Chilean Journal of Statistics
Vol. 4, No. 1, April 2013, 3–17

Sampling Theory
Research Paper

Improved estimation of population mean using population proportion of an auxiliary character

Ramkrishna S. Solanki and Housila P. Singh

School of Studies in Statistics, Vikram University, Ujjain, India

(Received: 20 January 2011 · Accepted in final form: 27 February 2012)

Abstract

This paper suggests a class of estimators for population mean of a study variable using known population proportion of an auxiliary character (attribute) which is highly correlated with the study variable. The suggested class of estimators is more general and includes the usual unbiased estimator (sample mean) and the estimators reported by Singh et al. (2007). Expressions of bias and mean square error (MSE) for the proposed class of estimators have been obtained under large sample approximation. Asymptotic optimum estimator (AOE) has been identified along with its mean square error formula. Regions of preferences have been investigated under which the proposed class of estimators is more efficient than the usual unbiased, Naik and Gupta (1996) and Singh et al. (2007) estimators. Two phase sampling version (i.e. population proportion is unknown) of the proposed class of estimators has been given along with its properties under large sample approximation. The theoretical results have been illustrated empirically by using some population data sets in both phases.

Keywords: Bias · Mean square error · Proportion · Study variable · Two phase sampling.

Mathematics Subject Classification: Primary 62D05.

1. Introduction

It is well known that the efficiencies of the estimators of population mean of a study variable y have been increased by the use of information available on an auxiliary variable x which is highly correlated with study variable y. Out of many ratio, product and regression methods of estimation are good examples in this context; see Singh (2003). However in many situations of practical importance, instead of an auxiliary variable x there exists an auxiliary attribute (say ϕ), which is highly correlated with the study variable y, such as

(1) height of the person (y) and gender (ϕ),
(2) amount of milk produced (y) and a particular breed of the cow (ϕ),
(3) amount of yield of wheat crop (y) and a particular variety of wheat (ϕ),
(4) use of drugs (y) and gender (ϕ); see Jhajj et al. (2006) and Shabbir and Gupta (2007).

*Corresponding author. Email: ramkssolanki@gmail.com

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

In all of these examples point biserial correlation, (see Kendall and Stuart, 1967) between study variable y and the auxiliary attribute ϕ exists. In such situations, taking the advantage of point biserial correlation between the study variable y and the auxiliary attribute ϕ, the efficient estimators of parameters of study can be constructed by using prior knowledge of the parameters of auxiliary attribute ϕ. So by taking in to consideration the point biserial correlation between a study variable y and an attribute ϕ, some authors such as Naik and Gupta (1996), Jhajj et al. (2006), Singh et al. (2007), Shabbir and Gupta (2007, 2010) and Abd-Elfattah et al. (2010) have paid their attention towards the improved estimation of population mean when the prior information of population proportion of units possessing the same attribute is available. This encourages authors to envisage a class of estimators in simple random sampling for population mean of a study variable y which is more general and efficient than other existing estimators by using information on an auxiliary attribute ϕ which is highly correlated with the study variable y.

Consider a sample of size n drawn by simple random sampling without replacement (SRSWOR) from a population of size N. Let y_i and ϕ_i denote the observations on the study variable y and attribute ϕ respectively for ith unit of the population, for $i = 1, \ldots, N$. Suppose there is a complete dichotomy in the population with respect to the presence ($\phi = 1$) or absence ($\phi = 0$) of an attribute ϕ. We take the following notations which we will be used along the paper.

The population mean of the study variable y is denoted by $\bar{Y} = (1/N)\sum_{i=1}^{N} y_i$, $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ denotes the sample mean of the study variable y, $A = \sum_{i=1}^{N} \phi_i$ is the total number of units in the population possessing attribute ϕ, $a = \sum_{i=1}^{n} \phi_i$ is the total number of units in the sample possessing attribute ϕ, $p = (A/N)$ is the proportion of units in the population possessing attribute ϕ, $\hat{p} = (a/n)$ is the proportion of units in the sample possessing attribute ϕ, $S_y^2 = [1/(N-1)]\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2$ is the population mean square of y, $S_\phi^2 = [1/(N-1)]\sum_{i=1}^{N}(\phi_i - p)^2$ is the population mean square of ϕ, $s_\phi^2 = [1/(n-1)]\sum_{i=1}^{n}(\phi_i - \hat{p})^2$ is the sample mean square of ϕ, $S_{y\phi} = [1/(N-1)]\left(\sum_{i=1}^{N} y_i\phi_i - Np\bar{Y}\right)$ is the population covariance between y and ϕ, $s_{y\phi} = [1/(n-1)]\left(\sum_{i=1}^{n} y_i\phi_i - n\hat{p}\bar{Y}\right)$ is the sample covariance between y and ϕ, $C_y^2 = \left(S_y^2/\bar{Y}^2\right)$ is the square of coefficient of variation of y, $C_p^2 = (S_\phi^2/p^2)$ is the square of coefficient of variation of ϕ, $\rho_{pb} = [S_{y\phi}/(S_y S_\phi)]$ is the point biserial correlation coefficient between y and ϕ, $K_p = \rho_{pb}\,(C_y/C_p)$ and $\theta = [(1/n) - (1/N)]$.

When population proportion p is known, Naik and Gupta (1996) suggested the following ratio and product estimators for population mean $\bar{Y}$ of study variable y respectively as

$$t_1 = \bar{y}\left(\frac{p}{\hat{p}}\right), \text{ (ratio estimator)},$$

$$t_2 = \bar{y}\left(\frac{\hat{p}}{p}\right), \text{ (product estimator)}.$$

Following Bahl and Tuteja (1991), Singh et al. (2007) suggested ratio and product exponential estimators for population mean $\bar{Y}$ of study variable y respectively as

$$t_3 = \bar{y}\exp\left(\frac{p-\hat{p}}{p+\hat{p}}\right), \text{ (ratio exponential estimator)},$$

$$t_4 = \bar{y}\exp\left(\frac{\hat{p}-p}{\hat{p}+p}\right), \text{ (product exponential estimator)}.$$

The goal of the paper is to envisaged an efficient class of estimators in the SRSWOR for the population mean $\bar{Y}$ of the study variable y which is more general and includes as particular cases other estimators such as usual unbiased estimator $\bar{y}$, Singh et al.'s (2007) ratio exponential estimators t_3, and product exponential estimator t_4 using the known population proportion p of an attribute ϕ which is highly correlated with the study variable y (see Section 2.1).

It is well known under SRSWOR that the variance of the usual unbiased estimator $\bar{y}$ (sample mean) is

$$\mathrm{Var}\left[\bar{y}\right] = \theta S_y^2 = \bar{Y}^2 \theta C_y^2.$$

The remaining part of the paper has been organized as follows. In Section 2 we have proposed a class of estimators for population mean along with its bias and MSE formulae up to the first degree of approximation. We have also made bias and efficiency comparisons of the proposed class of estimators with different estimators. The optimum choice of the scalar involved in the proposed class of estimators has been obtained. An estimator based on estimated optimum value is derived along with its mean square error formula. We have carried out an empirical study to judge the merits of the proposed class of estimators over other competitors. Section 3 described two phase sampling procedure and the description of the estimators envisaged by earlier authors. In Section 4 we have considered the proposed class of estimators in two phase sampling along with its properties. Finally Section 5 sketches some conclusions.

2. Estimation of Mean

2.1 The proposed class of estimators with known population proportion

We suggest the following class of estimators for population mean $\bar{Y}$ as

$$t_{(\alpha)} = \bar{y} \exp\left(\frac{\alpha\left(p-\hat{p}\right)}{p+\hat{p}}\right), \tag{1}$$

where α is a suitably chosen scalar. We note that

(1) for $\alpha = 0, t_{(\alpha)} = t_{(0)} = \bar{y}$, (usual unbiased estimator),

(2) for $\alpha = 1, t_{(\alpha)} = t_{(1)} = \bar{y} \exp\left(\frac{(p-\hat{p})}{p+\hat{p}}\right) = t_3$, (Singh et al., 2007),

(3) for $\alpha = -1, t_{(\alpha)} = t_{(-1)} = \bar{y} \exp\left(\frac{(\hat{p}-p)}{\hat{p}+p}\right) = t_4$, (Singh et al., 2007).

Thus the proposed class of estimators t_α is a generalized version of usual unbiased estimator $\bar{y}$ and Singh et al. (2007) estimators t_3 and t_4.

To obtain the bias and MSE of $t_{(\alpha)}$, we define

$$\bar{y} = \bar{Y}\left(1+e_0\right) \text{ and } \hat{p} = p\left(1+e_1\right),$$

such that

$$\mathrm{E}\left(e_0\right) = \mathrm{E}\left(e_1\right) = 0, \mathrm{E}\left(e_0^2\right) = \theta C_y^2, \mathrm{E}\left(e_1^2\right) = \theta C_p^2 \text{ and } \mathrm{E}\left(e_0 e_1\right) = \theta K_p C_p^2.$$

Expressing the suggested class of estimators $t_{(\alpha)}$ in Equation (1) in term of e's we have

$$\begin{aligned} t_{(\alpha)} &= \bar{Y}\left(1+e_0\right)\exp\left(\frac{-\alpha e_1}{2+e_1}\right) \\ &= \bar{Y}\left(1+e_0\right)\exp\left(\frac{-\alpha e_1}{2}\left[1+\frac{e_1}{2}\right]^{-1}\right) \\ &= \bar{Y}\left[1+e_0-\frac{\alpha e_1}{2}-\frac{\alpha e_0 e_1}{2}+\frac{\alpha(\alpha+2)}{8}e_1^2-\frac{\alpha e_1^3}{48}\left(\alpha^2+6\alpha+6\right)+\cdots\right]. \end{aligned}$$

Neglecting terms of e's having power greater than two of the above expression we have

$$t_{(\alpha)} \approxeq \bar{Y}\left[1+e_0-\frac{\alpha e_1}{2}-\frac{\alpha e_0 e_1}{2}+\frac{\alpha\left(\alpha+2\right)}{8}e_1^2\right],$$

or

$$\left(t_{(\alpha)}-\bar{Y}\right) \approxeq \bar{Y}\left[e_0-\frac{\alpha e_1}{2}-\frac{\alpha e_0 e_1}{2}+\frac{\alpha\left(\alpha+2\right)}{8}e_1^2\right]. \tag{2}$$

Taking expectation on both sides of Equation (2) we get the bias of $t_{(\alpha)}$ to the first degree of approximation as

$$B\left(t_{(\alpha)}\right) = \mathrm{E}\left(t_{(\alpha)}-\bar{Y}\right) = \bar{Y}\theta\left(\frac{\alpha C_p^2}{8}\right)\left(\alpha-4K_p+2\right). \tag{3}$$

It is interesting to note that if we set $\alpha = 2$, -2, 1 and -1 in Equation (3) we can easily get the biases of the estimators t_1, t_2, t_3 and t_4 respectively to the first degree of approximation.

Squaring both sides of Equation (2) and neglecting terms of e's having power greater than two we have

$$\left(t_{(\alpha)}-\bar{Y}\right)^2 \approxeq \bar{Y}^2\left[e_0+\frac{\alpha^2 e_1^2}{4}-\alpha e_0 e_1\right].$$

Taking expectation on both sides of above expression, we get the MSE of $t_{(\alpha)}$ up to first degree of approximation as

$$\begin{aligned} \mathrm{MSE}\left(t_{(\alpha)}\right) &= \mathrm{E}\left(t_{(\alpha)}-\bar{Y}\right)^2 = \bar{Y}^2\mathrm{E}\left[e_0+\frac{\alpha^2 e_1^2}{4}-\alpha e_0 e_1\right] \\ &= \bar{Y}^2\theta\left[C_y^2+\left(\frac{\alpha C_p^2}{4}\right)\left(\alpha-4K_p\right)\right]. \end{aligned} \tag{4}$$

It is interesting to note that if we set $\alpha = 0$, 2, -2, 1 and -1 in Equation (4) we get the mean square errors (MSEs) of the estimators $\bar{y}$, t_1, t_2, t_3 and t_4 respectively to the first degree of approximation.

2.2 Bias comparisons

In this subsection we have obtained the conditions under which the proposed class of estimators $t_{(\alpha)}$ is less biased than Naik and Gupta's (1996) estimators t_1, t_2 and Singh et al.'s (2007) estimators t_3, t_4.

From Equation (3) we have

(1) $\left|B\left(t_{(\alpha)}\right)\right| < |B(t_1)|$ if $\left|\frac{\alpha}{8}\left(\alpha - 4K_p + 2\right)\right| < |(1 - K_p)|$.

(2) $\left|B\left(t_{(\alpha)}\right)\right| < |B\left(t_2\right)|$ if $\left|\frac{\alpha}{8}\left(\alpha - 4K_p + 2\right)\right| < |(K_p)|$.

(3) $\left|B\left(t_{(\alpha)}\right)\right| < |B(t_3)|$ if $|\alpha\left(\alpha - 4K_p + 2\right)| < |(3 - 4K_p)|$.

(4) $\left|B\left(t_{(\alpha)}\right)\right| < |B\left(t_4\right)|$ if $|\alpha\left(\alpha - 4K_p + 2\right)| < |(4K_p - 1)|$.

2.3 Efficiency comparisons

In this subsection we have derived the conditions under which the proposed class of estimators $t_{(\alpha)}$ is more efficient than the usual unbiased estimator $\bar{y}$, Naik and Gupta's (1996) estimators t_1 and t_2, and Singh et al.'s (2007) estimators t_3 and t_4.

From Equation (4) we have

(1) $\mathrm{MSE}\left(t_{(\alpha)}\right) < \mathrm{Var}\left[\bar{y}\right]$ if $\min(0, 4K_p) < \alpha < \max\left(0, 4K_p\right)$.

(2) $\mathrm{MSE}\left(t_{(\alpha)}\right) < \mathrm{MSE}\left(t_1\right)$ if $\min\left(2, 2\left(2K_p - 1\right)\right) < \alpha < \max\left(2, 2\left(2K_p - 1\right)\right)$.

(3) $\mathrm{MSE}\left(t_{(\alpha)}\right) < \mathrm{MSE}\left(t_2\right)$ if $\min\left(-2, 2\left(1 + 2K_p\right)\right) < \alpha < \max\left(-2, 2\left(1 + 2K_p\right)\right)$.

(4) $\mathrm{MSE}\left(t_{(\alpha)}\right) < \mathrm{MSE}\left(t_3\right)$ if $\min\left(1, \left(4K_p - 1\right)\right) < \alpha < \max\left(1, \left(4K_p - 1\right)\right)$.

(5) $\mathrm{MSE}\left(t_{(\alpha)}\right) < \mathrm{MSE}\left(t_4\right)$ if $\min\left(-1, \left(4K_p + 1\right)\right) < \alpha < \max\left(-1, \left(4K_p + 1\right)\right)$.

2.4 Optimum choice of the scalar α

Differentiating $\mathrm{MSE}\left(t_{(\alpha)}\right)$ in Equation (4) with respect to α and equating it to zero, we get the optimum value of α as

$$\alpha = 2K_p = \alpha_o. \tag{5}$$

Substituting Equation (5) in Equation (1) we get the optimum estimator for $\bar{Y}$ as

$$t_{(\alpha_o)} = \bar{y}\exp\left(\frac{2K_p\left(p - \hat{p}\right)}{p + \hat{p}}\right). \tag{6}$$

From Equations (3), (4) and (5) we get the bias and MSE of the optimum estimator $t_{(\alpha_o)}$ to the first degree of approximation respectively as

$$\begin{aligned} B\left(t_{(\alpha_o)}\right) &= \bar{Y}\theta\left(\frac{K_p C_p^2}{2}\right)\left(1 - K_p\right), \\ \mathrm{MSE}\left(t_{(\alpha_o)}\right) &= \theta S_y^2\left(1 - \rho_{pb}^2\right). \end{aligned} \tag{7}$$

The expression in Equation (7) is equal to the variance of the linear regression estimator $\hat{\bar{y}}_{lr} = \left(\bar{y} + \hat{b}\left(p - \hat{p}\right)\right)$, where $\hat{b}\ (= s_{y\phi}/s_{\phi}^2)$ is the sample estimate of the population regression coefficient $\beta\ \left(= S_{y\phi}/S_{\phi}^2\right)$.

It is observed from Equation (6) that the optimum estimator $t_{(\alpha_o)}$ depends on the population parameter K_p which is assumed to be known for the efficient use of the optimum estimator $t_{(\alpha_o)}$. The value of the parameter K_p can be obtained either from the pilot sample survey or from the past experience; for instance see Reddy (1973, 1974) and Srivenkataramana and Tracy (1980). On the other hand if the value of the parameter K_p is not known then it is worth advisable to estimate K_p from the sample data at hand. An estimate of K_p (based on sample data) is given by

$$\hat{K_p} = \left(\frac{s_{y\phi p}}{s_{\phi}^2 \bar{y}}\right) = \frac{\hat{b}}{\hat{R}}, \tag{8}$$

where $\hat{R} = (p/\bar{y})$.

Replacing K_p by its estimate $\hat{K_p}$ defined in Equation (8) we get an estimator of population mean $\bar{Y}$ (based on estimated optimum value) as

$$t_{(\hat{\alpha}_0)} = \bar{y} \exp\left(\frac{2\hat{K_p}\,(p - \hat{p})}{p + \hat{p}}\right).$$

It can be easily shown, to the first degree of approximation that

$$\mathrm{MSE}\left(t_{\hat{\alpha}_0}\right) = \theta S_y^2 \left(1 - \rho_{pb}^2\right) = \mathrm{MSE}\left(t_{(\alpha_o)}\right).$$

From the above expression it is clear that the estimator $t_{(\hat{\alpha}_0)}$ (based on estimated optimum value) is as efficient as the optimum estimator $t_{(\alpha_o)}$.

2.5 Empirical study

In this subsection, we study the preceding theoretical results empirically on two different population data sets. The description of population data sets are summarized in Table 1. Using the conditions which we have obtained in Subsection 2.3 we calculated the ranges of scalar α (see Table 2) in which proposed class of estimators $t_{(\alpha)}$ is more efficient than the usual unbiased estimator $\bar{y}$, Naik and Gupta's ratio estimator (t_1), and Singh et al.'s (2007) ratio exponential estimator (t_3), for the two population data sets described in Table 1. To evaluate the performance of the proposed class of estimators $t_{(\alpha)}$ over the other existing estimators, we have computed the percent relative efficiencies (PREs) of $t_{(\alpha)}$ with respect to $\bar{y}$, t_1 and t_3 in certain range (-1.25, 2.25) of α for population data sets I and II using the following formula

$$\mathrm{PRE}\left(t_{(\alpha)}, \bullet\right) = \frac{\mathrm{MSE}\left(\bullet\right)}{\mathrm{MSE}\left(t_{(\alpha)}\right)} \times 100,$$

where $(\bullet)$ represent the estimators $\bar{y}$, t_1 and t_3.

In addition, to see the effect of scalar α on the bias of proposed class of estimators $t_{(\alpha)}$ we have calculated the following quantity

$$B_{(\alpha)} = \left|\frac{B\left(t_{(\alpha)}\right)}{\left(\frac{1}{8}\bar{Y}\theta C_p^2\right)}\right| = \left|\alpha\left(\alpha - 4K_p + 2\right)\right|,$$

for the population data sets I and II in certain range (-1.25, 2.25) of α. The findings are summarized in Table 3 and Figure 1.

Table 1. Description of population data sets.

	Population data sets (Source: Sukhatme and Sukhatme, 1970, p. 256)	
	I	II
$\bar{y}$:	Number of villages in the circles.	Area (in acres) under wheat crop in the circles.
ϕ:	A circle consisting more than five villages.	A circle consisting more than five villages.
N	89	89
n	23	23
$\bar{Y}$	3.360	1102
ρ_{pb}	0.766	0.643
C_y	0.604	0.65405
C_p	2.19012	2.19012
K_p	0.21125	0.19202

Table 2. Ranges of α in which $t_{(\alpha)}$ is more efficient than $\bar{y}$, t_1 and t_3 for population data sets I and II.

	Range of α	
Previous estimator	I	II
$\bar{y}$ (usual unbiased)	(0.00, 0.8450)	(0.00, 0.7681)
t_1	(-1.1550, 2.00)	(-1.2319, 2.00)
t_3 (Singh et al., 2007)	(-0.1550, 1.00)	(-0.2319, 1.00)

Table 2 exhibits that when $\alpha \in$(0.00, 0.8450)[0.00, 0.7681], $\alpha \in$(-1.1550, 2.00)[-1.2319, 2.00], $\alpha \in$(-0.1550, 1.00) [-0.2319, 1.00] the proposed class of estimators $t_{(\alpha)}$ is more efficient than the estimators $\bar{y}$, t_1 and t_3 in the data set I [II] respectively. The common range of the scalar α is (0.00, 0.8450) [0.00, 0.7681] in which $t_{(\alpha)}$ is to be superior than $\bar{y}$, t_1 and t_3 for data set I [II].

It is observed from Table 3 that the performance of proposed class of estimators $t_{(\alpha)}$ is better than the estimators $\bar{y}$, t_1 and t_3, if α lies between the corresponding range of the previous estimators which we have calculated in Table 2 for both the population data set I and II. Further it is observed from Table 3 that there is larger gain in efficiency by using the proposed class of estimators $t_{(\alpha)}$ over usual unbiased estimator $\bar{y}$, Naik and Gupta's (1996) ratio estimator t_1 and Singh et al.'s (2007) ratio exponential estimator t_3 in certain range of α. The maximum gain in efficiency is seen at the optimum value α_o=0.4225 (for population data set I) and α_o=0.3840 (for population data set II) of scalar α. The proposed class of estimators $t_{(\alpha)}$ is more efficient than Naik and Gupta's (1996) ratio estimator t_1 for wider range of α as compared to the usual unbiased estimator $\bar{y}$ and Singh et al.'s (2007) ratio exponential estimator t_3 in both the population data sets I and II. Further it is observed from Table 3 and Figure 1 that the magnitude of the quantity $B_{(\alpha)}$ is very low for negative values of α [-1.25$\leq \alpha \leq$-0.1550] and it is zero when α assumes the values -1.1550 and 0, for data set I, and -1.2319 and 0, for data set II. The magnitude of the quantity $B_{(\alpha)}$ increases as α increases from zero. There is slow increase in magnitude of $B_{(\alpha)}$ when $\alpha \in$ (0.00, 0.8450) while it increases in speedy manner when α goes beyond 0.8450. Thus we conclude that the proposed class of estimators is less biased or almost unbiased if α lies between -1.25 and 0.00, for both the population data sets. So if the primary concern of the study is not to obtain less biased estimators one should not pick such values of α. However this conclusion should not be extrapolated due to limited empirical study.

3. Two Phase Sampling

When the value of population proportion p is unknown, we usually apply two-phase (or double) sampling design to obtain a better estimator of the population mean $\bar{Y}$ of the study variable y. Let $\hat{p}'$ denote the proportion of units possessing attribute ϕ in the first

Table 3. PRE of $t_{(\alpha)}$ with respect to $\bar{y}$, t_1 and t_3 for different value of α in population data sets I and II.

	PRE $(t_{(\alpha)}, \bar{y})$		PRE $(t_{(\alpha)}, t_1)$		PRE $(t_{(\alpha)}, t_3)$		$B_{(\alpha)}$	
α	I	II	I	II	I	II	I	II
-1.25	10.41	12.39	89.44	97.96	15.71	20.44	0.12	0.02
-1.2319	10.63	12.65	91.32	100.00	16.04	20.87	0.09	0.00
-1.1550	11.64	13.84	100.00	109.41	17.57	22.83	0.00	0.09
-1.00	14.16	16.79	121.64	132.74	21.37	27.70	0.15	0.23
-0.75	20.28	23.86	174.23	188.63	30.61	39.37	0.30	0.36
-0.50	31.15	36.01	267.65	284.68	47.02	59.41	0.33	0.37
-0.25	52.64	58.36	452.31	461.43	79.45	96.30	0.23	0.25
-0.2319	54.92	60.60	471.91	479.17	82.90	100.00	0.21	0.23
-0.1550	66.25	71.37	569.27	564.32	100.00	117.77	0.15	0.17
0.00	100.00	100.00	859.30	790.66	150.95	165.01	0.00	0.00
0.25	195.67	157.01	1681.42	1241.37	295.37	259.07	0.35	0.37
0.3840	239.17	170.49	2055.17	1347.97	361.02	281.32	0.59	0.62
0.4225	241.99	169.29	2079.40	1338.47	365.28	279.33	0.67	0.70
0.50	230.95	160.19	1984.59	1266.59	348.62	264.33	0.83	0.87
0.75	130.58	103.95	1122.10	821.92	197.11	171.53	1.43	1.49
0.7681	124.09	100.00	1066.34	790.64	187.32	165.01	1.48	1.54
0.8450	100.00	84.59	859.30	668.82	150.95	139.58	1.69	1.75
1.00	66.25	60.60	569.27	479.16	100.00	100.00	2.15	2.23
1.25	37.54	37.19	322.55	294.08	56.66	61.37	3.01	3.10
1.50	23.64	24.52	203.17	193.91	35.69	40.47	3.98	4.10
1.75	16.11	17.19	138.47	135.93	24.32	28.37	5.08	5.22
2.00	11.64	12.65	100.00	100.00	17.57	20.87	6.31	6.46
2.25	8.78	9.66	75.44	76.42	13.25	15.95	7.66	7.83

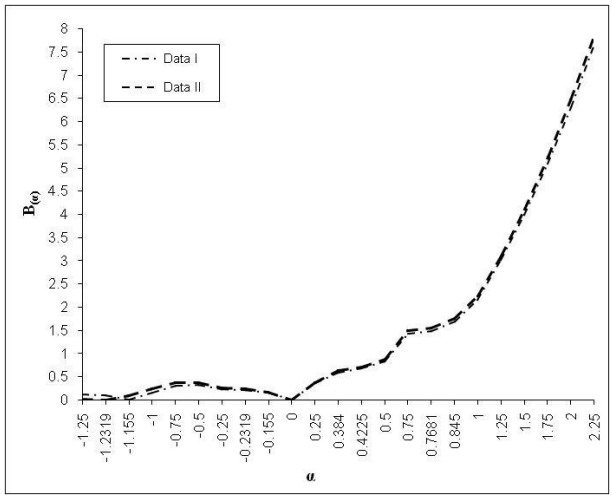

Figure 1. Effect of scalar α on the quantity $B_{(\alpha)}$ in population data sets I and II.

phase sample of size n', $\hat{p}$ denote the proportion of units possessing attribute ϕ in the second phase sample of size $n < n'$ and $\bar{y}$ denote the mean of the study variable y in the second phase sample, under two phase sampling design. In practice, the information of $\hat{p}'$ can be obtained with a little additional cost.

Thus, when the population proportion p is unknown, the usual double sampling ratio and product estimators for population mean $\bar{Y}$ are respectively defined by

$$t_{d_1} = \bar{y}\left(\frac{\hat{p}'}{\hat{p}}\right),$$

$$t_{d_2} = \bar{y}\left(\frac{\hat{p}}{\hat{p}'}\right).$$

Singh et al. (2007) suggested the ratio and product type exponential estimators in two phase for $\bar{Y}$ respectively as

$$t_{d_3} = \bar{y}\exp\left(\frac{\hat{p}'-\hat{p}}{\hat{p}'+\hat{p}}\right),$$

$$t_{d_4} = \bar{y}\exp\left(\frac{\hat{p}-\hat{p}'}{\hat{p}+\hat{p}'}\right).$$

In the following section we have suggested a class of estimators in two phase sampling (i.e. two phase sampling version of the estimator $t_{(\alpha)}$) for estimating population mean $\bar{Y}$ of the study variable y with its properties under large sample approximation. The following new notations will be used in remaining sections of the paper:

$$\theta' = \left(\frac{1}{n'}-\frac{1}{N}\right),\ \theta^* = \left(\frac{1}{n}-\frac{1}{n'}\right) \text{ and } \theta^\bullet = \left(\frac{1}{n}+\frac{1}{n'}-\frac{2}{N}\right).$$

4. Estimation of Mean in Two Phase Sampling

4.1 The suggested class of estimators when population proportion is unknown

We suggest a double sampling version of the proposed class of estimators $t_{(\alpha)}$ in Equation (1) as

$$t_{d_{(\alpha)}} = \bar{y}\exp\left(\frac{\alpha\left(\hat{p}'-\hat{p}\right)}{\hat{p}'+\hat{p}}\right). \tag{9}$$

It is to be mentioned that

(1) for $\alpha = 0, t_{d_{(\alpha)}} = t_{d_{(0)}} = \bar{y}$, (usual unbiased estimator),
(2) for $\alpha = 1, t_{d_{(\alpha)}} = t_{d_{(1)}} = t_{d_3} = \bar{y}\exp\left(\frac{\hat{p}'-\hat{p}}{\hat{p}'+\hat{p}}\right)$, (Singh et al., 2007),
(3) for $\alpha = -1, t_{d_{(\alpha)}} = t_{d_{(-1)}} = t_{d_4} = \bar{y}\exp\left(\frac{\hat{p}-\hat{p}'}{\hat{p}+\hat{p}'}\right)$, (Singh et al., 2007).

To obtain the bias and MSE of $t_{d_{(\alpha)}}$ we define

$$\bar{y} = \bar{Y}\left(1+e_0\right),\ \hat{p} = p\left(1+e_1\right),\ \hat{p}' = p\left(1+e_1'\right),$$

such that

$$\mathrm{E}\left[e_0\right] = \mathrm{E}\left[e_1\right] = \mathrm{E}\left[e_1'\right] = 0,$$

and

$$\mathrm{E}\left[e_0^2\right] = \theta C_y^2,\ \mathrm{E}\left[e_1^2\right] = \theta C_p^2,\ \mathrm{E}\left[e_1^{'2}\right] = \theta' C_p^2,$$

$$\mathrm{E}\left[e_0 e_1\right] = \theta \rho_{pb} C_y C_p,\ \mathrm{E}\left[e_0 e_1^{'}\right] = \theta \rho_{pb} C_y C_p \text{ and } \mathrm{E}\left[e_1 e_1^{'}\right] = \theta' C_p^2.$$

Now expressing $t_{d_{(\alpha)}}$ in Equation (9) in terms of e's we have

$$\begin{aligned} t_{d_{(\alpha)}} &= \bar{Y}\left(1+e_0\right)\exp\left(\alpha\frac{\left(e_1^{'}-e_1\right)}{\left(2+e_1^{'}+e_1\right)}\right) \\ &= \bar{Y}\left(1+e_0\right)\exp\left(\frac{\alpha}{2}\left(e_1^{'}-e_1\right)\left\{1+\left(\frac{e_1+e_1^{'}}{2}\right)\right\}^{-1}\right) \\ &= \bar{Y}\left[1+e_0-\frac{\alpha}{2}\left(e_1-e_1^{'}\right)-\frac{\alpha}{2}\left(e_0e_1-e_0e_1^{'}\right)+\frac{\alpha}{4}\left(e_1^2-e_1^{'2}\right)+\frac{\alpha^2}{8}\left(e_1-e_1^{'}\right)^2-\cdots\right]. \end{aligned}$$

Neglecting terms of e's having power greater than two, we have from the above expression that

$$t_{d_{(\alpha)}} \approxeq \bar{Y}\left[1+e_0-\frac{\alpha}{2}\left(e_1-e_1^{'}\right)-\frac{\alpha}{2}\left(e_0e_1-e_0e_1^{'}\right)+\frac{\alpha}{4}\left(e_1^2-e_1^{'2}\right)+\frac{\alpha^2}{8}\left(e_1-e_1^{'}\right)^2\right],$$

or

$$\left(t_{d_{(\alpha)}}-\bar{Y}\right) = \bar{Y}\left[e_0-\frac{\alpha}{2}\left(e_1-e_1^{'}\right)-\frac{\alpha}{2}\left(e_0e_1-e_0e_1^{'}\right)+\frac{\alpha}{4}\left(e_1^2-e_1^{'2}\right)+\frac{\alpha^2}{8}\left(e_1-e_1^{'}\right)^2\right]. \tag{10}$$

Taking the expectation of both sides of Equation (10), we get the bias of $t_{d_{(\alpha)}}$ to the first degree of approximation as

$$B\left(t_{d_{(\alpha)}}\right) = \bar{Y}\theta^*\left(\frac{\alpha C_p^2}{8}\right)\left(\alpha-4K_p+2\right). \tag{11}$$

It is to be mentioned that the biases of estimators t_{d_1}, t_{d_2}, t_{d_3} and t_{d_4} can be easily obtained from Equation (11) just by taking α = 2, -2, 1 and -1 respectively.

Squaring both sides of Equation (10) and neglecting terms of e's having power greater than two we have

$$\left(t_{d_{(\alpha)}}-\bar{Y}\right)^2 \approxeq \bar{Y}^2\left[e_0^2+\frac{\alpha^2}{4}\left(e_1-e_1^{'}\right)^2-\alpha\left(e_0e_1-e_0e_1^{'}\right)\right]. \tag{12}$$

Taking the expectation of both sides of Equation (12), we get the MSE of $t_{d_{(\alpha)}}$ to the first degree of approximation as

$$\mathrm{MSE}\left(t_{d_{(\alpha)}}\right) = \bar{Y}^2\left[\theta C_y^2+\theta^*\frac{\alpha C_p^2}{4}\left(\alpha-4K_p\right)\right]. \tag{13}$$

It is to be mentioned that the MSEs of the estimators $\bar{y}$, t_{d_1}, t_{d_2}, t_{d_3} and t_{d_4} can be easily obtained from Equation (13) just by taking α= 0, 2, -2, 1 and -1 respectively.

4.2 Bias comparisons

In this subsection, we have established the conditions under which the suggested class of estimators $t_{d_{(\alpha)}}$ is less biased than the estimators t_{d_1}, t_{d_2}, t_{d_3} and t_{d_4}. From Equation (11) we have

(1) $\left|B\left(t_{d_{(\alpha)}}\right)\right| < \left|B\left(t_{d_1}\right)\right|$ if $\left|\frac{\alpha}{8}\left(\alpha - 4K_p + 2\right)\right| < \left|\left(1 - K_p\right)\right|$.
(2) $\left|B\left(t_{d_{(\alpha)}}\right)\right| < \left|B\left(t_{d_2}\right)\right|$ if $\left|\frac{\alpha}{8}\left(\alpha - 4K_p + 2\right)\right| < \left|K_p\right|$.
(3) $\left|B\left(t_{d_{(\alpha)}}\right)\right| < \left|B\left(t_{d_3}\right)\right|$ if $\left|\alpha\left(\alpha - 4K_p + 2\right)\right| < \left|\left(3 - 4K_p\right)\right|$.
(4) $\left|B\left(t_{d_{(\alpha)}}\right)\right| < \left|B\left(t_{d_4}\right)\right|$ if $\left|\alpha\left(\alpha - 4K_p + 2\right)\right| < \left|\left(4K_p - 1\right)\right|$.

4.3 Efficiency comparisons

Once again in this subsection we have obtained the conditions under which the suggested class of estimators $t_{d_{(\alpha)}}$ is more efficient than the estimators $\bar{y}$, t_{d_1}, t_{d_2}, t_{d_3} and t_{d_4}. From Equation (13) we have

(1) $\mathrm{MSE}\left(t_{d_{(\alpha)}}\right) < \mathrm{Var}\left[\bar{y}\right]$ if $\min\left(0, 4K_p\right) < \alpha < \max\left(0, 4K_p\right)$.
(2) $\mathrm{MSE}\left(t_{d_{(\alpha)}}\right) < \mathrm{MSE}\left(t_{d_1}\right)$ if $\min\left(2, 2\left(2K_p - 1\right)\right) < \alpha < \max\left(2, 2\left(2K_p - 1\right)\right)$.
(3) $\mathrm{MSE}\left(t_{d_{(\alpha)}}\right) < \mathrm{MSE}\left(t_{d_2}\right)$ if $\min\left(-2, 2\left(1 + 2K_p\right)\right) < \alpha < \max\left(-2, 2\left(1 + 2K_p\right)\right)$.
(4) $\mathrm{MSE}\left(t_{d_{(\alpha)}}\right) < \mathrm{MSE}\left(t_{d_3}\right)$ if $\min\left(1, \left(4K_p - 1\right)\right) < \alpha < \max\left(1, \left(4K_p - 1\right)\right)$.
(5) $\mathrm{MSE}\left(t_{d_{(\alpha)}}\right) < \mathrm{MSE}\left(t_{d_4}\right)$ if $\min\left(-1, \left(4K_p + 1\right)\right) < \alpha < \max\left(-1, \left(4K_p + 1\right)\right)$.

4.4 Optimum choice of the scalar α

The MSE of $t_{d_{(\alpha)}}$ is minimized if we assume

$$\alpha = 2K_p = \alpha_o. \tag{14}$$

Thus the resulting minimum MSE of $t_{d_{(\alpha)}}$ is given by

$$\mathrm{MSE}_{\min}\left(t_{d_{(\alpha)}}\right) = \left(\theta - \theta^* \rho_{pb}^2\right) S_y^2 = \theta\left(1 - \rho_{pb}^2\right) S_y^2 + \theta' \rho_{pb}^2 S_y^2. \tag{15}$$

Substitution of Equation (14) in Equation (9), yields the optimum estimator (OE) in the class of estimators $t_{d_{(\alpha)}}$ as

$$t_{d_{(\alpha_o)}} = \bar{y} \exp\left(\frac{\alpha_o\left(\hat{p}' - \hat{p}\right)}{\hat{p}' + \hat{p}}\right),$$

with MSE

$$\mathrm{MSE}\left(t_{d_{(\alpha_o)}}\right) = \mathrm{MSE}_{\min}\left(t_{d_{(\alpha)}}\right) = \theta\left(1 - \rho_{pb}^2\right) S_y^2 + \theta' \rho_{pb}^2 S_y^2.$$

The value K_p can be made known in practice from the past data or experience gathered in due course of time. If K_p can not be made known then it is worth advisable to replace it by its estimate $\hat{K}_p$ as given in Equation (8). Thus the resulting estimator based on estimated optimum value $\hat{K}_p$ is given by

$$t_{d_{(\hat{\alpha}_o)}} = \bar{y} \exp\left(\frac{\hat{\alpha}_o\left(\hat{p}' - \hat{p}\right)}{\hat{p}' + \hat{p}}\right) = \bar{y} \exp\left(\frac{2\hat{K}_p\left(\hat{p}' - \hat{p}\right)}{\hat{p}' + \hat{p}}\right).$$

To the first degree of approximation, it can be shown that

$$\mathrm{MSE}\left(t_{d_{(\hat{\alpha}_o)}}\right) = \mathrm{MSE}_{\min}\left(t_{d_{(\alpha)}}\right) = \mathrm{MSE}\left(t_{d_{(\alpha_o)}}\right).$$

If the second phase sample is independent of the first phase sample then the bias and mean square error of the proposed class of estimators $t_{d_{(\alpha)}}$ are, respectively, given by

$$B^*\left(t_{d_{(\alpha)}}\right) = \left(\frac{\bar{Y}\alpha}{8}\right) C_p^2 \left[2\theta^* - 4\theta K_p + \alpha\theta^\bullet\right],$$

$$\mathrm{MSE}^*\left(t_{d_{(\alpha)}}\right) = \left[\theta C_y^2 + \left(\frac{\alpha C_p^2}{4}\right)(\alpha\theta^\bullet - 4\theta K_p)\right]. \tag{16}$$

The MSE^* of $t_{d_{(\alpha)}}$ in Equation (16) is minimized for

$$\alpha = \frac{2n^{'}(N-n)K_p}{(Nn^{'} + nN - 2nn^{'})} = \frac{2\theta K_p}{\theta^\bullet} = \alpha_o^*. \tag{17}$$

Thus the resulting minimum MSE of $t_{d_{(\alpha)}}$ in Equation (16) is given by

$$\mathrm{MSE}^*_{\min}\left(t_{d_{(\alpha)}}\right) = \theta S_y^2\left[1 - \frac{\theta}{\theta^\bullet}\rho_{pb}^2\right].$$

If the value of K_p is not known and also the close prior is not available then it is worth advisable to replace it by its estimate $\hat{K}_p$. Thus the estimate of the optimum value α at Equation (17) is

$$\hat{\alpha}_o^* = \frac{2\theta\hat{K}_p}{\theta^\bullet}.$$

Substitution of $\hat{\alpha}_o^*$ in place of α in the class of estimators $t_{d_{(\alpha)}}$ given in Equation (9), yields the estimator based on estimated optimum value $\hat{\alpha}_o^*$ as

$$t_{d_{(\hat{\alpha}_o^*)}} = \bar{y}\exp\left(\frac{2\theta\hat{K}_p}{\theta^\bullet}\frac{(\hat{p}^{'} - \hat{p})}{(\hat{p}^{'} + \hat{p})}\right).$$

It can be shown to the first degree of approximation that

$$\mathrm{MSE}^*\left(t_{d_{(\hat{\alpha}_o^*)}}\right) = \theta S_y^2\left[1 - \frac{\theta}{\theta^\bullet}\rho_{pb}^2\right] = \mathrm{MSE}^*_{\min}\left(t_{d_{(\alpha)}}\right). \tag{18}$$

From Equations (15) and (18) we have

$$\mathrm{MSE}\left(t_{d_{(\hat{\alpha}_o)}}\right) - \mathrm{MSE}^*\left(t_{d_{(\hat{\alpha}_o^*)}}\right) = S_y^2\rho_{pb}^2\frac{\theta^{'2}}{\theta^\bullet} \geq 0. \tag{19}$$

It is follows from Equation (19) that the proposed estimator $t_{d_{(\hat{\alpha}_o^*)}}$, which is based on independent sample, is more efficient than the estimator $t_{d_{(\hat{\alpha}_o)}}$, where the second phase sample is a subsample of the first phase sample.

4.5 Empirical study

To judge the merits of the proposed estimator $t_{d_{(\alpha)}}$, we have computed the PREs of $t_{d_{(\alpha)}}$ with respect to $\bar{y}$, t_{d_1} and t_{d_3} for the two population data sets which is summarized in Table 4.

We have also calculated the ranges of α in which proposed class of estimators $t_{d_{(\alpha)}}$ is more efficient than the estimators $\bar{y}$, t_{d_1} and t_{d_3} using the conditions which we have obtained in Subsection 4.3. The findings are summarized in Tables 5 and 6.

Table 4. Description of population data sets.

	Population data sets (Source: Sukhatme and Sukhatme, 1970, p. 256)	
	I	II
$\bar{y}$:	Number of villages in the circles.	Area (in acres) under wheat crop in the circles.
ϕ:	A circle consisting more than five villages.	A circle consisting more than five villages.
N	89	89
n	23	23
$n^{'}$	45	45
$\bar{Y}$	3.360	1102
ρ_{pb}	0.766	0.643
C_y	0.604	0.65405
C_p	2.19012	2.19012
K_p	0.21125	0.19202

Table 5. Ranges of α in which $t_{d_{(\alpha)}}$ is more efficient than $\bar{y}$, t_{d_1} and t_{d_3} for population data sets I and II.

	Range of α	
Previous estimator	I	II
$\bar{y}$ (usual unbiased)	(0.00, 0.8450)	(0.00, 0.7681)
t_{d_1}	(-1.1550, 2.00)	(-1.2319, 2.00)
t_{d_3} (Singh et al., 2007)	(-0.1550, 1.00)	(-0.2319, 1.00)

Table 5 demonstrates that when $\alpha \in$(0.00, 0.8450) [0.00, 0.7681], $\alpha \in$(-1.1550, 2.00) [-1.2319, 2.00], $\alpha \in$(-0.1550, 1.00) [-0.2319, 1.00] proposed class of estimators $t_{d_{(\alpha)}}$ is performing well if compared to the estimators $\bar{y}$, t_{d_1} and t_{d_3} in the data set I [II] respectively. The common range of the scalar α is (0.00, 0.8450) [0.00, 0.7681] in which $t_{d_{(\alpha)}}$ is superior than $\bar{y}$, t_{d_1} and t_{d_3} for data set I [II].

It is observed from Table 6 that the performance of the proposed class of estimators $t_{d_{(\alpha)}}$ is better than that of estimators $\bar{y}$, t_{d_1} and t_{d_3} if α follows corresponding range of inferior estimator which we have calculated in Table 5 for both data sets I and II. Further it is observed from Table 6 that there is larger gain in efficiency by using the proposed class of estimators $t_{d_{(\alpha)}}$ over usual unbiased estimator $\bar{y}$, ratio estimator t_{d_1} and Singh et al.'s (2007) ratio exponential estimator t_{d_3} in appreciable range of α. The maximum gain in efficiency is seen on the optimum value α_o=0.4225 (for population data set I) and α_o=0.3840 (for population data set II) of scalar α. The proposed class of estimators $t_{d_{(\alpha)}}$ is more efficient than the ratio estimator t_{d_1} for a wider range of α as compared to the usual unbiased estimator $\bar{y}$ and Singh et al.'s (2007) ratio exponential estimator t_{d_3} in both the population data sets I and II. Thus we have concluded that the proposed class of estimators $t_{(\alpha)}\left[t_{d_{(\alpha)}}\right]$ is performing consistently in single phase as well as in double phase sampling for these specific data sets.

Table 6. PRE of $t_{d_{(\alpha)}}$ with respect to $\bar{y}$, t_{d_1} and t_{d_3} for different value of α in population data sets I and II.

	$\text{PRE}\left(t_{d_{(\alpha)}}, \bar{y}\right)$		$\text{PRE}\left(t_{d_{(\alpha)}}, t_{d_1}\right)$		$\text{PRE}\left(t_{d_{(\alpha)}}, t_{d_3}\right)$	
α	I	II	I	II	I	II
-1.25	14.98	17.66	89.98	98.08	20.01	25.23
-1.2319	15.28	18.01	91.77	100.00	20.41	25.73
-1.1550	16.65	19.59	100.00	108.78	22.24	27.98
-1.00	20.01	23.43	120.16	130.13	26.73	33.48
-0.75	27.84	32.22	167.19	178.90	37.19	46.02
-0.50	40.69	46.05	244.40	255.70	54.36	65.78
-0.25	62.77	68.01	376.96	377.68	83.85	97.16
-0.2319	64.89	70.00	389.69	388.73	86.68	100.00
-0.1550	74.86	79.09	449.57	439.19	100.00	112.98
0.00	100.00	100.00	600.58	555.32	133.59	142.86
0.25	147.57	131.47	886.25	730.07	197.13	187.81
0.3840	162.24	137.47	974.35	763.40	216.73	196.39
0.4225	163.09	136.95	979.45	760.53	217.86	195.65
0.50	159.70	132.93	959.09	738.19	213.33	189.90
0.75	118.26	102.57	710.24	569.61	157.98	146.53
0.7681	114.68	100.00	688.73	555.32	153.20	142.86
0.8450	100.00	89.28	600.58	495.78	133.59	127.54
1.00	74.86	70.00	449.57	388.72	100.00	100.00
1.25	47.69	47.32	286.39	262.78	63.70	67.60
1.50	31.96	33.02	191.93	183.34	42.69	47.16
1.75	22.56	23.95	135.51	132.99	30.14	34.21
2.00	16.65	18.01	100.00	100.00	22.24	25.73
2.25	12.74	13.96	76.50	77.54	17.02	19.95

5. Conclusion

In this article we have considered the problem of estimating the population mean $\bar{Y}$ of the study variable y when the population proportion of an auxiliary character is known and unknown in SRSWOR. The bias and mean square error expressions of the proposed class of estimators have been obtained under large sample approximation in single phase as well as in double phase sampling. The bias and MSE expressions of usual unbiased estimator, the usual ratio estimator, the usual product estimator and Singh et al. (2007) estimators can be obtained from that of the proposed class of estimators just by putting the suitable values of the scalar. Thus the proposed class of estimators unifies several others previously defined. The realistic conditions under which the proposed class of estimators is better than the usual unbiased, ordinary ratio and product and Singh et al. (2007) estimators have been obtained in both the phases. The estimators based on estimated optimum values of the scalar have been obtained along with its approximate MSE formulae in both the phases. Numerical illustrations are given to through light on the merits of the proposed study.

Acknowledgements

Authors wish to thank the Executive Editor Víctor Leiva and referees for their helpful comments that aided in improving this article.

References

Abd-Elfattah, A.M., El-Sherpieny, E.A., Mohamed, S.M., Abdou, O.F., 2010. Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute. Applied Mathematics and Computation, 215, 4198–4202.

Bahl, S., Tuteja, R.K., 1991. Ratio and product type exponential estimator. Information and Optimization Sciences, 12, 159–163.

Jhajj, H.S., Sharma, M.K., Grover, L.K., 2006. A family of estimators of population mean using information on auxiliary attribute. Pakistan Journal of Statistics, 22, 43–50.

Kendall, M.G., Stuart, A., 1967. The Advanced Theory of Statistics. Second edition. Charles Griffin and Company Limited, London.

Naik, V.D., Gupta, P.C., 1996. A note on estimation of mean with known population proportion of an auxiliary character. Journal of the Indian Society of Agricultural Statistics, 48, 151–158.

Reddy, V.N., 1973. On ratio and product method of estimation. Sankhyā, Series B, 35, 307–316.

Reddy, V.N., 1974. On a transformed ratio method of estimation. Sankhyā, Series C, 36, 59–70.

Shabbir, J., Gupta, S., 2007. On estimating the finite population mean with known population proportion of an auxiliary variable. Pakistan Journal of Statistics, 23, 1–9.

Shabbir, J., Gupta, S., 2010. Estimation of the finite population mean in two phase sampling when auxiliary variable are attributes. Pakistan Journal of Statistics, 39, 121–129.

Singh, R., Chouhan, P., Sawan, N., Smarandache, F., 2007. Ratio-product Type Exponential Estimator for Estimating Finite Population Mean Using Information on Auxiliary Attribute. Renaissance High Press, USA. pp. 18–32.

Singh, S., 2003. Advanced Sampling Theory with Applications. How Michael "Selected" Amy. Kluwer Academic Publishers, The Netherlands.

Srivenkataramana, T., Tracy, D.S., 1980. An alternative to ratio method in sample survey. Annals of the Institute of Statistical Mathematics, 32, 111–120.

Sukhatme, P.V., Sukhatme, B.V., 1970. Sampling Theory of Surveys with Applications. Iowa State University Press, USA.