Study of general families of estimators in two-stage sampling with unequal size clusters under non-response

Manoj K. Chaudhary*†, V. K. Singh‡ and Cem Kadilar§

Abstract

This paper presents some salient features of the biased and unbiased estimators of population mean in two-stage sampling with unequal size clusters in presence of non-response. Paper proposes general families of factor-type estimators of population mean using information on an auxiliary variable under the condition that the non-response is observed on study variable only and on both study and auxiliary variables respectively.Biases and mean square errors of the families are obtained. Optimum estimators of the proposed families have been investigated in both the cases. An empirical study is also carried out to support the theoretical results.

Keywords: Two-stage sampling, unequal size clusters, auxiliary information, factor-type estimators, non-response.

1. Introduction

The two-stage sampling method has its own importance in sampling theory and estimation procedure. Bellhouse and Rao [2] discussed on the efficiency of prediction estimator in two-stage sampling. Khan and Muttlak [6] suggested an adjusted two-stage adaptive cluster sampling procedure in estimating the population mean. There are several authors who have utilized auxiliary information for improving the efficiency of the estimators in two-stage sampling at the estimation stage. Sahoo and Panda [8] have suggested a class of estimators of population total with varying probability in two-stage sampling. Sahoo *et al.* [7] introduced a class of predictive estimators for estimating population mean using two auxiliary variables in two-stage sampling. Ahmed [1] has proposed some estimators for finite population mean in two-stage sampling using multi-auxiliary

---

*Department of Statistics, Banaras Hindu University, Varanasi, India, Email: ritamanoj15@gmail.com

†Corresponding Author.

‡Department of Statistics, Banaras Hindu University, Varanasi, India

§Department of Statistics, Hacettepe University, Beytepe 06800, Ankara, Turkey

information. Eideh and Nathan [4] considered the effects of informative two-stage cluster sampling on estimation and prediction.

In mail surveys where questionnaires are sent to the respondents and it is expected to send back their returns within a specified time. It is seen that many of them do not reply and available questionnaires of returns are imperfect. Due to this unavailability of response, one may infer the biased estimate of the parameter of interest. Thus the problem of non-response is a serious issue in sample survey as well as in complete enumeration. In order to deal with the problem of non-response, Chaudhary and Singh [3] have suggested some families of factor-type estimators under different situations in two-stage sampling with equal size clusters in the presence of non-response.

In most of the situations, assumption of equal size clusters is impracticable. During the surveys, an experimenter frequently faces the population with unequal size clusters. Due to this reason, we have proposed two different sampling strategies for estimating the population mean in two-stage sampling with unequal size clusters utilizing the information on an auxiliary characteristic in the presence of non-response. First, we have discussed the case where non-response is observed on study variable and auxiliary variable is free from non-response. Secondly, we entertained the case where non-response is observed on both the variables.

2. Estimation procedure in presence of Non-Response

Let us suppose that the population of size M_0 is divided into N first-stage units $(f.s.u^{'s})$ such that the ith f.s.u has size M_i of second stage unit $(s.s.u^{'s})$. First of all we select a sample of size n from N $(f.s.u^{'s})$ with the help of simple random sampling without replacement (SRSWOR) scheme and then for each of the selected $f.s.u^{'s}$ at the first stage, a sub-sample of specified size of $s.s.u^{'s}$ is selected at the second-stage (i.e., m_i $s.s.u^{'s}$ are selected from the ith f.s.u.). It is observed that out of m_i $s.s.u^{'s}$, m_{i1} units respond and m_{i2} units do not respond. Adopting Hansen and Hurwitz [5] technique of sub-sampling of non-respondents, we now select a sub-sample of size h_{i2} from the m_{i2} non-respondent units with the help of SRSWOR scheme for the ith f.s.u. selected in the sample such that $m_{i2}=k_i h_{i2}(k_i \geq 1)$ and collect the information on all h_{i2} units by interview method.

Let X_0 and X_1 be the study and auxiliary variables with respective population means $\bar{X}_0$ and $\bar{X}_1$. First, we define the Hansen-Hurwitz estimator for population mean of ith f.s.u. $\bar{X}_{0i}$ without using auxiliary information as

$$(2.1) \qquad \bar{x}_{0HHi} = \frac{m_{i1}\bar{x}_{0i1} + m_{i2}\bar{x}_{0i2}^*}{m_i}$$

where $\bar{x}_{0i1}$ and $\bar{x}_{0i2}^*$ are the means per s.s.u. based on m_{i1} respondent units and h_{i2} non-respondent units respectively for the study variable. Taking average of $\bar{x}_{0HHi}^{'s}$ over all the $f.s.u^{'s}$ selected in the sample, we find the estimator of population mean $\bar{X}_0$ as

$$(2.2) \qquad T_{0HH} = \frac{1}{n}\sum_{i}^{n} \bar{x}_{0HHi}$$

2.1. Properties of the estimator T_{OHH}. We have

$$(2.3) \qquad E\left[T_{0HH}\right] = E\left[E\left(T_{0HH}|n\right)\right] = E\left[E\left(\frac{1}{n}\sum_{i}^{n}\bar{x}_{0HHi}|i\right)\right]$$

$$= E\left[\bar{X}_{0i}\right] = \bar{\bar{X}}_0 \neq \bar{X}_0$$

We can easily understand that $\bar{X}_0 = \frac{1}{M_0}\sum_{i=1}^{N}\sum_{j=1}^{M_i} x_{0ij}$, $M_0 = \sum_{i=1}^{N} M_i$ and $\bar{\bar{X}}_0 = \frac{1}{N}$ $\sum_{i=1}^{N}\frac{1}{M_i}\sum_{j=1}^{M_i} x_{0ij} = \frac{1}{N}\sum_{i=1}^{N}\bar{X}_0$ where $x_{0ij}(x_{1ij})$ represents the observation on j^{th} s.s.u. in the ith f.s.u. for the study (auxiliary) variable (i=1,2,...,N;j=1,2,...M_i). Therefore, T_{0HH} is biased estimator of $\bar{X}_0$ and its bias is given by

$$(2.4) \qquad B\left[T_{0HH}\right] = \bar{\bar{X}}_0 - \bar{X}_0$$

The expression for mean square error (MSE) of the estimator T_{0HH} can be obtained as

$$(2.5) \qquad MSE(T_{0HH}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_0^{*2}$$

$$+ \frac{1}{nN}\sum_{i=1}^{N}\left\{\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{0i}^2 + \frac{(k_i - 1)}{m_i}W_{i2}S_{0i2}^2\right\}$$

$$+ \left[\bar{\bar{X}}_0 - \bar{X}_0\right]^2$$

where S_{0i}^2 and S_{0i2}^2 are respectively the mean squares per s.s.u. of entire group and non-response group of study variable in the ith f.s.u. further $S_0^{*2} = \frac{1}{N-1}\sum_{i=1}^{N}\left[\bar{X}_{0i} - \bar{\bar{X}}_0\right]^2$ and $W_{i2} = \frac{M_{i2}}{M_i}$ (non-response rate in i^{th} f.s.u.).

2.2. The unbiased estimator. We have seen that the estimator T_{0HH} is a biased estimator of population mean $\bar{X}_0$. Let us define an estimator parallel to T_{0HH} given in equation (2.2)

$$(2.6) \qquad T_{0HH}^{'} = \frac{1}{n}\sum_{i}^{n}u_i\bar{x}_{0HHi}$$

where $u_i = \frac{M_i}{\bar{M}}$ and $\bar{M} = \frac{1}{N}\sum_{i=1}^{N}M_i$ We can easily understand that $T_{0HH}^{'}$ is an unbiased estimator of $\bar{X}_0$ because

$$(2.7)$$
$$E\left[T_{0HH}^{'}\right] = E\left[E\left(T_{0HH}^{'}|n\right)\right] = E\left[E\left(\frac{1}{n}\sum_{i}^{n}u_i\bar{x}_{0HHi}|i\right)\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n}u_i\bar{X}_{0i}\right] = \frac{1}{N}\sum_{i=1}^{N}u_i\bar{X}_{0i} = \bar{X}_0$$

Now, the variance of $T_{0HH}^{'}$ is expressed as

$$(2.8) \qquad V\left[T_{0HH}^{'}\right] = \left(\frac{1}{n} - \frac{1}{N}\right)S_0^{*'2}$$

$$+ \frac{1}{nN}\sum_{i=1}^{N}u_i^2\left\{\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{0i}^2 + \frac{(k_i - 1)}{m_i}W_{i2}S_{0i2}^2\right\}$$

where $S_0^{*'2} = \frac{1}{N-1}\sum_{i=1}^{N}\left(u_i\bar{X}_{0i} - \bar{X}_0\right)^2$

3. Proposed families of estimators

Due to Singh and Shukla [10], a family of Factor-Type Estimators (FTE) for estimating the population mean $\bar{X}_0$ using an auxiliary variable, can be defined as (if there is no non-response in the population)

$$(2.9) \qquad T_\alpha = \bar{x}_0 \left[\frac{(A+C)\bar{X}_1 + fB\bar{x}_1}{(A+fB)\bar{X}_1 + C\bar{x}_1} \right]$$

where $\bar{x}_0 = \frac{1}{n}\sum_{i=1}^{n}\bar{x}_{0i}$ and $\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n}\bar{x}_{1i}$ are the means per s.s.u. in the sample for study and auxiliary variables respectively and $\bar{x}_{0i} = \frac{1}{m_i}\sum_{j=1}^{m_i}x_{0ij}$ and $\bar{x}_{1i} = \frac{1}{m_i}\sum_{j=1}^{m_i}x_{1ij}$. Further $f = \frac{n}{N}$, $A = (\alpha-1)(\alpha-2)$, $B = (\alpha-1)(\alpha-4)$ and $C = (\alpha-2)(\alpha-3)(\alpha-4)$; $\alpha > 0$.

We shall now define some families of estimators utilizing an auxiliary variable in the presence of non-response. Let us consider the cases:

(i) When non-response is present on study variable and auxiliary variable is free from non-response.

(ii) When non-response is present on both study and auxiliary variables.

3.1. Non-Response on study variable only. Using T'_{0HH} discussed in subsection 2.2, we now define the family of FTE in two-stage sampling with unequal size clusters under non-response as

$$(2.10) \quad T_\alpha^{**} = T'_{0HH} \left[\frac{(A+C)\bar{X}_1 + fB\bar{x}_1}{(A+fB)\bar{X}_1 + C\bar{x}_1} \right]$$

where the notations have already been defined earlier.

3.1.1. *Properties of the proposed family.* In order to obtain the bias and MSE of T_α^{**}, we use the large sample approximation. Let us consider

$T_{0HHi} = \bar{X}_0(1+e_0)$, $\bar{x}_1 = \bar{X}_1(1+e_1)$ such that $E(e_0) = E(e_1) = 0$, $E(e_0^2) = \frac{V(T'_{0HH})}{\bar{X}_0^2}$, $E(e_1^2) = \frac{V(\bar{x}_1)}{\bar{X}_1^2}$

and $E(e_0 e_1) = \frac{Cov(T'_{0HH}, \bar{x}_1)}{\bar{X}_0 \bar{X}_1}$

The bias of T_α^{**} up to the first order of approximation can be obtained as

$$B(T_\alpha^{**}) = E(T_\alpha^{**}) - \bar{X}_0 = \bar{X}_0\phi(\alpha)\left[\phi_2(\alpha)\frac{V(\bar{x}_1)}{\bar{X}_1^2} - \frac{Cov(T'_{0HH}, \bar{x}_1)}{\bar{X}_0\bar{X}_1}\right]$$

$$= \bar{X}_0\phi(\alpha)\left[\frac{\phi_2(\alpha)}{\bar{X}_1^2}\left\{\left(\frac{1}{n} - \frac{1}{N}\right)S_1^{*'2} + \frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{1i}^2\right\}\right]$$

$$- \bar{X}_0\phi(\alpha)\left[\frac{1}{\bar{X}_0\bar{X}_1}\left\{\left(\frac{1}{n} - \frac{1}{N}\right)S_{01}^{*'} + \frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{01i}\right\}\right]$$

$$(2.11) \quad B(T_\alpha^{**}) = \frac{\phi(\alpha)}{\bar{X}_1}\left[\left(\frac{1}{n} - \frac{1}{N}\right)\left\{\phi_2(\alpha)RS_1^{*'2} - S_{01}^{*'}\right\}\right]$$

$$+ \frac{\phi(\alpha)}{\bar{X}_1}\left[\frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left\{\phi_2(\alpha)RS_{1i}^2 - S_{01i}\right\}\right]$$

where S_{1i}^2 is the mean square per s.s.u. of entire group of auxiliary variable in the ith f.s.u. and $S_{01}^{*'} = \frac{1}{N-1}\sum_{i=1}^N \left(u_i\bar{X}_{0i} - \bar{X}_0\right)\left(u_i\bar{X}_{1i} - \bar{X}_1\right)$, $S_1^{*'2} = \frac{1}{N-1}\sum_{i=1}^N \left(u_i\bar{X}_{1i} - \bar{X}_1\right)^2$, $S_{01i} = \frac{1}{M_i}\sum_{j=1}^{M_i}\left(x_{0ij} - \bar{X}_{0i}\right)\left(x_{1ij} - \bar{X}_{1i}\right)$, $R = \frac{\bar{X}_0}{\bar{X}_1}$, $\phi(\alpha) = \frac{C-fB}{A+fB+C}$, $\phi_2(\alpha) = \frac{C}{A+fB+C}$.

The MSE of T_α^{**} up to the first order of approximation, can be derived as

$$MSE\left(T_\alpha^{**}\right) = E\left[T_\alpha^{**} - \bar{X}_0\right]^2$$

$$= \bar{X}_0^2\left[\frac{V\left(T_{0HH}'\right)}{\bar{X}_0^2} + \frac{\phi\left(\alpha\right)^2}{\bar{X}_1^2}V\left(\bar{x}_1\right) - 2\frac{\phi(\alpha)}{\bar{X}_0\bar{X}_1}Cov\left(T_{0HH}', \bar{x}_1\right)\right]$$

$$= \bar{X}_0^2\left[\frac{1}{\bar{X}_0^2}\left(\frac{1}{n} - \frac{1}{N}\right)S_0^{*'2}\right]$$

$$+ \bar{X}_0^2\left[\frac{1}{\bar{X}_0^2}\frac{1}{nN}\sum_{i=1}^N u_i^2\left\{\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{0i}^2 + \frac{(k_i-1)}{m_i}\frac{M_{i2}}{M_i}S_{0i2}^2\right\}\right]$$

$$+ \bar{X}_0^2\left[\frac{\phi(\alpha)^2}{\bar{X}_1^2}\left\{\left(\frac{1}{n} - \frac{1}{N}\right)S_1^{*'2} + \frac{1}{nN}\sum_{i=1}^N u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{1i}^2\right\}\right]$$

$$- \bar{X}_0^2\left[\frac{2\phi(\alpha)}{\bar{X}_0\bar{X}_1}\left\{\left(\frac{1}{n} - \frac{1}{N}\right)S_{01}^{*'} + \frac{1}{nN}\sum_{i=1}^N u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{01i}\right\}\right]$$

$$(2.12) \quad MSE\left(T_\alpha^{**}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_0^{*'2} + \phi(\alpha)^2 R^2 S_1^{*'2} - 2\phi(\alpha)R S_{01}^{*'}\right]$$

$$+ \frac{1}{nN}\sum_{i=1}^N u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left[S_{0i}^2 + \phi(\alpha)^2 R^2 S_{1i}^2 - 2\phi(\alpha)R S_{01i}\right]$$

$$+ \frac{1}{nN}\sum_{i=1}^N u_i^2\frac{k_i-1}{m_i}W_{i2}S_{0i2}^2$$

3.1.2. *Optimum choice of α and minimum MSE.* On minimizing MSE of T_α^{**} with respect to α, we get the optimum choice of α. Differentiating $MSE(T_\alpha^{**})$ with respect to α and equating the derivative to zero, we get

$$(2.13) \quad \left(\frac{1}{n} - \frac{1}{N}\right)\left[2\phi'(\alpha)\phi(\alpha)R^2 S_1^{*'2} - 2\phi'(\alpha)R S_{01}^{*'}\right] + \frac{1}{nN}\sum_{i=1}^N u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)$$

$$\left[2\phi'(\alpha)\phi(\alpha)R^2 S_{1i}^2 - 2\phi'(\alpha)R S_{01i}\right] = 0$$

$$(2.14) \quad \phi(\alpha) = \frac{\left(\frac{1}{n} - \frac{1}{N}\right)R S_{01}^{*'} + \frac{1}{nN}\sum_{i=1}^N u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)R S_{01i}}{\left(\frac{1}{n} - \frac{1}{N}\right)R^2 S_1^{*'2} + \frac{1}{nN}\sum_{i=1}^N u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)R S_{1i}^2}$$

Since the right hand side term in the above expression can be treated as a constant for given population, so the equation reduces to $\phi(\alpha) = \frac{C-fB}{A+fB+C} = V(say)$.

The above expression is a cubic equation in α which gives three optimum values of α. On putting the value of $\phi(\alpha)$ from the above equation in the expression of MSE of T_α^{**}, we get the estimator in the defined family possesses minimum MSE.

3.2. Non-Response on both study and auxiliary variables. Similar to $T_{0HH}^{'}$,we first define the unbiased estimator of population mean $\bar{X}_1$ in the presence of non-response using the Hansen-Hurwitz [5] technique

$$(2.15) \quad T_{1HH}^{'} = \frac{1}{n} \sum_{i}^{n} u_i \bar{x}_{1HHi}$$

where $\bar{x}_{1HHi} = \frac{m_{i1}\bar{x}_{1i1} + m_{i2}\bar{x}_{1i2}}{m_i}$

Further, $\bar{x}_{1i1}$ and $\bar{x}_{1i2}^{*}$ are the means per s.s.u. based on m_{i1} respondent units and h_{i2} non-respondent units respectively for the auxiliary variable. Thus, a family of FTE for estimating the population mean $\bar{X}_0$ under non-response, can be defined as

$$(2.16) \quad T_{\alpha}^{(**)} = T_{0HH}^{'} \left[\frac{(A+C)\bar{X}_1 + fBT_{1HH}^{'}}{(A+fB)\bar{X}_1 + CT_{1HH}^{'}} \right]$$

Remark: It is remembered that the FTE define a family of estimators including a number of well-known estimators existing in literature, namely, ratio, product, dual to ratio and usual sample mean estimators, which can be obtained for $\alpha=1,2,3$ and 4 respectively. It is, therefore, easy to generate similar estimators from the families defined by T_{α}^{**} and $T_{\alpha}^{(**)}$ in the proposed sampling scheme, that is, in two-stage sampling under non-response.

3.2.1. *Properties of proposed family.* Using large sample approximation, we can obtain the bias and MSE of $T_{\alpha}^{(**)}$. Let us assume
$$T_{0HH}^{'} = \bar{X}_0 (1 + e_0), \ T_{1HH}^{'} = \bar{X}_1 \left(1 + e_1^{'}\right)$$
such that
$$E(e_0) = E(e_1^{'}) = 0, \ E(e_0^2) = \frac{V(T_{0HH}^{'})}{\bar{X}_0^2}, \ E(e_1^{'2}) = \frac{V(T_{1HH}^{'})}{\bar{X}_1^2} \text{ and } E\left(e_0 e_1^{'}\right) = \frac{Cov\left(T_{0HH}^{'}, T_{1HH}^{'}\right)}{\bar{X}_0 \bar{X}_1}$$
Thus, the bias of $T_{\alpha}^{(**)}$ up to the first order of approximation is obtained as
$$B\left(T_{\alpha}^{(**)}\right) = E\left(T_{\alpha}^{(**)}\right) - \bar{X}_0 = \bar{X}_0 \phi(\alpha) \left[\phi_2(\alpha) \frac{V(T_{1HH}^{'})}{\bar{X}_1^2} - \frac{Cov\left(T_{0HH}^{'}, T_{1HH}^{'}\right)}{\bar{X}_0 \bar{X}_1}\right]$$
$$= \bar{X}_0 \phi(\alpha) \left[\frac{\phi_2(\alpha)}{\bar{X}_1^2} \left\{\left(\frac{1}{n} - \frac{1}{N}\right) S_1^{*'2} + \frac{\phi_2(\alpha)}{\bar{X}_1^2 nN} \sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{1i}^2 + \frac{(k_i-1)}{m_i} \frac{M_{i2}}{M_i} S_{1i2}^2\right\}\right]$$
$$-\bar{X}_0 \phi(\alpha) \left[\frac{1}{\bar{X}_0 \bar{X}_1 nN} \sum_{i=1}^{N} u_i^2 \left\{\left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{01i} + \frac{(k_i-1)}{m_i} \frac{M_{i2}}{M_i} S_{01i2}\right\}\right]$$

$$(2.17) \quad B\left(T_{\alpha}^{(**)}\right) = \frac{\phi_{\alpha}}{\bar{X}_1} \left[\left(\frac{1}{n} - \frac{1}{N}\right) \left\{\phi_2(\alpha) RS_1^{*'2} - S_{01}^{*'}\right\} \right.$$
$$+ \frac{1}{nN} \sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \left\{\phi_2(\alpha) RS_{1i}^2 - S_{01i}\right\} \right]$$
$$\left. - \frac{1}{nN} \sum_{i=1}^{N} u_i^2 \frac{(k_i-1)}{m_i} W_{i2} \left\{\phi_2(\alpha) RS_{1i2}^2 - S_{01i2}\right\}$$

where S_{1i2}^2 is the mean square per s.s.u. of non-response group for auxiliary variable in the ith f.s.u. and $S_{01i2} = \frac{1}{M_{i2}-1} \sum_{j}^{M_{i2}} \left(x_{0ij} - \bar{X}_{0i2}\right) \left(x_{1ij} - \bar{X}_{1i2}\right)$.
Further $\bar{X}_{0i2}$ and $\bar{X}_{1i2}$ are the means per s.s.u. of the non-response group for study and auxiliary variables respectively in the ith f.s.u. Now we can obtain the MSE of $T_{\alpha}^{(**)}$ up to the first order of approximation as

$$MSE\left(T_\alpha^{(**)}\right) = E\left[T_\alpha^{(**)} - \bar{X}_0\right]^2$$

$$= \bar{X}_0^2\left[\frac{V\left(T_{0HH}^{'}\right)}{\bar{X}_0^2} + \frac{\phi\left(\alpha\right)^2}{\bar{X}_1^2}V\left(T_{1HH}^{'}\right) - 2\frac{\phi\left(\alpha\right)}{\bar{X}_0\bar{X}_1}Cov\left(T_{OHH}^{'},T_{1HH}^{'}\right)\right]$$

$$= \bar{X}_0^2\left[\frac{1}{\bar{X}_0^2}\left(\frac{1}{n} - \frac{1}{N}\right)S_0^{*'2}\right]$$

$$+ \bar{X}_0^2\left[\frac{1}{\bar{X}_0^2}\frac{1}{nN}\sum_{i=1}^{N}u_i^2\left\{\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{0i}^2 + \frac{(k_i-1)}{m_i}\frac{M_{i2}}{M_i}S_{0i2}^2\right\}\right]$$

$$+ \bar{X}_0^2\left[\frac{\phi(\alpha)^2}{\bar{X}_1^2}\left\{\left(\frac{1}{n} - \frac{1}{N}\right)S_1^{*'2}\right\}\right]$$

$$+ \bar{X}_0^2\left[\frac{\phi(\alpha)^2}{\bar{X}_1^2}\left(\frac{\phi(\alpha)^2}{\bar{X}_1nN}\right)\sum_{i=1}^{N}u_i^2\left\{\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{1i}^2 + \frac{(k_i-1)}{m_i}\frac{M_{i2}}{M_i}S_{1i2}^2\right\}\right]$$

$$- \bar{X}_0^2\left[\frac{\phi(\alpha)^2}{\bar{X}_1^2}\left(\frac{2\phi(\alpha)}{\bar{X}_0\bar{X}_1}\right)\left(\frac{1}{n} - \frac{1}{N}\right)S_{01}^{*'}\right]$$

$$- \bar{X}_0^2\left[\frac{\phi(\alpha)^2}{\bar{X}_1^2}\left(\frac{2\phi(\alpha)}{\bar{X}_0\bar{X}_1nN}\right)\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{01i} + \frac{(k_i-1)}{m_i}\frac{M_{i2}}{M_i}S_{01i2}\right]$$

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

(2.18)
$$+ \frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left[S_{0i}^2 + \phi(\alpha)^2R^2S_{1i}^2 - 2\phi(\alpha)RS_{01i}\right]$$

$$+ \frac{1}{nN}\sum_{i=1}^{N}u_i^2\frac{k_i-1}{m_i}W_{i2}\left[S_{0i2}^2 + \phi(\alpha)^2R^2S_{1i2}^2 - 2\phi(\alpha)RS_{01i2}\right]$$

3.2.2. *Optimum choice of α.* To obtain optimum estimator within the proposed family, we should obtain the optimum value of α. Differentiating $MSE\left(T_\alpha^{(**)}\right)$ with respect to α and equating the derivative to zero, we get
(2.19)

$$\left(\frac{1}{n} - \frac{1}{N}\right)\left[2\phi^{'}(\alpha)\phi(\alpha)R^2S_{1i}^{*'2} - 2\phi^{'}(\alpha)RS_{01}^{*'}\right]$$

$$+ \frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left[2\phi^{'}(\alpha)\phi(\alpha)R^2S_{1i}^2 - 2\phi^{'}(\alpha)RS_{01i}\right]$$

$$+ \frac{1}{nN}\sum_{i=1}^{N}u_i^2\frac{(k_i-1)}{m_i}W_{i2}\left[2\phi^{'}(\alpha)\phi(\alpha)R^2S_{1i2}^2 - 2\phi^{'}(\alpha)RS_{01i2}\right] = 0$$

(2.20) $\quad \phi(\alpha) = \dfrac{\left(\frac{1}{n} - \frac{1}{N}\right)RS_{01}^{*'} + \frac{1}{nN}\sum_{i=1}^{N}u_i^2\left\{\left(\frac{1}{m_i} - \frac{1}{M_i}\right)RS_{01i} + \frac{(k_i-1)}{m_i}W_{i2}R^2S_{1i2}^2RS_{01i2}\right\}}{\left(\frac{1}{n} - \frac{1}{N}\right)R^2S_1^{*'2} + \frac{1}{nN}\sum_{i=1}^{N}u_i^2\left\{\left(\frac{1}{m_i} - \frac{1}{M_i}\right)R^2S_{1i}^2 + \frac{(k_i-1)}{m_i}W_{i2}R^2S_{1i2}^2\right\}}$

which is a cubic equation in and hence giving three optimum values of α, for a given population.

4. Empirical Study

We have considered the $CO124$ population available in Sarndal *et al.* [9].There are 124 countries $(s.s.u^{'s})$ divided into 7 continents $(f.s.u^{'s})$ according to locations. Since 7th continent consists of only one country therefore, we placed it in the 6th continent. Here we considered the import and export in 1983 as study and auxiliary variables respectively. Thus the data are divided into 6 clusters. We have N, $M_0 = \sum_N^{i=1} M_i = 124$ and $\bar{M} = 20.67$.

The following tables 1 and 2 show the characteristics of the population.

Table 1. Cluster Sizes, Means and Mean-squares under Study and Auxiliary Variables (X_0, X_1)

Cluster No.	M_i	m_i	u_i	$\bar{X}_{0i}$	$\bar{X}_{1i}$	S_{oi}^2	S_{1i}^2
1	38	15	1.8387	2254.45	1901.11	15030175.51	1356941.51
2	14	6	0.6774	25533.14	22083.21	5599280338	3256997379.00
3	11	4	0.5323	3602.82	5835.46	19221733.92	74298983.08
4	33	13	1.5968	12156.79	12438.85	526648674.70	711462797.70
5	24	10	1.1613	28493.46	30698.00	938699360.40	1599448820.00
6	4	2	0.1936	26392.50	29360.50	1814997631.0	2376032656.00

Table 2. Cluster Covariances, Mean Ratios, Correlations and Mean-squares for Non-response Groups

Cluster No.	S_{01i}	R_{01i}	ρ_{01i}	$S_{0i2}^2 = \frac{4}{5}(S_{0i}^2)$	$S_{1i2}^2 = \frac{4}{5}(S_{1i}^2)$	$S_{o1i2} = \frac{4}{5}S_{01i}$
1	12035657.14	1.1859	0.865	12024140.41	108695553.21	9628525.71
2	3920918987.0	1.1562	0.989	4479424270.0	2605597903.0	3136735190.0
3	33223013.79	0.6174	0.967	15377387.14	59439186.46	26578411.03
4	586829812.3	0.9773	0.989	421318939.80	569170238.20	469463849.80
5	594135373.70	0.9282	0.506	750959488.30	1279559056.00	475308299.00
6	1557362451.0	0.8989	1.00	1451998105.0	1900826125.00	1245889961.0

Further, we have $\bar{X}_0$=13494.77, $\bar{X}_1$=13792.52 ,R=0.9784 , $S_0^{*'2}$=145036370.90 ,$S_1^{*'2}$=160383672.30,$S_{01}^{*'}$=151324910.80.
On the basis of the above data, using the equation (2.14) with $n = 2$, we get $\phi(\alpha)$=0.9112 for the estimator T_α^{**}.In the Table 3, we have depicted values of $MSE(T_\alpha^{**})$ for optimum choice of α and for α=1, 2, 3 and 4 for comparison purpose.

While using the family of estimators given by $T_\alpha^{(**)}$, we have different values of $\phi(\alpha)$ for different combinations of W_{i2} and K_i, due to expression (2.20). The Table 4 shows the values of , which have been obtained through (2.20).

Similar to Table 3, Table 5 presents the comparison of $MSE(T_\alpha^{(**)})$ for different combinations of α, W_{i2} and K_i.

Table 3. *MSE* Comparison for the Family T_α^{**}

W_{i2} (for all i)	α	K_i (for all i)		
		1.5	2.0	2.5
	1	17062949.00	19385874.64	21708800.28
	2	314134864.80	316457790.40	318780716.00
0.1	3	30200211.18	32523136.82	34846062.46
	4	84091284.54	86414210.18	88737135.82
	α_{opt}	16419911.76	18742837.40	21065763.04
	1	19385874.64	24031725.92	28677577.20
	2	316457790.40	321103641.70	325749493
0.2	3	32523136.82	37168988.10	41814839.38
	4	86414210.18	91060061.46	95705912.74
	α_{opt}	18742837.40	23388688.68	28034539.96
	1	21708800.28	28677577.20	35646354.11
	2	318780716.00	325749493.00	332718269.90
0.3	3	34846062.46	41814839.38	48783616.30
	4	88737135.82	95705912.74	102674689.70
	α_{opt}	21065763.04	28034539.96	35003316.87
	1	24031725.92	33323428.47	42615131.03
	2	321103641.70	330395344.20	339687046.80
0.4	3	37168988.10	46460690.66	55752392.22
	4	91060061.46	100351764.00	109643466.60
	α_{opt}	23388688.68	32680391.23	41972093.79

Table 4. Values of $\phi(\alpha)$ for the estimator T_α^{**}

W_{i2} (for all i)	K_i (for all i)		
	1.5	2.0	2.5
0.1	0.9090	0.9070	0.9049
0.2	0.9069	0.9030	0.8994
0.3	0.9049	0.8994	0.8904
0.4	0.9030	0.8962	0.8904

5. Conclusion

The problem of estimating the population mean in two-stage sampling with unequal size clusters under non-response utilizing the auxiliary information, has been discussed. We have proposed the families of estimators whenever the non-response is present on study variable only and on both study and auxiliary variables respectively. The proposed families can generate non-response versions of a number of well-known estimators existing in literature, namely, ratio, product, dual to ratio and usual mean estimators. The properties of the proposed families have been discussed along with its optimum estimators. The tables 3 and 5 reveal that MSE_s of the estimators of the families T_α^{**} and $T_\alpha^{(**)}$ for optimum α are smaller than that obtained for α=1, 2, 3 and 4.From Table 3,we observe that for a given value of the parameter, α,MSE, of T_α^{**}increases with increase in

Table 5. MSE Comparison for the Family $T_\alpha^{(**)}$

W_{i2} (for all i)	α	K_i (for all i)		
		1.5	2.0	2.5
0.1	1	15800784.47	16861545.58	17922306.70
	2	319778183.80	327744428.50	335710673.20
	3	29021484.61	30165683.67	31309882.73
	4	84091284.54	86414210.18	88737135.82
	a	15091783.93	16085095.53	17077029.71
0.2	1	16861545.58	18983067.81	21104590.04
	2	327744428.50	343676917.90	359609407.20
	3	3016683.67	32454081.79	34742479.92
	4	86414210.18	91060061.46	95705912.74
	a	16085095.53	18067684.89	20045507.25
0.3	1	17922306.70	21104590.04	24286873.38
	2	335710673.20	359609407.20	383508141.30
	3	31309882.73	34742479.92	38175077.10
	4	88737135.82	95705912.74	102674689.70
	a	17077029.71	20045507.25	23004634.46
0.4	1	18983067.81	23226112.27	27469156.72
	2	343676917.90	375541896.60	407406875.30
	3	32454081.79	37030878.04	41607674.29
	4	91060061.46	100351764.00	109643466.60
	a	18067684.89	22019184.66	25956095.25

non-response rate and also with smaller size of sub-samples of non-respondents. Further, for the estimator $T_\alpha^{(**)}$, the same trend is seen in Table 5. The result is also intuitively expected.

References

[1] Ahmed, M.S. *Some estimators for a finite population mean under two-stage sampling using multivariate auxiliary information*, Applied Mathematics and Computation, 153(2), 505 − 511, 2004.

[2] Bellhouse, D. R. and Rao, J. N. K. *On the efficiency of prediction estimator in two-stage sampling*, Journal of Statistical Planning and Inference, 13, 269 − 281, 1986.

[3] Chaudhary, M. K. and Singh, V. K. *Estimating the population mean in two-stage sampling with equal size clusters under non-response using auxiliary characteristic*, Mathematical Journal of Interdisciplinary Sciences, 2(1), 43 − 55, 2013.

[4] Eideh, A. and Nathan, G. *Two-stage informative cluster sampling-estimation and prediction with applications for small-area models*, Journal of Statistical Planning and Inference, 139(9), 3088 − 3101, 2009.

[5] Hansen, M. H. and Hurwitz, W. N. *The problem of non-response in sample surveys*, Journal of the American Statistical Association, 41, 517 − 529, 1946.

[6] Khan, A. and Muttlak, H. A. *Adjusted two-stage adaptive cluster sampling*, Applied Mathematics and Computation, 126(1), 83 − 95, 2002.

[7] Sahoo, L. N., Das, B. C. and Sahoo, J. *A class of predictive estimators in two stage sampling*, Journal of the Indian Society of Agricultural Statistics, 63(2), 175 − 180, 2009.

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

[9] Sarndal, C. E., Swensson, B. and Wretman, J. *Model Assisted Survey Sampling*, Springer-Verlag, New York, Inc. Appendix D, 662, 1992.

[10] Singh, V. K. and Shukla, D. *One parameter family of factor-type ratio estimators,* Metron,45(1 − 2), 273 − 283, 1997.