

A General Class of Estimators of the Population Mean in Survey Sampling Using Auxiliary Information with Sub Sampling the Non-Respondents

Housila P. Singh¹ · Sunil Kumar²

¹School of Studies in Statistics, Vikram University; ²School of Studies in Statistics, Vikram University

(Received December 2008; accepted January 2009)

Abstract

In this paper we have considered the problem of estimating the population mean \bar{Y} of the study variable y using auxiliary information in presence of non-response. Classes of estimators for \bar{Y} in the presence of non-response on the study variable y only and complete response on the auxiliary variable x is available, have been proposed in different situations *viz.*, (i) population mean \bar{X} is known, (ii) when population mean \bar{X} and variance S_x^2 are known; (iii) when population mean \bar{X} is not known; and (iv) when both population mean \bar{X} and variance S_x^2 are not known; single and two-phase (or double) sampling. It has been shown that various estimators including usual unbiased estimator and the estimators reported by Rao (1986), Khare and Srivastava (1993, 1995) and Tabasum and Khan (2006) are members of the proposed classes of estimators. The optimum values of the first phase sample size n' , second phase sample size n and the sub sampling fraction $1/k$ have been obtained for the fixed cost and the fixed precision. To illustrate foregoing, we have carried out an empirical investigation to reflect the relative performance of all the potentially competing estimators including the one due to Hansen and Hurwitz (1946) estimator, Rao (1986) estimator, Khare and Srivastava (1993, 1995) and Tabasum and Khan (2006) estimator.

Keywords: Asymptotic variance, study variable, auxiliary variable, non-response, double sampling.

1. Introduction

Consider a finite population of size N from which a random sample 's' of size n is drawn without replacement. Consider a characteristic of interest, say, y take the value y_i on the unit i ($i = 1, 2, \dots, N$).

In surveys on human populations, frequently n_1 units respond on the items under examination in the first attempt while remaining $n_2 (= n - n_1)$ units do not provide any response. When non-response occurs in the initial attempt, Hansen and Hurwitz (1946) proposed a double sampling

²Corresponding author: Research Scholar, School of Studies in Statistics, Vikram University, Ujjain-456010, M. P., India. E-mail: hpsujn@rediffmail.com, sunilbhoulgal06@gmail.com.

scheme for estimating population mean comprising the following steps: (i) a simple random sample of size n is selected and the questionnaire is mailed to the sampled units, (ii) a sub sample of size $r = n_2/k$, ($k \geq 1$) from the n_2 non-responding units in the initial attempt is contacted through personal interviews. In this procedure the population is supposed to be consisting of the response stratum of size N_1 and the non-response stratum of size $N_2 = (N - N_1)$. Let $\bar{Y} = \sum_{i=1}^N y_i/N$ and $S_y^2 = \sum_{i=1}^N (y_i - \bar{Y})^2/(N - 1)$ denote the population mean and the population variance of the survey variable y . Let $\bar{Y}_1 = \sum_{i=1}^{N_1} y_i/N_1$ and $S_{y(1)}^2 = \sum_{i=1}^{N_1} (y_i - \bar{Y}_1)^2/(N_1 - 1)$ denote the mean and variance of the response group. Similarly, let $\bar{Y}_2 = \sum_{i=1}^{N_2} y_i/N_2$ and $S_{y(2)}^2 = \sum_{i=1}^{N_2} (y_i - \bar{Y}_2)^2/(N_2 - 1)$ denote the mean and variance of the non-response group. The population mean can be written as $\bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2$, where $W_1 = (N_1/N)$ and $W_2 = (N_2/N)$. Let $\bar{y}_1 = \sum_{i=1}^{n_1} y_i/n_1$ and $\bar{y}_2 = \sum_{i=1}^{n_2} y_i/n_2$ denote the means of the n_1 responding units and the n_2 non-responding units. Further, let $\bar{y}_{2r} = \sum_{i=1}^r y_i/r$ denote the mean of the $r = n_2/k$ sub sampled units. Thus an unbiased estimator, due to Hansen and Hurwitz (1946) of the population mean \bar{Y} is given by

$$\bar{y}^* = w_1\bar{y}_1 + w_2\bar{y}_{2r}, \quad (1.1)$$

where $w_1 = n_1/n$ and $w_2 = n_2/n$ are responding and non-responding proportions in the sample. The variance of \bar{y}^* to terms of order n^{-1} , is given by

$$\text{Var}(\bar{y}^*) = \bar{Y}^2 \left\{ \left(\frac{\lambda'}{n} \right) C_y^2 + \lambda^* C_{y(2)}^2 \right\}, \quad (1.2)$$

where $\lambda' = (N - n)/N$, $\lambda^* = \{W_2(k - 1)\}/n$, $C_y = S_y/\bar{Y}$ and $C_{y(2)} = S_{y(2)}/\bar{Y}$ (see Cochran, 1977, p.371).

If the population size N is large enough so that the finite population correction(fpc) term can be ignored (*i.e.* $\lambda' \cong 1$), then (1.2) reduces to

$$\text{Var}(\bar{y}^*) = \left(\frac{\bar{Y}^2}{n} \right) \{C_y^2 + W_2(k - 1)C_{y(2)}^2\}. \quad (1.3)$$

Let the information for all the units be available on an auxiliary variable x which is correlated with the study variable y . Let x_i be the value of auxiliary variable x on unit i ($i = 1, 2, \dots, N$). Let $\bar{X} = \sum_{i=1}^N x_i/N$ and $S_x^2 = \sum_{i=1}^N (x_i - \bar{X})^2/(N - 1)$ denote the population mean and population variance of the auxiliary variable x .

In some situations, there may not be any non-response on the auxiliary characteristic. Family size, years of education, years of employment, and the like, are the above type of auxiliary variables, see Rao (1986, p.220). In such situations $\bar{x} = \sum_{i=1}^n x_i/n$ and $s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2/(n - 1)$ the sample mean and the sample variance of all the n units respectively are available. When the population mean \bar{X} of the auxiliary variable x is known, Rao (1986) suggested a ratio estimator for the population mean \bar{Y} of the study variable y as

$$t_R = \left(\frac{\bar{y}^*}{\bar{x}} \right) \bar{X} = \bar{y}^* u^{-1}, \quad (1.4)$$

where $u = \bar{x}/\bar{X}$.

Khare and Srivastava (1993) suggested a product estimator for \bar{Y} as

$$t_P = \left(\frac{\bar{y}^*}{\bar{X}} \right) \bar{x} = \bar{y}^* u. \quad (1.5)$$

Following Srivastava (1967) one can define a class of estimators for \bar{Y} as

$$t_s = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}} \right)^\alpha = \bar{y}^* u^{-\alpha}, \tag{1.6}$$

where α is a suitably chosen constant. We note that the estimator t_s reduces to the estimators \bar{y}^* , t_R and t_P for $\alpha = 0, 1$ and -1 respectively. Many other modifications of ratio and product estimators can be proposed like:

$$\begin{aligned} t_1 &= \bar{y}^* \{ \alpha + (1 - \alpha)u \}^{-1}, & t_2 &= \bar{y}^* \left(\frac{\alpha^* + 1}{\alpha^* + u} \right), & t_3 &= \bar{y}^* \left(\frac{\alpha^* - u}{\alpha^* - 1} \right), \\ t_4 &= \bar{y}^* \{ \alpha + (1 - \alpha)u^{-1} \}, & t_5 &= \bar{y}^* \{ \alpha + (1 - \alpha)u \}, & t_6 &= \bar{y}^* \{ 2 - u^\alpha \}, \\ t_7 &= \bar{y}^* \{ \alpha + (1 - \alpha)u^{\alpha_1} \}, \text{ etc.}, \end{aligned}$$

where α , α^* and α_1 are suitably chosen constants. A close look of the estimators t_R , t_P , t_s and t_j ($j = 1, \dots, 7$), suggests authors to define a class of estimators on the lines of Srivastava (1971) as

$$\bar{y}_h = \bar{y}^* h(u), \tag{1.7}$$

where $h(\cdot)$ is a parametric function such that $h(1) = 1$ and which satisfies certain regularity conditions. It can easily be seen that the estimators \bar{y}^* , t_R , t_P , t_s and t_j ($j = 1, \dots, 7$) are members of the class of the estimators \bar{y}_h .

To the first degree of approximation (ignoring fpc term), the variance of \bar{y}_h is given by

$$\text{Var}(\bar{y}_h) = \left(\frac{\bar{Y}^2}{n} \right) \{ C_y^2 + W_2(k - 1)C_{y(2)}^2 + C_x^2 h_1^2(1) + 2\rho C_y C_x h_1(1) \}, \tag{1.8}$$

where $R = \bar{Y}/\bar{X}$, $\rho = S_{xy}/(S_x S_y)$ is the correlation coefficient between y and x for the entire population, $S_{xy} = \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})/(N - 1)$; and $h_1(1)$ denotes the first order partial derivative of the function $h(u)$ with respect to u about the point 'unity'. The variance of \bar{y}_h at (1.8) is minimized for

$$h_1(1) = -\frac{\beta}{R} = -\rho \left(\frac{C_y}{C_x} \right), \tag{1.9}$$

where $\beta = S_{xy}/S_x^2$ is the population regression coefficient of y on x .

Substitution of (1.9) in (1.8) yields the minimum variance of \bar{y}_h as

$$\min \text{Var}(\bar{y}_h) = \left(\frac{\bar{Y}^2}{n} \right) \{ C_y^2(1 - \rho^2) + W_2(k - 1)C_{y(2)}^2 \}. \tag{1.10}$$

Thus we have the following theorem:

Theorem 1.1. *Up to terms of order n^{-1} ,*

$$\text{Var}(\bar{y}_h) \geq \left(\frac{\bar{Y}^2}{n} \right) \{ C_y^2(1 - \rho^2) + W_2(k - 1)C_{y(2)}^2 \}.$$

with equality holding if

$$h_1(1) = -\frac{\beta}{R}.$$

It is to be noted that the lower bound of the variance of \bar{y}_h at (1.10) is the asymptotic variance of linear regression estimator $\bar{y}_{lr} = \bar{y}^* + \hat{\beta}(\bar{X} - \bar{x})$, which shows that ratio estimators of the form (1.7) are asymptotically no more efficient than the linear regression estimator \bar{y}_{lr} , where $\hat{\beta} = \hat{S}_{xy}/s_x^2$, $s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2/(n-1)$ and the estimate \hat{S}_{xy} is based on the available data under the given sampling design.

It is easily shown that if we consider a wider class of estimators

$$\bar{y}_g = g(\bar{y}^*, u) \quad (1.11)$$

of population mean \bar{Y} , where the function $g(\bar{y}^*, u)$ satisfies $g(\bar{Y}, 1) = \bar{Y}$ and $g_1(\bar{Y}, 1) = 1$, $g_1(\bar{Y}, 1)$ denoting the first partial derivative of the function $g(\bar{Y}^*, u)$ with respect to \bar{y}^* about the point $(\bar{Y}, 1)$, the minimum variance of \bar{y}_g is equal to (1.10) and is not reduced. The difference estimator $\bar{y}_d = \bar{y}^* + \alpha(u-1)$ is a member of the class \bar{y}_g at (1.11) but not of the class \bar{y}_h at (1.7).

In this paper motivated by Srivastava and Jhajji (1981) we propose a class of estimators \bar{y}_f (say) which assumes the prior knowledge of both the population mean \bar{X} and the variance S_x^2 of the auxiliary variable x . Asymptotic bias and variance of the proposed class of estimators \bar{y}_f are obtained. A condition is obtained for the variance of \bar{y}_f is minimum. It has been shown that the minimum variance of the estimator \bar{y}_f is smaller than those which use only the ratio of sample mean to population mean of the auxiliary variable x (i.e. that of \bar{y}_h at (1.7)). Double sampling version of the proposed class of estimators \bar{y}_f is also given with its properties.

We assume that a simple random sample of size n is selected from a finite population of size N . For simplicity we assume that the population size N is very large as compared to sample size n so that finite population correction (fpc) terms are ignored see Srivastava and Jhajji (1981). We write

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2} = \frac{\mu_{20}}{\bar{Y}^2}, \quad C_x^2 = \frac{S_x^2}{\bar{X}^2} = \frac{\mu_{02}}{\bar{X}^2}, \quad \rho = \frac{\mu_{11}}{S_y S_x}, \quad \lambda = \frac{\mu_{12}}{\bar{Y} S_x^2}, \quad \gamma_1 = \frac{\mu_{03}}{S_x^3},$$

$$\beta_1 = \gamma_1^2, \quad \beta_2 = \frac{\mu_{04}}{S_x^4}, \quad C_{y(2)}^2 = \frac{S_{y(2)}^2}{\bar{Y}^2} = \frac{\mu_{20(2)}}{\bar{Y}^2}, \quad \mu_{20(2)} = \sum_{j=1}^{N_2} \frac{(y_j - \bar{Y}_2)^2}{N},$$

$\mu_{pq} = \sum_{j=1}^N (y_j - \bar{Y})^p (x_j - \bar{X})^q / N$, (p, q) being non-negative integers.

Let $w = \bar{y}^*/\bar{Y}$, $u = \bar{x}/\bar{X}$ and $v = s_x^2/S_x^2$. Then we have

$$E(w) = E(u) = E(v) = 1, \quad E\{(w-1)^2\} = \left(\frac{1}{n}\right) \{C_y^2 + W_2(k-1)C_{y(2)}^2\},$$

$$E(u-1)^2 = \frac{C_x^2}{n}, \quad E\{(w-1)(u-1)\} = \frac{\rho C_y C_x}{n}$$

and up to terms of order n^{-1} ,

$$E(v-1)^2 = \frac{\beta_2 - 1}{n}, \quad E\{(w-1)(v-1)\} = \frac{\lambda}{n}, \quad E\{(u-1)(v-1)\} = \frac{\gamma_1 C_x}{n}.$$

2. The Proposed Class of Estimators

Let $f(u, v)$ be a function of (u, v) such that $f(1, 1) = 1$ and such that it satisfies the following conditions:

- (i) Whatever be the sample chosen, (u, v) assume values in a bounded, closed convex subset Q , of the two dimensional real space containing the point $(1, 1)$.

- (ii) In Q , the function $f(u, v)$ is continuous and bounded.
- (iii) The first and second partial derivatives of $f(u, v)$ exist and are continuous and bounded in Q .

We propose the estimate of the population mean \bar{Y} , defined by

$$\bar{y}_f = \bar{y}^* f(u, v). \tag{2.1}$$

Under the conditions (i) and (ii) the bias and the variance of the estimator \bar{y}_f , exist since there are only a finite number of possible samples. Expanding $f(u, v)$ about the point $(u, v) = (1, 1)$ in a second order Taylor's series, we obtain

$$\begin{aligned} \bar{y}_f &= \bar{y}^* \left[f(1, 1) + (u - 1)f_1(1, 1) + (v - 1)f_2(1, 1) + \left(\frac{1}{2}\right) \{ (u - 1)^2 f_{11}(u^*, v^*) \right. \\ &\quad \left. + 2(u - 1)(v - 1)f_{12}(u^*, v^*) + (v - 1)^2 f_{22}(u^*, v^*) \} \right] \\ &= \bar{Y} \{ 1 + (w - 1) \} \left[1 + (u - 1)f_1(1, 1) + (v - 1)f_2(1, 1) + \left(\frac{1}{2}\right) \{ (u - 1)^2 f_{11}(u^*, v^*) \right. \\ &\quad \left. + 2(u - 1)(v - 1)f_{12}(u^*, v^*) + (v - 1)^2 f_{22}(u^*, v^*) \} \right] \\ &= \bar{Y} \left[1 + (w - 1) + (u - 1)f_1(1, 1) + (v - 1)f_2(1, 1) + (u - 1)(w - 1)f_1(1, 1) \right. \\ &\quad \left. + (v - 1)(w - 1)f_2(1, 1) + \left(\frac{1}{2}\right) \{ (u - 1)^2 f_{11}(u^*, v^*) + 2(u - 1)(v - 1)f_{12}(u^*, v^*) \right. \\ &\quad \left. + (v - 1)^2 f_{22}(u^*, v^*) \} + \left(\frac{1}{2}\right) (w - 1) \{ (u - 1)^2 f_{11}(u^*, v^*) + 2(u - 1)(v - 1)f_{12}(u^*, v^*) \right. \\ &\quad \left. + (v - 1)^2 f_{22}(u^*, v^*) \} \right], \tag{2.2} \end{aligned}$$

where $u^* = 1 + \theta(u - 1)$, $v^* = 1 + \theta(v - 1)$, $0 < \theta < 1$ and θ may depend on u and v , and $f_{ij}(\cdot)$, $(i, j = 1, 2)$ denotes the second order partial derivatives of the function $f(u, v)$.

Taking expectation and noting that $f_{ij}(u^*, v^*)$, $(i, j = 1, 2)$ are bounded, from (2.2) we get

$$E(\bar{y}_f) = \bar{Y} + o(n^{-1}).$$

Thus the bias of \bar{y}_f is of the order of n^{-1} and hence its contribution to the mean squared error of \bar{y}_f will be of the order of n^{-2} .

Up to the terms of order n^{-1} , the variance of \bar{y}_f is

$$\begin{aligned} \text{Var}(\bar{y}_f) &= E(\bar{y}_f - \bar{Y})^2 \\ &= \bar{Y}^2 E \{ (w - 1) + (u - 1)f_1(1, 1) + (v - 1)f_2(1, 1) \}^2 \\ &= \bar{Y}^2 E \{ (w - 1)^2 + (u - 1)^2 f_1^2(1, 1) + (v - 1)^2 f_2^2(1, 1) + 2(w - 1)(u - 1)f_1(1, 1) \\ &\quad + 2(w - 1)(v - 1)f_2(1, 1) + 2(u - 1)(v - 1)f_1(1, 1)f_2(1, 1) \} \\ &= \frac{\bar{Y}^2}{n} \{ C_y^2 + W_2(k - 1)C_{y(2)}^2 + C_x^2 f_1^2(1, 1) + (\beta_2 - 1)f_2^2(1, 1) \\ &\quad + 2\gamma_1 C_x f_1(1, 1)f_2(1, 1) + 2\rho C_y C_x f_1(1, 1) + 2\lambda f_2(1, 1) \} \tag{2.3} \end{aligned}$$

which is minimized for

$$\left. \begin{aligned} f_1(1, 1) = \varphi_0(\text{say}) &= \frac{\lambda\gamma_1 - \rho C_y(\beta_2 - 1)}{(\beta_2 - \beta_1 - 1)C_x} \\ f_2(1, 1) = \psi_0(\text{say}) &= \frac{\rho C_y \gamma_1 - \lambda}{\beta_2 - \beta_1 - 1} \end{aligned} \right\} \tag{2.4}$$

and the minimum variance is given by

$$\min \text{Var}(\bar{y}_f) = \left(\frac{\bar{Y}^2}{n}\right) \left\{ C_y^2(1 - \rho^2) + W_2(k - 1)C_{y(2)}^2 - \frac{(\rho C_y \gamma_1 - \lambda)^2}{\beta_2 - \beta_1 - 1} \right\}. \tag{2.5}$$

Now we state the following theorem:

Theorem 2.1. *Up to terms of order n^{-1} ,*

$$\text{Var}(\bar{y}_f) \geq \left(\frac{\bar{Y}^2}{n}\right) \left\{ C_y^2(1 - \rho^2) + W_2(k - 1)C_{y(2)}^2 - \frac{(\rho C_y \gamma_1 - \lambda)^2}{\beta_2 - \beta_1 - 1} \right\}.$$

with equality holding if

$$f_1(1, 1) = \varphi_0, \quad f_2(1, 1) = \psi_0,$$

where φ_0 and ψ_0 are defined in (2.4).

From (1.10) and (2.5) we have

$$\min \text{Var}(\bar{y}_h) - \min \text{Var}(\bar{y}_f) = \left(\frac{\bar{Y}^2}{n}\right) \left\{ \frac{(\rho C_y \gamma_1 - \lambda)^2}{\beta_2 - \beta_1 - 1} \right\} \geq 0$$

implying that

$$\min \text{Var}(\bar{y}_f) < \min \text{Var}(\bar{y}_h). \tag{2.6}$$

From (2.6) it is obvious that \bar{y}_f is more efficient than \bar{y}_h with equality if and only if $\rho C_y \gamma_1 = \lambda$, and hence than \bar{y}^* Rao (1986) estimator, t_R , t_P (Khare and Srivastava, 1993, estimator), t_s and t_j ; $j = 1, \dots, 7$.

This leads that the suggested class of estimators \bar{y}_f would be worth using when the relationship between y and x is markedly non-linear and $(\beta_2 - \beta_1 - 1)$ is small.

Some members of the class of estimators represented by \bar{y}_f are

$$\begin{aligned} \bar{y}_{f(1)} &= \bar{y}^* u^\alpha v^\beta, & \bar{y}_{f(2)} &= \bar{y}^* \{1 + \alpha(u - 1) + \beta(v - 1)\}, \\ \bar{y}_{f(3)} &= \bar{y}^* \{1 - \alpha(u - 1) - \beta(v - 1)\}^{-1}, & \bar{y}_{f(4)} &= \bar{y}^* \frac{1 + \alpha(u - 1)}{1 - \beta(v - 1)}, \text{ etc.} \end{aligned}$$

If α and β in above estimators are respectively given by the right hand sides of the equation in (2.4), $\bar{y}_{f(j)}$; $j = 1, \dots, 4$ attain the lower bound of the variance by (2.5).

Next, if we consider a wider class of estimators of \bar{Y} as

$$\bar{y}_F = F(\bar{y}^*, u, v), \tag{2.7}$$

where the function $F(\bar{y}^*, u, v)$ satisfies $F(\bar{Y}, 1, 1) = \bar{Y}$ and $F_1(\bar{Y}, 1, 1) = 1$, $F_1(\bar{Y}, 1, 1)$ denoting the partial derivative of $F(\bar{y}^*, u, v)$ with respect to \bar{y}^* , the minimum variance of \bar{y}_F is equal to (2.5)

and is not reduced. Also the regression type estimator $\bar{y}_{F(1)} = \bar{y}^* + \alpha(u - 1) + \beta(v - 1)$ is a member of the class represented by \bar{y}_F but not of \bar{y}_f . Thus, the minimum variance of the regression type estimator $\bar{y}_{F(1)}$ is attained by estimators from the class \bar{y}_f .

To obtain the bias of \bar{y}_f , we assume that the third order partial derivative of $f(u, v)$ also exist and continuous and bounded. Thus we have

$$\begin{aligned}
 (\bar{y}_f - \bar{Y}) \cong \bar{Y} & \left[(w - 1) + (u - 1)f_1(1, 1) + (v - 1)f_2(1, 1) + (u - 1)(w - 1)f_1(1, 1) \right. \\
 & + (v - 1)(w - 1)f_2(1, 1) + \frac{1}{2} \left\{ (u - 1)^2 f_{11}(1, 1) + 2(u - 1)(v - 1)f_{12}(1, 1) \right. \\
 & \left. \left. + (v - 1)^2 f_{22}(1, 1) \right\} \right]. \tag{2.8}
 \end{aligned}$$

Taking expectation of both sides of (2.8) we get the bias of \bar{y}_f to the first degree of approximation as

$$B(\bar{y}_f) = \left(\frac{\bar{Y}}{2n} \right) \left\{ 2\rho C_y C_x f_1(1, 1) + 2\lambda f_2(1, 1) + C_x^2 f_{11}(1, 1) + 2C_x \gamma_1 f_{12}(1, 1) + (\beta_2 - 1)f_{22}(1, 1) \right\} \tag{2.9}$$

which is same as obtained by Srivastava and Jhajj (1981) in complete response case. Thus we note that the presence of non-response on the study variable y has no influence over the bias.

3. Optimal Values of n and k

The cost function C is considered to be

$$C = c_0 n + c_1 n_1 + c_2 r, \tag{3.1}$$

where

- c_0 : is the initial cost for "setting up" the survey,
- c_1 : is the cost per unit for collecting and processing data obtained from n_1 responding units,
- c_2 : is the cost per unit for contacting the sub-sampled units and for obtaining and processing responses from them.

Usually c_2 is much larger than c_1 because extra efforts are needed for contacting the non-respondents and eliciting responses from them. From (3.1), the expected cost function C is

$$C^* = E(C) = n \left(c_0 + c_1 W_1 + \frac{c_2 W_2}{k} \right). \tag{3.2}$$

Now, minimizing the variance of an estimator Z ; ($Z = \bar{y}^*, \bar{y}_h, \bar{y}_f$) for a given cost C^* (say), is equivalent to minimizing VC for large N .

Let the variance of an estimator Z is given by

$$\text{Var}(Z) = \frac{1}{n} (V_1 - W_2 V_2) + \frac{W_2 V_2 k}{n}, \tag{3.3}$$

where V_1 and V_2 are the coefficient of $1/n$ and k/n respectively in the expressions of the variance of an estimator Z ; ($Z = \bar{y}^*, t_R, \bar{y}_h, \bar{y}_f$).

Then

$$\zeta = VC = W_2 k (c_0 + W_1 c_1) V_2 + (V_1 - W_2 V_2) \frac{c_2 W_2}{k} + \text{terms independent of } n \text{ and } k. \tag{3.4}$$

Differentiating ζ in (3.4) with respect to k and equating to zero, we obtain

$$k_{opt} = \left\{ \left(\frac{c_2}{c_0 + W_1 c_1} \right) \left(\frac{V_1}{V_2} - W_2 \right) \right\}^{\frac{1}{2}}.$$

If the variance of an estimator V is fixed, *i.e.* $V = V_0$, then from (3.3), we obtain

$$n_{opt} = n_0 \left\{ \left(1 + \frac{W_2 V_2}{V_1} \right) (k_{opt} - 1) \right\}^{\frac{1}{2}},$$

where $n_0 = (V_1/V_2)$.

Using the optimum n and k in (3.2), the minimum cost is given by

$$\min C = n_{opt} \left[(c_0 + c_1 W_1 + c_2 W_2) \left\{ \left(\frac{c_2}{c_0 + W_1 c_1} \right) \left(\frac{V_1}{V_2} - W_2 \right) \right\}^{-\frac{1}{2}} \right].$$

Now, if the cost is fixed, *i.e.* $C^* = C_0^*$, then

$$n_{opt} = C_0^* \left(c_0 + W_1 c_1 + \frac{W_2 c_2}{k_{opt}} \right)^{-1}.$$

Using the optimum n and k in (3.3), the minimum variance of Z is

$$\min \text{Var}(Z) = \left(\frac{V_2}{C_0^*} \right) \left(\sqrt{c_0 + c_1 W_1} \sqrt{\frac{V_1}{V_2} - W_2} + W_2 \sqrt{c_2} \right)^2. \quad (3.5)$$

The optimum n and k , along with the minimum (i) cost for given variance $V = V_0$ and (ii) variance for given cost $C^* = C_0^*$ for the estimators Z ; ($Z = \bar{y}^*, t_R, \bar{y}_h, \bar{y}_f$) can be easily obtained.

4. Empirical Study

To illustrate the performance of the proposed strategy, we consider three artificial populations. The values of the parameters are as follows:

Population I

$$C_y = 2.5, \quad C_{y(2)} = 2, \quad C_x = 2, \quad \rho = 0.7, \quad \rho_{(2)} = 0.6, \quad \lambda = 5, \quad \gamma_1 = 2, \quad \beta_1 = 4, \\ \beta_2 = 8, \quad W_2 = 0.25, \quad c_0 = \text{Rs.}1.00, \quad c_1 = \text{Rs.}1.50, \quad c_2 = \text{Rs.}4.00.$$

Population II

$$C_y = 3, \quad C_{y(2)} = 2.5, \quad C_x = 2.5, \quad \rho = 0.8, \quad \rho_{(2)} = 0.7, \quad \lambda = 4, \quad \gamma_1 = 1.5, \quad \beta_1 = 2.25, \\ \beta_2 = 4, \quad W_2 = 0.25, \quad c_0 = \text{Rs.}1.50, \quad c_1 = \text{Rs.}2.00, \quad c_2 = \text{Rs.}5.00.$$

Population III

$$C_y = 3.5, \quad C_{y(2)} = 1.5, \quad C_x = 3, \quad \rho = 0.9, \quad \rho_{(2)} = 0.8, \quad \lambda = 4, \quad \gamma_1 = 1, \quad \beta_1 = 1, \\ \beta_2 = 4, \quad W_2 = 0.25, \quad c_0 = \text{Rs.}2.00, \quad c_1 = \text{Rs.}3.00, \quad c_2 = \text{Rs.}6.00.$$

We have computed the (i) percent relative efficiencies of the estimators $\bar{y}^*, t_R, \bar{y}_h$ and \bar{y}_f with respect to conventional unbiased estimator \bar{y}^* ; (ii) the expected cost for fixed variance and the percent relative efficiency for the fixed cost. Findings are presented in Tables 4.1 and 4.2.

Table 4.1. Percent relative efficiency(PRE.(.)) of the different estimators with respect to \bar{y}^*

Estimator(s)	(1/k)				
	(1/5)	(1/4)	(1/3)	(1/2)	
Population I	\bar{y}^*	100.00	100.00	100.00	100.00
	t_R	141.38	148.00	157.14	170.59
	\bar{y}_h	142.61	149.49	159.04	173.13
	\bar{y}_f	159.22	170.11	185.92	210.91
Population II	\bar{y}^*	100.00	100.00	100.00	100.00
	t_R	160.53	172.44	190.19	219.48
	\bar{y}_h	160.69	172.67	190.49	219.94
	\bar{y}_f	164.33	177.36	196.99	229.99
Population III	\bar{y}^*	100.00	100.00	100.00	100.00
	t_R	315.22	345.05	384.48	440.21
	\bar{y}_h	316.59	346.77	387.83	443.25
	\bar{y}_f	343.60	381.92	433.01	506.32

Table 4.2. Percent relative efficiency and the expected cost

Estimator(s)	k_{opt}	n_{opt}	PRE.(., \bar{y}^*)	n_{opt}	Exp. Cost(in Rs.)
Population I		$C_0 = \text{Rs.}50.00$ (fixed)		$V_0 = 3 \times 10^{-1}$ (fixed)	
\bar{y}^*	3.07	20	100.00	26	26.27
t_R	1.94	19	213.90	8	12.34
\bar{y}_h	1.91	19	218.92	4	12.01
\bar{y}_f	1.49	18	321.84	4	7.61
Population II		$C_0 = \text{Rs.}100.00$ (fixed)		$V_0 = 2 \times 10^{-1}$ (fixed)	
\bar{y}^*	2.74	29	100.00	50	78.07
t_R	1.18	25	398.17	4	14.33
\bar{y}_h	1.17	25	403.86	4	13.95
\bar{y}_f	1.07	24	455.97	2	8.89
Population III		$C_0 = \text{Rs.}150.00$ (fixed)		$V_0 = 1.5 \times 10^{-1}$ (fixed)	
\bar{y}^*	5.39	33	100.00	164	175.23
t_R	2.04	30	575.62	13	37.21
\bar{y}_h	2.03	30	581.03	12	35.24
\bar{y}_f	1.79	29	712.88	3	8.45

It is observed from the Table 4.1 that the proposed class of estimators \bar{y}_f is more efficient than the usual unbiased estimator \bar{y}^* , t_R and \bar{y}_h in all the three populations for different values of k . The PRE.(.)increases as k decreases.

Table 4.2, demonstrated that the proposed class of estimators \bar{y}_f is more efficient than \bar{y}^* , t_R and \bar{y}_h when the cost is fixed. When variance is fixed in advance the estimator \bar{y}_f has less cost as compared to \bar{y}^* , t_R and \bar{y}_h .

5. Estimation of Population in Using Two-Phase Sampling in Presence of Non-Response

This section deals with the problem of estimating the population mean \bar{Y} of the study variable y in the presence of non-response on the study variable y only when the population mean \bar{X} and variance S_x^2 of the auxiliary variable x are not known. Usually, the double sampling procedure is used when it is necessary to use the auxiliary variable to improve the precision of the estimate but the population parameters such as \bar{X} and S_x^2 of the auxiliary variable x are not known. The first phase sample is used to estimate the population mean \bar{X} and variance S_x^2 of the auxiliary variable

x and the second phase sample is used to obtain the required information on the study variable y . We therefore use the double sampling procedure to estimate population mean \bar{X} and variance S_x^2 and then estimate the population mean \bar{Y} of the study variable y , see Cochran (1977).

In the two phase sampling procedure a preliminary large sample of size n' , called the first phase, is drawn from the given population by the method of simple random sampling without replacement (SRSWOR) and only the auxiliary variable x is measured on it. We assume that all the n' units supply information on the auxiliary variable x . Then a sub sample of size n is selected from the first phase sample by the method of SRSWOR and the study variable y is measured on it.

From the second phase sample of n units, let n_1 units supply information on the study variable y and n_2 units do not respond. From n_2 non-respondents; using Hansen and Hurwitz (1946) procedure we again select a sub sample of size $\{r = (n_2/k), k \geq 1\}$ units, assuming all the r units respond. Such a procedure can be used in a household survey where the household size is used as an auxiliary variable x and the family expenditure as the study variable y . Information on the family size can be obtained completely during the household listing, while there may be non-response on the household expenditure, Tabasum and Khan (2006, pp. 74-75). Let $\bar{y}^* = w_1\bar{y}_1 + w_2\bar{y}_{2r}$ be the unbiased estimator of the population mean \bar{Y} due to Hansen and Hurwitz (1946). Let \bar{x} and S_x^2 respectively denote the sample mean and the sample variance of x based on second phase sample of size n , and \bar{x}' and $S_x'^2$ denote respectively the sample mean and the sample variance of x based on first phase sample of size n' . For simplicity we assume throughout that the population size N is large as compared to the sample size n and n' so that finite population correction terms are ignored.

We consider the double sampling version of the class of the estimators \bar{y}_h at (1.7) as

$$\bar{y}_{hd} = \bar{y}^* h(a), \quad (5.1)$$

where $a = (\bar{x}/\bar{x}')$ and $h(a)$ is the function of ' a ' such that $h(1) = 1$. The function $h(a)$ also satisfies certain regularity conditions similar to those given in Srivastava (1971).

To the first degree of approximation, the bias and variance of \bar{y}_{hd} are respectively given by

$$B(\bar{y}_{hd}) = \left(\frac{\bar{Y}}{2}\right) \left(\frac{1}{n} - \frac{1}{n'}\right) \{C_x^2 h_1(1) + 2\rho C_y C_x h_{11}(1)\}, \quad (5.2)$$

$$\text{Var}(\bar{y}_{hd}) = \bar{Y}^2 \left[\frac{1}{n} C_y^2 + \lambda^* C_{y(2)}^2 + \left(\frac{1}{n} - \frac{1}{n'}\right) \{C_x^2 h_1^2(1) + 2\rho C_y C_x h_1(1)\} \right], \quad (5.3)$$

where $h_1(1) = \partial h(\cdot)/\partial a|_{a=1}$ and $h_{11}(1) = \partial^2 h(\cdot)/\partial a^2|_{a=1}$.

The variance of (5.3) is minimized for

$$h_1(1) = -\rho \left(\frac{C_y}{C_x}\right). \quad (5.4)$$

Thus the resulting minimum variance of \bar{y}_{hd} is given by

$$\min \text{Var}(\bar{y}_{hd}) = \bar{Y}^2 \left\{ \left(\frac{1}{n} - \frac{1}{n'}\right) (1 - \rho^2) C_y^2 + \lambda^* C_{y(2)}^2 + \frac{1}{n'} C_y^2 \right\}. \quad (5.5)$$

Thus we established the following theorem:

Theorem 5.1. *Up to the terms of order n^{-1} ,*

$$\text{Var}(\bar{y}_{hd}) \geq \bar{Y}^2 \left\{ \left(\frac{1}{n} - \frac{1}{n'}\right) (1 - \rho^2) C_y^2 + \lambda^* C_{y(2)}^2 + \frac{1}{n'} C_y^2 \right\}.$$

with equality holding if

$$h_1(1) = -\rho \left(\frac{C_y}{C_x} \right).$$

The class of estimators \bar{y}_{hd} at (5.1) is very large. The following estimators:

$$\begin{aligned} \bar{y}_{hd(1)} &= \bar{y}^* a^{-1} \text{ Khare and Srivastava (1993, 1995) and Tabasum and Khan (2006),} \\ \bar{y}_{hd(2)} &= \bar{y}^* a \text{ Khare and Srivastava (1993, 1995),} \\ \bar{y}_{hd(3)} &= \bar{y}^* a^{\alpha_1}, \quad \bar{y}_{hd(4)} = \bar{y}^* (2 - a^{\alpha_1}), \quad \bar{y}_{hd(5)} = \bar{y}^* \{1 + \alpha_1(a - 1)\}, \\ \bar{y}_{hd(6)} &= \bar{y}^* \{1 - \alpha_1(a - 1)\}^{-1}, \quad \bar{y}_{hd(7)} = \bar{y}^* \{\alpha_1 + (1 - \alpha_1)a^{\alpha_2}\}, \end{aligned}$$

etc. are members of the class of estimators \bar{y}_{hd} at (5.1) where α_1 and α_2 are suitably chosen constants. The asymptotic biases and variances of \bar{y}_{hd} ; $j = 1, \dots, 7$ can be easily obtained from (3.2) and (5.3) just putting the suitable values of the derivatives $h_1(1)$ and $h_{11}(1)$.

It is easily shown that if we consider a wider class of estimators of population mean \bar{Y} as

$$\bar{y}_{gd} = g(\bar{y}^*, a), \tag{5.6}$$

where the function $g(\bar{y}^*, a)$ satisfies $g(\bar{Y}, 1) = \bar{Y}$ and $g_1(\bar{Y}, 1) = 1$, $g_1(\bar{Y}, 1)$ denoting the first order partial derivative of $g(\cdot)$ with respect to \bar{y}^* , the minimum variance of \bar{y}_{gd} is equal to (5.5) and is not reduced. The regression type estimator $\bar{y}_{gd} = \bar{y}^* + \alpha_1(a - 1)$ is a member of the class \bar{y}_{gd} but not of the class (5.1).

Now we define a double sampling analogue of the class of estimators \bar{y}_f at (2.1) of the population mean \bar{Y} as

$$\bar{y}_{fd} = \bar{y}^* f(a, b), \tag{5.7}$$

where $a = (\bar{x}/\bar{x}')$, $b = (s_x^2/s_x'^2)$, $s_x'^2 = \sum_{i=1}^{n'} (x_i - \bar{x}')^2 / (n' - 1)$ and $f(a, b)$ is the function of (a, b) such that $f(1, 1) = 1$ and it also satisfies certain regularity conditions similar to those given for \bar{y}_f at (2.1).

To the first degree of approximation the bias and variance of \bar{y}_{fd} at (5.7) are respectively given by

$$\begin{aligned} B(\bar{y}_{fd}) &= \left(\frac{\bar{Y}}{2} \right) \left(\frac{1}{n} - \frac{1}{n'} \right) \{ 2\rho C_y C_x f_1(1, 1) + 2\lambda f_2(1, 1) + C_x^2 f_{11}(1, 1) \\ &\quad + 2\gamma_1 C_x f_{12}(1, 1) + (\beta_2 - 1) f_{22}(1, 1) \}, \end{aligned} \tag{5.8}$$

$$\begin{aligned} \text{Var}(\bar{y}_{fd}) &= \bar{Y}^2 \left[\left(\frac{1}{n} \right) \{ C_y^2 + W_2(k - 1) C_{y(2)}^2 \} + \left(\frac{1}{n} - \frac{1}{n'} \right) \{ 2\rho C_y C_x f_1(1, 1) \right. \\ &\quad \left. + 2\lambda f_2(1, 1) + C_x^2 f_1^2(1, 1) + (\beta_2 - 1) f_2^2(1, 1) + 2\gamma_1 C_x f_1(1, 1) f_2(1, 1) \right]. \end{aligned} \tag{5.9}$$

The variance of \bar{y}_{fd} at (5.9) is minimum when

$$f_1(1, 1) = \varphi_0 \quad \text{and} \quad f_2(1, 1) = \psi_0, \tag{5.10}$$

where φ_0 and ψ_0 are same as defined by (2.4).

Substitution (5.10) in (5.9) yields the minimum variance of \bar{y}_{fd} as

$$\min \text{Var}(\bar{y}_{fd}) = \min \text{Var}(\bar{y}_{hd}) - \left(\frac{1}{n} - \frac{1}{n'} \right) \bar{Y}^2 \left\{ \frac{(\rho C_y \gamma_1 - \lambda)^2}{\beta_2 - \beta_1 - 1} \right\}. \tag{5.11}$$

Thus we state the following theorem:

Theorem 5.2. *To the first degree of approximation,*

$$\text{Var}(\bar{y}_{fd}) \geq \min \text{Var}(\bar{y}_{hd}) - \left(\frac{1}{n} - \frac{1}{n'}\right) \bar{Y}^2 \left\{ \frac{(\rho C_y \gamma_1 - \lambda)^2}{\beta_2 - \beta_1 - 1} \right\}$$

with equality holding if

$$f_1(1, 1) = \varphi_0 \quad \text{and} \quad f_2(1, 1) = \psi_0.$$

It is observed from (5.8) that in two-phase sampling also the presence of non-response on study variable has no effect on bias. The bias expression (5.8) is same as that of Srivastava and Jhajj (1987) class of estimators when there is complete response on both variables y and x .

The class of estimators \bar{y}_{fd} is very wide. The following estimators:

$$\begin{aligned} \bar{y}_{fd(1)} &= \bar{y}^* a^{\alpha_1} b^{\alpha_2}, & \bar{y}_{fd(2)} &= \bar{y}^* \left\{ \frac{1 + \alpha_1(a + 1)}{1 + \alpha_2(b - 1)} \right\}, \\ \bar{y}_{fd(3)} &= \bar{y}^* \{1 + \alpha_1(a - 1) + \alpha_2(b - 1)\}, & \bar{y}_{fd(4)} &= \bar{y}^* \{1 - \alpha_1(a - 1) - \alpha_2(b - 1)\}, \end{aligned}$$

etc. are members of the class \bar{y}_{fd} , where α_1 and α_2 are suitably chosen constants. The biases and variances of these estimators $\bar{y}_{fd(j)}$; $j = 1, \dots, 4$ to the first degree of approximation can be easily obtained from (5.8) and (5.9) just by putting the suitable values of the derivatives $f_1(1, 1)$, $f_2(1, 1)$, $f_{11}(1, 1)$, $f_{12}(1, 1)$ and $f_{22}(1, 1)$.

Further we define a class wider than \bar{y}_{fd} at (5.7) for population mean \bar{Y} as

$$\bar{y}_{Fd} = F(\bar{y}^*, a, b),$$

where $F(\cdot)$ is the function of (\bar{y}^*, a, b) and such that

$$F(\bar{Y}, 1, 1) = \bar{Y} \Rightarrow F_1(\bar{Y}, 1, 1) = \left. \frac{\partial F(\cdot)}{\partial \bar{y}^*} \right|_{(\bar{Y}, 1, 1)} = 1.$$

It can be shown to the first degree of approximation that the minimum variance of \bar{y}_{Fd} is same as given by (5.11) and is not reduced. We note that the regression type estimator $\bar{y}_{Fd(1)} = \bar{y}^* + \alpha_1(a - 1) + \alpha_2(b - 1)$ is a member of the class \bar{y}_{Fd} but not of the class \bar{y}_{fd} at (5.7).

6. Efficiency Comparisons

For the purpose of comparison we give the variance expressions of Khare and Srivastava (1993, 1995) and Tabasum and Khan (2006) estimators $\bar{y}_{hd(1)} = \bar{y}^*(\bar{x}'/\bar{x})$ (ratio estimator) and $\bar{y}_{hd(2)} = \bar{y}^*(\bar{x}/\bar{x}')$ (product estimator) to the first degree of approximation respectively as

$$\text{Var}(\bar{y}_{hd(1)}) = \bar{Y}^2 \left[\left(\frac{1}{n}\right) \{C_y^2 + W_2(k - 1)C_{y(2)}^2\} + \left(\frac{1}{n} - \frac{1}{n'}\right) (C_x^2 - 2\rho C_y C_x) \right], \quad (6.1)$$

$$\text{Var}(\bar{y}_{hd(2)}) = \bar{Y}^2 \left[\left(\frac{1}{n}\right) \{C_y^2 + W_2(k - 1)C_{y(2)}^2\} + \left(\frac{1}{n} - \frac{1}{n'}\right) (C_x^2 + 2\rho C_y C_x) \right]. \quad (6.2)$$

Expressions in (6.1) and (6.2) can easily be obtained from (5.3) just by putting $h_1(1) = -1$ and 1 respectively.

From (1.3), (5.5), (5.11), (6.1) and (6.2) we have

$$\text{Var}(\bar{y}^*) - \min \text{Var}(\bar{y}_{hd}) = \left[\bar{Y}^2 \left(\frac{1}{n} - \frac{1}{n'} \right) \rho^2 C_y^2 \right] \geq 0, \tag{6.3}$$

$$\text{Var}(\bar{y}_{hd(1)}) - \min \text{Var}(\bar{y}_{hd}) = \left[\bar{Y}^2 \left(\frac{1}{n} - \frac{1}{n'} \right) (C_x - \rho C_y) \right] \geq 0, \tag{6.4}$$

$$\text{Var}(\bar{y}_{hd(2)}) - \min \text{Var}(\bar{y}_{hd}) = \left[\bar{Y}^2 \left(\frac{1}{n} - \frac{1}{n'} \right) (C_x + \rho C_y) \right] \geq 0, \tag{6.5}$$

$$\text{Var}(\bar{y}_{fd}) - \min \text{Var}(\bar{y}_{hd}) = \left[\bar{Y}^2 \left(\frac{1}{n} - \frac{1}{n'} \right) \left\{ \frac{(\rho\gamma_1 C_y - \lambda)^2}{\beta_2 - \beta_1 - 1} \right\} \right] \geq 0. \tag{6.6}$$

It is clear from (6.3) to (6.6) that the suggested class of estimators \bar{y}_{fd} is more efficient than usual unbiased estimator \bar{y}^* , Khare and Srivastava (1993, 1995) and Tabasum and Khan (2006) estimators $\bar{y}_{hd(1)}$ and $\bar{y}_{hd(2)}$ and the class of estimators \bar{y}_{hd} .

7. Determination of n , n' and k

Let D_0 be the total cost (fixed) of the survey apart from overhead cost. The expected total cost of the survey apart from overhead cost is given by

$$D' = D_1 n' + n \left(D_1 + D_2 W_1 + D_3 \frac{W_2}{k} \right), \tag{7.1}$$

where D_1' is the cost per unit of identifying and observing auxiliary character; D_1 is the cost per unit of mailing questionnaire/visiting the unit at the second phase; D_2 is the cost per unit of collecting and processing data obtained from n_1 responding units; D_3 is the cost per unit of obtaining and processing data (after extra effort) from the sub-sampled units; and $W_1 = (N_1/N)$ and $W_2 = (N_2/N)$ denote the response and non-response rate in the population.

Let the variance of an estimator t in presence of non-response using two-phase sampling scheme be represented by

$$\text{Var}(t) = \frac{V_0}{n} + \frac{V_1}{n'} + \frac{k}{n} V_2 + \text{terms independent of } n, n' \text{ and } k, \tag{7.2}$$

where V_0 , V_1 and V_2 are respectively the coefficient of the terms of $1/n$, $1/n'$ and k/n in the expressions of variance of $t = \bar{y}^*$, $\bar{y}_{hd(1)}$, \bar{y}_{hd} and \bar{y}_{fd} .

For Fixed Cost $D' \leq D^*$

We define a function for minimizing the variance for the fixed cost $D' \leq D^*$ and to obtain the optimum values of n , n' and k , we define a function by

$$\phi = \text{Var}(t) + \lambda_i \left\{ D_1' n' + n \left(D_1 + W_1 D_2 + \frac{W_2 D_3}{k} \right) - D^* \right\}, \tag{7.3}$$

where λ_i is Lagrange's multiplier.

Now differentiating ϕ with respect to n , n' and k , and equating to zero, we get

$$n' = \sqrt{\frac{V_1}{\lambda_i D_1'}}, \tag{7.4}$$

$$n = \sqrt{\frac{V_0 + kV_2}{\lambda_i \left(D_1 + D_2W_1 + \frac{D_3W_2}{k} \right)}} \quad (7.5)$$

and

$$\frac{n}{k} = \sqrt{\frac{V_2}{\lambda_i D_3 W_2}}. \quad (7.6)$$

Putting the value of n from (7.5) in (7.6), we have

$$k_{opt} = \sqrt{\frac{V_0 D_3 W_2}{(D_1 + D_2 W_1) V_2}}. \quad (7.7)$$

Substituting the values of n' and n from (7.4) and (7.5) and using k_{opt} from (7.7) in (7.1) respectively, we have

$$\sqrt{\lambda_i} = \frac{1}{D^*} \left\{ \sqrt{D'_1 V_1} + \sqrt{\left(D_1 + D_2 W_1 + \frac{D_3 W_2}{k_{opt}} \right) (V_0 + k_{opt} V_2)} \right\}. \quad (7.8)$$

Thus the minimum value of $\text{Var}(t)$ is given as

$$\min \text{Var}(t) = \frac{1}{D^*} \left\{ \sqrt{D'_1 V_1} + \sqrt{\left(D_1 + D_2 W_1 + \frac{D_3 W_2}{k_{opt}} \right) (V_0 + k_{opt} V_2)} \right\}^2. \quad (7.9)$$

For Fixed Variance $V = V^*$

From (7.7), we see that the optimum value of k is independent of the total cost or specified precision. Let V^* be the variance, fixed in advance, so we have

$$V^* = \frac{V_0}{n} + \frac{V_1}{n'} + \frac{k}{n} V_2. \quad (7.10)$$

Thus, we define a function φ^* for obtaining the optimum values of n , n' and k as

$$\varphi^* = \left\{ D'_1 n' + n \left(D_1 + D_2 W_1 + \frac{D_3 W_2}{k} \right) \right\} + \mu_i (\text{Var}(t) - V^*), \quad (7.11)$$

where μ_i is Lagrange's multiplier.

Now differentiating (7.11) with respect to n , n' and k , and equating to zero we have

$$n' = \sqrt{\frac{\mu_i V_1}{D'_1}}, \quad (7.12)$$

$$n = \sqrt{\frac{\mu_i (V_0 + k V_2)}{D_1 + D_2 W_1 + \frac{D_3 W_2}{k}}}, \quad (7.13)$$

$$\frac{n}{k} = \sqrt{\frac{\mu_i V_2}{D_3 W_2}}. \quad (7.14)$$

Substituting the values of n from (7.13) in (7.14), we have

$$k_{opt} = \sqrt{\frac{V_0 D_3 W_2}{(D_1 + D_2 W_1) V_2}}. \quad (7.15)$$

Table 8.1. Percent relative efficiency(PRE.(.)) of the different estimators of \bar{Y} with respect to \bar{y}^* .

	Estimator(s)	(1/k)			
		(1/5)	(1/4)	(1/3)	(1/2)
Population I $n = 30, n' = 50$	\bar{y}^*	100.00	100.00	100.00	100.00
	$\bar{y}_{hd(1)}$	109.62	110.78	112.24	114.17
	\bar{y}_{hd}	123.36	124.82	126.73	129.25
	\bar{y}_{fd}	126.79	128.73	131.26	134.65
Population II $n = 50, n' = 70$	\bar{y}^*	100.00	100.00	100.00	100.00
	$\bar{y}_{hd(1)}$	112.13	113.70	115.75	118.51
	\bar{y}_{hd}	112.21	113.85	115.92	118.77
	\bar{y}_{fd}	112.71	114.42	116.59	119.58
Population III $n = 30, n' = 50$	\bar{y}^*	100.00	100.00	100.00	100.00
	$\bar{y}_{hd(1)}$	136.29	138.31	140.59	143.13
	\bar{y}_{hd}	137.57	139.81	142.29	145.12
	\bar{y}_{fd}	139.44	141.82	144.46	147.48

Table 8.2. Percent relative efficiency and the expected cost.

Estimator(s)	k_{opt}	n'_{opt}	n_{opt}	PRE.(., \bar{y}^*)	n'_{opt}	n_{opt}	Exp. Cost(in Rs.)
Population I		$C_0 = \text{Rs.}30.00$ (fixed)			$V_0 = 3 \times 10^{-1}$ (fixed)		
		$D'_1 = 0.25, D_1 = 1, D_2 = 1.5, D_3 = 4$					
\bar{y}^*	1.57	-	9	100.00	-	29	100.47
$\bar{y}_{hd(1)}$	1.03	25	7	182.73	47	14	54.86
\bar{y}_{hd}	1.02	27	6	186.11	47	11	54.27
\bar{y}_{fd}	0.82	32	7	218.48	49	10	46.38
Population II		$C_0 = \text{Rs.}100.00$ (fixed)			$V_0 = 2 \times 10^{-1}$ (fixed)		
		$D'_1 = 0.50, D_1 = 1.5, D_2 = 2, D_3 = 5$					
\bar{y}^*	1.41	-	26	100.00	-	48	187.49
$\bar{y}_{hd(1)}$	0.67	64	14	131.36	91	20	142.68
\bar{y}_{hd}	0.67	64	14	132.01	90	20	142.02
\bar{y}_{fd}	0.63	60	15	137.43	74	18	136.39
Population III		$C_0 = \text{Rs.}75.00$ (fixed)			$V_0 = 4 \times 10^{-1}$ (fixed)		
		$D'_1 = 0.75, D_1 = 2, D_2 = 3, D_3 = 6$					
\bar{y}^*	2.71	-	16	100.00	-	33	158.52
$\bar{y}_{hd(1)}$	1.06	43	8	155.12	58	10	102.19
\bar{y}_{hd}	1.06	43	8	155.38	58	10	102.02
\bar{y}_{fd}	0.94	62	18	308.09	42	13	51.45

Now, putting the values of n, n' and k , from (7.13), (7.12) and (7.15) in (7.11), we get

$$\sqrt{\mu_i} = \frac{1}{V^*} \left\{ \sqrt{D'_1 V_1} + \sqrt{\left(D_1 + D_2 W_1 + \frac{D_3 W_2}{k_{opt}} \right) (V_0 + k_{opt} V_2)} \right\}. \tag{7.16}$$

Thus, the minimum expected total cost would be given by

$$\min D^* = \frac{1}{V^*} \left\{ \sqrt{D'_1 V_1} + \sqrt{\left(D_1 + D_2 W_1 + \frac{D_3 W_2}{k_{opt}} \right) (V_0 + k_{opt} V_2)} \right\}^2. \tag{7.17}$$

8. Empirical Study

To show the performance of the proposed strategy, we consider same three artificial populations, which are given in Section 4.

Table 8.1 shows that the estimator \bar{y}_{fd} is more efficient than the usual unbiased estimator \bar{y}^* , $\bar{y}_{hd(1)}$ and \bar{y}_{hd} in all three populations for different values of k . The PRE's increases as the value of k decreases.

Table 8.2, exhibits that in case when cost is fixed in advance, the proposed estimator \bar{y}_{fd} is more efficient than \bar{y}^* , $\bar{y}_{hd(1)}$ and \bar{y}_{hd} . When variance is fixed in advance, the proposed estimator \bar{y}_{fd} has less cost as compared to \bar{y}^* , t_R and \bar{y}_h .

References

- Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed., John Wiley & Sons, New York.
- Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys, *Journal of the American Statistical Association*, **41**, 517–529.
- Khare, B. B. and Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response, *National Academic Science and Letters*, **16**, 111–114.
- Khare, B. B. and Srivastava, S. (1995). Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response, In *Proceedings of the Indian National Science Academy*, **65(A)**, 195–203.
- Rao, P. S. R. S. (1986). Ratio estimation with sub sampling the non-respondents, *Survey Methodology*, **12**, 217–230.
- Srivastava, S. K. (1967). An estimator using auxiliary information in sample surveys, *Bulletin of the Calcutta Statistical Association*, **16**, 121–132.
- Srivastava, S. K. (1971). A generalized estimator for the mean of a finite population using multi auxiliary information, *Journal of the American Statistical Association*, **66**, 404–407.
- Srivastava, S. K. and Jhaji, H. S. (1981). A class of estimators of the population mean in survey sampling using auxiliary information, *Biometrika*, **68**, 341–343.
- Srivastava, S. K. and Jhaji, H. S. (1987). Improved estimation in two-phase and successive sampling, *Journal of Indian Society of Agricultural Statistics*, **25**, 71–75.
- Tabasum, R. and Khan, I. A. (2006). Double sampling ratio estimator for the population mean in presence of non-response, *Assam Statistical Review*, **20**, 73–83.