Modified Regression Type Estimators in the Presence of Non-Response in Two Phase sampling

Grace Chumba[1], John Kung'u[2]

[1] Kabarak University, School of Education, Private Bag – 20157, Kabarak – Kenya
[2] Laikipia University, School of Science and Applied Technology, Box 1100-20300, Nyahururu-Kenya

Abstract:

It is a common experience in sample survey that data cannot always be collected for all units selected in the sample at the first attempt and even after some call-backs. An estimate obtained from such incomplete data may be misleading because of the non-response in the data. In addition, the population mean of the auxiliary variable from the previous census may not be available. In this paper, Modified regression type estimators proposed by Tum *et al.* (2014) in single phase sampling, assuming complete response, have been proposed to estimate the population mean of the study variable in the presence of non-response under two phase sampling scheme. The expression of mean squared errors (MSE) based on the proposed estimators have been derived under two phase sampling to the first degree of approximation. A comparison of the proposed estimators with the usual unbiased estimator and existing estimators under two phase sampling scheme have been carried out. The proposed Modified regression type estimators have been found to be the most efficient compared to the existing estimators and they are recommended for use in practice.

Keywords: Modified regression type estimators, Study variable, Auxiliary variables, Mean Square Error, Non-response.

1. Introduction

It is a common experience in sample survey that data cannot always be collected for all units selected in the sample. For example in a household survey, the families may not be found at home at the first attempt and some may refuse to cooperate with the interviewer even if contacted. In general, during surveys, it is observed that information in most cases is not obtained at the first attempt even after some call-backs. The failure to measure or get information from some of the units in the selected sample is referred to as non-response. An estimate obtained from such incomplete data may be misleading because of the non-response. There are various practical reasons for this incomplete information due to non-response like unwillingness of the respondent to answer some particular questions, accidental loss of information caused by unknown factors and failure on the part of investigator to collect the correct information, etc. The usual approach to face the non-response is

to re-contact the non-respondents and obtain the information. Hansen and Hurwitz (1946) were the first authors to contract the problem of non-response while conducting mail surveys. They suggested a technique, known as 'sub-sampling of non-respondents', to deal with the problem of non-response and its adjustments. In fact they developed an unbiased estimator for population mean in the presence of non-response by dividing the population into two groups, viz. response group and non-response group. To avoid bias due to non-response, they suggested taking a sub-sample of the non-responding units. Neyman (1938)

developed the concept of double phase sampling in sample survey. This sampling scheme is used to obtain the information about auxiliary variable cheaply from a larger sample at first phase and relatively small sample at the second stage. Khare and Srivastava (1993,1995) proposed ratio, product and regression estimators when population

mean of auxiliary variable X is known and non-response occur only in the study variable in two phase sampling. Singh *et al*. (2010) proposed Exponential ratio and Exponential product estimators when population mean of auxiliary variable X is known and non-response occur only in the study variable in two phase sampling. Kumar and Bhougal (2011) proposed Exponential ratio-product estimator when population mean of auxiliary variable X is known and non-response occur only in the study variable in two phase sampling. Khare and Srivastava (1993, 1995) also proposed ratio, product and regression estimators when population mean of auxiliary variable X is unknown and non-response occurs in both the study variable and auxiliary variable in two phase sampling. Singh *et al*. (2010) also proposed Exponential ratio and Exponential product estimators when population mean of auxiliary variable X is unknown and non-response occur occurs in both the study variable and auxiliary variable in two phase sampling. Kumar and Bhougal (2011) also proposed Exponential ratio-product estimator when population mean of auxiliary variable X is unknown and non-response occurs in both the study variable and auxiliary variables in two phase sampling. Tum *et al.* (2014) proposed a regression type estimator that combined regression estimator and ratio - product type exponential estimator to form a new modified regression estimator. This study will focus on Tum *et al.* (2014) in presence of non-response in the study variable only and in both the study variable and auxiliary variables and develop their mean squared errors by incorporate Arora and Lal (1989) approach in writing down the mean squared error.

2. Notation

Consider a finite population of size N and a random sample of size n drawn without replacement. In surveys on human populations, frequently n_1 units respond on the items under examination in the first attempt while remaining $n_2 (= n - n_1)$ units do not provide any response. When non-response occurs in the initial attempt, Hansen and Hurwitz (1946) proposed a double sampling scheme for estimating population mean comprising the following steps:

i) a simple random sample of size n is selected and the questionnaire is mailed to the sampled units;
ii) a sub sample of size $r = (n_2/k)$, $(k \geq 1)$ from the n_2 non-responding units in the initial attempt is contacted through personal interviews.

Consider a finite population of size N. We draw a sample of size n from a population by using simple random sample without replacement (SRSWOR) sampling scheme. Let y_i and x_i be the observations on the study variable (y) and the auxiliary variable (x) respectively. Let $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$ denote the population mean and the population variance of the survey variable y. When information on $\bar{X}$ is unknown then double sampling or two phase sampling is suitable to estimate the population mean. In the first phase sample we select a sample of size n' by SRSWOR from a population to observe x. In the second phase, we select a sample of size n from n' such that $n < n'$ by SRSWOR also. Non-response occurs on second phase in which n_1 units respond and n_2 do not out of n units. From n_2 non-respondents a sample of $r = (n_2/k)$, $(k > 1)$ units are selected, where k is the inverse sampling rate at the second phase sample of size n. Let $\bar{Y}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} y_i$ and $S_{y(1)}^2 = \frac{1}{N_1 - 1}\sum_{i=1}^{N_1}(y_i - \bar{Y}_1)^2$ denote the mean and variance of the response group. Similarly, $\bar{Y}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} y_i$ and $S_{y(2)}^2 = \frac{1}{N_2 - 1}\sum_{i=1}^{N_2}(y_i - \bar{Y}_2)^2$ denote the mean and variance of the non-response group. The population mean can be written as $\bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2$, where $W_1 = (N_1/N)$ and $W_2 = (N_2/N)$. Let $\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i$ and $\bar{y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_i$ denote the means of the n_1 responding units and the n_2 non-responding units. Further, let $\bar{y}_{2r} = \frac{1}{r}\sum_{i=1}^{r} y_i$ denote the mean of the $r = n_2/k$ sub sampled units.

Thus, an unbiased estimator, due to Hansen and Hurwitz (1946) of the population mean $\bar{Y}$ of the study variable y is given by,

$$\overline{y}^* = w_1 \overline{y}_1 + w_2 \overline{y}_{2r} \qquad (1)$$

Where $w_1 = (n/n_1)$, $w_2 = (n/n_2)$ are responding and non-responding proportions in the sample. The variance of $\overline{y}^*$ to terms of order n^{-1} is given by,

$$V\left(\overline{y}^*\right) = \overline{Y}^2 \left\{ \theta C_y^2 + \lambda C_{y(2)}^2 \right\} \qquad (2)$$

where $\theta = \dfrac{N-n}{Nn}$, $\lambda = \dfrac{W_2(k-1)}{n}$,

$C_y^2 = \left(S_y^2 / \overline{Y}^2\right)$ and $C_{y(2)}^2 = \left(S_{y(2)}^2 / \overline{Y}^2\right)$

Let $\overline{x}_1 = \dfrac{1}{n_1}\sum\limits_{i=1}^{n_1} x_i$ and $\overline{x}_2 = \dfrac{1}{n_2}\sum\limits_{i=1}^{n_2} x_i$ denote the means of the auxiliary variable of n_1 responding units and the n_2 non-responding units. Further, let

$\overline{x}_{2r} = \dfrac{1}{r}\sum\limits_{i=1}^{r} x_i$ denote the mean of the $r = n_2/k$ sub sampled units. Thus, an unbiased estimator, due to Hansen and Hurwitz (1946) of the population mean $\overline{Y}$ of the study variable y is given by.

$$\overline{x}^* = w_1 \overline{x}_1 + w_2 \overline{x}_{2r} \qquad (3)$$

where $w_1 = (n/n_1)$, $w_2 = (n/n_2)$ are responding and non-responding proportions in the sample. The variance of $\overline{y}^*$ to terms of order n^{-1} is given by,

$$V\left(\overline{x}^*\right) = \overline{X}^2 \left\{ \theta C_x^2 + \lambda C_{x(2)}^2 \right\} \qquad (4)$$

where $C_x^2 = \left(S_x^2 / \overline{X}^2\right)$, $C_{x(2)}^2 = \left(S_{x(2)}^2 / \overline{X}^2\right)$,

$S_x^2 = \dfrac{1}{N-1}\sum\limits_{i=1}^{N}\left(x_i - \overline{X}\right)^2$ and $S_{x(2)}^2 = \dfrac{1}{N_2-1}\sum\limits_{i=1}^{N_2}\left(x_i - \overline{X}_2\right)^2$

The expectation includes,

$\overline{y} = \overline{Y}\left(1+e_y\right)$ $\overline{x} = \overline{X}\left(1+e_x\right)$ $\overline{z} = \overline{Z}\left(1+e_z\right)$

$E(\overline{e}_0) = E(\overline{e}_1) = 0$

$E(\overline{e}_0^2) = V\left(\overline{y}^*\right) = \overline{Y}^2 \left\{ \theta C_y^2 + \lambda C_{y(2)}^2 \right\}$

$E(\overline{e}_1^2) = V\left(\overline{x}^*\right) = \overline{X}^2 \left\{ \theta C_x^2 + \lambda C_{x(2)}^2 \right\}$

$E(\overline{e}_1 \overline{e}_0) = V\left(\overline{y}^*, \overline{x}^*\right) = \overline{Y}\,\overline{X} \left\{ \theta \rho_{yx} C_x C_y + \lambda \rho_{yx(2)} C_{x(2)} C_{y(2)} \right\}$

Where $\rho_{yx} = S_{yx} / S_y S_x$, $\rho_{yx(2)} = S_{yx(2)} / S_{y(2)} S_{x(2)}$

$\rho_2 = \dfrac{S_{2xy}}{S_{2y}S_{2x}}$ $k_{yx} = \rho_{yx}\dfrac{C_y}{C_x}$ $k_{yx2} = \rho_{yx2}\dfrac{C_{y2}}{C_{x2}}$

$S_{yx} = \dfrac{1}{N-1}\sum\limits_{i=1}^{N}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)$ $S_{yx(2)} = \dfrac{1}{N_2-1}\sum\limits_{i=1}^{N_2}\left(y_i - \overline{Y}_2\right)\left(x_i - \overline{X}_2\right)$

$S_{2yx} = \dfrac{1}{N-1}\sum\limits_{i=1}^{N}\left(y_i - \overline{Y}_2\right)\left(x_i - \overline{X}_2\right)$ $S_{2x} = \dfrac{1}{N-1}\sum\limits_{i=1}^{N}\left(x_i - \overline{X}_2\right)^2$

$S_{2y} = \dfrac{1}{N-1}\sum\limits_{i=1}^{N}\left(y_i - \overline{Y}_2\right)^2$ $\boldsymbol{\Theta} = \dfrac{N-n}{Nn}$,

$\theta' = \dfrac{n'-n}{n'n}$, $\lambda = \dfrac{N-n'}{Nn'}$ and $\delta = \dfrac{W_2(k-1)}{n}$

$A^{-1} = \dfrac{1}{|A|}\left(C^T\right)_{ij} = \dfrac{Adj(A)}{|A|}$

$\dfrac{|R|_{yx_q}}{|R|_{x_q}} = \left(1 - \rho^2_{y.x_q}\right)$ \qquad Arora and Lal (1989)

$$(5)$$

The following notations will be used in deriving the mean square errors of proposed estimator

$|R|_{yx_p}$ Determinant of population correlation matrix of variables y, x_1, x_2

$|R|_{x_p}$ Determinant of population correlation matrix of variables x_1, x_2

3.1 CASE 1: When there is incomplete information on the study variable *y* and complete information on the auxiliary variable *x*.

When the population mean $\overline{X}$ of the auxiliary variable 'x' is known, the two phase ratio estimator by Khare and Srivastava (1993,1995) is given by,

$$t_{R1} = \overline{y}^* \frac{\overline{x}'}{\overline{x}} \qquad (6)$$

The MSE of the estimator t_{R1};

$$MSE(t_{R1}) = \overline{Y}^2 \left[\theta C_y^2 + (\theta - \lambda) C_y^2 \left(1 - 2k_{yx}\right) C_x^2 + \delta C_{y(2)}^2 \right]$$
$$(7)$$

When the population mean $\overline{X}$ of the auxiliary variable 'x' is known, the two phase product estimator by Khare and Srivastava (1993,1995) is given by,

$$t_{P1} = \overline{y}^* \frac{\overline{x}}{\overline{x}'} \qquad (8)$$

The MSE of the estimator t_{P1};

$$MSE(t_{P1}) = \overline{Y}^2 \left[\theta C_y^2 + (\theta - \lambda) C_y^2 \left(1 + 2k_{yx}\right) C_x^2 + \delta C_{y(2)}^2 \right] \quad (9)$$

When the population mean $\overline{X}$ of the auxiliary variable 'x' is known, the two phase regression estimator by Khare and Srivastava (1993,1995) is given by,

$$t_{RE1} = \bar{y}^* + \beta\left(\bar{x}' - \bar{x}\right) \qquad (10)$$

The MSE of the estimator t_{RE1};

$$MSE(t_{RE1}) = \bar{Y}^2\left(\theta C_y^2 + (\theta - \lambda)\left(1 - \rho_{yx}^2\right) + \delta C_{y(2)}^2\right) \qquad (11)$$

When the population mean $\bar{X}$ of the auxiliary variable 'x' is known, the two phase exponential ratio estimator by Singh *et al*. (2010) is given by,

$$t_{ER1} = \bar{y}^* \exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right) \qquad (12)$$

The MSE of the estimator t_{ER1};

$$MSE(t_{ER1}) = \bar{Y}^2\left[\theta C_y^2 + (\theta - \lambda)\left\{C_y^2 + \frac{1}{2}(1 - 2k_{yx})C_x^2\right\} + \delta C_{y(2)}^2\right] \qquad (13)$$

When the population mean $\bar{X}$ of the auxiliary variable 'x' is known, the two phase exponential product estimator by Singh *et al*. (2010) is given by,

$$t_{EP1} = \bar{y}^* \exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right) \qquad (14)$$

The MSE of the estimator t_{EP1};

$$MSE(t_{EP1}) = \bar{Y}^2\left[\theta C_y^2 + (\theta - \lambda)\left\{C_y^2 + \frac{1}{2}(1 + 2k_{yx})C_x^2\right\} + \delta C_{y(2)}^2\right] \qquad (15)$$

When the population mean $\bar{X}$ of the auxiliary variable 'x' is known, the two phase exponential ratio-product estimator by Kumar and Bhougal (2011) is given by,

$$t_{ERP1} = \bar{y}^* \exp\left(\alpha\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right) + (1 - \alpha)\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right)\right) \qquad (16)$$

The MSE of the estimator t_{ERP1};

$$MSE(t_{ERP1}) = \bar{Y}^2\left[\theta\left\{C_y^2 + \frac{B_2}{A_2}\left(\frac{B_2}{A_2} - 2k_{yx}\right)C_x^2\right\} + \delta C_{y(2)}^2\right] \qquad (17)$$

Where

$$A_2 = \theta C_x^2, \quad B_2 = \theta k_{yx} C_x^2 \quad \text{and} \quad \alpha = \frac{A_2 + 2B_2}{2A_2}$$

3.1.1 Proposed Modified Regression Type Estimator in Two Phase Sampling in the Presence of Non-Response in the Study Variable Only.

The modified regression estimator proposed by Tum *et al*. (2014) is given by,

$$t_{ERP} = \left(\bar{y} + \beta\left(\bar{X} - \bar{x}\right)\right)\left(\alpha\exp\left(\frac{\bar{Z} - \bar{z}}{\bar{Z} + \bar{z}}\right) + (1 - \alpha)\exp\left(\frac{\bar{z} - \bar{Z}}{\bar{z} + \bar{Z}}\right)\right) \qquad (18)$$

We will consider the above estimator in presence of non-response in the study variable only and in both the study variable and auxiliary variables. When there is incomplete information on the study variable y and complete information on the auxiliary variable x, our proposed Modified regression type estimator in the presence of non-response will be given by,

$$t_7 = \left(\bar{y}_2^* + \beta\left(\bar{X} - \bar{x}_2\right)\right)\left(\alpha\exp\left(\frac{\bar{Z} - \bar{z}_2}{\bar{Z} + \bar{z}_2}\right) + (1 - \alpha)\exp\left(\frac{\bar{z}_2 - \bar{Z}}{\bar{Z} + \bar{z}_2}\right)\right) \qquad (19)$$

Substituting (5) equation in (19) we get,

$$t_7 = \left(\bar{Y}\left(1 + \bar{e}_{y_2}^*\right) - \beta\bar{e}_{x_2}\right)\left(\begin{array}{c}\alpha\exp\left(\frac{\bar{Z} - \bar{Z}\left(1 + \bar{e}_{z_2}\right)}{\bar{Z} + \bar{Z}\left(1 + \bar{e}_{z_2}\right)}\right) \\ + (1 - \alpha)\exp\left(\frac{\bar{Z}\left(1 + \bar{e}_{z_2}\right) - \bar{Z}}{\bar{Z} + \bar{Z}\left(1 + \bar{e}_{z_2}\right)}\right)\end{array}\right) \qquad (20)$$

Simplifying (20), we get

$$t_7 = \left(\bar{e}_{y_2}^* + \bar{Y} - \beta\bar{e}_{x_2}\right)\left(\begin{array}{c}\alpha\exp\left(-\frac{\bar{e}_{z_2}}{2}\left(1 + \frac{\bar{e}_{z_2}}{2}\right)^{-1}\right) \\ + (1 - \alpha)\exp\left(\frac{\bar{e}_{z_2}}{2}\left(1 + \frac{\bar{e}_{z_2}}{2}\right)^{-1}\right)\end{array}\right) \qquad (21)$$

Expanding and ignoring the second and higher terms for each expansion of product and after simplification we can write

$$t_7 = \bar{Y} + \bar{e}_{y_2}^* - \beta\bar{e}_{x_2} + \bar{Y}\left(\frac{1}{2} - \alpha\right)\bar{e}_{z_2} \qquad (22)$$

The mean squared error of t_7 is given by,

$$MSE(t_7) = E\left(t_7 - \bar{Y}\right)^2 = E\left(\bar{e}_{y_2}^* - \beta\bar{e}_{x_2} + \bar{Y}\left(\frac{1}{2} - \alpha\right)\bar{e}_{z_2}\right)^2 \qquad (23)$$

Squaring the right sides of (23) and taking expectation and using (5), we get,

$$MSE(t_7) = \theta\bar{Y}^2 C_y^2 + \lambda\bar{Y}^2 C_{y(2)}^2 + \beta^2\theta\bar{X}^2 C_x^2$$

$$+ \bar{Y}^2\left(\frac{1}{2} - \alpha\right)^2\theta\bar{Z}^2 C_z^2 + 2\bar{Y}\left(\frac{1}{2} - \alpha\right)\theta\bar{X}\bar{Y}C_y C_z\rho_{yz} \qquad (24)$$

$$- 2\beta\theta\bar{X}\bar{Y}C_y C_x\rho_{yx} - 2\bar{Y}\left(\frac{1}{2} - \alpha\right)\beta\theta\bar{X}\bar{Z}C_x C_z\rho_{xz}$$

Expanding (24) by opening brackets we get,

$$MSE(t_7) = \theta \bar{Y}^2 C_y^2 + \lambda \bar{Y}^2 C_{y(2)}^2 + \beta^2 \theta \bar{X}^2 C_x^2 + \frac{1}{4} \bar{Y}^2 \theta \bar{Z}^2 C_z^2$$

$$- \alpha \bar{Y}^2 \theta \bar{Z}^2 C_z^2 + \bar{Y}^2 \alpha^2 \theta \bar{Z}^2 C_z^2 + \bar{Y}^2 \theta \bar{X} C_y C_z \rho_{yz} \qquad (25)$$

$$- \alpha 2 \bar{Y}^2 \theta \bar{Z} C_y C_z \rho_{yz} - 2 \beta \theta \bar{X} \bar{Y} C_y C_x \rho_{yx} - \bar{Y} \beta \theta \bar{X} \bar{Z} C_x C_z \rho_{xz}$$

$$+ \alpha 2 \bar{Y} \beta \theta \bar{X} \bar{Z} C_x C_z \rho_{xz}$$

Differentiating (25) with respect to α and β and equating to zero gives

$$\beta = \frac{\bar{Y} C_y}{\bar{X} C_x \left| R_{2\times2} \right|} (-1)^{1+1} \left| R_{yx} \right|_{yxz}$$

$$\alpha = \frac{1}{2} + \frac{C_y}{\bar{Z} C_z \left| R_{2\times2} \right|} (-1)^{2+1} \left| R_{yz} \right|_{yxz} \qquad (26)$$

Using normal equations that are used to find the optimum values of α and β (26) can be written in simplified form as

$$MSE(t_7) = E\left[\bar{e}_y^* \left(\bar{e}_y^* - \beta \bar{e}_x + \bar{Y} \left(\frac{1}{2} - \alpha \right) \bar{e}_z \right) \right] \qquad (27)$$

Simplifying and taking expectation, we get,

$$MSE(t_7) = \theta \bar{Y}^2 C_y^2 + \lambda \bar{Y}^2 C_{y(2)}^2$$
$$- \beta \theta \bar{X} \bar{Y} C_y C_x \rho_{yx} + \bar{Y} \alpha \theta \bar{Y} \bar{Z} C_y C_z \rho_{yz} \qquad (28)$$

Substituting the optimum value in (26), we get,

$$MSE(t_7) = \theta \bar{Y}^2 C_y^2 + \lambda \bar{Y}^2 C_{y(2)}^2$$
$$- \left(\frac{\bar{Y} C_y}{\bar{X} C_x \left| R_{2\times2} \right|} (-1)^{1+1} \left| R_{yx} \right|_{yxz} \right) \theta \bar{X} \bar{Y} C_y C_x \rho_{yx} \qquad (29)$$
$$+ \bar{Y} \left(\frac{1}{2} - \left(\frac{1}{2} + \frac{C_y}{\bar{Z} C_z \left| R_{2\times2} \right|} (-1)^{2+1} \left| R_{yz} \right|_{yxz} \right) \right) \theta \bar{Y} \bar{Z} C_y C_z \rho_{yz}$$

Simplifying (29) we get

$$MSE(t_7) = \theta \bar{Y}^2 C_y^2 \left(1 - \frac{(-1)^{1+1} \left| R_{yx} \right|_{yxz}}{\left| R_{2\times2} \right|} \rho_{yx} - \frac{(-1)^{2+1} \left| R_{yz} \right|_{yxz}}{\left| R_{2\times2} \right|} \rho_{yz} \right) + \lambda \bar{Y}^2 C_{y(2)}^2 \qquad (30)$$

We can rewrite (30) as,

$$MSE(t_7) = \frac{\theta \bar{Y}^2 C_y^2}{\left| R_{2\times2} \right|} \left(\left| R_{2\times2} \right| + (-1)^1 \left| R_{yx} \right|_{yxz} \rho_{yx} + (-1)^2 \left| R_{yz} \right|_{yxz} \rho_{yz} \right) + \lambda \bar{Y}^2 C_{y(2)}^2 \qquad (31)$$

We can also rewrite (31) as,

$$MSE(t_7) = \frac{\theta \bar{Y}^2 C_y^2}{\left| R_{2\times2} \right|} \left| R \right|_{yxz} + \lambda \bar{Y}^2 C_{y(2)}^2 \qquad (32)$$

Using Arora and Lal (1989) in (32) we get.

$$MSE(t_7) = \theta \bar{Y}^2 C_y^2 \left(1 - \rho_{y.xz}^2 \right) + \lambda \bar{Y}^2 C_{y(2)}^2 \qquad (33)$$

3.2 CASE 2: When there is incomplete information on both the study variable *y* and the auxiliary variable *x*.

Khare and Srivastava (1993, 1995) proposed ratio and product estimators under case 2. When the population mean $\bar{X}$ of the auxiliary variable '*x*' is unknown, the two phase ratio estimator is given by,

$$t_{R2} = \bar{y}^* \frac{\bar{x}'}{\bar{x}^*} \qquad (34)$$

The MSE of the estimator t_{R2};

$$MSE(t_{R2}) = \bar{Y}^2 \left(\begin{array}{c} \theta' \left\{ C_y^2 + \left(1 - 2k_{yx} \right) C_x^2 \right\} + \\ \delta \left\{ C_{y(2)}^2 + \left(1 - 2k_{yx(2)} \right) C_{x_2}^2 \right\} + \lambda' C_y^2 \end{array} \right) \qquad (35)$$

When the population mean $\bar{X}$ of the auxiliary variable '*x*' is unknown, the two phase product estimator is given by,

$$t_{P2} = \bar{y}^* \frac{\bar{x}^*}{\bar{x}'} \qquad (36)$$

The MSE of the estimator t_{P2};

$$MSE(t_{P2}) = \bar{Y}^2 \left(\begin{array}{c} \theta' \left\{ C_y^2 + \left(1 + 2k_{yx} \right) C_x^2 \right\} \\ + \delta \left\{ C_{y(2)}^2 + \left(1 + 2k_{yx(2)} \right) C_{x_2}^2 \right\} + \lambda' C_y^2 \end{array} \right) \qquad (37)$$

When the population mean $\bar{X}$ of the auxiliary variable '*x*' is unknown, the two phase regression estimator by Khare and Srivastava (1993, 1995) is given by,

$$t_{RE2} = \bar{y}^* + \beta \left(\bar{x}' - \bar{x}^* \right) \qquad (38)$$

The MSE of the estimator t_{RE2};

$$MSE(t_{RE2}) = \bar{Y}^2 \left(\begin{array}{c} \theta' C_y^2 + \lambda \left(1 - \rho^2 \right) C_y^2 \\ + \delta \left(C_{y(2)}^2 + k_{yx} \left(k_{yx} - 2k_{yx(2)} \right) C_{x(2)}^2 \right) \end{array} \right) \qquad (39)$$

When the population mean $\bar{X}$ of the auxiliary variable '*x*' is unknown, the two phase exponential ratio estimator by Singh *et al*. (2010) is given by,

$$t_{ER2} = \bar{y}^* \exp \left(\frac{\bar{x}' - \bar{x}^*}{\bar{x}' + \bar{x}^*} \right) \qquad (40)$$

The MSE of the estimator t_{ER2};

$$MSE(t_{ER2}) = \bar{Y}^2 \left(\begin{array}{c} \theta' \left\{ C_y^2 + \left(1 - 4k_{yx} \right) \frac{C_x^2}{4} \right\} \\ + \delta \left\{ C_{y(2)}^2 + \left(1 - 4k_{yx(2)} \right) \frac{C_{x_2}^2}{4} \right\} + \lambda' C_y^2 \end{array} \right) \qquad (41)$$

When the population mean $\bar{X}$ of the auxiliary variable '*x*' is unknown, the two phase exponential

product estimator by Singh *et al.* (2010) is given by,

$$t_{EP2} = \bar{y}^* \exp\left(\frac{\bar{x}^{-*} - \bar{x}'}{\bar{x}^{-*} + \bar{x}'}\right) \tag{42}$$

The MSE of the estimator t_{EP2};

$$MSE(t_{EP2}) = \bar{Y}^2 \left(\begin{array}{l} \theta'\left\{C_y^2 + \left(1 + 4k_{yx}\right)\dfrac{C_x^2}{4}\right\} \\ + \delta\left\{C_{y(2)}^2 + \left(1 + 4k_{yx(2)}\right)\dfrac{C_{x_2}^2}{4}\right\} + \lambda' C_y^2 \end{array} \right)$$

(43)

When the population mean $\bar{X}$ of the auxiliary variable 'x' is unknown, the two phase exponential ratio-product estimator by Kumar and Bhougal (2011) is given by,

$$t_{ERP2} = \bar{y}^* \exp\left(\alpha\left(\frac{\bar{x}' - \bar{x}^{-*}}{\bar{x}' + \bar{x}^{-*}}\right) + (1-\alpha)\left(\frac{\bar{x}^{-*} - \bar{x}'}{\bar{x}^{-*} + \bar{x}'}\right) \right)$$

(44)

The MSE of the estimator t_{ERP2};

$$MSE(t_{ERP2}) = \bar{Y}^2 \left(\begin{array}{l} \theta'\left\{C_y^2 + \dfrac{B_2}{A_2}\left(\dfrac{B_2}{A_2} - 2k_{yx}\right)C_x^2\right\} \\ + \delta\left\{C_{y(2)}^2 + \dfrac{B_2}{A_2}\left(\dfrac{B_2}{A_2} - 2k_{yx}\right)C_x^2\right\} + \lambda' C_y^2 \end{array} \right)$$

(45)

Where $\qquad A_2 = \theta C_x^2 + \delta C_{x(2)}^2$,

$$B_2 = \theta k_{yx} C_x^2 + \delta k_{yx_2} C_{x(2)}^2 \qquad \alpha = \frac{A_2 + 2B_2}{2A_2}$$

3.2.1 Modified Regression Estimator with Non-Response in both study variable and the auxiliary variable

Following Tum *et al.* (2014), we define the following estimator for $\bar{Y}$ in the presence of non-response in the study variable and auxiliary variables as:

$$t_8 = \left(\bar{y}^{-*} + \beta\left(\bar{X} - \bar{x}^{-*}\right)\right)\left(\alpha\exp\left(\frac{\bar{Z} - \bar{z}^{-*}}{\bar{Z} + \bar{z}^{-*}}\right) + (1-\alpha)\exp\left(\frac{\bar{z}^{-*} - \bar{Z}}{\bar{Z} + \bar{z}^{-*}}\right)\right)$$

(46)

Substituting (5) equation in (46) we get,

$$t_8 = \left(\bar{Y}\left(1 + e_y^{*}\right) - \beta\bar{e}_x^{-*}\right)\left(\begin{array}{l} \alpha\exp\left(\dfrac{\bar{Z} - \bar{Z}\left(1 + \bar{e}_z^{*}\right)}{\bar{Z} + \bar{Z}\left(1 + \bar{e}_z^{*}\right)}\right) \\ + (1-\alpha)\exp\left(\dfrac{\bar{Z}\left(1 + \bar{e}_z^{*}\right) - \bar{Z}}{\bar{Z} + \bar{Z}\left(1 + \bar{e}_z^{*}\right)}\right) \end{array} \right) \tag{47}$$

Opening brackets and ignoring the second and higher terms for each expansion of product and after simplification we can write t_8 as

$$t_8 = \bar{Y} + \bar{e}_y^{*} - \beta\bar{e}_x^{*} + \bar{Y}\left(\frac{1}{2} - \alpha\right)\bar{e}_z^{*} \tag{48}$$

The mean squared error of t_8 is given by,

$$MSE(t_8) = E\left(t_8 - \bar{Y}\right)^2 = E\left(\bar{e}_y^{*} - \beta\bar{e}_x^{*} + \bar{Y}\left(\frac{1}{2} - \alpha\right)\bar{e}_z^{*}\right)^2 \tag{49}$$

Squaring the right sides of (49), we get,

$$MSE(t_8) = E\left(\begin{array}{l} \bar{e}_y^{*2} + \beta^2\bar{e}_x^{*2} + \bar{Y}^2\left(\dfrac{1}{2} - \alpha\right)^2\bar{e}_z^{*2} + 2\bar{Y}\left(\dfrac{1}{2} - \alpha\right)\bar{e}_z^{*}\bar{e}_y^{*} \\ - 2\bar{e}_y^{*}\beta\bar{e}_x^{*} - 2\bar{Y}\left(\dfrac{1}{2} - \alpha\right)\bar{e}_z^{*}\beta\bar{e}_x^{*} \end{array} \right) \tag{50}$$

Taking expectation and using (5) in (50), we get,

$$\begin{aligned}
MSE(t_8) = {} & \theta\bar{Y}^2 C_y^2 + \lambda\bar{Y}^2 C_{y(2)}^2 + \theta\bar{X}^2\beta^2 C_x^2 + \lambda\bar{X}^2\beta^2 C_{x(2)}^2 \\
& + \bar{Y}^2\left(\frac{1}{2} - \alpha\right)^2\left(\theta\bar{Z}^2 C_z^2 + \lambda\bar{Z}^2 C_{z(2)}^2\right) \\
& + 2\bar{Y}\left(\frac{1}{2} - \alpha\right)\left(\bar{Y}\bar{X}\left(\theta\rho_{yz}C_z C_y + \lambda\rho_{yz(2)}C_{z(2)}C_{y(2)}\right)\right) \\
& - 2\beta\,\bar{X}\bar{Y}\left(\theta\rho_{yx}C_x C_y + \lambda\rho_{yx(2)}C_{x(2)}C_{y(2)}\right) \\
& - 2\bar{Y}\beta\left(\frac{1}{2} - \alpha\right)\left(\bar{Y}\bar{X}\left(\theta\rho_{xz}C_z C_x + \lambda\rho_{xz(2)}C_{x(2)}C_{z(2)}\right)\right)
\end{aligned} \tag{51}$$

Expanding and differentiating (51) with respect to α, α^*, β and β^* and equating to zero gives

$$\beta = \frac{\bar{Y}C_y}{\bar{X}C_x\left|R_{2\times2}\right|}(-1)^{1+1}\left|R_{yx}\right|_{yxz}$$

$$\beta^* = \frac{\bar{Y}C_{y(2)}}{\bar{X}C_{x(2)}\left|R^*_{2\times2}\right|}(-1)^{1+1}(-1)^{1+1}\left|R_{yx}^*\right|_{yxz}$$

$$\alpha = \frac{1}{2} + \frac{C_y}{\bar{Z}C_z\left|R_{2\times2}\right|}(-1)^{2+1}\left|R_{yz}\right|_{yxz}$$

$$\alpha^* = \frac{1}{2} + \frac{C_{y(2)}}{\bar{Z}C_{z(2)z}\left|R^*_{2\times2}\right|}(-1)^{2+1}\left|R_{yz}^*\right|_{yxz} \tag{52}$$

Using normal equations that are used to find the optimum values of α *and* β, (52) can be written in simplified form as,

$$MSE(t_8) = E\left(\bar{e}_y^{*}\left(\bar{e}_y^{*} - \beta\bar{e}_x^{*} + \bar{Y}\left(\frac{1}{2} - \alpha\right)\bar{e}_z^{*}\right)\right) \tag{53}$$

Expanding and Taking expectation and using (5) in (53) we get

$$MSE(t_8) = \theta\bar{Y}^2 C_y^2 + \lambda\bar{Y}^2 C_{y(2)}^2 - \beta\bar{X}\bar{Y}\theta\rho_{yx}C_x C_y$$
$$- \lambda\beta^*\bar{X}\bar{Y}\rho_{yx(2)}C_{x(2)}C_{y(2)} + \bar{Y}\left(\frac{1}{2}-\alpha\right)\theta\bar{Y}\bar{X}\rho_{yz}C_z C_y \quad (54)$$
$$+ \bar{Y}\left(\frac{1}{2}-\alpha^*\right)\lambda\bar{Y}\bar{X}\rho_{yz(2)}C_{z(2)}C_{y(2)}$$

Substituting the optimum value in (52), we get

$$MSE(t_8) = \theta\bar{Y}^2 C_y^2 - \left(\frac{\bar{Y}C_y}{\bar{X}C_x|R_{2\times2}|}(-1)^{1+1}\left|R_{yx}\right|_{yxz}\right)\theta\bar{X}\bar{Y}C_y C_x\rho_{yx}$$
$$+ \bar{Y}\left(\frac{1}{2}-\left(\frac{1}{2}+\frac{C_y}{\bar{Z}C_z|R_{2\times2}|}(-1)^{2+1}\left|R_{yz}\right|_{yxz}\right)\right)\theta\bar{Y}\bar{Z}C_y C_z\rho_{yz}$$
$$+ \lambda\bar{Y}^2 C_{y(2)}^2 - \left(\frac{\bar{Y}C_{y(2)}}{\bar{X}C_{x(2)}|R^*_{2\times2}|}(-1)^{1+1}(-1)^{1+1}\left|R_{yx}^*\right|_{yxz}\right)\lambda\bar{X}\bar{Y}C_{y(2)}C_{x(2)}\rho_{yx(2)} \quad (55)$$
$$+ \bar{Y}\left(\frac{1}{2}-\left(\frac{1}{2}+\frac{C_{y(2)}}{\bar{Z}C_{z(2)z}|R^*_{2\times2}|}(-1)^{2+1}\left|R_{yz}^*\right|_{yxz}\right)\right)\lambda\bar{Y}\bar{Z}C_{y(2)}C_{z(2)}\rho_{yz(2)}$$

Simplifying (55) we get

$$MSE(t_8) = \theta\bar{Y}^2 C_y^2\left(1 - \frac{(-1)^{1+1}\left|R_{yx}\right|_{yxz}}{|R_{2\times2}|}\rho_{yx} - \frac{(-1)^{2+1}\left|R_{yz}\right|_{yxz}}{|R_{2\times2}|}\rho_{yz}\right)$$
$$+ \lambda\bar{Y}^2 C_{y(2)}^2\left(1 - \frac{(-1)^{1+1}\left|R_{yx}^*\right|_{yxz}}{|R^*_{2\times2}|}\rho_{yx(2)} - \frac{(-1)^{2+1}\left|R_{yz}^*\right|_{yxz}}{|R^*_{2\times2}|}\rho_{yz(2)}\right) \quad (56)$$

Or

$$MSE(t_8) = \frac{\theta\bar{Y}^2 C_y^2}{|R_{2\times2}|}|R|_{yxz} + \frac{\lambda\bar{Y}^2 C_{y(2)}^2}{|R^*_{2\times2}|}|R^*|_{yxz} \quad (57)$$

Using Arora and Lal (1989) in (57) we get,

$$MSE(t_8) = \theta\bar{Y}^2 C_y^2\left(1-\rho_{y.xz}^2\right) + \lambda\bar{Y}^2 C_{y(2)}^2\left(1-\rho_{y.xz}^{*2}\right) \quad (58)$$

4. Simulation, Result and Discussion

In this section, we carried out some data simulation experiments using R statistical program to compare the performance of modified regression estimators in two phase sampling with non-response in either the study variable only or both the study variable and the auxiliary variable with already existing estimators in two phase sampling for finite population. In the simulated population, the study variable and auxiliary variables are normally distributed.

i) Study variable N = 500, $n_1 = 75$, $n = 50$ $\mu = 75$ $\sigma = 10$

ii) For ratio estimator the auxiliary variable is strongly positively correlated with the study variable and the line passes through the origin.

N=500, $n_1 = 75$, $n = 50$ $\mu = 55$ $\sigma = 9.4$
$\rho_{yx} = 0.7018$

iii) For regression estimator the auxiliary variable was strongly positively correlated with the study variable and the regression line does not pass through the origin.

N =500, $n_1 = 75$ and $n = 50$. $\mu = 22$
$\sigma = 3.29$ $\rho_{yx} = 0.8114$

iv) For product estimator the auxiliary variable was strongly negatively correlated with the study variable.

N=500, $n_1 = 75$ and $n = 50$, $\mu = 34$ $\sigma = 5.10$
$\rho_{yx} = -0.7477$

The percent relative efficiencies (PREs) of different proposed estimators were computed with respect to a usual unbiased estimator $\bar{y}^*$ for different values of k.

$$PRE(\hat{t}) = \frac{Var(\bar{y}^*)}{MSE(\hat{t})}*100 \quad (59)$$

Table 1: Percent relative efficiency of different $\bar{Y}$ estimators with respect to $\bar{y}^*$ when the non-response is on the study variable only in two phase sampling.

Estimators	$1/k$			
	$1/5$	$1/4$	$1/3$	$1/2$
$\bar{y}^*$	100	100	100	100
t_{R1}	110	114	116	117
t_{P1}	101	101	102	105
t_{RE1}	129	140	143	157
t_{ER1}	126	135	138	150
t_{EP1}	106	107	110	111

t_{ERP1}	126	136	139	151
t_7 (**Proposed**)	**139**	**155**	**161**	**183**

The Percent relative efficiencies of all the estimators in the Table I increases as k increases. Our proposed estimator (t_7) is the most efficient among all the estimators since it has the highest Percent relative efficiency when the non-response is on the study variable only in two phase sampling **Table 2**: Percent relative efficiency of different $\bar{Y}$ estimators with respect to $\bar{y}^*$ when the non-response is on the study variable and also in auxiliary variable in two phase sampling.

Estimators	$1/k$			
	$1/5$	$1/4$	$1/3$	$1/2$
$\bar{y}^*$	100	100	100	100
t_{R2}	104	105	133	119
t_{P2}	102	105	106	106
t_{RE2}	202	205	213	219
t_{ER2}	176	180	191	181
t_{EP2}	154	138	133	124
t_{ERP2}	197	174	182	180
t_8 (**Proposed**)	**269**	**268**	**273**	**261**

Our proposed estimator (t_8) is the most efficient among all the estimators since it has the highest Percent relative efficiency when the non-response

are on the study variable and the auxiliary variable in two phase sampling.

5.0 Conclusion

Based on these study results, the proposed estimators t_7 and t_8 are recommended for use in practice when non-response is on the study variable only and on both the study variable and the auxiliary variable in two phase sampling respectively.

References

[1] Hansen, M. H. and Hurwitz, W. N., 1946. The problem of non-response in sample surveys. *J. Amer. Statist. Assoc.*, 41, 517-529.

[2] Neyman, J. (1938). Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.*, 33, 101-116.

[3] Khare, B. B. and Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response, *The National Academy of Sciences, Letters, India*, 16, 111–114.

[4] Khare, B. B. and Srivastava, S. (1995). Study of conventional and alternative two phase sampling ratio, product and regression estimators in presence of non-response, *Proceedings of the National Academy of Sciences,*

[5] Singh, H. P. and Kumar, S. (2010). Estimation of mean in presence of non-response using two phase sampling scheme, *Statistical Papers*, 50, 559–582.

[6] Kumar1, S and Bhougal*, S (2011) Estimation of the Population Mean in Presence of Non-Response, Vol. 18, No. 4, 537–548, Communications of the Korean Statistical Society*

[7] Tum, E.C., Kung'u, J. and Odongo, L. (2014) A New Regression Type Estimator with Two Auxiliary Variables for Single-Phase Sampling. *Open Journal of Statistics*, **4**, 789-796.

[8] Arora, S. and Lal, B. (1989) *New Mathematical Statistics*. Satya Prakashan, New Delhi.