Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 67(3) 2013 305-317

Two Stage Sampling with Two-phases at the Second Stage of Sampling for Estimation of Finite Population Mean under Random Response Mechanism

U.C. Sud, Kaustav Aditya, Hukum Chandra and Rajender Parsad
Indian Agricultural Statistics Research Institute, New Delhi

Received 25 January 2012; Revised 29 April 2013; Accepted 30 April 2013

SUMMARY

The problem of estimation of finite population mean in the presence of the random response has been considered when the sampling design is two-stage with two phases at the second stage. Three different types of estimators, based on subsampling of nonrespondents, collecting data on the subsample through specialized efforts, are developed. Expressions for the variances of the estimators along with unbiased variance estimators are developed. Optimum values of sample sizes are obtained by considering a suitable cost function. The percentage reduction in the expected cost of the proposed estimators is studied empirically.

Keywords: Cost function, Nonresponse, Random response, Population mean, Subsampling, Two-stage sampling, Percentage reduction in the expected cost.

1. INTRODUCTION

For large or medium scale surveys we are often faced with the scenario that the sampling frame of ultimate stage units is not available and the cost of construction of the frame is very high. Sometimes the population elements are scattered over a wide area resulting in a widely scattered sample. Therefore, not only the cost of enumeration of units in such a sample may be very high, the supervision of field work may also be very difficult. For such situations, two-stage or multi-stage sampling designs are very effective. It is also the case that, in many human surveys, information is not obtained from all the units in surveys. The problem of nonresponse persist even after call backs. The estimates obtained from incomplete data may be biased particularly when the respondents differ from the nonrespondents. Hansen and Hurwitz (1946) proposed a technique for adjusting for nonresponse to address the problem of bias. The technique consists of selecting a subsample of nonrespondents. Through specialized efforts data are collected from the nonrespondents so as to obtain an estimate of nonresponding units in the population. Foradori (1961) studied the subsampling of the nonrespondents technique to estimate the population total in two stages using unequal probability sampling. Srinath (1971) used a different procedure for selecting the subsample of respondents where the subsampling procedure varied according to the nonresponse rates.

Oh and Schereun (1983) attempted to compensate for nonresponse by weighing adjustment. Platek and Gray (1983) used quassi-randomisation technique for estimation in the presence of nonresponse. Kalton and Karsprzyk (1986) tried the imputation technique. Tripathi and Khare (1997) extended the subsampling of nonrespondents approach to multivariate case. Okafor and Lee (2000) extended the approach to double sampling for ratio and regression estimation. Okafor (2001, 2005) further extended the approach in the

Corresponding author: Kaustav Aditya
E-mail address: kaustav@iasri.res.in

context of element sampling and two-stage sampling respectively on two successive occasions. Chhikara and Sud (2009) used the approach for estimation of population and domain totals in the context of item nonresponse. Again, Sud *et al.* (2012) considered the problem of estimation of finite population mean in the presence of nonresponse under two stage sampling design when the response mechanism was assumed to be deterministic. It may be mentioned that the weighting and imputation procedures aim at eliminating the bias caused by nonresponse. However, these procedures are based on certain assumptions on the response mechanism. When these assumptions do not hold good the resulting estimate may be seriously biased. Further, when the nonresponse is confounded *i.e*. the response probability is dependent on the survey character, it becomes difficult to eliminate the bias entirely. Rancourt *et al.* (1994) provided a partial correction for the situation. Hansen and Hurwitz's subsampling approach although costly, is free from any assumptions. When the bias caused by nonresponse is serious this technique is very effective *i.e*. one does not have to go for 100 percent response, which can be very expensive.

In what follows, different estimators of population mean using two-stage sampling designs are developed in Section 2 based on the technique of subsampling the nonrespondents, where nonresponse mechanism is assumed to follow Bernoulli distribution within each selected primary stage units (psus). However, it is assumed that the response mechanism is deterministic at the primary stage unit (psu) level *i.e*. the entire population of psus can be assumed to be divided into responding and nonresponding groups. Also given are expressions for variance of the estimators and unbiased variance estimators. Optimum values of sample sizes are obtained by minimizing the expected cost for a fixed variance. The results are empirically illustrated in Section 5.3.

2. THEORETICAL DEVELOPMENTS

Let the finite population U under consideration consists of N known psus labelled 1 through N. Let the i-th psu comprise M second stage units (ssus). Let y_{ij} be the value of study character pertaining to j-th ssu in the i-th psu, i=1, 2,…, N, j=1, 2,…, M. The objective is to estimate the population mean which is defined as,

$$\bar{Y} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}.$$

Case1. Let n psus be selected by simple random sampling without replacement (srswor) from N and within each selected psu, m ssus are also selected by srswor design. Further, out of m ssus, m_{i1} ssus respond while m_{i2} ssus do not respond, $m_{i1}+m_{i2}= m$. A subsample of size h_{i2} is selected from m_{i2} by srswor and data are collected on the subsampled units through specialized efforts, $m_{i2} = h_{i2}f_{i2}$, i = 1, 2, …, n. Here at the second stage m_{i1} responding and m_{i2} nonresponding units are being generated as a result of m independent Bernoulli trials, one for each element i in m with constant probability θ_i of "success", *i.e.* the response. So, we have $\Pr(i \in m_{i1} \mid m) = \theta_{i|m} = \theta_i$ and $\Pr(i \,\&\, j \in m_{i1}| m) = \theta_{ij|m} = \theta_i^2$.

Theorem 2.1 An unbiased estimator of $\bar{Y}$ is given by

$$\bar{y}_r' = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{h_{i2}}), \tag{2.1}$$

with variance

$$V(\bar{y}_r') = \left(\frac{1}{n}-\frac{1}{N}\right)S_b^2 + \frac{1}{Nn}\sum_{i=1}^{N}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2 + \frac{1}{Nmn}\sum_{i=1}^{N}(1-\theta_i)(f_{i2}-1)S_{iM}^2. \tag{2.2}$$

An unbiased variance estimator of (2.2) is given as

$$\hat{V}(\bar{y}_r') = \left(\frac{1}{n}-\frac{1}{N}\right)s_b'^2 + \frac{(N-1)}{Nn(n-1)}\sum_{i=1}^{n}\left(\frac{1}{m}-\frac{1}{M}\right)\frac{s_{im}'^2}{\alpha} + \frac{1}{Nn}\sum_{i=1}^{n}\frac{m_{i2}}{m^2}(f_{i2}-1)\frac{s_{im}'^2}{\alpha} + \frac{(N-n)}{NMn(n-1)}\sum_{i=1}^{n}\left[D-\frac{M}{m}\right]\frac{s_{im}'^2}{\alpha}, \tag{2.3}$$

where

$$S_b^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(\bar{Y}_{iM}-\bar{Y})^2, \bar{Y}_{iM} = \frac{1}{M}\sum_{j=1}^{M}Y_{ij},$$

$$S_{iM}^2 = \frac{1}{(M-1)}\sum_{j=1}^{M}(Y_{ij}-\bar{Y}_{iM})^2, \quad D = \frac{m_{i1}^2}{m^2}+\frac{m_{i2}^2}{m^2} \text{ and}$$

$$\alpha = \frac{(m+1)M}{m} - \frac{m_{i2}}{m^2(m-1)}(f_{i2}-1) + \frac{[D-m]}{M(m-1)}.$$

Define,

$$s_b'^2 = \frac{1}{(n-1)}\left(\sum_{i=1}^{n}\bar{y}_{im}'^2 - n\bar{y}_r'^2\right), \bar{y}_{im}' = \frac{1}{m}(m_{i1}\bar{y}_{m_{il}} + m_{i2}\bar{y}_{h_{i2}}),$$

$$s_{im}'^2 = \frac{1}{(m-1)}(\sum_{j=1}^{m_{i1}} y_{ij}^2 + \frac{m_{i2}}{h_{i2}}\sum_{j=1}^{h_{i2}} y_{ij}^2 - m\bar{y}_{im}'^2), \bar{y}_{im} = \frac{1}{m}\sum_{j=1}^{m} y_{ij}.$$

Proof: By definition, we have

$$\begin{aligned}
E(\bar{y}_r') &= E_1E_2E_3E_4\left[E_5\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{h_{i2}})}{m}\right\}\right] \\
&= E_1E_2E_3\left[E_4\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{m_{i2}})}{m}\right\}\right] \\
&= E_1E_2\left[E_3\left\{\frac{1}{n}\sum_{i=1}^{n}(\frac{m_{i1}}{m}\bar{y}_{im} + \frac{m_{i2}}{m}\bar{y}_{im})\right\}\right] \\
&= E_1\left[E_2\left\{\frac{1}{n}\sum_{i=1}^{n}(\theta_i\bar{y}_{im} + (1-\theta_i)\bar{y}_{im})\right\}\right] \\
&= E_1\left[\frac{1}{n}\sum_{i=1}^{n}(\theta_i\bar{Y}_{iM} + (1-\theta_i)\bar{Y}_{iM})\right] \\
&= E_1\left\{\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{iM}\right\} = \frac{1}{N}\sum_{i=1}^{N}\bar{Y}_{iM} = \bar{Y}.
\end{aligned}$$

This shows that $\bar{y}_r'$ is an unbiased estimator of the population mean $\bar{Y}$. Here, E_5 represents the conditional expectations of all possible samples of size h_{i2} drawn from m_{i2}, E_4 is the conditional expectation of all possible samples of size m_{i1}, m_{i2} respectively drawn from m by keeping m_{i1}, m_{i2} fixed, E_3 is the conditional expectation arising out of m independent Bernoulli trials leading to m_{i1} success and m_{i2} failures, $m_{i1}+ m_{i2} = m$, E_2 is the conditional expectation of all possible samples of size m drawn from M and E_1 refers to expectation arising out of all possible samples of size n drawn from a population of size N.

Similarly, we can write

$$\begin{aligned}
V(\bar{y}_r') = {} & V_1\{E_2E_3E_4E_5(\bar{y}_r')\} + E_1V_2\{E_3E_4E_5(\bar{y}_r')\} \\
& + E_1E_2V_3\{E_4E_5(\bar{y}_r')\} + E_1E_2E_3V_4\{E_5(\bar{y}_r')\} \\
& + E_1E_2E_3E_4\{V_5(\bar{y}_r')\}.
\end{aligned}$$

Various terms are expressed as below.

$$V_1\{E_2E_3E_4E_5(\bar{y}_r')\} = \left(\frac{1}{n} - \frac{1}{N}\right)S_b^2,$$

$$E_1V_2\{E_3E_4E_5(\bar{y}_r')\} = \frac{1}{Nn}\sum_{i=1}^{N}\left(\frac{1}{m} - \frac{1}{M}\right)S_{iM}^2,$$

$$E_1E_2V_3\{E_4E_5(\bar{y}_r')\} = 0,$$

$$E_1E_2E_3V_4\{E_5(\bar{y}_r')\} = 0,$$

$$E_1E_2E_3E_4\{V_5(\bar{y}_r')\} = \frac{1}{Nn}\sum_{i=1}^{N}\frac{(1-\theta_i)}{m}(f_{i2}-1)S_{iM}^2.$$

Here, V_1, V_2, V_3, V_4 and V_5 are defined similarly as E_1, E_2, E_3, E_4 and E_5. Hence, by adding all the terms we get,

$$\begin{aligned}
V(\bar{y}_r') = {} & \left(\frac{1}{n} - \frac{1}{N}\right)S_b^2 + \frac{1}{Nn}\sum_{i=1}^{N}\left(\frac{1}{m} - \frac{1}{M}\right)S_{iM}^2 \\
& + \frac{1}{Nmn}\sum_{i-1}^{N}(1-\theta_i)(f_{i2}-1)S_{iM}^2.
\end{aligned}$$

To obtain an unbiased variance estimator, we proceed as follows,

Consider, $s_b'^2 = \frac{1}{(n-1)}\left(\sum_{i=1}^{n}\bar{y}_{im}'^2 - n\bar{y}_r'^2\right)$ where,

$$\bar{y}_{im}' = \frac{1}{m}(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{h_{i2}}).$$

It can be shown that,

$$\begin{aligned}
E_1E_2E_3E_4E_5(s_b'^2) = {} & S_b^2 - \frac{1}{N(n-1)}\sum_{i=1}^{N}\left(\frac{1}{m} - \frac{1}{M}\right)S_{iM}^2 \\
& + \frac{1}{N}\sum_{i=1}^{N}\frac{(1-\theta_i)}{m}(f_{i2}-1)S_{iM}^2 - \frac{n}{N(n-1)M} \\
& \times\sum_{i=1}^{N}\left\{\frac{2\theta_i(1-\theta_i)}{m} + (1-\theta_i)^2 + \theta_i^2 - \frac{M}{m}\right\}S_{iM}^2 \text{ and}
\end{aligned}$$

$$\begin{aligned}
E_2E_3E_4E_5(s_{im}'^2) = {} & \frac{(m+1)M}{m}S_{iM}^2 - \frac{(1-\theta_i)}{m(m-1)}(f_{i2}-1)S_{iM}^2 \\
& + \frac{1}{(m-1)}\left\{\frac{2\theta_i(1-\theta_i)}{m} + (1+\theta_i)^2 + \theta_i^2 - m\right\}\frac{S_{iM}^2}{M}.
\end{aligned}$$

Therefore,

$$\hat{S}_b^2 = s_b'^2 + \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{1}{m}-\frac{1}{M}\right)\frac{s_{im}'^2}{\alpha} - \frac{1}{n}\sum_{i=1}^{n}\frac{m_{i2}}{m^2}(f_{i2}-1)\frac{s_{im}'^2}{\alpha} + \frac{1}{M(n-1)}\sum_{i=1}^{n}(D-\frac{M}{m})\frac{s_{im}'^2}{\alpha}$$

and

$$\hat{S}_{iM}^2 = \frac{s_{im}'^2}{\alpha}$$

Substituting the estimated values of S_b^2 and S_{iM}^2 in Eq. (2.2) we get the required result.

As the total cost of the survey is proportional to the optimum size of the sample, we determine the optimum values of n, m and f_{i2} by minimizing the expected cost for a fixed variance. To achieve this, consider the following cost function

$$C = C_1 n + C_2\sum_{i=1}^{n} m_{i1} + C_3\sum_{i=1}^{n} h_{i2}$$

where,

C : Total cost,

C_1 : Per unit travel and miscellaneous cost between the psus,

C_2 : Cost per unit of collecting the information on the study character in the first attempt,

C_3 : Cost per unit of collecting the information by expensive method after the first attempt to obtain information failed.

The cost function considered above is suitable for situations prevailing in mail surveys. In these surveys the first attempt to collect information from the respondents is made through e-mail/postal mail. Many of the respondents may not send the required information through mails. To collect information, a subsample of nonrespondents may be obtained for data collection by specialized effort, say, personal interview.

The expected cost in this case is,

$$C' = E(C) = n\left[C_1 + \frac{C_2}{N}\sum_{i=1}^{N} m\theta_i + \frac{C_3}{N}\sum_{i=1}^{N}\frac{m(1-\theta_i)}{f_{i2}}\right].$$

To minimize the expected cost consider the function $\phi = C' + \lambda\{V(\bar{y}'') - V_0\}$. Here, λ is the Lagrangian multiplier. Also, we determine V_0 by fixing the coefficient of variation, say equal to 5%. To obtain closed form expressions for the various sample sizes we have considered $m_{i2} = h_{i2}f_2$ in place of $m_{i2} = h_{i2}f_{i2}$, $i = 1, 2, \ldots, n$. Differentiating with respect to n, m, λ and f_2 equating the resultant derivatives to '0' and simplifying give the optimum values as,

$$n_{opt} = \frac{k}{\left(V_0 + \frac{S_b^2}{N}\right)},$$

$$m_{opt} = \frac{-(B-D_1)+\sqrt{(B-D_1)^2+4AE}}{2A} \text{ and}$$

$$f_{2opt} = \pm\sqrt{\frac{C_3\sum_{i=1}^{N}(1-\theta_i)\left[mS_b^2 - \frac{1}{N}\sum_{i=1}^{N}(1-\theta_i)S_{iM}^2 + \frac{1}{N}\sum_{i=1}^{N}\frac{M-m}{M}S_{iM}^2\right]}{\left[\frac{C_1}{m}+\frac{C_2}{N}\sum_{i=1}^{N}\theta_i\right]\sum_{i=1}^{N}(1-\theta_i)S_{iM}^2}}.$$

Keeping in view the fact that the sample sizes are positive values, we took only the positive values of m_{opt} and f_{2opt}.

$$m_{opt} = \frac{-(B-D_1)-\sqrt{(B-D_1)^2+4AE}}{2A} \text{ and}$$

$$f_{2opt} = \sqrt{\frac{C_3\sum_{i=1}^{N}(1-\theta_i)\left[mS_b^2 - \frac{1}{N}\sum_{i=1}^{N}(1-\theta_i)S_{iM}^2 + \frac{1}{N}\sum_{i=1}^{N}\frac{M-m}{M}S_{iM}^2\right]}{\left[\frac{C_1}{m}+\frac{C_2}{N}\sum_{i=1}^{N}\theta_i\right]\sum_{i=1}^{N}(1-\theta_i)S_{iM}^2}}.$$

where,

$$k = S_b^2 + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2 + \frac{1}{N}\sum_{i=1}^{N}\frac{(1-\theta_i)}{m}(f_2-1)S_{iM}^2,$$

$$D_1 = \left[\frac{C_2}{N}\sum_{i=1}^{N}\theta_i + \frac{C_3}{N}\sum_{i=1}^{N}\frac{(1-\theta_i)}{f_2}\right]\sum_{i=1}^{N}(1-\theta_i)S_{iM}^2,$$

$$A = C_3\sum_{i=1}^{N}\frac{(1-\theta_i)}{f_2^2}S_b^2 - \sum_{i=1}^{N}\frac{S_{iM}^2}{N},$$

$$B = C_3\sum_{i=1}^{N}\frac{(1-\theta_i)}{f_2^2 N}\{\sum_{i=1}^{N}(f_2-1)S_{iM}^2 + \sum_{i=1}^{N}S_{iM}^2\},$$

$$\text{E} = \sum_{i=1}^{N} S_{iM}^2 \text{ and } V_0 = 0.0025 \times \bar{Y}^2 .$$

Case 2. Consider the situation that a sample of n psus is drawn from N and within each selected psu a sample of m ssus is drawn by srswor design. Let there be no nonresponse in n_1 psus. In the remaining n_2 psus, m_{i1} ssus respond while m_{i2} ssus do not respond. A subsample of h_{i2} units is selected from m_{i2} by srswor and data are collected through specialized efforts, $m_{i2} = h_{i2} f_{i2}$, $i = 1, 2, \ldots, n_2$.

In this context, we state the Theorem 2.2 as below

Theorem2.2. The estimator

$$\bar{y}_r'' = \frac{1}{n}\left\{\sum_{i=1}^{n_1} \bar{y}_{im} + \sum_{i=1}^{n_2} \frac{(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{h_{i2}})}{m}\right\} \qquad (2.4)$$

is unbiased for $\bar{Y}$, with variance

$$V(\bar{y}_r'') = \left(\frac{1}{n} - \frac{1}{N}\right) S_b^2 + \frac{1}{Nn}\sum_{i=1}^{N}\left(\frac{1}{m} - \frac{1}{M}\right) S_{iM}^2 + \frac{1}{Nn}\sum_{i=1}^{N_2} \frac{(1-\theta_i)}{m}(f_{i2} - 1) S_{iM}^2 . \qquad (2.5)$$

An unbiased variance estimator of the estimator $\bar{y}_r''$ is given as

$$\hat{V}(\bar{y}_r'') = \left(\frac{1}{n} - \frac{1}{N}\right) s_b''^2 + \frac{N-1}{Nn(n-1)}\sum_{i=1}^{n}\left(\frac{1}{m} - \frac{1}{M}\right)\frac{s_{im}'^2}{\alpha} + \frac{1}{Nn}\sum_{i=1}^{n_2}\frac{m_{i2}}{m^2}(f_{i2} - 1)\frac{s_{im}'^2}{\alpha} + \frac{N-n}{Nn(n-1)} \times \left\{\sum_{i=1}^{n_2}\left[D - \frac{M}{m}\right]\frac{s_{im}'^2}{M\alpha} - \sum_{i=1}^{n_1}\left(\frac{1}{m} - \frac{1}{M}\right) s_{im}^2\right\}, \qquad (2.6)$$

where

$$s_b''^2 = \frac{1}{(n-1)}\left[\sum_{i=1}^{n_1} \bar{y}_{im}^2 + \sum_{i=1}^{n_2}\frac{1}{m^2}(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{h_{i2}})^2 - n\bar{y}_r''^2\right]$$

and $s_{im}^2 = \frac{1}{(m-1)}\left(\sum_{j=1}^{m} \bar{y}_{ij}^2 - m\bar{y}_{im}^2\right), \bar{y}_{im} = \frac{1}{m}\sum_{j=1}^{m} y_{ij}.$

Proof: By definition, we have

$$E(\bar{y}_r'') = E_1E_2E_3E_4\left[E_5\frac{1}{n}\left\{\sum_{i=1}^{n_1}\bar{y}_{im} + \sum_{i=1}^{n_2}\frac{(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{h_{i2}})}{m}\right\}\right]$$

$$= E_1E_2E_3\left[E_4\frac{1}{n}\left\{\sum_{i=1}^{n_1}\bar{y}_{im} + \sum_{i=1}^{n_2}\frac{(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{m_{i2}})}{m}\right\}\right]$$

$$= E_1E_2\left[E_3\frac{1}{n}\left\{\sum_{i=1}^{n_1}\bar{y}_{im} + \sum_{i=1}^{n_2}\left(\frac{m_{i1}}{m}\bar{y}_{im} + \frac{m_{i2}}{m}\bar{y}_{im}\right)\right\}\right]$$

$$= E_1\left[E_2\frac{1}{n}\left\{\sum_{i=1}^{n_1}\bar{y}_{im} + \sum_{i=1}^{n_2}(\theta_i\bar{y}_{im} + (1-\theta_i)\bar{y}_{im})\right\}\right]$$

$$= E_1\left[\frac{1}{n}\left\{\frac{n_1}{N_1}\sum_{i=1}^{N_1}\bar{Y}_{iM} + \frac{n_2}{N_2}\sum_{i=1}^{N_2}\bar{Y}_{iM}\right\}\right]$$

$$= \frac{1}{N}\left\{\sum_{i=1}^{N_1}\bar{Y}_{iM} + \sum_{i=1}^{N_2}\bar{Y}_{iM}\right\} = \bar{Y} .$$

Thus, $\bar{y}_r''$ is an unbiased estimator of the population mean $\bar{Y}$. Here, E_5 represents conditional expectations of all possible samples of size h_{i2} drawn from m_{i2}, E_4 is the conditional expectation of all possible samples of size m_{i1}, m_{i2} respectively drawn from m by keeping m_{i1}, m_{i2} fixed, E_3 is the conditional expectation arising out of m independent Bernoulli trials leading to m_{i1} success and m_{i2} failures, $m_{i1} + m_{i2} = m$, E_2 is the conditional expectation of all possible samples of size m drawn from M and E_1 arises out of selection of all possible samples of size n from N.

To obtain the variance we proceed as follows:

$$V(\bar{y}_r'') = V_1\{E_2E_3E_4E_5(\bar{y}_r'')\} + E_1V_2\{E_3E_4E_5(\bar{y}_r'')\} + E_1E_2V_3\{E_4E_5(\bar{y}_r'')\} + E_1E_2E_3V_4\{E_5(\bar{y}_r'')\} + E_1E_2E_3E_4\{V_5(\bar{y}_r'')\}$$

where,

$$V_1\{E_2E_3E_4E_5(\bar{y}_r'')\} = \left(\frac{1}{n} - \frac{1}{N}\right)S_b^2,$$

$$E_1V_2\{E_3E_4E_5(\bar{y}_r'')\} = \frac{1}{Nn}\sum_{i=1}^{N}\left(\frac{1}{m} - \frac{1}{M}\right)S_{iM}^2,$$

$$E_1E_2V_3E_4E_5(\bar{y}_r'')\} = 0,$$

$$E_1E_2E_3V_4E_5(\bar{y}_r'')\} = 0$$

and $E_1E_2E_3E_4V_5(\bar{y}_r'') = \frac{1}{Nn}\sum_{i=1}^{N_2}\frac{(1-\theta_i)}{m}(f_{i2} - 1)S_{iM}^2 .$

Here, V_1, V_2, V_3, V_4 and V_5 are defined similarly as E_1, E_2, E_3, E_4 and E_5. By adding the above three terms we get the required result. To obtain an unbiased variance estimator, consider,

$$s_b''^2 = \frac{1}{(n-1)}\left[\sum_{i=1}^{n_1}\bar{y}_{im}^2 + \sum_{i=1}^{n_2}\frac{1}{m^2}(m_{i1}\bar{y}_{m_{i1}} + m_{i2}\bar{y}_{h_{i2}})^2 - n\bar{y}_r''^2\right]$$

Taking the expectations and simplifying we get,

$$E_1E_2E_3E_4E_5E_6(s_b''^2) = S_b^2 - \frac{1}{N(n-1)}\sum_{i=1}^{N}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2$$

$$+\frac{1}{N}\sum_{i=1}^{N_2}\frac{(1-\theta_i)}{m}(f_{i2}-1)S_{iM}^2 + \frac{n}{N(n-1)}\left[\sum_{i=1}^{N_1}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2\right.$$

$$\left.-\sum_{i=1}^{N_2}\left\{\frac{2\theta_i(1-\theta_i)}{m}+(1-\theta_i)^2+\theta_i^2-\frac{M}{m}\right\}\frac{S_{iM}^2}{M}\right],$$

and also we have

$$E_3E_4E_5E_6(s_{im}^2) = S_{iM}^2,$$

$$E_3E_4E_5E_6(s_{im}'^2) = \frac{(m+1)M}{m}S_{iM}^2 - \frac{(1-\theta_i)}{m(m-1)}(f_{i2}-1)S_{iM}^2$$

$$+\frac{1}{(m-1)}\left\{\frac{2\theta_i(1-\theta_i)}{m}+(1-\theta_i)^2+\theta_i^2-m\right\}\frac{S_{iM}^2}{M}.$$

Here E_6 represents conditional expectations of all possible samples of size h_{i2} drawn from m_{i2}, E_5 is the conditional expectation of all possible samples of size m_{i1}, m_{i2} respectively drawn from m by keeping m_{i1}, m_{i2} fixed, E_4 is the conditional expectation arising out of m independent Bernoulli trials leading to m_{i1} success and m_{i2} failures, $m_{i1} + m_{i2} = m$, E_3 is the conditional expectation of all possible samples of size m drawn from M, E_2 arising out of selection of all possible samples of size n_1, n_2 drawn from N_1, N_2 keeping n_1, n_2 fixed and E_1 is the expectation arises out of randomness of n_1 and n_2, $n_1 + n_2 = n$, $N_1 + N_2 = N$.

Thus,

$$\hat{S}_b^2 = s_b''^2 + \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{1}{m}-\frac{1}{M}\right)\frac{s_{im}'^2}{\alpha} - \frac{1}{n}\sum_{i=1}^{n_2}\frac{m_{i2}}{m^2}(f_{i2}-1)\frac{s_{im}'^2}{\alpha}$$

$$+\frac{1}{(n-1)}\left\{\sum_{i=1}^{n_2}(D-\frac{M}{m})\frac{s_{im}'^2}{M\alpha} - \sum_{i=1}^{n_1}\left(\frac{1}{m}-\frac{1}{M}\right)s_{im}^2\right\}.$$

For the psus with no nonresponse $\hat{S}_{iM}^2 = s_{im}^2$ while in psus with nonresponse problem $\hat{S}_{iM}^2 = \frac{s_{im}'^2}{\alpha}$. Substituting the estimated values in the variance expression in Eq. (2.5) we obtain the required result.

To determine the optimum values of n, m and f_{i2} we proceed as earlier *i.e*. minimization of expected cost subject to fixed variance. The relevant cost function in this case is,

$$C = C_1n_2 + C_2n_1m + C_2\sum_{i=1}^{n_2}m_{i1} + C_3\sum_{i=1}^{n_2}h_{i2},$$

where, C, C_1, C_2 and C_3 are same as defined earlier. The expected cost is,

$$C'' = E(C)$$

$$= \frac{n}{N}\left[C_1N_2 + C_2N_1m + C_2\sum_{i=1}^{N_2}m\theta_i + C_3\sum_{i=1}^{N_2}\frac{(1-\theta_i)m}{f_{i2}}\right]$$

To minimize the expected cost consider the function $\phi = E(C) + \lambda\{V(\bar{y}_r'') V_0\}$, where λ is the Lagrangian multiplier. To obtain closed form expressions for the various sample sizes we have considered $m_{i2} = h_{i2}f_2$ in place of $m_{i2} = h_{i2}f_{i2}$, i = 1, 2, …, n_2. The optimum values obtained through minimization are as follows:

$$n_{opt} = \frac{k_1}{(V_0 + \frac{S_b^2}{N})}, m_{opt} = \sqrt{\frac{B_2}{A_2}} \text{ and } f_{2opt} = \sqrt{\frac{B_1}{A_1}}.$$

Keeping in view the fact that the sample sizes are positive values, we took only the positive values of m_{opt} and f_{2opt} which are

$$m_{opt} = \sqrt{\frac{B_2}{A_2}} \text{ and } f_{opt} = \sqrt{\frac{B_1}{A_1}}.$$

where,

$$k_1 = S_b^2 + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2 + \frac{1}{N}\sum_{i=1}^{N_2}\frac{(1-\theta_i)}{m}(f_2-1)S_{iM}^2,$$

$$A_1 = \left(C_2N_1 + C_2\sum_{i=1}^{N_2}\theta_i\right)\sum_{i=1}^{N_2}(1-\theta_i)S_{iM}^2,$$

$$B_1 = C_3\sum_{i=1}^{N_2}(1-\theta_i)\left[\sum_{i=1}^{N_2}S_{iM}^2 - \sum_{i=1}^{N_2}(1-\theta_i)S_{iM}^2\right],$$

$$A_2 = C_2N_1 + C_2\sum_{i=1}^{N_2}\theta_i + C_3\sum_{i=1}^{N_2}\frac{(1-\theta_i)}{f_2}\left[NS_b^2 - \frac{1}{M}\sum_{i=1}^{N}S_{iM}^2\right],$$

$$B_2 = C_1N_2\left[\sum_{i=1}^{N}S_{iM}^2 + \sum_{i=1}^{N_2}(1-\theta_i)(f_2-1)S_{iM}^2\right],$$

and $V_0 = 0.0025\times\overline{Y}^2$.

Case 3. Let a sample of n psus is drawn from N, within each selected psu a sample of m ssus is drawn by srswor. Let there be no nonresponse in n_1 psus. In the n_2 psus m_{i1} ssus respond while m_{i2} ssus do not respond. A subsample of h_{i2} units is selected by srswor from m_{i2} and data are collected through specialized efforts, let there be complete nonresponse in the n_3 psus. Further a subsample of h_3 psus is drawn out of n_3 psus and data are collected through specialized efforts on each of m ssus in the selected h_3 psus. Here $n_3 = f_3h_3$ and $m_{i2} = h_{i2}f_{i2}$, $i = 1, 2, \ldots, n_2$. In this context we state the following theorem.

Theorem 2.3.The unbiased estimator of $\overline{Y}$ is

$$\overline{y}_r''' = \frac{1}{n}\{\sum_{i=1}^{n_1}\overline{y}_{im} + \sum_{i=1}^{n_2}\frac{1}{n}(m_{i1}\overline{y}_{m_{i1}} + m_{i2}\overline{y}_{h_{i2}}) + \frac{n_3}{h_3}\sum_{i=1}^{h_3}\overline{y}_{im}\} \tag{2.7}$$

with variance

$$V(\overline{y}_r'') = \left(\frac{1}{n}-\frac{1}{N}\right)S_b^2 + \frac{1}{Nn}\left(\frac{1}{m}-\frac{1}{M}\right)\left\{\sum_{i=1}^{N_1}S_{iM}^2 + \sum_{i=1}^{N_2}S_{iM}^2 + f_3\sum_{i=1}^{N_3}S_{iM}^2\right\} + \frac{1}{Nn}\sum_{i=1}^{N_2}\frac{(1-\theta_i)}{m}(f_{i2}-1)S_{iM}^2 + \frac{N_3}{Nn}(f_3-1)S_{bN_3}^2 \tag{2.8}$$

where, $S_{bN_3}^2 = \frac{1}{N_3-1}\sum_{i=1}^{N_3}(\overline{Y}_{iM}-\overline{Y}_{N_3})^2$, where, $\overline{Y}_{N_3} = \frac{1}{N_3}\sum_{i=1}^{N_3}\overline{Y}_{iM}$,

An unbiased estimator of variance is,

$$\hat{V}(\overline{y}_r''') = \left(\frac{1}{n}-\frac{1}{N}\right)s_b'''^2 + \frac{1}{Nn}\left(\frac{1}{m}-\frac{1}{M}\right)\sum_{i=1}^{n_1}s_{im}^2 + \frac{1}{Nn}\sum_{i=1}^{n_2}\frac{m_{i2}}{m^2}(f_{i2}-1)\frac{s_{im}'^2}{\alpha} + \frac{(N-1)}{Nn(n-1)}\left(\frac{1}{m}-\frac{1}{M}\right)\sum_{i=1}^{n_2}\frac{s_{im}'^2}{\alpha} + \frac{(N-n)}{Nn(n-1)}\sum_{i=1}^{n_2}\left[D-\frac{M}{m}\right]\frac{s_{im}'^2}{M\alpha} + \frac{n_3(f_3-1)}{n}\frac{(N-1)}{N(n-1)} \times\left\{s_{bh_3}^2 - \left(\frac{1}{m}-\frac{1}{M}\right)\frac{1}{h_3}\sum_{i=1}^{h_3}s_{im}^2\right\} + \frac{1}{n^2}\left[f_3 - \frac{(N-n)(n-f_3)}{N(n-1)}\right] \times\sum_{i=1}^{h_3}\left(\frac{1}{m}-\frac{1}{M}\right)s_{im}^2, \tag{2.9}$$

where, $s_b'''^2 = \frac{1}{(n-1)}\left[\sum_{i=1}^{n_1}\overline{y}_{im}^2 + \sum_{i=1}^{n_2}\frac{1}{m^2}(m_{i1}\overline{y}_{m_{i1}} + m_{i2}\overline{y}_{h_{i2}})^2 + \frac{n_3}{h_3}\sum_{i=1}^{n_3}\overline{y}_{im}^2 - n\overline{y}_r'''^2\right]$ and $s_{bh_3}^2 = \frac{1}{(h_3-1)}\sum_{i=1}^{h_3}(\overline{y}_{im}-\overline{y}_{h_3})^2$,

where, $\overline{y}_{h_3} = \frac{1}{h_3}\sum_{i=1}^{h_3}\overline{y}_{im}$ and $S_b^2, \overline{y}_{im}, s_{im}^2, s_{im}'^2, \alpha, D$ etc. are defined earlier.

Proof: By definition,

$$E(\overline{y}_r''') = E_1E_2E_3E_4 \times E_5\left[E_6\frac{1}{n}\left\{\sum_{i=1}^{n_1}\overline{y}_{im} + \sum_{i=1}^{n_2}\frac{(m_{i1}\overline{y}_{m_{i1}} + m_{i2}\overline{y}_{h_{i2}})}{m} + \frac{n_3}{h_3}\sum_{i=1}^{h_3}\overline{y}_{im}\right\}\right]$$

$$= E_1E_2E_3E_4 \times \left[E_5\frac{1}{n}\left\{\sum_{i=1}^{n_1}(\overline{y}_{im}) + \sum_{i=1}^{n_2}\frac{(m_{i1}\overline{y}_{m_{i1}} + m_{i2}\overline{y}_{m_{i2}})}{m} + \frac{n_3}{h_3}\sum_{i=1}^{h_3}(\overline{y}_{im})\right\}\right]$$

$$= E_1E_2\left[E_3\frac{1}{n}\left\{\sum_{i=1}^{n_1}\overline{y}_{im} + \sum_{i=1}^{n_2}(\theta_i\overline{y}_{im} + (1-\theta_i)\overline{y}_{im}) + \frac{n_3}{h_3}\sum_{i=1}^{h_3}\overline{y}_{im}\right\}\right]$$

$$= E_1\left[E_2\frac{1}{n}\left\{\sum_{i=1}^{n_1}\overline{Y}_{iM} + \sum_{i=1}^{n_2}(\theta_i\overline{Y}_{iM} + (1-\theta_i)\overline{Y}_{iM}) + \frac{n_3}{h_3}\sum_{i=1}^{h_3}\overline{Y}_{iM}\right\}\right]$$

$$= E_1\left[\frac{1}{n}\left\{\sum_{i=1}^{n_1}\overline{Y}_{iM} + \sum_{i=1}^{n_2}\overline{Y}_{iM} + \sum_{i=1}^{n_3}\overline{Y}_{iM}\right\}\right]$$

$$= E_1\left[\frac{1}{n}\sum_{i=1}^{n_1}\overline{Y}_{iM}\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\overline{Y}_{iM} = \overline{Y}.$$

Hence, $\bar{Y}_r'''$ is an unbiased estimator of the population mean $\bar{Y}$. Where, E_6 represents conditional expectations of all possible samples of size h_{i2} drawn from, E_5 is the conditional expectation of all possible samples of size m_{i1}, m_{i2} respectively drawn from m by keeping m_{i1}, m_{i2} fixed, E_4 is the conditional expectation arising out of m independent Bernoulli trials leading to m_{i1} success and m_{i2} failures, $m_{i1}+m_{i2}=m$, E_3 is the conditional expectation of all possible samples of size m drawn from M, E_2 arises out of selection of all possible samples of size h_3 drawn from n_3 while E_1 arises out of selection of all possible samples of size n from N.

To obtain the variance we proceed as follows:

$$V(\bar{y}_r''') = V_1\{E_2E_3E_4E_5E_6\ (\bar{y}_r''')\} + E_1V_2(E_3E_4E_5E_6(\bar{y}_r''')\} + E_1E_2V_3\{E_4E_5E_6(\bar{y}_r''')\} + E_1E_2E_3V_4\{E_4E_6(\bar{y}_r''')\} + E_1E_2E_3E_4V_5\{E_6(\bar{y}_r''')\} + E_1E_2E_3E_4E_5\{V_6(\bar{y}_r''')\}.$$

Here, V_1, V_2, V_3, V_4, V_5 and V_6 are defined similarly as E_1, E_2, E_3, E_4, E_5 and E_6.

Hence, we have,

$$V_1\{E_2E_3E_4E_5E_6(\bar{y}_r''')\} = (\frac{1}{n}-\frac{1}{N})S_b^2,$$

$$E_1V_2\{E_3E_4E_5E_6(\bar{y}_r''')\} = \frac{N_3}{Nn}(f_3-1)S_{bN_3}^2,$$

$$E_1E_2V_3\{E_4E_5E_6(\bar{y}_r''')\} = \frac{1}{Nn}\left(\frac{1}{m}-\frac{1}{M}\right)\left\{\sum_{i=1}^{N_1}S_{iM}^2+\sum_{i=1}^{N_2}S_{iM}^2+f_3\sum_{i=1}^{N_3}S_{iM}^2\right\},$$

$$E_1E_2E_3V_4\{E_5E_6(\bar{y}_r''')\}=0,$$

$$E_1E_2E_3E_4V_5\{E_6(\bar{y}_r''')\}=0,$$

$$E_1E_2E_3E_4E_5\{V_6(\bar{y}_r''')\} = \frac{1}{Nn}\sum_{i=1}^{N_2}\frac{(1-\theta_i)}{m}(f_{i2}-1)S_{iM}^2.$$

Thus, by adding all the terms we obtain the required variance of the estimator *i.e.*

$$V(\bar{y}_r''')=\left(\frac{1}{n}-\frac{1}{N}\right)S_b^2+\frac{1}{Nn}\left(\frac{1}{m}-\frac{1}{M}\right)\left\{\sum_{i=1}^{N_2}S_{iM}^2+\sum_{i=1}^{N_2}S_{iM}^2+f_3\sum_{i=1}^{N_3}S_{iM}^2\right\}+\frac{1}{Nn}\sum_{i=1}^{N_2}\frac{(1-\theta_i)}{m}(f_{i2}-1)S_{iM}^2+\frac{N_3}{Nn}(f_3-1)S_{bN_3}^2$$

To obtain the unbiased estimator of variance, consider,

$$s_b'''^2 = \frac{1}{(n-1)}\left[\sum_{i=1}^{n_1}\bar{y}_{im}^2+\sum_{i=1}^{n_2}\frac{1}{m^2}(m_{i1}\bar{y}_{m_{i1}}+m_{i2}\bar{y}_{h_{i2}})^2+\frac{n_3}{h_3}\sum_{i=1}^{h_3}\bar{y}_{im}^2-n\bar{y}_r'''^2\right],$$

where,

$$E_1E_2E_3E_4E_5E_6E_7(s_b'''^2) = S_b^2+\frac{1}{N}\sum_{i=1}^{N_1}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2+\frac{1}{N}\sum_{i=1}^{N}\frac{(1-\theta_i)}{m}(f_{i2}-1)S_{iM}^2+\frac{n-f_3}{N(n-1)}\sum_{i=1}^{N_3}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2-\frac{1}{N(n-1)}\sum_{i=1}^{N_2}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2$$

$$-\frac{N_3}{N(n-1)}(f_3-1)S_{bN_3}^2-\frac{n}{N(n-1)}\sum_{i=1}^{N_2}\left\{\frac{2\theta_i(1-\theta_i)}{m}+(1-\theta_i)^2+\theta_i^2-\frac{M}{n}\right\}\frac{S_{iM}^2}{M},$$

$$E_4E_5E_6E_7(s_{im}^2) = S_{iM}^2,$$

$$E_4E_5E_6E_7(s_{im}'^2) = \frac{(m+1)M}{m}S_{iM}^2-\frac{(1-\theta_i)}{m(m-1)}(f_{i2}-1)S_{iM}^2+\frac{1}{(m-1)}\left\{\frac{2\theta_i(1-\theta_i)}{m}+(1-\theta_i)^2+\theta_i^2-m\right\}\frac{S_{iM}^2}{M},\ \text{and}$$

$$E_1E_2E_3E_4(s_{bh_3}^2) = S_{bN_3}^2+\frac{1}{N_3}\sum_{i=1}^{N_3}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2.$$

Here, E_7 represents conditional expectations of all possible samples of size h_{i2} drawn from m_{i2}, E_6 is the conditional expectation of all possible samples of m_{i1}, m_{i2} respectively drawn from m by keeping m_{i1}, m_{i2} fixed, E_5 is the conditional expectation arising out of m independent Bernoulli trials leading to m_{i1} success and m_{i2} failures, $m_{i1} + m_{i2} = m$, E_4 is the conditional expectation of all possible samples of size m drawn from M, E_3 arises out of selection of all possible samples of size h_3 drawn from n_3, E_2 arises out of selection of all possible samples of size n_1 ,n_2 and n_3 drawn from N_1, N_2 and N_3 keeping n_1, n_2 and n_3 fixed and E_1 refers to the expectation arises out of

randomness of n_1, n_2 and n_3, where $n_1 + n_2 + n_3 = n$ and $N_1 + N_2 + N_3 = N$.

Thus,

$$\hat{S}_b^2 = s_b'''^2 - \frac{1}{n}\left(\frac{1}{m}-\frac{1}{M}\right)\sum_{i=1}^{n_1} s_{im}^2 - \frac{1}{n}\sum_{i=1}^{n_2}\frac{m_{i2}}{m^2}(f_{i2}-1)\frac{s_{im}'^2}{\alpha}$$

$$+\frac{1}{(n-1)}\sum_{i=1}^{n_2}(D-\frac{M}{m})\frac{s_{im}'^2}{M\alpha}+\frac{1}{n(n-1)}\sum_{i=1}^{n_2}\left(\frac{1}{m}-\frac{1}{M}\right)\frac{s_{im}'^2}{\alpha}$$

$$+\frac{n_3}{n(n-1)}(f_3-1)\left(s_{bh_3}^2-\frac{1}{h_3}\sum_{i=1}^{h_3}\left(\frac{1}{m}-\frac{1}{M}\right)s_{im}^2\right)$$

$$-\frac{n-f_3}{n(n-1)}\sum_{i=1}^{h_3}\left(\frac{1}{m}-\frac{1}{M}\right)s_{im}^2.$$

Here again, for the psus with no nonresponse $\hat{S}_{iM}^2 = s_{im}^2$ while in psus when there is nonresponse problem $\hat{S}_{iM}^2 = \frac{s_{im}'^2}{\alpha}$ and $\hat{S}_{bN_3}^2 = s_{bh_3}^2 - \frac{1}{h_3}\sum_{i=1}^{h_3}\left(\frac{1}{m}-\frac{1}{M}\right)s_{im}^2$. Substituting the estimated values in the Eq. (2.8) we get the required expression in Eq. (2.9).

To determine the optimum values of n, m, f_{i2} and f_3 we proceed as follows,

The cost function in this case is given as,

$$C = C_1 nm + C_2 n_1 m + C_2\sum_{i=1}^{n_2} m_{i1} + C_3(\sum_{i=1}^{n_2} h_{i2} + n_3 h_3)$$

where, C, C_1, C_2 and C_3 are same as defined earlier.

The expected cost is,

$$C''' = E(C)$$

$$= n\left[C_1 m + C_2\frac{N_1}{N}m + \frac{C_2}{N}\sum_{i=1}^{N_2} m\theta_i + \frac{C_3}{N}\sum_{i=1}^{N_2}\frac{(1-\theta_i)m}{f_{i2}} + C_3\frac{N_3}{Nf_3}\right].$$

To minimize the expected cost consider the function, $\phi = E(C) + \lambda\{V(\bar{y}_r''') - V_0\}$. To obtain closed form expressions for the various sample sizes we have considered $m_{i2} = h_{i2} f_2$ in place of $m_{i2} = h_{i2} f_{i2}$, $i = 1, 2, \ldots, n_2$ and also to overcome the problem arising due to simultaneous minimisation of n, m, f_2, f_3 we assume that $n_3 = f_2 h_3$. Thus, minimization gives the optimum values as

$$n_{opt} = \frac{k_2}{(V_0 + \frac{S_b^2}{N})},\ m_{opt} = \pm\sqrt{\frac{G_2}{D_2}} \text{ and}$$

$$f_{2opt} = \frac{-B_2 \pm \sqrt{B_2^2 - 4A_2G_3}}{2A_2}$$

Keeping in view the fact that the sample sizes are positive values, we took only the positive values of m_{opt} and f_{2opt}

$$m_{opt} = \sqrt{\frac{G_2}{D_2}} \text{ and } f_{2opt} = \frac{-B_2 + \sqrt{B_2^2 - 4A_2G_3}}{2A_2},$$

where,

$$k_2 = S_b^2 + \frac{N_3}{N}(f_2-1)S_{bN_3}^2 + \frac{1}{N}\left(\frac{1}{m}-\frac{1}{M}\right)\times$$

$$\left\{\sum_{i=1}^{N_1} S_{iM}^2 + \sum_{i=1}^{N_2} S_{iM}^2 + f_3\sum_{i=1}^{N_3} S_{iM}^2\right\} + \frac{1}{N}\sum_{i=1}^{N_2}\frac{(1-\theta_i)}{m}(f_2-1)S_{iM}^2,$$

$$A_2 = N(C_1 + C_2\frac{N_1}{N} + \frac{C_2}{N}\sum_{i=1}^{N_2}\theta_i)\sum_{i=1}^{N_2}(1-\theta_i)S_{iM}^2,$$

$$B_2 = -C_3\sum_{i=1}^{N_2}(1-\theta_i)\sum_{i=1}^{N_2} S_{iM}^2,$$

$$G_3 = -C_3\sum_{i=1}^{N_2}(1-\theta_i)\{\sum_{i=1}^{N_1} S_{iM}^2 + \sum_{i=1}^{N_2} S_{iM}^2 - \sum_{i=1}^{N_2}(1-\theta_i)S_{iM}^2\},$$

$$G_2 = \sum_{i=1}^{N_2}(1-\theta_i)[N_3 S_{bN_3}^2 + (\frac{1}{m}-\frac{1}{M})S_{iM}^2],$$

$$D_2 = N_3 n_3\sum_{i=1}^{N_2}(1-\theta_i)S_{iM}^2 \text{ and } V_0 = 0.0025 \times \bar{Y}^2.$$

Control Case. The following estimator was also considered for efficiency comparison purpose. Here we assume that a srswor sample of n psus is selected from N and within each selected psu a srswor sample of m ssus are selected. Data are collected through specialised efforts *i.e.* there is no nonresponse. Then we give the following Theorem.

Theorem 2.4. The estimator

$$\bar{y} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_{im} \qquad (2.10)$$

is unbiased for $\bar{Y}$, with variance

$$V(\bar{y}) = \left(\frac{1}{n}-\frac{1}{N}\right)S_b^2 + \frac{1}{Nn}\sum_{i=1}^{N}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2. \qquad (2.11)$$

An unbiased estimator of variance is,

$$\hat{V}(\bar{y}) = \left(\frac{1}{n}-\frac{1}{N}\right)s_b^2 + \frac{1}{Nn}\sum_{i=1}^{n}\left(\frac{1}{m}-\frac{1}{M}\right)s_{im}^2, \qquad (2.12)$$

where,

$$s_b^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(\bar{y}_{im}-\bar{y})^2, \bar{y}_{im} = \frac{1}{m}\sum_{j=1}^{m} y_{ij}, \text{ and}$$

$$s_{im}^2 = \frac{1}{(m-1)}\left(\sum_{j=1}^{m} y_{ij}^2 - m\bar{y}_{im}^2\right).$$

Proof: By definition,

$$E(\bar{y}) = E_1\left[E_2\left(\frac{1}{n}\sum_{i=1}^{n}\bar{y}_{im}\right)\right] = E_1\left(\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{iM}\right)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\bar{Y}_{iM} = \bar{Y}.$$

Thus, $\bar{y}$ is an unbiased estimator of the population mean $\bar{Y}$. Here, E_2 denotes the conditional expectation pertaining to all possible samples of size m drawn from M and E_1 is the conditional expectation pertaining to all possible samples of size n drawn from N.

Again we see that, $V(\bar{y}) = V_1E_2(\bar{y}) + E_1V_2(\bar{y})$, where

$$V_1E_2(\bar{y}) = \left(\frac{1}{n}-\frac{1}{N}\right)S_b^2 \text{ and}$$

$$E_1V_2(\bar{y}) = \frac{1}{Nn}\sum_{i=1}^{N}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2.$$

To obtain unbiased estimator of variance, we have,

$$E(s_b^2) = E_1E_2(s_b^2) = S_b^2 + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2 \text{ and}$$
$$E(s_{im}^2) = S_{iM}^2.$$

Now substituting $\hat{S}_b^2 = s_b^2 - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}-\frac{1}{M}\right)s_{im}^2$ and $\hat{S}_{iM}^2 = s_{im}^2$ in the variance expression we obtain the required result.

The cost function in this case is, $C = C_1n + C_3nm$, where, C, C_1 and C_3 have been defined earlier. To obtain optimum values of n and m we minimize the cost by fixing the variance. The optimum values are as follows,

$$n_{opt} = \frac{S_b^2 + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{m}-\frac{1}{M}\right)S_{iM}^2}{\left(V_0 + \frac{S_b^2}{N}\right)}, \text{ and}$$

$$m_{opt} = \sqrt{\frac{C_1\sum_{i=1}^{N}\frac{S_{iM}^2}{N}}{C_3\left(S_b^2 - \frac{1}{MN}\sum_{i=1}^{N}S_{iM}^2\right)}}.$$

3. EMPIRICAL ILLUSTRATION

For the purpose of empirical illustration we consider the MU284 data given in Sarndal *et al.* (1992). Using this data a population with N=27 psus and M=10 ssus was generated by combining the adjacent 10 units and allocating them to the respective psus. In our analysis we considered two target variables from this MU284 data. These variables are denoted by P85 and P75, and described as the human population (in thousand) of 270 municipalities of Sweden in 1985 and 1975 respectively. Here, we used four different values of θ_i for each psus. Further, various combinations of cost components C_1, C_2 and C_3 were considered. We computed the percentage reduction in expected cost (%RIEC) as well as optimum values of sample sizes of different estimators described in previous with respect to controlled estimator $\bar{y}$. The values of %RIEC, optimum sample sizes and various combinations of cost components C_1, C_2 and C_3 are reported in Table 3.1 and Table 3.2. In particular, Table 3.1 and Table 3.2 present the values for P85 and P75 respectively. Note that the percentage reduction in expected cost for case i(i=1, 2, 3) is computed as %RIEC = $\frac{(C-C^{(i)})}{C}\times 100$ where C is the total cost for Control Case and $C^{(i)}$ (i = 1, 2, 3) is the expected cost

Table 3.1. The optimum values of sample sizes along with percentage reduction in expected cost (%RIEC) of $\overline{y}_r', \overline{y}_r'', \overline{y}_r'''$ over $\overline{y}_r$ for the variable P85.

Cost			Control ($\overline{y}_r$)		First estimator ($\overline{y}_r'$)				Second estimator ($\overline{y}_r''$)				Third estimator ($\overline{y}_r'''$)			
C_1	C_2	C_3	n	m	n	m	f_2	%RIEC	n	m	f_2	%RIEC	n	m	f_2	%RIEC
25	2	45	14	3	23	8	3.06	36.19	23	8	2.73	64.15	18	5	1.68	49.34
25	2	50	14	3	24	8	3.22	36.44	24	8	2.88	64.71	19	5	1.79	49.90
25	2	55	14	3	25	8	3.38	36.66	24	8	3.02	65.21	19	5	1.90	50.31
25	4	45	14	3	22	8	2.66	22.96	21	6	1.93	60.73	18	5	1.65	42.68
25	4	50	14	3	22	8	2.80	23.42	22	6	2.04	61.43	19	5	1.75	43.29
25	4	55	14	3	23	8	2.94	23.82	23	6	2.14	62.05	19	5	1.86	43.90
30	2	45	14	4	23	8	2.87	38.15	23	9	2.73	63.91	17	5	1.52	49.67
30	2	50	14	3	23	8	3.03	38.38	23	9	2.88	64.47	18	5	1.62	50.14
30	2	55	14	3	24	8	3.17	38.58	24	9	3.02	64.97	18	5	1.71	50.59
30	4	45	14	4	21	8	2.53	26.61	21	6	1.93	60.53	17	5	1.49	43.46
30	4	50	14	3	22	8	2.67	27.02	22	6	2.04	61.24	18	5	1.59	44.17
30	4	55	14	3	22	8	2.80	27.38	22	6	2.14	61.86	18	5	1.68	44.82
35	2	45	14	4	22	8	2.71	39.58	22	9	2.73	63.69	17	5	1.39	49.94
35	2	50	14	4	22	8	2.86	39.79	22	9	2.88	64.26	17	5	1.48	50.36
35	2	55	14	3	23	8	2.99	39.98	23	9	3.02	64.76	18	5	1.56	50.75
35	4	45	14	4	21	8	2.42	29.32	21	7	1.93	60.36	16	5	1.37	44.32
35	4	50	14	4	21	8	2.56	29.69	21	7	2.04	61.06	17	5	1.46	44.94
35	4	55	14	3	22	8	2.68	30.02	22	7	2.14	61.68	17	5	1.54	45.52

for the case *i.* That is, $C^{(1)} = C'$, $C^{(2)} = C''$ and $C^{(3)} = C'''$ are the expected cost for case 1, 2 and 3 respectively. The empirical analysis reported in the paper was done using SAS 9.3 software.

The results for variable P85 reported in Table 3.1 reveal that the %RIEC is maximum for the second estimator followed by the third estimator and least in the first estimator. The %RIEC increases with increase in travel and miscellaneous cost (C_1) for the first and third estimator and decreases in case of the second estimator. The %RIEC decreases with increase in data collection cost at first attempt (C_2) for the all the estimators. The %RIEC increases for all the estimators with the increase in cost per unit of collecting the information by expensive method after the first attempt to obtain information failed (C_3). It is also seen that, for given C_1 and C_2 as C_3 increases the rate of increase of %RIEC is maximum in the second estimator followed by the third estimator and least in the first estimator.

Turning now to Table 3.2 for the results of variable P75. In Table 3.2 again similar to variable P85 we observe that the %RIEC is maximum for the second estimator followed by the third estimator and least in the first estimator. However, in contrast the %RIEC

Table 3.2. The optimum values of sample sizes along with percentage reduction in expected cost (%RIEC) of $\bar{y}_r', \bar{y}_r'', \bar{y}_r'''$ over $\bar{y}_r$ for the variable P75.

Cost			Control ($\bar{y}_r$)		First estimator ($\bar{y}_r'$)				Second estimator ($\bar{y}_r''$)				Third estimator ($\bar{y}_r'''$)			
C_1	C_2	C_3	n	m	n	m	f_2	%RIEC	n	m	f_2	%RIEC	n	m	f_2	%RIEC
25	2	45	15	8	23	13	2.12	45.10	23	8	2.78	66.92	13	8	0.80	50.54
25	2	50	15	8	23	13	2.23	45.32	23	8	2.93	67.41	13	8	0.85	51.24
25	2	55	15	8	24	13	2.34	45.52	24	8	3.08	67.84	13	8	0.90	51.89
25	4	45	15	8	21	13	1.84	34.09	21	7	1.97	63.42	13	8	0.78	45.38
25	4	50	15	8	21	13	1.94	34.48	21	7	2.07	64.08	13	8	0.83	46.29
25	4	55	15	8	22	13	2.04	34.82	22	7	2.18	64.65	13	8	0.88	47.11
30	2	45	15	9	22	13	1.99	46.71	22	9	2.78	66.81	12	8	0.72	50.28
30	2	50	15	9	22	13	2.10	46.92	22	9	2.93	67.30	13	8	0.77	50.92
30	2	55	15	8	23	13	2.20	47.11	23	9	3.08	67.74	13	8	0.81	51.52
30	4	45	15	9	20	13	1.75	37.07	20	8	1.97	63.34	12	8	0.70	45.67
30	4	50	15	9	21	13	1.85	37.43	21	8	2.07	64.00	12	8	0.75	46.49
30	4	55	15	8	21	13	1.94	37.75	21	8	2.18	64.57	13	8	0.80	47.25
35	2	45	15	10	21	13	1.88	47.89	21	9	2.78	66.71	12	8	0.66	50.06
35	2	50	15	9	22	13	1.98	48.09	22	9	2.93	67.21	12	8	0.70	50.65
35	2	55	15	9	22	13	2.08	48.27	22	9	3.08	67.64	12	8	0.74	51.21
35	4	45	15	10	20	13	1.68	39.28	20	7	1.97	63.26	12	8	0.65	45.86
35	4	50	15	9	20	13	1.77	39.63	20	7	2.07	63.92	12	8	0.69	46.63
35	4	55	15	9	21	13	1.86	39.92	21	7	2.18	64.49	12	8	0.73	47.32

increases with increase in travel and miscellaneous cost (C_1) for the first estimator and decreases in case of the second and third estimator. Moreover, like for P85, the %RIEC decreases with increase in data collection cost at first attempt (C_2) for the all the estimators in case of P75 too. We also noticed a identical pattern between P85 and P75 with respect to C_3. Overall, results for two variable are almost identical.

4. DISCUSSION AND CONCLUSION

The %RIEC is maximum for the second estimator because there is partial nonresponse in the second stage for the first estimator for whole sample size n, whereas for the second estimator, there is partial nonresponse in the second stage only for a part of the sample size (*i.e.* n_2 psus) and there is complete response in the other part (*i.e*. n_1 psus) where as for the third estimator, there is full response in n_1 psus, partial nonresponse in the second stage for n_2 psus and complete nonresponse at the first stage for n_3 psus ($n_1 + n_2 + n_3 = n$). Thus the first estimator is more costly than other two estimators and the third estimator is more costly than the second estimator.

To summarize, all the three estimators, of population mean in the presence of nonresponse based on subsampling of the nonrespondents, have better % RIEC as compared to the estimator based only on interview method of data collection resulting in 100%

response. Among all the three estimators, the second estimator has the maximum %RIEC followed by the third estimator and %RIEC is least in the first estimator. Hence, the second estimator was found best among all the three estimators in respect of the criterion of %RIEC.

ACKNOWLEDGEMENTS

Authors are grateful to the Associate Editor and referees for their valuable comments.

REFERENCES

Chhikara, R.S. and Sud, U.C. (2009). Estimation of population and domain totals under two-phase sampling in the presence of nonresponse. *J. Ind. Soc. Agril. Statist.*, **63**, 297-304.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd Edition. John Wiley & Sons, Inc, New York.

Foradori, G.T. (1961). Some nonresponse sampling theory for two stage designs. Institute of Statistics, North Carolina State College.

Hansen, M.H. and Hurwitz, W.N. (1946). The problem of nonresponse in sample surveys. *Jour. Amer. Statist. Assoc.*, **41**, 517-529.

Okafor, F.C. (2001). Treatment of nonresponse in successive sampling. *Statistica,* **61(2)**, 195-204.

Okafor, F.C. (2005). Subsampling the nonrespondents in two-stage sampling over successive occasions. *J. Ind. Statist. Assoc.*, **43**(1), 33-49.

Okafor, F.C. and Lee, H. (2000). Double sampling for ratio and regression estimation with subsampling the nonrespondents. *Survey Methodology*, **26(2)**, 183-188.

Kalton, G. and Kasprzyk, D. (1986).The treatment of missing survey data. *Survey Methodology*, **12**, 1-16.

Oh, H.L. and Scheuren, F.J. (1983). Weighting adjustment for unit nonresponse. In: W.G. Madow, I. Olkin, and B. Rubin (eds.), *Incomplete Data in Sample Surveys*, Vol. 2. Academic Press, New York, 143-184.

Rancourt, E., Lee, H. and Särndal, C.E. (1994). Bias corrections for survey estimates from data with ratio imputed values for confounded nonresponse. *Survey Methodology*, **20**, 137-147.

Singh, R. and Mangat, N.P.S. (1996). *Elements of Survey Sampling*. Kluwer Academic Publishers.

Sarndal, C.E., Swenson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Srinath, K.P. (1971). Multiphase sampling in nonresponse problems. *Jour. Amer. Statist. Assoc.*, **66**, 583-586.

Sud, U.C., Aditya, K., Chandra, H. and Parsad, R. (2012). Two stage sampling for estimation of population mean with sub-sampling of non-respondents. *J. Ind. Soc. Agril. Statist.*, **64,** 343-347.

Tripathi, T.P. and Khare, B.B. (1997). Estimation of mean vector in presence of nonresponse. *Comm. Statist. - Theory Methods*, **26(9)**, 2255-2269.