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*

Follow this and additional works at: [http://digitalcommons.wayne.edu/jmasm](http://digitalcommons.wayne.edu/jmasm)

Part of the [Applied Statistics Commons](http://digitalcommons.wayne.edu/jmasm), [Social and Behavioral Sciences Commons](http://digitalcommons.wayne.edu/jmasm), and the [Statistical Theory Commons](http://digitalcommons.wayne.edu/jmasm)

Recommended Citation


DOI: 10.22237/jmasm/1478002260

Available at: [http://digitalcommons.wayne.edu/jmasm/vol15/iss2/13](http://digitalcommons.wayne.edu/jmasm/vol15/iss2/13)
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, India

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.
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 \( k \)th unit of the finite population \( U = (U_1, U_2, U_3, \ldots, U_N) \). 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:

\( \bar{x}_m, \bar{y}_m, \bar{z}_m, \bar{z}_n \): 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.

\( \rho_{xy}, \rho_{yz}, \rho_{xz} \): 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 \( m \).

\( \beta_{yz} \): Population regression coefficient between the variables \( y \) and \( z \).

\( b_{yx}(n), b_{yz}(m), b_{zx}(m) \): 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

\[
\bar{Y}_r = \frac{\bar{y}_m}{\bar{x}_m} \bar{X}
\]

(1)

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

\[
t_1 = \frac{\bar{y}_m}{\bar{x}_m} \bar{x}_n
\]

(2)

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

\[
t_2 = \bar{y}_m \left( \frac{\bar{x}_n}{\bar{x}_m} \right)^{\alpha}
\]

(3)

where \( \alpha \) 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

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

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

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

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

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

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

$$t_6 = \frac{\bar{y}_m}{k \frac{\bar{x}_n}{\bar{x}_m}} + (1-k) \frac{\bar{x}_m}{\bar{x}_n}$$

Proposed Estimator

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

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

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

**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:

\[
\begin{align*}
\bar{y}_m &= \bar{Y}(1+e_1), \quad x_m = \bar{X}(1+e_2), \quad x_n = \bar{X}(1+e_3), \\
\bar{z}_m &= \bar{Z}(1+e_4), \quad S_{yz}(m) = S_{yz}(1+e_5), \quad \bar{s}_z^2(m) = S_z^2(1+e_6)
\end{align*}
\]

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:

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

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:
EFFICIENT AND UNBIASED ESTIMATION OF POPULATION MEAN

\[
E(e_i^2) = f_m C_y^2, \quad E(e_i^3) = f_m C_x^2, \quad E(e_i^4) = f_i C_x^2, \\
E(e_4^2) = f_m C_z^2, \quad E(e_4^3) = f_m \rho_{yx} C_y C_x, \quad E(e_4^4) = f_i \rho_{yx} C_y C_x, \\
E(e_4^4) = f_m \rho_{yx} C_y C_z, \quad E(e_4^4) = f_i \rho_{yx} C_y C_z, \quad E(e_4^4) = f_m \rho_{xy} C_y C_z, \\
E(e_4^4) = f_m \rho_{xy} C_x C_z, \quad E(e_4^4) = f_i \rho_{xy} C_x C_z, \quad E(e_4^4) = f_m \frac{\mu_{012}}{ZS_{yz}}, \\
E(e_4^4) = f_i \rho_{xy} C_x C_z, \quad E(e_4^4) = f_m \frac{\mu_{003}}{ZS_z^2}, \quad E(e_4^4) = f_i \rho_{xy} C_x C_z, \\
\] (11)

where

\[
f_m = \frac{1}{m} + \frac{1}{N}, \quad f_i = \frac{1}{n} + \frac{1}{N}, \\
\mu_{pqr} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{X})^p (y_i - \bar{Y})^q (z_i - \bar{Z})^r
\]

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

\[
B(T_R) = E(T_R - \bar{Y}) = f_m \beta_{yz} \left( \frac{\mu_{003}}{S_z^2} - \frac{\mu_{012}}{S_{yz}} \right) + f_2 P \left( \bar{Y} C_x^2 - \bar{Y} \rho_{yx} C_y C_x + \bar{Z} \rho_{yz} C_y C_z + f_2 C_x^2 \right) \\
- d_2 f_2 \bar{Y} C_x^2
\]

(12)

\[
M(T_R) = E(T_R - \bar{Y}) = \bar{Y}^2 f_m C_y^2 + \bar{Z}^2 \beta_{yz}^2 f_m C_z^2 - 2 \bar{Y} \bar{Z} \beta_{yz} f_m \rho_{yx} C_y C_z + P^2 \bar{Y}^2 \bar{Z}^2 f_2 C_x^2 \\
+ 2 P \bar{Y} f_2 \left( \bar{Z} \beta_{yz} \rho_{yx} C_y C_z - \bar{Y} \rho_{yx} C_y C_x \right)
\]

(13)

where

\[
P = \sum_{i=1}^{3} i d_i, \quad f_2 = \frac{1}{m} - \frac{1}{n}
\]

(14)
Minimization of mean square error in (13) with respect to $P$ yields its optimum value as

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

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

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

Further, from (14) and (15),

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

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

$$B(T_R) = 0 \quad (18)$$

Thus, from (12),

$$Kd_i \left( 2 - \frac{f_2 C_i^2}{\beta \rho_{yz} C_y C_x} \right) + 3Kd_3 = M \quad (19)$$

where

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

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

177
EFFICIENT AND UNBIASED ESTIMATION OF POPULATION MEAN

\[
\begin{pmatrix}
1 & 1 & 1 \\
1 & 2 & 3 \\
K & 2K - \bar{Y}_2C_i^2 & 3K
\end{pmatrix}
\begin{pmatrix}
d_1 \\
d_2 \\
d_3
\end{pmatrix}
= \begin{pmatrix}
1 \\
R \\
M
\end{pmatrix}
\]

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

\[
d_1 = \frac{3}{2} + \frac{1}{2f_2\bar{Y}C_x^2}\left[M - R(K + f_2\bar{Y}C_x^2)\right]
\]

\[
d_2 = \frac{1}{f_2\bar{Y}C_x^2}(RK - M)
\]

\[
d_3 = \frac{1}{2f_2\bar{Y}C_x^2}\left[(M - RK) + (R - 1)f_2\bar{Y}C_x^2\right]
\]

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

\[
T_R = \left[\frac{3}{2} + \frac{1}{2f_2\bar{Y}C_x^2}\left(M - R(K + f_2\bar{Y}C_x^2)\right)\right]\bar{Y}_m^*(\frac{\bar{X}_n}{\bar{X}_m})
\]

\[
+ \frac{1}{f_2\bar{Y}C_x^2}(RK - M)\bar{Y}_m^*\left(\frac{\bar{X}_n}{\bar{X}_m}\right)^2
\]

\[
+ \left[\frac{1}{2f_2\bar{Y}C_x^2}\left((M - RK) + (R - 1)f_2\bar{Y}C_x^2\right)\right]\bar{Y}_m^*\left(\frac{\bar{X}_n}{\bar{X}_m}\right)^3
\]

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

\[
V(T_R) = f_mS_y^2\left(1 - \rho_{x\bar{Y}}^2\right) - f_2S_y^2\rho_{x\bar{Y}}^2\left(\rho_{x\bar{Y}} - \rho_{x\bar{x}}\rho_{x\bar{Y}}\right)^2
\]

\[\text{(23)}\]

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:
MAJI ET AL.

\[ \begin{align*}
E(e_1^2) &= f_mC_y^2, \\
E(e_2^2) &= f_mC_y^2, \\
E(e_3^2) &= f_tC_z^2, \\
E(e_4^2) &= f_mC_z^2, \\
E(e_5e_2) &= f_m\rho_{yx}C_yC_s, \\
E(e_4e_4) &= f_m\rho_{yx}C_yC_z, \\
E(e_5e_3) &= f_m\rho_{yz}C_yC_z, \\
E(e_3e_3) &= E(e_5e_3) = E(e_4e_4) = 0
\end{align*} \quad (24) \]

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

\[
T_R = \frac{1}{f_1 + 2f_m} \left[ (f_1 + f_m)(3-G) - \frac{1}{\overline{YC}_x^2}(WG-A) \right] \overline{y}_m \left( \frac{\overline{x}_n}{\overline{x}_m} \right) \\
+ \frac{1}{f_1 + 2f_m} \left[ \frac{2}{\overline{YC}_x^2}(WG-A) - f_1(3-G) \right] \overline{y}_m \left( \frac{\overline{x}_n}{\overline{x}_m} \right)^2
\]

\[
+ \left[ 1 - \frac{1}{f_1 + 2f_m} \left\{ f_m(3-G) + \frac{1}{\overline{YC}_x^2}(WG-A) \right\} \right] \overline{y}_m \left( \frac{\overline{x}_n}{\overline{x}_m} \right)^3
\]  

with variance up to the first order of approximations as

\[
V(T_R) = f_mS_y^2 \left( 1 - \rho_{yz}^2 \right) - \frac{f_m^2x_1}{f_1 + f_m} S_y^2 \left( \rho_{yx} - \rho_{yz}\rho_{xz} \right)^2 \quad (25)
\]

where

\[
W = f_1\overline{YC}_x^2 + f_m\overline{YC}_x^2 - f_m\overline{Y}\rho_{yx}C_yC_s + f_m\overline{Z}\beta_{yz}\rho_{yx}C_sC_z
\]

\[
G = \frac{f_m}{f_1 + f_m} \left( \rho_{yx} - \rho_{yz}\rho_{xz} \right) C_y + f_m\beta_{yz}\rho_{yx}C_yC_z
\]

\[
A = f_m\beta_{yz} \left( \frac{H_{112} - H_{003}}{S_y^2} \right)
\]

**Remark 2:** The unique value of the scalars \( d_i \) depend on unknown population parameters such as \( \beta_{yz}, \beta_{xz}, \mu_{112}, \mu_{003}, C_y, C_z, \overline{x}_x, \overline{Y}, \rho_{yx}, \) and \( \rho_{xz} \). 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:**

\[
M(t_1) = \bar{Y}^2 \left[ f_m C_y^2 + f_x C_x^2 - 2 f_z \rho_{yz} C_x C_z \right]
\]

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

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

\[
M(t_4) = \bar{Y}^2 \left[ f_m C_y^2 + f_x C_x^2 + f_z C_z^2 + 2 f_z \rho_{yz} C_y C_x + 2 f_z \rho_{yz} C_y C_z \right]
\]

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

\[
M(t_6) = S_y^2 \left[ f_m - f_z \rho_{yz}^2 \right]
\]

**Case II:**

\[
M(t_1) = \bar{Y}^2 \left[ f_m C_y^2 + f_m C_x^2 + f_z C_z^2 - 2 f_z \rho_{yz} C_x C_z \right]
\]

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

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

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

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

\[
M(t_6) = f_m S_y^2 \left[ 1 - \theta \rho_{yz}^2 \right]
\]
where $\theta = \frac{f_m}{f_1 + f_m}$.

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.
EFFICIENT AND UNBIASED ESTIMATION OF POPULATION MEAN

Table 1. Parametric values of different populations

<table>
<thead>
<tr>
<th>Population</th>
<th>N</th>
<th>n</th>
<th>m</th>
<th>$\bar{Y}_r$</th>
<th>$\rho_{yx}$</th>
<th>$\rho_{yz}$</th>
<th>$\rho_{xz}$</th>
<th>$C_y$</th>
<th>$C_x$</th>
<th>$C_z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>80</td>
<td>34</td>
<td>6</td>
<td>5182.60</td>
<td>0.9100</td>
<td>0.9400</td>
<td>0.9900</td>
<td>0.3500</td>
<td>0.9400</td>
<td>0.7500</td>
</tr>
<tr>
<td>II</td>
<td>34</td>
<td>10</td>
<td>7</td>
<td>201.41</td>
<td>0.9300</td>
<td>0.9000</td>
<td>0.8300</td>
<td>0.7400</td>
<td>0.7600</td>
<td>0.6100</td>
</tr>
<tr>
<td>III</td>
<td>50</td>
<td>20</td>
<td>12</td>
<td>5.69</td>
<td>0.7418</td>
<td>0.5677</td>
<td>0.7346</td>
<td>0.5060</td>
<td>0.0560</td>
<td>0.0919</td>
</tr>
<tr>
<td>IV</td>
<td>25</td>
<td>10</td>
<td>7</td>
<td>183.84</td>
<td>0.9100</td>
<td>0.7500</td>
<td>0.7108</td>
<td>0.0546</td>
<td>0.0546</td>
<td>0.0488</td>
</tr>
<tr>
<td>V</td>
<td>55</td>
<td>30</td>
<td>18</td>
<td>17.08</td>
<td>0.5400</td>
<td>0.5100</td>
<td>0.7108</td>
<td>0.0265</td>
<td>0.0700</td>
<td>0.0265</td>
</tr>
</tbody>
</table>

Table 2. PREs of different estimators (Case I)

<table>
<thead>
<tr>
<th>Population</th>
<th>$\bar{y}_m$</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$T_R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
<td>380.6032</td>
<td>*</td>
<td>*</td>
<td>602.4841</td>
<td>380.6032</td>
<td>859.4502</td>
</tr>
<tr>
<td>II</td>
<td>100</td>
<td>147.7505</td>
<td>148.5310</td>
<td>566.9582</td>
<td>*</td>
<td>156.4518</td>
<td>202.8928</td>
</tr>
<tr>
<td>III</td>
<td>100</td>
<td>128.6936</td>
<td>140.7686</td>
<td>159.0693</td>
<td>*</td>
<td>140.7686</td>
<td>201.7059</td>
</tr>
<tr>
<td>IV</td>
<td>100</td>
<td>122.5372</td>
<td>126.6649</td>
<td>178.8188</td>
<td>*</td>
<td>120.9633</td>
<td>165.7976</td>
</tr>
<tr>
<td>V</td>
<td>100</td>
<td>120.9633</td>
<td>120.9751</td>
<td>131.9087</td>
<td>*</td>
<td>120.9633</td>
<td>165.7976</td>
</tr>
</tbody>
</table>

Table 3. PREs of different estimators (Case II)

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

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

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

where $M(T)$ denotes the MSE/Minimum MSE of an estimator $T$. 

182
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 $\rho_{yx}$, $\rho_{yz}$, $\rho_{xz}$, 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 $\rho_{yx}$, $\rho_{yz}$, and PRE, respectively, and that $\rho_{xz}$ is assumed to be 0.5.

![Graphical Representation](image_url)

**Figure 1.** PRE of $T_R$ (Case I)
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 $\rho_{xz}$, the PRE of the proposed estimator is increasing with increasing values of $\rho_{yx}$ and $\rho_{yz}$. 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


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


EFFICIENT AND UNBIASED ESTIMATION OF POPULATION MEAN


