STATISTICA, anno LXXV, n. 4, 2015

USE OF DOUBLE SAMPLING SCHEME IN ESTIMATING THE MEAN OF STRATIFIED POPULATION UNDER NON-RESPONSE

Manoj K. Chaudhary [1]
Department of Statistics, Banaras Hindu University, Varanasi, India
Amit Kumar
Department of Statistics, Banaras Hindu University, Varanasi, India

1. Introduction

Auxiliary information can be used to improve the efficiency of the estimator of a particular population parameter. The effectiveness of the estimation procedure with the use of auxiliary information closely depends upon the method in which the estimator has been proposed, *i.e.*, the form in which the functions of auxiliary information have been considered. There is a lot of works in which the auxiliary information are used to enhance the precision of the estimator. Upadhyaya and Singh (1999) have suggested a class of estimators for estimating the population mean in simple random sampling. Kadilar and Cingi (2003) and Shabbir and Gupta (2005) expanded these estimators for estimating the population mean of stratified population. Singh et al. (2012) have proposed a general family of estimators for estimating the population mean in systematic sampling.

We know that the problem of non-response is inherent in the population of every survey. The non-response error is not so important if the characteristics of the non-responding units are similar to those of the responding units. But, it is noticed that such similarity of characteristics between two types of units (responding and non-responding units) is not always obtained in practice. Hansen and Hurwitz (1946) introduced a technique of sub-sampling of non-respondents in order to adjust the non-response in mail surveys. Khoshnevisan et al. (2007) have proposed a general family of estimators of population mean using known coefficients of some population parameters in simple random sampling. Chaudhary *et al.* (2009) have proposed a combined-type family of estimators with the use of an auxiliary variable for mean of a stratified population in the presence of non-response adopting Khoshnevisan et al. (2007). Recently, Chaudhary and Singh (2013) have proposed factor type families of estimators of population mean in two-stage sampling with equal size clusters under non-response.

[1] Corresponding Author e-mail: ritamanoj15@gmail.com

It is well known fact that when the parametric values of auxiliary variable are not known, one can utilize the double sampling (or two-phase sampling scheme) in improving the estimation procedure. Two-phase sampling is very effective in terms of cost as well as applications. This sampling scheme is used to obtain the information about auxiliary variable inexpensively from a larger sample at first phase and comparatively small sample at the second phase. Okafor and Lee (2000) have discussed the method of double sampling scheme for estimating the family expenditure in a household survey under non-response. Khare and Sinha (2004) have suggested the estimators for population ratio using two-phase sampling scheme in the presence of non-response.

In the light of above context, we have suggested a family of combined-type estimators of population mean in stratified random sampling using two-phase sampling scheme under non-response whenever the population mean of auxiliary variable is unknown. The properties of the family along with its optimum property have been discussed.

2. Notations, Sampling Strategy and Estimation Procedure

y_{ij}: Observation on the j^{th} unit in the i^{th}stratum under study variable.

$$(i = 1, 2, ..., k; j = 1, 2, ..., N_i) .$$

x_{ij} : Observation on the j^{th} unit in the i^{th}stratum under auxiliary variable.

$$(i = 1, 2, ..., k; j = 1, 2, ..., N_i) .$$

N_{i1}: Population size of the response group in the i^{th}stratum.

N_{i2}: Population size of non-response group in the i^{th}stratum.

$\overline{Y} = \sum_{i=1}^{k} p_i \overline{Y}_i$: Population mean under study variable.

$\overline{X} = \sum_{i=1}^{k} p_i \overline{X}_i$: Population mean under auxiliary variable.

$\overline{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$: Population mean of the i^{th}stratum under study variable.

$\overline{X}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$: Population mean of the i^{th}stratum under auxiliary variable.

$\overline{Y}_{i2} = \frac{1}{N_{i2}} \sum_{j}^{N_{i2}} y_{ij}$: Population mean of the non-response group in the i^{th} stratum under study variable.

$\overline{x}_i = \frac{1}{n_i} \sum_{j}^{n_i} x_{ij}$: Sample mean of the i^{th}stratum under auxiliary variable.

$\overline{x}_i^{'} = \frac{1}{n_i^{'}} \sum_{j}^{n_i^{'}} x_{ij}$: Mean based on first phase sample in the i^{th}stratum under auxiliary variable.

$\overline{y}_{ni1} = \frac{1}{n_{i1}} \sum_{j}^{n_{i1}} y_{ij}$: Sample mean of the response group in the i^{th} stratum under study variable.

$\overline{y}_{hi2} = \frac{1}{h_{i2}} \sum_{j}^{h_{i2}} y_{ij}$: Mean based on h_{i2} non-respondent units in the i^{th} stratum under study variable.

$S_{Yi}^2 = \frac{1}{N_i-1} \sum_{j=1}^{N_i} \left(y_{ij} - \overline{Y}_i\right)^2$: Population mean square of the i^{th}stratum under study variable.

$S_{Xi}^2 = \frac{1}{N_i-1} \sum_{j=1}^{N_i} \left(x_{ij} - \overline{X}_i\right)^2$: Population mean square of the i^{th}stratum under auxiliary variable.

$S_{Yi2}^2 = \frac{1}{N_{i2}-1} \sum_{j}^{N_{i2}} \left(y_{ij} - \overline{Y}_{i2}\right)^2$: Population mean square of the non- response group in the i^{th}stratum under study variable.

$W_{i1} = \frac{N_{i1}}{N_i}$: Response rate in the i^{th}stratum.

$W_{i2} = \frac{N_{i2}}{N_i}$: Non-response rate in the i^{th}stratum.

Let us suppose that a population of size N is divided into k strata. Let the size of the i^{th} stratum be N_i $(i = 1, 2, ..., k)$ such that$\sum_{i=1}^{k} N_i = N$. Let a sample of size n be selected from the entire population in such a way that n_i units are selected from the i^{th} stratum and we have$\sum_{i=1}^{k} n_i = n$. Let Ybe the study variable with population mean$\overline{Y} = \sum_{i=1}^{k} p_i \overline{Y}_i$($\overline{Y}_i$ being the mean of the i^{th} stratum based on N_i units and $p_i = \frac{N_i}{N}$) and we assume that the non-response is observed on study variable. It is observed that there are n_{i1} respondent units and n_{i2} non-respondent units in n_i units. Using Hansen and Hurwitz (1946) technique of sub-sampling of non-respondents, we select a sub-sample of h_{i2} units out of n_{i2} non-respondent units such that$n_{i2} = L_i h_{i2}$, $L_i \geq 1$and collect the information from all h_{i2} units. Thus, an unbiased estimator of population mean $\overline{Y}$ is given by

$$\overline{y}_{st}^* = \sum_{i=1}^{k} p_i \overline{y}_i^* \tag{1}$$

where $\overline{y}_i^* = \frac{n_{i1}\overline{y}_{ni1} + n_{i2}\overline{y}_{hi2}}{n_i}$,
$\overline{y}_{ni1}$ and $\overline{y}_{hi2}$ are the means based on n_{i1} respondent units and h_{i2}sub-sampled non-respondent units respectively. The variance of $\overline{y}_{st}^*$ is given as

$$V\left(\overline{y}_{st}^*\right) = \sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2 S_{Yi}^2 + \sum_{i=1}^{k} \frac{(L_i - 1)}{n_i} W_{i2} p_i^2 S_{Yi2}^2 \tag{2}$$

where $S_{Yi}^2$and S_{Yi2}^2 are the population mean squares of the entire group and non-response group respectively in i^{th} stratum for study variable. W_{i2} is the population non-response rate in the i^{th} stratum.

In order to improve the efficiency of an estimator, one can utilize the auxiliary information at the estimation stage. In this sequence, Chaudhary et al. (2009)

have proposed a family of combined-type estimators of population mean using an auxiliary variable X in stratified random sampling under the condition that the non-response is observed on study variable and auxiliary variable is free from non-response, given as

$$T_C = \overline{y}^*_{st}\left[\frac{a\overline{X}+b}{\alpha\left(a\overline{x}_{st}+b\right)+(1-\alpha)\left(a\overline{X}+b\right)}\right]^g \tag{3}$$

where $\overline{x}_{st}=\sum_{i=1}^{k}p_i\overline{x}_i$, $\overline{X}=\sum_{i=1}^{k}p_i\overline{X}_i$ (population mean of auxiliary variable), $a\neq 0$ and $b$are either real numbers or functions of known parameters of auxiliary variable. α and $g$are the constants and to be determined. $\overline{x}_i$ and $\overline{X}_i$ are respectively the mean based on n_i units and mean based on N_i units in the i^{th} stratum for auxiliary variable. The bias and mean square error (MSE) of T_C up to the first order of approximation are respectively given by

$$B\left(T_C\right)=\frac{1}{\bar{Y}}\sum_{i=1}^{k}f_ip_i^2\left[\frac{g(g+1)}{2}\alpha^2\lambda^2R^2S_{Xi}^2-\alpha\lambda gR\rho_iS_{Xi}S_{Yi}\right] \tag{4}$$

and

$$MSE\left(T_C\right)=\sum_{i=1}^{k}f_ip_i^2\left[S_{Yi}^2+\alpha^2\lambda^2g^2R^2S_{Xi}^2-2\alpha\lambda gR\rho_iS_{Xi}S_{Yi}\right]+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2S_{Yi2}^2 \tag{5}$$

where $f_i=\left(\frac{1}{n_i}-\frac{1}{N_i}\right)$, $\lambda=\frac{a\overline{X}}{a\overline{X}+b}$, $R=\frac{\overline{Y}}{\overline{X}}$, S_{Xi}^2 is the population mean square of auxiliary variable for the i^{th} stratum and ρ_i is the population correlation coefficient between $Y$and X in the i^{th} stratum.

3. Proposed Family of Estimators

If the population mean of auxiliary variable $\overline{X}$ is known, one can easily use the family of estimators shown in equation (3) for estimating the population mean of study variable$\overline{Y}$. But if the population mean $\overline{X}$ is unknown, it is not easy to adopt the present form of the considered family of estimators. Thus the double sampling scheme (or two-phase sampling scheme) can be utilized to estimate the population mean$\overline{Y}$. Adopting double sampling scheme first the estimate of $\overline{X}$may be generated from a large first phase sample of size $n_i^{'}$ drawn from the N_i units by simple random sampling without replacement (SRSWOR) scheme for the i^{th} stratum. Secondly, a smaller second phase sample of size n_i is drawn from $n_i^{'}$ units by SRSWOR. It is noted that the non-response is observed on study variable at second phase sample. Out ofn_i units, let n_{i1}units respond and n_{i2} units do not respond at the second phase. Applying Hansen and Hurwitz (1946) technique of sub-sampling of non-respondents, a sub-sample of $h_{i2}\,(=n_{i2}/L_i, L_i\geq 1)$ units is selected from the sample of n_{i2} non-respondents and the information is collected from all of them.

Let us assume that the full information is supplied on the $n_i^{'}$ units at the first phase for the auxiliary variable $X$and the study variable goes through the non-response. Thus the usual combined ratio and product estimators of population mean $\overline{Y}$using two-phase sampling scheme in stratified random sampling under non-response are respectively given by

$$T_1^* = \frac{\overline{y}_{st}^*}{\overline{x}_{st}} \overline{x}_{st}^{'} \tag{6}$$

and

$$T_2^* = \frac{\overline{y}_{st}^*}{\overline{x}_{st}^{'}} \overline{x}_{st} \tag{7}$$

where $\overline{x}_{st}^{'} = \sum_{i=1}^{k} p_i \overline{x}_i^{'}$ and $\overline{x}_i^{'}$ is the mean based on $n_i^{'}$ units for auxiliary variable.

The mean square errors (MSEs) of the estimators T_1^* and T_2^* up to the first order of approximation are respectively given by

$$MSE\left(T_1^*\right) = \sum_{i=1}^{k} f_i^{'} p_i^2 S_{Yi}^2 + \sum_{i=1}^{k} f_i^* p_i^2 \left(S_{Yi}^2 + R^2 S_{Xi}^2 - 2R\rho_i S_{Xi} S_{Yi}\right) + \sum_{i=1}^{k} \frac{(L_i - 1)}{n_i} W_{i2} p_i^2 S_{Yi2}^2 \tag{8}$$

and

$$MSE\left(T_2^*\right) = \sum_{i=1}^{k} f_i^{'} p_i^2 S_{Yi}^2 + \sum_{i=1}^{k} f_i^* p_i^2 \left(S_{Yi}^2 + R^2 S_{Xi}^2 + 2R\rho_i S_{Xi} S_{Yi}\right) + \sum_{i=1}^{k} \frac{(L_i - 1)}{n_i} W_{i2} p_i^2 S_{Yi2}^2 \tag{9}$$

where $f_i^{'} = \left(\frac{1}{n_i^{'}} - \frac{1}{N_i}\right)$ and $f_i^* = \left(\frac{1}{n_i} - \frac{1}{n_i^{'}}\right)$.

In the presence of above circumstances, a family of combined-type estimators of mean $\overline{Y}$ of stratified population using two-phase sampling in the presence of non-response is given by

$$T_C^{'} = \overline{y}_{st}^* \left[\frac{a\overline{x}_{st}^{'} + b}{\alpha\left(a\overline{x}_{st} + b\right) + (1-\alpha)\left(a\overline{x}_{st}^{'} + b\right)}\right]^g \tag{10}$$

In order to obtain the bias and MSE of $T_C^{'}$, we use large sample approximation. Let us assume that

$$\overline{y}_{st}^* = \overline{Y}\left(1 + e_0\right), \overline{x}_{st} = \overline{X}\left(1 + e_1\right), \overline{x}_{st}^{'} = \overline{X}\left(1 + e_1^{'}\right).$$

Form the above, we have

$$E\left(e_0\right) = E\left(e_1\right) = E\left(e_1^{'}\right) = 0,$$

$$E\left(e_0^2\right)=\frac{1}{\overline{Y}^2}\sum_{i=1}^{k}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2S_{Yi}^2+\frac{(L_i-1)}{n_i}W_{i2}p_i^2S_{Yi2}^2\right],$$

$$E\left(e_1^2\right)=\frac{1}{\overline{X}^2}\sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2S_{Xi}^2,$$

$$E\left(e_1^{'2}\right)=\frac{1}{\overline{X}^2}\sum_{i=1}^{k}\left(\frac{1}{n_i^{'}}-\frac{1}{N_i}\right)p_i^2S_{Xi}^2, E\left(e_0e_1\right)=\frac{1}{\overline{XY}}\sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2\rho_iS_{Yi}S_{Xi},$$

$$E\left(e_0e_1^{'}\right)=\frac{1}{\overline{XY}}\sum_{i=1}^{k}\left(\frac{1}{n_i^{'}}-\frac{1}{N_i}\right)p_i^2\rho_iS_{Yi}S_{Xi}, E\left(e_1e_1^{'}\right)=\frac{1}{\overline{X}^2}\sum_{i=1}^{k}\left(\frac{1}{n_i^{'}}-\frac{1}{N_i}\right)p_i^2S_{Xi}^2.$$

Putting the above assumptions into equation (10), $T_C^{'}$ can be expressed as

$$T_C^{'}=\overline{Y}\left(1+e_0\right)\left(1+\lambda e_1^{'}\right)^g\left[1+\lambda\left\{\alpha e_1+(1-\alpha)e_1^{'}\right\}\right]^{-g} \tag{11}$$

On expanding equation (11) and neglecting the terms ofe_0, $e_1$and $e_1^{'}$having power greater than two, we get

$$\begin{aligned}T_C^{'}-\overline{Y}=\overline{Y}\Big[&-g\lambda\left\{\alpha e_1+(1-\alpha)e_1^{'}\right\}+\frac{g(g+1)}{2}\lambda^2\Big\{\alpha^2e_1^2+(1-e)^2e_1^{'2}\\&+2\alpha(1-\alpha)e_1e_1^{'}\Big\}+g\lambda e_1^{'}-g^2\lambda^2e_1^{'}\left\{\alpha e_1+(1-\alpha)e_1^{'}\right\}\\&+\frac{g(g-1)}{2}\lambda^2e_1^{'2}+e_0-g\lambda e_0\left\{\alpha e_1+(1-\alpha)e_1^{'}\right\}+g\lambda e_1^{'}e_0\Big]\end{aligned} \tag{12}$$

Taking expectation both the sides of equation (12), we get bias of $T_C^{'}$ up to the first order of approximation as

$$\begin{aligned}B\left(T_C^{'}\right)=\frac{1}{\overline{Y}}g\lambda\sum_{i=1}^{k}p_i^2\Big[&f_i\left\{\frac{(g+1)}{2}\lambda\alpha^2R^2S_{Xi}^2-2R\rho_iS_{Yi}S_{Xi}\right\}\\&+f_i^{'}\Big\{\Big(\frac{(g+1)}{2}\lambda(1-\alpha)^2R^2+(g+1)\lambda\alpha(1-\alpha)R^2-g\lambda R^2\\&+\frac{(g-1)}{2}\lambda R^2\Big)S_{Xi}^2+R\alpha\rho_iS_{Yi}S_{Xi}\Big\}\Big]\end{aligned} \tag{13}$$

Squaring both the sides of the equation (12) and then taking the expectation by neglecting the terms ofe_0, $e_1$and $e_1^{'}$having power greater than two, we get MSE of $T_C^{'}$ up to the first order of approximation as

$$\begin{aligned}MSE\left(T_C^{'}\right)=&g^2\lambda^2R^2\alpha^2\sum_{i=1}^{k}f_i^*p_i^2S_{Xi}^2-2g\lambda R\alpha\sum_{i}f_i^*p_i^2\rho_iS_{Xi}S_{Yi}\\&+\sum_{i=1}^{k}f_ip_i^2S_{Yi}^2+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2S_{Yi2}^2\end{aligned} \tag{14}$$

or

$$MSE\left(T_C^{'}\right)=\sum_{i=1}^{k}f_i^{'}p_i^2S_{Yi}^2+\sum_{i=1}^{k}f_i^{*}p_i^2\left(S_{Yi}^2+g^2\lambda^2R^2\alpha^2S_{Xi}^2\right.$$
$$\left.-2g\lambda R\alpha\rho_iS_{Xi}S_{Yi}\right)+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2S_{Yi2}^2 \tag{15}$$

3.1 Optimum Choice of α

In this section, we choose the optimum value of α for which the MSE of the proposed family would remain its minimum. On differentiating $MSE\left(T_C^{'}\right)$ with respect to α and equating the derivative to zero, we get

$$\frac{\partial MSE\left(T_C^{'}\right)}{\partial\alpha}=2g^2\lambda^2R^2\alpha\sum_{i=1}^{k}f_i^{*}p_i^2S_{Xi}^2-2g\lambda R\sum_{i}^{k}f_i^{*}p_i^2\rho_iS_{Xi}S_{Yi}=0 \tag{16}$$

$$\Rightarrow\alpha_{opt}=\frac{\sum_i^k f_i^{*}p_i^2\rho_iS_{Xi}S_{Yi}}{g\lambda R\sum_{i=1}^{k}f_i^{*}p_i^2S_{Xi}^2} \tag{17}$$

On substituting the value of α_{opt} from equation (17) into equation (14) or equation (15), we get the minimum MSE of $T_C^{'}$.

3.2 Cost of the Survey and Optimumn_i,$n_i^{'}$, L_i

Let $c_i^{'}$ be the cost per unit associated with the first phase sample of size $n_i^{'}$ and c_{i0} be the unit cost of first attempt on study variable with second phase sample of sizen_i. Let $c_{i1}$and c_{i2} be respectively the cost per unit of enumerating the n_{i1}respondent units and h_{i2} non-respondent units. Then the total cost for the i^{th} stratum is given by

$$C_i=c_i^{'}n_i^{'}+c_{i0}n_i+c_{i1}n_{i1}+c_{i2}h_{i2}\quad\forall i=1,2,...,k.$$

Now, we obtain the expected cost per stratum as

$$E\left(C_i\right)=c_i^{'}n_i^{'}+n_i\left(c_{i0}+c_{i1}W_{i1}+c_{i2}\frac{W_{i2}}{L_i}\right).$$

Thus the total cost over all the strata is represented as

$$C_0=\sum_{i=1}^{k}E\left(C_i\right)$$

$$=\sum_{i=1}^{k}\left[c_i^{'}n_i^{'}+n_i\left(c_{i0}+c_{i1}W_{i1}+c_{i2}\frac{W_{i2}}{L_i}\right)\right]. \tag{18}$$

Let us consider the Lagrange function

$$\phi = MSE\left(T_C^{'}\right) + \mu C_0 \tag{19}$$

where μ is Lagrange's multiplier.

In order to get the optimum values ofn_i, $n_i^{'}$andL_i, we differentiate ϕ with respect ton_i, $n_i^{'}$and L_i respectively and equate the derivatives to zero. Thus, we have

$$\frac{\partial \phi}{\partial n_i} = -\frac{p_i^2}{n_i^2}\left(S_{Yi}^2 + g^2\lambda^2R^2\alpha^2S_{Xi}^2 - 2g\lambda R\alpha\rho_i S_{Xi}S_{Yi}\right) - \frac{(L_i-1)}{n_i^2}W_{i2}p_i^2S_{Yi2}^2$$

$$+\mu\left(c_{i0} + c_{i1}W_{i1} + c_{i2}\frac{W_{i2}}{L_i}\right) = 0 \tag{20}$$

$$\frac{\partial \phi}{\partial n_i^{'}} = \frac{p_i^2}{n_i^{'2}}\left(g^2\lambda^2R^2\alpha^2S_{Xi}^2 - 2g\lambda R\alpha\rho_i S_{Xi}S_{Yi}\right) + \mu c_i^{'} = 0 \tag{21}$$

$$\frac{\partial \phi}{\partial L_i} = \frac{p_i^2}{n_i}W_{i2}S_{Yi2}^2 - \mu n_i c_{i2}\frac{W_{i2}}{L_i^2} = 0 \tag{22}$$

From equations (20), (21) and (22), we respectively get

$$n_i = \frac{p_i\sqrt{S_{Yi}^2 + g^2\lambda^2R^2\alpha^2S_{Xi}^2 - 2g\lambda R\alpha\rho_i S_{Xi}S_{Yi} + (L_i-1)W_{i2}S_{Yi2}^2}}{\sqrt{\mu\left(c_{i0} + c_{i1}W_{i1} + c_{i2}\frac{W_{i2}}{L_i}\right)}} \tag{23}$$

$$n_i^{'} = \frac{p_i\sqrt{2g\lambda R\alpha\rho_i S_{Xi}S_{Yi} - g^2\lambda^2R^2\alpha^2S_{Xi}^2}}{\sqrt{\mu c_i^{'}}} \tag{24}$$

and

$$\sqrt{\mu} = \frac{p_iL_iS_{Yi2}}{n_i\sqrt{c_{i2}}} \tag{25}$$

Putting the value of $\sqrt{\mu}$ from equation (25) into equation (23), we get

$$L_{i(opt)} = \frac{\sqrt{c_{i2}}B_i}{S_{Yi2}A_i} \tag{26}$$

where $A_i = \sqrt{c_{i0} + c_{i1}W_{i1}}$ and

$$B_i = \sqrt{S_{Yi}^2 + g^2\lambda^2R^2\alpha^2S_{Xi}^2 - 2g\lambda R\alpha\rho_i S_{Xi}S_{Yi} - W_{i2}S_{Yi2}^2}.$$

On substituting the value of $L_{i(opt)}$ from equation (26) into equation (23), we can express n_i as

$$n_i = \frac{p_i\sqrt{B_i^2 + \frac{\sqrt{c_{i2}}B_iW_{i2}S_{Yi2}}{A_i}}}{\sqrt{\mu}\sqrt{A_i^2 + \frac{\sqrt{c_{i2}}A_iW_{i2}S_{Yi2}}{B_i}}} \tag{27}$$

In obtaining the value of $\sqrt{\mu}$ in terms of total costC_0, we put the values of $n_i^{'}$, $L_{i(opt)}$ and n_i from equations (24), (26) and (27) into equation (18). Thus, we have

TABLE 1
Particulars of Data

Stratum No.	N_i	$n_i^{'}$	n_i	$\overline{Y}_i$	$\overline{X}_i$	S_{Yi}^2	S_{Xi}^2	ρ_i	S_{Yi2}^2
1	73	65	26	40.85	39.56	6369.10	6624.44	0.999	618.88
2	70	25	10	27.83	27.57	1051.07	1147.01	0.998	240.91
3	97	48	19	25.79	25.44	2014.97	2205.40	0.999	265.52
4	44	11	5	20.64	20.36	538.47	485.27	0.997	83.69

$$\sqrt{\mu} = \frac{1}{C_0} \sum_{i=1}^{k} \left[\sqrt{c_i^{'}} \sqrt{2g\lambda R\alpha\rho_i S_{Xi} S_{Yi} - g^2\lambda^2 R^2 \alpha^2 S_{Xi}^2} + p_i \left(A_i B_i + \sqrt{c_{i2}} W_{i2} S_{Yi2} \right) \right] \tag{28}$$

Substituting the value of $\sqrt{\mu}$ from equation (28) into equations (27) and (24), we respectively get the optimum values of n_i and $n_i^{'}$

$$n_{i(opt)} = C_0 p_i \sqrt{B_i^2 + \frac{\sqrt{c_{i2}} B_i W_{i2} S_{Yi2}}{A_i}} \times \left(\sqrt{A_i^2 + \frac{\sqrt{c_{i2}} A_i W_{i2} S_{Yi2}}{B_i}} \right.$$
$$\left. \times \sum_{i=1}^{k} \left[\sqrt{c_i^{'}} \sqrt{2g\lambda R\alpha\rho_i S_{Xi} S_{Yi} - g^2\lambda^2 R^2 \alpha^2 S_{Xi}^2} + p_i \left(A_i B_i + \sqrt{c_{i2}} W_{i2} S_{Yi2} \right) \right] \right)^{-1} \tag{29}$$

$$n_{i(opt)}^{'} = C_0 p_i \sqrt{2g\lambda R\alpha\rho_i S_{Xi} S_{Yi} - g^2\lambda^2 R^2 \alpha^2 S_{Xi}^2} \left(\sqrt{c_i^{'}} \right.$$
$$\left. \times \sum_{i=1}^{k} \left[\sqrt{c_i^{'}} \sqrt{2g\lambda R\alpha\rho_i S_{Xi} S_{Yi} - g^2\lambda^2 R^2 \alpha^2 S_{Xi}^2} + p_i \left(A_i B_i + \sqrt{c_{i2}} W_{i2} S_{Yi2} \right) \right] \right)^{-1} \tag{30}$$

4. Empirical Study

The theoretical study of the proposed family can easily be comprehended by an empirical analysis. To support the theoretical results, we have used the data considered by Chaudhary et al. (2012). There are 284 municipalities divided into four strata having respective sizes 73, 70, 97, and 44. The data relate to the population size (in thousands) in the year 1985 as study variable and the population size (in thousands) in the year 1975 as auxiliary variable. Particulars are given in Table 1.

Table 2 represents the MSE and percent relative efficiency (PRE) of T_1^*, T_2^* and $T_C^{'}$ (at $\alpha_{(opt)}$, $a = 1$, $b = 1$ and $g = 1$)with respect to $\overline{y}_{st}^*$ for the different choices of W_{i2} and L_i.

TABLE 2
MSE and PRE of T_1^*, T_2^* and $T_C^{'}$ with respect to $\overline{y}_{st}^*$

W_{i2}	L_i	$V(\overline{y}_{st}^*)$	$MSE(T_1^*)$	$MSE(T_2^*)$	$MSE\left(T_C^{'}\right)$	$PRE(T_1^*)$	$PRE(T_2^*)$	$PRE\left(T_C^{'}\right)$
	2	34.42	23.62	375.17	6.28	145.72	9.17	548.09
0.1	2.5	34.67	24.01	375.56	6.54	144.40	9.23	530.12
	3	34.92	24.41	375.96	6.79	143.06	9.29	514.29
	3.5	35.18	24.8	376.35	7.04	141.85	9.35	499.72
	2	34.92	24.41	375.96	6.79	143.06	9.29	514.29
0.2	2.5	35.43	25.2	376.75	7.3	140.60	9.40	485.34
	3	35.94	25.99	377.54	7.8	138.28	9.52	460.77
	3.5	36.44	26.78	378.33	8.31	136.07	9.63	438.51
	2	35.43	25.2	376.75	7.3	140.60	9.40	485.34
0.3	2.5	36.19	26.38	377.93	8.06	137.19	9.58	449.01
	3	36.95	27.57	379.12	8.82	134.02	9.75	418.93
	3.5	37.71	28.75	380.3	9.57	131.17	9.92	394.04
	2	35.94	25.99	377.54	7.8	138.28	9.52	460.77
0.4	2.5	36.95	27.57	379.12	8.82	134.02	9.75	418.93
	3	37.96	29.14	380.69	9.83	130.27	9.97	386.16
	3.5	38.97	30.72	382.27	10.84	126.86	10.19	359.50

From the Table 2, it is revealed that the optimum estimator of the proposed family certainly provides the better estimates as compared to $\overline{y}^*_{st}$, T^*_1 andT^*_2. It is obvious that the increment in population non-response rate W_{i2} or inverse sampling rate L_i would cause a further increment in the MSE of the estimators. Thus, from the above table (Table 2), it is also revealed that the efficiency of the estimators decreases with an increase in non-response rate W_{i2} as well as with an increase in inverse sampling rate L_i.

5. CONCLUSION

The use of auxiliary information certainly improves the precision of the estimator at the estimation stage. The situations in which the parametric values of auxiliary variable(s) are not known, one can use its estimates by using double sampling scheme in estimating the parameters of study variable. To cope up the situations, we have proposed a family of combined-type estimators of population mean in stratified random sampling using double sampling scheme under non-response. The optimum combined-type estimator of the family has been obtained and comparison of it with the usual mean, usual combined ratio and usual combined product estimators has also been made. The optimum values of first phase sample size$n^{'}_{i}$, second phase sample size n_i and inverse sampling rate L_i for the different strata through the proposed family, have been determined under the cost survey. From Table 2, it is observed that the optimum combined-type estimator of the proposed family provides better estimates than the usual mean, usual combined ratio and usual combined product estimators. It is also observed that precision of the optimum combined-type estimator as well as usual mean, usual combined ratio and usual product estimators decreases with increase in non-response rate and inverse sampling rate. The results are intuitively expected.

Acknowledgements

The authors are thankful to the learned referees for their constructive suggestions regarding improvement of the present paper.

References

M. K. Chaudhary, R. Singh, R. K. Shukla, M. Kumar (2009). *A family of estimators for estimating population mean in stratified sampling under non-response*. Pakistan Journal of Statistics and Operation Research, 5, no. 1, pp. 47–54.

M. K. Chaudhary, V. K. Singh (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, no. 1, pp. 43–55.

M. K. Chaudhary, V. K. Singh, R. K. Shukla (2012). *Combined-type family of estimators of population mean in stratified random sampling under non-*

response. Journal of Reliability and Statistical Studies (JRSS), 5, no. 2, pp. 133–142.

M. H. Hansen, W. N. Hurwitz (1946). *The problem of non-response in sample surveys*. Journal of The American Statistical Association, , no. 41, pp. 517–529.

C. Kadilar, H. Cingi (2003). *Ratio estimators in stratified random sampling*. Biometrical journal, 2, no. 45, pp. 218–225.

B. B. Khare, R. R. Sinha (2004). *Estimation of finite population ratio using two-phase sampling scheme in the presence of non-response*. Aligarh J. Stat., 24, 43-56., 24, pp. 43–56.

M. Khoshnevisan, R. Singh, P. Chauhan, N. Sawanand, F. Smarandache (2007). *A general family of estimators for estimating population mean using known value of some population parameter(s)*. Far East Journal of Theoretical Statistics, 22, pp. 181–191.

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

J. Shabbir, S. Gupta (2005). *Improved ratio estimators in stratified sampling*. American Journal of Mathematical and Management Sciences, 25, no. 3-4, pp. 293–311.

R. Singh, S. Malik, M. K. Chaudhary, H. K. Verma, A. A. Adewara (2012). *A general family of ratio-type estimators in systematic sampling*. Journal of Reliability and Statistical Studies, 5, no. 1, pp. 73–82.

L. N. Upadhyaya, H. P. Singh (1999). *Use of transformed auxiliary variable in estimating the finite population mean*. Biometrical Journal, 41, no. 5, pp. 627–636.

Summary

The present paper focuses on the use of double sampling scheme in stratified random sampling for estimating the population mean in the presence of non-response. Motivated by Khoshnevisan et al. (2007), we have proposed a family of combined-type estimators of population mean utilizing the information on an auxiliary variable with the use of double sampling scheme under non-response. The optimum property of the proposed family has been discussed. An empirical study has also been carried out in the support of theoretical results.

Keywords: Double sampling scheme; stratified random sampling; auxiliary variable; population mean; non-response