1-(\gamma-1)g]}-\frac{\hat{K}(1+L')[\alpha+1-2(1+L')\hat{K}]}{\gamma g[\alpha+1-(\gamma-1)g]}\right.$$

$$+\frac{\hat{K}(1+L')[2(1+L')\hat{K}-(\gamma-1)g]}{\alpha[\alpha+1-(\gamma-1)g]}\left[\frac{(1+L')\xi}{1+L'\xi}\right]^{\alpha}$$

$$\left.+\frac{\hat{K}(1+L')[\alpha+1-2(1+L')\hat{K}]}{\gamma g[\alpha+1-(\gamma-1)g]}[\{(1+L')\xi+g(\xi-1)\}/(1+L')\xi]^{\gamma}\right\} \tag{2.19}$$

where $\hat{K}=\dfrac{\bar{X}s_{xy}}{\bar{y}s_x^2}$ denotes an estimator of K.

To find the mean square error of the estimator, let us define $\eta=(\hat{K}/K)-1$, where $E(\eta)=O(n^{-1})$, then, expressing t_1^* in terms of e_0, e_1 and η, we have

$$t_1^*-\bar{Y}=\bar{Y}\{e_0-Ke_1-K\eta e_1-Ke_0e_1+K^2e_1^2\}. \tag{2.20}$$

One can easily check that the mean square error of t_1^* to the first order of approximation is same as that given in (2.18). In addition, t_1^* is biased with magnitude of bias given by

$$B(t_1^*) = \theta \bar{Y}\left(C_{11}^2 C_{02} + C_{03}C_{11} - C_{12}C_{02}\right)/C_{02}^2 \tag{2.21}$$

where

$$C_{ls} = \frac{(1/N)\sum_{i=1}^{N}(y_i - \bar{Y})^l (x_i - \bar{X})^s}{\bar{Y}^l \bar{X}^s}.$$

Further, the sampling bias of the simple regression estimator $\bar{y}_l = \bar{y} + b(\bar{X} - \bar{x})$, where $b = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2$, can also be obtained as

$$B(\bar{y}_l) = \theta \bar{Y}\left(C_{03}C_{11} - C_{12}C_{02}\right)/C_{02}^2 . \tag{2.22}$$

Comparing t_1^* with respect to $\bar{y}_l$, we observe that the former is more efficient than the latter if the following condition is satisfied:

$$C_{11}^2 C_{02} + 2\left(C_{03}C_{11} - C_{12}C_{02}\right) < 0 . \tag{2.23}$$

3. SPECIFIC ESTIMATOR OF THE CLASS

In this section, we assume that the weights $\delta_1 \in R$ and $\delta_2, \delta_3 \in R^+ \cup \{0\}$ where R denotes the set of real numbers and R^+ be the set of positive real numbers. The mean square errors of $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}_3$ to the first degree of approximation are respectively given by

$$MSE(\bar{y}_1) = \theta \bar{Y}^2 C_y^2 , \tag{3.1}$$

$$MSE(\bar{y}_2) = \theta \bar{Y}^2 \left\{C_y^2 + \alpha(1+L')^{-2}\left[\alpha - 2K(1+L')\right]C_x^2\right\}, \tag{3.2}$$

$$MSE(\bar{y}_3) = \theta \bar{Y}^2 \left\{C_y^2 + \gamma g(1+L')^{-2}\left[\gamma g - 2K(1+L')\right]C_x^2\right\}. \tag{3.3}$$

Notice that $MSE(\bar{y}_1)$ does not depend on the value of L. Now we are desired to find the optimum value of L such that $\delta_2 MSE(\bar{y}_2) + \delta_3 MSE(\bar{y}_3)$ attains its minimum. It is easy to verify that the minimum value occurs if the following equation holds:

$$(1+L')^{-1}\left(\frac{\alpha^2\delta_2 + \gamma^2 g^2 \delta_3}{\alpha\delta_2 + \gamma g \delta_3}\right) = K . \tag{3.4}$$

Solving the expression (3.4) with respect to restrictions in (2.13), we obtain the optimum values of δ_i $(i = 1, 2, 3)$ and L as follows:

$$\delta_1 = \frac{(\alpha - 1)(\gamma + 1)(\alpha - \gamma g)^2}{\alpha\gamma\left[(\gamma+1)g + (1-\alpha)\right]^2} , \tag{3.5}$$

$$\delta_2 = \frac{(\alpha + \gamma)(\gamma + 1)g^2}{\alpha\left[(\gamma+1)g + (1-\alpha)\right]^2} , \tag{3.6}$$

$$\delta_3 = \frac{(\alpha+\gamma)(1-\alpha)}{\gamma[(\gamma+1)g+(1-\alpha)]^2}, \tag{3.7}$$

$$L = \bar{X}\left[\frac{(\alpha+\gamma)g}{K[(\gamma+1)g+(1-\alpha)]}-1\right]. \tag{3.8}$$

And thus, we get the almost optimum unbiased estimators t_2 (say), given by

$$t_2 = \bar{y}\left\{\frac{(\alpha-1)(\gamma+1)(\alpha-\gamma g)^2}{\alpha\gamma[(\gamma+1)g+(1-\alpha)]^2} + \frac{(\alpha+\gamma)(\gamma+1)g^2}{\alpha[(\gamma+1)g+(1-\alpha)]^2}\left[\frac{(\alpha+\gamma)g\xi}{(\alpha+\gamma)g\xi-K\{(\gamma+1)g+(1-\alpha)\}(\xi-1)}\right]^{\alpha} + \frac{(\alpha+\gamma)(1-\alpha)}{\gamma[(\gamma+1)g+(1-\alpha)]^2}\left[\frac{(\alpha+\gamma)\xi+K\{(\gamma+1)g+(1-\alpha)\}(\xi-1)}{(\alpha+\gamma)\xi}\right]^{\gamma}\right\} \tag{3.9}$$

with the same variance as given in (2.18).

Remark 4: It is obvious to see that for the cases of $\alpha=\gamma=1$ and $\alpha=\gamma=-1$, the resulting almost optimum unbiased estimators are identical with Reddy (1973) estimator t_3 (say), given by

$$t_3 = \bar{y}\xi[\xi-K(\xi-1)]^{-1} \tag{3.10}$$

which is demonstrated as a particular case of the suggested class.

4. HIGHER ORDER APPROXIMATION

In deriving the expected values and the mean square errors in the previous sections, it is assumed that the contribution of terms involving powers in e_0 and e_1 higher than the second is negligible. We shall now retain the terms in e_0 and e_1 up to the fourth, and proceed to obtain a better approximation to the expected value and the mean square error of the estimator t_1. For simplicity, it is assumed that the population size N is very large as compared to sample size n, so that finite population correction terms can be ignored and $g=0$.

For the optimum values of δ_1, δ_2 and δ_3 given in section 2, up to the second degree of approximation, it follows that

$$t_1 - \bar{Y} = \bar{Y}\{e_0 - Ke_1 - Ke_0e_1 + K^2e_1^2 + K^2e_0e_1^2 - \tau e_1^3 - \tau e_0e_1^3 + \omega e_1^4\} \tag{4.1}$$

where

$$\tau = \frac{(\alpha+2)}{3(1+L')}K^2$$

and

$$\omega = \frac{(\alpha+2)(\alpha+3)}{12(1+L')^2}K^2.$$

Squaring (4.1) and ignoring terms involving powers in e_0 and e_1 higher than the fourth, we have

$$MSE(t_1) = \bar{Y}^2 E\{e_0^2 - 2Ke_0e_1 + K^2e_1^2 + 4K^2e_0e_1^2 - 2Ke_0^2e_1 - 2K^3e_1^3 + 3K^2e_0^2e_1^2 - 2(2K^3 + \tau)e_0e_1^3 + K(K^3 + 2\tau)e_1^4\}. \quad (4.2)$$

It is seen that the contribution of higher order terms depends on the values of the moments and product-moments of $\bar{x}$ and $\bar{y}$. Sukhatme and Sukhatme (1970) have obtained the expected values of third and fourth order moments and product-moments of two variables. In case of bivariate normal population, it follows that

$$C_{12} = C_{21} = C_{03} = C_{30} = 0, \quad C_{22} = C_{02}C_{20} + 2C_{11}^2,$$

$$C_{13} = 3C_{11}C_{02}, \quad C_{04} = 3C_{02}^2.$$

And then we have

$$E(e_0e_1^2) = E(e_0^2e_1) = E(e_1^3) = 0, \quad E(e_0^2e_1^2) = \frac{1}{n^2}(C_{20}C_{02} + 2C_{11}^2),$$

$$E(e_0e_1^3) = \frac{3}{n^2}C_{11}C_{02}, \quad E(e_1^4) = \frac{3}{n^2}C_{02}^2.$$

Substituting these expected values in (4.2), after some simple algebra we get

$$MSE(t_1) = \frac{1}{n}\bar{Y}^2C_y^2(1 - \rho^2)\left(1 + \frac{3}{n}\rho^2C_y^2\right) \quad (4.3)$$

which does not depend on the value of L.

Moreover, the bias of t_1, up to the second degree of approximation, is

$$B(t_1) = \frac{3}{n^2}C_x^4(\omega - \tau K). \quad (4.4)$$

It may be noted that if $L = 0$, i.e., Tracy et al. (1999) estimator $\hat{Y}_{80}^{*}$ is biased with amount of bias given by

$$B(\hat{Y}_{80}^{*}) = \frac{(\alpha + 2)(\alpha + 3 - 4K)K^2C_x^4}{4n^2}. \quad (4.5)$$

Now we are attempted to find the optimum value of L in order to provide an unbiased estimator. It is obvious to see that bias is zero if $\omega = \tau K$, i.e., when the following equation holds:

$$\alpha + 3 - 4(1 + L')K = 0. \quad (4.6)$$

This leads to

$$L = \bar{X}\left(\frac{\alpha + 3}{4K} - 1\right). \quad (4.7)$$

Putting the optimum value of L in (2.17), we obtain the almost unbiased estimator t_4 (say) as

$$t_4 = \bar{y}\left[1 - \frac{(\alpha + 3)^2}{8\alpha(\alpha + 1)} + \frac{(\alpha + 3)^2}{8\alpha(\alpha + 1)}\left\{\frac{(\alpha + 3)\xi}{4K + (\alpha + 3 - 4K)\xi}\right\}^{\alpha}\right]. \quad (4.8)$$

It can easily be proved that the proposed almost unbiased estimator t_4 is more efficient than the usual unbiased $\bar{y}$, ratio $\bar{y}_R$, product $\bar{y}_P$ estimators, as these are particular cases of the suggested class in (2.2). Moreover, it is interesting to note that

t_4 and $\hat{Y}_{80}^{(u)}$ utilize the same amount of information and the proposed estimator t_4 is superior to Tracy et al. (1999) estimator $\hat{Y}_{80}^{(u)}$ in the sense of bias criterion.

Furthermore, the mean square error of the usual linear regression estimator $\bar{y}_{lr} = \bar{y} + B(\bar{X} - \bar{x})$, where $B = S_{xy}/S_x^2$, is given by

$$MSE(\bar{y}_{lr}) = \frac{1}{n}\bar{Y}^2 C_y^2 (1-\rho^2) \tag{4.9}$$

which is equal to the first degree mean square error of t_4.

The mean square error of t_4 can, therefore, be expressed as

$$MSE(t_1) = MSE(\bar{y}_{lr})\left(1 + \frac{3}{n}\rho^2 C_y^2\right). \tag{4.10}$$

Equation (4.10) shows that the contribution of the third and fourth degree terms to the MSE of the estimator t_4 is $3\rho^2 C_y^2/n$ times the value of mean square error of usual linear regression estimator. Unless n is small, the contribution can, therefore, be considered to be negligible.

5. EMPIRICAL STUDY

In this section, we carry out an empirical study to see the performance of the proposed estimator t_4 with respect to the unbiased estimator $\bar{y}$, ratio estimator $\bar{y}_R$ and Tracy et al. (1999) estimator $\hat{Y}_{80}^{(u)}$ with $\alpha = 1$. For this purpose, we consider the data given in Cochran (1977, p.172) dealing with data

y : the estimated production in bushels of peach

x : the number of peach trees in an orchard.

The required values are given as follows:

$S_y^2 = 6409$, $S_x^2 = 3898$, $S_{xy} = 4434$, $N = 256$

$\bar{Y} = 56.47$, $\bar{X} = 44.45$, $\rho = 0.887$

and we take $n = 2, 4, 8, 16, 32, 64$.

The first degree and second degree biases and mean square errors of the unbiased estimator $\bar{y}$, ratio estimator $\bar{y}_R$, Tracy et al. (1999) estimator $\hat{Y}_{80}^{(u)}$, and the proposed estimator t_4 are presented in Tables 1 and 2 respectively. The symbol 'I' in the suffix indicates the first degree approximation, while 'II' represents the second degree approximation.

Table 1. The first degree biases and MSEs of the mentioned estimators

n	$B_I(\bar{y})$	$B_I(\bar{y}_R)$	$B_I(\hat{Y}_{80}^{(u)})$	$B_I(t_4)$	$MSE_I(\bar{y})$	$MSE_I(\bar{y}_R)$	$MSE_I(\hat{Y}_{80}^{(u)})$	$MSE_I(t_4)$
2	0	5.8341	0	0	3204.5000	717.8039	683.2987	683.2987
4	0	2.9171	0	0	1602.2500	358.9019	341.6494	341.6494
8	0	1.4585	0	0	801.1250	179.4510	170.8247	170.8247
16	0	0.7293	0	0	400.5625	89.7255	85.4123	85.4123
32	0	0.3646	0	0	200.2813	44.8627	42.7062	42.7062
64	0	0.1823	0	0	100.1406	22.4314	21.3531	21.3531

Table 2. The second degree biases and MSEs of the mentioned estimators

n	$B_{II}(\bar{y})$	$B_{II}(\bar{y}_R)$	$B_{II}(\hat{Y}_{80}^{(w)})$	$B_{II}(t_4)$	$MSE_{II}(\bar{y})$	$MSE_{II}(\bar{y}_R)$	$MSE_{II}(\hat{Y}_{80}^{(w)})$	$MSE_{II}(t_4)$
2	0	23.0991	0.2450	0	3204.5000	3046.2261	2304.0032	2304.0032
4	0	7.2333	0.0613	0	1602.2500	941.0075	746.8255	746.8255
8	0	2.5376	0.0153	0	801.1250	324.9774	272.1187	272.1187
16	0	0.9990	0.0038	0	400.5625	126.1071	110.7359	110.7359
32	0	0.4321	0.0010	0	200.2813	53.9581	49.0370	49.0370
64	0	0.1992	0.0002	0	100.1406	24.7052	22.9358	22.9358

It is observed from the above tables that the proposed estimator t_4 is almost unbiased whether we use first or second degree approximation and also, there is a considerable reduction in the mean square error of this estimator from that of ratio estimator $\bar{y}_R$. And as expected, the constructed estimator t_4 performs better than the unbiased estimator $\bar{y}$, ratio estimator $\bar{y}_R$ and Tracy et al. (1999) estimator $\hat{Y}_{80}^{(w)}$. Thus in such situation one should go for the almost unbiased estimator t_4.

ACKNOWLEDGEMENT

The authors are grateful to the referees and the editor for their helpful comments in improving the content of this paper.

REFERENCES

Adhvaryu, D. and Gupta, P. C. (1983). On some alternative sampling strategies using auxiliary information. Metrika, 30, 217-226.

Bandyopadhyay, S. (1980). Improved ratio and product estimators. Sankhya C, 42, 45-59.

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

Gupta, P. C. (1978). On some quadratic and higher degree ratio and product estimator. Jour. Ind. Soc. Agril. Stat., 30, 71-80.

Kothwala, N. H. and Gupta, P. C. (1988). A study of second order approximation of some ratio type strategies. Biometrical journal, 30, 3, 369-377.

Mohanty, S. and Das, M. N. (1971). Use of transformation in sampling. Jour. Ind. Soc. Agril. Stat., 23, 83-87.

Murthy, M. N. (1964). Product method of estimation. Sankhya A, 26, 294-307.

Murthy, M. N. (1967). Sampling Theory and Methods. Statistical Publishing Society, Calcutta.

Reddy, V. N. (1973). On ratio and product methods of estimation. Sankhya B, 35, 307-317.

Reddy, V. N. (1978). A study on the prior knowledge on certain population parameters. Sankhya C, 40, 29-37.

Robson, D. S. (1957). Applications of multivariate polykays to the theory of unbiased ratio-type estimation. Jour. Amer. Stat. Assoc., 52, 511-522.

Sahai, A. (1979). An efficient varient of the product and ratio estimators. Statistica Neerlandica, 33, 27-35.

Singh, S. and Singh, R. (1993).Almost filtration of bias precipitates: a new approach. Jour. Ind. Soc. Agril. Stat., 45, 214-218.

Srivastava, S. K. (1967). An estimator using auxiliary information. Bull. Cal. Stat. Assoc., 16, 121-132.

Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. Biometrika, 67, 199-204.

Srivenkataramana, T. and Tracy, D. S. (1984). Positive and negative valued auxiliary variates in surveys. Metron, 42 (3-4), 3-14.

Sukhatme, P. V. and Sukhatme, B. V. (1970). Sampling theory of surveys with application. 2nd edition, Iowa State University Press, Ames.

Tracy, D. S., Singh, H. P. and Singh, R. (1999). Constructing an unbiased estimator of population mean in finite populations using auxiliary information. Statistical Papers, 40, 363-368.