Communications of the Korean Statistical Society
2011, Vol. 18, No. 4, 537–548
DOI: 10.5351/CKSS.2011.18.4.537

Estimation of the Population Mean in Presence of Non-Response

Sunil Kumar[1,a], Sandeep Bhougal[b]

[a]Department of Statistics, University of Jammu
[b]School of Mathematics, Shri Mata Vaishno Devi University

Abstract

In this paper following Singh *et al.* (2008), we propose a modified ratio-product type exponential estimator to estimate the finite population mean $\bar{Y}$ of the study variable y in presence of non-response in different situations viz. (i) population mean $\bar{X}$ is known, and (ii) population mean $\bar{X}$ is unknown. The expressions of biases and mean squared error of the proposed estimators have been obtained under large sample approximation using single as well as double sampling. Some realistic conditions have been obtained under which the proposed estimator is more efficient than usual unbiased estimators, ratio estimators, product estimators and exponential ratio and product estimators reported by Rao (1986) and Singh *et al.* (2010) are found to be more efficient in many situations.

Keywords: Study variable, auxiliary variable, bias, mean squared error, exponential estimator, non-response.

1. Introduction

The problem of non-response in sample surveys is common and is more prevalent in mail surveys than in personal interview surveys. Hansen and Hurwitz (1946) have given a sampling plan that calls for taking a sub sample of non-respondents after the first mail attempt and then enumerating the sub sample by personal interview (see Srinath, 1971). In estimating population parameters like the mean, total or ratio, sample survey experts sometimes use auxiliary information to improve precision of the estimators. Further, various authors like Cochran (1977), Rao (1986, 1987), Khare and Srivastava (1993, 1995, 1997), Okafor and Lee (2000), Tabasum and Khan (2004, 2006), Singh and Kumar (2008, 2009a, 2009b, 2010) and Singh *et al.* (2010) studied the problem of non-response under double (two-stage) sampling.

Consider a finite population of size N and a random sample of size n drawn without replacement. 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 scheme to estimate the population mean:

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.

[1] Corresponding author: Assistant Professor, Department of Statistics, University of Jammu, J & K, India.
E-mail: sunilbhougal06@gmail.com

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 $\bar{Y}$ of the study variable y is given by

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

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

$$\mathrm{Var}\,(\bar{y}^*) = \bar{Y}^2\left\{\left(\frac{1-f}{n}\right)C_y^2 + \frac{W_2(k-1)}{n}C_{y(2)}^2\right\}, \tag{1.2}$$

where $C_y^2 = (S_y^2/\bar{Y}^2)$, $C_{y(2)}^2 = (S_{y(2)}^2/\bar{Y}^2)$.

Let x_i $(i = 1, 2, \ldots, N)$ denote an auxiliary variable correlated with the study variable y_i $(i = 1, 2, \ldots, N)$. The population mean of the auxiliary variable x is $\bar{X} = \sum_{i=1}^{N} x_i/N$. Let $\bar{X}_1 = \sum_{i=1}^{N_1} x_i/N_1$ and $\bar{X}_2 = \sum_{i=1}^{N_2} x_i/N_2$ denote the population means of the response and non-response groups (or strata). Let $\bar{x} = \sum_{i=1}^{n} x_i/n$ denote the mean of all the n units. Let $\bar{x}_1 = \sum_{i=1}^{n_1} x_i/n_1$ and $\bar{x}_2 = \sum_{i=1}^{n_2} x_i/n_2$ denote the means of the n_1 responding units and n_2 non-responding units. Further, let $\bar{x}_{2r} = \sum_{i=1}^{r} x_i/r$ denote the mean of the $r(= n_2/k)$, $k > 1$ sub-sampled units. With this background we define an unbiased estimator of the population mean $\bar{X}$ as

$$\bar{x}^* = w_1\bar{x}_1 + w_2\bar{x}_{2r}. \tag{1.3}$$

The variance of $\bar{x}^*$ is given by

$$\mathrm{Var}\,(\bar{x}^*) = \bar{X}^2\left\{\left(\frac{1-f}{n}\right)C_x^2 + \frac{W_2(k-1)}{n}C_{x(2)}^2\right\}, \tag{1.4}$$

where $C_x^2 = (S_x^2/\bar{X}^2)$, $C_{x(2)}^2 = (S_{x(2)}^2/\bar{X}^2)$, $S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2/(N-1)$ and $S_{x(2)}^2 = \sum_{i=1}^{N_2}(x_i - \bar{X}_2)^2/(N_2 - 1)$.

In some situations, there may not be any non-response on the auxiliary variables. Family size, years of education, and years of employment are the above type of auxiliary variables, see Rao (1986, p.220). 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

$$t_1^* = \bar{y}^*\left(\frac{\bar{X}}{\bar{x}^*}\right). \tag{1.5}$$

Khare and Srivastava (1993) suggested a product estimator for the population mean $\bar{Y}$ of the study variable y as

$$t_2^* = \bar{y}^*\left(\frac{\bar{x}^*}{\bar{X}}\right). \tag{1.6}$$

An exponential ratio and product type estimators for the population mean $\bar{Y}$ of the study variable y are

$$t_3^* = \bar{y}^* \exp\left(\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right) \tag{1.7}$$

and

$$t_4^* = \bar{y}^* \exp\left(\frac{\bar{x}^* - \bar{X}}{\bar{x}^* + \bar{X}}\right). \tag{1.8}$$

The objective of this paper is to suggest a ratio-product type exponential estimator for estimating the finite population mean in the presence of non-response in different situations viz. (i) population mean $\bar{X}$ is known, and (ii) population mean $\bar{X}$ is unknown. The expressions of biases and mean squared errors of the proposed estimators, up to the first order of approximation, have been obtained. The results obtained are depicted with the help of numerical illustration.

2. Proposed Estimators

In general, the linear regression estimator is more efficient than the ratio (product) estimator except when the regression line of y on x passes through the neighborhood of the origin, in which case the efficiencies of these estimators are almost equal. In addition, in many practical situations the regression line does not pass through the neighborhood of the origin. In these situations, the ratio estimator does not perform as good as the linear regression estimator.

Singh *et al.* (2008) proposed a ratio-product type exponential estimator by following Singh and Ruize Espejo (2003) for estimating the finite population mean. Following Singh *et al.* (2008), we propose following class of ratio-product estimators for estimating population mean $\bar{Y}$ of the study variable y in presence of non-response, as

$$t_5^* = \bar{y}^* \left\{\alpha \exp\left(\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right) + (1-\alpha) \exp\left(\frac{\bar{x}^* - \bar{X}}{\bar{x}^* + \bar{X}}\right)\right\}, \tag{2.1}$$

where α is a real constant to be determined such that the MSE of t_5^* is minimum.

For $\alpha = 0, 1$ the class of estimators respectively reduce to the estimator t_4^* and t_3^* respectively. To obtain the bias and variance of t_5^*, we write

$$\bar{y}^* = \bar{Y}(1+\epsilon_0); \qquad \bar{x}^* = \bar{X}(1+\epsilon_1),$$

such that

$$E(\epsilon_0) = E(\epsilon_1) = 0$$

and

$$E\left(\epsilon_0^2\right) = \mathrm{Var}(\bar{y}^*) = \bar{Y}^2 \left\{\left(\frac{1-f}{n}\right) C_y^2 + \frac{W_2(k-1)}{n} C_{y(2)}^2\right\},$$

$$E\left(\epsilon_1^2\right) = \mathrm{Var}(\bar{x}^*) = \bar{X}^2 \left\{\left(\frac{1-f}{n}\right) C_x^2 + \frac{W_2(k-1)}{n} C_{x(2)}^2\right\},$$

$$E\left(\epsilon_0\epsilon_1\right) = \mathrm{Cov}(\bar{y}^*, \bar{x}^*) = \bar{Y}\bar{X} \left\{\left(\frac{1-f}{n}\right) \rho_{yx} C_y C_x + \frac{W_2(k-1)}{n} \rho_{yx(2)} C_{y(2)} C_{x(2)}\right\},$$

where $\rho_{yx} = S_{yx}/(S_x S_y)$; $\rho_{yx(2)} = S_{yx(2)}/(S_{x(2)} S_{y(2)})$; $S_{yx} = 1/(N-1)\sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})$; and $S_{yx(2)} = 1/(N_2 - 1)\sum_{i=1}^{N_2}(y_i - \bar{Y}_2)(x_i - \bar{X}_2)$.

Now, expressing t_5^* in terms of ϵ's we have

$$
\begin{aligned}
t_5^* &= \bar{Y}(1+\epsilon_0)\left[\alpha \exp\left\{\frac{\bar{X} - \bar{X}(1+\epsilon_1)}{\bar{X} + \bar{X}(1+\epsilon_1)}\right\} + (1-\alpha)\exp\left\{\frac{\bar{X}(1+e_1) - \bar{X}}{\bar{X}(1+e_1) + \bar{X}}\right\}\right] \\
&= \bar{Y}(1+\epsilon_0)\left[\alpha \exp\left(\frac{-\epsilon_1}{2+\epsilon_1}\right) + (1-\alpha)\exp\left(\frac{\epsilon_1}{2+\epsilon_1}\right)\right] \\
&= \bar{Y}(1+\epsilon_0)\left[\alpha \exp\left\{\frac{-\epsilon_1}{2}\left(1+\frac{\epsilon_1}{2}\right)^{-1}\right\} + (1-\alpha)\exp\left\{\frac{\epsilon_1}{2}\left(1+\frac{\epsilon_1}{2}\right)^{-1}\right\}\right].
\end{aligned} \tag{2.2}
$$

Expanding the right hand side of (2.2) and neglecting the terms involving powers of ϵ's greater than two, we have

$$
\begin{aligned}
t_5^* &= \bar{Y}\left[1 + \epsilon_0 + \frac{\epsilon_1}{2} - \alpha\epsilon_1 + \frac{\epsilon_1^2}{4} + \frac{\epsilon_0\epsilon_1}{2} - \alpha\epsilon_0\epsilon_1\right] \\
\left(t_5^* - \bar{Y}\right) &= \bar{Y}\left[\epsilon_0 + \frac{\epsilon_1}{2} - \alpha\epsilon_1 + \frac{\epsilon_1^2}{4} + \frac{\epsilon_0\epsilon_1}{2} - \alpha\epsilon_0\epsilon_1\right].
\end{aligned} \tag{2.3}
$$

Taking expectations of both sides of (2.3), we get the bias of the estimator t_5^* as

$$
B\left(t_5^*\right) = \frac{\bar{Y}}{4}\left[\left(\frac{1-f}{n}\right)\left\{1 + 2(1-2\alpha)k_{yx}\right\}C_x^2 + \frac{W_2(k-1)}{n}\left\{1 + 2(1-2\alpha)k_{yx(2)}\right\}C_{x(2)}^2\right], \tag{2.4}
$$

where $k_{yx} = \rho_{yx}C_y/C_x$; $k_{yx(2)} = \rho_{yx(2)}C_{y(2)}/C_{x(2)}$.

Squaring both sides of (2.3) and neglecting terms of ϵ's involving power greater than two, we have

$$
\begin{aligned}
\left(t_5^* - \bar{Y}\right)^2 &= \bar{Y}^2\left\{\epsilon_0 + \epsilon_1\left(\frac{1}{2} - \alpha\right)\right\}^2 \\
&= \bar{Y}^2\left\{\epsilon_0^2 + \epsilon_1^2\left(\frac{1}{4} + \alpha^2 - \alpha\right) + 2\epsilon_0\epsilon_1\left(\frac{1}{2} - \alpha\right)\right\} \\
\left(t_5^* - \bar{Y}\right)^2 &= \bar{Y}^2\left\{\epsilon_0^2 + \epsilon_1^2\left(\frac{1}{4} + \alpha^2 - \alpha\right) + \epsilon_0\epsilon_1(1-2\alpha)\right\}.
\end{aligned} \tag{2.5}
$$

Taking expectations of both sides of (2.5), we get the exact mean squared error(MSE) of t_5^* and approximation (to the first degree of approximation) MSE of t_5^*, as

$$
\begin{aligned}
\mathrm{MSE}\left(t_5^*\right) = \bar{Y}^2\Bigg[&\left(\frac{1-f}{n}\right)\left\{C_y^2 + C_x^2\left(\frac{1}{4} + \alpha^2 - \alpha\right) + \rho_{yx}C_yC_x(1-2\alpha)\right\} \\
&+ \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2 + C_{x(2)}^2\left(\frac{1}{4} + \alpha^2 - \alpha\right) + \rho_{yx(2)}C_{y(2)}C_{x(2)}(1-2\alpha)\right\}\Bigg].
\end{aligned} \tag{2.6}
$$

Minimization of (2.6) with respect to 'α' yields its optimum value as

$$
\alpha = \frac{A + 2B}{2A} = \alpha_0(\text{say}), \tag{2.7}
$$

where $A = \{(1-f)/n\}C_x^2 + \{W_2(k-1)\}/n\,C_{x(2)}^2$, $B = \{(1-f)/n\}k_{yx}C_x^2 + \{W_2(k-1)\}/n\,k_{yx(2)}C_{x(2)}^2$.

Substitute the optimum value of 'α' from (2.7) in (2.1) yields the optimum estimator as

$$t^*_{5(opt)} = \bar{y}^* \left\{ \alpha_0 \exp\left(\frac{\bar{X}-\bar{x}^*}{\bar{X}+\bar{x}^*}\right) + (1-\alpha_0)\exp\left(\frac{\bar{x}^*-\bar{X}}{\bar{x}^*+\bar{X}}\right) \right\}. \tag{2.8}$$

The exact MSE of the optimum estimator $t^*_{5(opt)}$ is given by

$$\begin{aligned} \mathrm{MSE}\left(t^*_{5(opt)}\right) = \min \mathrm{MSE}\left(t^*_5\right) = \bar{Y}^2 \Bigg[&\left(\frac{1-f}{n}\right)\left\{C_y^2 + \frac{B}{A}\left(\frac{B}{A} - 2k_{yx}\right)C_x^2\right\} \\ &+ \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2 + \frac{B}{A}\left(\frac{B}{A} - 2k_{yx(2)}\right)C_{x(2)}^2\right\}\Bigg]. \end{aligned} \tag{2.9}$$

3. Efficiency Comparisons

In this section, the conditions for which the proposed estimator t^*_5 is better than the usual unbiased estimator $\bar{y}^*$, t^*_1, t^*_2, t^*_3 and t^*_4 have been obtained. The MSE's of these estimators to the first degree of approximation are derived as

$$\mathrm{Var}\,(\bar{y}^*) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)C_y^2 + \frac{W_2(k-1)}{n}C_{y(2)}^2\right], \tag{3.1}$$

$$\mathrm{MSE}\,(t^*_1) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)\left\{C_y^2 + \left(1-2k_{yx}\right)C_x^2\right\} + \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2 + \left(1-2k_{yx(2)}\right)C_{x(2)}^2\right\}\right], \tag{3.2}$$

$$\mathrm{MSE}\,(t^*_2) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)\left\{C_y^2 + \left(1+2k_{yx}\right)C_x^2\right\} + \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2 + \left(1+2k_{yx(2)}\right)C_{x(2)}^2\right\}\right], \tag{3.3}$$

$$\mathrm{MSE}(t^*_3) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)\left\{C_y^2 + \frac{C_x^2}{4}\left(1-4k_{yx}\right)\right\} + \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2 + \frac{C_{x(2)}^2}{4}\left(1-4k_{yx(2)}\right)\right\}\right], \tag{3.4}$$

$$\mathrm{MSE}\,(t^*_4) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)\left\{C_y^2 + \frac{C_x^2}{4}\left(1+4k_{yx}\right)\right\} + \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2 + \frac{C_{x(2)}^2}{4}\left(1+4k_{yx(2)}\right)\right\}\right]. \tag{3.5}$$

To compare the efficiency of the proposed estimator t^*_5 with the existing estimators, from (2.9) and (3.1)–(3.5), we have

$$\mathrm{Var}\,(\bar{y}^*) - \mathrm{MSE}\left(t^*_{5(opt)}\right) = \left(\frac{1-f}{n}\right)k_{yx}C_x^2 + \frac{W_2(k-1)}{n}k_{yx(2)}C_{x(2)}^2 \geq 0, \tag{3.6}$$

$$\mathrm{MSE}\,(t^*_1) - \mathrm{MSE}\left(t^*_{5(opt)}\right) = \left(\frac{1-f}{n}\right)\left(1-k_{yx}\right)C_x^2 + \frac{W_2(k-1)}{n}\left(1-k_{yx(2)}\right)C_{x(2)}^2 \geq 0, \tag{3.7}$$

$$\mathrm{MSE}\,(t^*_2) - \mathrm{MSE}\left(t^*_{5(opt)}\right) = \left(\frac{1-f}{n}\right)\left(1+k_{yx}\right)C_x^2 + \frac{W_2(k-1)}{n}\left(1+k_{yx(2)}\right)C_{x(2)}^2 \geq 0, \tag{3.8}$$

$$\mathrm{MSE}(t^*_3) - \mathrm{MSE}\left(t^*_{5(opt)}\right) = \left(\frac{1-f}{n}\right)\left(1-2k_{yx}\right)C_x^2 + \frac{W_2(k-1)}{n}\left(1-2k_{yx(2)}\right)C_{x(2)}^2 \geq 0, \tag{3.9}$$

$$\mathrm{MSE}\,(t^*_4) - \mathrm{MSE}\left(t^*_{5(opt)}\right) = \left(\frac{1-f}{n}\right)\left(1+2k_{yx}\right)C_x^2 + \frac{W_2(k-1)}{n}\left(1+2k_{yx(2)}\right)C_{x(2)}^2 \geq 0. \tag{3.10}$$

From (3.6)–(3.10), we conclude that the proposed estimator t_5^* outperforms the usual unbiased estimator $\bar{y}^*$, t_1^*, t_2^*, t_3^* and t_4^*. If α does not coincide with α_0, *i.e.* $\alpha = \alpha_0$, then from (3.1), (3.2), (3.3), (3.4), (3.5) and (2.6), we envisaged that the suggested estimator t_5^* is better than

i) the usual unbiased estimator $\bar{y}^*$ if

$$\begin{cases} \text{either} & \frac{1}{2} < \alpha < \frac{1}{2}\left(1 + 4k_{yx}\right) \quad \text{and} \quad \frac{1}{2} < \alpha < \frac{1}{2}\left(1 + 4k_{yx(2)}\right), \\ \text{or} & \frac{1}{2}\left(1 + 4k_{yx}\right) < \alpha < \frac{1}{2} \quad \text{and} \quad \frac{1}{2}\left(1 + 4k_{yx(2)}\right) < \alpha < \frac{1}{2}, \end{cases} \tag{3.11}$$

ii) the ratio estimator t_1^* if

$$\begin{cases} \text{either} & \frac{3}{2} < \alpha < \frac{1}{2}\left(4k_{yx} - 1\right) \quad \text{and} \quad \frac{3}{2} < \alpha < \frac{1}{2}\left(4k_{yx(2)} - 1\right), \\ \text{or} & \frac{1}{2}\left(4k_{yx} - 1\right) < \alpha < \frac{3}{2} \quad \text{and} \quad \frac{1}{2}\left(4k_{yx(2)} - 1\right) < \alpha < \frac{3}{2}, \end{cases} \tag{3.12}$$

iii) the product estimator t_2^* if

$$\begin{cases} \text{either} & \frac{1}{2} < \alpha < \frac{1}{2}\left(4k_{yx} + 3\right) \quad \text{and} \quad \frac{1}{2} < \alpha < \frac{1}{2}\left(4k_{yx(2)} + 3\right), \\ \text{or} & \frac{1}{2}\left(4k_{yx} + 3\right) < \alpha < \frac{1}{2} \quad \text{and} \quad \frac{1}{2}\left(4k_{yx(2)} + 3\right) < \alpha < \frac{1}{2}, \end{cases} \tag{3.13}$$

iv) the exponential ratio type estimator t_3^* if

$$\begin{cases} \text{either} & 1 < \alpha < k_{yx} \quad \text{and} \quad 1 < \alpha < k_{yx(2)}, \\ \text{or} & k_{yx} < \alpha < 1 \quad \text{and} \quad k_{yx(2)} < \alpha < 1, \end{cases} \tag{3.14}$$

v) the exponential product type estimator t_4^* if

$$\begin{cases} \text{either} & 0 < \alpha < 1 + 2k_{yx} \quad \text{and} \quad 0 < \alpha < 1 + 2k_{yx(2)}, \\ \text{or} & 1 + 2k_{yx} < \alpha < 0 \quad \text{and} \quad 1 + 2k_{yx(2)} < \alpha < 0. \end{cases} \tag{3.15}$$

The proposed class of ratio product estimator t_5^* is more efficient than $\bar{y}^*$, t_1^*, t_2^*, t_3^* and t_4^* respectively, if (3.10), (3.11), (3.12), (3.13) and (3.14) respectively hold true.

4. Empirical Study

To see the performance of the suggested estimators of the population mean, we consider a natural dataset considered by Khare and Srivastava (1993). The description of the population is given below:

A list of 70 villages in India along their population in 1981 and cultivated areas (in acres) in the same year is considered (Singh and Choudhary, 1986). Here the cultivated area (in acres) is taken as the main study variable and the population of the village is taken as the auxiliary variable. The parameters of the population are as follows:

$$\bar{Y} = 981.29, \quad \bar{X} = 1755.53, \quad C_y = 0.6254, \quad C_x = 0.8009, \quad \bar{Y}_2 = 597.29, \quad \bar{X}_2 = 1100.24,$$
$$C_{y(2)} = 0.4087, \quad C_{x(2)} = 0.5739, \quad \rho_{yx} = 0.778, \quad \rho_{yx(2)} = 0.445, \quad R = 0.558971,$$
$$k_{yx} = 0.6075, \quad k_{yx(2)} = 0.3169, \quad W_2 = 0.20, \quad N = 70, \quad n = 35.$$

Table 1: Percent-relative efficiency(PRE) of different estimators

PRE $(\cdot, \bar{y}^*)$	$(1/k)$			
	(1/5)	(1/4)	(1/3)	(1/2)
PRE $\left(t_1^*, \bar{y}^*\right)$	92.52	99.18	109.01	125.00
PRE $\left(t_2^*, \bar{y}^*\right)$	22.43	22.29	22.12	21.91
PRE $\left(t_3^*, \bar{y}^*\right)$	167.29	176.31	189.09	208.59
PRE $\left(t_4^*, \bar{y}^*\right)$	43.67	43.31	42.88	42.32
PRE $\left(t_{5(opt)}^*, \bar{y}^*\right)$	167.58	176.35	189.38	211.05

We have computed the percent-relative efficiencies(PRE's) of various suggested estimators with respect to the usual unbiased estimator $\bar{y}^*$ for various values of k, by using the formulae

$$\text{PRE}\,(t_i^*, \bar{y}^*) = \frac{\text{Var}\,(\bar{y}^*)}{\text{MSE}\,(*)} \times 100; \quad i = 1, 2, 3, 4 \text{ and } 5(\text{opt}).$$

It is to be envisaged from Table 1 that the PRE's of the ratio type estimators t_1^*, t_3^* and t_5^* increase while the PRE's of the product type estimators t_2^* and t_4^* decrease as the value of k increases. Further, it has been observed that the estimator t_5^* is the best among $\bar{y}^*$, t_1^*, t_2^*, t_3^* and t_4^*. Thus, the suggested estimator t_5^* is to be recommended for its use in practice.

5. Double (Two-Stage) Sampling

In many of the large scale sample surveys a multi-stage sampling design is generally used for selection of a sample and data are collected for several items. However, it is noted that information in most cases are not obtained at the first attempt even after some call-backs. For the estimate of population mean $\bar{X}$ of the auxiliary variable x, a large first phase sample of size n' is selected from a population of size N units by simple random sampling without replacement(SRSWOR). A smaller second phase sample of size 'n' is selected from n' by SRSWOR and the variable 'y' under study is measured on it; however, take a sub-sample of the non-respondents and re-conduct them if there is non-response in the second phase sample.

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 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

$$m = \frac{n_2}{k}, \quad k > 1 \text{ (see, Tabasum and Khan, 2004).}$$

When the population mean '$\bar{X}$' of the auxiliary variable 'x' is unknown, the two phase ratio and product type estimator are

$$t_6^* = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right), \quad \text{(by Khare and Srivastava, 1995; Okafor and Lee, 2000; Tabasum and Khan, 2004),} \tag{5.1}$$

$$t_7^* = \bar{y}^* \left(\frac{\bar{x}^*}{\bar{x}'}\right), \quad \text{(by Khare and Srivastava, 1995; Okafor and Lee, 2000; Tabasum and Khan, 2004),} \tag{5.2}$$

where $\bar{x}'$ denote the sample mean of x based on first phase sample of size n'.

The two phase ratio and product type exponential estimator

$$t_8^* = \bar{y}^* \exp\left[\frac{\bar{x}' - \bar{x}^*}{\bar{x}' + \bar{x}^*}\right], \quad \text{(by Singh } et\ al.\text{, 2010)} \tag{5.3}$$

and

$$t_9^* = \bar{y}^* \exp\left[\frac{\bar{x}^* - \bar{x}'}{\bar{x}^* + \bar{x}'}\right], \quad \text{(by Singh } et\ al., \text{ 2010).} \tag{5.4}$$

The MSE of the estimators t_6^*, t_7^*, t_8^* and t_9^* is given by

$$\text{MSE}\left(t_6^*\right) = \bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{C_y^2+\left(1-2k_{yx}\right)C_x^2\right\}+\frac{W_2(k-1)}{n}\left\{C_{y(2)}^2+\left(1-2k_{yx(2)}\right)C_{x(2)}^2\right\}+\left(\frac{1}{n'}-\frac{1}{N}\right)C_y^2\right], \tag{5.5}$$

$$\text{MSE}\left(t_7^*\right) = \bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{C_y^2+\left(1+2k_{yx}\right)C_x^2\right\}+\frac{W_2(k-1)}{n}\left\{C_{y(2)}^2+\left(1+2k_{yx(2)}\right)C_{x(2)}^2\right\}+\left(\frac{1}{n'}-\frac{1}{N}\right)C_y^2\right], \tag{5.6}$$

$$\text{MSE}\left(t_8^*\right) = \bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{C_y^2+\left(1-4k_{yx}\right)\frac{C_x^2}{4}\right\}+\frac{W_2(k-1)}{n}\left\{C_{y(2)}^2+\left(1-4k_{yx(2)}\right)\frac{C_{x(2)}^2}{4}\right\}+\left(\frac{1}{n'}-\frac{1}{N}\right)C_y^2\right], \tag{5.7}$$

$$\text{MSE}\left(t_9^*\right) = \bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{C_y^2+\left(1+4k_{yx}\right)\frac{C_x^2}{4}\right\}+\frac{W_2(k-1)}{n}\left\{C_{y(2)}^2+\left(1+4k_{yx(2)}\right)\frac{C_{x(2)}^2}{4}\right\}+\left(\frac{1}{n'}-\frac{1}{N}\right)C_y^2\right]. \tag{5.8}$$

6. Suggested Estimator

We define a class of ratio-product estimator for estimating the population mean $\bar{Y}$ of the study variable y in presence of non-response, as

$$t_{10}^* = \bar{y}^*\left[\alpha_1 \exp\left(\frac{\bar{x}' - \bar{x}^*}{\bar{x}' + \bar{x}^*}\right) + (1-\alpha_1)\exp\left(\frac{\bar{x}^* - \bar{x}'}{\bar{x}^* + \bar{x}'}\right)\right], \tag{6.1}$$

where α_1 is a real constant to be determined such that MSE of t_{10}^* is minimum.

For $\alpha_1 = 1$, t_{10}^* reduces to estimator t_8^*, and for $\alpha_1 = 0$, t_{10}^* reduces to t_9^*, respectively. Let $\bar{x}' = \bar{X}(1+\epsilon_2)$ such that $E(\epsilon_2) = 0$ and

$$E\left(\epsilon_2^2\right) = \left(\frac{1}{n'}-\frac{1}{N}\right)C_x^2; \quad E(\epsilon_1\epsilon_2) = \left(\frac{1}{n'}-\frac{1}{N}\right)C_x^2; \quad E(\epsilon_0\epsilon_2) = \left(\frac{1}{n'}-\frac{1}{N}\right)\rho_{yx}C_yC_x.$$

Expressing (6.1) in terms of ϵ's, we get

$$\left(t_{10}^* - \bar{Y}\right) = \bar{Y}\left\{\epsilon_0 + \left(\frac{\epsilon_1-\epsilon_2}{2}\right) + \alpha_1(\epsilon_2-\epsilon_1) + \alpha_1\epsilon_0\epsilon_2 - \alpha_1\epsilon_0\epsilon_1 + \frac{(\epsilon_1-\epsilon_2)^2}{4} + \frac{\epsilon_0\epsilon_1}{2} - \frac{\epsilon_0\epsilon_2}{2}\right\}. \tag{6.2}$$

Taking expectations of both sides of (6.2), we get the bias of the proposed estimator t_{10}^*, up to the first degree of approximation, as

$$B\left(t_{10}^*\right) = \bar{Y}\left[\frac{1}{4}\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2\left\{1-4\left(\alpha_1-\frac{1}{2}\right)k_{yx}\right\} + \frac{W_2(k-1)}{n}\frac{C_{x(2)}^2}{4}\left\{1-4\left(\alpha_1-\frac{1}{2}\right)k_{yx(2)}\right\}\right]. \tag{6.3}$$

From (6.2), we have

$$\left(t_{10}^* - \bar{Y}\right) \cong \bar{Y}\left\{\epsilon_0 + \left(\frac{\epsilon_1-\epsilon_2}{2}\right) + \alpha_1(\epsilon_2-\epsilon_1)\right\}^2. \tag{6.4}$$

Squaring both sides of (6.4) and then taking expectations, we get the MSE of the proposed estimator t^*_{10}, up to the first order of approximation as

$$\mathrm{MSE}\left(t^*_{10}\right) = \bar{Y}^2\Bigg[\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{C_y^2+\left(\alpha_1-\frac{1}{2}\right)\left(\left(\alpha_1-\frac{1}{2}\right)-2k_{yx}\right)C_x^2\right\}+\left(\frac{1}{n'}-\frac{1}{N}\right)C_y^2 + \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2+\left(\alpha_1-\frac{1}{2}\right)\left(\left(\alpha_1-\frac{1}{2}\right)-2k_{yx(2)}\right)C_{x(2)}^2\right\}\Bigg], \tag{6.5}$$

which is minimum when

$$\alpha_1 = \frac{A_{(2)}+2B_{(2)}}{2A_{(2)}} = \alpha_{10}(\text{say}),$$

where

$$A_{(2)} = \left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2+\frac{W_2(k-1)}{n}C_{x(2)}^2; \quad B_{(2)} = \left(\frac{1}{n}-\frac{1}{n'}\right)k_{yx}C_x^2+\frac{W_2(k-1)}{n}k_{yx(2)}C_{x(2)}^2.$$

Thus, the resulting minimum MSE of t^*_{10} is given by

$$\mathrm{MSE}\left(t^*_{10(opt)}\right) = \bar{Y}^2\Bigg[\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{C_y^2+\frac{B_{(2)}}{A_{(2)}}\left(\frac{B_{(2)}}{A_{(2)}}-2k_{yx}\right)C_x^2\right\} + \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2+\frac{B_{(2)}}{A_{(2)}}\left(\frac{B_{(2)}}{A_{(2)}}-2k_{yx(2)}\right)C_{x(2)}^2\right\}+\left(\frac{1}{n'}-\frac{1}{N}\right)C_y^2\Bigg]. \tag{6.6}$$

7. Efficiency Comparisons

From (3.1), (5.5), (5.6), (5.7), (5.8) and (6.6), we have

$$\mathrm{Var}\left(\bar{y}^*\right)-\mathrm{MSE}\left(t^*_{10(opt)}\right) = \left(\frac{1}{n}-\frac{1}{n'}\right)k_{yx}C_x^2+\frac{W_2(k-1)}{n}k_{yx(2)}^2C_{x(2)}^2 \geq 0, \tag{7.1}$$

$$\mathrm{MSE}\left(t^*_6\right)-\mathrm{MSE}\left(t^*_{10(opt)}\right) = \left(\frac{1}{n}-\frac{1}{n'}\right)\left(1-k_{yx}\right)C_x^2+\frac{W_2(k-1)}{n}\left(1-k_{yx(2)}\right)C_{x(2)}^2 \geq 0, \tag{7.2}$$

$$\mathrm{MSE}\left(t^*_7\right)-\mathrm{MSE}\left(t^*_{10(opt)}\right) = \left(\frac{1}{n}-\frac{1}{n'}\right)\left(1+k_{yx}\right)C_x^2+\frac{W_2(k-1)}{n}\left(1+k_{yx(2)}\right)C_{x(2)}^2 \geq 0, \tag{7.3}$$

$$\mathrm{MSE}\left(t^*_8\right)-\mathrm{MSE}\left(t^*_{10(opt)}\right) = \left(\frac{1}{n}-\frac{1}{n'}\right)\left(1-2k_{yx}\right)C_x^2+\frac{W_2(k-1)}{n}\left(1-2k_{yx(2)}\right)C_{x(2)}^2 \geq 0, \tag{7.4}$$

$$\mathrm{MSE}\left(t^*_9\right)-\mathrm{MSE}\left(t^*_{10(opt)}\right) = \left(\frac{1}{n}-\frac{1}{n'}\right)\left(1+2k_{yx}\right)C_x^2+\frac{W_2(k-1)}{n}\left(1+2k_{yx(2)}\right)C_{x(2)}^2 \geq 0. \tag{7.5}$$

From (7.1)–(7.5), it is envisaged that the proposed estimator t^*_{10} is more efficient than the estimators t^*_6, t^*_7, t^*_8 and t^*_9 respectively. If α_1 does not coincide with α_{10}, *i.e.* $\alpha_1 \neq \alpha_{10}$, then from (7.1), (7.2), (7.3), (7.4), (7.5) and (6.5), we envisaged that the suggested estimator t^*_{10} is better than

i) the usual unbiased estimator $\bar{y}^*$ if

$$\begin{cases} \text{either} & \frac{1}{2} < \alpha_1 < \frac{1}{2}\left(1+4k_{yx}\right) \quad \text{and} \quad \frac{1}{2} < \alpha_1 < \frac{1}{2}\left(1+4k_{yx(2)}\right), \\ \text{or} & \frac{1}{2}\left(1+4k_{yx}\right) < \alpha_1 < \frac{1}{2} \quad \text{and} \quad \frac{1}{2}\left(1+4k_{yx(2)}\right) < \alpha_1 < \frac{1}{2}, \end{cases} \tag{7.6}$$

ii) the ratio estimator t_6^* if

$$\begin{cases} \text{either} \quad 0 < \alpha_1 < \left(1 + k_{yx}\right) \quad \text{and} \quad 0 < \alpha_1 < \left(1 + k_{yx(2)}\right), \\ \text{or} \quad \left(1 + k_{yx}\right) < \alpha_1 < 0 \quad \text{and} \quad \left(1 + k_{yx(2)}\right) < \alpha_1 < 0, \end{cases} \tag{7.7}$$

iii) the product estimator t_7^* if

$$\begin{cases} \text{either} \quad -\frac{1}{2} < \alpha_1 < \frac{1}{2}\left(4k_{yx} + 3\right) \quad \text{and} \quad -\frac{1}{2} < \alpha_1 < \frac{1}{2}\left(4k_{yx(2)} + 3\right), \\ \text{or} \quad \frac{1}{2}\left(4k_{yx} + 3\right) < \alpha_1 < -\frac{1}{2} \quad \text{and} \quad \frac{1}{2}\left(4k_{yx(2)} + 3\right) < \alpha_1 < -\frac{1}{2}, \end{cases} \tag{7.8}$$

iv) the exponential ratio type estimator t_8^* if

$$\begin{cases} \text{either} \quad 1 < \alpha_1 < 2k_{yx} \quad \text{and} \quad 1 < \alpha_1 < 2k_{yx(2)}, \\ \text{or} \quad 2k_{yx} < \alpha_1 < 1 \quad \text{and} \quad 2k_{yx(2)} < \alpha_1 < 1, \end{cases} \tag{7.9}$$

v) the exponential product type estimator t_9^* if

$$\begin{cases} \text{either} \quad 0 < \alpha_1 < \left(1 + 2k_{yx}\right) \quad \text{and} \quad 0 < \alpha_1 < \left(1 + 2k_{yx(2)}\right), \\ \text{or} \quad \left(1 + 2k_{yx}\right) < \alpha_1 < 0 \quad \text{and} \quad \left(1 + 2k_{yx(2)}\right) < \alpha_1 < 0. \end{cases} \tag{7.10}$$

The proposed class of ratio product estimator t_{10}^* is more efficient than $\bar{y}^*$, t_6^*, t_7^*, t_8^* and t_9^* respectively, if (7.6), (7.7), (7.8), (7.9) and (7.10) respectively hold true.

8. Empirical Study

To look closely the excellence of the suggested estimators, we consider the following data set:
Source: Khare and Srivastava (1995, p.201).

The population of 100 consecutive trips (after leaving 20 outlier values) measured by two fuel meters for a small family car in normal usage given by Lewisi *et al.* (1991) has been taken into consideration. The measurements of turbine meter (in ml.) is considered as main variable y and the measurements of displacement meter (in cm^3) is considered as auxiliary variable x. We treat 25% last values as non-response units. The values of the parameters are as follows:

$\bar{Y} = 3500.12$, $\bar{X} = 260.84$, $C_y = 0.5941$, $C_x = 0.5996$, $\bar{Y}_2 = 3401.08$, $\bar{X}_2 = 259.96$,
$C_{y(2)} = 0.5075$, $C_{x(2)} = 0.5168$, $\rho_{yx} = 0.985$, $\rho_{yx(2)} = 0.995$, $W_2 = 0.25$,
$k_{yx} = 0.9759$, $k_{yx(2)} = 0.9771$, $R = 13.4187$, $N = 100$, $n = 30$, $n' = 50$.

Here, we have calculated the percent relative efficiencies(PRE's) of different suggested estimators with respect to usual unbiased estimator $\bar{y}^*$ for different values of k by using the following formulae

$$\text{PRE}\,(t_i^*, \bar{y}^*) = \frac{\text{Var}\,(\bar{y}^*)}{\text{MSE}\,(*)} \times 100; \quad i = 6, 7, 8, 9 \text{ and } 10(\text{opt}).$$

Table 2 exhibits that the PRE's of estimators t_6^*, t_9^* and $t_{10(opt)}^*$ decreases as the value of k increases, while the PRE's of estimators t_7^* and t_8^* decreases as the value of k increases. Further, it is to be noted that the estimator $t_{10(opt)}^*$ is the best among $\bar{y}^*$, t_6^*, t_7^*, t_8^* and t_9^*. Thus, the suggested estimator $t_{10(opt)}^*$ is to be recommended for its use in practice.

Table 2: Percent relative efficiency of the different estimators of $\bar{Y}$ with respect to $\bar{y}^*$.

$\text{PRE}(\cdot, \bar{y}^*)$	$(1/k)$			
	(1/5)	(1/4)	(1/3)	(1/2)
$\text{PRE}\left(t_6^*, \bar{y}^*\right)$	439.59	385.97	331.72	276.81
$\text{PRE}\left(t_7^*, \bar{y}^*\right)$	29.82	30.66	31.87	33.75
$\text{PRE}\left(t_8^*, \bar{y}^*\right)$	146.25	149.49	154.08	161.05
$\text{PRE}\left(t_9^*, \bar{y}^*\right)$	66.44	65.24	63.69	61.64
$\text{PRE}\left(t_{10(opt)}^*, \bar{y}^*\right)$	440.53	386.68	332.22	277.15

9. Conclusion

The present article considers the problem for estimating the finite population mean $\bar{Y}$ of the study variable y in presence of non-response in different situations viz. (i) population mean $\bar{X}$ is known, and (ii) population mean $\bar{X}$ is unknown. Using the Hansen and Hurwitz (1946) procedure of sub-sampling the non-respondents for both the cases where the population mean of the auxiliary character is known and not known in advance, an ratio-product type estimator respectively have been proposed and their properties are studied. The optimum mean squared error's(MSE's) of the proposed estimators are also obtained. The relative performance of the proposed estimators is compared with the conventional estimators. The proposed estimators are efficient and should work very well in practical surveys.

Acknowledgement

Authors wish to thank the learned referees for their critical and constructive comments regarding improvement of the paper.

References

Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed., John Wiley and 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, *The National Academy of Sciences, Letters, India*, **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, *Proceedings of the National Academy of Sciences, India*, **65(A)**, 195–203.

Khare, B. B. and Srivastava, S. (1997). Transformed ratio type estimators for the population mean in the presence of non-response, *Communications in Statistics - Theory and Methods*, **26**, 1779–1791.

Lewisi, P. A., Jones, P. W., Polak, J. W. and Tillotson, H. T. (1991). The problem of conversion in method comparison studies, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **40**, 105–112.

Okafor, F. C. and Lee, H. (2000). Double sampling for ratio and regression estimation with sub-sampling the non-respondents, *Survey Methodology*, **26**, 183–188.

Rao, P. S. R. S. (1986). Ratio estimation with sub sampling the non-respondents, *Survey Methodology*, **12**, 217–230.

Rao, P. S. R. S. (1987). Ratio and regression estimates with sub sampling the non- respondents. Paper presented at a special contributed session of the International Statistical Association Meeting, Sept., 2–16, Tokyo, Japan.

Singh, D. and Choudhary, F. S. (1986). *Theory and Analysis of Sample Survey Designs*, Wiley Eastern Limited, New Delhi, p.108.

Singh, H. P. and Kumar, S. (2008). A regression approach to the estimation of finite population mean in presence of non-response, *Australian and New Zealand Journal of Statistics*, **50**, 395–408.

Singh, H. P. and Kumar, S. (2009a). A general class of estimators of the population mean in survey sampling using auxiliary information with sub sampling the non-respondents, *The Korean Journal of Applied Statistics*, **22**, 387–402.

Singh, H. P. and Kumar, S. (2009b). A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information, *SORT*, **33**, 71–84.

Singh, H. P. and Kumar, S. (2010). Estimation of mean in presence of non-response using two phase sampling scheme, *Statistical Papers*, **50**, 559–582.

Singh, H. P., Kumar, S. and Kozak, M. (2010). Improved estimation of finite population mean when sub-sampling is employed to deal with non-response, *Communication in Statistics - Theory and Methods*, **39**, 791–802.

Singh, H. P. and Ruiz Espejo, M. (2003). On linear regression and ratio-product estimation of a finite population mean, *Journal of the Royal Statistical Society. Series D (The Statistician)*, **52**, 59–67.

Singh, R., Chauhan, P. and Sawan, N. (2008). On linear combination of ratio and product type exponential estimator for estimating the finite population mean, *Statistics in Transition - New Series*, **9**, 105–115.

Srinath, K. P. (1971). Multiphase sampling in non-response problems, *Journal of the American Statistical Association*, **66**, 583–586.

Tabasum, R. and Khan, I. A. (2004). Double sampling for ratio estimation with non- response, *Journal of the Indian Society of Agricultural Statistics*, **58**, 300–306.

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.

Received July 2010; Accepted May 2011