Improved estimation of population mean under two-phase sampling with subsampling the non-respondents

Housila P. Singh, Sunil Kumar*

School of Studies in Statistics, Vikram University, Ujjain 456010, Ujjain, MP, India

ARTICLE INFO

Article history:
Received 28 January 2010
Received in revised form
28 February 2010
Accepted 3 March 2010
Available online 11 March 2010

Keywords:
Study variate
Auxiliary variate
Bias
Variance
Two-phase sampling
Non-response

ABSTRACT

This paper considers the problem of estimating the population mean $\bar{Y}$ of the study variate $y$ with two auxiliary variates $x$ and $z$ in the presence of non-response using two-phase (double) sampling procedure. Four classes of combined regression and ratio estimators have been defined in four different situations and their properties are studied under large sample approximation. Comparisons of the suggested classes of estimators with usual unbiased estimator $\bar{Y}$ reported by Hansen and Hurwitz (1946), Khare and Srivastava's (1995) estimator, Okafor and Lee's (2000) estimator and Tabasum and Khan's (2004, 2006) estimator have been made. The results obtained are demonstrated with the help of an empirical study.

1. Introduction

In surveys covering human populations, it is observed that information in most cases are not obtained at the first attempt even after some callbacks. An estimate obtained from such incomplete data may be misleading especially when the respondents differ from the non-respondents which results in uncontrollable bias. Hansen and Hurwitz (1946) considered the problem of non-response while estimating the population mean by taking a sub-sample from the non-respondent group with the help of some extra efforts and an estimator was suggested by combining the information available from response and non-response groups. It is well known that in sample surveys precision in estimating the population mean $\bar{Y}$ may be increased by using information on single or multiple auxiliary variables. Using Hansen and Hurwitz (1946) technique several authors including Cochran (1977), Rao (1986, 1987), Khare and Srivastava (1993, 1995, 1997), Okafor and Lee (2000), Lundström and Särndal (2001), Särndal and Lundström (2005), Tabasum and Khan (2004, 2006), Singh and Kumar (2008a, b, c) and Singh and Kumar (2009a, b) have suggested improvement in the estimation procedure for population mean in the presence of non-response using auxiliary variate. Olkin (1958), Mohanty (1967), Rao and Mudholkar (1967) and Srivastava (1971) and others have made the extension of the ratio estimate to the case where multiple auxiliary variables are used to increase the precision. The reader is also refereed to Wu and Luan (2003), Kadilar and Cingi (2004), Sahoo et al. (2003) and Singh et al. (2007). Keeping this in view and using the Hansen and Hurwitz (1946) technique we have suggested four classes of combined regression and ratio estimators and studied their properties.
2. The suggested classes of combined regression and ratio estimators

Let $y$ and $(x,z)$ be the study character and auxiliary characters having $j$th value $(y_j,x_j,z_j)$ $j=1,2,...,N$ with population means $\bar{Y}, \bar{X}$ and $\bar{Z}$ respectively. If the information on an auxiliary variable $x$ highly correlated with $y$ is readily available for all the units of the population, it is well known that regression and ratio type estimators could be used for increased efficiency, incorporating the population mean $\bar{X}$. However, in certain practical situations the population mean $\bar{X}$ is not known a priori in which case the technique of double (two-phase) sampling can be useful. To estimate the population mean $\bar{X}$ of the auxiliary variable, a large first phase sample of size $n'$ units is selected from $N$ units from the population by simple random sampling without replacement (SRSWOR). A smaller second phase sample of size $n$ is selected from $n'$ by SRSWOR and the character $y$ is measured on it. If there is non-response in the second phase sample we may use an estimator obtained from only the respondents or take a sub-sample of the non-respondents and recontact them.

Let us assume that at the first phase, all the $n'$ units supplied information on the auxiliary variable $x$. At the second phase from sample of size $n$, let $n_1$ units supply information on $y$ and $n_2$ refuse to respond. Using Hansen and Hurwitz (1946) approach to sub-sampling, from the $n_2$ non-respondents a sub-sample of size $m$ units is selected at random and is enumerated by direct interview, such that $m = n_2/k$, $k > 1$. Here it is assumed that response is obtained for all the $m$ units.

In the presentation of the further text of this paper we follow the notations and arguments used earlier by Okafor and Lee (2000). We assume that the whole population (denoted by $A$) is stratified into two strata; one is the stratum (denoted by $A_1$) of $N_1$ units, which would respond on the first call at the second phase and the other stratum (denoted by $A_2$) consists of $N_2$ units, which will not respond on the first call at the second phase but will respond on the second call. We note that $N_1$—“responding” and $N_2$—“non-responding” units or the strata sizes $N_1$ and $N_2$ are not known in advance, see Tripathi and Khare (1997, p. 2257). Let the first and second phase samples be denoted by $a'$ and $a$ respectively, and let $a_1 = a \cap A_1$ and $a_2 = a \cap A_2$. The sub-sample of $a_2$ will be denoted by $a_2m$. Summation over the units in a set $s$ will be denoted by $\sum_s$.

When there is non-response on the study variable $y$ as well as on the auxiliary variable $x$, the conventional two-phase ratio and regression estimators for population mean $\bar{Y}$ are defined as

\[
t^*_R = \frac{\bar{Y}^*/\bar{X}^*}{\bar{X}^*}, \quad (\text{Ratio estimator})
\]

\[
t^*_x = \bar{Y}^* + \hat{\beta}_{yx}^* (\bar{X}^* - \bar{X}^*), \quad (\text{Regression estimator})
\]

where $\bar{X}^*$ and $\bar{Y}^*$ are the Hansen–Hurwitz estimators for population means $\bar{X}$ and $\bar{Y}$ respectively and are defined by $\bar{X}^* = \bar{w}_1\bar{x}_1 + \bar{w}_2\bar{x}_2$ and $\bar{Y}^* = \bar{w}_1\bar{y}_1 + \bar{w}_2\bar{y}_2$ with $\bar{w}_1 = (n_1/n)$, $\bar{w}_2 = (n_2/n)$, $(\bar{x}_1, \bar{y}_1)$ and $(\bar{x}_2, \bar{y}_2)$ denote the sample means of $(x,y)$ characters based on $n_1$ and $m$ units respectively, $\hat{\beta}_{yx}$ is the mean of auxiliary variate $x$ based on $n'$ units selected in the first phase sampling, $\hat{\beta}_{yx}^* = \hat{\beta}_{yx}/\hat{s}_x^2$ is an estimator of population regression coefficient $\beta_{yx} = S_{xy}/S_x^2$ of $y$ on $x$, $s_{xy} = (1/(n-1))\sum_{j=1}^n(x_j - \bar{x})^2$, $s_{x}^2 = (1/(n-1))\sum_{j=1}^n(x_j - \bar{x})^2$ (see Okafor and Lee (2000, p. 185)),

\[
S_{xy} = (1/(N-1))\sum_{j=1}^N(x_j - \bar{X})(y_j - \bar{Y}), \quad \hat{s}_x^2 = (1/(N-1))\sum_{j=1}^N(x_j - \bar{X})^2.
\]

It is to be mentioned that the estimators $t^*_R$ and $t^*_x$ were first proposed by Khare and Srivastava (1995) and revisited by Okafor and Lee (2000). Further Tabasum and Khan (2004) suggested the estimator $t^*_z$ and studied its properties.

When there occurs non-response only on the study variate $y$ and the complete information on the auxiliary variate $x$ is available in both first phase and second phase samples of size $n'$ and $n$ respectively, then the conventional double sampling ratio and regression estimators are defined by

\[
t_R = (\bar{Y}^*/\bar{X})\bar{x}, \quad (\text{Ratio estimator})
\]

\[
t_x = \bar{Y} + \hat{\beta}_{yx}^* (\bar{X} - \bar{x}), \quad (\text{Regression estimator})
\]

where $\bar{x}$ is the sample mean of the auxiliary character $x$ based on second phase sample of size $n$. $\hat{\beta}_{yx}^* = \hat{\beta}_{yx}/\hat{s}_x^2$ is the estimate of the population regression coefficient $\beta_{yx}$, $s_x^2 = (1/(n-1))\sum_{j=1}^n(x_j - \bar{x})^2$. It is to be noted that the estimators $t_R$ and $t_x$ were first envisaged by Khare and Srivastava (1995). Further Tabasum and Khan (2006) suggested the ratio estimator $t_2$ and studied its properties.

Now suppose that information on another auxiliary variable $z$ is available. Then motivated by Mohanty (1967) the idea of ratio and regression estimates has been combined together when there are two auxiliary variates $x$ and $z$ which are correlated to the study variate $y$ in presence of non-response using double (two-phase) sampling. Four different classes of combined regression and ratio estimators have been proposed with their properties in four different situations.

\textit{Situation I: We assume that the population mean $\bar{Z}$ of the second auxiliary variable $z$ be known from the previous census and we want to estimate the population mean $\bar{X}$ of the auxiliary variable $x$ using double sampling. Suppose there is non-response on the study variate $y$ as well as on the auxiliary variate $x$ and $z$ in the second phase sample of size $n$. With this background we consider the following class of combined regression and ratio estimators of population mean $\bar{Y}$ as}

\[
t^{(1)}_{R(z)} = \left\{ \bar{Y}^* + \hat{\beta}_{yx}^* (\bar{X}^* - \bar{x}^*) \right\} \frac{\bar{Z}}{\bar{Z} + \alpha(\bar{Z}^* - \bar{Z})},
\]

where $\alpha$ is a suitably chosen constant.
The bias and variance of the estimator $t^{(1)}_{RR(2)}$ to the first degree of approximation as

$$
B(t^{(1)}_{RR(2)}) = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) \sum \sigma C_z^2 (x-K_y) + \left( \frac{1}{n} - \frac{1}{N} \right) \sum \sigma C_z^2 (x-K_y+K_{xy}K_{xz}) \right\}
$$

$$
- \beta_{xy} \left( \frac{N}{N-2} \right) \left( \frac{1}{n} - \frac{1}{N} \right) \left( \frac{\sigma_{11}}{\sigma_{11}^2} - \frac{\sigma_{00}}{\sigma_{00}^2} \right) \right\}
$$

where

$$
W_2 = N_2/N, \quad C_y = S_y/\sqrt{\sum}, \quad C_y(z) = S_y(z)/\sqrt{\sum}, \quad C_x = S_x/\sqrt{\sum}, \quad C_x(z) = S_x(z)/\sqrt{\sum},
$$

$$
C_z = S_z/\sqrt{\sum}, \quad C_z(z) = S_z(z)/\sqrt{\sum}, \quad \rho_{xy} = S_{xy}/S_xS_y, \quad \rho_{xy(z)} = S_{xy(z)}/S_xS_y(z),
$$

$$
\rho_{xy} = S_{xy}/S_xS_z, \quad \rho_{xy(z)} = S_{xy(z)}/S_xS_z(z), \quad \rho_{xz} = S_{xz}/S_xS_z,
$$

$$
\rho_{xz} = S_{xz}/S_xS_z(z), \quad \rho_{xy} = S_{xy}/S_xS_z, \quad \rho_{xz} = S_{xz}/S_xS_z(z), \quad K_{xy} = \rho_{xy}(C_y/C_x), \quad K_{xy} = \rho_{xy}(C_y/C_x),
$$

$$
K_{xz} = \rho_{xz}(C_x/C_z), \quad K_{xz} = \rho_{xz}(C_x/C_z(z), \quad K_{xz} = \rho_{xz}(C_x(z)/C_z(z),\quad K_{xz} = \rho_{xz}(C_x(z)/C_z(z),\quad K_{xz} = \rho_{xz}(C_x(z)/C_z(z),
$$

$$
S_{y_z} = (1/(N-1)) \sum (y_i - \bar{y})^2, \quad S_{y_z} = (1/(N-1)) \sum (z_i - \bar{z})^2,
$$

$$
S_{y_z} = (1/(N-1)) \sum (y_i - \bar{y})(z_i - \bar{z}), \quad S_{y_z} = (1/(N-1)) \sum (x_i - \bar{x})(z_i - \bar{z}),
$$

$$
S_{y_z} = (1/(N-2)) \sum (y_i - \bar{y})^2, \quad S_{y_z} = (1/(N-2)) \sum (x_i - \bar{x})^2,
$$

$$
S_{y_z} = (1/(N-2)) \sum (z_i - \bar{z})^2, \quad S_{y_z} = (1/(N-2)) \sum (y_i - \bar{y})(z_i - \bar{z}),
$$

$$
S_{y_z} = (1/(N-2)) \sum (x_i - \bar{x})(z_i - \bar{z}), \quad \mu_{y(z)} = \frac{1}{N_2} \sum (x_i - \bar{x})^y(z_i - \bar{z})^z,
$$

$$
\mu_{y(z)} = \frac{1}{N} \sum (x_i - \bar{x})^y(z_i - \bar{z})^z, \quad (r,s) being non-negative integers,
$$

$$
\bar{x}_2 = \frac{1}{N_2} \sum x_i, \quad \bar{y}_2 = \frac{1}{N_2} \sum y_i, \quad and \quad \bar{z}_2 = \frac{1}{N_2} \sum z_i,
$$

$$
\text{Var}(t^{(1)}_{RR(2)}) = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) \sum S^2_y + 2\sigma_2 + 2\beta_{xy}S^2_{y(z)} \right\} + \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) \sum S^2_y + 2\sigma_2 + 2\beta_{xy}S^2_{y(z)} \right\}
$$

$$
+ \frac{W_2(k-1)}{n} \left\{ S^2_{y(z)} + \beta_{xy}S^2_{y(z)}(\beta_{xy}S^2_{y(z)} + 2\sigma_2 + 2\beta_{xy}S^2_{y(z)}) \right\}
$$

where

$$
A^* = \beta_{xz} - \beta_{yx}S_{xz}, \quad B^* = \beta_{xz(z)} - \beta_{yx}S_{xz(z)}, \quad B^* = \beta_{xz}S_{z^2}/S_z^2,
$$

$$
\beta_{xz} = S_{xz}/S_{z^2}, \quad \beta_{yx} = S_{yx}/S_y, \quad \beta_{yx} = S_{yx}/S_y, \quad \beta_{yx} = S_{yx}/S_y, \quad \beta_{yx} = S_{yx}/S_y,
$$

$$\beta_{yx(z)} = S_{yx(z)}/S_y(z), \quad \beta_{yx(z)} = S_{yx(z)}/S_y(z), \quad \beta_{yx(z)} = S_{yx(z)}/S_y(z),\quad \beta_{yx(z)} = S_{yx(z)}/S_y(z).
$$

The $\text{Var}(t^{(1)}_{RR(2)})$ is minimum when

$$\lambda = \{A^*/(R^+D^*)\} = \lambda_{10} \quad \text{(say)},$$

$$\lambda = \{A^*/(R^+D^*)\} = \lambda_{10} \quad \text{(say)},$$

where

$$M^* = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) \beta_{yx}S^2_x + \left( \frac{1}{n} - \frac{1}{N} \right) A^* S^2_x + \frac{W_2(k-1)}{n} B^* S^2_x \right\}.
Thus the resulting minimum variance of \(t_{RR(2)}^{(1)}\) is given by

\[
\min \text{Var}(t_{RR(2)}^{(1)}) = \text{Var}(t^*_R) - M^* / D^*.
\]  
(8)

where

\[
\text{Var}(t^*_R) = \left( \frac{1}{n} \right) \left[ \frac{1}{n} - \frac{1}{N} \right] S^2_y + \left( \frac{1}{n} - \frac{1}{N} \right) S^2_y (1 - \rho^2_{xy}) + \frac{W_2(k-1)}{n} \left\{ S^2_{y(2)} + \beta_{yx} S^2_{x(2)} (\beta_{yx} - 2 \beta_{y(2)}) \right\}.
\]
(9)

is an approximate variance of the regression estimator \(t^*_R\) defined at (2).

The variance of the ratio estimator \(t^*_R\) to the first degree of approximation is given by

\[
\text{Var}(t^*_R) = \left( \frac{1}{n} \right) \left[ \frac{1}{n} - \frac{1}{N} \right] S^2_y + \left( \frac{1}{n} - \frac{1}{N} \right) S^2_y (1 - \rho^2_{xy}) + \frac{W_2(k-1)}{n} \left\{ S^2_{y(2)} + \frac{1}{n} \left( S^2_{y(2)} + R^2 S^2_y - 2 R S_{y(y)} + 2 \beta_{yx} S^2_{x(2)} - 2 \beta_{y(2)} S^2_{x(2)} \right) \right\}.
\]
(10)

Setting \(\alpha = 1\) in (5), we get an estimator for population mean \(\bar{Y}\) as

\[
t_{RR(1)}^{(1)} = (\bar{y}^* + \hat{\beta}_{yx} (x - \bar{x}^*)) (\bar{z} / \bar{z}^*).
\]
(11)

To the first degree of approximation, the variance of \(t_{RR(1)}^{(1)}\) is given by

\[
\text{Var}(t^*_R) = \frac{(1 - \rho^2_{xy})}{n} S^2_y + \frac{W_2(k-1)}{n} \left\{ S^2_{y(2)} - 2 \beta_{yx} S^2_{x(2)} (\beta_{yx} - 2 \beta_{y(2)}) + (R^2 + \beta_{yx} S^2_{x(2)} - 2 \beta_{y(2)} S^2_{x(2)}) \right\}.
\]
(12)

The variance of usual unbiased estimator \(\bar{Y}^*\) is given by

\[
\text{Var}(\bar{Y}^*) = \left( \frac{1}{n} \right) S^2_y + \frac{W_2(k-1)}{n} S^2_{y(2)}.
\]
(13)

From (8), (9), (12) and (13), we have

\[
\text{Var}(\bar{Y}^*) - \min \text{Var}(t_{RR(2)}^{(1)}) = \left\{ \left( \frac{1}{n} \right) S^2_y (1 - \rho^2_{xy}) + \frac{W_2(k-1)}{n} \beta_{yx} S^2_{x(2)} (\beta_{yx} - 2 \beta_{y(2)}) + (R^2 + M^*) / D^* \right\} > 0 \quad \text{if} \quad \beta_{y(2)} > \beta_{yx} / 2,
\]
(14)

\[
\text{Var}(t^*_R) - \min \text{Var}(t_{RR(2)}^{(1)}) = (M^* / D^*) > 0,
\]
(15)

\[
\text{Var}(t_{RR(1)}^{(1)}) - \min \text{Var}(t_{RR(2)}^{(1)}) = \frac{(R^2 D^* - M^*)^2}{D^*} > 0.
\]
(16)

It is observed from (14)–(16) that the proposed estimator \(t_{RR(2)}^{(1)}\) with \(\alpha = \alpha_{10}\) is better than the usual unbiased estimator \(\bar{Y}^*\) if \(\beta_{y(2)} > \beta_{yx} / 2\), the regression estimator \(t^*_R\) and ratio estimator \(t_{RR(2)}^{(1)}\). If \(\alpha \neq \alpha_{10}\), then from (9), (12) and (13), it is observed that the suggested class of estimator \(t_{RR(2)}^{(1)}\) is better than:

(i) the usual unbiased estimator \(\bar{Y}^*\) if

\[
M^* - \frac{\sqrt{M^* + D^* P^*}}{R^* D^*} < \alpha < \frac{M^* + \sqrt{M^* + D^* P^*}}{R^* D^*}
\]

where

\[
P^* = \left\{ \left( \frac{1}{n} \right) S^2_y (1 - \rho^2_{xy}) + \frac{W_2(k-1)}{n} \beta_{yx} S^2_{x(2)} (2 \beta_{yx} - 2 \beta_{y(2)}) \right\}.
\]

(ii) the regression estimator \(t^*_R\) if

\[
0 < \alpha < 2 (M^* / (R^* D^*)).
\]

(iii) the ratio estimator \(t_{RR(2)}^{(1)}\) if

\[
M^* - \frac{\sqrt{M^* + D^* F^*}}{R^* D^*} < \alpha < \frac{M^* + \sqrt{M^* + D^* F^*}}{R^* D^*}
\]

where \(F^* = (D^* + 2 R^* M^*)\).

From (9) and (12), we note that \(\text{Var}(t_{RR(1)}^{(1)}) < \text{Var}(t^*_R)\) if

\[
\beta_{y(2)} < R^* / 2 \quad \text{and} \quad \beta_{yx(2)} < R^* / 2.
\]

\[
\begin{align*}
\bar{T}^{(1)}_{RR(2)} &= \bar{T}^*_u, \\
\bar{T}^{(1)}_{RR(3)} &= \bar{T}^*_u + (1-\alpha)u^*, \\
\bar{T}^{(3)}_{RR(2)} &= \frac{\bar{T}^*_u + (1-\alpha)u^*}{Zu + (1-\alpha)}, \\
\bar{T}^{(1)}_{RR(5)} &= \frac{\bar{T}^*_u + (1-\alpha)u^*}{Zu + (1-\alpha)}.
\end{align*}
\]

where \( \bar{T}^*_u \) is a suitably chosen constant. However, it is easily shown that if we consider a class of estimators wider than (17), keeping the form of the estimators \( \bar{T}^{(1)}_{RR(2)} \) and \( \bar{T}^{(1)}_{RR(3)} \) (j = 1 to 10) in view and motivated by Srivastava (1971), we define a class of estimators of population mean \( \bar{Y} \) as

\[
T^* = \bar{T}^*_u T^*(u^*) = (\bar{Y}^* + \hat{\beta}^{*}_{\bar{Y}}(X - \bar{X}^*)) T^*(u^*)
\]

where \( T^*(u^*) \) is a function of \( u^* \) such that \( T^*(1) = 1 \) and satisfying the following conditions:

(i) whatever be the sample chosen, \( u^* \) assumes values in a bounded, closed convex subset \( G \) of one dimensional real space containing the point 1.

(ii) in \( G \), the function \( T^*(u^*) \) is continuous and bounded.

(iii) the first and second partial derivatives of \( T^*(u^*) \) exist and are continuous and bounded in \( G \).

By Taylor’s expansion, it is observed that the bias of \( T^* \) is of order \( n^{-1} \).

Denoting the first order partial derivative of \( T^*(u^*) \) with respect to \( u^* \), at the point \( u^* = 1 \), by

\[
T^*_1 (1) = \frac{\partial T^*}{\partial u^*} |_{u^* = 1},
\]

we obtain the variance of \( T^* \) to the first degree of approximation, as

\[
\text{Var}(T^*) = \text{Var}(\bar{T}^*_u) + \left( \frac{1}{n} \right) R^* D^* T^*_1 (1) + 2R^* M^* T^*_2 (1) = \text{Var}(\bar{T}^*_u) + \left( \frac{1}{n} \right) R^* \left[ R^* D^* T^*_1 (1) + 2M^* \right] T^*_1 (1)
\]

which is minimum when

\[
T^*_1 (1) = -(M^*/R^* D^*).
\]

Thus the resulting minimum variance of \( T^* \) is given by

\[
\text{min} \text{Var}(T^*) = \text{Var}(\bar{T}^*_u) - (M^*/D^*).
\]

Theorem 2.1. To the first degree of approximation

\[
\text{Var}(T^*) \geq \text{Var}(\bar{T}^*_u) - (M^*/D^*)
\]

with equality holding if

\[
T^*_1 (1) = -(M^*/R^* D^*).
\]

It is observed that the class of estimators \( T^* \) does not include the following difference type estimator

\[
\bar{Y}_{dd} = \bar{T}^*_u + d(\bar{Z} - \bar{X}^*) = (\bar{Y}^* + \hat{\beta}^{*}_{\bar{Y}}(\bar{X} - \bar{X}^*)) + d(1-u^*)
\]

where \( d \) is a suitably chosen constant. However, it is easily shown that if we consider a class of estimators wider than (17), defined by

\[
T_{F} = F(\bar{T}^*_u, u^*)
\]

of population mean \( \bar{Y} \), where \( F(\cdot) \) is a function of \( \bar{T}^*_u \) and \( u^* \) such that

\[
F(\bar{Y}, 1) = F(\bar{T}^*_u, u^*) |_{\bar{Y} = \bar{X} = Z} = \bar{Y}.
\]

The minimum asymptotic variance of \( T_{F} \) is equal to (19) and is not reduced. The estimator \( \bar{Y}_{dd} \) is member of the class (21) and attains minimum variance for optimum value of constant \( d \) in (20).
Situation II: When there is non-response on the study variate \( y \) as well as auxiliary variates \( x \) and \( z \) in the second phase sample of size \( n \) under double (two-phase) sampling and the population means \( \bar{X} \) and \( \bar{Z} \) of the two auxiliary variates are unknown. With this background we consider the following class of combined regression and ratio estimators of population mean \( \bar{Y} \) as

\[
t^{(2)}_{(RR)} = \left\{ \bar{Y}^* + \hat{y}_x^* (\bar{X} - \bar{x}^*) \right\} \frac{\bar{Z}}{[\bar{Z} + \gamma (\bar{Z} - \bar{Z}^*)]},
\]

(22)

where \( \bar{Z}^* \) is the mean of second auxiliary variate \( z \) based on \( n' \) units selected in the first phase sample and \( \gamma \) is a suitably chosen constant.

The bias and variance of \( t^{(2)}_{(RR)} \) to the first degree of approximation are

\[
B(t^{(2)}_{(RR)}) = \left[ \left( \frac{1}{n} - \frac{1}{n'} \right) \bar{Y}^* C_z^2 (\bar{y}' - \bar{y}_x + \bar{K}_y \bar{K}_x) + \frac{W_2 (k-1)}{n} \bar{Y}^* (\bar{y}' - \bar{y}_x \bar{K}_x + \bar{K}_y \bar{K}_x) \right] - \beta_{yx} \left\{ \left( \frac{N}{N-2} \right) \left( \frac{1}{n} - \frac{1}{n'} \right) \left( \frac{\mu_{11}}{\mu_{20}} - \frac{\mu_{20}}{\mu_{11}} \right) + \frac{W_2 (k-1)}{n} \left( \frac{\mu_{11}}{\mu_{20}} - \frac{\mu_{20}}{\mu_{11}} \right) \right\}.
\]

(23)

\[
\text{Var}(t^{(2)}_{(RR)}) = \left[ \left( \frac{1}{n} - \frac{1}{n'} \right) S_y^2 + \left( \frac{1}{n} - \frac{1}{n'} \right) \left( S_y^2 (1-\beta_{yx}^2) + \gamma R^* (\gamma R^* - 2A^*) S_y^2 \right) \right] + \frac{W_2 (k-1)}{n} \left( S_y^2 (\beta_{yx}^2 + \beta_{yx} S_{y2} (\beta_{yx} - \beta_{yx} \bar{K}_x), \gamma R^* (\gamma R^* - 2B^*) S_{y2} \right) \right].
\]

(24)

which is minimum when

\[ \gamma = (N*/(R^* D^*)) = \gamma^*_1 \] (say),

where

\[
N^* = \left\{ \left( \frac{1}{n} - \frac{1}{n'} \right) A^* S_y^2 + \frac{W_2 (k-1)}{n} B^* S_y^2 \right\}.
\]

Thus the resulting minimum variance of \( t^{(2)}_{(RR)} \) is given by

\[
\text{min} \text{Var}(t^{(2)}_{(RR)}) = \text{Var}(t^*_{(R)}) - (N^*/D^*),
\]

(25)

where \( \text{Var}(t^*_{(R)}) \) is defined in (9). Substituting \( \gamma = 1 \) in (22), we get

\[
t^{(2)}_{(RR)} = \left\{ \bar{Y}^* + \hat{y}_x^* (\bar{X} - \bar{x}^*) \right\} \left( \bar{Z} / \bar{Z}^* \right)
\]

with the appropriate variance as

\[
\text{Var}(t^{(2)}_{(RR)}) = \left[ \left( \frac{1}{n} - \frac{1}{n'} \right) S_y^2 + \left( \frac{1}{n} - \frac{1}{n'} \right) \left( S_y^2 (1-\beta_{yx}^2) + R^* (\gamma R^* - 2A^*) S_y^2 \right) \right] + \frac{W_2 (k-1)}{n} \left( S_y^2 (\beta_{yx}^2 + \beta_{yx} S_{y2} (\beta_{yx} - \beta_{yx} \bar{K}_x), R^* (\gamma R^* - 2B^*) S_{y2} \right) \right].
\]

(26)

From (9), (13), (25) and (26), we have

\[
\text{Var}(\bar{Y}^*) - \text{min} \text{Var}(t^{(2)}_{(RR)}) = \left\{ \left( \frac{1}{n} - \frac{1}{n'} \right) \beta_{yx}^2 S_y^2 + \frac{W_2 (k-1)}{n} \beta_{yx} S_{y2} (2 \beta_{yx} - \beta_{yx} + (N^*/D^*) \right\} > 0 \text{ if } \beta_{yx} > \beta_{yx}/2,
\]

(27)

\[
\text{Var}(t^*_{(R)}) - \text{min} \text{Var}(t^{(2)}_{(RR)}) = (N^*/D^*) > 0,
\]

(28)

\[
\text{Var}(t^{(2)}_{(RR)}) - \text{min} \text{Var}(t^{(2)}_{(RR)}) = \frac{(N^* - R^* D^*)^2}{D^*} > 0.
\]

(29)

It is noted from (27)–(29) that the proposed estimator \( t^{(2)}_{(RR)} \) with \( \gamma = \gamma^*_1 \) is better than the usual unbiased estimator \( \bar{Y}^* \) if \( \beta_{yx} > \beta_{yx}/2 \), the regression estimator \( t^*_{(R)} \) and the ratio estimator \( t^{(2)}_{(RR)} \). If \( \gamma \) does not coincide with \( \gamma^*_1 \) i.e. \( \bar{Y} \), then from (9), (13) and (25), we envisaged that the suggested class of estimator \( t^{(2)}_{(RR)} \) is better than

(i) the usual unbiased estimator \( \bar{Y}^* \) if

\[
N^* - \sqrt{N^*+D^*C^*} < \gamma < N^* + \sqrt{N^*+D^*C^*} - \frac{R^* D^*}{R^* D^*},
\]

where

\[
C^* = \left\{ \left( \frac{1}{n} - \frac{1}{n'} \right) \beta_{yx}^2 S_y^2 + \frac{W_2 (k-1)}{n} \beta_{yx} S_{y2} (2 \beta_{yx} - \beta_{yx}) \right\}.
\]
(ii) the regression estimator $t^*_g$ if
\[ 0 < \gamma < (2N^*/R_*D^*), \]
(iii) the ratio estimator $t^*_{RR(1)}$ if
\[ \frac{N^* - \sqrt{N^* + D_*C_*}}{R_*D_*} < \gamma < \frac{N^* + \sqrt{N^* + D_*C_*}}{R_*D_*} \]

where $F^* = (D^* + 2R^*M^*)$.

From (9) and (25), we note that $Var(t^*_{RR(1)}) < Var(t^*_g)$ if
\[ A^* < R^*/2 \quad \text{and} \quad B^* < R^*/2. \]

**Remark 2.2.** For estimating the population mean $\bar{Y}$ one can define different estimators like the estimator $t^*_{RR(2)}$. With $v^* = (\bar{Z}^*/\bar{X})$ the following are few examples:

\[
t^*_{RR(2)} = t^*_g v^* = t^*_g \{(1 - \gamma)v^* + (1 - \gamma)\bar{X}^*\},
\]

where $t^*_g$ is a function of $\bar{X}^*$. However, it is easily shown that if we consider a class of estimators wider than (30),

Motivated by Srivastava (1971), we define a class of estimators of population mean $\bar{Y}$ as

\[ T_g = t^*_g T_g(v^*) = (\bar{Y}^* + \hat{\beta}_g (\bar{X} - \bar{X}^*)) T_g(v^*), \]

where $T_g(v^*)$ is a function of $v^*$ such that $g(1) = 1$ satisfying the conditions similar to those given in situation I. It can be easily seen that the bias of $T_g$ is of order $n^{-1}$. Denoting the first order partial derivative of $T_g(v^*)$ with respect to $v^*$, at the point $v^* = 1$, by

\[ T_g(1) = \frac{\partial T_g(v^*)}{\partial v^*} \bigg|_{v^* = 1}, \]

we obtain the variance of $T_g$ to the first degree of approximation, as

\[ Var(T_g) = Var(t^*_g) + \left(1 - \frac{1}{n}\right) R^* \{R^*D^* T_g(1) + 2N^*\} T_g(1) \]

which is minimum when

\[ T_g(1) = -(N^*/R^*D^*). \]

Thus the resulting minimum variance of $T_g$ is given by

\[ \min Var(T_g) = Var(t^*_g) - (N^*/D^*). \]

Now we state the following theorem:

**Theorem 2.2.** To the first degree of approximation

\[ Var(T_g) \geq Var(t^*_g) - (N^*/D^*) \]

with equality holding if

\[ T_g(1) = -(N^*/R^*D^*). \]

It is noted that the class of estimators $T_g$ does not include the following difference type estimator

\[ \bar{Y}^*_{ID} = t^*_g + c(1 - v^*) = (\bar{Y}^* + \hat{\beta}_g (\bar{X} - \bar{X}^*)) + c(1 - v^*) \]

where $c$ is a suitably chosen constant. However, it is easily shown that if we consider a class of estimators wider than (30), defined by

\[ T_C = G(\bar{Y}^*, v^*) \]

of population mean $\bar{Y}$, where $G(\cdot)$ is a function of $t^*_g$ and $v^*$ such that

\[ G(\bar{Y}, 1) = G(t^*_g, v^*)_{(\bar{Y}, \bar{X}, \bar{Z})} = \bar{Y} \]
the minimum asymptotic variance of $T_C$ is equal to (32) and is not reduced. The estimator $\overline{Y}_{1D}^*$ is a member of the class (34) and attains minimum variance for optimum values of constant $c$ in (33).

**Situation III:** Let the population mean $\overline{X}$ of the auxiliary variable $x$ be unknown and the population mean $\overline{Z}$ of the second auxiliary variable $z$ be known. Double (two-phase) sampling is used to estimate population mean $\overline{X}$ and assumed that the complete information on auxiliary variable $x$ is available in both the phases. Further let there be non-response on study variable $y$. With this background we suggest a class of combined regression and ratio estimators of the population $\overline{Y}$ as

$$t_{RR(0)}^{(3)} = \left\{ \overline{Y} + \hat{\beta}_{yx}(\overline{X} - \overline{x}) \right\} \frac{Z}{(Z + \delta(\overline{Z} - \overline{Z}))},$$

where $\delta$ is a suitably chosen constant.

To the first degree of approximation, the bias and variance of $t_{RR(0)}^{(3)}$ are respectively given by

$$B(t_{RR(0)}^{(3)}) = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) \text{Var} C_2^2(\delta - K_{yz}) + \left( \frac{1}{n} - \frac{1}{N} \right) \text{Var} C_2^2(\delta - K_{yz} + A_0 K_{xz}) - \beta_{yx} \left( \frac{N}{N - 2} \right) \left( \frac{1}{n} - \frac{1}{N} \right) \frac{\mu_{21}}{\mu_{11}} \right\} \frac{H_{30} \delta}{H_{20}}$$

$$Var(t_{RR(0)}^{(3)}) = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) \left\{ S_y^2 + 2 \alpha R^*(\alpha R^* - 2 \beta_{yz}) S_z^2 \right\} + \left( \frac{1}{n} - \frac{1}{N} \right) \left\{ S_y^2(1 - \rho_{yz}^2) + 2 \alpha R^*(\alpha R^* - 2 A^*) S_z^2 \right\} + \frac{W_2(k-1)}{n} S_y^2 \right\}$$

which is minimum when

$$\delta = \{Q^*/(R^* S^*)\} = \delta_{11}$$

where

$$Q^* = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) A^* + \left( \frac{1}{n} - \frac{1}{N} \right) \beta_{yz} \right\}, \quad S^* = \left( \frac{1}{n} - \frac{1}{N} \right), \quad A_0 = \beta_{yx}(\overline{X}/\overline{Y})$$

Thus the resulting minimum variance of $t_{RR(0)}^{(3)}$ is given by

$$\min Var(t_{RR(0)}^{(3)}) = Var(t_0) - (\overline{Q}^*/S^*) S_y^2.$$  

(38)

where

$$Var(t_0) = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) S_y^2 + \left( \frac{1}{n} - \frac{1}{N} \right) S_y^2(1 - \rho_{yz}^2) + \frac{W_2(k-1)}{n} S_y^2 \right\}$$

(39)

is an approximate variance of the regression estimator $t_0$ defined at (4). The variance of the ratio estimator $t_R$ (defined at (3)) to the first degree of approximation is given by

$$Var(t_R) = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) S_y^2 + \left( \frac{1}{n} - \frac{1}{N} \right) (S_y^2 + R^2S_z^2 - 2RS_y) + \frac{W_2(k-1)}{n} S_y^2 \right\}.$$  

(40)

Putting $\delta=1$ in (35), we get an estimator for population mean $\overline{Y}$ as

$$t_{RR1}^{(3)} = \{\overline{Y}^* + \hat{\beta}_{yx}(\overline{X} - \overline{x})\}/\overline{Z}$$

with the appropriate variance as

$$Var(t_{RR1}^{(3)}) = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) S_y^2 + \left( \frac{1}{n} - \frac{1}{N} \right) S_y^2(1 - \rho_{yz}^2) + \frac{W_2(k-1)}{n} S_y^2 \right\}.$$  

(41)

From (13), (38), (39) and (41), we have

$$Var(\overline{Y}^*) - \min Var(t_{RR(0)}^{(3)}) = \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) \rho_{yx}^2 S_y^2 + (Q^*/S^*) S_y^2 \right\} > 0,$$  

(42)

$$Var(t_0) - \min Var(t_{RR(0)}^{(3)}) = (Q^*/S^*) S_y^2 > 0,$$  

(43)

$$Var(t_{RR1}^{(3)}) - \min Var(t_{RR(0)}^{(3)}) = \frac{(Q^*/S^*)^2 S_y^2}{S^2} > 0.$$  

(44)

Thus, it is envisaged from (42)–(44) that the proposed estimator $t_{RR(0)}^{(3)}$ with $\delta=\delta_{11}$ is better than the usual unbiased estimator $\overline{Y}^*$, the regression estimator $t_0$ and the ratio estimator $t_{RR1}^{(3)}$. If $\delta$ does not coincide with $\delta_{11}$ i.e. $\delta \neq \delta_{11}$, then from (13), (37) and (39), we envisaged that the suggested class of estimator $t_{RR(0)}^{(3)}$ is better than

(i) the usual unbiased estimator $\overline{Y}^*$ if

$$Q^* - \sqrt{Q^*/R^* S^*} < \delta < Q^* + \sqrt{Q^*/R^* S^*}$$

(45)
where

\[ H^* = \left( \frac{1}{n} - \frac{1}{n'} \right) \beta_{yxz}^2 \]

(iii) the ratio estimator \( t_{R^*} \)

\[ 0 < \delta < 2Q^*/R^*S^* \]

(ii) the regression estimator \( t_j \)

\[ 0 < \delta < 2Q^*/R^*S^* \]

Remark 2.3. With \( u = (\bar{Z}/\bar{Z}) \) the following estimators:

\[ t_{R^*}^{(1)} = t_j u^3, \quad t_{R^*}^{(2)} = t_j (\delta u^{-1} + (1-\delta)u^{-2}), \quad t_{R^*}^{(3)} = t_j (\delta u + (1-\delta)u^2), \]

\[ t_{R^*}^{(4)} = t_j \left[ \frac{1}{(1-\delta)u} \right], \quad t_{R^*}^{(5)} = t_j \left( \frac{1}{u} \right), \quad t_{R^*}^{(6)} = t_j \left( \frac{1}{u} \right), \]

\[ t_{R^*}^{(7)} = t_j (2 - u^2), \quad t_{R^*}^{(8)} = t_j \left( \frac{1}{1 - u} \right), \quad t_{R^*}^{(9)} = t_j \left( \frac{1}{1 + u} \right), \]

\[ t_{R^*}^{(10)} = t_j (\delta - \bar{Z}/(\delta - \bar{Z})), \text{ etc. (where } (\delta, a, b) \text{ are suitably chosen scalars) are members of the class of estimators of population mean } \bar{Y} \text{ defined by} \]

\[ T_h = t_j T_h(u) = \{ \bar{Y}^* + \hat{\beta}_{yxz}^* (\bar{X} - \bar{X}) \} T_h(u) \] (45)

where \( T_h(u) \) is a function of \( u \) such that \( T_h(1) = 1 \) satisfying the conditions similar to those given in situation I. It can be easily observed that the bias of \( T_h \) is of order \( n^{-1} \). Denoting the first order partial derivative of \( T_h(u) \) with respect to \( u \), at the point \( u = 1 \), by \( T_h(1) = \bar{v} T_h(u)/\bar{v} u | u = 1 \), we obtain the variance of \( T_h \) to the first degree of approximation, as

\[ \text{Var}(T_h) = \text{Var}(t_j) + \left( \frac{1-f}{n} \right) R^* \{ R^* T_h(1)S_{y}^2 + 2S_{yx} \} T_h(1) \] (46)

which is minimum when

\[ T_h(1) = -\left( \hat{\beta}_{yxz}/R^* \right). \]

Thus the resulting minimum variance of \( T_h \) is given by

\[ \text{min Var}(T_h) = \text{Var}(t_j) - \left( \frac{1-f}{n} \right) \beta_{yxz}^2 S_{y}^2 \] (47)

Thus we define the following theorem:

**Theorem 2.3.** To the first degree of approximation

\[ \text{Var}(T_h) \geq \text{Var}(t_j) - \left( \frac{1-f}{n} \right) \beta_{yxz}^2 S_{y}^2 \]

with equality holding if

\[ T_h(1) = -\left( \hat{\beta}_{yxz}/R^* \right). \]

It is to be mentioned that the class of estimators \( T_h \) does not include the following difference type estimator

\[ \bar{Y}_d = t_i + q(1-u) = \{ \bar{Y}^* + \hat{\beta}_{yxz}^* (\bar{X} - \bar{X}) \} + q(1-u) \] (48)

where \( q \) is a suitably chosen constant. However, it is easily shown that if we consider a class of estimators wider than (45), defined by

\[ T_d = H(\bar{Y}^*, u) \]

of population mean \( \bar{Y} \), where \( H() \) is a function of \( t_i \) and \( u \) such that

\[ H(\bar{Y}, 1) = H(t_i, u)|_{\bar{Y}, \bar{X}, \bar{Z}} = \bar{Y}. \]
Thus, the minimum asymptotic variance of \( T_{ii} \) is equal to (47) and is not reduced. The estimator \( T_{ii} \) is a member of the class (45) and attains minimum variance for optimum values of constant \( q \) in (48).

**Situation IV:** Let both the population means \( \bar{X} \) and \( \bar{Z} \) of the auxiliary variable \( x \) and \( z \) respectively be unknown and to estimate them double (two-phase) sampling is used. It is assumed that the complete information on auxiliary variables \( x \) and \( z \) are available in both the phrases while there is non-response on study variable \( y \). With this background we suggest a class of combined regression and ratio estimators for population mean \( \bar{Y} \) of the study variate \( y \) as

\[
\hat{t}_{RR(4)}^{(4)} = \left\{ \hat{Y} + \hat{\beta}_{yx} \hat{X} \right\} \frac{z}{\bar{z} + \zeta(\bar{z} - \bar{X})}
\]  

(49)

where \( \zeta \) is a suitably chosen constant.

To the first degree of approximation, the bias and variance of \( \hat{t}_{RR(4)}^{(4)} \) are respectively given by

\[
B(t_{RR(4)}^{(4)}) = \left\{ \left( \frac{1}{n} - \frac{1}{m} \right) \bar{Y} \right\} C_z^2 (\zeta - K_{zy} + K_{xy} K_{xz}) - \beta_{yx} \left( \frac{N}{N - 2} \right) \left\{ \frac{1}{n} - \frac{1}{m} \right\} \left( \frac{\mu_{z1}}{\mu_{t1}} - \frac{\mu_{z0}}{\mu_{t0}} \right)
\]

(50)

\[
\text{Var}(t_{RR(4)}^{(4)}) = \left\{ \left( \frac{1}{n} - \frac{1}{m} \right) S_y^2 \right\} + \left\{ \left( \frac{1}{n} - \frac{1}{m} \right) \left\{ S_y^2 (1 - \rho_{yx}^2) + \zeta R^* (\zeta R^* - 2A^*) S_z^2 \right\} + \frac{W_2 (k - 1)}{n} S_{y(t)}^2 \right\}
\]

(51)

The \( \text{Var}(t_{RR(4)}^{(4)}) \) is minimum when

\( \zeta = A^*/R^* = \zeta_1 \) (say).

Thus the resulting minimum variance of \( t_{RR(4)}^{(4)} \) is given by

\[
\min \text{Var}(t_{RR(4)}^{(4)}) = \text{Var}(t_i) - \left( \frac{1}{n} - \frac{1}{m} \right) A^*_z S_z^2
\]

(52)

where \( \text{Var}(t_i) \) is defined at (37).

Putting \( \zeta = 1 \) in (49), we get an estimator for population mean \( \bar{Y} \) as

\[
\hat{t}_{RR(4)}^{(4)} = \left\{ \hat{Y} + \hat{\beta}_{yx} \hat{X} \right\} \frac{z}{\bar{z} / \bar{X}}
\]

(53)

with the appropriate variance as

\[
\text{Var}(t_{RR(4)}^{(4)}) = \left\{ \left( \frac{1}{n} - \frac{1}{m} \right) S_y^2 \right\} + \left\{ \left( \frac{1}{n} - \frac{1}{m} \right) \left\{ S_y^2 (1 - \rho_{yx}^2) + (R^* - 2A^*) S_z^2 \right\} + \frac{W_2 (k - 1)}{n} S_{y(t)}^2 \right\}
\]

(54)

From (13), (39), (52) and (54), we have

\[
\text{Var}(\hat{Y}) - \min \text{Var}(t_{RR(4)}^{(4)}) = \left\{ \left( \frac{1}{n} - \frac{1}{m} \right) \rho_{yx}^2 S_y^2 \right\} + \left\{ \left( \frac{1}{n} - \frac{1}{m} \right) A^*_z S_z^2 \right\} > 0
\]

(55)

\[
\text{Var}(t_i) - \min \text{Var}(t_{RR(4)}^{(4)}) = \left( \frac{1}{n} - \frac{1}{m} \right) A^*_z S_z^2 > 0
\]

(56)

\[
\text{Var}(t_{RR(1)}^{(4)}) - \min \text{Var}(t_{RR(4)}^{(4)}) = \left( \frac{1}{n} - \frac{1}{m} \right) (R^* - A^*)^2 S_z^2 > 0
\]

(57)

It is observed from (55)–(57) that the proposed estimator \( t_{RR(4)}^{(4)} \) with \( \zeta = \zeta_1 \) is better than the usual unbiased estimator \( \bar{Y} \), the regression estimator \( t_i \) and the ratio estimator \( t_{RR(1)}^{(4)} \). If \( \zeta \) does not coincide with \( \zeta_1 \) i.e. \( \zeta \neq \zeta_1 \), then from (13), (39) and (54), we noted that the suggested class of estimator \( t_{RR(4)}^{(4)} \) is better than

(i) the usual unbiased estimator \( \bar{Y} \) if

\[
\frac{A^* S_z^2 - \sqrt{A^*_z S_z^2 + \rho_{yx}^2 S^2}}{R^* S_z^2} < \zeta < \frac{A^* S_z^2 + \sqrt{A^*_z S_z^2 + \rho_{yx}^2 S^2}}{R^* S_z^2}
\]

(ii) the regression estimator \( t_i \) if

\[ 0 < \zeta < (2A^*/R^*) \]

(iii) the ratio estimator \( t_{RR(1)}^{(4)} \) if

\[
\frac{A^* - \sqrt{A^*_z + R^* P}}{R^*} < \zeta < \frac{A^* + \sqrt{A^*_z + R^* P}}{R^*}
\]

where \( P = R^* - 2A^* \).

From (39) and (54), we have \( \text{Var}(t_{RR(4)}^{(4)}) < \text{Var}(t_i) \) if \( R^* < 2A^* \).
Remark 2.4. With \(w = (\zeta/\zeta')\) the following estimator \(t_{RR(1)}^{(4)}\), one can define the following estimators:

\[
t_{RR(1)}^{(4)(1)} = t_{w}^{*}, \quad t_{RR(2)}^{(4)(2)} = t_{w}(1/z'w^{-1} + (1 - z)w^{-2}), \quad t_{RR(3)}^{(4)(3)} = t_{w}(z'w + (1 - z)w^{2}), \]

\[
t_{RR(4)}^{(4)(4)} = t_{w} \left( \frac{z' + (1 - z)w}{z'w + (1 - z)w^{2}} \right), \quad t_{RR(5)}^{(4)(4)} = t_{w} \left( \frac{z' + (1 - z)w}{z'w + (1 - z)w^{2}} \right), \quad t_{RR(6)}^{(4)(4)} = t_{w} \left( \frac{z' + (1 - z)w}{z'w + (1 - z)w^{2}} \right), \]

\[
t_{RR(7)}^{(4)(4)} = t_{w}(2w - z), \quad t_{RR(8)}^{(4)(4)} = t_{w} \left[ \frac{1 + a(w - 1)}{1 + b(w - 1)} \right], \quad t_{RR(9)}^{(4)(4)} = t_{w} \left[ \frac{z'w^{-1} + (1 - z)w}{z'w + (1 - z)w^{2}} \right], \]

\[
t_{RR(10)}^{(4)(4)} = t_{w}(z' - z'), \quad \text{etc. where } (\delta, a, b) \text{ are suitably chosen scalars. Considering the form of the estimators } t_{RR(1)}^{(4)} \text{ and } t_{RR(2)}^{(4)} \text{ (} j = 1 \text{ to } 10 \text{) we state class of estimators of population mean } \bar{Y} \text{ defined by}
\]

\[
T_{1} = T_{1}(w) = (\bar{Y}^{*} + \beta^{*}(X - X))T_{1}(w)
\]

which is minimum when

\[
T_{1}(w) = -(A*/R^{*})
\]

where \(T_{1}(w) = \partial T_{1}(w)/\partial w|_{w = 1} - 1\).

Thus the resulting minimum variance of \(T_{1}\) is given by

\[
\text{min Var}(T_{1}) = \text{Var}(t_{w}) - \left( \frac{1}{n} - \frac{1}{n} \right) A^{*} S^{2}_{x}.
\]

Thus we established the following theorem:

**Theorem 2.4.** To terms of order \(n^{-1}\)

\[
\text{Var}(T_{1}) \geq \text{Var}(t_{w}) - \left( \frac{1}{n} - \frac{1}{n} \right) A^{*} S^{2}_{x}
\]

with equality holding if

\[
T_{1}(w) = -(A*/R^{*}).
\]

We note that the class of estimators \(T_{1}\) does not include the following difference type estimator

\[
\bar{Y}_{ID}^{*} = t_{1} + m(1 - w) = (\bar{Y}^{*} + \beta^{*}(X - X)) + m(1 - w)
\]

where \(m\) is a suitably chosen constant. However, it is easily shown that if we consider a class of estimators wider than (58), defined by

\[
T_{1} = L(t_{1}, w)
\]

of population mean \(\bar{Y}\), where \(L(\cdot)\) is a function of \(t_{1}\) and \(w\) such that

\[
L(\bar{Y}, 1) = L(t_{1}, w)|_{(\bar{Y}, X, Z)} = \bar{Y}
\]

the minimum asymptotic variance of \(T_{1}\) is equal to (60) and is not reduced. The estimator \(\bar{Y}_{ID}^{*}\) is a member of the class (58) and attains minimum variance for optimum values of constant \(m\) in (61).

3. **Empirical study**

To illustrate the results we consider the data considered earlier by Khare and Sinha (2007). The description of the population is given below:

The data on physical growth of upper socio-economic group of 95 school children of Varanasi under an ICMR study, Department of Paediatrics, B.H.U., during 1983–84 have been taken under study. The first 25% units (i.e. 24 children) have been treated as non-responding units for all the variables \(y, x\) and \(z\) as per the situations I and II. However, for situations III and IV, the first 25% units (i.e. 24 children) were assumed as non-responding units only for the variable \(y\). Here we have taken the study characters and the auxiliary characters as follows:

\(y\): Weight (in kg) of the children, \(x\): skull circumference (in cm) of the children, \(z\): chest circumference (in cm) of the children.
Table 1
Percent relative efficiencies (PREs) of estimators with respect to $\bar{y}^*$ for different values of $k$.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Auxiliary variate(s) used</th>
<th>$(1/k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$(1/5)$</td>
</tr>
<tr>
<td>$\bar{y}^*$</td>
<td>No</td>
<td>100.00</td>
</tr>
<tr>
<td>$t^*_1$</td>
<td>Single auxiliary variate x is used</td>
<td>116.94</td>
</tr>
<tr>
<td>$t^*_2$</td>
<td></td>
<td>112.91</td>
</tr>
<tr>
<td>$t_1$</td>
<td></td>
<td>104.31</td>
</tr>
<tr>
<td>$t_5$</td>
<td></td>
<td>104.07</td>
</tr>
<tr>
<td>$t_{RR(y)}^{(1)}$</td>
<td>Two auxiliary variates x and z are used</td>
<td>225.43</td>
</tr>
<tr>
<td>$t_{RR(y)}^{(2)}$</td>
<td></td>
<td>212.16</td>
</tr>
<tr>
<td>$t_{RR(z)}^{(1)}$</td>
<td></td>
<td>138.96</td>
</tr>
<tr>
<td>$t_{RR(z)}^{(2)}$</td>
<td></td>
<td>137.05</td>
</tr>
<tr>
<td>$t_{RR(1)}^{(1)}$</td>
<td></td>
<td>200.87</td>
</tr>
<tr>
<td>$t_{RR(1)}^{(2)}$</td>
<td></td>
<td>181.33</td>
</tr>
<tr>
<td>$t_{RR(1)}^{(3)}$</td>
<td></td>
<td>136.36</td>
</tr>
<tr>
<td>$t_{RR(1)}^{(4)}$</td>
<td></td>
<td>127.07</td>
</tr>
</tbody>
</table>

For the population, we have $\bar{y} = 19.4968$, $X = 51.1726$, $Z = 55.8611$, $C_y = 0.15613$, $C_x = 0.03006$, $C_z = 0.05860$, $C_{yz} = 0.12075$, $C_{xz} = 0.02478$, $C_{xy} = 0.05402$, $\rho_{yz} = 0.328$, $\rho_{yx} = 0.477$, $\rho_{xz} = 0.846$, $\rho_{zx} = 0.729$, $\rho_{xy} = 0.297$, $\rho_{xz} = 0.570$, $N = 95$, $n = 35$, $n' = 70$, $W_2 = 0.25$. We have computed the percent relative efficiencies (PREs) of different estimators of population mean $\bar{y}$ with respect to usual unbiased estimator $\bar{y}^*$ for different values of $k$, by using the formulae

$$PRE(t, \bar{y}^*) = \frac{Var(\bar{y}^*)}{Var(t)} \times 100,$$

where $t = t^*_1$, $t^*_2$, $t_1$, $t_5$, $t_{RR(x)}, t_{RR(y)}, t_{RR(z)}, t_{RR(1)}, t_{RR(2)}, t_{RR(3)}, t_{RR(4)}$ and $t_{RR(1)}^{(j)}$.

From Table 1, it is seen that the usual unbiased estimator $\bar{y}^*$ is less efficient than all other estimators as it does not use auxiliary information. The PREs of estimators $t^*_1$, $t^*_2$ and $t_{RR(1)}^{(2)}$ decrease as the value of $k$ increases while the PREs of other estimators (i.e. $t_1$, $t_5$, $t_{RR(x)}, t_{RR(y)}, t_{RR(z)}, t_{RR(1)}, t_{RR(2)}, t_{RR(3)}$, and $t_{RR(4)}$) increase as the value of $k$ increases. It is observed that the estimators based on single auxiliary variable $x$ (i.e. $t^*_1$, $t^*_2$, $t_1$ and $t_5$) are less efficient than the proposed class of combined regression and ratio estimators based on two auxiliary variables $x$ and $z$. It is further observed that the estimators $t_{RR(1)}^{(j)}$, $j = 1, 2, 3, 4$ (which are free from constants) are better than the ratio $(t_{RR}^* + t_5)$ and regression $(t^*_1 + t_1)$ estimators with substantial gain in efficiency. The estimator $t_{RR(0)}^{(1)}$ is the best in the sense of having largest efficiency.

Acknowledgement

The authors are thankful to the referee for his valuable suggestions regarding improvement of the paper.

Appendix A

To obtain the bias and variance of the estimator $t_{RR(2)}^{(1)}$ given in (5) we write $\bar{y}^* = \bar{y}(1 + e_0)$, $\bar{X} = \bar{X}(1 + e_1)$, $\bar{X} = \bar{X}(1 + e_1)$, $\bar{X} = \bar{X}(1 + e_2)$, $\bar{X} = \bar{X}(1 + e_2)$, $\hat{\beta}_{yz} = S_{y\bar{x}} = S_{y\bar{x}}$, $S_{y\bar{x}} = S_{y\bar{x}}(1 + e_3)$, $S_{x\bar{z}} = S_{x\bar{z}}(1 + e_4)$ such that

$$E(e_{i0}) = 0 \quad \forall i = 0 \text{ to } 4,$$

$$E(e_{10}^2) = \left( \frac{1-f}{n} \right) C_{x}^2 + \frac{W_x(k-1)}{n} C_{y\bar{z}}^2,$$

$$E(e_{20}^2) = \left( \frac{1-f}{n} \right) C_{x}^2 + \frac{W_x(k-1)}{n} C_{y\bar{z}}^2,$$

$$E(e_{i1}^2) = \left( \frac{1-f}{n} \right) C_{x}^2 + \frac{W_x(k-1)}{n} C_{y\bar{z}}^2,$$

$$E(e_{i2}^2) = \left( \frac{1-f}{n} \right) C_{x}^2 + \frac{W_x(k-1)}{n} C_{y\bar{z}}^2,$$

$$E(e_{i3}^2) = \left( \frac{1-f}{n} \right) C_{x}^2 + \frac{W_x(k-1)}{n} C_{y\bar{z}}^2.$$
\[
E(e_0,e_1) = \left( \frac{1-f}{n} \right) \rho_{y,C_x} + \frac{W_2(k-1)}{n} \rho_{xy(2),C_xC_2},
\]
\[
E(e_0,e_1') = \left( \frac{1-f}{n} \right) \rho_{y,C_x},
E(e_1,e_2) = \left( \frac{1-f}{n} \right) \rho_{e_x,C_z},
E(e_1,e_1') = \left( \frac{1-f}{n} \right) C_1^2.
\]
\[
E(e_0,e_2) = \left( \frac{1-f}{n} \right) \rho_{y,C_x} + \frac{W_2(k-1)}{n} \rho_{yx(2),C_xC_2},
\]
\[
E(e_1,e_3) = \frac{N(N-n)}{(N-1)(N-2)} \frac{\mu_{21}}{nXS_y} + \frac{W_2(k-1)}{n} \frac{\mu_{21(2)}}{X_{S_y}},
\]
\[
E(e_1',e_3) = \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{21}}{nXS_y},
E(e_1',e_4) = \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{30}}{nXS_y^2},
\]
\[
E(e_1,e_4) = \frac{N(N-n)}{(N-1)(N-2)} \frac{\mu_{30}}{nXS_y^2} + \frac{W_2(k-1)}{n} \frac{\mu_{30(2)}}{X_{S_y^2}}.
\]
Expanding \( t_{RR(\alpha)}^{(1)} \) in terms of \( \Upsilon \), we have
\[
t_{RR(\alpha)}^{(1)} = (1 + e_0) + \beta_{yx} \Upsilon (1 + e_3)(e_0 - e_1)(1 + e_4)^{-1}(1 + e_2)^{-1} = (1 + e_0) + A_0(1 + e_3)(e_0 - e_1)(1 + e_4)^{-1}(1 + e_2)^{-1}
\]
where \( A_0 = \beta_{yx} \Upsilon / \Upsilon \).

We assume that \(|e_2| < 1 \) and \(|e_4| < 1 \) so that \((1 + e_2)^{-1} \) and \((1 + e_4)^{-1} \) are expandable in terms of \( e \)'s. Expanding the right hand side (rhs) of (A.1) in terms of \( e \)'s and neglecting terms of \( e \)'s having power greater than two, we have
\[
t_{RR(\alpha)}^{(1)} - \Upsilon = (1 + e_0 + A_0e_3)(e_0 - e_1 + e_4)^{-1}(1 + e_2) - [A_0e_3(e_0 - e_1)e_4 + e_0e_3 - e_1e_3]
\]
Taking expectation of both sides of (A.2), we obtain the bias of \( t_{RR(\alpha)}^{(1)} \) to the first degree of approximation as given in (6).

Squaring both sides of (A.2) and neglecting terms of \( e \)'s having power greater than two we have
\[
t_{RR(\alpha)}^{(2)} - \Upsilon^2 = (1 + e_0 + A_0e_3)(e_0 - e_1)^2 - A_0e_3(e_0e_1 + e_0e_3 + e_0e_4 + e_0e_2 - e_1e_2) - 2A_0(e_0e_1 - e_0e_1) - 2A_0(e_0e_1 - e_0e_1)
\]
Taking expectations of both sides of (A.3), we obtain the variance of \( t_{RR(\alpha)}^{(1)} \) to the first degree of approximation as given by (7).

Appendix B

To obtain the bias and variance of the estimator \( t_{RR(\gamma)}^{(2)} \) defined in (22), we write \( \Upsilon = Z(1 + e_2) \), such that
\[
E(e_0') = 0, \ E(e_0'e_2') = \left( \frac{1-f}{n'} \right) C_2^2, \ E(e_0'e_2') = \left( \frac{1-f}{n'} \right) C_2^2,
\]
\[
E(e_0'e_2') = \left( \frac{1-f}{n'} \right) \rho_{y,C_x} C_x, \ E(e_0'e_2') = \left( \frac{1-f}{n'} \right) \rho_{yx} C_x C_x,
\]
Expressing \( t_{RR(\gamma)}^{(2)} \) in terms of \( e \)'s we have
\[
t_{RR(\gamma)}^{(2)} = (1 + e_0) + \beta_{yx} \Upsilon (e_0 - e_1)(1 + e_3)(1 + e_4)^{-1}(1 + e_2)^{-1} + (1 + e_2)^{-1} - (1 + e_2)^{-1}.
\]
Expanding \( t_{RR(\gamma)}^{(2)} \) and retaining terms of \( e \)'s having power up to two, then subtracting \( \Upsilon \) both sides and squaring both sides and taking expectations we get the bias and variance of \( t_{RR(\gamma)}^{(2)} \) given at (23) and (24) respectively.

Appendix C

To obtain the bias and variance of the estimator \( t_{RR(\gamma)}^{(3)} \) given in (35) we write \( \Upsilon = Z(1 + e_1) \), \( \Upsilon = Z(1 + e_2) \), \( \beta_{yx} = s_y/s_x^2 \), \( s_y = S_y(1 + e_3) \), \( s_x = S_x(1 + e_4) \) such that
\[
E(e_i) = 0 \forall i = 1 \text{ to } 4
\]
Expanding $\ell_{RB}^{(3)}$ in terms of the first degree of approximation and squaring both sides and then taking expectation of both sides, we obtain the bias and variance of $\ell_{RB}^{(3)}$ given at (36) and (37) respectively.

References


Naik, V.D., Gupta, P.C., 1991a. A comparative study among the estimators obtained from weighted (arithmetic, geometric and harmonic) means of mean per unit estimator and ratio/product estimator. Metron XLIX (1–4), 233–244.


Tripathi, T.P., 1980. A general class of estimator’s population ratio. Sankhya 42 (C), 63–75.