Journal of Scientific Research
Banaras Hindu University, Varanasi

Vol. 62, 2018 : 213-222
ISSN : 0447-9483

STUDY ON THE USE OF DOUBLE SAMPLING SCHEME FOR ESTIMATING THE POPULATION MEAN IN TWO-STAGE...

***Manoj K. Chaudhary and Saurabh**
Department of Statistics, Banaras Hindu University, Varanasi, India.
***Corresponding author:e-mail: ritamanoj15@gmail.com**

Abstract

In the present paper, we have proposed some ratio and product-type estimators for estimating the population mean in two-stage sampling using the information on an auxiliary variable under non-response. The study has been presented under the situation in which the population mean of auxiliary variable is not known and hence the idea of double sampling scheme has been adopted in proposing the estimators. The expressions for biases and mean square errors of the proposed estimators have been derived. An empirical study has also been carried out to demonstrate the theoretical results.

Keywords: Two-stage sampling, equal size clusters, double sampling scheme, auxiliary information, non-response

Introduction

Two-stage sampling scheme is a particular kind of multi-stage sampling where the units are chosen in two different stages. In cluster sampling, once the clusters are selected, all the units of the selected clusters are enumerated. But, 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 auxiliary information may be utilized to enhance the precision of the estimates in two-stage sampling also. There are several uses of auxiliary information in increasing the precision of the estimators under the two-stage sampling scheme. Sahoo and Panda (1997) have proposed a main class of estimators for estimating the population total in two-stage sampling. Bellhouse and Rao (1986) have mentioned that the prediction estimator may be only marginally better than the classical estimator in Probability Proportional to Size (PPS) sampling in two-stage sampling design scheme.

It is well known fact that the non-response is inherent in the mail surveys. The problem of non-response was first tackled by Hansen and Hurwitz (1946) and they suggested a technique of sub-sampling of non-respondents to cope up the trouble of non-response. In order to suggest the estimators in presence of non-response. Rao (1986, 1987, 1990) proposed some alternate ratio and regression-type estimators. Khare (1992) discussed the problem of two-phase sampling regression estimator in the presence of non-response. Srivastava (1993) has studied several ratio, product and regression-type estimators under fixed and super population model approaches discussing non-response. Chaudhary and Singh (2013) have proposed some families of factor-type estimators of population mean in two-stage sampling using the information on an auxiliary variable under non-response.

In this article, we have tried to estimate the population mean in two-stage sampling scheme utilizing the information on an auxiliary variable under non-response. Some ratio and product-type estimators have been proposed by adopting the idea of double sampling scheme under the situation in which the information about population mean of auxiliary variable is not available and non-response is observed on study variable only. The properties of the proposed estimators have been discussed in detail.

2. Sampling Strategy and Proposed Estimators

Let a population of NM units be divided into N first-stage units (f.s.u'^{s}) each having M second-stage units (s.s.u'^{s}). Let Y and X be the study and auxiliary variables with respective means $\bar{Y}_{..}$ and $\bar{X}_{..}$. Let y_{ij} and x_{ij} be the observations on j^{th} s.s.u. in the i^{th} f.s.u. under study and auxiliary variables respectively $(i = 1,2,...,N),(j = 1,2,...,M)$. Further, it is assumed that only the study variable is suffering from non-response. Under the above circumstances, first a sample of n f.s.u'^{s} is selected from the N f.s.u'^{s} by the method of simple random sampling without replacement (SRSWOR) scheme. Adopting the idea of double sampling scheme, secondly, a larger first-phase sample of m' s.s.u'^{s} is selected from each of the f.s.u'^{s} selected at the first stage with the help of SRSWOR. We further select a smaller second-phase sample of m s.s.u'^{s} from the m' s.s.u'^{s} $(m < m')$ for each of the f.s.u'^{s} selected at the first stage by SRSWOR. At this stage, it is observed that out of m s.s.u'^{s}, m_{i1} units respond and m_{i2} units do not respond on Y in the i^{th} f.s.u. Following the Hansen and Hurwitz's (1946) technique of sub-sampling of non-respondents, we now select a sub-sample of h_{i2} units from m_{i2} non-respondent units by SRSWOR $(h_{i2} = m_{i2}/k_i\ , k_i \geq 1)$ and collect the information from all the h_{i2} units. Moreover, we have

$$\bar{Y}_{..} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} = \frac{1}{N}\sum_{i=1}^{N}\bar{Y}_{i.} \quad \text{and} \quad \bar{X}_{..} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij} = \frac{1}{N}\sum_{i=1}^{N}\bar{X}_{i.}$$

where $\bar{Y}_{i.} = \frac{1}{M}\sum_{j=1}^{M} y_{ij}$ and $\bar{X}_{i.} = \frac{1}{M}\sum_{j=1}^{M} x_{ij}$

Now, the estimator of $\bar{X}_{..}$ at the first phase is given as

$$\bar{x}'_{..} = \frac{1}{n}\sum_{i}^{n} \bar{x}'_{i.} \tag{1}$$

where $\bar{x}'_{i.} = \frac{1}{m'}\sum_{j}^{m'} x_{ij}$

At the second phase, the estimators of population means $\bar{Y}_{..}$ and $\bar{X}_{..}$ are respectively given by

$$\bar{y}^{*}_{nm} = \frac{1}{n}\sum_{i}^{n} \bar{y}^{*}_{im} \tag{2}$$

and $$\bar{x}_{nm} = \frac{1}{n}\sum_{i}^{n} \bar{x}_{im} \tag{3}$$

where $\bar{y}^{*}_{im} = \frac{m_{i1}\bar{y}_{mi1} + m_{i2}\bar{y}_{hi2}}{m}$, $\bar{x}_{im} = \frac{1}{m}\sum_{j}^{m} x_{ij}$, $\bar{y}_{mi1}$ and $\bar{y}_{hi2}$ are the means based on m_{i1} respondent units and h_{i2} non-respondent units respectively for the study variable.

We now propose some separate and combined-type ratio and product estimators of population mean $\bar{Y}_{..}$ in two-stage sampling under the situation in which population mean of auxiliary variable is not known and non-response is observed on study variable only as

$$T_1 = \frac{1}{n}\sum_{i}^{n} \bar{y}^{*}_{(im)R} \tag{4}$$

$$T_2 = \frac{1}{n}\sum_{i}^{n} \bar{y}^{*}_{(im)P} \tag{5}$$

$$T_3 = \frac{\bar{y}^{*}_{nm}}{\bar{x}_{nm}} \bar{x}'_{..} \tag{6}$$

$$T_4 = \frac{\bar{y}^{*}_{nm}}{\bar{x}'_{..}} \bar{x}_{nm} \tag{7}$$

where $\bar{y}^{*}_{(im)R} = \frac{\bar{y}^{*}_{im}}{\bar{x}_{im}} \bar{x}'_{i.}$ and $\bar{y}^{*}_{(im)P} = \frac{\bar{y}^{*}_{im}}{\bar{x}'_{i.}} \bar{x}_{im}$.

2.1 Properties of the Proposed Estimators

In order to get the biases and mean square errors (MSE) of the proposed estimators, we use the theory of large sample approximation. Let

$\bar{y}^*_{im} = \bar{Y}_{i.}(1+e_0)$, $\bar{x}_{im} = \bar{x}'_{i.}(1+e_1)$, $\bar{x}'_{i.} = \bar{X}_{i.}(1+e_2)$, $\bar{y}^*_{nm} = \bar{Y}_{..}(1+e_3)$, $\bar{x}_{nm} = \bar{x}'_{..}(1+e_4)$, $\bar{x}'_{..} = \bar{X}_{..}(1+e_5)$

such that $E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = E(e_5) = 0$,

$$E(e_0^2) = \frac{1}{\bar{Y}_{i.}^2}\left[\left(\frac{1}{m} - \frac{1}{M}\right)S_{Yi}^2 + \frac{W_{i2}(k_i - 1)}{m}S_{Yi2}^2\right],\ E(e_1^2) = \frac{1}{\bar{X}_{i.}^2}\left[\left(\frac{1}{m} - \frac{1}{m'}\right)S_{Xi}^2\right],$$

$$E(e_2^2) = \frac{1}{\bar{X}_{i.}^2}\left[\left(\frac{1}{m'} - \frac{1}{M}\right)S_{Xi}^2\right]$$

$$E(e_3^2) = \frac{1}{\bar{Y}_{..}^2}\left[\left(\frac{1}{n} - \frac{1}{N}\right)S_{bY}^2 + \frac{1}{nN}\left(\frac{1}{m} - \frac{1}{M}\right)\sum_{i=1}^{N}S_{Yi}^2 + \frac{1}{nmN}\sum_{i=1}^{N}W_{i2}(k_i - 1)S_{Yi2}^2\right],$$

$$E(e_4^2) = \frac{1}{\bar{X}_{..}^2}\left[\left(\frac{1}{n} - \frac{1}{N}\right)S_{bX}^2 + \frac{1}{nN}\left(\frac{1}{m} - \frac{1}{m'}\right)\sum_{i=1}^{N}S_{Xi}^2\right],$$

$$E(e_5^2) = \frac{1}{\bar{X}_{..}^2}\left[\left(\frac{1}{n} - \frac{1}{N}\right)S_{bX}^2 + \frac{1}{nN}\left(\frac{1}{m'} - \frac{1}{M}\right)\sum_{i=1}^{N}S_{Xi}^2\right]$$

$$E(e_0 e_1) = \frac{1}{\bar{X}_{i.}\bar{Y}_{i.}}\left[\left(\frac{1}{m} - \frac{1}{m'}\right)\rho_{XYi}S_{Xi}S_{Yi}\right] \quad \text{and}$$

$$E(e_3 e_4) = \frac{1}{\bar{Y}_{..}\bar{X}_{..}}\left[\left(\frac{1}{n} - \frac{1}{N}\right)S_{bXY} + \frac{1}{nN}\left(\frac{1}{m} - \frac{1}{m'}\right)\sum_{i=1}^{N}\rho_{XYi}S_{Xi}S_{Yi}\right]$$

where $S_{Yi}^2 = \frac{1}{M-1}\sum_{j=1}^{M}\left(y_{ij} - \bar{Y}_{i.}\right)^2$, $S_{Xi}^2 = \frac{1}{M-1}\sum_{j=1}^{M}\left(x_{ij} - \bar{X}_{i.}\right)^2$,

$S_{XYi} = \frac{1}{M-1}\sum_{j=1}^{M}\left(y_{ij} - \bar{Y}_{i.}\right)\left(x_{ij} - \bar{X}_{i.}\right)$, $\rho_{XYi} = \frac{S_{XYi}}{S_{Xi}S_{Yi}}$, $S_{bY}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i.} - \bar{Y}_{...}\right)^2$,

$S_{bX}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_{i.} - \bar{X}_{..}\right)^2$, $S_{bXY} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_{i.} - \bar{X}_{..}\right)\left(\bar{Y}_{i.} - \bar{Y}_{..}\right)$, S_{Yi2}^2 is the mean square of the non-response group in the i^{th} f.s.u. for the study variable and W_{i2} is the non-response rate in the i^{th} f.s.u.

Now, the bias of the estimator T_1 is obtained as

$$Bias(T_1) = E\left\{\frac{1}{n}\sum_{i}^{n} E\left(\frac{\bar{y}_{im}^{*}}{\bar{x}_{im}}\bar{x}_{i.}' \,/\, i\right)\right\} - \bar{Y}_{..}$$

$$= \frac{1}{N}\sum_{i=1}^{N} E\left\{\bar{Y}_{i.}(1+e_0)\frac{\bar{x}_{i.}'}{\bar{x}'(1+e_1)} / i\right\} - \bar{Y}_{..}$$

$$= \frac{1}{N}\sum_{i=1}^{N} E\left\{\bar{Y}_{i.}(1+e_0)(1-e_1+e_1^2-...)/ i\right\} - \bar{Y}_{..}$$

Solving the above expression on neglecting the terms having powers of e_0 and e_1 greater than two, we get

$$Bias(T_1) = \frac{1}{N}\sum_{i=1}^{N}\left[E(e_1^2) - E(e_0 e_1)\right]$$

Hence, the expression for required bias up to the first order of approximation is given by

$$Bias(T_1) = \frac{1}{N}\left(\frac{1}{m} - \frac{1}{m'}\right)\sum_{i=1}^{N}\bar{Y}_{i.}\left(\frac{S_{Xi}^2}{\bar{X}_{i.}^2} - \frac{\rho_{XYi}S_{Xi}S_{Yi}}{\bar{X}_{i.}\bar{Y}_{i.}}\right) \tag{8}$$

Similarly, the biases of the estimators T_2, T_3 and T_4 up to the first order of approximation are respectively given as

$$Bias(T_2) = \frac{1}{N}\left(\frac{1}{m} - \frac{1}{m'}\right)\sum_{i=1}^{N}\left(\frac{\rho_{XYi}S_{Xi}S_{Yi}}{\bar{X}_{i.}}\right) \tag{9}$$

$$Bias(T_3) = \bar{Y}_{..}\left[\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{S_{bX}^2}{\bar{X}_{..}^2} - \frac{S_{bXY}}{\bar{X}_{..}\bar{Y}_{..}}\right) + \frac{1}{nN}\left(\frac{1}{m} - \frac{1}{m'}\right)\sum_{i=1}^{N}\left(\frac{S_{Xi}^2}{\bar{X}_{..}^2} - \frac{\rho_{XYi}S_{Xi}S_{Yi}}{\bar{X}_{..}\bar{Y}_{..}}\right)\right] \tag{10}$$

and

$$Bias(T_4) = \bar{Y}_{..}\left[\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{S_{bXY}}{\bar{X}_{..}\bar{Y}_{..}}\right) + \frac{1}{nN}\left(\frac{1}{m} - \frac{1}{m'}\right)\sum_{i=1}^{N}\left(\frac{\rho_{XYi}S_{Xi}S_{Yi}}{\bar{X}_{..}\bar{Y}_{..}}\right)\right] \tag{11}$$

Now, the MSE of the estimator T_1 is obtained as

$$MSE\,(T_1) = E\,\{MSE\,(T_1/n)\} + MSE\,\{E(T_1/n)\}$$

$$= E\left\{MSE\left\{\frac{1}{n}\sum_{i}^{n}(\bar{y}_{(im)R}^{*}/i)\right\}\right\} + MSE\left\{\frac{1}{n}\sum_{i}^{n}E\,(\bar{y}_{(im)R}^{*}/i)\right\}$$

$$= E\left\{\frac{1}{n^2}\sum_{i}^{n}MSE\,(\bar{y}_{(im)R}^{*})\right\} + MSE\left\{\frac{1}{n}\sum_{i}^{n}E\,(\bar{y}_{(im)R}^{*}/i)\right\}$$

Since, $MSE(\bar{y}^*_{(im)P}) =$

$$\left(\frac{1}{m'}-\frac{1}{M}\right)S^2_{Yi}+\left(\frac{1}{m}-\frac{1}{m'}\right)\left(S^2_{Yi}+R^2_{XYi}S^2_{Xi}+2R_{XYi}\rho_{XYi}S_{Xi}S_{Yi}\right)+\left(k_i-1\right)\frac{W_{i2}S^2_{Yi2}}{m}$$

Thus, the MSE of T_1 up to the first order of approximation is given as

$$MSE(T_1)=\frac{1}{nN}\sum_{i=1}^{N}\left\{\left(\frac{1}{m'}-\frac{1}{M}\right)S^2_{Yi}+\left(\frac{1}{m}-\frac{1}{m'}\right)\left(S^2_{Yi}+R^2_{XYi}S^2_{Xi}-2R_{XYi}\rho_{XYi}S_{Xi}S_{Yi}\right)+\left(k_i-1\right)\frac{W_{i2}S^2_{Yi2}}{m}\right\}+\left(\frac{1}{n}-\frac{1}{N}\right)S^2_{bY} \quad (12)$$

where $R_{XYi}=\bar{Y}_{i.}/\bar{X}_{i.}$.

Similarly, the mean square error of T_2 is obtained as

$$MSE(T_2)=E\ \{MSE(T_2/n)\}+MSE\{E\ (T_2/n)\}$$

$$=E\ \left\{MSE\left\{\frac{1}{n}\sum_{i}^{n}(\bar{y}^*_{(im)P}/i)\right\}\right\}+MSE\left\{\frac{1}{n}\sum_{i}^{n}E\ (\bar{y}^*_{(im)P}/i)\right\}$$

$$=E\ \left\{\frac{1}{n^2}\sum_{i}^{n}MSE(\bar{y}^*_{(im)P})\right\}+MSE\left\{\frac{1}{n}\sum_{i}^{n}E\ (\bar{y}^*_{(im)P}/i)\right\}$$

Since, $MSE(\bar{y}^*_{(im)P}) =$

$$\left(\frac{1}{m'}-\frac{1}{M}\right)S^2_{Yi}+\left(\frac{1}{m}-\frac{1}{m'}\right)\left(S^2_{Yi}+R^2_{XYi}S^2_{Xi}+2R_{XYi}\rho_{XYi}S_{Xi}S_{Yi}\right)+\left(k_i-1\right)\frac{W_{i2}S^2_{Yi2}}{m}$$

Therefore, the MSE of T_2 up to first order of approximation is represented as

$$MSE(T_2)=\frac{1}{nN}\sum_{i=1}^{N}\left\{\left(\frac{1}{m'}-\frac{1}{M}\right)S^2_{Yi}+\left(\frac{1}{m}-\frac{1}{m'}\right)\left(S^2_{Yi}+R^2_{XYi}S^2_{Xi}+2R_{XYi}\rho_{XYi}S_{Xi}S_{Yi}\right)+\left(k_i-1\right)\frac{W_{i2}S^2_{Yi2}}{m}\right\}+\left(\frac{1}{n}-\frac{1}{N}\right)S^2_{bY} \quad (13)$$

Note: The expressions for MSE of the estimators $\bar{y}^*_{(im)R}$ and $\bar{y}^*_{(im)P}$ have been obtained by using large sample approximations.

The MSE of the estimator T_3 is envisaged as

$$MSE(T_3)=E\left[T_3-\bar{Y}_{..}\right]^2$$

$$=E\left[\bar{Y}_{..}(1+e_3)\frac{\bar{x}'_{..}}{\bar{x}'_{..}(1+e_4)}-\bar{Y}_{..}\right]^2$$

$$=E\left[\bar{Y}_{..}\ (1+e_3-e_4+\ldots)-\bar{Y}_{..}\right]^2$$

Expanding right hand side (RHS) of the above expression and then neglecting the terms involving powers of e_3 and e_4 greater than two, we get the MSE of T_3 up to the first order of approximation as

$$MSE(T_3)=\frac{1}{nN}\sum_{i=1}^{N}\left(\frac{1}{m'}-\frac{1}{M}\right)S_{Yi}^2+\frac{1}{nN}\left(\frac{1}{m}-\frac{1}{m'}\right)\sum_{i=1}^{N}\left(S_{Yi}^2+R_{XY}^2S_{Xi}^2-2R_{XY}\rho_{XYi}S_{Xi}S_{Yi}\right)$$

$$\left(\frac{1}{n}-\frac{1}{N}\right)\left(S_{bY}^2+R_{XY}^2S_{bX}^2-2R_{XY}S_{bXY}\right)+\frac{1}{nN}\sum_{i=1}^{N}\left(\frac{k_i-1}{m}\right)W_{i2}S_{Yi2}^2 \qquad (14)$$

where $R_{XY}=\overline{Y}_{..}/\overline{X}_{..}$

Now, the MSE of T_4 is represented as

$$MSE(T_4)=E\left[T_4-\overline{Y}_{..}\right]^2$$

$$=E\left[\overline{Y}_{..}(1+e_3)\frac{\overline{x}'_{..}(1+e_4)}{\overline{x}'_{..}}-\overline{Y}_{..}\right]^2$$

$$=E\left[\overline{Y}_{..}(1+e_3+e_4+\ldots)-\overline{Y}_{..}\right]^2$$

Now, we get the MSE of T_4 up to the first order of approximation on expanding RHS of the above expression and then neglecting the terms having powers of e_3 and e_4 higher than two as

$$MSE(T_4)=\frac{1}{nN}\sum_{i=1}^{N}\left(\frac{1}{m'}-\frac{1}{M}\right)S_{Yi}^2+\frac{1}{nN}\left(\frac{1}{m}-\frac{1}{m'}\right)\sum_{i=1}^{N}\left(S_{Yi}^2+R_{XYi}^2S_{Xi}^2+2R_{XY}\rho_{XYi}S_{Xi}S_{Yi}\right)+$$

$$\left(\frac{1}{n}-\frac{1}{N}\right)\left(S_{bY}^2+R_{XY}^2S_{bX}^2+2R_{XY}S_{bXY}\right)+\frac{1}{nN}\sum_{i=1}^{N}\left(\frac{k_i-1}{m}\right)W_{i2}S_{Yi2}^2 \qquad (15)$$

3. Empirical Study

In order to appreciate the applications of the results obtained in this paper and to examine the behavior of the estimators, we have illustrated the theoretical outcomes with some empirical data. We have used the fictitious data considered by Chaudhary and Singh (2013). In this data set, 20 random numbers were selected from random number table [Rao et al. (1966)] in bunches of 25 clusters. The four-digit random numbers were transformed into two-digit numbers by inserting the decimal after two digits so as to reduce the magnitudes of the numbers. Hence, a population of 25 clusters of size 20 each was constituted by assuming the random numbers, so selected, as study variable. In order to generate corresponding values of the auxiliary variable, the same process was again repeated. Thus, we have N = 25, M = 20, NM = 500. To illustrate the theoretical results, we assume m =5 and m'=10 and n = 10.

Table 1 represents the 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.)	$\bar{Y}_{i.}$	$\bar{X}_{i.}$	S_{Yi}^2	S_{Xi}^2	S_{XYi}	R_{XYi}	ρ_{XYi}
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

Here, we assume $S_{Yi2}^2 = \frac{4}{5}\left(S_{Yi}^2\right)$, $S_{Xi2}^2 = \frac{4}{5}\left(S_{Xi}^2\right)$ and $S_{XYi2} = \frac{4}{5}\left(S_{XYi}\right)$.

Further, we get $\bar{Y}.. = 49.8470$, $\bar{X}.. = 50.8752$, $S_{bY}^2 = 42.3954$, $S_{bX}^2 = 55.1814$ $S_{bXY} = 44.8808$ and $R_{XY} = 0.9798$.

Table 2 shows the variance and MSE of the estimators $\bar{y}_{nm}^*$, T_1, T_2, T_3 and T_4 for the different choices of k_i at non-response rate $W_{i2} = 0.2$ for all i .

Table 2: Var./MSE of $\bar{y}_{nm}^*$, T_1, T_2, T_3 and T_4 for

W_{i2}	k_i	$Var./MSE$				
		$\bar{y}_{nm}^*$	T_1	T_2	T_3	T_4
0.2	2	18.79 (100)	11.17 (168.24)	42.43 (44.29)	9.07 (207.13)	50.66 (37.09)
	3	21.65 (100)	14.03 (154.34)	45.29 (47.80)	11.93 (181.47)	53.52 (40.45)
	4	24.51 (100)	16.88 (145.14)	48.14 (50.90)	14.79 (165.73)	56.37 (43.46)

[Figures in parentheses represent the percentage relative efficiency (PRE) of the proposed estimators with respect to the usual mean estimator $\bar{y}_{nm}^*$]

4. Concluding Remarks

We have proposed some separate and combined-type ratio and product estimators for estimating the population mean in two-stage sampling using the information on an auxiliary variable under the situation in which the information about the population mean of auxiliary variable is not available and non-response is observed on study variable only. The expressions for the biases and mean square errors of the proposed estimators up to the first order of approximation have been derived. To support the theoretical results, an empirical study has also been carried out. From Table

2, it is revealed that the estimators T_1 and T_3 provide better estimates than the usual mean estimator $\bar{y}^*_{nm}$ while the estimators T_2 and T_4 do not provide better estimates than $\bar{y}^*_{nm}$. Though, the estimators T_2 and T_4 would perform well than $\bar{y}^*_{nm}$ if the study and auxiliary variables are negatively correlated. It is also revealed that the MSE of the proposed estimators increases with the increase in sub sampling rate k_i. The results are intuitively anticipated.

References

1. Bellhouse, D. R. and Rao, J. N. K. (1986). On the efficiency of prediction estimators in two-stage sampling, *Journal of Statistical Planning and Inference,* 13, 269-281.
2. Chaudhary, M. K. and Singh, V. K. (2013). Estimating the population mean in two-stage sampling with equal size clusters under non-response using auxiliary characteristic, *Mathematical Journal of Interdisciplinary Sciences*, 2(1), 43-55.
3. 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.
4. Khare, B. B. (1992). Non-response and the estimation of population mean in its presence in sample surveys, in *Seminar Proceedings, 12th Orientation Course (Science Stream), Academic staff College, Banaras Hindu University*, Varanasi, India.
5. Rao, C. R., Mitra, S. K. and Mathai, A. (1966). Formulae and tables for statistical work, *Statistical Publishing Society,* Calcutta, India.
6. Rao, P. S. R. S. (1986). Ratio estimators with sub sampling the non-respondents, *Survey Methodology*, 12 (2), 217-230.
7. Rao, P. S. R. S. (1987). Ratio and regression estimators with sub sampling the non-respondents, *paper presented at a special contributed session of the International Association Meetings*, September 2-16, Tokyo, Japan.
8. Rao, P. S. R. S. (1990). Regression estimators with sub sampling of non-respondents, in *Data Quality Control : Theory and Pragmatics*, Liepins, E. G. and Uppuluri, V. R. R. (eds.), Marcel Dekker, New York, 191-208.
9. 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.
10. Srivastava, S. (1993). Some problems on the estimation of population mean using auxiliary character in presence of non-response in sample surveys, *Unpublished Ph. D. Thesis submitted to Banaras Hindu University,* Varanasi, India.