International Journal of Statistics and Applied Mathematics 2025; 10(6): 195-202

ISSN: 2456-1452
Maths 2025, 10(6): 195-202

https://www.mathsjournal.com
Received: 06-04-2025
Accepted: 08-05-2025

Archana Panigrahi
Department of Statistics, Ravenshaw University, Cuttack, Odisha, India

Amiya Ojha
Department of Statistics, Ravenshaw University, Cuttack, Odisha, India

LN Sahoo
Former Professor, Department of Statistics, Utkal University, Bhubaneswar, Odisha, India

Corresponding Author:
Archana Panigrahi
Department of Statistics, Ravenshaw University, Cuttack, Odisha, India

A reducible product-type estimator in two-phase sampling using an additional auxiliary variable

Archana Panigrahi, Amiya Ojha and LN Sahoo

DOI: https://www.doi.org/10.22271/maths.2025.v10.i6c.2072

Abstract
The objective of this paper is to develop product methods of estimating population means under two-phase sampling framework when information on an additional (second) auxiliary is readily available. This issue has been addressed by introducing a new approach, called a Redesigned Approach that brought out a reducible (generalized) product-type estimator comprising a family/class of estimators. Efficiency comparisons of some specific selected estimators of the class together with the classical two-phase product estimator have been made. Some design-based properties of the proposed estimator have also been investigated and finally an empirical study has been included to understand the effectiveness of the proposed estimation technique quantitatively.

Keywords: Auxiliary variable, bias, efficiency, product estimator, two-phase sampling

1. Introduction

Consider a finite population $U = \{1, 2, \dots, i, \dots, N\}$. Let y and x be the study variable and an auxiliary variable taking values y_i and x_i respectively on the ith unit $(i = 1, 2, \dots, N)$. Let $\overline{Y} = \sum_{i=1}^{N} y_i/N$, $\overline{X} = \sum_{i=1}^{N} x_i/N$ be the population means and $S_y^2 = \sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2/(N-1)$, $S_x^2 = \sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2/(N-1)$ be the population variances of y and x, and $S_{yx} = \sum_{i=1}^{N}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)/(N-1)$ be the population covariance between y and x. It has already been established that the ratio method of estimation fails to estimate population mean $\overline{Y}$ with acceptable gain in precision when y and x are negatively correlated. However, in such a situation the product method of estimation works well depending on the strength of inverse relationship between the variables.

The classical product estimator of $\overline{Y}$ essentially needs previous knowledge of the mean $\overline{X}$ or total $X\ (= N\overline{X})$. However, there are situations where neither $\overline{X}$ nor X is known in advance. The usual procedure is then to apply the two-phase or double sampling technique. When simple random sampling without replacement (SRSWOR) is entertained at each phase, the two-phase sampling scheme will be as follows:

- A large preliminary sample called the first phase sample s_1 $(s_1 \subset U)$ of fixed size n_1 is drawn to observe x.
- Given s_1, a second phase sample s_2 $(s_2 \subset s_1)$ of fixed size n_2 is drawn to observe y only.

Let $\bar{x}_1 = \frac{1}{n_1}\sum_{i\in s_1} x_i$ be the sample mean of x based on s_1; $\bar{y}_2 = \frac{1}{n_1}\sum_{i\in s_1} y_i$ and $\bar{x}_2 = \frac{1}{n_2}\sum_{i\in s_2} x_i$ be the sample means of y and x respectively based on s_2. Then the two-phase sampling classical product estimator for $\overline{Y}$ is defined by

$$t_P = \bar{y}_2 \frac{\bar{x}_2}{\bar{x}_1}.$$

The estimator, although biased, for large sample sizes the bias is customarily negligible. The approximate expression for the MSE is given by

$$M(t_P) = \overline{Y}^2\left[\theta_2 C_y^2 + (\theta_2 - \theta_1)\left(C_x^2 + 2C_{yx}\right)\right], \quad (1)$$

where $\theta_1 = \frac{1}{n_1} - \frac{1}{N}$, $\theta_2 = \frac{1}{n_2} - \frac{1}{N}$, $C_y^2 = S_y^2/\overline{Y}^2$, $C_x^2 = S_x^2/\overline{X}^2$ and $C_{yx} = S_{yx}/\overline{Y}\,\overline{X}$. Hence, t_P performs better than the mean per unit estimator $\bar{y}_2$ when $\rho_{yx}C_y/C_x < -1/2$, ρ_{yx} being the correlation coefficient between y and x.
Substantial MSE reduction over t_P can be brought either by remodeling the sampling scheme or by restructuring the estimator. But one of the simplest courses of action to achieve this is to reshape t_P by engaging one or more additional auxiliary variables. In this paper, the goal is to consider some alterations over t_P with the aid of an additional auxiliary variable z to compose more précised product-type estimators.

2. The Role of An Additional Auxiliary Variable

Let us consider a survey situation where no prior information on $\overline{X}$ or X is obtainable but the values $z_1, z_2, \ldots, z_N$ of a secondary auxiliary variable z are accessible for the entire population such as the population mean $\overline{Z} = \sum_{i=1}^N z_i/N$ or total $Z = N\overline{Z}$ is known precisely. It is also expected that z is highly correlated with x. For instance, while estimating the infant mortality rate in a district, the infant mortality and number of educated married women for each village are likely to be unknown. But village-wise figures on female population or total population are readily available from census records. Then y, x and z are respectively the infant mortality, number of educated married women and female or total population. Here, the correlation between y and x is negative whereas the same between x and z is likely to be positive.
The two-phase sampling mechanism involving n_1 and n_2 in this context is such that the preliminary sample s_1 is used to collect measurements on (x, z) whereas the second phase sample s_2 is used to collect measurements on y only. The key idea behind this is to make reasonable estimates for $\overline{X}$ or X based on the measured values of $(x_i, z_i), i \epsilon s_1$. Of course, the precision of such an estimate is influenced by the strength of correlation between x and z.
Groundwork for the estimation of $\overline{Y}$ under the above background was instituted for the first time by Chand (1975) [1] and Sukhatme and Chand (1977) [4]. They have grown a chaining principle under which estimators for $\overline{Y}$ are designed from a classical two-phase estimator simply replacing first phase sample mean $\bar{x}_1$ by a better estimator of $\overline{X}$ considering z as an auxiliary variable and exploiting data on (x, z) for s_1. Afterwards, the said chaining technique was deliberated in great depth by Kiregyera (1980, 1984) [2, 3] and consequently inspired several authors to construct large varieties of estimators. In these works, although more concentration was given for the ratio and regression-types estimators, only a few attempts have been made for producing product-type estimators [see for example, Sahoo *et al.* (2006) [5], Sharma *et al.* (2014) [12]. But the present study emphasizes creation of estimators considering the two-phase product estimator t_P as the base.
Considering the kind of correlation that z have with x *i.e*., either positive or negative or both, selection of an $s_1 -$ based estimator for $\overline{X}$ alternative to $\bar{x}_1$ can be made in line of Chand (1975) [1]. Hence, the ratio estimator $\bar{x}_1\overline{Z}/\bar{z}_1$, the product estimator $\bar{x}_1\bar{z}_1/\overline{Z}$ and the regression estimator $\bar{x}_1 - b_{xz(1)}\left(\bar{z}_1 - \overline{Z}\right)$ are likely to be more preferrable to $\bar{x}_1$. Accordingly, the following three estimators would be taken into consideration:

Ratio-in-product estimator: $t_{RP} = \bar{y}_2\frac{\bar{x}_2\bar{z}_1}{\bar{x}_1\overline{Z}}$

Product-in-product estimator: $t_{PP} = \bar{y}_2\frac{\bar{x}_2\overline{Z}}{\bar{x}_1\bar{z}_1}$

Regression-in-product estimator: $t_{RGP} = \bar{y}_2\frac{\bar{x}_2}{\left[\bar{x}_1 - b_{xz(1)}\left(\bar{z}_1 - \overline{Z}\right)\right]}$,

where $\bar{z}_1 = \frac{1}{n_1}\sum_{i\in s_1} z_i$ and $b_{xz(1)} = \frac{\sum_{i\in s_1}(x_i - \bar{x}_1)(z_i - \bar{z}_1)}{\sum_{i\in s_1}(z_i - \bar{z}_1)^2}$ is the sample regression coefficient of x on z for s_1.

The asymptotic expressions for the MSEs of t_{RP}, t_{PP} and t_{RGP} are as follows:

$$M(t_{RP}) = M(t_P) + \overline{Y}^2\theta_1\left(C_z^2 + 2C_{yz}\right) \quad (2)$$

$$M(t_{PP}) = M(t_P) + \overline{Y}^2\theta_1\left(C_z^2 - 2C_{yz}\right) \quad (3)$$

$$M(t_{RGP}) = M(t_P) + \overline{Y}^2\theta_1\left(\rho_{xz}^2 C_x^2 + 2\rho_{yz}\rho_{xz}C_y C_x\right), \quad (4)$$

where $C_z^2 = S_z^2/\overline{Z}^2$, $C_{yz} = S_{yz}/\overline{Y}\,\overline{Z}$ such that $S_z^2 = \sum_{i=1}^N\left(z_i - \overline{Z}\right)^2/(N-1)$ and $S_{yz} = \sum_{i=1}^N\left(y_i - \overline{Y}\right)\left(z_i - \overline{Z}\right)/(N-1)$; ρ_{yz} and ρ_{xz} are respectively correlation coefficients between y, z and x, z. Hence, t_{RP}, t_{PP} and t_{RGP} are likely to be more efficient than t_P if

$$\rho_{yz}\frac{C_y}{C_z} < -\frac{1}{2},\ \ \rho_{yz}\frac{C_y}{C_z} > \frac{1}{2}\ \text{ and } \rho_{yz} < -\frac{1}{2}\rho_{xz}\frac{C_x}{C_y}, \quad (5)$$

respectively. These conditions denote that selection of ratio or product estimator for the unknown mean $\overline{X}$ in terms of z depends on the strength and magnitude of ρ_{yz} whereas selection of regression estimator depends on those of both ρ_{yz} and ρ_{xz}.

In lieu of $\bar{x}_1$, $\bar{x}_1\frac{\bar{Z}}{\bar{z}_1}$, $\bar{x}_1\frac{\bar{z}_1}{\bar{Z}}$ and $\bar{x}_1 - b_{xz(1)}(\bar{z}_1 - \bar{Z})$ as estimators of $\bar{X}$, Sahoo *et al.* (2006) [5] considered more generally a difference estimator $\bar{x}_1 - d(\bar{z}_1 - \bar{Z})$ and accordingly defined a generalized estimator for $\bar{Y}$ by

$$t = \bar{y}_2\frac{\bar{x}_2}{[\bar{x}_1 - d(\bar{z}_1 - \bar{Z})]}.$$

The estimators t_P, t_{RP}, t_{PP} and t_{RGP} run out as its special cases when $d = 0$, $\frac{\bar{x}_1}{\bar{z}_1}$, $-\frac{\bar{x}_1}{\bar{Z}}$ and $b_{xz(1)}$ respectively.

3. A Reducible Estimator Under a Redesigned Approach

Many authors of course used Chand-Kiregyera (C-K) approach *i.e.*, replacement of $\bar{x}_1$ by z −based estimators in a classical estimator to compose various estimators for $\bar{Y}$. But they simply recommended estimators without giving any explanation on the technique adopted for their construction. However, the purpose here is to gain better improvements over t_P, in respect of efficiency, reformatting the Chand-Kiregyera approach. This approach may be called a *Redesigned Approach*, as explained below, in which the target is to use available ancillary information on z at each phase.

As said earlier, t_{RP}, t_{PP} and t_{RGP} are generated when $\bar{x}_1$ in the standard two-phase product estimator t_P is replaced by $\bar{x}_1\frac{\bar{Z}}{\bar{z}_1}$, $\bar{x}_1\frac{\bar{z}_1}{\bar{Z}}$ and $\bar{x}_1 - b_{xz(1)}(\bar{z}_1 - \bar{Z})$ respectively with the understanding that the later estimators are better than the former to estimate $\bar{X}$ under certain conditions. On the other hand, we do admit that $\bar{x}_2$ offers a less efficient estimated value of $\bar{X}$ than $\bar{x}_1$. Hence, this school of thought encourages an alternative option for $\bar{x}_2$ in respect of the second covariate z. But, to generalize our estimation technique we would like to engage a difference estimator $\bar{x}_2 - \eta(\bar{z}_2 - \bar{z}_1)$ in place of $\bar{x}_2$. At the same time, we also restrict to $\bar{x}_1 - \delta(\bar{z}_1 - \bar{Z})$ as alternative to $\bar{x}_1$. These arrangements provide the following reducible or generalized product-type estimator:

$$\ell^{(G)} = \bar{y}_2\frac{\bar{x}_2 - \eta(\bar{z}_2 - \bar{z}_1)}{\bar{x}_1 - \delta(\bar{z}_1 - \bar{Z})}.$$

The coefficients η and δ used in $\ell^{(G)}$ are either suitable picked out constants or random variables converging to some finite values. But in the usual practice, they decide to maximize precision of the estimator.

The flexibility characteristic of $\ell^{(G)}$ brings many estimators based on either one or two supplementary variables after suitable selections of the coefficients. This means that, it generates a class or family of product-type estimators for $\bar{Y}$. For the simplest case when $\eta = 0$ and $\delta = 0$, $\ell^{(G)} = t_P$, the base estimator. But for $\eta = 0$ and $\delta = d$, $\ell^{(G)} = t$, the generalized estimator defined above. This means for $\eta = 0$, $\ell^{(G)} \longrightarrow t_{RP}, t_{PP}$ and t_{RGP} when $\delta = \frac{\bar{x}_1}{\bar{z}_1}, -\frac{\bar{x}_1}{\bar{Z}}$ and $b_{xz(1)}$ respectively. Nevertheless, the following estimators are also some more specific cases of $\ell^{(G)}$ for appropriate options of the coefficients:

$$t_{11} = \bar{y}_2\frac{\bar{x}_2\bar{z}_1}{\bar{x}_1\bar{z}_2},\ t_{12} = \bar{y}_2\frac{\bar{x}_2\bar{z}_2}{\bar{x}_1\bar{z}_1},\ t_{13} = \bar{y}_2\frac{\bar{x}_2 - b_{xz(2)}(\bar{z}_2 - \bar{z}_1)}{\bar{x}_1},\ t_{14} = \bar{y}_2\frac{\bar{x}_2(\bar{z}_1)^2}{\bar{x}_1\bar{z}_2\bar{Z}}$$

$$t_{15} = \bar{y}_2\frac{\bar{x}_2\bar{z}_2\bar{Z}}{\bar{x}_1(\bar{z}_1)^2},\ t_{16} = \bar{y}_2\frac{\bar{x}_2 - b_{xz(2)}(\bar{z}_2 - \bar{z}_1)}{\bar{x}_1 - b_{xz(1)}(\bar{z}_1 - \bar{Z})},$$

where $b_{xz(2)} = \frac{\sum_{i\in s_2}(x_i - \bar{x}_2)(z_i - \bar{z}_2)}{\sum_{i\in s_2}(z_i - \bar{z}_2)^2}$.

4. Comparison of Some Selected Estimators

For establishing effectiveness of our redesigned approach over C-K approach, we shall now make MSE comparisons between some selected estimators considered/elicited above. But to make the study more attainable, we limit it to two situations. In the first case it is assumed that $\rho_{xz} > 0$ so that only ratio estimators are considered in place of $\bar{x}_2$ or $\bar{x}_1$ or both. Considering structural resemblance, here we compare t_{11} and t_{14} with t_P and t_{RP}. In the second case, ρ_{xz} is presumed to be either positive or negative. On this ground, we are restricted to those estimators considering only regression estimators for either $\bar{x}_2$ or $\bar{x}_1$ or both and simply compare t_{13} and t_{16} with t_P and t_{RGP}.

Asymptotic MSE expressions of the comparable estimators are presented below:

$$M(t_{11}) = M(t_P) + \bar{Y}^2(\theta_2 - \theta_1)(C_z^2 - 2C_{yz} - 2C_{xz}) \quad (6)$$

$$M(t_{14}) = M(t_P) + \bar{Y}^2[\theta_2(C_z^2 - 2C_{yz} - 2C_{xz}) + 2\theta_1(2C_{yz} + C_{xz})] \quad (7)$$

$$M(t_{13}) = M(t_P) - \bar{Y}^2(\theta_2 - \theta_1)(\rho_{xz}^2C_x^2 + 2\rho_{yz}\rho_{xz}C_yC_x) \quad (8)$$

$$M(t_{16}) = M(t_P) - \bar{Y}^2(\theta_2 - 2\theta_1)(\rho_{xz}^2C_x^2 + 2\rho_{yz}\rho_{xz}C_yC_x) \quad (9)$$

Throughout the present work, it is assumed that $\theta_2 - 2\theta_1 > 0$ *i.e.*, $n_2 < \frac{n_1}{2}$. This condition of course is a very mild restriction satisfied in many survey situations and can be decided by the sampler at the planning stage without any appreciable increase in cost.
In the following discussions, we obtain certain sufficient conditions to judge superiority of the selected estimators coming out under redesigned approach over their respective counterparts coming under C-K approach in respect of MSE. However, it may be remarked here that drawing out of necessary conditions is difficult.

4.1. Comparison of t_{11} and t_{14} with t_P and t_{RP}

From (1), (2) and (6), it is straightforwardly derived that $M(t_{11}) < M(t_P)$ and $M(t_{11}) < M(t_{RP})$ if

$$\rho_{xz}\frac{C_x}{C_z} > \frac{1}{2} - \rho_{yz}\frac{C_y}{C_z} \text{ and } \rho_{xz}\frac{C_x}{C_z} > \left(\frac{k_1}{2} - k_2\rho_{yz}\frac{C_y}{C_z}\right), \quad (10)$$

respectively, where $k_1 = \frac{\theta_2 - 2\theta_1}{\theta_2 - \theta_1}$ and $k_2 = \frac{\theta_2}{\theta_2 - \theta_1} (> 1)$.

Note that $0 < k_1 < 1$ for $n_2 < \frac{n_1}{2}$ and $M(t_{RP}) < M(t_P)$ if $\rho_{yz}\frac{C_y}{C_z} < -\frac{1}{2}$. Hence, we may conclude that when t_{RP} is superior to t_P, t_{11} is superior to both t_P and t_{RP} if

$$\rho_{xz}\frac{C_x}{C_z} > \max\left(\frac{1}{k_1}, \frac{1}{4}\right)$$

$$\Rightarrow \rho_{xz}\frac{C_x}{C_z} > \frac{1}{k_1}. \quad (11)$$

Exactly in a similar way and omitting details of the derivations, we also deduce that when t_{RP} is more efficient than t_P, t_{14} is more efficient than both t_P and t_{RP} if

$$\rho_{xz}\frac{C_x}{C_z} > \max\left(\frac{1}{k_2}, \frac{1}{4}\right). \quad (12)$$

Considering (6) and (7) it has been established that $M(t_{11}) >$ or $< M(t_{14})$ in accordance with

$$\rho_{yz}\frac{C_y}{C_z} < \text{ or } \rho_{yz}\frac{C_y}{C_z} > -\frac{1}{2}. \quad (13)$$

4.2 Comparison of t_{13} and t_{16} with t_P and t_{RGP}

From the MSE expressions (1), (4), (8) and (9), we directly see that both t_{13} and t_{16} would be more efficient than both t_P and t_{RGP} if

$$\rho_{yz} > -\frac{1}{2}\rho_{xz}\frac{C_x}{C_y}. \quad (14)$$

But, under this condition t_{RGP} is less efficient than t_P. Hence, both t_{13} and t_{16} are superior to t_P and t_{RGP} just when t_{RGP} is inferior to t_P. In this sense the estimators t_{13} and t_{16} may be considered as complementary to t_{RGP}.
In passing, to compare achievability of t_{13} and t_{16} in respect of MSE, it is found from (8) and (9) that $M(t_{13}) <$ or $> M(t_{16})$ if

$$\rho_{yz} > -\frac{1}{2}\rho_{xz}\frac{C_x}{C_y} \text{ or } \rho_{yz} < -\frac{1}{2}\rho_{xz}\frac{C_x}{C_y}. \quad (15)$$

5. Some Design-Based Properties of $\ell^{(G)}$

To study conventional efficacy of the proposed reducible product estimator, we need expressions for their design-based bias and MSE. As the exact derivation of these expressions under a finite population set-up is not possible, we depend on the asymptotic expressions. However, omitting details to save, derived asymptotic bias and MSE expressions of $\ell^{(G)}$ are as follows:

$$B\left(\ell^{(G)}\right) = \overline{Y}\left[(\theta_2 - \theta_1)\left(C_{yx} - \eta D C_{yz}\right) + \theta_1 \delta D\left(\delta D C_z^2 + C_{yz} - C_{xz}\right)\right] \quad (16)$$

$$M\left(\ell^{(G)}\right) = M(t_P) + \overline{Y}^2 D\left[(\theta_2 - \theta_1)\eta\left(\eta D C_z^2 - 2C_{yz} - 2C_{xz}\right)\right.$$

$$\left. + \theta_1\delta\left(\delta D C_z^2 + 2C_{yz}\right)\right], \quad (17)$$

where $D = \overline{Z}/\overline{X}$.

As is known, the asymptotic bias of t_P is given by

$B(t_P) = \overline{Y}(\theta_2 - \theta_1)C_{yx}$,

it cannot be zero as $\rho_{yx} \neq 0$. Alternatively, expression (16) is not simple to provide practicable conclusions on the bias of $\ell^{(G)}$. However, in the following, we shall just derive some sufficient conditions for which $B\left(\ell^{(G)}\right) = 0$. Because here our intention is to achieve improvement over t_P in a certain sense.
Assuming that $\delta \neq 0$, from (16) we have $B\left(\ell^{(G)}\right) = 0$ when $C_{yx} - \eta D C_{yz} = 0$ and $\delta D C_z^2 + C_{yz} - C_{xz} = 0$. This means that $\ell^{(G)}$ is asymptotically unbiased if

$$\eta = \frac{\beta_{xy}}{\beta_{zy}} \text{ and } \delta = \beta_{xz} - \frac{\beta_{yz}}{R}, \tag{18}$$

where $\beta_{xy} = S_{yx}/S_y^2$, $\beta_{yz} = S_{yz}/S_z^2$ etc. and $R = \overline{Y}/\overline{X}$.

The MSE expression provisionally determines some possible ranges or intervals for η and δ in order that $\ell^{(G)}$ would be better than t_P. From (17), $M\left(\ell^{(G)}\right) < M(t_P)$ when $Q_1 = \eta\left(\eta D C_z^2 - 2C_{yz} - 2C_{xz}\right) < 0$ and $Q_2 = \delta\left(\delta D C_z^2 + 2C_{yz}\right) < 0$.

These conditions of course hold if the roots of the quadratic equations $Q_1 = 0$ in η and $Q_2 = 0$ in δ are real and distinct, and η and δ lie between them. Finally, this leads to restrictions

$$0 < \eta + \delta \leq 2\beta_{xz} \text{ or } 2\beta_{xz} \leq \eta + \delta < 0. \tag{19}$$

Above obtained ranges for η and δ provide guidelines to decide their values to improve accuracy of $\ell^{(G)}$ compared to t_P.
Conditions (18) and (19) depending exclusively on β_{yz}, β_{xz}, β_{xy}, β_{zy} and R would be competent enough to provide suitable values for the coefficients to make $\ell^{(G)}$ more productive than t_P on the grounds of bias and efficiency. Sometimes this of course may not be feasible in the absence of known values of the said parameters. However, prior knowledge from past data or surveys or experience or even guessed values having close approximations to the true values may be very helpful for this purpose. It may also be noted that derived ranges for η and δ in (19) are only necessary but not sufficient in the sense that other reasonable ranges can be sorted out for the purpose.

6. Selections of Optimal Coefficients
It is obvious that proper selections of the coefficients make the proposed reducible estimator more purposeful. Discussions on the ranges of determinations in the preceding section may be helpful to some extent in this regard. But what is more desirable is to obtain the best values, *i.e.*, optimal values of the coefficients η and δ which minimize MSE of $\ell^{(G)}$. Hence, with the aid of the usual optimizing technique, these optimal coefficients obtained from (17) are as follows:

$$\hat{\eta} = \beta_{xz} + \frac{\beta_{yz}}{R}, \hat{\delta} = -\frac{\beta_{yz}}{R}. \tag{20}$$

Using these optimal values, the minimum value of $M\left(\ell^{(G)}\right)$ is obtained as

$$M_{min}\left(\ell^{(G)}\right) = M(t_P) - \overline{Y}^2 C_y^2\left[(\theta_2 - \theta_1)\left(\rho_{yz} + \frac{C_x}{C_y}\rho_{xz}\right)^2 + \theta_1\rho_{yz}^2\right]. \tag{21}$$

$M_{min}\left(\ell^{(G)}\right)$ may be interpreted as the minimum MSE bounds of $\ell^{(G)}$. An estimator whose MSE is equal to the said bound is designated as minimum MSE bound estimator of $\ell^{(G)}$. One such estimator that can be generated using optimal values $\hat{\eta}$ and $\hat{\delta}$ in $\ell^{(G)}$, is a product-type estimator defined by

$$\ell_P^{(G)} = \bar{y}_2\frac{\bar{x}_2 - \left(\beta_{xz} + \frac{\beta_{yz}}{R}\right)(\bar{z}_2 - \bar{z}_1)}{\bar{x}_1 + \frac{\beta_{yz}}{R}.(\bar{z}_1 - \bar{Z})}.$$

As in the case of range determination, here the optimal coefficients of η and δ also require known values of the parameters β_{yz}, β_{xz} and R otherwise the optimum estimator $\ell_P^{(G)}$ cannot be computed from the survey data. But on most occasions the parameters are unknown, and the normal practice is therefore to estimate them using available data on the second-phase sample s_2.

Let $b_{yz(2)} = \frac{\sum_{i\in s_2}(y_i - \bar{y}_2)(z_i - \bar{z}_2)}{\sum_{i\in s_2}(z_i - \bar{z}_2)^2}$, $b_{xz(2)} = \frac{\sum_{i\in s_2}(x_i - \bar{x}_2)(z_i - \bar{z}_2)}{\sum_{i\in s_2}(z_i - \bar{z}_2)^2}$ and $r_2 = \frac{\bar{y}_2}{\bar{x}_2}$ respectively be the consistent estimators of β_{yz}, β_{xz} and R based on s_2. Then for computational purposes the optimum estimator shall be defined in the following manner:

$$\hat{\ell}_P^{(G)} = \bar{y}_2\frac{\bar{x}_2 - \left(b_{xz(2)} + \frac{b_{yz(2)}}{r_2}\right)(\bar{z}_2 - \bar{z}_1)}{\bar{x}_1 + \frac{b_{yz(2)}}{r_2}\left(\bar{z}_1 - \overline{Z}\right)}.$$

Note that the use of sample estimates in preference to the respective unknown parameters does not make any change in the asymptotic MSE expressions of the resulting estimator *i.e.*, $M_{min}\left(\ell^{(G)}\right) = M\left(\hat{\ell}_P^{(G)}\right)$.

7. Comparison of $\ell^{(G)}$ With t

As said earlier, the generalized estimator t constructed under the Chand-Kiregyera approach produces a system of estimators covering t_P, t_{RP}, t_{PP} and t_{RGP} as its perspective members. This system also remains as a subclass of estimators of the wider classes of estimators coming out of $\ell^{(G)}$ for $\eta = 0$ and $\delta = d$. But to uphold that our redesigned technique is better than the Chand-Kiregyera method, there is a need to comparison of $\ell^{(G)}$ with t at least regarding MSE. Considering $\eta = 0$ and $\delta = d$, asymptotic expression for the MSE of t can be directly obtained from (17) as

$$M(t) = M(t_P) + \overline{Y}^2 \theta_1 dD\left(dDC_z^2 + 2C_{yz}\right). \quad (22)$$

Hence, we have,

$$M(t) - M\left(\ell^{(G)}\right) = -\overline{Y}^2 D\left[(\theta_2 - \theta_1)\eta\left(\eta DC_z^2 - 2C_{yz} - 2C_{xz}\right)\right.$$

$$\left.+\theta_1(\delta - d)\left\{(\delta + d)DC_z^2 + 2C_{yz}\right\}\right]. \quad (23)$$

$\ell^{(G)}$ is therefore more efficient than t if the expression within the square brackets of (23) is negative which of course results under multiple conditions. But, as in the previous discussions, here we report only some sufficient conditions. Hence, $M\left(\ell^{(G)}\right) \leq M(t)$ if

$$0 < \eta \leq 2\left(\beta_{xz} + \frac{\beta_{yz}}{R}\right) \text{ and } d < \delta \leq -2\left(\frac{\beta_{yz}}{R} + d\right) \quad (24)$$

or

$$2\left(\beta_{xz} + \frac{\beta_{yz}}{R}\right) \leq \eta < 0 \text{ and } -2\left(\frac{\beta_{yz}}{R} + d\right) \leq \delta < d. \quad (25)$$

But if $\delta = d$, then the first two ranges of (24) and (25) remain as sufficient conditions to make $\ell^{(G)}$ more effective than t.

Note that the above derived sufficient conditions favoring $\ell^{(G)}$ of course difficult to check on many occasions. But they clearly indicate that there is scope for improving upon the formulated estimation strategy over that considered in Chand (1975) [1] and Kiregyera (1980, 1984) [2, 3]. However, this dilemma has been clarified below for optimum choices of the coefficients.

The optimum value of d minimizing $M(t)$ is $\hat{d} = -\frac{\beta_{yz}}{R}$, and the resulting minimum MSE bound and the minimum MSE bound estimator are respectively given below according as β_{yz} and R are known or estimated:

$$M_{min}(t) = M(t_P) - \overline{Y}^2 \theta_1 C_y^2 \rho_{yz}^2 \quad (26)$$

$$t^{(G)} = \bar{y}_2 \frac{\bar{x}_2}{\left[\bar{x}_1 + \frac{\beta_{yz}}{R}(\bar{z}_1 - \overline{Z})\right]} \text{ or } \hat{t}^{(G)} = \bar{y}_2 \frac{\bar{x}_2}{\left[\bar{x}_1 + \frac{b_{yz(2)}}{r_2}(\bar{z}_1 - \overline{Z})\right]}.$$

As we see that $M_{min}\left(\ell^{(G)}\right) < M_{min}(t)$, t is less efficient than both $\ell^{(G)}$ regarding minimum MSE bound criterion.

It has already been established that the four estimators t_P, t_{RP}, t_{PP} and t_{RGP}, and the estimators t_{11}, t_{13}, t_{14} and t_{16} considered earlier, are some distinct cases of $\ell^{(G)}$. These estimators are therefore less efficient than their minimum MSE bound estimator $\hat{\ell}_P^{(G)}$. On the same ground, t_P, t_{RP}, t_{PP} and t_{RGP} being some cases of t are always less efficient than $\hat{t}^{(G)}$. Further it is also noted that $M\left(\hat{\ell}_P^{(G)}\right) < M\left(\hat{t}^{(G)}\right)$. Hence, these derived results lead to a conclusion that that t may be inferior to $\ell^{(G)}$.

8. An Alternative Reducible Estimator Under the Redesigned Approach

Let us once again recall the methodology adopted in section 3 to compose $\ell^{(G)}$ under the redesigned approach. As another option, one should also like to prefer the difference estimator $\bar{x}_2 - \omega\left(\bar{z}_2 - \overline{Z}\right)$ instead of $\bar{x}_2 - \eta(\bar{z}_2 - \bar{z}_1)$ to replace $\bar{x}_2$ in t_P. This mechanism produces the following alternative estimator:

$$\ell_A^{(G)} = \bar{y}_2 \frac{\bar{x}_2 - \omega\left(\bar{z}_2 - \overline{Z}\right)}{\bar{x}_1 - \lambda\left(\bar{z}_1 - \overline{Z}\right)}.$$

The estimator also arises as a particular case from the product-type estimator defined in Sharma *et al.* (2014) [12] where no explanation on the construction of the estimator has been provided. But here it is considered under our redesigned approach.

Note that when $\eta = \omega = 0$ and $\delta = \lambda = d$, $\ell^{(G)} = \ell_A^{(G)} = t$, and when $\eta = \omega = 0$ and $\delta = \lambda = 0$, $\ell^{(G)} = \ell_A^{(G)} = t_P$. These results imply that the classes of estimators launched by $\ell^{(G)}$ and $\ell_A^{(G)}$ are certainly overlapping. While analyzing design-based properties of $\ell_A^{(G)}$, the authors established that the class possesses the same minimum MSE bound as that of $\ell^{(G)}$ but with a different MSE bound estimator defined by $\hat{\ell}_{AP}^{(G)} = \bar{y}_2 \frac{\bar{x}_2 - \left(b_{xz(2)} + \frac{b_{yz(2)}}{r_2}\right)(\bar{z}_2 - \bar{Z})}{\bar{x}_1 - b_{xz(2)}(\bar{z}_1 - \bar{Z})}$.

In view of these results, $\ell_A^{(G)}$ is not considered in the present study.

9. Empirical Study

In the usual practice, the foregoing theoretical discussions/comparisons make the identification of a better estimator among others difficult. Hence, at this stage an empirical study has been carried out using data sets of 10 populations as described below for a quantitative analysis of the performance of different estimators. Nevertheless, to make the study manageable, we selected populations with $\rho_{xz} > 0$ so that ratio-in-product and regression-in-product type estimators would be taken into consideration. Hence, the estimators come under attention are t_P, t_{RP}, t_{RGP}, t_{11}, t_{13}, t_{14}, t_{16}, $\hat{t}^{(G)}$ and $\hat{\ell}_P^{(G)}$.

Description of the Populations

- **Population 1** [Montgomery, Peck and Vining (2012, p.556) [9]]: N = 32 automobiles, y = miles/gallon, x = displacement, z = horsepower
- **Population 2** [Gujarati and Porter (2009, p.51) [7]]: N = 27 years, y = civilian unemployment rate, x = civilian labor force participation rate, z = average hourly earnings
- **Population 3** [Bhuyan (2005, p.76) [6]]: N = 28 two times milking cows, y = daily milk production, x = weight of cow after lactation period, z = initial weight
- **Population 4** [Bhuyan (2005, p.77) [6]]: N = 28 three times milking cows, y = daily milk production, x = weight of cow after lactation period, z = initial weight
- **Population 5** [Steel and Torrie (1960, p.282) [13]]: N = 30 locations, y = Log of leaf burns in secs, x = nitrogen percentage, z = chlorine percentage
- **Population 6** [Montgomery, Peck and Vining (2012, p.558) [9]]: N = 27 Belle Ayr Liquefaction Runs, y = oil yield, x = coal total, z = carbon di oxide
- **Population 7** [Johnson and Wichern (2007, p.215) [8]]: N = 20 healthy females, y = sweat rate, x = potassium content, z = sodium content
- **Population 8** [Morrison (1990, p.470) [10]]: N = 26 lighter and heavier underweight young males, y = pigment creatinine, x = Phosphate level, z = calcium level
- **Population 9** [Bhuyan (2005, p.4) [6]]: N = 28 married couples of middle-class families, y = number of ever born children, x = education level of mother, z = education level of father
- **Population 10** [Rawlings, Pantula and Dickey (1998, p.396) [11]]: N = 40 plots (depth 1 and 2), y = sand percentage, x = clay percentage, z = silt percentage

To examine the relative performance of the selected estimators, their percentage relative efficiencies (PREs) compared to the conventional estimator $\bar{y}_2$ whose variance is given by $V(\bar{y}_2) = \bar{Y}^2 \theta_2 C_y^2$, have been computed. These computed values for different values of n_1 and n_2, meeting the restriction $n_2 < \frac{n_1}{2}$, are compiled in table 1.

After scrutiny of the tabulated figures, the findings are summarized in the following manner:

- For all populations being taken into consideration, $\hat{\ell}_P^{(G)}$ gains the maximum precision amongst all in conformity with the theoretical findings.
- Both t_{RP} and t_{RGP} are more efficient than t_P. But, as is anticipated, efficacy of $\hat{t}^{(G)}$ is better than t_P, t_{RP} and t_{RGP} in all cases.
- t_{RGP} seems to be less efficient than t_{RP} for 3 populations. This means that selection of regression estimator in place of ratio estimator as a substitute of $\bar{x}_1$ cannot always enhance efficiency in estimation.
- The estimators t_{11} and t_{14} are undoubtedly preferable to t_{RP}, and both t_{13} and t_{16} are better than t_{RGP}.
- Performance of $\hat{t}^{(G)}$ is better than t_{13} and t_{16} in 4 cases only whereas in other cases it appears to be worse than at least one of them.

Table 1: PREs of Different Estimators

Est.	**Population (n_1, n_2)**									
	1 (12,5)	**2 (10,4)**	**3 (10,4)**	**4 (10,4)**	**5 (12,5)**	**6 (11,5)**	**7 (10,4)**	**8 (10,4)**	**9 (10,4)**	**10 (14,6)**
t_P	166.79	173.49	119.05	122.11	138.11	141.02	130.68	109.14	106.85	152.43
t_{RP}	249.73	177.03	123.22	125.61	141.42	161.37	132.75	120.59	117.83	162.61
t_{11}	252.14	181.51	125.44	132.72	144.85	168.16	136.74	131.09	125.64	179.50
t_{14}	252.92	181.13	128.93	137.90	142.86	163.55	138.11	131.28	130.62	174.28
t_{RGP}	248.27	182.66	127.62	126.68	142.32	160.05	134.29	145.85	116.03	171.32
t_{13}	256.12	199.23	132.58	137.33	149.44	167.14	140.91	152.32	120.21	178.92
t_{16}	263.50	187.15	131.78	130.24	151.92	164.85	144.28	161.58	122.39	175.83
$\hat{t}^{(G)}$	276.76	192.74	129.55	133.36	154.04	162.39	148.57	151.09	139.91	174.94
$\hat{\ell}_P^{(G)}$	306.59	213.75	142.75	147.73	187.16	179.14	164.01	178.25	188.43	192.44

10. Conclusions

Reviewing discussed analytical and empirical results of the present work, it may be finally concluded that the planned estimation method, *i.e*., redesigned approach in connection with the reducible estimator $\ell^{(G)}$ has a greater scope than the Chand-Kiregyera approach and can be successfully used in many survey situations.

11. References

1. Chand L. Some ratio-type estimators are based on two or more auxiliary variables. Unpublished PhD dissertation. Ames: Iowa State University, 1975.
2. Kiregyera B. A chain ratio-type estimator in finite population double sampling using two auxiliary variables. Metrika. 1980; 27:217-223. DOI: 10.1007/BF01893599
3. Kiregyera B. Regression-type estimators using two auxiliary variables and the model of double sampling. Metrika. 1984; 31:215-226. Available from: https://doi.org/10.1007/BF01915203
4. Sukhatme BV, Chand L. Multivariate ratio type estimator. In: Proceedings of the Social Statistics Section, American Statistical Association, 1977, p. 927-931.
5. Sahoo J, Sahoo LN, Nayak SR. Some product-type estimators in double sampling. International Journal of Agricultural and Statistical Sciences. 2006; 2(2):131-135.
6. Bhuyan KC. Multivariate Analysis and its Applications. Kolkata: New Central Book Agency (P) Ltd, 2005.
7. Gujarati DN, Porter DC. Basic Econometrics. 5th Ed. New York: McGraw-Hill, 2009.
8. Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. Upper Saddle River (NJ): Pearson Education, Inc., 2007.
9. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis. 5th Ed. Hoboken (NJ): John Wiley & Sons, Inc., 2012.
10. Morrison DF Multivariate Statistical Methods. 3rd Ed. New York: McGraw-Hill, 1990.
11. Rawlings JO, Pantula SG, Dickey DA. Applied Regression Analysis. 2nd Ed. New York: Springer-Verlag, 1998.
12. Sharma BK, Singh HP, Tailor R. A generalized product method of estimation in two-phase sampling. Research and Reviews: Journal of Statistics. 2014; 3(2):10-17. Available from: www.stmjournals.com
13. Steel GDR, Torrie H. Principles and Procedures of Statistics, New York: McGraw-Hill, 1960.