Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 48 (5) (2019), 1522 – 1546
DOI : 10.15672/hujms.465609

Research Article

Homogeneous imputation under two phase probability proportional to size sampling

Muhammad Umair Sohail*[1], Javid Shabbir[1], Cem Kadilar[2]

[1] *Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan*
[2] *Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey*

Abstract

In this paper, we consider the problem of missing complete at random (MCAR) values in two phase probability proportional to size (*pps*) sampling for the estimation of population mean. A class of estimators is considered by the suitable use of auxiliary information with the traditional estimators for imputing the missing values. Theoretically, bias and mean squared errors of the proposed estimators are obtained up to the first order approximation. Two numerical studies are carried out for relative comparison of the proposed estimators with mean estimator under two phase *pps* sampling for each situation.

Mathematics Subject Classification (2010). 62D05

Keywords. missing values, two phase sampling, imputation, probability proportional to size, auxiliary information

1. Introduction

In field of survey sampling, researchers utilize different statistical tools and models for the selection of the sample units from a target population. The utilization of such statistical tools depends upon the availability of observation units in the given population. Nowadays, different probability and non-probability models are available in literature for the selection of units from the population (say Ω). In probability sampling scheme like simple random sampling (SRS) and systematic sampling (SS); every unit in the population is considered same with respect to size, so they have the same chance of selection in the sample. When the units have unequal probability of selection, then the probability begin proportion to size of the auxiliary information associated with the particular unit, is called probability proportional to size (*pps*) sampling. Availability of the suitable auxiliary information is the necessary condition for the selection of sample units in the sample in *pps* sampling, because we assign the selection probabilities on the behalf of the auxiliary variable.

In many real life situations the problem occurs if we have no auxiliary information regarding the variable of interest. In such cases, multi-phase sampling is a reliable procedure for obtaining the auxiliary information before observing the study variable. Readers may

*Corresponding Author.
Email addresses: umairsohailch@gmail.com (M.U. Sohail), javidshabbir@gmail.com (J. Shabbir), cemkadilar@gmail.com (C. Kadilar)
Received: 29.09.2018; Accepted: 19.02.2019

referred to read [4, 8, 10, 12] and [11]. The main focus of our present study is to combine the features of two phase and *pps* sampling for the estimation of population parameters, when the units are varying in size.

The problem occurs when the study or/and auxiliary have the missing values that lead to the misleading inference about the parameters of interest. These missing values can create the problem when sample units are difficult to follow-up or expensive to observe them repeatedly or at regular period of time. In comparison to follow-up visits, imputation is a well grounded procedure for imputing non-response without any specified cost and time. Several imputation strategies are available in literature to impute the missing value in efficient manners. [9] provide the idea about the nature of missing values by suggesting missing complete at random values (MCAR) and missing at random (MAR). [1] and [2] provide the efficient imputation models by utilizing the known parametric values of the auxiliary information. [3] considered the hot deck imputation under ranking set sampling. Many other researchers, such as [5, 6] and [13] consider this problem in an effective way.

The main focus of our study is to handle the problem of MCAR values which are usually occurred in most of the social science and demographic studies, where the respondents are reluctant to response to the certain items of the questionnaire. The brief discretion of this study is given bellow:

1.1. Statement of the problem

For a population (Ω) of size N units, with variate values of the study variable $(Y_j = Y_1, Y_2, Y_3 \ldots\ Y_N)$ and the auxiliary variables $(X_{1j} = X_{11}, X_{12}, X_{13}, \cdots X_{1N}$ and $X_{2j} = X_{21}, X_{22}, X_{23}, \cdots, X_{2N}\ )$, a random sample $(s^{'})$ of size m is drawn from Ω at the first phase. From the selected m units, the auxiliary information of X_1 or/and X_2 are obtained. At second phase, the sample s of n units is selected from the preselected m units, then the information is obtained on the study and the auxiliary variable respectively. Let r be the total number of the respondents, that can belongs to the sub-set of G in sample (s) and $(n-r)$ are those, who refuse to response relevant to the study variable from the subset G^c. Such that, $s = G \cup G^c$. Its also assumed that $\bar{y}_r = \frac{1}{r}\sum_{j=1}^{r} y_j$ be the sample mean of the study variable (Y) form G at the second phase. Let $\bar{x}_1 = \frac{1}{n}\sum_{j=1}^{n} x_{1j}$ and $\bar{x}_2 = \frac{1}{n}\sum_{j=1}^{n} y_{2j}$ be the unbiased sample mean corresponding to the population mean of $\bar{X}_1$ and $\bar{X}_2$, respectively. Let ρ_{yx_1} and ρ_{yx_2} be the correlation coefficient between the study variable and the auxiliary variables. It is also assume that, the first auxiliary variable (X_{1j}) has low correlation with Y_j than the second auxiliary variable (X_{2j}). So, X_{2j} is used at the estimation stage and X_{1j} is used at the design stage of the study.

Now, we define the four different possible situation under which the non-response is occurred in two phase *pps* sampling as follows:

1.1.1. Pps sampling in both phases. Let we have an auxiliary variable (X_{1j}) correlated (small in degree) with the study variable (Y_j). For the better estimation of Y_j, we wish to measure another auxiliary variable (X_{2j}), which has high correlation with the study variable at first phase by utilizing the selection probabilities of X_{1j}. The availability of response is discussed as:

Situation 1: Assume that we measure an auxiliary variable (X_{2j}) at first phase and full response is available about it. At second phase, the study variable and the auxiliary variable are measured accordingly. Assume that, the non-response is occurred only in the study variable.

Situation 2: Suppose that the full response about X_{2j} is not available at first phase, only $r^{'}$ units can provide the response out of m units $(r^{'} < m)$. At the second phase, we again face the problem of non-response in both the study and auxiliary variables receptively, only r out of $r^{'}$ units $(r < r^{'})$ can provide the response.

1.1.2. SRS on first phase and *pps* sampling on second phase. Let the selection probabilities of the study variable are not available, but we can visually understand that the units are varying in size. Then, our focus is to use the pps sampling by obtaining the selection (measuring the two auxiliary variables (X_{1j} and X_{2j}), which are selected by SRS) at the first phase. The auxiliary variable (X_{1j}) is used for obtaining the selection probabilities of sample units for second phase. The availability of response is define as follow:

Situation 3: As like situation 1: Let, complete response about X_1 and X_2 are obtained at first phase by using the SRS scheme. On the basis of first phase auxiliary information, we select the sample units at second phase for the study variable by using the selection probabilities by *pps* sampling and assume that only the non-response be occurred in the study variable .

Situation 4: As like situation 2: Let, the non-response be occurred during observing X_{1j} and X_{2j} at first phase, only r' units can provide the response. We utilized such limited information for the selection of sample units for the study variable at second phase and we assume that the non-response is occurred in the study variable and in the auxiliary variable as well.

For each of the above mentioned situations, we consider four different imputation procedures for each and totally sixteen procedures to consider the comprehensive examination of missing values in two phase *pps* sampling.

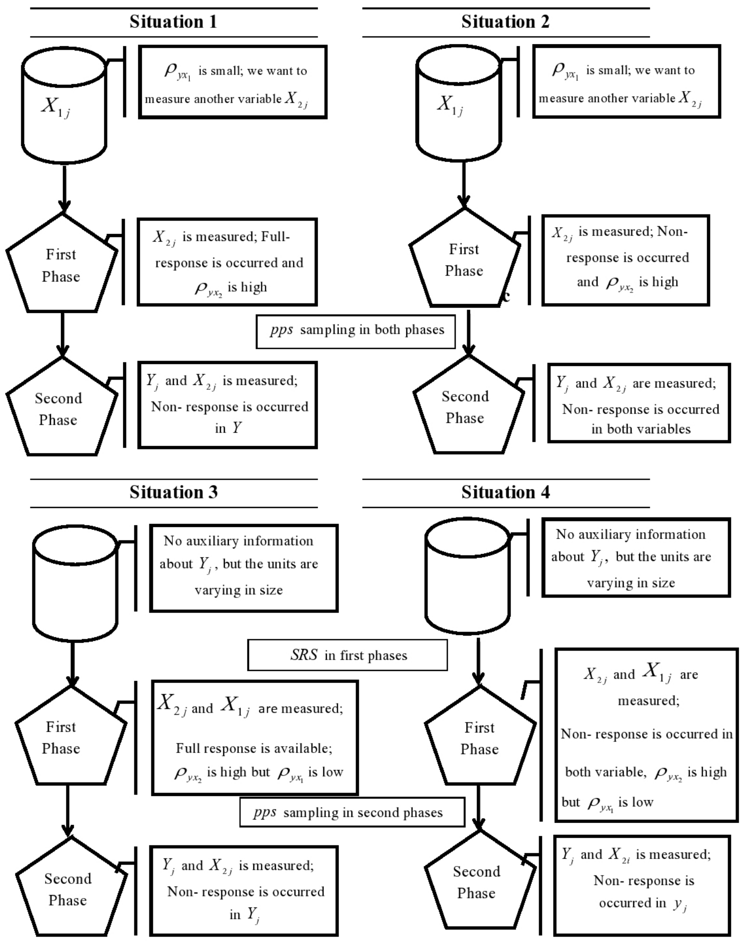

Figure 1. Illustartion of four possible situtations of non-response in two phase *pps* sampling

The rest of the study discusses the main points are as follow: In Section 3, we consider some traditional imputation procedure under two phase *pps* sampling. In Section 4, we proposed a modified class of estimators for imputing the missing values under two phase *pps* sampling. For practical application of the proposed estimators, numerical results are discussed comprehensively in Section 5 by considering different response rates in two phase *pps* sampling, given in Appendix. There are final remarks in Section 6.

2. Notations and expectations

Let $\bar{Y} = \sum_{j=1}^{N} Y_j/N$ and $\bar{X}_2 = \sum_{j=1}^{N} X_{2j}/N$ be the population mean of the study and the second auxiliary variable respectively. For evaluating the mathematical expressions for bias and mean squared error of the modified estimators for each of the specified situations, we define following useful notations under large sample approximation, as:

Situation 1. Following [14], let $u_j = y_j/(NP_j)$ and $v_{2j} = x_{2j}/(NP_j)$, where $P_j = X_{1j}/\sum_{j=1}^{N} X_{1j}$ and also let $\bar{v}_{2m}^* = \sum_{j=1}^{m} v_{2j}/m$ be the sample mean of the auxiliary information at first phase, $\bar{v}_{2n} = \sum_{j=1}^{n} v_{2j}/n$ and $\bar{u}_r = \sum_{j=1}^{r} u_j/r$ be the sample mean of the auxiliary variable and the study variable at second phase respectively.
Let

$$\zeta_0 = \frac{\bar{u}_r}{\bar{Y}} - 1, \quad \zeta_1 = \frac{\bar{v}_{2n}}{\bar{v}_{2m}^*} - 1, \quad \zeta_2 = \frac{\bar{v}_{2m}^*}{\bar{X}_2} - 1, \; E(\zeta_0) = E(\zeta_1) = E(\zeta_2) = 0.$$

up to the first order of approximation, we have

$$\begin{aligned} E(\zeta_0^2) =& r^{-1}C_u^2, \quad E(\zeta_1^2) = n^{-1}C_v^2 \quad E(\zeta_2^2) = m^{-1}C_v^2, E(\zeta_0\zeta_1) = n^{-1}\rho_{uv}C_uC_v, \\ E(\zeta_0\zeta_2) =& m^{-1}\rho_{uv}C_uC_v, \quad E(\zeta_1\zeta_2) = m^{-1}C_v^2. \end{aligned}$$

Situation 2. Let r' be the total number of respondents $(r' < m)$. So, $\bar{v}_{2r'}^* = \sum_{j=1}^{r'} v_{2j}/r'$ be the sample mean of the available auxiliary information at first phase. It is also assumed that r be the respondent units at second phase $(r < r')$. So, $\bar{v}_{2r} = \sum_{j=1}^{r} v_{2j}/r$ be the sample mean of the auxiliary variable at second phase.
Let

$$\zeta_1' = \frac{\bar{v}_{2r}}{\bar{v}_{2r'}^*} - 1, \quad \zeta_2' = \frac{\bar{v}_{2r'}^*}{\bar{X}_2} - 1, \quad E(\zeta_1') = E(\zeta_2') = 0.$$

up to the first order of approximation, we have

$$\begin{aligned} E(\zeta_1^{'2}) &= r^{-1}C_v^2 \quad E(\zeta_2^{'2}) = r'^{-1}C_v^2 \quad E(\zeta_0\zeta_1') = r^{-1}\rho_{uv}C_uC_v, \\ E(\zeta_0\zeta_2') &= r'^{-1}\rho_{uv}C_uC_v, \quad E(\zeta_1'\zeta_2') = r'^{-1}C_v^2 \end{aligned}$$

For the first two situations, we used following expressions are:

$$\begin{aligned} C_u^2 =& \frac{\sigma_u^2}{\bar{Y}^2}, \quad C_v^2 = \frac{\sigma_v^2}{\bar{X}^2}, \quad \rho_{uv} = \frac{\sigma_{uv}}{\sigma_u\sigma_v}, nVar(\bar{y}) = \sigma_u^2 = \sum_{j=1}^{N} P_i(u_j - \bar{Y})^2, \\ \sigma_v^2 =& \sum_{j=1}^{N} P_i(v_{2j} - \bar{X})^2, \rho_{uv} = \frac{1}{\sigma_v\sigma_u}\sum_{j=1}^{N} P_i(v_{2j} - \bar{X})(u_j - \bar{Y}). \end{aligned}$$

Situation 3. Let $u_j^* = y_j/(mP_j^*)$ and $v_{2j}^* = x_{2j}/(mP_j^*)$, where $P_j^* = x_{1j}/\sum_{j=1}^{m} x_{1j}$ and also let $\bar{x}_{2m}^{**} = \sum_{j=1}^{m} x_{2j}/m$ be the sample mean of the auxiliary information which are selected by SRS at first phase. It is also assume that $\bar{v}_{2n}^* = \sum_{j=1}^{n} v_{2j}^*/n$ and $\bar{u}_r^* = \sum_{j=1}^{r} u_j^*/r$ be the sample mean of X_{2j} and Y_j at the second phase respectively.
Let

$$\zeta_0' = \frac{\bar{u}_r^*}{\bar{X}_2} - 1, \zeta_1'' = \frac{\bar{v}_{2n}^*}{\bar{x}_{2m}^{**}} - 1, \quad \zeta_2'' = \frac{\bar{x}_{2m}^{**}}{\bar{X}_2} - 1, \; E(\zeta_0') = E(\zeta_1'') = E(\zeta_2'') = 0.$$

up to the first order of approximation, we have

$$E(\zeta_0^{'2}) = n^{-1}C_u^{2^*}, \quad E(\zeta_1^{''2}) = n^{-1}C_v^{2^*} \quad E(\zeta_2^{''2}) = m^{-1}C_v^{2^*},$$
$$E(\zeta_0^{'}\zeta_1^{''}) = n^{-1}\rho_{uv}^{*}C_u^{*}C_v^{*}, \quad E(\zeta_0^{'}\zeta_2^{''}) = m^{-1}\rho_{uv}^{*}C_u^{*}C_v^{*}, \quad E(\zeta_1^{''}\zeta_2^{''}) = m^{-1}C_v^{2^*}$$

where

$$C_u^{2^*} = \frac{\sigma_u^{2^*}}{\bar{Y}^2}, \quad C_v^{2^*} = \frac{\sigma_v^{2^*}}{\bar{X}^2}, \quad \rho_{uv}^{*} = \frac{\sigma_{uv}^{*}}{\sigma_u^{*}\sigma_v^{*}}, \quad mVar(\bar{y}) = \sigma_u^{2^*} = \sum\nolimits_{j=1}^{m} P_i^{*}(u_j^{*} - \bar{Y})^2 + \sigma_y^2,$$
$$\sigma_v^{2^*} = \sum\nolimits_{j=1}^{m} P_i^{*}(v_{2j}^{*} - \bar{X})^2 + \sigma_x^2, \quad \rho_{uv}^{*} = \frac{1}{\sigma_u^{*}\sigma_v^{*}}\sum\nolimits_{j=1}^{m} P_i^{*}(v_{2j}^{*} - \bar{X})(u_j^{*} - \bar{Y}),$$

Situation 4. Let $u_j^{**} = y_j/(r^{'}P_j^{**})$ and $v_{2j}^{**} = x_{2j}/(r^{'}P_j^{**})$, where $P_j^{**} = x_{1j}/\sum_{j=1}^{r^{'}} x_{1j}$ and also let $\bar{x}_{2r^{'}}^{**} = \sum_{j=1}^{r^{'}} x_{2j}/r^{'}$ be the sample mean of the auxiliary information at first phase, $\bar{v}_{2r}^{*} = \sum_{j=1}^{r} v_{2j}^{**}/n$ and $\bar{u}_r^{**} = \sum_{j=1}^{r} u_j^{**}/r$ be the sample mean of X_{2j} and Y_j at the second phase respectively.
Let

$$\zeta_0^{''} = \frac{\bar{u}_r^{**}}{\bar{Y}} - 1, \ \zeta_1^{'''} = \frac{\bar{v}_{2r}^{*}}{\bar{x}_{2r^{'}}^{**}} - 1, \ \zeta_2^{'''} = \frac{\bar{x}_{2r^{'}}^{**}}{\bar{X}_2} - 1, \ E(\zeta_0^{''}) = E(\zeta_1^{'''}) = E(\zeta_2^{'''}) = 0.$$

up to first order of approximation, we have

$$E(\zeta_0^{''2}) = r^{-1}C_u^{**2}, \quad E(\zeta_1^{'''2}) = r^{-1}C_v^{**2} \quad E(\zeta_2^{'''2}) = r^{'-1}C_v^{**2},$$
$$E(\zeta_0^{''}\zeta_1^{'''}) = r^{-1}\rho_{uv}^{**}C_u^{**}C_v^{**}, \quad E(\zeta_0^{''}\zeta_2^{'''}) = r^{'-1}\rho_{uv}^{**}C_u^{**}C_v^{**}, \quad E(\zeta_1^{'''}\zeta_2^{'''}) = r^{'-1}C_v^{**2},$$

where

$$C_u^{**2} = \frac{\sigma_u^{**2}}{\bar{Y}^2}, \quad C_v^{**2} = \frac{\sigma_v^{**2}}{\bar{X}^2}, \quad \rho_{uv}^{**} = \frac{\sigma_{uv}^{**}}{\sigma_u^{**}\sigma_v^{**}},$$
$$r^{'}Var(\bar{y}) = \sigma_u^{**2} = \sum\nolimits_{j=1}^{r^{'}} P_i^{**}(u_j^{**} - \bar{Y})^2 + \sigma_y^2, \sigma_v^{**2} = \sum\nolimits_{j=1}^{r^{'}} P_i^{**}(v_{2j}^{**} - \bar{X})^2 + \sigma_x^2,$$
$$\rho_{uv}^{**} = \frac{1}{\sigma_u^{**}\sigma_v^{**}}\sum\nolimits_{j=1}^{r^{'}} P_i^{**}(v_{2j}^{*} - \bar{X})(u_j^{*} - \bar{Y}), r < n < r^{'} < m.$$

3. Available imputation method in literature

For the above mentioned situations, we reformulate [1] imputation procedure under two phase *pps* sampling scheme, as:

3.1. Situation 1

The missing values are imputed as by using the mean imputation procedure as

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon G \\ \bar{u}_r & \text{if } j\epsilon G^c \end{cases} \tag{3.1}$$

The point estimator for population mean is given by:

$$\hat{\bar{Y}}_M^{(1)} = \frac{1}{n}\left\{\sum_{j=1}^{r} u_j + \sum_{j=1}^{n-r} u_j\right\} = \bar{u}_r.$$

The variance of $\hat{\bar{Y}}_M^{(1)}$ is given by

$$Var(\hat{\bar{Y}}_M^{(1)}) \cong r^{-1}\bar{Y}^2 C_u^2.$$

We rewrite the ratio estimators for imputing the missing values under two phase *pps* sampling, as:

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j \epsilon\, G \\ \frac{1}{1-g_1}\left[\frac{\bar{u}_r}{\bar{v}_{2n}}\bar{v}^*_{2m} - g_1\bar{u}_r\right] & \text{if } j \epsilon\, G^c \end{cases} \tag{3.2}$$

where $g_1 = \frac{r}{n}$. The point estimator for the given procedure in (3.2) is given as:

$$\hat{\bar{Y}}^{(1)}_R = \frac{\bar{u}_r}{\bar{v}_{2n}}\bar{v}^*_{2m}.$$

The bias and mean squared error are given by

$$Bias(\hat{\bar{Y}}^{(1)}_R) \cong \Pi_{nm}\,\bar{Y}\Big(C_v^2 - \rho_{uv}C_uC_v\Big)$$

and

$$MSE(\hat{\bar{Y}}^{(1)}_R) \cong r^{-1}\bar{Y}^2C_u^2 + \Pi_{nm}\,\bar{Y}^2\Big(C_v^2 - 2\rho_{uv}C_uC_v\Big),$$

where $\Pi_{nm} = \Big(\frac{1}{n} - \frac{1}{m}\Big)$.

3.2. Situation 2

For the second situation, the imputation procedures are defined as
The mean estimator is defined as

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j \epsilon\, G \\ \bar{u}_r & \text{if } j \epsilon\, G^c \end{cases} \tag{3.3}$$

The point estimator for population mean ($\bar{Y}$) is given by:

$$\hat{\bar{Y}}^{(2)}_M = \frac{1}{n}\Bigg\{\sum_{j=1}^{r} u_j + \sum_{j=1}^{n-r} u_j\Bigg\} = \bar{u}_r.$$

The variance of the mean estimator is

$$Var(\hat{\bar{Y}}^{(2)}_M) \cong r^{-1}\bar{Y}^2C_u^2.$$

The ratio estimators for imputing the missing values, is given as:

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j \epsilon\, G \\ \frac{1}{1-g_1}\left[\frac{\bar{u}_r}{\bar{v}_{2r}}\bar{v}^*_{2r'} - g_1\bar{u}_r\right] & \text{if } j \epsilon\, G^c \end{cases} \tag{3.4}$$

The point estimator for the given procedure in (3.4) is

$$\hat{\bar{Y}}^{(2)}_R = \frac{\bar{u}_r}{\bar{v}_{2r}}\bar{v}^*_{2r'}.$$

The bias and mean square error of $\hat{\bar{Y}}^{(2)}_R$ are

$$Bias(\hat{\bar{Y}}^{(2)}_R) \cong \Pi_{rr'}\,\bar{Y}\Big(C_v^2 - \rho_{uv}C_uC_v\Big)$$

and

$$MSE(\hat{\bar{Y}}^{(2)}_R) \cong r^{-1}\bar{Y}^2C_u^2 + \Pi_{rr'}\,\bar{Y}^2\Big(C_v^2 - 2\rho_{uv}C_uC_v\Big),$$

where $\Pi_{rr'} = \Big(\frac{1}{r} - \frac{1}{r'}\Big)$.

3.3. Situation 3

- Under mean method of imputation, the missing values are imputed as

$$\hat{Y}_j = \begin{cases} u_j^* & \text{if } j \epsilon G \\ \bar{u}_r^* & \text{if } j \epsilon G^c \end{cases} \tag{3.5}$$

The point estimator is given as

$$\hat{\bar{Y}}_M^{(3)} = \frac{1}{n}\left\{ \sum_{j=1}^{r} u_j + \sum_{j=1}^{n-r} u_j \right\} = \bar{u}_r^*.$$

The variance of $\hat{\bar{Y}}_M^{(3)}$ is given by

$$Var(\hat{\bar{Y}}_M^{(3)}) \cong r^{-1}\bar{Y}^2 C_u^{2*}.$$

The ratio estimators for imputing the missing values, is as:

$$\hat{Y}_j = \begin{cases} u_j^* & \text{if } j \epsilon G \\ \frac{1}{1-g_1}\left[\frac{\bar{u}_r^*}{\bar{v}_{2n}^*}\bar{x}_{2m}^{**} - g_1 \bar{u}_r^* \right] & \text{if } j \epsilon G^c \end{cases} \tag{3.6}$$

The point estimator for the strategy in given (3.6) is given by

$$\hat{\bar{Y}}_R^{(3)} = \frac{\bar{u}_r^*}{\bar{v}_{2n}^*}\bar{x}_{2m}^{**}.$$

The bias and mean square error of $\hat{\bar{Y}}_R^{(3)}$ is

$$Bias(\hat{\bar{Y}}_R^{(3)}) \cong \Pi_{nm}\,\bar{Y}\Big(C_v^{*2} - \rho_{uv}^* C_u^* C_v^*\Big)$$

and

$$MSE(\hat{\bar{Y}}_R^{(3)}) \cong r^{-1}\bar{Y}^2 C_u^{*2} + \Pi_{nm}\,\bar{Y}^2\Big(C_v^{*2} - 2\rho_{uv}^* C_u^* C_v^*\Big)$$

3.4. Situation 4

The average imputation procedure is defined as

$$\hat{Y}_j = \begin{cases} u_j^{**} & \text{if } j \epsilon G \\ \bar{u}_r^{**} & \text{if } j \epsilon G^c \end{cases} \tag{3.7}$$

The point estimator for population mean ($\bar{Y}$) is given by:

$$\hat{\bar{Y}}_M^{(4)} = \frac{1}{n}\left\{ \sum_{j=1}^{r} Y_j + \sum_{j=1}^{n-r} Y_j \right\} = \bar{u}_r^{**}.$$

The variance of the mean imputation procedure is given by

$$Var(\hat{\bar{Y}}_M^{(4)}) \cong r^{-1}\bar{Y}^2 C_u^{**2}.$$

We rewrite the ratio estimators for imputing the missing values, is as:

$$\hat{Y}_j = \begin{cases} u_j^{**} & \text{if } j \epsilon G \\ \frac{1}{1-g_1}\left[\frac{\bar{u}_r^{**}}{\bar{v}_{2r}^*}\bar{x}_{2r'}^{**} - g_1 \bar{u}_r^{**} \right] & \text{if } j \epsilon G^c \end{cases} \tag{3.8}$$

The point estimator for the given procedure in (3.8) is given as:

$$\hat{\bar{Y}}_R^{(4)} = \frac{\bar{u}_r^{**}}{\bar{v}_{2r}^*}\bar{x}_{2r'}^{**}.$$

The bias and mean square error of $\hat{\bar{Y}}_{R}^{(4)}$ is

$$Bias(\hat{\bar{Y}}_{R}^{(4)}) \cong \Pi_{rr'}\bar{Y}\Big(C_v^{**2} - \rho_{uv}^{**}C_u^{**}C_v^{**}\Big)$$

and

$$MSE(\hat{\bar{Y}}_{R}^{(4)}) \cong r^{-1}\bar{Y}^2C_u^{**2} + \Pi_{rr'}\,\bar{Y}^2\Big(C_v^{**2} - 2\rho_{uv}^{**}C_u^{**}C_v^{**}\Big).$$

4. Modified imputation procedures

In this section, we modified ratio type estimators for imputing missing values that could be occurred in two phase *pps* sampling. The estimation of population parameters is quite laborious when the complete information is not known. Especially, if the sample units are varying in size, then the traditional sampling procedures are not effective for selecting s from Ω. In such situation, *pps* sampling is an decent procedure for the selection of s by the proper use of supplementary information. Under *pps* sampling scheme, the estimation or inference of the population parameters is more credible and reliable as compared to traditional sample selection procedures, when the units are varying in size.

In many real life situations like in economics and other social science studies, where population units are varying in size and we have no auxiliary information in hand for the selection of s from Ω, then multi-phase sampling is a well known procedure, which provides suitable auxiliary information regarding study variable prior to observing it. Behind this argument, we consider the combine version of two phase and *pps* sampling schemes for the estimation of finite population mean, when the units are varying in size and we have no auxiliary information is in hand. For the detailed consideration of missing values, we defined four different modified imputation procedure but seems to be similar for the estimation of population mean in two phase *pps* sampling. The imputation procedures for each of the previously defined situations in 1.1 is given as follow:

4.1. Situation 1: Full information on X_2 is available at first phase and X_1 is known in advance

■

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon\, G \\ \frac{1}{(1-g_1)}\Big[\Delta_1\bar{u}_r\frac{\bar{v}_{2m}^*}{\bar{v}_{2n}} - g_1\bar{u}_r\Big] & \text{if } j\epsilon\, G^c \end{cases} \tag{4.1}$$

where Δ_1 is the suitably chosen constant by minimizing the resultant mean squared error. The point estimator for the population mean is defined as:

$$\hat{\bar{Y}}_1^{(1)} = \Delta_1\bar{u}_r\frac{\bar{v}_{2m}^*}{\bar{v}_{2n}}$$

Rewriting $\hat{\bar{Y}}_1^{(1)}$ in term of errors, we have

$$\hat{\bar{Y}}_1^{(1)} = \Delta_1\bar{Y}\Big(1 - \zeta_1 + \zeta_1^2 + \zeta_2 - \zeta_1\zeta_2 + \zeta_0 - \zeta_0\zeta_1 + \zeta_0\zeta_2\Big)$$

Expanding and keeping terms up to first order of approximation, the bias and mean square error of $\hat{\bar{Y}}_1^{(1)}$ is given as

$$E(\hat{\bar{Y}}_1^{(1)} - \bar{Y}) \;\;\cong\;\; \bar{Y}\Big(\Delta_1 - 1\Big) + \Pi_{nm}\,\Delta_1\Big(C_v^2 - \rho_{uv}C_uC_v\Big)$$

and

$$MSE(\hat{\bar{Y}}_1^{(1)}) \cong \frac{1}{r}\bar{Y}^2\Delta_1^2C_u^2 - 2\,\Pi_{nm}\,\bar{Y}\Delta_1\Big\{\Big(\frac{3}{2}\Delta_1 - 1\Big)C_v^2 -2\Big(\Delta_1 - \frac{1}{2}\Big)\rho_{uv}C_uC_v\Big\} \tag{4.2}$$

The optimum value of Δ_1 is obtained by setting $\frac{\partial MSE(\hat{\bar{Y}}_1^{(1)})}{\partial \Delta_1} = 0$, as follow:

$$\Delta_{1(opt.)} = \frac{\Big(n^*C_v^2 - C_u\rho_{yx}n^*C_v - nm\Big)r}{3\,n^*rC_v^2 - 4\,rC_u\rho_{uv}n^*C_v - nm\Big(C_u^2 + r\Big)},$$

where $n^* = n - m$.
Substituting the optimum value of Δ_1 in (4.2), the minimum mean squared error of $\hat{\bar{Y}}_1^{(1)}$ is

$$MSE(\hat{\bar{Y}}_1^{(1)})_{min.} \cong \frac{\Big[C_v\Big\{n^*\Lambda + \Gamma C_v - 2\,nC_um\rho_{uv}\Big\}n^*r - n^2C_u^2m^2\Big]\bar{Y}^2}{3\Big[\Big\{n^*\Big(C_v^2 - \frac{4}{3}\,C_vC_u\rho_{uv}\Big) - \frac{1}{3}\,nm\Big\}r - \frac{1}{3}\,nC_u^2m\Big]nm},$$

where $\Lambda = C_v^3 - 2\,C_v^2C_u\rho_{uv}$ and $\Gamma = nC_u^2\rho_{uv}^2 + m\Big(n - C_u^2\rho_{uv}^2\Big)$.
■

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon\, G \\ \frac{1}{(1-g_1)}\Big[\bar{u}_r\Big(\frac{\bar{v}_{2m}^*}{\bar{v}_{2n}}\Big)^{\Delta_2} - g_1\bar{u}_r\Big] & \text{if } j\epsilon\, G^c \end{cases} \tag{4.3}$$

where Δ_2 is a suitably chosen constant. The point estimator is defined as

$$\hat{\bar{Y}}_2^{(1)} = \bar{u}_r\Big(\frac{\bar{x}_{2m}^*}{\bar{x}_{2n}}\Big)^{\Delta_2}$$

In term of error, $\hat{\bar{Y}}_2^{(1)}$ is rewritten as

$$\hat{\bar{Y}}_2^{(1)} = \bar{Y}\Big(1 - \Delta_2\zeta_1 - \frac{1}{2}\Delta_2(\Delta_2 - 1)\zeta_1^2 + \Delta\zeta_2 - \Delta_2^2\zeta_1\zeta_2 + \zeta_0 +\frac{1}{2}\Delta_2(\Delta_2 - 1)\zeta_2^2 - \Delta_2\zeta_0\zeta_1 + \Delta_2\zeta_0\zeta_2\Big)$$

Expanding and keeping terms up to first order of approximation, the bias and mean square error of $\hat{\bar{Y}}_2^{(1)}$ are given by

$$E(\hat{\bar{Y}}_2^{(1)} - \bar{Y}) \cong \bar{Y}\,\Pi_{nm}\,\Delta_2\Big\{\frac{1}{2}\Big(\Delta_2 + 1\Big)C_v^2 - \rho_{uv}C_uC_v\Big\}$$

and

$$MSE(\hat{\bar{Y}}_2^{(1)}) \cong \frac{1}{r}\bar{Y}^2C_u^2 + \Pi_{nm}\,\bar{Y}^2\Big(\Delta_2^2C_x^2 - 2\Delta_2\rho_{uv}C_uC_v\Big) \tag{4.4}$$

The optimum value of Δ_2 is obtained as $\frac{\partial MSE(\hat{\bar{Y}}_2^{(1)})}{\partial \Delta_2} = 0$, then

$$\Delta_{2(opt.)} = \frac{\rho_{uv}C_u}{C_v}$$

Substituting the optimum value of Δ_2 in (4.4), we set the minimum mean squared error of $\hat{\bar{Y}}_2^{(1)}$ as follow

$$MSE(\hat{\bar{Y}}_2^{(1)})_{min.} = \frac{1}{r}C_u^2\bar{Y}^2\Big\{\Big(1 - g_1\rho_{uv}^2\Big) + \lambda\rho_{uv}^2\Big\},$$

where $\lambda = \frac{r}{m}$.

■

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon G \\ \frac{1}{(1-g_1)}\left[\frac{\bar{u}_r\ \bar{v}^*_{2m}}{\Delta_3\bar{v}_{2n}+(1-\Delta_3)\bar{v}^*_{2m}} - g_1\bar{u}_r\right] & \text{if } j\epsilon G^c \end{cases} \tag{4.5}$$

where Δ_3 is a suitably chosen unknown value. The point estimator for the population mean is defined as:

$$\hat{\hat{Y}}_3^{(1)} = \frac{\bar{u}_r\ \bar{v}^*_{2m}}{\Delta_3\bar{v}_{2n} + (1-\Delta_3)\bar{v}^*_{2m}}$$

The $\hat{\hat{Y}}_3^{(1)}$ in term of errors can also be written as

$$\begin{aligned} \hat{\hat{Y}}_3^{(1)} &= \bar{Y}\Big(1 - \Delta_3\zeta_1 + \Delta_3\zeta_2 - \zeta_2 + \Delta_3^2\zeta_1^2 + \Delta_3^2\zeta_2^2 + \zeta_1^2 - 2\Delta_3^2\zeta_1\zeta_2 \\ &\quad +2\Delta_3\zeta_1\zeta_2 - 2\Delta_3\zeta_2^2 + \zeta_2 - \Delta_1\zeta_1\zeta_2 + \Delta_3\zeta_2^2 - \zeta_2^2 + \zeta_0 - \Delta_3\zeta_0\zeta_1 \\ &\quad +\Delta_3\zeta_0\zeta_2 - \zeta_0\zeta_2 + \zeta_0\zeta_2\Big) \end{aligned}$$

Expanding and keeping terms up to first order of approximation, the bias and mean square error of $\hat{\hat{Y}}_3^{(1)}$ are given by

$$E(\hat{\hat{Y}}_3^{(1)} - \bar{Y}) \cong \bar{Y}\Pi_{nm}\,\Delta_3\left(\Delta_3 C_v^2 - \rho_{uv}C_uC_v\right)$$

and

$$MSE(\hat{\hat{Y}}_3^{(1)}) \cong \frac{1}{r}\bar{Y}^2C_u^2 + \Pi_{nm}\bar{Y}^2\Big(\Delta_3^2C_v^2 - 2\Delta_3\rho_{uv}C_xC_v\Big) \tag{4.6}$$

The optimum value of Δ_3 is obtained as $\frac{\partial MSE(\hat{\hat{Y}}_3^{(1)})}{\partial\Delta_3} = 0$, then

$$\Delta_{3(opt.)} = \frac{\rho_{uv}C_u}{C_v}$$

Substituting the optimum value of Δ_3 in (4.6), the minimum mean squared error of $\hat{\hat{Y}}_3^{(1)}$ is

$$MSE(\hat{\hat{Y}}_3^{(1)})_{min.} = \frac{1}{r}C_u^2\bar{Y}^2\Big\{\Big(1 - g_1\rho_{uv}^2\Big) + \lambda\rho_{uv}^2\Big\}$$

■

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon G \\ \frac{1}{(1-g_1)}\left[\Delta_4\bar{u}_r + (1-\Delta_4)\bar{u}_r\left(\frac{\bar{v}^*_{2m}}{\bar{v}_{2n}}\right) - g_1\bar{u}_r\right] & \text{if } j\epsilon G^c \end{cases} \tag{4.7}$$

where Δ_4 is an unknown constant. The point estimator for the given procedure in (4.7) is defined as:

$$\hat{\hat{Y}}_4^{(1)} = \Delta_4\bar{u}_r + (1-\Delta_4)\bar{u}_r\left(\frac{\bar{v}^*_{2m}}{\bar{v}_{2n}}\right)$$

Rewriting $\hat{\hat{Y}}_4^{(1)}$ in term of error, we have

$$\begin{aligned} \hat{\hat{Y}}_4^{(1)} &= \Delta_4\bar{Y}(1+\zeta_0) + (1-\Delta_4)\bar{Y}\Bigg(1 - \zeta_1 + \zeta_1^2 + \zeta_2 - \zeta_1\zeta_2 + \zeta_0 \\ &\quad -\zeta_0\zeta_1 + \zeta_0\zeta_2\Bigg) \end{aligned}$$

expanding and keeping terms up to first order approximation, the bias and mean square error of $\hat{\hat{Y}}_4^{(1)}$ is given as

$$E(\hat{\hat{Y}}_4^{(1)} - \bar{Y}) \cong (\Delta_4 - 1)\bar{Y} + (1-\Delta_4)\,\bar{Y}\Pi_{nm}\left(C_v^2 - \rho_{uv}C_uC_v\right)$$

and

$$MSE(\hat{\bar{Y}}_4^{(1)}) \cong \frac{1}{r}\bar{Y}^2C_v^2 + \Big(\Delta_4 - 1\Big)\Pi_{nm}\bar{Y}^2\Big\{\Delta_4C_v^2 + 2\rho_{uv}C_vC_u\Big\} \tag{4.8}$$

The optimum value of Δ_4 is obtained as $\frac{\partial MSE(\hat{\bar{Y}}_4^{(1)})}{\partial \Delta_4} = 0$, then

$$\Delta_{4(opt.)} = 1 - \frac{C_u\rho_{uv}}{C_v}$$

Substituting the optimum value of Δ_4 in (4.8), the minimum mean squared error of $\hat{\bar{Y}}_4^{(1)}$ is

$$MSE(\hat{\bar{Y}}_4^{(1)})_{min.} = \frac{1}{r}C_u^2\bar{Y}^2\Big\{\Big(1 - g_1\rho_{uv}^2\Big) + \lambda\rho_{uv}^2\Big\}$$

4.2. Situation 2: Non-response is occurred in X_2 at first phase

■

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon\, G \\ \frac{1}{(1-g_1)}\Big[\omega_1\bar{u}_r\frac{\bar{v}_{2r}^*}{\bar{v}_{2r}} - g_1\bar{u}_r\Big] & \text{if } j\epsilon\, G^c \end{cases} \tag{4.9}$$

where ω_1 is a suitably chosen constant by minimizing the mean squared error. The point estimator for the population mean is defined as

$$\hat{\bar{Y}}_1^{(2)} = \omega_1\bar{u}_r\frac{\bar{v}_{2r}^*}{\bar{v}_{2r}}$$

Rewriting $\hat{\bar{Y}}_1^{(2)}$ in term of errors, we have

$$\hat{\bar{Y}}_1^{(2)} = \omega_1\bar{Y}\Big(1 - \zeta_1^{'} + \zeta_1^{'2} + \zeta_2^{'} - \zeta_1^{'}\zeta_2^{'} + \zeta_0 - \zeta_0\zeta_1^{'} + \zeta_0\zeta_2^{'}\Big)$$

Expanding and keeping terms up to first order of approximation, the bias and mean square error of $\hat{\bar{Y}}_1^{(2)}$ is given as

$$E(\hat{\bar{Y}}_1^{(2)} - \bar{Y}) \cong \bar{Y}(\omega_1 - 1) + \Pi_{rr'}\,\omega_1\Big(C_v^2 - \rho_{uv}C_uC_v\Big)$$

and

$$\begin{aligned} MSE(\hat{\bar{Y}}_1^{(2)}) \quad \cong \quad & \frac{1}{r}\bar{Y}^2\omega_1^2C_u^2 - 2\Pi_{rr'}\bar{Y}\omega_1\Big\{\Big(\frac{3}{2}\omega_1 - 1\Big)C_v^2 \\ & -(2\omega_1 - 1)\rho_{uv}C_uC_v\Big\} \end{aligned} \tag{4.10}$$

The optimum value of ω_1 is obtained by setting $\frac{\partial MSE(\hat{\bar{Y}}_1^{(2)})}{\partial \omega_1} = 0$, as follow

$$\omega_{1(opt.)} = \frac{r^*C_v^2 - C_u\rho_{yx}r^*C_v - rr^{'}}{3r^*{C_v}^2 - 4\,C_u\rho_{uv}r^*C_v - r^{'}\Big({C_u}^2 + r\Big)},$$

where $r^* = r - r^{'}$.

Substituting the optimum value of ω_1 in (4.10), the minimum mean squared error of $\hat{\bar{Y}}_1^{(2)}$ is

$$MSE(\hat{\bar{Y}}_1^{(2)})_{min.} \cong \frac{\Big[r^*C_v\Big\{C_v\Big(r^*C_v\Lambda^{'} + \Gamma^{'}\Big) - 2\,rC_ur^{'}\rho_{uv}\Big\} - rC_u^2{r^{'}}^2\Big]\bar{Y}^2}{3\Big\{r^*C_v^2 - \frac{4}{3}\,C_u\rho_{uv}r^*C_v - \frac{1}{3}\,r^{'}\Big(C_u^2 + r\Big)\Big\}r^{'}r},$$

where $\Lambda^{'} = C_v - 2\,C_v\rho_{uv}$ and $\Gamma^{'} = rC_u^2\rho_{uv}^2 + \Big(r - C_u^2\rho_{uv}^2\Big)r^{'}$.

■

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon\, G \\ \frac{1}{(1-g_1)}\left[\bar{u}_r\left(\frac{\bar{v}_{2r}^*}{\bar{v}_{2r}}\right)^{\omega_2} - g_1\bar{u}_r\right] & \text{if } j\epsilon\, G^c \end{cases} \tag{4.11}$$

where ω_2 is a suitably chosen constant. The point estimator is defined as:

$$\hat{\hat{Y}}_2^{(2)} = \bar{u}_r\left(\frac{\bar{v}_{2m}^*}{\bar{v}_{2n}}\right)^{\omega_2}$$

The $\hat{\hat{Y}}_2^{(2)}$ in term of errors can also be written as

$$\begin{aligned}\hat{\hat{Y}}_2^{(2)} &= \bar{Y}\Big(1 - \omega_2\zeta_1^{'} - \frac{\omega_2(\omega_2-1)}{2}\zeta_1^{'2} + \omega\zeta_2^{'} - \omega_2^2\zeta_1^{'}\zeta_2^{'} + \zeta_0 \\ &\quad + \frac{\omega_2(\omega_2-1)}{2}\zeta_2^{'2} - \omega_2\zeta_0\zeta_1^{'} + \omega_2\zeta_0\zeta_2^{'}\Big)\end{aligned}$$

Keeping terms up to first order, the bias and mean square error of $\hat{\hat{Y}}_2^{(2)}$ is given as

$$E(\hat{\hat{Y}}_2^{(2)} - \bar{Y}) \;\cong\; \bar{Y}\,\Pi_{rr'}\,\omega_2\Big\{\frac{1}{2}\Big(\omega_2+1\Big)C_v^2 - \rho_{uv}C_uC_v\Big\}$$

and

$$MSE(\hat{\hat{Y}}_2^{(2)}) \;\cong\; \frac{1}{r}\bar{Y}^2C_u^2 + \Pi_{rr'}\,\bar{Y}^2\Big(\omega_2^2C_x^2 - \omega_2\rho_{uv}C_uC_v\Big) \tag{4.12}$$

The optimum value of ω_2 is as $\frac{\partial MSE(\hat{\hat{Y}}_2^{(2)})}{\partial\omega_2} = 0$, then

$$\omega_{2(opt.)} = \frac{\rho_{uv}C_u}{C_v}$$

Substituting the optimum value of ω_2 in (4.12), the minimum mean squared error of $\hat{\hat{Y}}_2^{(2)}$ is

$$MSE(\hat{\hat{Y}}_2^{(2)})_{min.} \cong \frac{1}{r}\bar{Y}^2C_u^2\left(1 + \lambda^{'}\rho_{uv}^2\right),$$

where $\lambda^{'} = \frac{r^*}{r^{'}}$.

■

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon\, G \\ \frac{1}{(1-g_1)}\left[\frac{\bar{u}_r\ \bar{v}_{2r'}^*}{\omega_3\bar{v}_{2r}+(1-\omega_3)\bar{v}_{2r'}^*} - g_1\bar{u}_r\right] & \text{if } j\epsilon\, G^c \end{cases} \tag{4.13}$$

where ω_3 is a suitably chosen unknown value. The point estimator for the population mean is defined as:

$$\hat{\hat{Y}}_3^{(2)} = \frac{\bar{u}_r\ \bar{v}_{2r'}^*}{\omega_3\bar{v}_{2r} + (1-\omega_3)\bar{v}_{2r'}^*}$$

In term of error, the $\hat{\hat{Y}}_3^{(2)}$ can be rewrite as

$$\begin{aligned}\hat{\hat{Y}}_3^{(2)} &= \bar{Y}\Big(1 - \omega_3\zeta_1 + \omega_3\zeta_1 - \zeta_2 + \omega_3^2\zeta_1^2 + \omega_3^2\zeta_2^2 + \zeta_1^2 - 2\omega_3^2\zeta_1\zeta_2 \\ &\quad -2\omega_3\zeta_1\omega_2 + 2\omega_3\zeta_2^2 + \zeta_2 - \omega_1\zeta_1\zeta_2 + \omega_3\zeta_2^2 - \zeta_2^2 + \zeta_0 - \omega_3\zeta_0\zeta_1 \\ &\quad +\omega_3\zeta_0\zeta_2 - \zeta_0\zeta_2 + \zeta_0\zeta_2\Big)\end{aligned}$$

Expanding terms up to first order of approximation, the bias and mean square error of $\hat{\hat{Y}}_3^{(2)}$ is given as

$$E(\hat{\hat{Y}}_3^{(2)} - \bar{Y}) \;\cong\; \bar{Y}\Pi_{rr'}\,\omega_3\left(\omega_3C_v^2 - \rho_{uv}C_uC_v\right)$$

and

$$MSE(\hat{\bar{Y}}_3^{(2)}) \cong \frac{1}{r}\bar{Y}^2 C_u^2 + \Pi_{rr'}\bar{Y}^2\Big(\omega_3^2 C_v^2 - 2\omega_3\rho_{uv}C_uC_v\Big) \tag{4.14}$$

The ω_3 is obtained as $\frac{\partial MSE(\hat{\bar{Y}}_3^{(2)})}{\partial\omega_3} = 0$, then

$$\omega_{3(opt.)} = \frac{C_u\rho_{uv}}{C_v}$$

Substituting the optimum value of ω_3 in (4.14), the minimum mean squared error of $\hat{\bar{Y}}_3^{(2)}$ is

$$MSE(\hat{\bar{Y}}_3^{(2)})_{min.} \cong \frac{1}{r}\bar{Y}^2 C_u^2\left(1 + \lambda^{'}\rho_{uv}^2\right)$$

■

$$\hat{Y}_j = \begin{cases} u_j & \text{if } j\epsilon\, G \\ \frac{1}{(1-g_1)}\left[\omega_4\bar{u}_r + (1-\omega_4)\bar{u}_r\left(\frac{\bar{v}_{2r}^*}{\bar{v}_{2r}}\right) - g_1\bar{u}_r\right] & \text{if } j\epsilon\, G^c \end{cases} \tag{4.15}$$

where ω_4 is an unknown constant value. The point estimator for the given procedure in (4.15) is defined as:

$$\hat{\bar{Y}}_4^{(2)} = \omega_4\bar{u}_r + (1-\omega_4)\bar{u}_r\left(\frac{\bar{v}_{2m}^*}{\bar{v}_{2n}}\right)$$

Rewriting $\hat{\bar{Y}}_4^{(2)}$ in term of error, we have

$$\begin{aligned}\hat{\bar{Y}}_4^{(2)} &= \bar{Y}\omega_4(1+\zeta_0) + (1-\omega_4)\bar{Y}\Big(1 - \zeta_1^{'} + \zeta_1^{'2} + \zeta_2^{'} - \zeta_1^{'}\zeta_2^{'} + \zeta_0 \\ &\quad -\omega_2\zeta_0\zeta_1^{'} + \omega_2\zeta_0\zeta_2^{'}\Big)\end{aligned}$$

Keeping terms up to first order of approximation, the bias and mean square error of $\hat{\bar{Y}}_4^{(2)}$ is given as

$$E(\hat{\bar{Y}}_4^{(2)} - \bar{Y}) \cong (\omega_4 - 1)\bar{Y} + (1-\omega_4)\bar{Y}\Pi_{rr'}\left(C_v^2 - \rho_{uv}C_uC_v\right)$$

and

$$MSE(\hat{\bar{Y}}_4^{(2)}) \cong \frac{1}{r}\bar{Y}^2 C_v^2 + \Pi_{rr'}\Big\{\omega_4\Big(\omega_4 - 1\Big)C_v^2 + 2\Big(\omega_4 - 1\Big)\rho_{uv}C_vC_u\Big\} \tag{4.16}$$

The ω_4 is obtained as $\frac{\partial MSE(\hat{\bar{Y}}_4^{(2)})}{\partial\omega_4} = 0$, then

$$\omega_{4(opt.)} = 1 - \frac{C_u\rho_{uv}}{C_v}$$

Substituting the optimum value of ω_4 in (4.16), the minimum mean squared error of $\hat{\bar{Y}}_4^{(2)}$ is

$$MSE(\hat{\bar{Y}}_4^{(2)})_{min.} \cong \frac{1}{r}\bar{Y}^2 C_u^2\left(1 + \lambda^{'}\rho_{uv}^2\right)$$

4.3. Situation 3: Response on X_1 and X_2 is obtained from First Phase

■

$$\hat{Y}_j = \begin{cases} u_j^* & \text{if } j \epsilon\, G \\ \frac{1}{(1-g_1)}\left[\varphi_1 \bar{u}_r \frac{\bar{x}_{2m}^{**}}{\bar{v}_{2n}^{*}} - g_1 \bar{u}_r^*\right] & \text{if } j \epsilon\, G^c \end{cases} \tag{4.17}$$

where φ_1 is a suitably chosen constant by minimizing the resultant mean squared error. The point estimator for the population mean is defined as:

$$\hat{\bar{Y}}_1^{(3)} = \varphi_1 \bar{u}_r^* \frac{\bar{x}_{2m}^{**}}{\bar{v}_{2n}^{*}}$$

Rewriting $\hat{\bar{Y}}_1^{(3)}$ in term of error as

$$\hat{\bar{Y}}_1^{(3)} = \varphi_1 \bar{Y}\Big(1 - \zeta_1^{''} + \zeta_1^{''2} + \zeta_2^{''} - \zeta_1^{''}\zeta_2^{''} + \zeta_0^{'} - \zeta_0\zeta_1 + \zeta_0^{'}\zeta_2^{''}\Big)$$

Expanding and keeping terms up to first order of approximation, the bias and mean square error of $\hat{\bar{Y}}_1^{(3)}$ is given as

$$E(\hat{\bar{Y}}_1^{(3)} - \bar{Y}) \;\cong\; \bar{Y}(\varphi_1 - 1) + \Pi_{nm}\,\bar{Y}\,\varphi_1\Big(C_v^{*2} - \rho_{uv}^{*} C_u^{*} C_v^{*}\Big)$$

and

$$\begin{aligned} MSE(\hat{\bar{Y}}_1^{(3)}) \;\cong\; & \frac{1}{r}\bar{Y}^2\Delta_1^2 C_u^{*2} - 2\Pi_{nm}\bar{Y}^2\varphi_1\Big\{\Big(\frac{3}{2}\varphi_1 - 1\Big)C_v^{*2} \\ & -(2\varphi_1 - 1)\rho_{uv}^{*} C_u^{*} C_v^{*}\Big\} \end{aligned} \tag{4.18}$$

The optimum values of φ_1 is obtained as $\frac{\partial MSE(\hat{\bar{Y}}_1^{(3)})}{\partial \varphi_1} = 0$, then

$$\varphi_{1(opt.)} = \frac{\Big(n^* C_v^{*2} - C_u^{*}\rho_{yx}^{*} n^* C_v^{*} - nm\Big)r}{3\,n^* r C_v^{*2} - 4\,r C_u^{*}\rho_{uv}^{*} n^* C_v^{*} - nm\,(C_u^{*2} + r)} \tag{4.19}$$

Substituting (4.19) in (4.18), the minimum mean squared error of $\hat{\bar{Y}}_1^{(3)}$ is

$$MSE(\hat{\bar{Y}}_1^{(3)})_{min.} \cong \frac{\Big[C_v^{*}\Big\{n^*\Lambda + \Gamma C_v^{*} - 2\,n C_u^{*} m \rho_{uv}^{*}\Big\}n^* r - n^2 C_u^{*2} m^2\Big]\bar{Y}^2}{3\Big[\Big\{n^*\Big(C_v^{*2} - \frac{4}{3}\,C_v^{*} C_u^{*}\rho_{uv}^{*}\Big) - \frac{1}{3}\,nm\Big\}r - \frac{1}{3}\,n C_u^{*2} m\Big]nm}$$

■

$$\hat{Y}_j = \begin{cases} u_j^* & \text{if } j \epsilon\, G \\ \frac{1}{(1-g_1)}\left[\bar{u}_r^*\left(\frac{\bar{x}_{2m}^{**}}{\bar{v}_{2n}^{*}}\right)^{\varphi_2} - g_1 \bar{u}_r^*\right] & \text{if } j \epsilon\, G^c \end{cases} \tag{4.20}$$

where φ_2 is the suitably chosen constant value. The point estimator is defined as:

$$\hat{\bar{Y}}_2^{(3)} = \bar{u}_r^* \left(\frac{\bar{x}_{2m}^{*}}{\bar{v}_{2n}^{**}}\right)^{\varphi_2}$$

Rewriting $\hat{\bar{Y}}_2^{(3)}$ in term of error, we have

$$\begin{aligned} \hat{\bar{Y}}_2^{(3)} \;=\; & \bar{Y}\Big(1 - \varphi_2\zeta_1^{''} - \frac{1}{2}\varphi_2(\varphi_2 - 1)\zeta_1^{''2} + \varphi\zeta_2^{''} - \varphi_2^2\zeta_1^{''}\zeta_2^{''} + \zeta_0^{'} \\ & + \frac{1}{2}\varphi_2(\varphi_2 - 1)\zeta_2^{''2} - \varphi_2\zeta_0^{'}\zeta_1^{''} + \varphi_2\zeta_0^{'}\zeta_2^{''}\Big) \end{aligned}$$

Expanding and keeping terms up to first order of approximation, the bias and mean square error of $\hat{\bar{Y}}_2^{(3)}$ is given as

$$E(\hat{\bar{Y}}_2^{(3)} - \bar{Y}) \;\cong\; \bar{Y}\Pi_{nm}\,\varphi_2\Big\{\frac{1}{2}\Big(\varphi_2 + 1\Big)C_v^{*2} - \rho_{uv}^{*} C_u^{*} C_v^{*}\Big\}$$

and

$$MSE(\hat{\bar{Y}}_2^{(3)}) \cong \frac{1}{r}\bar{Y}^2 C_u^{*2} + \Pi_{nm}\bar{Y}^2\left(\varphi_2^2 C_v^{*2} - \varphi_2\rho_{uv}^* C_u^* C_v^*\right) \quad (4.21)$$

The optimum value of φ_2 is obtained as $\frac{\partial MSE(\hat{\bar{Y}}_2^{(3)})}{\partial\varphi_2} = 0$, then

$$\varphi_{2(opt.)} = \frac{\rho_{uv}^* C_u^*}{C_v^*} \quad (4.22)$$

Substituting (4.22) in (4.21), the minimum mean squared error of $\hat{\bar{Y}}_2^{(3)}$ is

$$MSE(\hat{\bar{Y}}_2^{(3)})_{min.} = \frac{1}{r}C_u^{*2}\bar{Y}^2\Big\{\Big(1 - g_1\rho_{uv}^{*2}\Big) + \lambda\rho_{uv}^{*2}\Big\}$$

■

$$\hat{Y}_j = \begin{cases} u_j^* & \text{if } j\epsilon G \\ \frac{1}{(1-g_1)}\left[\frac{\bar{u}_r^*\ \bar{x}_{2m}^{**}}{\varphi_3\bar{v}_{2n}^* + (1-\varphi_3)\bar{x}_{2m}^{**}} - g_1\bar{u}_r^*\right] & \text{if } j\epsilon G^c \end{cases} \quad (4.23)$$

where φ_3 is the suitably chosen unknown value. The point estimator for the population mean is defined as:

$$\hat{\bar{Y}}_3^{(3)} = \frac{\bar{u}_r^*\ \bar{x}_{2m}^{**}}{\varphi_3\bar{v}_{2n}^* + (1-\varphi_3)\bar{x}_{2m}^{**}} \quad (4.24)$$

Rewriting (4.24) in term of error, we have

$$\begin{aligned}\hat{\bar{Y}}_3^{(3)} \quad = \quad & \varphi_3\bar{Y}\Big(1 - \varphi_3\zeta_1^{''} + \varphi_3\zeta_1^{''} - \zeta_2{''} + \varphi_3^2\zeta_1^{''2} + \varphi_3^2\zeta_2^{''2} + \zeta_1^{''2} - 2\varphi_3^2\zeta_1^{''}\zeta_2^{''} \\ & -2\varphi_3\zeta_1^{''}\zeta_2^{''} + 2\varphi_3\zeta_2^{''2} + \zeta_2^{''} - \varphi_1\zeta_1^{''}\zeta_2^{''} + \Delta_3\zeta_2^{''2} - \zeta_2^{''2} + \zeta_0^{'} - \Delta_3\zeta_0^{'}\zeta_1^{''} \\ & +\varphi_3\zeta_0^{'}\zeta_2^{''} - \zeta_0^{'}\zeta_2^{''} + \zeta_0^{'}\zeta_2^{''}\Big)\end{aligned}$$

keeping terms up to first order of approximation, the bias and mean square error of $\hat{\bar{Y}}_3^{(3)}$ is given as

$$E(\hat{\bar{Y}}_3^{(3)} - \bar{Y}) \cong \Pi_{nm}\bar{Y}\ \varphi_3\left(\varphi_3 C_v^{*2} - \rho_{uv}^* C_u^* C_v^*\right)$$

and

$$MSE(\hat{\bar{Y}}_3^{(3)}) \cong \frac{1}{r}\bar{Y}^2 C_u^{*2} + \Pi_{nm}\bar{Y}^2\Big(\varphi_3^2 C_v^{*2} - 2\varphi_3\rho_{uv}^* C_x^* C_v^*\Big) \quad (4.25)$$

The optimum value of φ_3 is obtained as $\frac{\partial MSE(\hat{\bar{Y}}_3^{(3)})}{\partial\varphi_3} = 0$, then

$$\varphi_{3(opt.)} = \frac{\rho_{uv}^* C_u^*}{C_v^*}$$

Substituting the optimum value of φ_3 in (4.25), the minimum mean squared error of $\hat{\bar{Y}}_3^{(3)}$ is

$$MSE(\hat{\bar{Y}}_3^{(3)})_{min.} = \frac{1}{r}C_u^{*2}\bar{Y}^2\Big\{\Big(1 - g_1\rho_{uv}^{*2}\Big) + \lambda\rho_{uv}^{*2}\Big\}$$

■

$$\hat{Y}_j = \begin{cases} u_j^* & \text{if } j\epsilon G \\ \frac{1}{(1-g_1)}\left[\varphi_4\bar{u}_r^* + (1-\varphi_4)\bar{u}_r^*\left(\frac{\bar{x}_{2m}^*}{\bar{v}_{2n}^*}\right) - g_1\bar{u}_r^*\right] & \text{if } j\epsilon G^c \end{cases} \quad (4.26)$$

where φ_4 is an unknown constant. The point estimator for the given procedure in (4.26) is defined as:

$$\hat{\bar{Y}}_4^{(3)} = \varphi_4\bar{u}_r^* + (1-\varphi_4)\bar{u}_r^*\left(\frac{\bar{x}_{2m}^{**}}{\bar{v}_{2n}^*}\right)$$

Rewriting $\hat{\hat{Y}}_4^{(3)}$ in term of error, we have

$$\hat{\hat{Y}}_4^{(3)} = \varphi_4 \bar{Y}(1+\zeta_0^{'}) + (1-\varphi_4)\bar{Y}\Big(1-\zeta_1^{''}+\zeta_1^{''2}+\zeta_2^{''}-\zeta_1^{''}\zeta_2^{''}+\zeta_0^{'} -\zeta_0^{'}\zeta_1^{''}+\zeta_0^{'}\zeta_2^{''}\Big)$$

Expanding and keeping terms up to first order of approximation, the bias and mean square error of $\hat{\hat{Y}}_4^{(3)}$ is given as

$$E(\hat{\hat{Y}}_4^{(3)} - \bar{Y}) \cong (\varphi_4 - 1)\bar{Y} + (1-\varphi_4)\bar{Y}\Pi_{nm}\left(C_v^{*2} - \rho_{uv}^{*} C_y^{*} C_v^{*}\right)$$

and

$$MSE(\hat{\hat{Y}}_4^{(3)}) \cong \frac{1}{r}\bar{Y}^2 C_v^{*2} + \Pi_{nm}\bar{Y}^2\left\{\varphi_4(\varphi_4-1)C_v^{*2} + 2(\varphi_4-1)\rho_{uv}^{*}C_v^{*}C_u^{*}\right\} \tag{4.27}$$

The optimum value of φ_4 is obtained as $\frac{\partial MSE(\hat{\hat{Y}}_4^{(3)})}{\partial\varphi_4} = 0$, then

$$\varphi_{4(opt.)} = 1 - \frac{C_u^{*}\rho_{uv}^{*}}{C_v^{*}}$$

Substituting the optimum value of φ_4 in (4.27), the minimum mean squared error of $\hat{\hat{Y}}_4^{(3)}$ is

$$MSE(\hat{\hat{Y}}_4^{(3)})_{min.} = \frac{1}{r}C_u^2\bar{Y}^2\Big\{\Big(1-g_1\rho_{uv}^2\Big) + \lambda\rho_{uv}^2\Big\}$$

4.4. Situation 4: Non-response in X_1 and X_2 at first phase

■

$$\hat{Y}_j = \begin{cases} u_j^{**} & \text{if } j\epsilon\, G \\ \frac{1}{(1-g_1)}\left[\gamma_1\bar{u}_r^{**}\frac{\bar{x}_{2r'}^{**}}{\bar{v}_{2r}^{*}} - g_1\bar{u}_r^{**}\right] & \text{if } j\epsilon\, G^c \end{cases} \tag{4.28}$$

where γ_1 is a suitably chosen constant that makes the MSE minimum. The point estimator for the population mean is defined as:

$$\hat{\hat{Y}}_1^{(4)} = \gamma_1\bar{u}_r^{**}\frac{\bar{x}_{2r'}^{**}}{\bar{v}_{2r}^{*}} \tag{4.29}$$

In term of error, the (4.29) can be written as

$$\hat{\hat{Y}}_1^{(4)} = \gamma_1\bar{Y}\Big(1-\zeta_1^{'''}+\zeta_1^{'''2}+\zeta_2^{'''}-\zeta_1^{'''}\zeta_2^{'''}+\zeta_0^{''}-\zeta_0^{''}\zeta_1^{'''}+\zeta_0^{''}\zeta_2^{'''}\Big)$$

Keeping terms up to first order of approximation, the bias and mean square error of $\hat{\hat{Y}}_1^{(4)}$ is given as

$$E(\hat{\hat{Y}}_1^{(4)} - \bar{Y}) \cong \bar{Y}(\gamma_1 - 1) + \Pi_{rr'}\,\bar{Y}\,\gamma_1\Big(C_v^{**2} - \rho_{uv}^{**}C_u^{**}C_v^{**}\Big)$$

and

$$MSE(\hat{\hat{Y}}_1^{(4)}) \cong \frac{1}{r}\bar{Y}^2\Delta_1^2 C_u^{**2} - 2\Pi_{rr'}\bar{Y}^2\gamma_1\Big\{\Big(\frac{3}{2}\gamma_1 - 1\Big)C_v^{**2} -(2\gamma_1-1)\rho_{uv}^{**}C_u^{**}C_v^{**}\Big\} + \bar{Y}^2(\gamma_1-1)^2 \tag{4.30}$$

The optimum value of γ_1 is obtained by $\frac{\partial MSE(\hat{\hat{Y}}_1^{(2)})}{\partial\gamma_1} = 0$, as follow

$$\gamma_{1(opt.)} = \frac{r^{*}C_v^{**2} - C_u^{**}\rho_{yx}r^{*}C_v^{**} - rr^{'}}{3r^{*}C_v^{**2} - 4\,C_u^{**}\rho_{uv}^{**}r^{*}C_v^{**} - r^{'}\,(C_u^{**2} + r)} \tag{4.31}$$

Substituting (4.31) in (4.30), the minimum mean squared error of $\hat{\bar{Y}}_1^{(4)}$ is

$$MSE(\hat{\bar{Y}}_1^{(4)})_{min.} \cong \frac{\left[r^* C_v^{**}\left\{C_v^{**}\left(r^* C_v^{**}\Lambda' + \Gamma'\right) - 2\, r C_u^{**} r' \rho_{uv}^{**}\right\} - r C_u^{**2} r'^2\right]\bar{Y}^2}{3\left\{r^* C_v^{**2} - \frac{4}{3} C_u^{**}\rho_{uv}^{**} r^{**} C_v^{**} - \frac{1}{3} r'\left(C_u^{**2} + r\right)\right\} r' r}$$

■

$$\hat{Y}_j = \begin{cases} u_j^{**} & \text{if } j \epsilon\, G \\ \frac{1}{(1-g_1)}\left[\bar{u}_r^{**}\left(\frac{\bar{x}_{2r'}^{**}}{\bar{v}_{2r}^*}\right)^{\gamma_2} - g_1 \bar{u}_r^{**}\right] & \text{if } j \epsilon\, G^c \end{cases} \quad (4.32)$$

where γ_2 is the suitably chosen constant. The point estimator is defined as:

$$\hat{\bar{Y}}_2^{(4)} = \bar{u}_r\left(\frac{\bar{x}_{2r'}^{**}}{\bar{v}_{2n}^*}\right)^{\gamma_2} \quad (4.33)$$

Rewriting (4.33) in term of error, we have

$$\begin{aligned}\hat{\bar{Y}}_2^{(4)} &= \bar{Y}\Big(1 - \gamma_2\zeta_1^{'''} - \frac{1}{2}\gamma_2(\gamma_2 - 1)\zeta_1^{'''2} + \gamma\zeta_2^{'''} - \gamma_2^2\zeta_1^{'''}\zeta_2^{'''} + \zeta_0^{''} \\ &\quad + \frac{1}{2}\gamma_2(\gamma_2 - 1)\zeta_2^{'''2} - \omega_2\zeta_0^{''}\zeta_1^{'''} + \gamma_2\zeta_0^{''}\zeta_2^{'''}\Big)\end{aligned}$$

expanding and keeping terms up to first order approximation, the bias and mean square error of $\hat{\bar{Y}}_2^{(4)}$ is given as

$$E(\hat{\bar{Y}}_2^{(4)} - \bar{Y}) \;\cong\; \bar{Y}\Pi_{rr'}\,\gamma_2\left\{\frac{1}{2}\left(\gamma_2 - 1\right)C_v^{**2} - \rho_{uv}^{**}C_u^{**}C_v^{**}\right\}$$

and

$$MSE(\hat{\bar{Y}}_2^{(4)}) \cong \frac{1}{r}\bar{Y}^2 C_u^{**2} + \Pi_{rr'}\bar{Y}^2\left(\gamma_2^2 C_v^{**2} - \gamma_2\rho_{uv}^{**}C_u^{**}C_v^{**}\right) \quad (4.34)$$

The optimum value of γ_2 is obtained as $\frac{\partial MSE(\hat{\bar{Y}}_2^{(4)})}{\partial\gamma_2} = 0$, then

$$\gamma_{2(opt.)} = \frac{\rho_{uv}^{**}C_u^{**}}{C_v^{**}} \quad (4.35)$$

Substituting (4.35) in (4.34), the minimum mean squared error of $\hat{\bar{Y}}_2^{(4)}$ is

$$MSE(\hat{\bar{Y}}_2^{(4)})_{min.} \cong \frac{1}{r}\bar{Y}^2 C_u^{**2}\left(1 + \lambda'\rho_{uv}^{**2}\right)$$

■

$$\hat{Y}_j = \begin{cases} u_j^{**} & \text{if } j \epsilon\, G \\ \frac{1}{(1-g_1)}\left[\frac{\bar{u}_r^{**}\;\bar{x}_{2r'}^{**}}{\gamma_3\bar{v}_{2r}^* + (1-\gamma_3)\bar{x}_{2r'}^{**}} - g_1\bar{u}_r^{**}\right] & \text{if } j \epsilon\, G^c \end{cases} \quad (4.36)$$

where γ_3 is the suitably chosen unknown value. The point estimator for the population mean is defined as:

$$\hat{\bar{Y}}_3^{(4)} = \frac{\bar{u}_r^{**}\;\bar{x}_{2r'}^{**}}{\gamma_3\bar{v}_{2r}^* + (1-\gamma_3)\bar{x}_{2r'}^{**}}$$

Rewriting the $\hat{\bar{Y}}_3^{(4)}$ in term of error, we have

$$\begin{aligned}\hat{\bar{Y}}_3^{(4)} &= \bar{Y}\Big(1 - \gamma_3\zeta_1^{'''} + \gamma_3\zeta_1^{'''} - \zeta_2^{'''} + \omega_3^2\zeta_1^{'''2} + \gamma_3^2\zeta_2^{'''2} + \zeta_1^{'''2} - 2\gamma_3^2\zeta_1^{'''}\zeta_2^{'''} \\ &\quad - 2\gamma_3\zeta_1^{'''}\zeta_2^{'''} + 2\gamma_3\zeta_2^{'''2} + \zeta_2^{'''} - \gamma_1\zeta_1^{'''}\zeta_2^{'''} + \gamma_3\zeta_2^{'''} - \zeta_2^{'''2} \\ &\quad + \zeta_0^{'''} - \gamma_3\zeta_0^{''}\zeta_1^{'''} + \gamma_3\zeta_0^{''}\zeta_2^{'''} - \zeta_0^{''}\zeta_2^{'''} + \zeta_0^{''}\zeta_2^{'''}\Big)\end{aligned}$$

The bias and mean square error of $\hat{\hat{Y}}_4^{(4)}$ is given as

$$E(\hat{\hat{Y}}_3^{(4)} - \bar{Y}) \cong \bar{Y}\Pi_{rr'}\gamma_3\left(\gamma_3 C_v^{**2} - \rho_{uv}^{**}C_u^{**}C_v^{**}\right)$$

and

$$MSE(\hat{\hat{Y}}_3^{(4)}) \cong \frac{1}{r}\bar{Y}^2 C_u^{**2} + \Pi_{rr'}\bar{Y}^2\left(\gamma_3^2 C_v^{**2} - 2\gamma_3\rho_{uv}^{**}C_x^{**}C_v^{**}\right) \tag{4.37}$$

The suitable value of γ_3 is obtained by setting $\frac{\partial MSE(\hat{\hat{Y}}_3^{(4)})}{\partial\omega_3} = 0$, as follow

$$\gamma_{3(opt.)} = \frac{C_u^{**}\rho_{uv}^{**}}{C_v^{**}}$$

Substituting the optimum value of γ_3 in (4.37), the minimum mean squared error of $\hat{\hat{Y}}_3^{(4)}$ is

$$MSE(\hat{\hat{Y}}_2^{(4)})_{min.} \cong \frac{1}{r}\bar{Y}^2 C_u^{**2}\left(1 + \lambda'\rho_{uv}^{**2}\right)$$

■

$$\hat{Y}_j = \begin{cases} u_j^{**} & \text{if } j\epsilon G \\ \frac{1}{(1-g_1)}\left[\gamma_4\bar{u}_r^{**} + (1-\gamma_4)\bar{u}_r^{**}\left(\frac{\bar{x}_{2r'}^{**}}{\bar{v}_{2r}^{*}}\right) - g_1\bar{u}_r^{**}\right] & \text{if } j\epsilon G^c \end{cases} \tag{4.38}$$

where γ_4 is an unknown constant. The point estimator for the given procedure in (4.38) is defined as:

$$\hat{\hat{Y}}_4^{(4)} = \gamma_4\bar{u}_r^{**} + (1-\gamma_4)\bar{u}_r^{**}\left(\frac{\bar{x}_{2r'}^{**}}{\bar{v}_{2r}^{*}}\right) \tag{4.39}$$

Rewriting (4.39) in term of error, we have

$$\begin{aligned}\hat{\hat{Y}}_4^{(4)} &= \bar{Y}\gamma_4\bar{Y}(1+\zeta_0'') + (1-\gamma_4)\bar{Y}\Big(1 - \zeta_1''' + \zeta_1'''^2 + \zeta_2''' - \zeta_1'''\zeta_2''' \\ &\quad + \zeta_0'' - \gamma_2\zeta_0''\zeta_1''' + \gamma_2\zeta_0''\zeta_2'''\Big)\end{aligned}$$

The bias and mean square error of $\hat{\hat{Y}}_4^{(4)}$ is given as

$$E(\hat{\hat{Y}}_4^{(4)} - \bar{Y}) \cong (\gamma_4 - \omega_4 - 1)\bar{Y} + (1-\gamma_4)\bar{Y}\Pi_{rr'}\left(C_v^{**2} - \rho_{uv}^{**}C_u^{**}C_v^{**}\right)$$

and

$$MSE(\hat{\hat{Y}}_4^{(4)}) \cong \frac{1}{r}\bar{Y}^2 C_u^{**2} + \Pi_{rr'}\left\{\gamma_4\left(\gamma_4 - 1\right)C_v^{**2} + 2(\gamma_4 - 1)\rho_{uv}^{**}C_v^{**}C_u^{**}\right\} \tag{4.40}$$

The optimum values of γ_4 is obtained by $\frac{\partial MSE(\hat{\hat{Y}}_4^{(4)})}{\partial\gamma_4} = 0$, then

$$\gamma_{4(opt.)} = 1 - \frac{C_u^{**}\rho_{uv}^{**}}{C_v^{**}}$$

Substituting the optimum value of γ_4 in (4.40), the minimum mean squared error of $\hat{\hat{Y}}_4^{(4)}$ is

$$MSE(\hat{\hat{Y}}_4^{(4)})_{min.} \cong \frac{1}{r}\bar{Y}^2 C_u^{**2}\left(1 + \lambda'\rho_{uv}^{**2}\right) \tag{4.41}$$

5. Application

In this section, we discuss the numerical findings of the modified class of estimators under two phase *pps* sampling scheme by using the two real life data sets at varying response rate. The data description and method of bootstrapping for the previously predefined situations are defined as follow:

5.1. Situtation 1 and 2

Population 1: Source: [7]
y= Output in (000) rupees, x_1 = Number of Workers and x_2 = Fixed Capital in (000) rupees.

$N = 80, \quad \bar{Y} = 84443.509, \quad \bar{X} = 1338.756, \quad C_u = 0.0609, \ C_v = 0.0274, \ \rho_{uv} = 0.8520.$

Population 2: Source: [14]
y= Estimated number of fish caught by marine recreational fishermen in year 1995, x_1= estimated number of fish caught by marine recreational fishermen in year 1994 and x_2= estimated number of fish caught by marine recreational fishermen in year 1993.

$N = 69, \quad \bar{Y} = 4699.529, \quad \bar{X} = 5218.194, \quad C_u = 0.0401, \quad C_v = 0.0335, \quad \rho_{uv} = 0.6483.$

5.2. Situtation 3 and 4

For the 3^{rd} and 4^{th} situation, we have no auxiliary information regarding the study variable in advance. In such circumstances, we select the sample at first phase by SRS and at second phase by *pps* sampling. The values of C_v^*, C_v^{**}, $C_u^{**}, C_u^{**}, \rho_{uv}^*$ and ρ_{uv}^{**} are obtained under bootstrap approach by using the population 1 and 2. Repeat the process of the selection of the units 10000 (say H) times. The selection procedure of s from Ω is define as follows:

First we select the m or r' units at first phase by SRS from N units of Ω. Then, form the selected s, we select the n or r units by *pps* sampling, repeating the procedure H times and then obtain the mean value of the C_v^*, C_v^{**}, $C_u^{**}, C_u^{**}, \rho_{uv}^*$ and ρ_{uv}^{**} , and utilized such values for the relative comparison of the modified estimators.

We use the following expression for calculating the percentage relative efficiencies of the modified imputation strategies under two phase *pps* sampling than their counterpart, as follow:

$$PRE(k) = \frac{Var(\hat{\hat{Y}}_M^q)}{MSE(\hat{\hat{Y}}_k^q)}, \quad \text{for } k = 1-4 \text{ and } q = 1-4 \tag{5.1}$$

At the fixed response rate, the PRE's are reported in Table 1 by using the population 1 and 2 respectively for the given situations.

Table 1. PRE(k) of the modified estimators

Pop.	PRE(1)	PRE(2)	PRE(3)	PRE(4)
	Sitautation 1 ($m = 50$, $n = 25$, $r = 15$)			
1	123.3124	127.8408	127.8408	127.8408
2	101.5064	109.3101	109.3101	109.3101
	Situtation 2 ($r' = 45$, $n = 25$, $r = 15$)			
1	167.8081	193.7795	193.7795	193.7795
2	101.7872	123.3457	123.3457	123.3457
	Situtation 3 ($m = 45$, $n = 25$, $r = 15$)			
1	119.6057	121.6906	121.6906	121.6906
2	100.9423	107.3416	107.3416	107.3416
	Situtation 4 ($m = 35$, $n = 25$, $r = 15$)			
1	138.6255	139.7692	139.7692	139.7692
2	100.8386	109.1891	109.1891	109.1891

Table 2. PRE(.) for situtation 1

			Population 1				Population 2			
m	n	r	PRE 1	PRE 2	PRE 3	PRE 4	PRE 1	PRE 2	PRE 3	PRE 4
50	25	15	123.3124	127.8408	127.8408	127.8408	101.5064	109.3101	109.3101	109.3101
		10	115.7463	116.9844	116.9844	116.9844	102.3078	106.0199	106.0199	106.0199
		5	110.7542	107.8275	107.8275	107.8275	104.4806	102.9220	102.9220	102.9220
	15	12	152.0836	168.4970	168.4970	168.4970	102.2976	118.9042	118.9042	118.9042
		8	131.2926	137.1764	137.1764	137.1764	103.3361	111.8557	111.8557	111.8557
		4	117.8840	115.6746	115.6746	115.6746	106.0891	105.5961	105.5961	105.5961
	10	9	177.7732	209.4966	209.4966	209.4966	103.6909	125.6932	125.6932	125.6932
		6	143.4654	153.4787	153.4787	153.4787	105.0115	115.7775	115.7775	115.7775
		3	123.6739	121.0979	121.0979	121.0979	108.6179	107.3119	107.3119	107.3119
30	20	15	119.0352	122.1719	122.1719	122.1719	101.4779	107.6399	107.6399	107.6399
		10	113.2487	113.7640	113.7640	113.7640	102.2620	104.9668	104.9668	104.9668
		5	109.6618	106.4389	106.4389	106.4389	104.4173	102.4232	102.4232	102.4232
	12	9	139.0610	148.5147	148.5147	148.5147	103.0671	114.6470	114.6470	114.6470
		6	125.2947	127.8408	127.8408	127.8408	104.3524	109.3101	109.3101	109.3101
		3	117.0233	112.2194	112.2194	112.2194	107.9233	104.4480	104.4480	104.4480
	8	6	152.1251	166.4609	166.4609	166.4609	105.2893	118.5042	118.5042	118.5042
		4	133.0224	136.2717	136.2717	136.2717	107.1245	111.6194	111.6194	111.6194
		2	122.8976	115.3517	115.3517	115.3517	112.3877	105.4907	105.4907	105.4907
20	15	12	115.3985	116.9844	116.9844	116.9844	101.8798	106.0199	106.0199	106.0199
		8	111.3978	110.7162	110.7162	110.7162	102.8156	103.9343	103.9343	103.9343
		4	109.6971	105.0856	105.0856	105.0856	105.4654	101.9292	101.9292	101.9292
	12	9	124.1934	127.8408	127.8408	127.8408	102.7708	109.3101	109.3101	109.3101
		6	117.1377	116.9844	116.9844	116.9844	104.0202	106.0199	106.0199	106.0199
		3	113.6617	107.8275	107.8275	107.8275	107.5550	102.9220	102.9220	102.9220
	8	6	140.0611	148.5147	148.5147	148.5147	104.8962	114.6470	114.6470	114.6470
		4	126.9469	127.8408	127.8408	127.8408	106.7269	109.3101	109.3101	109.3101
		2	120.5821	112.2194	112.2194	112.2194	111.9856	104.4480	104.4480	104.4480

Table 3. PRE(.) for situtation 2

			Population 1				Population 2			
r'	n	r	PRE 1	PRE 2	PRE 3	PRE 4	PRE 1	PRE 2	PRE 3	PRE 4
45	25	15	167.8081	193.7795	193.7795	193.7795	101.7872	123.3457	123.3457	123.3457
		10	188.9941	229.6779	229.6779	229.6779	103.3534	128.3393	128.3393	128.3393
		5	216.6503	281.9013	281.9013	281.9013	108.4160	133.7543	133.7543	133.7543
	15	12	179.8966	213.8326	213.8326	213.8326	102.5535	126.2941	126.2941	126.2941
		8	199.0975	248.0595	248.0595	248.0595	104.5885	130.4518	130.4518	130.4518
		4	223.3483	295.3316	295.3316	295.3316	111.0223	134.8926	134.8926	134.8926
	10	9	193.9084	238.5151	238.5151	238.5151	103.8983	129.3869	129.3869	129.3869
		6	210.4242	269.6394	269.6394	269.6394	106.6993	132.635	132.635	132.635
		3	230.6701	310.1056	310.1056	310.1056	115.4288	136.0504	136.0504	136.0504
25	20	15	132.8834	140.9184	140.9184	140.9184	101.5704	112.8111	112.8111	112.8111
		10	157.9589	177.165	177.165	177.165	102.9807	120.5318	120.5318	120.5318
		5	194.9796	238.5151	238.5151	238.5151	107.8771	129.3869	129.3869	129.3869
	12	9	164.1675	186.7733	186.7733	186.7733	103.4939	122.2045	122.2045	122.2045
		6	186.1982	223.0661	223.0661	223.0661	106.1959	127.5133	127.5133	127.5133
		3	215.6277	276.8651	276.8651	276.8651	114.8066	133.3043	133.3043	133.3043
	8	6	186.1982	223.0661	223.0661	223.0661	106.1959	127.5133	127.5133	127.5133
		4	204.7075	256.2633	256.2633	256.2633	110.4448	131.3164	131.3164	131.3164
		2	228.2551	301.069	301.069	301.069	123.7506	135.3534	135.3534	135.3534
15	10	8	141.1905	151.2321	151.2321	151.2321	103.5664	115.2723	115.2723	115.2723
		6	159.0118	177.165	177.165	177.165	105.4951	120.5318	120.5318	120.5318
		4	182.1334	213.8326	213.8326	213.8326	109.6182	126.2941	126.2941	126.2941
	8	6	159.0118	177.165	177.165	177.165	105.4951	120.5318	120.5318	120.5318
		4	182.1334	213.8326	213.8326	213.8326	109.6182	126.2941	126.2941	126.2941
		2	213.9539	269.6394	269.6394	269.6394	122.7517	132.635	132.635	132.635
	6	4	182.1334	213.8326	213.8326	213.8326	109.6182	126.2941	126.2941	126.2941
		3	196.5896	238.5151	238.5151	238.5151	113.9058	129.3869	129.3869	129.3869
		2	213.9539	269.6394	269.6394	269.6394	122.7517	132.635	132.635	132.635

Table 4. PRE(.) for situtation 3

			Population 1				Population 2			
m	n	r	PRE 1	PRE 2	PRE 3	PRE 4	PRE 1	PRE 2	PRE 3	PRE 4
45	25	15	119.6057	121.6906	121.6906	121.6906	100.9423	107.3416	107.3416	107.3416
		10	113.5742	113.4658	113.4658	113.4658	102.1460	104.7511	104.7511	104.7511
		5	109.7927	106.3086	106.3086	106.3086	105.2532	102.3332	102.3332	102.3332
	15	12	148.0786	158.4113	158.4113	158.4113	105.5349	117.4324	117.4324	117.4324
		8	130.4249	132.5194	132.5194	132.5194	107.2176	110.9650	110.9650	110.9650
		4	120.4511	113.9926	113.9926	113.9926	112.4339	105.2015	105.2015	105.2015
	10	9	174.7657	196.6369	196.6369	196.6369	114.7185	125.611	125.611	125.611
		6	145.5935	148.6959	148.6959	148.6959	116.9741	115.6501	115.6501	115.6501
		3	133.0882	119.5711	119.5711	119.5711	126.397	107.2814	107.2814	107.2814
30	20	15	117.0176	118.5384	118.5384	118.5384	100.3176	106.2655	106.2655	106.2655
		10	111.4644	111.6438	111.6438	111.6438	100.5683	104.1267	104.1267	104.1267
		5	107.3274	105.5083	105.5083	105.5083	102.6324	102.0046	102.0046	102.0046
	12	9	136.9852	142.7462	142.7462	142.7462	104.0726	114.0326	114.0326	114.0326
		6	123.9916	124.9023	124.9023	124.9023	105.3895	108.9407	108.9407	108.9407
		3	116.0199	111.0531	111.0531	111.0531	109.1232	104.2475	104.2475	104.2475
	8	6	151.3950	160.5782	160.5782	160.5782	112.0256	119.3809	119.3809	119.3809
		4	133.9495	133.5281	133.5281	133.5281	113.8538	112.1130	112.1130	112.1130
		2	126.5166	114.3838	114.3838	114.3838	121.2343	105.6933	105.6933	105.6933
20	15	12	112.4419	112.8691	112.8691	112.8691	100.2681	104.1939	104.1939	104.1939
		8	108.5124	108.2521	108.2521	108.2521	100.3426	102.7052	102.7052	102.7052
		4	105.4670	103.9598	103.9598	103.9598	101.6140	101.3509	101.3509	101.3509
	12	9	120.5222	121.9266	121.9266	121.9266	100.3985	107.1552	107.1552	107.1552
		6	113.6749	113.5814	113.5814	113.5814	101.4925	104.6421	104.6421	104.6421
		3	108.7208	106.3603	106.3603	106.3603	103.3270	102.2875	102.2875	102.2875
	8	6	135.8765	140.1535	140.1535	140.1535	103.8653	113.3375	113.3375	113.3375
		4	123.4585	123.6419	123.6419	123.6419	105.1021	108.4416	108.4416	108.4416
		2	115.7355	110.5456	110.5456	110.5456	109.1980	104.1017	104.1017	104.1017

Table 5. PRE(.) for situtation 4

			Population 1				Population 2			
r'	n	r	PRE 1	PRE 2	PRE 3	PRE 4	PRE 1	PRE 2	PRE 3	PRE 4
35	25	15	138.6255	139.7692	139.7692	139.7692	100.8386	109.1891	109.1891	109.1891
		10	153.3283	155.1452	155.1452	155.1452	101.9139	111.5064	111.5064	111.5064
		5	172.0155	174.7142	174.7142	174.7142	103.1092	114.3143	114.3143	114.3143
	15	12	154.2364	160.3843	160.3843	160.3843	102.7001	115.0852	115.0852	115.0852
		8	170.0231	178.7351	178.7351	178.7351	104.4348	118.2009	118.2009	118.2009
		4	189.8148	202.1161	202.1161	202.1161	106.2724	121.7501	121.7501	121.7501
	10	9	170.9669	184.4274	184.4274	184.4274	109.0338	121.6416	121.6416	121.6416
		6	186.0077	204.0766	204.0766	204.0766	108.4835	124.3379	124.3379	124.3379
		3	204.8966	229.1396	229.1396	229.1396	114.6697	127.8079	127.8079	127.8079
25	20	15	125.8821	127.1147	127.1147	127.1147	100.8819	107.3775	107.3775	107.3775
		10	144.283	146.8567	146.8567	146.8567	101.5221	111.2629	111.2629	111.2629
		5	169.1765	173.9383	173.9383	173.9383	102.9534	116.1377	116.1377	116.1377
	12	9	154.4358	161.9805	161.9805	161.9805	104.7686	116.5543	116.5543	116.5543
		6	171.8914	183.2653	183.2653	183.2653	109.7789	120.6860	120.6860	120.6860
		3	193.9579	210.8838	210.8838	210.8838	111.3464	124.2724	124.2724	124.2724
	8	6	176.3413	193.3752	193.3752	193.3752	104.8743	124.9147	124.9147	124.9147
		4	192.0527	214.8227	214.8227	214.8227	108.0107	128.138	128.138	128.138
		2	211.0831	240.7031	240.7031	240.7031	117.4007	131.5687	131.5687	131.5687
15	10	8	136.0069	140.3463	140.3463	140.3463	100.1882	112.5078	112.5078	112.5078
		6	151.2405	158.5835	158.5835	158.5835	101.1974	116.5307	116.5307	116.5307
		4	170.4139	182.3315	182.3315	182.3315	102.7417	121.3813	121.3813	121.3813
	8	6	152.8711	161.8102	161.8102	161.8102	103.1252	118.1632	118.1632	118.1632
		4	172.8094	187.5118	187.5118	187.5118	106.1244	123.1759	123.1759	123.1759
		2	199.6618	223.3881	223.3881	223.3881	115.3648	128.6149	128.6149	128.6149
	6	4	175.4890	193.1868	193.1868	193.1868	112.8095	126.2304	126.2304	126.2304
		3	188.4202	211.0872	211.0872	211.0872	116.7172	129.1315	129.1315	129.1315
		2	203.8417	232.8731	232.8731	232.8731	125.5094	132.5675	132.5675	132.5675

In Table 1, we observe that the performance of modified imputation strategies under two phase *pps* sampling. For all the previously mentioned situations 1-4, the estimation procedure is more effective and reliable as compare to the simple mean estimator. In Appendix, we consider the comprehensive examination of the suggested imputation procedure with varying response rate at first and second phase respectively. The percentage relative efficiencies are reported in Table 2 and 3 is for first two situations, when the probability of selection of the observation units is known in advance. For last two, when the selection probabilities are obtained through bootstrapping, then PREs reported in Table 4 and 5. In all the reported numerical findings in Appendix, the performance of modified imputation strategies is better than their counterpart.

6. Conclusion

Imputation of missing values in sample surveys is a good practice, for dealing non-response problem, in term of cost and duration. Several strategies have been proposed for the purpose of bias reduction and efficient imputation. The current research dealt with problem of non-response under four possible scenarios with respect to non-response occurrence under two phase *pps* sampling. A modified class of estimators is developed using the available auxiliary information on both phases. The theoretical findings suggest that the proposed class of estimators performs better then their counterparts under certain constrains. Numerical studies are given to support the theoretical finding. The suggest class of estimator is an efficient and might be cost effective alternative to the situations where two phase sampling is feasible. This research can be extended for the stratified and clustered populations.

Acknowledgment. Authors are thankful to the editor and reviewers for their constructive review.

References

[1] M. Ahmed, O. Al-Titi, Z. Al-Rawi and W. Abu-Dayyeh, *Estimation of a population mean using different imputation methods* , Statistics in Transition **7** (6), 1247-1264, 2006.

[2] D. F. Heitjan and S. Basu, *Distinguishing missing at random and missing completely at random*, The Amer. Stat. **50** (3), 207-213, 1996.

[3] C. N. B. Herrera and A. I. Al-Omari, Ranked set estimation with imputation of the missing observations: the median estimator, Rev. Invest. Opera. **32** (1), 30-37, 2011.

[4] J. K. Kim, A. Navarro and W. A. Fuller, *Replication variance estimation for two-phase stratified sampling*, J. Amer. Statist. Assoc. **101** (473), 312-320, 2006.

[5] R. J. Little and D. B. Rubin, *Statistical analysis with missing data*, John Wiley and Sons, 2014.

[6] C. Mohamed , S. A. Sedory and S. Singh, *Imputation using higher order moments of an auxiliary variable*, Comm. Statist. Simulation Comput. **46** (8), 6588-6617, 2017.

[7] M. N. Murthy, *Sampling theory and methods*, Stat. Pub. Soc. 204/1, Barrackpore Tunk Road Calcutta, India, 1976.

[8] J. N. Rao and R. Sitter, *Variance estimation under two-phase sampling with application to imputation for missing data*, Biometrika **82** (2), 453-460, 1995.

[9] D. B. Rubin, *Inference and missing data*, Biometrika **63** (3), 581-592, 1976.

[10] J. Shabbir and S. Gupta, *Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes*, Hacet. J. Math. Stat. **39** (1), 121-129, 2010.

[11] H. P. Singh and S. Kumar, *Estimation of mean in presence of non-response using two phase sampling scheme*, Statist. Papers **51** (3), 559-582, 2010.

[12] H. P. Singh, S. Kumar and M. Kozak, *Improved estimation of finite population mean using sub-sampling to deal with non response in two-phase sampling scheme*, Comm. Statist. Theory Methods **39** (5), 791-802, 2010.

[13] M. U. Sohail, J. Shabbir and S. Ahmed, *Modified class of ratio and regression type estimators for imputing scrambling response*, Pakistan J. Statist., **33** (4), 277-300, 2017.

[14] S. Singh, *Advanced Sampling Theory With Applications: How Michael Selected Amy*, Volume 2. Springer Science and Business Media, 2003.