Generalized Ratio-cum-Product Estimator for Finite Population Mean under Two-Phase Sampling Scheme

Gajendra Kumar Vishwakarma
Indian Institute of Technology (ISM) Dhanbad, vishwagk@rediffmail.com

Sayed Mohammed Zeeshan
VIT Bhopal University, zeeshan008x52@gmail.com

Generalized Ratio-cum-Product Estimator for Finite Population Mean under Two-Phase Sampling Scheme

Gajendra Kumar Vishwakarma
Indian Institute of Technology (ISM) Dhanbad
Dhanbad, India

Sayed Mohammed Zeeshan
VIT Bhopal University
Bhopal, India

A method to lower the MSE of a proposed estimator relative to the MSE of the linear regression estimator under two-phase sampling scheme is developed. Estimators are developed to estimate the mean of the variate under study with the help of auxiliary variate (which are unknown but it can be accessed conveniently and economically). The mean square errors equations are obtained for the proposed estimators. In addition, optimal sample sizes are obtained under the given cost function. The comparison study has been done to set up conditions for which developed estimators are more effective than other estimators with novelty. The empirical study is also performed to supplement the claim that the developed estimators are more efficient.

Keywords: Statistical methods, correlation, point estimation, two-phase sampling, bias, MSE

Introduction

In sampling surveys, a number of sampling techniques required information about an auxiliary variate, say *X*, to increase the efficiency of the estimator (population mean) of the variate, say *Y*, the character of interest for estimation. There may be cases where such auxiliary information is not available but can be obtained relatively easily (i.e. at a comparatively low cost in terms of time and money). In such cases, it is worthwhile (suitable) to draw a relatively large sample from a population and enumerate it for the auxiliary variate *X*, and then take either an independent sample, or a sub-sample of the first sample for measuring the variable *Y*. This technique of taking samples in two phases is known as two-phase sampling. Some notable contributions in this direction were made by Cochran (1940), Robson

(1957), Murthy (1964), Singh (1967), Sahai (1979), Bahl and Tuteja (1991), Singh and Espejo (2003), Kadilar and Cingi (2005), Singh and Tailor (2005), Singh and Vishwakarma (2007, 2008), Shabbir et al. (2014), Vishwakarma and Sayed (2017), and many others.

In general, the two-phase regression estimator is more efficient than the two-phase ratio (or product) estimators except when the regression line of the variable under study on the auxiliary variable passes through the neighborhood of the origin. Most of the ratio and product type estimators recently developed are simply a modification of other existing estimators available in the literature. This has led to the accumulation of a large number of the ratio as well as product type estimators with cumbersome structure over the passage of time. Often these estimators require the knowledge of other population parameters in advance or has to guess it with the experience gathered over the period of time in sample survey or estimate it through pilot survey or the sample itself and in optimum case the MSE of the proposed estimator is found generally equivalent to the MSE of the regression estimator (in two-phase). Moving in this direction, we have proposed estimators and shown that how in optimal case its minimum MSE of the proposed estimators is more efficient than regression estimator. One more aspect of the proposed method is the important role played by the Bias of the estimator in improving MSE which was neglected before in the survey literature works in the area of ratio and product estimators.

Consider $V = (V_1, , V_2,\ldots, V_N)$ to be a finite population composed of N units and (y_j, x_j), where j = 1, 2,…, N, indicate the variates on V_j (the jth unit) of the population V. In the case when the population mean of auxiliary variate is unknown, but it can be accessed conveniently and economically, population mean of study variate is obtained using two-phase sampling method. In two-phase sampling a large preliminary sample of size n' is selected, known as the first-phase sample and then from the first-phase sample a sub-sample of size n is drawn known as the second-phase sample. Note the drawing of units from the population is done using simple random sampling without replacement (SRSWOR) scheme.

The variance of the $\bar{y}$ (usual unbiased estimator) is

$$V\left(\bar{y}\right) = f_1 \bar{Y}^2 C_y^2. \tag{1}$$

The two-phase sampling version of ratio and product estimators are:

$$\bar{y}_{\text{Rd}} = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) \tag{2}$$

3

$$\bar{y}_{Pd} = \bar{y}\left(\frac{\bar{x}}{\bar{x}'}\right) \tag{3}$$

where

$\bar{x}'$ fist-phase sample mean of auxiliary variate,

$\bar{y}$ second-phase sample mean of study variate,

$\bar{x}$ second-phase sample mean of auxiliary variate.

The MSEs of $\bar{y}_{Rd}$ and $\bar{y}_{Pd}$ are:

$$\text{MSE}\left(\bar{y}_{Rd}\right) = \bar{Y}^2\left[f_1 C_y^2 + f_3\left(C_x^2 - 2\rho_{yx}C_y C_x\right)\right], \tag{4}$$

$$\text{MSE}\left(\bar{y}_{Pd}\right) = \bar{Y}^2\left[f_1 C_y^2 + f_3\left(C_x^2 + 2\rho_{yx}C_y C_x\right)\right]. \tag{5}$$

where

$$f_1 = \left(\frac{1}{n} - \frac{1}{N}\right), f_2 = \left(\frac{1}{n'} - \frac{1}{N}\right), f_3 = f_1 - f_2 = \left(\frac{1}{n} - \frac{1}{n'}\right), C_y^2 = \frac{S_y^2}{\bar{Y}^2},$$

$$C_x^2 = \frac{S_x^2}{\bar{X}^2}, \rho_{yx} = \frac{S_{yx}}{S_y S_x}, S_y^2 = \frac{1}{(N-1)}\sum_{j=1}^{N}\left(Y_j - \bar{Y}\right)^2,$$

$$S_x^2 = \frac{1}{(N-1)}\sum_{j=1}^{N}\left(X_j - \bar{X}\right)^2, S_{yx} = \frac{1}{(N-1)}\sum_{j=1}^{N}\left(Y_j - \bar{Y}\right)\left(X_j - \bar{X}\right)$$

The two-phase sampling version of difference estimators is:

$$\bar{y}_D = \bar{y} + D\left(\bar{x} - \bar{x}'\right). \tag{6}$$

The minimum MSE of $\bar{y}_D$ is:

$$\text{MSE}\left(\bar{y}_D\right)_{\min} = \bar{Y}^2 C_Y^2\left(f_1 - f_3\rho_{yx}^2\right). \tag{7}$$

The two-phase sampling version of exponential ratio and product estimators (see Singh & Vishwarma, 2007) are:

$$t_1 = \bar{y}\exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right) \tag{8}$$

$$t_2 = \bar{y}\exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right) \tag{9}$$

The MSEs of t_1 and t_2 are:

$$\mathrm{MSE}(t_1) = \bar{Y}^2\left[f_1C_y^2 + f_3\left\{\frac{1}{4}C_x^2 - \rho_{yx}C_yC_x\right\}\right], \tag{10}$$

$$\mathrm{MSE}(t_2) = \bar{Y}^2\left[f_1C_y^2 + f_3\left\{\frac{1}{4}C_x^2 + \rho_{yx}C_yC_x\right\}\right]. \tag{11}$$

Proposed Estimators

Motivated by the above two-phase sampling scheme's estimators the following estimator is suggested:

$$T = \bar{y}\exp\left[\frac{(\bar{x}'/\bar{x})^\alpha - 1}{(\bar{x}'/\bar{x})^\alpha + 1}\right], \tag{12}$$

where α is a real constant.

Moreover, T reduces to a set of estimators $\{\bar{y}, t_1, t_2\}$ for different values of α such that $\bar{y}$ for $\alpha = 0$, t_1 for $\alpha = 1$, and t_2 for $\alpha = -1$.

In order to calculate the bias and MSE of T, let us consider

$$\left.\begin{aligned}
&\bar{y} = \bar{Y}(1+e_0), \bar{x} = \bar{X}(1+e_1), \bar{x}' = \bar{X}(1+e_1')\\
&\mathrm{E}(e_0) = \mathrm{E}(e_1) = \mathrm{E}(e') = 0\\
&\mathrm{E}(e_0^2) = f_1C_y^2, \mathrm{E}(e_1^2) = f_1C_x^2, \mathrm{E}(e_1'^2) = f_2C_x^2,\\
&\mathrm{E}(e_0e_1) = f_1\rho_{yx}C_yC_x, \mathrm{E}(e_0e_1') = f_2\rho_{yx}C_yC_x, \mathrm{E}(e_1e_1') = f_2C_x^2
\end{aligned}\right\} \tag{13}$$

Expressing Equation (12) in terms of errors up to the second degree, we obtain:

5

$$T - \bar{Y}$$

$$= \bar{Y} \left[e_0 + \frac{\alpha}{4} \left(2e_1' - 2e_1 - e_1'^2 + e_1^2 + 2e_1'e_0 - 2e_1e_0 \right) + \frac{\alpha^2}{8} \left(e_1'^2 + e_1^2 - 2e_1e_1' \right) \right] \quad (14)$$

Taking the expectations in Equation (14) the bias of T obtained is:

$$\text{Bias}(T) = f_3 \bar{Y} \left[\frac{\alpha^2}{8} C_X^2 + \frac{\alpha}{4} \left(C_X^2 - 2\rho_{yx} C_Y C_X \right) \right]. \quad (15)$$

Again, from Equation (14), by neglecting the error terms of second degree, we have:

$$T - \bar{Y} = \bar{Y} \left[e_0 + \frac{\alpha}{2} \left(e_1' - e_1 \right) \right]. \quad (16)$$

Now squaring Equation (16) and taking expectation the MSE of T is obtained as:

$$\text{MSE}(T) = \bar{Y}^2 \left(f_1 C_Y^2 - \alpha f_3 \rho_{yx} C_Y C_X + \frac{\alpha^2}{4} f_3 C_X^2 \right) \quad (17)$$

Optimal Value of α

The optimal values of α, for which the MSE of the T gets minimized, is calculated by using the following conditions:

$$\frac{\partial}{\partial \alpha} \text{MSE}(T) = 0. \quad (18)$$

On solving Equation (18), we have

$$\alpha^* = 2\rho_{yx} \frac{C_Y}{C_X}, \quad (19)$$

where α^* denote the optimal values of α. Thus, putting optimal values of α in Equation (17), we get:

$$\text{MSE}(T)_{\min} = \bar{Y}^2 C_Y^2 \left(f_1 - f_3 \rho_{yx}^2 \right). \tag{20}$$

Note: MSE(T)min = MSE($\bar{y}$D)min.

Proposed Generalized Estimator

The proposed estimator is generalized so that, in the process, the proposed generalized estimator improves on the proposed estimator T or difference estimator y_D (or regression estimator in two-phase sampling). Thus, our proposed generalized estimator is

$$T_g = k\bar{y} \exp\left[\frac{(\bar{x}'/\bar{x})^\alpha - 1}{(\bar{x}'/\bar{x})^\alpha + 1} \right], \tag{21}$$

where k is again constant so that it can be minimized further.

Now, expressing Equation (21) in terms of errors up to the second degree, we obtain

$$T_g - \bar{Y} = (k-1)\bar{Y} + k\bar{Y}\left[e_0 + \frac{\alpha}{4}\left(2e_1' - 2e_1 - e_1'^2 + e_1^2 + 2e_1'e_0 - 2e_1e_0 \right) \right. \tag{22}$$
$$\left. + \frac{\alpha^2}{8}\left(e_1'^2 + e_1^2 - 2e_1e_1' \right) \right]$$

Squaring Equation (22),

$$\left(T_g - \bar{Y} \right)^2 = (k-1)^2 \bar{Y}^2 + k^2\bar{Y}^2 \left\{ e_0 + \frac{\alpha}{2}\left(e_1' - e_1 \right) \right\}^2$$
$$+ 2k(k-1)\bar{Y}^2 \left[e_0 + \frac{\alpha}{4}\left(2e_1' - 2e_1 - e_1'^2 + e_1^2 + 2e_1'e_0 - 2e_1e_0 \right) \right. \tag{23}$$
$$\left. + \frac{\alpha^2}{8}\left(e_1'^2 + e_1^2 - 2e_1e_1' \right) \right]$$

and taking the expectation of Equation (23), the MSE of T_g is obtained as:

$$\text{MSE}(T_g) = (k-1)^2 \bar{Y}^2 + k^2 \text{MSE}(T) + 2k(k-1)\bar{Y}\,\text{Bias}(T). \tag{24}$$

7

Minimizing MSE(T_g), the optimum value of k is found by differentiating (24) with respect to k and equating the resultant equation with zero:

$$K_\text{opt} = \frac{\bar{Y}^2 + \bar{Y}\,\text{Bias}(T)}{\bar{Y}^2 + \text{MSE}(T) + 2\bar{Y}\,\text{Bias}(T)}. \tag{25}$$

Using (25) in (24),

$$\text{MSE}(T_\text{g})_\text{min} = \frac{\bar{Y}^2\left(\text{MSE}(T) - \text{Bias}(T)^2\right)}{\bar{Y}^2 + \text{MSE}(T) + 2\bar{Y}\,\text{Bias}(T)}. \tag{26}$$

The MSE(T_g) depends upon MSE(T) and Bias(T). Now for $\alpha^* = 2\rho_{yx}C_Y / C_X$ we know that MSE(T) will further attain minimum. Hence, the final minimum MSE using α^* will turn out to be

$$\text{MSE}(T_\text{g})_\text{min}\,|\,\alpha^* = \frac{\bar{Y}^2\left[\text{MSE}(T)_\text{min} - \left\{\text{Bias}(T)\,|\,\alpha^*\right\}^2\right]}{\bar{Y}^2 + \text{MSE}(T)_\text{min} + 2\bar{Y}\left\{\text{Bias}(T)\,|\,\alpha^*\right\}}, \tag{27}$$

where Bias(T) $|\,\alpha^*$ is the value of Bias(T) at $\alpha^* = 2\rho_{yx}C_Y / C_X$.

Theorem 1. The relative efficiency of the proposed generalized estimator T_g with respect to MSEs of estimator T and regression estimator in two-phase sampling or difference estimator y_dd under SRSWOR, is inversely proportional to the sample size n' and n. In other words, as $n' \to N$ and $n \to N$ the value of relative efficiency (*RE*) tends to 1, i.e. $RE \to 1$.

Proof. From the definition of relative efficiency *RE*, we get

$$RE = \frac{\text{MSE}(\square)}{\text{MSE}(T_\text{g})_\text{min}\,|\,\alpha^*},$$

where MSE($\cdot$) stands for MSE(T)$_\text{min}$ or MSE(y_dd). On simplifying we get

$$RE = \dfrac{1}{\bar{Y}^2 \left(1 - \dfrac{\text{Bias}(T) \mid \alpha^*}{\text{MSE}(\square)}\right)} \left(\bar{Y}^2 + \text{MSE}(\square) + 2\bar{Y}\,\text{Bias}(T) \mid \alpha^*\right).$$

Now as $n' \to N$ and $n \to N$ we have $f_1 \to 0$, $f_2 \to 0$, and $f_3 \to 0$. As a result, $\text{Bias}(T) \mid \alpha^* / \text{MSE}(\cdot) \to 0$, $\text{MSE}(\cdot) \to 0$, and therefore $RE \to 1$.

Efficiency Comparisons

For making efficiency comparisons of the estimators which are developed with the existing estimators discussed above in the article, we have from (1), (4), (5), (7), (10), (11), and (20):

$$\text{MSE}(T) < \text{V}(\bar{y}) \quad \text{iff} \quad \alpha < 4\rho_{xy}\frac{C_Y}{C_X}, \tag{28}$$

$$\text{MSE}(T) < \text{MSE}(\bar{y}_{\text{Rd}}) \quad \text{iff} \quad \alpha < 4\rho_{xy}\frac{C_Y}{C_X} - 2, \tag{29}$$

$$\text{MSE}(T) < \text{MSE}(\bar{y}_{\text{Pd}}) \quad \text{iff} \quad \alpha < 4\rho_{xy}\frac{C_Y}{C_X} + 2, \tag{30}$$

$$\text{MSE}(T) < \text{MSE}(t_1) \quad \text{iff} \quad \alpha < 4\rho_{xy}\frac{C_Y}{C_X} - 1, \tag{31}$$

$$\text{MSE}(T) < \text{MSE}(t_2) \quad \text{iff} \quad \alpha < 4\rho_{xy}\frac{C_Y}{C_X} + 1. \tag{32}$$

Now, in case of comparison study between $\text{MSE}(T_g)_{\min} \mid \alpha^*$ and $\text{MSE}(T)$ we get that $\text{MSE}(T) \leq \text{MSE}(T_g)_{\min} \mid \alpha^*$ only if

$$\left[\text{MSE}(T)_{\min} + \bar{Y}\,\text{Bias}(T) \mid \alpha^*\right]^2 \geq 0. \tag{33}$$

Thus, $\text{MSE}(T)$ will be always smaller than $\text{MSE}(T_g)$. Therefore, T_g will be more efficient than T because Equation (33) will always be satisfied.

9

Determination of *n'* and *n* (for Fixed Cost *c*)

Suppose c is the total cost spent in the survey to obtain information. Total cost of the survey (expected) is given as:

$$c = c_1 n' + c_2 n, \tag{34}$$

where

$c_1 = $ cost per unit spent to obtain information on first-phase sample

$c_2 = $ cost per unit spent to obtain information on second-phase sample

The expressions for the MSEs of $\bar{y}_{Rd}$, $\bar{y}_{Pd}$, $\bar{y}_{dd}$, t_1, t_2, and T can be expressed or rewritten by

$$\text{MSE}(\square) = \left[\frac{1}{n} A_1 + \frac{1}{n'} A_2 \right], \tag{35}$$

where

$$A_1 = \bar{Y}^2 \left[C_Y^2 + \frac{\alpha^2}{4} C_X^2 - \alpha \rho_{yx} C_Y C_X \right]$$

$$A_2 = \bar{Y}^2 \left[\alpha \rho_{yx} C_Y C_X - \frac{\alpha^2}{4} C_X^2 \right]$$

MSE($\cdot$) stands for MSE of $\bar{y}_{Rd}$, $\bar{y}_{Pd}$, $\bar{y}_{dd}$, t_1, t_2, and T, and A_1 and A_2 are coefficients of n^{-1} and n'^{-1}. Note: For $\alpha = 2$, -2, $2\rho_{xy}(C_Y / C_X)$, 1, and -1 become coefficients for MSE for $\bar{y}_{Rd}$, $\bar{y}_{Pd}$, $\bar{y}_D = T$, t_1, and t_2, respectively.

Let us define functions ϕ as

$$\phi = \text{MSE}(\square) + \lambda \left(c_1 n' + c_2 n - c \right), \tag{36}$$

where λ is the Lagrange's multiplier.

Partially differentiating (36) with respect to n' and n, equating it to zero, and using (33), we get the optimum n' and optimum n for the proposed estimator T as

$$n'_{\text{opt}} = \frac{c\sqrt{A_2 c_2}}{c_1\sqrt{A_2 c_2} + c_2\sqrt{A_1 c_1}} \tag{37}$$

and

$$n_{\text{opt}} = \frac{c\sqrt{A_1 c_1}}{c_1\sqrt{A_2 c_2} + c_2\sqrt{A_1 c_1}}. \tag{38}$$

Hence, on substituting (37) and (38) in (35), the optimum MSE of T is

$$\text{MSE}(T)_{\text{opt}} = \left(\frac{c_1\sqrt{A_2 c_2} + c_2\sqrt{A_1 c_1}}{c\sqrt{A_1 c_1}}\right) A_1 + \left(\frac{c_1\sqrt{A_2 c_2} + c_2\sqrt{A_1 c_1}}{c\sqrt{A_2 c_2}}\right) A_2. \tag{39}$$

Using it for MSE of T_g we get

$$\text{MSE}(T_g)_{\text{opt}} \mid \alpha^* = \frac{\bar{Y}^2 \left(\text{MSE}(T)_{\text{opt}} - \text{Bias}(T)_{\text{opt}}^2 \mid \alpha^*\right)}{\bar{Y}^2 + \text{MSE}(T)_{\text{opt}} + 2\bar{Y}\,\text{Bias}(T)_{\text{opt}} \mid \alpha^*}, \tag{40}$$

where $\text{Bias}(T)_{\text{opt}} \mid \alpha^*$ is the value of Bias(T) at $\alpha^* = 2\rho_{yx} C_Y / C_X$ and

$$f_3 = \left(\frac{1}{n_{\text{opt}}} - \frac{1}{n'_{\text{opt}}}\right).$$

Empirical Study

In order to check or confirm the gain in the efficiency of the recommended estimators, five data sets were used:

Data Set I

The source of this datasets is Dobson (1990, p. 47). The auxiliary variate X is taken as initial white blood cell count and the study variate Y is taken as survival time of a leukemia patient. The parameters given are:

$$C_X = 0.7493; C_Y = 0.2017; \rho_{xy} = -0.54709; \bar{Y} = 0.6860; N = 30; n' = 12; n = 4$$

Data Set II

The source of this datasets is Steel and Torrie (1960, p. 282). The auxiliary variate X is taken as percentage of chlorine and the study variate Y is taken as log of leaf burn in sacs. The parameters given are:

$$C_X = 0.7493; C_Y = 0.4803; \rho_{xy} = -0.4996; \bar{Y} = 0.6860; N = 30; n' = 12; n = 4$$

Data Set III

The source of this datasets is Das (1988). The auxiliary variate X is taken as number of laborers in agricultural farms for 1961 and the study variate Y is taken as number of laborers in agricultural farms for 1971. The parameters given are:

$$C_X = 1.6198; C_Y = 1.4451; \rho_{xy} = 0.7213; \bar{Y} = 39.0680; N = 278; n' = 70; n = 30$$

Data Set IV

The source of this datasets is Murthy (1967, p. 228). The auxiliary variate X is taken as number of workers and the study variate Y is taken as output. The parameters given are:

$$C_X = 0.9484; C_Y = 0.3542; \rho_{xy} = 0.9150; \bar{Y} = 5182.64; N = 80; n' = 30; n = 10$$

Table 1. Percentage relative efficiencies of various estimators of $\bar{Y}$

	Data Sets				
Estimators	**I**	**II**	**III**	**IV**	**V**
$\bar{y}$	100.00	100.00	100.00	100.00	100.00
$\bar{y}_{Rd}$	*	*	130.03	36.64	72.33
$\bar{y}_{Pd}$	14.11	59.77	*	*	*
t_1	*	*	146.34	200.42	297.97
t_2	53.02	115.14	*	*	*
$T = \bar{y}_D$	123.00	123.76	149.98	276.15	307.77
T_g	123.02	125.08	158.04	284.23	313.62

Data Set V

The source of this datasets is Murthy (1967, p. 228). The auxiliary variate X is taken as fixed capital and the study variate Y is taken as output. The parameters given are:

$$C_X = 0.7507; C_Y = 0.3542; \rho_{xy} = 0.9413; \bar{Y} = 5182.64; N = 80; n' = 30; n = 10$$

In order to check whether the suggested estimators are effective even when cost is fixed we consider two data sets.

Data Set 1

The data frame consists of as many as 70 villages (known as Tehsil) which are located in some province of India. The number of people living there, corresponding to each Tehsil, are considered as the auxiliary variate and the area under cultivation (in acres) is taken as the study variate. The parameters required are (Srivastava, 1993):

$$C_X = 0.8009; C_Y = 0.6253; \rho_{xy} = 0.7780; \bar{X} = 1755.53; \bar{Y} = 981.29;$$
$$N = 70; n = 25; n' = 40$$

Table 2. Mean square error of various estimators of $\bar{Y}$

Data set	Estimators	n_{opt}	n'_{opt}	MSE(·)
		$c_2 = 10$, $c_2 = 50$, $c_0 = 1025$ (fixed)		
1	$\hat{\bar{Y}}_{Rd}$	24.81	40.93	7690.23
	$\hat{\bar{Y}}_{ReMd}$	21.53	57.32	5705.54
	T	21.23	58.80	5494.97
	T_g	21.23	58.80	5433.86
2	$\hat{\bar{Y}}_{Pd}$	5.00	6.50	0.3598
	$\hat{\bar{Y}}_{PeMd}$	5.57	5.06	0.3674
	T	4.77	7.05	0.3491
	T_g	4.77	7.05	0.3478

Data Set 2

For demonstrating the performance in the case when correlation is negative, the data set is sourced from Maddala (1977, p. 282). The auxiliary variate X is taken as the deflated prices of veal and the study variate Y is taken as the consumption per capita. The parameters required of the data set are:

$$C_X = 0.1645; C_Y = 0.3901; \rho_{xy} = -0.508; \bar{X} = 75.79; \bar{Y} = 61.59;$$
$$N = 56; n = 15; n' = 25$$

From Table 2, it can be concluded that T and T_g perform better than other estimators with novelty.

Conclusion

The present study deals with the estimation of unknown mean $\bar{Y}$ under SRSWOR in two-phase sampling scheme. From Table 1, it is observed that for all the populations, the PREs of T and T_g are more than that of $\bar{y}$, $\bar{y}_{Rd}$, $\bar{y}_{Pd}$, $\bar{y}_{dd}$, t_1, and t_2. In addition, T and T_g increase efficiency in comparison to $\bar{y}$, $\bar{y}_{Rd}$, $\bar{y}_{Pd}$, $\bar{y}_{dd}$, t_1, and t_2 for fixed cost $c < c_0$. Hence, T and T_g should be preferred in practice. It can be seen that T_g is more efficient than the difference estimator or two-phase regression estimator. It can also be extended to stratified sampling, two-stage sampling and other sampling designs or sampling scheme when study variate is contaminated with non-response. Also, this idea can be extended to multi-auxiliary variate for any sampling design. So, the proposed estimators T and T_g outperform the other existing estimators of the sampling literature, and hence can be preferred for practical applications.

References

Bahl, S., & Tuteja, R. K. (1991). Ratio and product type exponential estimator. *Information and Optimization Sciences, 12*(1), 159-163. doi: 10.1080/02522667.1991.10699058

Cochran, W. G. (1940). The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science, 30*(2), 262-275. doi: 10.1017/S0021859600048012

Das, A. K. (1988). *Contribution to the theory of sampling strategies based on auxiliary information* [Unpublished doctoral dissertation]. Bidhan Chandra Agricultural University.

Dobson, A. J. (1990). *An introduction to generalized linear models*. New York: Chapman & Hall.

Kadilar, C., & Cingi, H. (2005). A new estimator using two auxiliary variables. *Applied Mathematics and Computation, 162*(2), 901-908. doi: 10.1016/j.amc.2003.12.130

Maddala, G. S. (1977). *Econometrics*. New York: McGraw Hill.

Murthy, M. N. (1964). Product method of estimation. *Sankhyā: The Indian Journal of Statistics, Series A, 26*(1), 69-74.

Murthy, M. N. (1967). *Sampling theory and methods*. Calcutta, India: Statistical Publishing Society.

Robson, D. S. (1957). Application of multivariate polykays to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association, 52*(280), 511-522. doi: 10.1080/01621459.1957.10501407#.XKGTtO8gTuc

Sahai, A. (1979). An efficient variant of the product and ratio estimators. *Statistica Neerlandica, 33*(1), 27-35. doi: 10.1111/j.1467-9574.1979.tb00659.x

Shabbir, J., Haq, A., & Gupta, S. (2014). A new difference-cum-exponential type estimator of finite population mean in simple random sampling. *Revista Colombiana de Estadistica, 37*(1), 199-211. doi: 10.15446/rce.v37n1.44366

Singh, H. P., & Espejo, M. R. (2003). On linear regression and ratio-product estimation of a finite population mean. *Journal of the Royal Statistical Society: Series D (The Statistician), 52*(1), 59-67. doi: 10.1111/1467-9884.00341

Singh, H. P., & Tailor, R.(2005). Estimation of finite population mean using known correlation coefficient between auxiliary characters. *Statistica, 65*(4), 407-418. doi: 10.6092/issn.1973-2201/1300

Singh, H. P., & Vishwakarma, G. K.(2007). Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian Journal of Statistics, 36*(3), 217-225. doi: 10.17713/ajs.v36i3.333

Singh, H. P., & Vishwakarma, G. K. (2008). Some estimators of finite population mean using auxiliary information in sample surveys. *Journal of Applied Statistical Sciences, 16*(4), 11-22.

Singh, M. P. (1967). Ratio-cum-product method of estimation. *Metrika, 12*, 34-42. doi: 10.1007/BF02613481

Srivastava, S. (1993). *Some problems on the estimation of population mean using auxiliary character in presence of non-response in sample surveys* [Unpublished doctoral thesis]. Banaras Hindu University, Varanasi.

Steel, R. G. D., & Torrie, J. H.(1960). *Principles and procedures of statistics*. New York: McGraw Hill.

Vishwakarma, G. K., & Sayed, M. Z. (2017). A class of ratio-cum-product type of exponential estimators under simple random sampling. *The Philippine Statistician, 66*(1), 41-48.