11-1-2016

Efficient and Unbiased Estimation Procedure of Population Mean in Two-Phase Sampling

Reba Maji
Indian School of Mines, Dhanbad, India, rebamaji09@gmail.com

Arnab Bandyopadhyay
Asansol Engineering College, Asansol, India, arnabbandyopadhyay4@gmail.com

G. N. Singh
Indian School of Mines, Dhanbad, India, gnsingh_ism@yahoo.com

Efficient and Unbiased Estimation Procedure of Population Mean in Two-Phase Sampling

Reba Maji
Indian School of Mines
Dhanbad, India

Arnab Bandyopadhyay
Asansol Engineering College
Asansol, Inida

G. N. Singh
Indian School of Mines
Dhanbad, India

In this paper, an unbiased regression-ratio type estimator has been developed for estimating the population mean using two auxiliary variables in double sampling. Its properties are studied under two different cases. Empirical studies and graphical simulation have been done to demonstrate the efficiency of the proposed estimator over other estimators.

Keywords: Double sampling, study variable, auxiliary variable, chain-type, regression, bias, variance, efficiency

Introduction

The use of supplementary information on auxiliary variable for estimating the finite population mean of the variable under study has played an eminent role in sampling theory and practices. Auxiliary information may be truthfully utilized at the planning, design, and estimation stages to develop improved estimation procedures in sample surveys. Ratio, product, and regression methods of estimation are good examples in this context. Use of auxiliary information at the estimation stage was introduced during the 1930's with a comprehensive theory provided by Cochran (1940). Sometimes, information on auxiliary variable may be readily available for all the units of a population; for example, tonnage (or seat capacity) of each vehicle or ship is known in survey sampling of transportation, and number of beds available in different hospitals may be known well in advance in health care surveys. If such information is lacking, it is sometimes relatively cheap to take a large preliminary sample where an auxiliary variable alone is measured. Such practice is applicable in two-phase (or double) sampling. Two-phase sampling happens to be a powerful and cost-effective (economical) technique to generate reliable estimates of the unknown population parameters of the auxiliary variables in a first phase sample.

Reba Maji is in the Department of Applied Mathematics. Email at:
rebamaji09@gmail.com.

In order to construct an efficient estimator of the population mean of the auxiliary variable in a first-phase (preliminary) sample, Chand (1975) introduced the technique of chaining another auxiliary variable with the first auxiliary variable by using the ratio estimator in the first phase sample. This estimator is known as the chain-type ratio estimator. This work was further extended by Kiregyera (1980; 1984), Sahoo and Sahoo (1993), Tracy, Singh, and Singh (1996), Singh and Espejo (2007), Gupta and Shabbir (2007), Dash and Mishra (2011), Shukla, Pathak, and Thakur (2012), and Choudhury and Singh (2012), among others, who proposed various chain-type ratio and regression estimators. It may be noted that the most of these estimation procedures of the population mean in two-phase sampling are biased which becomes a serious drawback for their practical applications.

Encouraged and fascinated with the work discussed earlier, we have proposed an unbiased regression-ratio type estimator of the population mean and studied its properties under two different structures of two-phase sampling. Performances of the proposed estimator have been examined through empirical and graphical means of comparisons. Suitable recommendations to the survey statistician are made.

Methodology

Sample Structure and Some Existing Estimation Procedures

Let y_k, x_k, and z_k be the values of the study variable y, first auxiliary variable x, and second auxiliary variable z, respectively, associated with the kth unit of the finite population U = $(U_1, U_2, U_3,…, U_N)$. The intent is to estimate the population mean $\bar{Y}$ of the study variable y in the presence of auxiliary variables x and z when the population mean $\bar{X}$ of x is unknown but information on z is readily available for all the units of population.

To estimate $\bar{Y}$, a first-phase sample S' (S' $\subset$ U) of size n is drawn via a simple random sampling without replacement (SRSWOR) scheme from the entire population U and observed for the auxiliary variable x to furnish the estimate of $\bar{X}$. Next, a second-phase sample S of size m ($m \leq n$) is drawn by SRSWOR according to the following rules to observe the study variable y:

Case I: Second-phase sample is drawn as a subsample of the first-phase sample

Case II: Second-phase sample is drawn independently of the first-phase sample

The case where the second sample is drawn independent of the first was considered by Bose (1943).

In the sections below, we use the following notations:

$\bar{x}_m, \bar{x}_n, \bar{y}_m, \bar{z}_m, \bar{z}_n$: Sample mean of the respective variables of the sample sizes shown in subscripts.

$\bar{X}, \bar{Y}, \bar{Z}$: Population mean of x, y, and z, respectively.

$\rho_{yx}, \rho_{yz}, \rho_{xz}$: Correlation coefficient between the variables shown in subscripts.

C_x, C_y, C_z: Coefficient of variance of x, y, and z respectively.

S_{yz}: Population covariance between y and z.

S_z^2: Population mean square of z.

$s_{yz}(m)$: Sample covariance between y and z based on the sample of size m.

$s_z^2(m)$: Sample mean square of z based on the sample of size m.

β_{yz}: Population regression coefficient between the variables y and z.

$b_{xz}(n), b_{yz}(m), b_{yx}(m)$: Sample regression coefficient between the variables shown in subscripts and based on samples of the size indicated in braces.

To estimate the population mean $\bar{Y}$, the classical ratio estimator is presented as

$$\bar{y}_r = \frac{\bar{y}_m}{\bar{x}_m} \bar{X} \tag{1}$$

If $\bar{X}$ is unknown, we estimate $\bar{Y}$ under the two-phase sampling set up as

$$t_1 = \frac{\bar{y}_m}{\bar{x}_m} \bar{x}_n \tag{2}$$

S. K. Srivastava (1971) generalized the ratio method of estimation, and its structure in two-phase sampling is given as

$$t_2 = \bar{y}_m \left(\frac{\bar{x}_n}{\bar{x}_m} \right)^\alpha \tag{3}$$

where α is a real scalar which can be suitably determined by minimizing the mean square error (MSE) of the estimator.

The way in which the estimate of $\bar{Y}$ is improved using the auxiliary information on x can also be extended to improve the estimate of $\bar{X}$ in the first-phase sample if another auxiliary variable, z, closely related to x but remotely related to y is used. Thus, assuming that the population mean of the auxiliary variable z is known, Chand (1975) proposed a chain-type ratio estimator as

$$t_3 = \frac{\bar{y}_m}{\bar{x}_m} \frac{\bar{x}_n}{\bar{z}_n} \bar{Z} \tag{4}$$

Similarly, for negative correlation between the variables y and x, the chain-type product estimator is defined as

$$t_4 = \bar{y}_m \frac{\bar{x}_m}{\bar{x}_n} \frac{\bar{z}_n}{\bar{Z}} \tag{5}$$

Kiregyera (1984) suggested the chain linear regression estimator in double sampling as

$$t_5 = \bar{y}_m + b_{yx}(m)\left[\bar{x}_n + b_{xz}(n)(\bar{Z} - \bar{z}_n) - \bar{x}_m\right] \tag{6}$$

Singh and Espejo (2007) considered a ratio-product type estimator in double sampling as

$$t_6 = \bar{y}_m\left[k\frac{\bar{x}_n}{\bar{x}_m} + (1-k)\frac{\bar{x}_m}{\bar{x}_n}\right] \tag{7}$$

Proposed Estimator

The suggested unbiased regression-ratio type estimator for estimating the population mean $\bar{Y}$ is

$$T_R = \sum_{i=1}^{3} d_i \bar{y}_m^* \left(\frac{\bar{x}_n}{\bar{x}_m}\right)^i \tag{8}$$

where $\bar{y}_m^* = \bar{y}_m + b_{yz}(m)(\bar{Z} - \bar{z}_m)$ and the d_i ($i = 1, 2, 3$) are real scalars suitably chosen so that

$$\sum_{i=1}^{3} d_i = 1 \tag{9}$$

Remark 1: The estimator T_R is proposed under the following conditions:

1. The sum of the weights is one.
2. The weights of the linear form are chosen such that the approximate bias is zero.
3. The approximate variance attains minimum.

Properties of the Estimator T_R

Note from (8) that the proposed estimator T_R is biased for $\bar{Y}$. Following Remark 1, it may be made unbiased for $\bar{Y}$. The variance V(.) up to the first order of approximations are derived under large sample approximations using the following transformations:

$$\bar{y}_m = \bar{Y}\left(1+e_1\right), \quad \bar{x}_m = \bar{X}\left(1+e_2\right), \quad \bar{x}_n = \bar{X}\left(1+e_3\right)$$
$$\bar{z}_m = \bar{Z}\left(1+e_4\right), \quad s_{yz}(m) = S_{yz}\left(1+e_5\right), \quad s_z^2(m) = S_z^2\left(1+e_6\right)$$

where $E(e_i) = 0$ and $|e_i| < 1$ for all $i = 1,\ldots, 6$.

Under the above transformations the estimator T_R takes the following form:

$$T_R = \sum_{i=1}^{3} d_i \left\{\bar{Y}\left(1+e_1\right) - \beta_{yz}\bar{Z}e_4\left(1+e_5\right)\left(1+e_6\right)^{-1}\right\} \times \left\{\left(1+e_3\right)\left(1+e_2\right)^{-1}\right\} \tag{10}$$

The bias and mean square error of the estimator was derived separately for the Cases I and II of the two-phase sampling structure.

Case I: When the second phase sample is drawn as a subsample of the first phase sample.

In this case we have the following expected values of the sample statistics:

$$\left.\begin{aligned}
&\mathrm{E}\left(e_1^2\right)=f_m C_y^2, && \mathrm{E}\left(e_s^2\right)=f_m C_x^2, && \mathrm{E}\left(e_3^2\right)=f_1 C_x^2, \\
&\mathrm{E}\left(e_4^2\right)=f_m C_z^2, && \mathrm{E}\left(e_1 e_2\right)=f_m \rho_{yx} C_y C_x, && \mathrm{E}\left(e_1 e_3\right)=f_1 \rho_{yx} C_y C_x, \\
&\mathrm{E}\left(e_1 e_4\right)=f_m \rho_{yz} C_y C_z, && \mathrm{E}\left(e_2 e_3\right)=f_1 C_x^2, && \mathrm{E}\left(e_2 e_4\right)=f_m \rho_{xz} C_x C_z, \\
&\mathrm{E}\left(e_3 e_4\right)=f_1 \rho_{xz} C_x C_z, && \mathrm{E}\left(e_4 e_5\right)=f_m \frac{\mu_{012}}{\bar{Z} S_{yz}}, && \mathrm{E}\left(e_4 e_6\right)=f_m \frac{\mu_{003}}{\bar{Z} S_z^2}
\end{aligned}\right\} \quad (11)$$

where

$$f_m = \frac{1}{m} - \frac{1}{N}, \quad f_1 = \frac{1}{n} - \frac{1}{N},$$

$$\mu_{pqr} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^p \left(y_i - \bar{Y}\right)^q \left(z_i - \bar{Z}\right)^r$$

and p, q, $r \geq 0$ are integers.

Expanding the terms of (10) binomially and using the results from (11), we have derived the expression of bias and mean square error of the estimator T_R up to the first order of approximations as

$$\begin{aligned}
\mathrm{B}\left(T_R\right) &= \mathrm{E}\left(T_R - \bar{Y}\right) \\
&= f_m \beta_{yz}\left(\frac{\mu_{003}}{S_z^2} - \frac{\mu_{012}}{S_{yz}}\right) + f_2 P\left(\bar{Y} C_x^2 - \bar{Y}\rho_{yx}C_y C_x + \bar{Z}\beta_{yz}\rho_{xz}C_x C_z\right) \quad (12) \\
&\quad - d_2 f_2 \bar{Y} C_x^2
\end{aligned}$$

$$\begin{aligned}
\mathrm{M}\left(T_R\right) &= \mathrm{E}\left(T_R - \bar{Y}\right) \\
&= \bar{Y}^2 f_m C_y^2 + \bar{Z}^2 \beta_{yz}^2 f_m C_z^2 - 2\bar{Y}\bar{Z}\beta_{yz}f_m \rho_{yz}C_y C_z + P^2 \bar{Y}^2 f_2 C_x^2 \quad (13) \\
&\quad + 2P\bar{Y}f_2\left(\bar{Z}\beta_{yz}\rho_{xz}C_x C_z - \bar{Y}\rho_{yx}C_y C_x\right)
\end{aligned}$$

where

$$P = \sum_{i=1}^{3} i d_i, \quad f_2 = \frac{1}{m} - \frac{1}{n} \quad (14)$$

Minimization of mean square error in (13) with respect to P yields its optimum value as

$$P = \left(\rho_{yx} - \rho_{yz}\rho_{xz}\right)\frac{C_y}{C_x} \qquad (15)$$

Substituting the optimum value of P in (13), we obtain the minimum mean square error of T_R as

$$\text{Min.M}\left(T_R\right) = f_m S_y^2 \left(1 - \rho_{yz}^2\right) - f_2 S_y^2 \left(\rho_{yx} - \rho_{yz}\rho_{xz}\right)^2 \qquad (16)$$

Further, from (14) and (15),

$$P = \sum_{i=1}^{3} id_i = \left(\rho_{yx} - \rho_{yz}\rho_{xz}\right)\frac{C_y}{C_x} \qquad (17)$$

which we will denote with R.

From (9) and (17), it may be noted that the two equations in three unknowns are not sufficient to find the unique values of the d_i ($i = 1, 2, 3$). In order to get unique values of the d_i, impose the linear constraint

$$\text{B}\left(T_R\right) = 0 \qquad (18)$$

Thus, from (12),

$$Kd_1 + \left(2K - \bar{Y}f_2 C_x^2\right)d_2 + 3Kd_3 = M \qquad (19)$$

where

$$K = f_2\left(\bar{Y}C_x^2 - \bar{Y}\rho_{yx}C_y C_x + \bar{Z}\beta_{yz}\rho_{xz}C_x C_z\right), \quad M = f_m\beta_{yz}\left(\frac{\mu_{012}}{S_{yz}} - \frac{\mu_{003}}{S_z^2}\right)$$

Equations (9), (17), and (19) can be written in matrix form as

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ K & 2K - \bar{Y}f_2 C_x^2 & 3K \end{pmatrix} \times \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \begin{pmatrix} 1 \\ R \\ M \end{pmatrix} \tag{20}$$

Solving (20), the unique values of the d_i are

$$\left. \begin{aligned} d_1 &= \frac{3}{2} + \frac{1}{2 f_2 \bar{Y} C_x^2} \left[M - R \left(K + f_2 \bar{Y} C_x^2 \right) \right] \\ d_2 &= \frac{1}{f_2 \bar{Y} C_x^2} \left(RK - M \right) \\ d_3 &= \frac{1}{2 f_2 \bar{Y} C_x^2 \left[\left(M - RK \right) + \left(R - 1 \right) f_2 \bar{Y} C_x^2 \right]} \end{aligned} \right\} \tag{21}$$

From (21), substituting the values of d_1, d_2, and d_3 into (8) yields the unbiased optimum regression-ratio type estimator as

$$\begin{aligned} T_R &= \left[\frac{3}{2} + \frac{1}{2 f_2 \bar{Y} C_x^2} \left\{ M - R \left(K + f_2 \bar{Y} C_x^2 \right) \right\} \right] \bar{y}_m^* \left(\frac{\bar{x}_n}{\bar{x}_m} \right) \\ &\quad + \frac{1}{f_2 \bar{Y} C_x^2} \left(RK - M \right) \bar{y}_m^* \left(\frac{\bar{x}_n}{\bar{x}_m} \right)^2 \\ &\quad + \left[\frac{1}{2 f_2 \bar{Y} C_x^2} \left\{ \left(M - RK \right) + \left(R - 1 \right) f_2 \bar{Y} C_x^2 \right\} \right] \bar{y}_m^* \left(\frac{\bar{x}_n}{\bar{x}_m} \right)^3 \end{aligned} \tag{22}$$

whose variance up to the first degree of approximations is given by

$$V \left(T_R \right) = f_m S_y^2 \left(1 - \rho_{yz}^2 \right) - f_2 S_y^2 \left(\rho_{yx} - \rho_{yz} \rho_{xz} \right)^2 \tag{23}$$

Case II: When the second-phase sample is drawn independently of the first-phase sample.

In this case, the expected values of the sample statistics are:

$$\left.\begin{array}{lll} \mathrm{E}\left(e_1^2\right) = f_m C_y^2, & \mathrm{E}\left(e_1^2\right) = f_m C_x^2, & \mathrm{E}\left(e_3^2\right) = f_1 C_x^2, \\[2mm] \mathrm{E}\left(e_4^2\right) = f_m C_z^2, & \mathrm{E}\left(e_1 e_2\right) = f_m \rho_{yx} C_y C_x, & \mathrm{E}\left(e_1 e_4\right) = f_m \rho_{yz} C_y C_z, \\[2mm] \mathrm{E}\left(e_2 e_4\right) = f_m \rho_{xz} C_x C_z, & \mathrm{E}\left(e_4 e_5\right) = f_m \dfrac{\mu_{012}}{\overline{Z} S_{yz}}, & \mathrm{E}\left(e_4 e_6\right) = f_m \dfrac{\mu_{003}}{\overline{Z} S_z^2}, \\[2mm] \mathrm{E}\left(e_1 e_3\right) = \mathrm{E}\left(e_2 e_3\right) = \mathrm{E}\left(e_3 e_4\right) = 0 \end{array}\right\} \quad (24)$$

Proceeding as in Case I, the unbiased optimum regression-ratio type estimator is obtained as

$$\begin{aligned} T_R = {} & \frac{1}{f_1 + 2f_m}\left[\left(f_1 + f_m\right)\left(3 - G\right) - \frac{1}{\overline{Y} C_x^2}\left(WG - A\right)\right] \overline{y}_m^* \left(\frac{\overline{x}_n}{\overline{x}_m}\right) \\[2mm] & + \frac{1}{f_1 + 2f_m}\left[\frac{2}{\overline{Y} C_x^2}\left(WG - A\right) - f_1\left(3 - G\right)\right] \overline{y}_m^* \left(\frac{\overline{x}_n}{\overline{x}_m}\right)^2 \\[2mm] & + \left[1 - \frac{1}{f_1 + 2f_m}\left\{f_m\left(3 - G\right) + \frac{1}{\overline{Y} C_x^2}\left(WG - A\right)\right\}\right] \overline{y}_m^* \left(\frac{\overline{x}_n}{\overline{x}_m}\right)^3 \end{aligned} \quad (25)$$

with variance up to the first order of approximations as

$$\mathrm{V}\left(T_R\right) = f_m S_y^2 \left(1 - \rho_{yz}^2\right) - \frac{f_m^2}{f_1 + f_m} S_y^2 \left(\rho_{yx} - \rho_{yz}\rho_{xz}\right)^2 \quad (26)$$

where

$$W = f_1 \overline{Y} C_x^2 + f_m \overline{Y} C_x^2 - f_m \overline{Y} \rho_{yx} C_y C_x + f_m \overline{Z} \beta_{yz} \rho_{xz} C_x C_z$$

$$G = \frac{f_m}{f_1 + f_m}\left(\rho_{yx} - \rho_{yz}\rho_{xz}\right)\frac{C_y}{C_x}$$

$$A = f_m \beta_{yz}\left(\frac{\mu_{012}}{S_{yz}} - \frac{\mu_{003}}{S_z^2}\right)$$

Remark 2: The unique value of the scalars d_i depend on unknown population parameters such as β_{xz}, β_{yz}, μ_{012}, μ_{003}, C_x, C_y, C_z, $\overline{X}$, $\overline{Y}$, ρ_{yx}, and ρ_{xz}. Thus, to make the estimator practicable, these unknown population parameters may be estimated with

their respective sample estimates or from past data or guessed from experience gathered over time. Such problems are also considered by Reddy (1978), Tracy et al. (1996), and Singh and Espejo (2007).

Results

Efficiency Comparison

To examine the performance of our proposed estimator, we have considered some contemporary estimators of population mean which are discussed in a previous section. The mean square errors/minimum mean square errors of the estimators t_i ($i = 1, 2,\ldots, 6$) are given below for both cases of two-phase sampling structure considered in this paper:

Case I:

$$M(t_1) = \bar{Y}^2 \left[f_m C_y^2 + f_2 C_x^2 - 2 f_2 \rho_{yx} C_y C_x \right]$$

$$\text{Min.}M(t_2) = S_y^2 \left[f_m - f_2 \rho_{yx}^2 \right]$$

$$M(t_3) = \bar{Y}^2 \left[f_m C_y^2 + f_2 C_x^2 + f_1 C_z^2 - 2 f_2 \rho_{yx} C_y C_x - 2 f_1 \rho_{yz} C_y C_z \right]$$

$$M(t_4) = \bar{Y}^2 \left[f_m C_y^2 + f_2 C_x^2 + f_1 C_z^2 + 2 f_2 \rho_{yx} C_y C_x + 2 f_1 \rho_{yz} C_y C_z \right]$$

$$M(t_5) = S_y^2 \left[f_m \left(1 - \rho_{yx}^2\right) + f_1 \left\{ \rho_{yx}^2 \left(1 + \rho_{xz}^2\right) - 2 \rho_{yx} \rho_{xz} \rho_{yz} \right\} \right]$$

$$M(t_6) = S_y^2 \left[f_m - f_2 \rho_{yx}^2 \right]$$

Case II:

$$M(t_1) = \bar{Y}^2 \left[f_m C_y^2 + f_m C_x^2 + f_1 C_x^2 - 2 f_m \rho_{yx} C_y C_x \right]$$

$$\text{Min.}M(t_2) = f_m S_y^2 \left[1 - \theta \rho_{yx}^2 \right]$$

$$M(t_3) = \bar{Y}^2 \left[f_m C_y^2 + f_m C_x^2 + f_1 C_z^2 - 2 f_m \rho_{yx} C_y C_x - 2 f_1 \rho_{xz} C_x C_z \right]$$

$$M(t_4) = \bar{Y}^2 \left[f_m C_y^2 + f_m C_x^2 + f_1 C_x^2 + f_1 C_z^2 + 2 f_m \rho_{yx} C_y C_x - 2 f_1 \rho_{xz} C_x C_z \right]$$

$$M(t_5) = S_y^2 \left[f_m - f_2 \rho_{yx}^2 - f_1 \rho_{yx}^2 \rho_{xz}^2 \right]$$

$$M(t_6) = f_m S_y^2 \left[1 - \theta \rho_{yx}^2 \right]$$

where $\theta = \dfrac{f_m}{f_1 + f_m}$.

The superiority of the suggested estimator has been demonstrated over the estimators t_i $(i = 1, 2, \ldots, 6)$ through numerical illustrations and graphical interpretation.

Numerical Illustrations

Five natural population data sets were selected to illustrate the efficiency of the proposed estimator. The source of the populations, the nature of the variables y, x, z and the values of the various parameters are as follows:

Population I: (Murthy, 1967)
y: Area under wheat in 1964.
x: Area under wheat in 1963.
z: Cultivated area in 1961.

Population II: (Sukhatme & Sukhatme, 1970)
y: Area (acres) under wheat in 1937.
x: Area (acres) under wheat in 1936.
z: Total cultivated area (acres) in 1931.

Population III: (S. K. Srivastava, 1971)
y: yield per plant.
x: Height of the plant.
z: Base diameter.

Population IV: (Anderson, 1958)
y: Head length of second son.
x: Head length of first son.
z: Head breadth of second son.

Population V: (R. S. Srivastava, Srivastava, & Khare, 1989)
y: measurement of weight of children.
x: Mid-arm circumference of children.
z: Skull circumference of children.

Table 1. Parametric values of different populations

Population	N	n	m	$\bar{Y}r$	ρ_{yx}	ρ_{yz}	ρ_{xz}	C_y	C_x	C_z
I	80	34	6	5182.60	0.9100	0.9400	0.9900	0.35000	0.94000	0.75000
II	34	10	7	201.41	0.9300	0.9000	0.8300	0.74000	0.76000	0.61000
III	50	20	12	5.69	0.7418	0.5677	0.2063	0.23830	0.09198	0.11260
IV	25	10	7	183.84	0.7108	0.6932	0.7346	0.05460	0.05260	0.04880
V	55	30	18	17.08	0.5400	0.5100	-0.0800	0.12690	0.07000	0.02650

Table 2. PREs of different estimators (Case I)

Population	$\bar{y}_m$	t_1	t_2	t_3	t_4	t_5	t_6	T_R
I	100	*	380.6032	*	*	602.4841	380.6032	859.4502
II	100	147.7505	148.5310	566.9582	*	557.1949	148.5310	591.1479
III	100	128.6936	140.7686	159.0693	*	156.4518	140.7686	202.8928
IV	100	122.5372	126.6649	178.8188	*	190.0258	126.6649	201.7059
V	100	120.9633	120.9751	131.9087	*	118.3101	120.9751	165.7976

Table 3. PREs of different estimators (Case II)

Population	$\bar{y}_m$	t_1	t_2	t_3	t_4	t_5	t_6	T_R
I	100	*	426.8195	*	*	577.2247	426.8195	862.0124
II	100	*	214.2050	286.3820	*	330.5549	214.2050	590.4718
III	100	*	159.5903	139.4500	*	143.0017	159.5903	242.1787
IV	100	*	146.8638	120.7684	*	158.6188	146.8638	202.5066
V	100	*	126.1805	116.6850	*	121.0860	126.1805	200.0489

The values of various parameters obtained from the above populations are presented in Table 1.

To have a tangible idea about the performance of the proposed estimator T_R, the percent relative efficiencies (PREs) of T_R and other estimators were computed with respect to the sample mean estimator $\bar{y}_m$, and the results are demonstrated in Tables 2-3. The PRE of an estimator T with respect to a sample mean estimator $\bar{y}$ is defined as

$$\text{PRE} = \frac{\text{V}(\bar{y})}{\text{M}(T)} \times 100 \tag{27}$$

where $\text{M}(T)$ denotes the MSE/Minimum MSE of an estimator T.

Graphical Interpretation

The performance of the proposed estimator is illustrated by means of pictorial representation for different choices of correlations. This could not only improve the readability of the results but also allows the comparison of a much denser grid of different correlation values. For $N = 100$, $n = 50$, $m = 20$, and different values of ρ_{yx}, ρ_{yz}, ρ_{xz}, the PREs of the proposed estimator T_R with respect to $\bar{y}_m$ are computed and presented in Figures 1-2. Note that the X-axis, Y-axis, and Z-axis are denoting ρ_{yx}, ρ_{yz}, and PRE, respectively, and that ρ_{xz} is assumed to be 0.5.
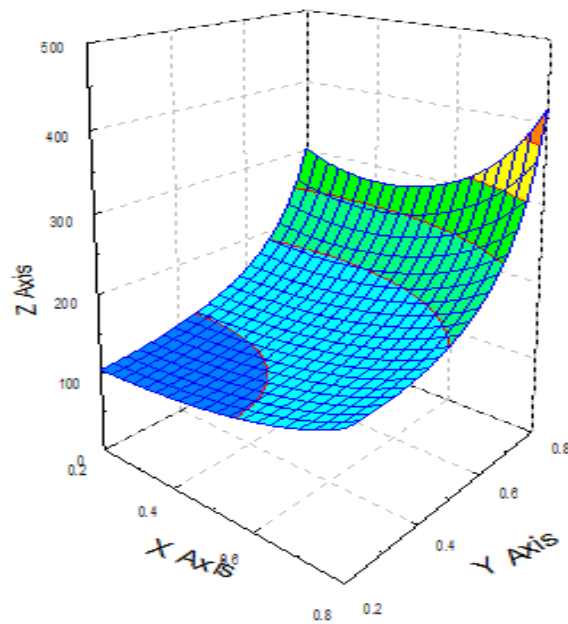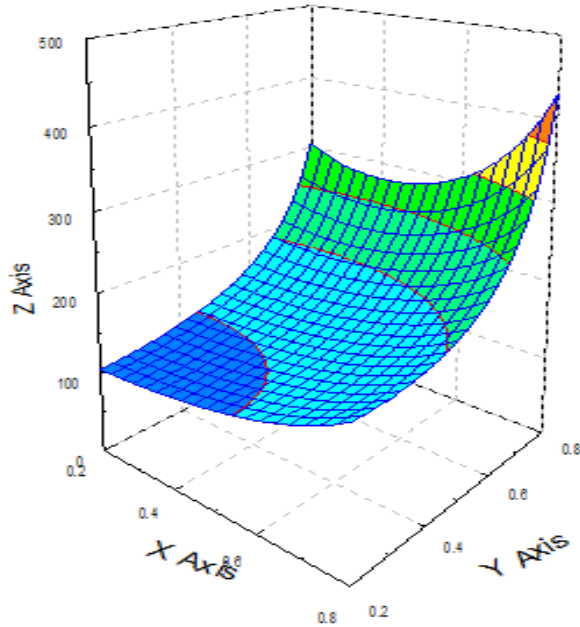

Figure 1. PRE of T_R (Case I)

Figure 2. PRE of T_R (Case II)

Conclusion

From Table 2 and Table 3, it may be observed that, under different structures of two-phase sampling set up, the suggested estimator T_R is superior to the existing ones. It can also be noted that, for high positive values of correlation coefficients, the estimator T_R yields impressive gains in efficiencies over the conventional estimators of population mean.

From Figures 1 and 2, it is observed that, for fixed values of ρ_{xz}, the PRE of the proposed estimator is increasing with increasing values of ρ_{yx} and ρ_{yz}. This phenomenon indicates that suggested estimator could perform satisfactorily if highly positive correlated auxiliary variables are available.

Therefore, the proposed estimator T_R is more justified in comparison with the previous work of similar nature. Hence, it may be recommended to the survey practitioners for their use in real life problems.

References

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. New York: John Wiley & Sons.

Bose, C. (1943). Note on the sampling error in the method of double sampling. *Sankhyā: The Indian Journal of Statistics, 6*(3), 329-330.

Chand, L. (1975). *Some ratio type estimators based on two or more auxiliary variables* (Unpublished doctoral thesis). Iowa State University, Ames, IA.

Choudhury, S., & Singh, B. K. (2012). A class of chain ratio-product type estimators with two auxiliary variables under double sampling scheme. *Journal of the Korean Statistical Society, 41*(2), 247-256. doi: 10.1016/j.jkss.2011.09.002

Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for ratio gain to total produce. *The Journal of Agricultural Science, 30*(02), 262-275. doi: 10.1017/S0021859600048012

Dash, P. R., & Misra, G. (2011). An improved class of estimators in two-phase sampling using two auxiliary variables. *Communication in Statistics – Theory and Methods, 40*(24), 4347-4352. doi: 10.1080/03610920903094683

Gupta, S., & Shabbir, J. (2007). On the use of transformed auxiliary variables in estimating population mean by using two auxiliary variables. *Journal of Statistical Planning and Inference, 137*(5), 1606-1611. doi: 10.1016/j.jspi.2006.09.008

Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika, 27*(1), 217-223. doi: 10.1007/BF01893599

Kiregyera, B. (1984). Regression type estimators using two auxiliary variables and the model of double sampling from finite populations. *Metrika, 31*(1), 215-226. doi: 10.1007/BF01915203

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

Reddy, V. N. (1978). A study on the use of prior knowledge on certain population parameters in estimation. *Sankhyā: The Indian Journal of Statistics, 40*(1), 29-37.

Sahoo, J., & Sahoo, L. N. (1993). A class of estimators in two-phase sampling using two auxiliary variables. *Journal of the Indian Statistical Association, 37*, 107-114.

Shukla, D., Pathak, S., & Thakur, N. S. (2012). Estimation of population mean using two auxiliary sources in sample surveys. *Statistics in Transition – new series, 13*(1), 21-36.

Singh, H. P., & Espejo, M. R. (2007). Double sampling ratio-product estimator of a finite population mean in sampling surveys. *Journal of Applied Statistics, 34*(1), 74-85. doi: 10.1080/02664760600994562

Srivastava, R. S., Srivastava, S. P., & Khare, B. B. (1989). Chain ratio type estimator for ratio of two population means using auxiliary character. *Communications in Statistics – Theory and Methods, 18*(10), 3917-3926. doi: 10.1080/03610928908830131

Srivastava, S. K. (1971). Generalized estimator for the mean of a finite population using multi auxiliary information. *Journal of the American Statistical Association, 66*(334), 404-407. doi: 10.1080/01621459.1971.10482277

Sukhatme, P. V., & Sukhatme, B. V. (1970). *Sampling theory of surveys with application*. Ames, IA: Iowa State University Press.

Tracy, D. S., Singh, H. P., & Singh, R. (1996). An alternative to the ratio-cum-product estimator in sample surveys. *Journal of Statistical Planning and Inference, 53*(3), 375-387. doi: 10.1016/0378-3758(95)00136-0