DOI: 10.15415/mjis.2013.21004

Estimating the Population Mean in Two-Stage Sampling with Equal Size Clusters under Non-Response Using Auxiliary Characteristic

MANOJ K. CHAUDHARY and V. K. SINGH

Department of Statistics, Banaras Hindu University, Varanasi, India

E-mail: ritamanoj15@gmail.com

Abstract The present paper has been devoted to the study of estimating the population mean in two-stage sampling with equal size clusters under non-response using an auxiliary variable. This paper focuses on the study of general families of factor-type estimators of population mean considering two different cases in which non-response is observed on study variable only and on both study and auxiliary variables respectively. The optimum properties of the proposed families in both the cases are discussed. The empirical study is also carried out in support of the theoretical results.

Keywords: Two-stage sampling, Population mean, Auxiliary variable, Factor-type estimators, Non-response.

1. INTRODUCTION

In cluster sampling, all the elements of the selected clusters are enumerated. Though cluster sampling is economical under certain circumstances, it is generally less efficient than sampling of individual units directly. A compromise between cluster sampling and direct sampling of units can be achieved by selecting a sample of clusters and surveying only a sample of units in each sampled cluster instead of completely enumerating all the units in the sampled clusters. Thus, the procedure of first selecting clusters and then choosing a specified number of units from each selected cluster is known as two-stage sampling. The clusters which form the units of the sampling at the first stage are called the first-stage units and the elements or groups of elements within clusters which form the units of sampling at the second stage are called second-stage units.

To improve the efficiency of surveys, auxiliary information may be used in the sampling design or the estimation of parameters. There are several uses of auxiliary information for improving the efficiency of the estimators

Mathematical Journal of Interdisciplinary Sciences
Vol. 2, No. 1,
September 2013
pp. 43–55

at the estimation stage in two-stage sampling. Sahoo and Panda (1997) have defined a main class of estimators of population total in two-stage sampling. Sahoo and Panda (1999a, 1999b) extended the results to the situation when two-auxiliary variables are available in estimating the population mean of the study variable. Srivastava and Garg (2009) have proposed a general family of estimators for population mean using multi-auxiliary information in two-stage sampling.

In the present paper, our aim is to present some new estimation strategies in two-stage sampling design when first-stage units are of equal size and the knowledge on an auxiliary characteristic is available. The population is assumed to be suffered from the problem of non-response. We have defined two different families of estimators, based on Factor-Type Estimators (FTE) proposed by Singh and Shukla (1987), assuming (i) non-response occurs only in terms of study variable and (ii) non-response is observed both on study and auxiliary variables. Properties of defined strategies are discussed and have been illustrated with the help of empirical data.

2. SAMPLING STRATEGY AND ESTIMATION PROCEDURE UNDER NON-RESPONSE

Let us consider a population of size NM divided into N first-stage units (f.s.u's) each having M second-stage units (s.s.u's). A sample of size n is selected from the N f.s.u's with the help of simple random sampling without replacement (SRSWOR) scheme. From each of selected f.s.u's, a random sample of size m of s.s.u's, is drawn from M s.s.u's with the help of SRSWOR scheme. It is observed that the non-response occurs at second stage only and out of m s.s.u's, there are m_{i1} respondent and m_{i2} non-respondent units for the i^{th} f.s.u. $(i = 1,2,...,N)$.

In presence of non-response, using Hansen and Hurwitz (1946) procedure of sub sampling of non-respondents, we select a sub-sample of size h_{i2} units from the m_{i2} non-respondent units with the help of SRSWOR scheme for the i^{th} f.s.u. $(i = 1,2,...,N)$ selected in the sample such that $m_{i2} = k_i h_{i2} \left(k_i > 1\right)$ and information are observed on all the h_{i2} units by interview method. This process is done for all the n f.s.u's selected in the sample.

Let us suppose that X_0 be the study variable with population mean $\bar{X}_0$. First of all, we define an unbiased estimator of population mean for i^{th} f.s.u.$\left(\bar{X}_{oi}\right)$ in the presence of non-response as:

$$\bar{x}_{0HHi} = \frac{m_{i1}\bar{x}_{0i1} + m_{i2}\bar{x}_{0i2}^{*}}{m} \tag{2.1}$$

where $\bar{x}_{0i1}$ and $\bar{x}_{0i2}^{\,*}$ are the means per s.s.u. based on m_{i1} respondent units and h_{i2} non-respondent units respectively for the study variable. Thus, the estimator of population mean $\bar{X}_o$ under non-response is given by:

$$T_{\mathrm{OHH}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}}^{\mathrm{n}} \bar{\mathrm{x}}_{\mathrm{0HHi}}. \tag{2.2}$$

It is observed that the estimator T_{OHH} is an unbiased estimator of population mean $\bar{X}_o$. Thus, the variance of T_{OHH} can be obtained as:

$$\begin{aligned} V\left[T_{OHH}\right] &= V\left[E\left(T_{OHH}|n\right)\right] + E\left[V\left(T_{OHH}|n\right)\right] \\ &= \left(\frac{1}{n}-\frac{1}{N}\right)S_0^{*^2} + \frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right)\bar{S}_0^{\,2} + \frac{1}{nmN}\sum_{i=1}^{N}\left(k_i-1\right)W_{i2}S_{0i2}^2 \end{aligned} \tag{2.3}$$

where $S_0^{*^2} = \frac{1}{N-1}\sum_{i=1}^{N}\left[\bar{X}_{0i} - \bar{X}_0\right]^2$, $\bar{S}_0^{\,2} = \frac{1}{N}\sum_{i=1}^{N} S_{0i}^{\,2}$, $W_{i2} = \frac{M_{i2}}{M}$ (M_{i2} be the number of second stage units in the non-response group for the i^{th} f.s.u.). $S_{0i}^{\,2}$ and $S_{0i2}^{\,2}$ are respectively the mean squares of entire group and non-response group of study variable for the i^{th} f.s.u.

3. PROPOSED FAMILIES OF ESTIMATORS

In order to get the improved estimate of the population mean, we utilize the auxiliary information. Let us suppose that X_1 be the auxiliary variable with population mean $\bar{X}_1$. Due to Singh and Shukla (1987), a family of factor-type estimators for estimating the population mean $\bar{X}_0$ using auxiliary information, can be defined as (if the population is free from non-response):

$$T_\alpha = \bar{x}_0\left[\frac{\left(A+C\right)\bar{X}_1 + fB\bar{x}_1}{\left(A+fB\right)\bar{X}_1 + C\bar{x}_1}\right] \tag{2.4}$$

where $\bar{x}_0$ and $\bar{x}_1$ are the sample mean estimators based on *nm* s.s.u's for study and auxiliary variables respectively and:

$$f = \frac{n}{N}, A = \left(\alpha-1\right)\left(\alpha-2\right), B = \left(\alpha-1\right)\left(\alpha-4\right),$$

$$C = \left(\alpha-2\right)\left(\alpha-3\right)\left(\alpha-4\right); \alpha > 0 \text{ (to be determined).}$$

We shall now consider two different sampling strategies under non-response utilizing the concept of FTE.

(i) Non-response is present on study variable and information on auxiliary variable is obtained for all selected s.s.u's.

(ii) Non-response is present on both study and auxiliary variables at the second stage.

3.1 Non-Response on Study Variable Only

In this case, the family of factor-type estimators for estimating the population mean, $\bar{X}_o$ can be defined as:

$$T_\alpha^* = T_{OHH}\left[\frac{(A+C)\bar{X}_1 + fB\bar{x}_1}{(A+fB)\bar{X}_1 + C\bar{x}_1}\right] \tag{2.5}$$

where T_{OHH} is given by (2.2).

Remark 1: Putting different values of α in (2.5), we may generate some well known estimators of the family. For example, for $\alpha = 4, T_4^* = T_{OHH}$; for $\alpha = 1$, $T_1^* = T_{0HH}\frac{\bar{X}_1}{\bar{x}_1}$ a ratio-type estimator in two-stage sampling in presence of non-response.

3.1.1 Properties of Proposed Family

In order to find the bias and mean square error (MSE) of T_α^*, we use large sample approximation. Let us assume that:

$$T_{OHH} = \bar{X}_0(1+e_1), \bar{x}_1 = \bar{X}_1(1+e_2)$$

such that

$$E(e_1) = E(e_2) = 0$$

$$E(e_1^2) = \frac{E(T_{OHH} - \bar{X}_0)^2}{\bar{X}_0^2} = \frac{V(T_{OHH})}{\bar{X}_0^2},\ E(e_2^2) = \frac{E(\bar{x}_1 - \bar{X}_1)^2}{\bar{X}_1^2} = \frac{V(\bar{x}_1)}{\bar{X}_1^2} \text{ and}$$

$$E(e_1e_2) = \frac{E(T_{OHH} - \bar{X}_0)(\bar{x}_1 - \bar{X}_1)}{\bar{X}_0\bar{X}_1} = \frac{Cov(T_{OHH}, \bar{x}_1)}{\bar{X}_0\bar{X}_1}$$

Now, we can easily see that

$$V[T_{OHH}] = \left(\frac{1}{n} - \frac{1}{N}\right)S_0^{*2} + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\bar{S}_0^{\,2} + \frac{1}{nmN}\sum_{i=1}^{N}(k_i - 1)W_{i2}S_{0i2}^2,$$

$$V(\bar{x}_1) = \left(\frac{1}{n} - \frac{1}{N}\right)S_1^{*2} + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\bar{S}_1^{\,2},$$

$$Cov(T_{OHH}, \bar{x}_1) = \left(\frac{1}{n} - \frac{1}{N}\right)S_{01}^{\ *} + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\bar{S}_{01} \text{ due to singh (1998)}$$

where

$$S_1^{*2} = \frac{1}{N-1}\sum_{i=1}^{N}\left[\overline{X}_{1i} - \overline{X}_1\right]^2,\ \overline{S}_1^{\,2} = \frac{1}{N}\sum_{i=1}^{N} S_{1i}^{\,2},$$

$$S_{01}^{*} = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{X}_{0i} - \overline{X}_0)(\overline{X}_{1i} - \overline{X}_1),\ \overline{S}_{01} = \frac{1}{N}\sum_{i=1}^{N} S_{01i},$$

$$S_{01i} = \frac{1}{M-1}\sum_{j=1}^{M}(x_{0ij} - \overline{X}_{0i})(x_{1ij} - \overline{X}_{1i}),$$

$\overline{X}_{1i}$ and $S_{1i}^{\,2}$ are respectively the mean and mean square of entire group of auxiliary variable for the i^{th} f.s.u. and $x_{0ij}(x_{1ij})$ represents the observation on j^{th} s.s.u. in the i^{th} f.s.u. for the study(auxiliary) variable (i = 1,2...,N; j = 1,2,...M).

Thus, the bias of the family T_α^* up to the first order of approximation is given as:

$$B\left(T_\alpha^*\right) = \frac{\overline{X}_0\left(C - fB\right)}{\left(A + fB + C\right)}\left[\frac{C}{\left(A + fB + C\right)}\frac{V\left(\bar{x}_1\right)}{\overline{X}_1^{\,2}} - \frac{Cov\left(T_{OHH}, \bar{x}_1\right)}{\overline{X}_0\overline{X}_1}\right].$$

Let

$$\phi_1\left(\alpha\right) = \frac{fB}{A + fB + C}, \phi_2\left(\alpha\right) = \frac{C}{A + fB + C}\ \textit{and}\ \phi\left(\alpha\right) = \phi_2\left(\alpha\right) - \phi_1\left(\alpha\right) = \frac{C - fB}{A + fB + C}$$

then

$$B\left(T_\alpha^*\right) = \overline{X}_0\phi\left(\alpha\right)\left[\phi_2\left(\alpha\right)\frac{V\left(\bar{x}_1\right)}{\overline{X}_1^{\,2}} - \frac{Cov\left(T_{OHH}, \bar{x}_1\right)}{\overline{X}_0\overline{X}_1}\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\frac{\phi\left(\alpha\right)}{\overline{X}_1}\left[\phi_2\left(\alpha\right)RS_1^{*2} - S_{01}^{*}\right] + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\frac{\phi\left(\alpha\right)}{\overline{X}_1}\left[\phi_2\left(\alpha\right)R\overline{S}_1^{\,2} - \overline{S}_{01}\right] \quad (2.6)$$

where $R = \dfrac{\overline{X}_0}{\overline{X}_1}$.

Since T_α^* gives a biased estimate of $\overline{X}_o$, therefore, we can obtain MSE of T_α^* up to the first order of approximation as:

$$M\left(T_\alpha^*\right) = \overline{X}_0^{\,2}\left[\frac{V\left(T_{OHH}\right)}{\overline{X}_0^{\,2}} + \frac{\phi^2\left(\alpha\right)}{\overline{X}_1^{\,2}}V\left(\bar{x}_1\right) - 2\frac{\phi\left(\alpha\right)}{\overline{X}_0\overline{X}_1}Cov\left(T_{OHH}, \bar{x}_1\right)\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_0^{*2} + \phi^2(\alpha) R^2 S_1^{*2} - 2\phi(\alpha) R S_{01}^{*}\right]$$

$$+ \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\left[\overline{S}_0^{*2} + \phi^2(\alpha) R^2 \overline{S}_1^{2} - 2\phi(\alpha) R \overline{S}_{01}\right] + \frac{1}{nmN}\sum_{i=1}^{N}(k_i - 1) W_{i2} S_{0i2}^2. \quad (2.7)$$

3.1.2 Optimum Value of α

In order to find the optimum estimator of the proposed family, we minimize $M\left(T_\alpha^*\right)$ with respect to α. On differentiating $M\left(T_\alpha^*\right)$ with respect to α and equating the derivative to zero, we get:

$$\left(\frac{1}{n} - \frac{1}{N}\right)\left[2\phi'(\alpha)\phi(\alpha) R^2 S_1^{*2} - 2\phi'(\alpha) R S_{01}^{*}\right] + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\left[2\phi'(\alpha)\phi(\alpha) R^2 \overline{S}_1^{2} - 2\phi'(\alpha) R \overline{S}_{01}\right] = 0 \quad (2.8)$$

$$\Rightarrow \qquad \phi(\alpha) = \frac{\left(\frac{1}{n} - \frac{1}{N}\right) R S_{01}^{*} + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right) R \overline{S}_{01}}{\left(\frac{1}{n} - \frac{1}{N}\right) R^2 S_1^{*2} + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right) R^2 \overline{S}_1^{2}} = V(\text{say}). \quad (2.9)$$

Since $\phi(\alpha) = \frac{C - fB}{A + fB + C}$, thus from equation (2.9), we have

$$\frac{C - fB}{A + fB + C} = V, \quad (2.10)$$

where V, being a function of parameters of the population, is a constant for a given population. The above expression is a cubic equation in α and, therefore, on solving, we may get at the most three real and positive optimum values of the parameter α for which $M\left(T_\alpha^*\right)$ would attain its minimum.

3.2 Non-Response on Both Study and Auxiliary Variables

If the auxiliary variable is also subjected to non-response, then similar to study variable, an unbiased estimator of population mean $\overline{X}_1$, based upon the sub-samples of non-respondents, may be defined as:

$$T_{1HH} = \frac{1}{n}\sum_{i}^{n} \overline{x}_{1HHi} \quad (2.11)$$

where $\bar{x}_{1HHi} = \dfrac{m_{i1}\bar{x}_{1i1} + m_{i2}\bar{x}_{1i2}^{*}}{m}$. Further, $\bar{x}_{1i1}$ and $\bar{x}_{1i2}^{*}$ are the means per s.s.u. based on m_{i1} respondent units and h_{i2} non-respondent units respectively for the auxiliary variable.

Thus, based upon factor-type estimation technique, the family of estimators for population mean $\bar{X}_0$ is defined as:

$$T_{\alpha}^{(*)} = T_{OHH}\left[\frac{(A+C)\bar{X}_1 + fBT_{1HH}}{(A+fB)\bar{X}_1 + CT_{1HH}}\right] \quad (2.12)$$

Remark 2: Letting $\alpha = 1$, we have $T_1^{(*)} = T_{OHH}\dfrac{\bar{X}_1}{T_{1HH}}$, a ratio-type estimator in two-stage sampling when both the variables are subjected to non-response. Similarly, for $\alpha = 4$, we get $T_4^{(*)} = T_{OHH}$, which is same as T_4^{*}, defined in sub-section 3.1. For other choices of α, similarly, one may get non-response versions of some existing estimators defined in two-stage sampling utilizing auxiliary characteristic.

3.2.1 Properties of the Proposed Family

We use large sample approximation for finding the bias and MSE of $T_{\alpha}^{(*)}$. Let us again assume that:

$$T_{OHH} = \bar{X}_0(1+e_1) \text{ and } T_{1HH} = \bar{X}_1(1+e_2')$$

such that

$$E(e_1) = E(e_2') = 0,$$

$$E(e_1^2) = \frac{V(T_{OHH})}{\bar{X}_0^2},\ E(e_2'^2) = \frac{V(T_{1HH})}{\bar{X}_1^2}$$

and

$$E(e_1, e_2') = \frac{Cov(T_{OHH}, T_{1HH})}{\bar{X}_o\bar{X}_1}$$

where

$$V(T_{1HH}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_1^{*2} + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\bar{S}_1^{2} + \frac{1}{nmN}\sum_{i=1}^{N}(k_i - 1)W_{i2}S_{1i2}^2$$

$$Cov(T_{OHH}, T_{1HH}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_{o1}^{*} + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\bar{S}_{01} + \frac{1}{nmN}\sum_{i=1}^{N}(k_i - 1)W_{i2}S_{01i2}.$$

These results are due to Singh (1998). Further, S_{1i2}^2 is the mean square of the non-response group for the auxiliary variable in i^{th} f.s.u. and

$$S_{01i2} = \frac{1}{M_{i2}-1}\sum_{j}^{M_{i2}}(x_{0ij} - \overline{X}_{0i2})(x_{1ij} - \overline{X}_{1i2}).$$

where $\overline{X}_{0i2}$ and $\overline{X}_{1i2}$ are respectively the means of study and auxiliary variables of non-response group for the i^{th} f.s.u.

Thus, the bias up to first order of approximation can be obtained as:

$$\begin{aligned} B\left(T_\alpha^{(*)}\right) &= E\left(T_\alpha^{(*)}\right) - \overline{X}_0 \\ &= \overline{X}_0\phi(\alpha)\left[\frac{\phi_2(\alpha)}{\overline{X}_1^{\,2}}V(T_{1HH}) - \frac{Cov(T_{OHH},T_{1HH})}{\overline{X}_0\overline{X}_1}\right] \\ &= \frac{\phi(\alpha)}{\overline{X}_1}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\left\{\phi_2(\alpha)RS_1^{*2} - S_{01}^{\;*}\right\}\right. \\ &\quad + \frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right)\left\{\phi_2(\alpha)R\overline{S}_1^{\,2} - \overline{S}_{01}\right\} \\ &\quad \left. + \frac{1}{nmN}\sum_{i=1}^{N}(k_i-1)W_{i2}\left\{\phi_2(\alpha)RS_{1i2}^{\;2} - S_{01i2}\right\}\right]. \end{aligned} \tag{2.13}$$

Now, the MSE up to the first order of approximation, can be obtained as:

$$\begin{aligned} M\left(T_\alpha^{(*)}\right) &= E\left[T_\alpha^{(*)} - \overline{X}_0\right]^2 \\ &= \overline{X}_0^{\,2}\left[\frac{V(T_{OHH})}{\overline{x}_0^{\,2}} + \frac{\phi^2(\alpha)}{\overline{X}_1^{\,2}}V(T_{1HH}) - 2\frac{\phi(\alpha)}{\overline{X}_0\overline{X}_1}Cov(T_{OHH},T_{1HH})\right] \\ &= \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_0^{*2} + \phi^2(\alpha)R^2S_1^{*2} - 2\phi(\alpha)RS_{01}^{\;*}\right] \\ &\quad + \frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right)\left[\overline{S}_0^{\,2} + \phi^2(\alpha)R^2\overline{S}_1^{\,2} - 2\phi(\alpha)R\overline{S}_{01}\right] \\ &\quad + \frac{1}{nmN}\sum_{i=1}^{N}(k_i-1)W_{i2}\left[S_{0i2}^{\;2} + \phi^2(\alpha)R^2S_{1i2}^{\;2} - 2\phi(\alpha)RS_{01i2}\right] \end{aligned} \tag{2.14}$$

3.2.2 Optimum Value of α

On minimizing $M\left(T_{\alpha}^{(*)}\right)$ with respect to α, we get the normal equation as:

$$\left(\frac{1}{n}-\frac{1}{N}\right)\left[2\phi^{'}(\alpha)\phi(\alpha)R^2S_1^{*2}-2\phi^{'}(\alpha)RS_{01}^{*}\right]$$
$$+\frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right)\left[2\phi^{'}(\alpha)\phi(\alpha)R^2\overline{S}_1^{2}-2\phi^{'}(\alpha)R\overline{S}_{01}\right] \quad (2.15)$$
$$+\frac{1}{nmN}\sum_{i=1}^{N}(k_i-1)W_{i2}\left[2\phi^{'}(\alpha)\phi(\alpha)R^2S_{1i2}^{2}-2\phi^{'}(\alpha)RS_{01i2}\right]=0$$

$$\Rightarrow \phi(\alpha)=\frac{\left(\frac{1}{n}-\frac{1}{N}\right)RS_{01}^{*}+\frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right)R\overline{S}_{01}+\frac{1}{nmN}\sum_{i=1}^{N}(k_i-1)W_{i2}RS_{01i2}}{\left(\frac{1}{n}-\frac{1}{N}\right)R^2S_1^{*2}+\frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right)R^2\overline{S}_1^{2}+\frac{1}{nmN}\sum_{i=1}^{N}(k_i-1)W_{i2}R^2S_{1/2}^{2}}$$
$$=V(\text{say}) \quad (2.16)$$

Thus, equation (2.16) can be solved for getting optimum values of α (at the most three real and positive roots of the equation), given the value of the constant V. The optimum values of α, when substituted in the expression (2.14), gives the minimum MSE of $T_{\alpha}^{(*)}$.

4. EMPIRICAL STUDY

In order to understand the applications of the results obtained in this paper and to observe the behavior of the estimators, it is essential to illustrate whatever has been discussed in previous sections with some empirical data. However, due to non-availability of suitable empirical data for the purpose, we have considered fictitious data in the following manner:

We selected 20 random numbers from random number table (Rao et al, 1966) in the bunches of 25 clusters. The four-digit random numbers were converted into two-digit numbers by placing the decimal after two digits so as to reduce the magnitudes of the numbers. The 500 numbers, so selected, were assumed to be values of the study variable in the population of 25 clusters of size 20 each. The same procedure was again repeated in order to generate corresponding values of the auxiliary variable in the population. Thus, we have N = 25, M = 20, NM = 500.

Taking m =5 and n = 10, we have illustrated the results.

Table 1 shows values of some of the parameters of clusters in the population.

Table 1: Cluster Means, Mean-squares, Covariances, Mean Ratios, Correlation Coefficients under Study and Auxiliary Variables

Cluster No.(f.s.u.)	$\bar{X}_{0i}$	$\bar{X}_{1i}$	S_{0i}^2	S_{1i}^2	S_{01i}	R_{01i}	ρ_{01i}
1	50.7540	51.0130	914.8451	986.13	870.1267	0.9949	0.964
2	50.5410	53.1795	1082.508	1142.36	1033.5187	0.9907	0.978
3	58.5375	60.1635	596.7012	492.29	460.3504	0.9730	0.894
4	59.4430	64.8130	603.4674	435.85	442.9060	0.9171	0.909
5	42.5170	39.1020	848.9938	737.07	700.4796	1.0873	0.932
6	61.2870	67.0730	683.1054	616.70	515.8315	0.9137	0.837
7	43.7235	44.8255	987.1914	711.09	685.1946	0.9754	0.913
8	51.2645	55.1395	852.0943	797.20	621.9776	0.9297	0.794
9	46.4745	50.4095	728.9443	767.63	648.6315	0.9219	0.913
10	59.9545	56.7095	1112.4090	975.88	861.0674	1.0572	0.870
11	46.8555	49.4525	851.2535	824.78	699.8488	0.9475	0.879
12	53.2215	51.2295	862.5906	814.66	696.2039	1.0389	0.874
13	39.1335	45.7105	891.1400	920.70	805.2625	0.8561	0.936
14	49.9330	49.7260	1086.7420	996.17	888.9422	1.0042	0.899
15	44.8975	44.7420	850.0099	767.11	722.2890	1.0035	0.942
16	47.1550	50.0750	815.736	931.02	783.1795	0.9417	0.946
17	56.8980	54.3925	1199.152	945.81	983.4855	1.0461	0.972
18	39.3300	36.8195	838.8203	640.68	649.9297	1.0682	0.933
19	52.8070	54.9235	900.905	1027.30	874.0159	0.9615	0.956
20	39.9350	39.3340	1152.004	952.92	971.3947	1.0153	0.976
21	43.0405	42.8095	1127.564	884.75	918.1830	1.0054	0.968
22	51.0170	50.4215	903.9001	969.57	854.9900	1.0118	0.961
23	53.6635	56.0410	773.7046	676.10	644.2793	0.9576	0.938
24	50.1515	49.4265	765.2139	755.95	659.9247	1.0147	0.913
25	53.6385	56.5140	889.826	792.80	773.8380	0.9491	0.970

For the study purpose, we have assumed that:

$$S_{0i2}{}^2 = \frac{4}{5}\left(S_{0i}^2\right),\ S_{1i2}{}^2 = \frac{4}{5}\left(S_{1i}^2\right),\ S_{01i2} = \frac{4}{5}\left(S_{01i}\right)$$

Further, from the above data, we get:

$$\bar{X}_0 = 49.8470,\ \bar{X}_1 = 50.8752,\ S_0^{*2} = 42.3954, =\ S_1^{*2}\, 55.1814,\ \bar{S}_0{}^2 = 892.7529,$$

$$\bar{S}_1^2 = 887.7658,\ S_{01}^* = 44.8808,\ \bar{S}_{01} = 750.6340, R = 0.9798.$$

From the above values of the parameters, equation (2.9) gives $\phi(\alpha) = 0.8564$ for the estimator T_{α}^{*}. Table 2 shows the values of $M\left(T_{\alpha}^{*}\right)$ for $\alpha = 1,4$ and α_{opt} for $W_{i2} = 0.1$ (0.1) 0.4 in each cluster.

Table 2: MSE of T_{α}^{*} for α =1, 4 and α_{opt}, W_{i2}=0.1 (0.1) 0.4 and k_i =2 (0.5) 3.5 for all i

W_{i2}	α	k_i			
		2	**2.5**	**3**	**3.5**
0.1	1	5.9847	6.6989	7.4131	8.1273
	4	17.3634	18.0776	18.7918	19.5060
	α_{opt}	5.6557	6.3699	7.0841	7.7983
0.2	1	7.4131	8.8415	10.2699	11.6983
	4	18.7918	20.2202	21.6486	23.0770
	α_{opt}	7.0841	8.5125	9.9409	11.3693
0.3	1	8.8415	10.9842	13.1268	15.2694
	4	20.2202	22.3628	24.5054	26.6480
		8.5125	10.6551	12.7977	14.9403
0.4	1	10.2699	13.12676	15.9836	18.8404
	4	21.6486	24.5054	27.3623	30.2191
	α_{opt}	9.9409	12.7977	15.6545	18.5114

Table 3: Values of $\phi(\alpha)$ for the Estimator T_{α}^{*} with Different Values of W_{i2} and k_i

W_{i2}	k_i			
	2	2.5	3	3.5
0.1	0.8619	0.8644	0.8667	0.8688
0.2	0.8666	0.8708	0.8745	0.8777
0.3	0.8708	0.8761	0.8806	0.8844
0.4	0.8745	0.8806	0.8855	0.8896

Table 4 below shows the values of $M\left(T_{\alpha}^{(*)}\right)$ for α =1, 4 and α_{opt} with W_{i2} = 0.1 (0.1) 0.4 k_i and = 2 (0.5) 3.5 for all i.

Table 4: MSE of T_{α}^{*} for α =1, 4 and α_{opt} with W_{i2} =0.1 (0.1) 0.4 and k_i =2 (0.5) 3.5 for all i.

W_{i2}	α	k_i			
		2	2.5	3	3.5
0.1	1	4.8946	5.0637	5.2329	5.4020
	4	17.3634	18.0776	18.7918	19.5060
	α_{opt}	4.5662	4.7353	4.9042	5.0730
0.2	1	5.2329	5.5711	5.9094	6.2477
	4	18.7918	20.2202	21.6486	23.0770
	α_{opt}	4.9042	5.2415	5.5782	5.9144
0.3	1	5.5711	6.0785	6.5859	7.0933
	4	20.2202	22.3628	24.5054	26.6480
	α_{opt}	5.2415	5.7464	6.2502	6.7532
0.4	1	5.9094	6.5859	7.2625	7.9390
	4	21.6486	24.5054	27.3623	30.2191
	α_{opt}	5.5782	6.2502	6.9208	7.5903

5. CONCLUSIONS

We have suggested two different general families of factor-type estimators for estimating the population mean in two-stage sampling with equal size clusters under non-response using an auxiliary variable. The suggested families can generate a number of well known estimators on different choices of α. A comparison of values of MSE of the estimator T_{α}^{*} in the table 2 reveals that for a given value of the parameter, α, MSE increases with increase in non-response rate and also with smaller size of sub-samples of non-respondents. The result is also intuitively expected. Further, the same trend is exhibited in table 4 for the estimator $T_{\alpha}^{(*)}$. In both the cases, MSEs of the estimators for optimum are slightly smaller than that obtained for α=1, implying that ratio-type estimators are almost as much précised as the optimum estimator. A comparison of values obtained in Table 4 makes it clear that the efficiency of the strategy is almost unaffected by the size of the sub-sample of non-respondents, whatever be the non-response rate. This might be due to the assumption of large sample approximation while deriving the MSE of the estimator.

REFERENCES

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. http://dx.doi.org/10.1080/01621459.1946.10501894

Rao, C. R., Mitra, S. K. and Mathai, A. (1966) : *Formulae and Tables for Statistical Work*, Statistical publishing Society, Calcutta, India.

Sahoo, L. N. and Panda, P. (1997) : A class of estimators in two-stage sampling with varying probabilities, *South African Statistical journal*, 31, 151-160.

Sahoo, L. N. and Panda, P. (1999a) : A class of estimators using auxiliary information in two-stage sampling, *Australian & New Zealand journal of Statistics*, 41 (4), 405-410. http://dx.doi.org/10.1111/1467-842X.00094

Sahoo, L. N. and Panda, P. (1999b) : A predictive regression type estimator in two-stage sampling, *Journal of The Indian Society of Agricultural Statistics*, 52 (3), 303-308.

Singh, L. B. (1998) : *Some Classes of Estimators for Finite Population Mean in Presence of Non-response*, Ph. D. Thesis submitted to Banaras Hindu University, Varanasi, India.

Singh, V. K. and Shukla, D. (1987) : One parameter family of factor-type ratio estimators, *Metron*, 45 (1-2), 273-283.

Srivastava, M. and Garg, N. (2009) : A general class of estimators of a finite population mean using multi-auxiliary information under two-stage sampling scheme, Journal of Reliability and Statistical Studies, 1(2), 103-118.