Efficient Method of Estimating the Finite Population Mean Based on Two Auxiliary Variables in the Presence of Non-Response Under Stratified Sampling

Housila P. Singh[1] and Pragati Nigam[2,*]

[1]*School of Studies in Statistics, Vikram University, Ujjain, Madhya Pradesh, India*
[2]*Mandsaur University, Mandsaur, Madhya Pradesh, India*
E-mail: pragatinigam1@gmail.com
[*]*Corresponding Author*

Received 16 October 2020; Accepted 06 March 2021;
Publication 18 June 2021

Abstract

This article addresses the problem of estimating the population mean using information on two auxiliary variables in presence of non-response on study variable only under stratified random sampling. A class of estimators has been defined. We have derived the bias and mean squared error up to first order of approximation. Optimum conditions are obtained in which the suggested class of estimators has minimum mean squared error. In addition to Chaudhury et al. (2009) estimator, many estimators can be identified as a member of the suggested class of estimators. It has been shown that the suggested class of estimators is better than the Chaudhury et al. (2009) estimator and other estimators. Results of the present study are supported through numerical illustration.

Keywords: Study variable, auxiliary variable, finite population, non-response, bias, mean squared error.

Journal of Reliability and Statistical Studies, Vol. 14, Issue 1 (2021), 223–242.
doi: 10.13052/jrss0974-8024.14111

1 Introduction

In many surveys, auxiliary information is usually used to improve the precision or accuracy of the estimator of the population mean under the supposition that all the observations in the sample are available. However in many surveys covering human population, information is in most cases not obtained from all the units in the surveys even after call-backs. For example, the selected families may not be at home at the first attempt and some may refuse to cooperate with the interviewer even if contacted. This is true in mail surveys in which questionnaire are mailed to the sampled respondents who are requested to send back their returns by some dead line. As many respondents do not reply, available sample of returns is incomplete. The resulting incompleteness, called non response [Sukhatme et al. (1984, pp. 484–485)]. Incompleteness or non-response in the form of absence, censoring, or grouping is a troubling issue of several data sets. Statisticians have identified for some time that failure to account for the stochastic nature of incompleteness or non-response can spoil the nature of data. An estimate derived from incomplete data may be misleading especially when the respondents differ from the non-respondents because the estimate can be biased. Hansen and Hurwitz (1946) suggested a method for adjusting for non-response to address the bias problem. Their idea is to select a sub-sample from the non-respondents to obtain an estimate for the sub-population represented by the non-respondents [Okafor and Lee (2000, p. 183)].

When the population mean of the auxiliary variable is known; Cochran (1977), using Hansen and Hurwitz (1946) technique, envisaged the ratio and regression estimators of the population mean of the study variable in which information on the auxiliary variable is obtained from all the sample units, while some sample units failed to supply information on the study. Later various authors including Rao (1986), Khare and Srivastava (1993, 1997), Tripathi and Khare (1997), Okafor and Lee (2000), Tabasum and Khan (2004, 2006), Singh and Kumar (2008, 2009), Singh et al. (2010), Khare et al. (2013) have paid their attention towards the estimation of the population mean of the study variable using information on auxiliary variable in presence of non-response. Singh and Khalid (2015) suggested exponential chain dual to ratio and regression type estimators of the population mean in two-phase sampling. Further Chaudhary et al. (2011), Haq and Shabbir (2013), Sanaullah et al. (2015) and Saleem et al. (2018) envisaged some improved estimators of the population mean of the study variable using auxiliary information for stratified random sampling under non-response.

Let a finite population $U = (U_1, U_2, \ldots, U_N)$ of size N be stratified into L strata (homogeneous). Let N_h be the size of the h^{th} stratum

$(h = 1, 2, \ldots, L)$: $\sum_{h=1}^{L} N_h = N$. Let (y_{hi}, x_{hi}, z_{hi}) be the values on the i^{th} unit of the h^{th} stratum of the study variable *y* and auxiliary variables (*x*, *z*) respectively. Corresponding to the population means $(\bar{Y}_h, \bar{X}_h, \bar{Z}_h)$, let $(\bar{y}_h, \bar{x}_h, \bar{z}_h)$ be the sample means of the h^{th} stratum respectively. In practice it is usually not possible to gather information on all the variables/units selected in the sample $n_h(\sum_{h=1}^{L} n_h = n)$. In this paper we have studied the situation when non-response occurs only on the study variable *y* whereas the two auxiliary variables (*x*, *z*) are observed with complete response.

Let $n_{h(1)}$ units from a sample of size n_h respond and $n_{h(2)}$ units do not. Employing Hansen and Hurwitz (1946) method of sub-sampling the non-respondents, a sub-sample of size $r_h(r_h = \frac{n_{h(2)}}{f_h}, f_h > 1)$ from $n_{h(2)}$ non-respondent group is selected at random and $\frac{1}{f_h}$ denotes the sampling fraction among the non-respondent group in the h^{th} stratum. In practice, r_h is generally not integer and has to be rounded. In accordance with most of the current literature on this research topic, we suppose that the followed-up $r_h(\subset n_{h(2)})$ units respond on the second call. Further, let *d* denotes a dummy variable taking value d_{hi} on the i^{th} population unit of stratum *h* and has h^{th} stratum population mean $\bar{D}_h$. Hereafter, *d* may stand for if, *x* or for a second auxiliary variable z (i.e. d_h may stand for y_h, x_h and z_h in stratified sampling). Let

$$\bar{d}_{n_{h(1)}} = \frac{1}{n_{h(1)}} \sum_{i=1}^{n_{h(1)}} d_{hi(1)}, \quad \bar{d}_{rh(2)} = \frac{1}{r_h} \sum_{i=1}^{r_h} d_{hi(2)}$$

and

$$\bar{d}_h^* = \frac{n_{h(1)}}{n_h} \bar{d}_{nh(2)} + \frac{n_{h(2)}}{n_h} \bar{d}_{rh(2)}, \tag{1}$$

where $\bar{d}_{nh(2)}$ is the mean of $n_{h(1)}$ respondents on first call and $d_{rh(2)}$ is the mean of r_h units respond on the second call and $\bar{d}_h^*$ denotes the unbiased Hansen and Hurwitz (1946) estimator of $\bar{D}_h$ for stratum *h*.

Thus we define an unbiased estimator of the population mean $\bar{D} = \sum_{h=1}^{L} W_h \bar{D}_h$ as

$$\bar{d}_{st}^* = \sum_{h=1}^{L} W_h \bar{d}_h^* \tag{2}$$

and the variance/*MSE* of $\bar{d}_{st}^*$ is given by

$$V(\bar{d}_{st}^*) = \sum_{h=1}^{L} \delta_h W_h^2 S_{dh}^2 + \sum_{h=1}^{L} \delta_h^* W_h^2 S_{dh(2)}^2, \tag{3}$$

where $S_{dh}^2 = \frac{\sum_{i=1}^{N_h}(d_{hi}-\bar{D}_h)^2}{(N_h-1)}$ and $S_{dh(2)}^2 = \frac{\sum_{i=1}^{N_{h(2)}}(d_{hi}-\bar{D}_{h(2)})^2}{(N_{h(2)}-1)}$ are respectively mean square of entire group and non-response group of variable *d* in the population for the h^{th} stratum, $W_h = \frac{N_h}{N}$, $\delta_h = (\frac{1}{n_h} - \frac{1}{N_h})$, $\delta_h^* = \frac{(f_h-1)W_{h(2)}}{n_h}$, $W_{h(2)} = \frac{N_{h(2)}}{N_h}$, $f_h = \frac{n_{h(2)}}{r_h}$ and $N_{h(2)}$ being the size of the non-response group of the population in the h^{th} stratum.

For obtaining the bias and mean squared errors (*MSE*s) of the proposed estimators we below give the values of the required expectations:

We write

$$\bar{y}_{st}^* = \bar{Y}(1+e_0^*),\ \bar{x}_{st} = \bar{X}(1+e_1),\ \bar{z}_{st} = \bar{Z}(1+e_2)$$

such that

$$E(e_0^*) = E(e_1) = E(e_2) = 0$$

and

$$E(e_0^{*2}) = \frac{1}{\bar{Y}^2}\sum_{h=1}^{L} W_h^2(\delta_h S_{yh}^2 + \delta_h^* S_{yh(2)}^2) = V_{020}^*,$$

$$E(e_1^2) = \frac{1}{\bar{X}^2}\sum_{h=1}^{L} W_h^2 \delta_h S_{xh}^2 = V_{200},$$

$$E(e_2^2) = \frac{1}{\bar{Z}^2}\sum_{h=1}^{L} W_h^2 \delta_h S_{zh}^2 = V_{002},$$

$$E(e_0^* e_1) = \frac{1}{\bar{X}\bar{Y}}\sum_{h=1}^{L} W_h^2 \delta_h S_{yxh} = V_{110},$$

$$E(e_0^* e_2) = \frac{1}{\bar{Y}\bar{Z}}\sum_{h=1}^{L} W_h^2 \delta_h S_{yzh} = V_{011},$$

$$E(e_1 e_2) = \frac{1}{\bar{X}\bar{Z}}\sum_{h=1}^{L} W_h^2 \delta_h S_{xzh} = V_{101},$$

where

$$S_{yxh} = \frac{1}{(N_h-1)}\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)(y_{hi} - \bar{Y}_h) = \rho_{yxh} S_{xh} S_{yh},$$

$$S_{yzh} = \frac{1}{(N_h-1)}\sum_{i=1}^{N_h}(y_{hi}-\bar{Y}_h)(z_{hi}-\bar{Z}_h) = \rho_{yzh}S_{yh}S_{zh},$$

$$S_{xzh} = \frac{1}{(N_h-1)}\sum_{i=1}^{N_h}(x_{hi}-\bar{X}_h)(z_{hi}-\bar{Z}_h) = \rho_{xzh}S_{xh}S_{zh},$$

$$\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h}y_{hi},\ \bar{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h}x_{hi},\ \bar{Z}_h = \frac{1}{N_h}\sum_{i=1}^{N_h}z_{hi},$$

$$\bar{Y}_{h(2)} = \frac{1}{N_{h(2)}}\sum_{i=1}^{N_{h(2)}}y_{hi},\ \bar{X}_{h(2)} = \frac{1}{N_{h(2)}}\sum_{i=1}^{N_{h(2)}}x_{hi},\ \bar{Z}_{h(2)} = \frac{1}{N_{h(2)}}\sum_{i=1}^{N_{h(2)}}z_{hi},$$

$(\rho_{yxh}, \rho_{xzh}, \rho_{yzh})$ are the correlation coefficients between the subscripted variables of entire population.

2 Suggested Class of Estimators for Estimating Population Mean in Stratified Sampling in Presence of Non-Response

When non-response occurs only on the study variable *y* (i.e. incomplete information is available on the study variable *y* in the h^{th} stratum while complete information on the sample of size n_h is available for the auxiliary variables (*x*,*z*)), we define the following class of estimators for population mean $\bar{Y}$ as

$$t_{(M_1,M_2)} = \bar{y}^*_{st}\left[\begin{array}{l} M_1\left\{\dfrac{a_x\bar{X}+b_x}{\alpha_x(a_x\bar{x}_{st}+b_x)+(1-\alpha_x)(a_x\bar{X}+b_x)}\right\}^{g_x} \\ \left\{\dfrac{a_z\bar{Z}+b_z}{\alpha_z(a_z\bar{z}_{st}+b_z)+(1-\alpha_z)(a_z\bar{Z}+b_z)}\right\}^{g_z} \\ +M_2\exp\left\{\dfrac{-\alpha_x h_x a_x(\bar{x}_{st}-\bar{X})}{(2-\alpha_x)(a_x\bar{X}+b_x)+\alpha_x(a_x\bar{x}_{st}+b_x)}\right\} \\ \exp\left\{\dfrac{-\alpha_z h_z a_z(\bar{z}_{st}-\bar{Z})}{(2-\alpha_z)(a_z\bar{Z}+b_z)+\alpha_z(a_z\bar{z}_{st}+b_z)}\right\}\end{array}\right], \quad (4)$$

where $(g_x, g_z, h_x, h_z, \alpha_x, \alpha_z, a_x, a_z, b_x, b_z)$ are suitably chosen scalars and (M_1, M_2) are constants to be determined such that *MSE* of $t_{(M_1,M_2)}$ is minimum. For $(M_1, g_z, M_2) = (1, 0, 0)$ the class of estimators $t_{(M_1,M_2)}$ reduces to the family of estimators due to Chaudhary et al. (2009).

Using the standard procedure we obtained the bias and *MSE* of $t_{(M_1,M_2)}$ to the first degree of approximation, respectively given by

$$\begin{aligned} B(t_{M_1,M_2}) &= \bar{Y}\left[\begin{array}{l} M_1\left\{\begin{array}{l} 1-(g_x\alpha_x\upsilon_x)V_{110}-(g_z\alpha_z\upsilon_z)V_{011} \\ \quad +(g_x\alpha_x\upsilon_x)(g_z\alpha_z\upsilon_z)V_{101} \\ \quad +\dfrac{g_x(g_x+1)}{2}(\alpha_x\upsilon_x)^2V_{200} \\ \quad +\dfrac{g_z(g_z+1)}{2}(\alpha_z\upsilon_z)^2V_{002} \end{array}\right\} \\ +M_2\left\{\begin{array}{l} 1-\dfrac{(h_x\alpha_x\upsilon_x)}{2}V_{110} \\ \quad -\dfrac{(h_z\alpha_z\upsilon_z)}{2}V_{011} \\ \quad +\dfrac{(h_x\alpha_x\upsilon_x)(h_z\alpha_z\upsilon_z)}{4}V_{101} \\ \quad +\dfrac{h_x(h_x+2)(\alpha_x\upsilon_x)^2}{8}V_{200} \\ \quad +\dfrac{h_z(h_z+2)(\alpha_z\upsilon_z)^2}{8}V_{002} \end{array}\right\}-1 \end{array}\right] \\ &= \bar{Y}[M_1E_4+M_2E_5-1], \end{aligned} \tag{5}$$

$$\begin{aligned} MSE(t_{(M_1,M_2)}) &= \bar{Y}^2[1+M_1^2E_1 \\ &\quad +M_2^2E_2+2M_1M_2E_3-2M_1E_4-2M_2E_5], \end{aligned} \tag{6}$$

where

$$E_1 = \begin{bmatrix} 1+V_{020}^*+(\alpha_x\upsilon_x)^2g_x(2g_x+1)V_{200}+(\alpha_z\upsilon_z)^2g_z(2g_z+1)V_{002} \\ -4(g_x\alpha_x\upsilon_x)V_{110} \\ -4(g_z\alpha_z\upsilon_z)V_{011}+4(g_x\alpha_x\upsilon_x)(g_z\alpha_z\upsilon_z)V_{101} \end{bmatrix},$$

$$E_2 = \begin{bmatrix} 1+V_{020}^*+\dfrac{(\alpha_x\upsilon_x)^2h_x(h_x+1)}{2}V_{200}+\dfrac{(\alpha_z\upsilon_z)^2h_z(h_z+1)}{2}V_{002} \\ -2(h_x\alpha_x\upsilon_x)V_{110}-2(h_z\alpha_z\upsilon_z)V_{011} \\ +(h_x\alpha_x\upsilon_x)(h_z\alpha_z\upsilon_z)V_{101} \end{bmatrix},$$

$$E_3 = \begin{bmatrix} 1 + V_{020}^* + \frac{(2g_x + h_x)(2g_x + h_x + 2)(\alpha_x \upsilon_x)^2}{8} V_{200} \\ + \frac{(2g_z + h_z)(2g_z + h_z + 2)(\alpha_z \upsilon_z)^2}{8} V_{002} \\ - \frac{(2g_x + h_x)(\alpha_x \upsilon_x)}{2} V_{110} - \frac{(2g_z + h_z)(\alpha_z \upsilon_z)}{2} V_{011} \\ + \frac{(2g_x + h_x)(2g_z + h_z)(\alpha_x \upsilon_x)(\alpha_z \upsilon_z)}{4} V_{101} \end{bmatrix},$$

$$E_4 = \begin{bmatrix} 1 - (g_x \alpha_x \upsilon_x) V_{110} - (g_z \alpha_z \upsilon_z) V_{011} + (g_x \alpha_x \upsilon_x)(g_z \alpha_z \upsilon_z) V_{101} \\ + \frac{g_x(g_x + 1)}{2} (\alpha_x \upsilon_x)^2 V_{200} + \frac{g_z(g_z + 1)}{2} (\alpha_z \upsilon_z)^2 V_{002} \end{bmatrix},$$

$$E_5 = \begin{bmatrix} 1 - \frac{(h_x \alpha_x \upsilon_x)}{2} V_{110} - \frac{(h_z \alpha_z \upsilon_z)}{2} V_{011} + \frac{(h_x \alpha_x \upsilon_x)(h_z \alpha_z \upsilon_z)}{4} V_{101} \\ + \frac{h_x(h_x + 2)(\alpha_x \upsilon_x)^2}{8} V_{200} + \frac{h_z(h_z + 1)(\alpha_z \upsilon_z)^2}{8} V_{002} \end{bmatrix}.$$

$$\upsilon_x = \frac{a_x \bar{X}}{a_x \bar{X} + b_x} \quad \text{and} \quad \upsilon_z = \frac{a_z \bar{Z}}{a_z \bar{Z} + b_z}.$$

The $MSE(t_{(M_1, M_2)})$ at (6) is minimized for

$$\left. \begin{aligned} M_1 &= \frac{(E_2 E_4 - E_3 E_5)}{(E_1 E_2 - E_3^2)} = M_{10} \ (say) \\ M_2 &= \frac{(E_1 E_5 - E_3 E_4)}{(E_1 E_2 - E_3^2)} = M_{20} \ (say) \end{aligned} \right\}. \tag{7}$$

Substitution of (7) in (6) yields the minimum *MSE* of $t_{(M_1, M_2)}$ as

$$MSE_{\min}(t_{(M_1, M_2)}) = \bar{Y}^2 \left[1 - \frac{(E_2 E_4^2 - 2E_3 E_4 E_5 + E_1 E_5^2)}{(E_1 E_2 - E_3^2)} \right]. \tag{8}$$

Thus we arrived at the following theorem.

Theorem 2.1. The *MSE* of the suggested class of estimators $t_{(M_1, M_2)}$ is greater than equal to the minimum *MSE* of $t_{(M_1, M_2)}$ i.e.

$$MSE(t_{(M_1, M_2)}) \geq \bar{Y}^2 \left[1 - \frac{(E_2 E_4^2 - 2E_3 E_4 E_5 + E_1 E_5^2)}{(E_1 E_2 - E_3^2)} \right]$$

with equality holding if

$$M_1 = M_{10} \quad \text{and} \quad M_2 = M_{20}.$$

A large number of estimators can be generated from the proposed class of estimators $t_{(M_1,M_2)}$ for suitable values of scalars involved in it. Some members of the proposed class of estimators $t_{(M_1,M_2)}$ are discussed below.

2.1 Some Members of the Proposed Class of Estimators

Case I. Putting $M_1 = M$ and $M_2 = (1 - M)$ in (4) we get a class of estimators for population mean $\bar{Y}$ as

$$t_M = \bar{y}^*_{st} \begin{bmatrix} M \left\{ \dfrac{a_x\bar{X} + b_x}{\alpha_x(a_x\bar{x}_{st} + b_x) + (1-\alpha_x)(a_x\bar{X} + b_x)} \right\}^{g_x} \\ \left\{ \dfrac{a_z\bar{Z} + b_z}{\alpha_z(a_z\bar{z}_{st} + b_z) + (1-\alpha_z)(a_z\bar{Z} + b_z)} \right\}^{g_z} \\ +(1-M)\exp\left\{ \dfrac{-\alpha_x h_x a_x(\bar{x}_{st} - \bar{X})}{(2-\alpha_x)(a_x\bar{X} + b_x) + \alpha_x(a_x\bar{x}_{st} + b_x)} \right\} \\ \exp\left\{ \dfrac{-\alpha_z h_z a_z(\bar{z}_{st} - \bar{Z})}{(2-\alpha_z)(a_z\bar{Z} + b_z) + \alpha_z(a_z\bar{z}_{st} + b_z)} \right\} \end{bmatrix}. \tag{9}$$

Inserting $M_1 = M$ and $M_2 = (1 - M)$ in (5) and (6) we get the bias and *MSE* of t_M to the first degree of approximation as

$$B(t_M) = \bar{Y} \begin{bmatrix} M \begin{Bmatrix} (H_x - G_x)V_{110} + (H_z - G_z)V_{011} \\ +(G_xG_z - H_xH_z)V_{101} \\ +\left(\dfrac{g_x(g_x+1)}{2} - \dfrac{h_z(h_z+2)}{8}\right)(\alpha_x\upsilon_x)^2V_{200} \\ +\left(\dfrac{g_z(g_z+1)}{2} - \dfrac{h_z(h_z+2)}{8}\right)(\alpha_z\upsilon_z)^2V_{002} \end{Bmatrix} \\ \dfrac{h_x(h_x+2)(\alpha_x\upsilon_x)^2}{8}V_{200} + \dfrac{h_z(h_z+2)(\alpha_z\upsilon_z)^2}{8}V_{002} \\ -H_xV_{110} - H_zV_{011} + H_xH_zV_{101} \end{bmatrix}, \tag{10}$$

$$MSE(t_M) = \bar{Y}^2[1 + E_2 - 2E_5 + M^2(E_1 + E_2 - 2E_3) - 2M(E_2 - E_3 + E_4 - E_5)], \tag{11}$$

where

$$H_x = \frac{(h_x \alpha_x \upsilon_x)}{2},\ H_z = \frac{(h_z \alpha_z \upsilon_z)}{2},$$
$$G_x = (g_x \alpha_x \upsilon_x) \quad \text{and} \quad G_z = (g_z \alpha_z \upsilon_z).$$

The $MSE(t_M)$ at (11) is minimum when

$$M = \frac{(E_2 - E_3 + E_4 - E_5)}{(E_1 + E_2 - 2E_3)} = M_0\ (say). \tag{12}$$

Thus the resulting minimum *MSE* of t_M is given by

$$MSE_{\min}(t_M) = \bar{Y}^2\left[1 + E_2 - 2E_5 - \frac{(E_2 - E_3 + E_4 - E_5)^2}{(E_1 + E_2 - 2E_3)}\right]. \tag{13}$$

Now we arrived at the following theorem.

Theorem 2.2. To the first degree if approximation,

$$MSE(t_M) \geq \bar{Y}^2\left[1 + E_2 - 2E_5 - \frac{(E_2 - E_3 + E_4 - E_5)^2}{(E_1 + E_2 - 2E_3)}\right]$$

with equality holding if

$$M = \frac{(E_2 - E_3 + E_4 - E_5)}{(E_1 + E_2 - 2E_3)}.$$

Case II. If we set $M_2 = 0$ in (4) we get the class of estimators for $\bar{Y}$ as

$$t_{M_1} = M_1 \bar{y}^*_{st}\left\{\frac{a_x\bar{X} + b_x}{\alpha_x(a_x\bar{x}_{st} + b_x) + (1 - \alpha_x)(a_x\bar{X} + b_x)}\right\}^{g_x} \left\{\frac{a_z\bar{Z} + b_z}{\alpha_z(a_z\bar{z}_{st} + b_z) + (1 - \alpha_z)(a_z\bar{Z} + b_z)}\right\}^{g_z}. \tag{14}$$

Putting $M_2 = 0$ in (5) and (6) we get the bias and *MSE* of t_{M_1} to the first degree of approximation as

$$B(t_{M_1}) = \bar{Y}(M_1E_4 - 1), \tag{15}$$

$$MSE(t_{M_1}) = \bar{Y}^2(1 + M_1^2E_1 - 2M_1E_4). \tag{16}$$

The $MSE(t_{M_1})$ at (16) is minimum when

$$M_1 = \frac{E_4}{E_1} = M_{1(opt)}. \tag{17}$$

Thus the resulting minimum *MSE* of t_{M_1} is given by

$$MSE_{\min}(t_{M_1}) = \bar{Y}^2 \left(1 - \frac{E_4^2}{E_1}\right). \tag{18}$$

Now we arrived at the following theorem.

Theorem 2.3. The *MSE* of t_{M_1} is greater than equal to the minimum *MSE* of t_{M_1} i.e.

$$MSE(t_{M_1}) \geq \bar{Y}^2 \left(1 - \frac{E_4^2}{E_1}\right)$$

with equality holding if

$$M_1 = \frac{E_4}{E_1}.$$

Case III. If we put $g_z = 0$ in (14) we get an improved version of Chaudhary et al. (2009) class of estimators as

$$t_{M_{1(1)}} = M_1 \bar{y}_{st}^* \left\{ \frac{a_x \bar{X} + b_x}{\alpha_x (a_x \bar{x}_{st} + b_x) + (1 - \alpha_x)(a_x \bar{X} + b_x)} \right\}^{g_x}. \tag{19}$$

Inserting $g_z = 0$ in (15) and (16) we get the bias and *MSE* of $t_{M_{1(1)}}$ to the first degree of approximation as

$$B(t_{M_{1(1)}}) = \bar{Y}(M_1 E_4^* - 1), \tag{20}$$

$$MSE(t_{M_{1(1)}}) = \bar{Y}^2 (1 + M_1^2 E_1^* - 2 M_1 E_4^*), \tag{21}$$

where

$$E_1^* = [1 + V_{020}^* + (\alpha_x \upsilon_x)^2 g_x (g_x + 1) V_{200} - 4(g_x \alpha_x \upsilon_x) V_{110}],$$

$$E_4^* = [1 - (g_x \alpha_x \upsilon_x) V_{110} + \frac{g_x (g_x + 1)}{2} (\alpha_x \upsilon_x)^2 V_{200}].$$

The $MSE(t_{M_{1(1)}})$ at (21) is minimum when

$$M_1 = \frac{E_4^*}{E_1^*}.$$

Thus the resulting minimum *MSE* of $t_{M_{1(1)}}$ is given by

$$MSE_{\min}(t_{M_{1(1)}}) = \bar{Y}^2\left(1 - \frac{E_4^{*2}}{E_1^*}\right) \tag{22}$$

We now established the following theorem.

Theorem 2.4. The *MSE* of $t_{M_{1(1)}}$ is greater than equal to the minimum *MSE* of $t_{M_{1(1)}}$ i.e.

$$MSE(t_{M_{1(1)}}) \geq \bar{Y}^2\left(1 - \frac{E_4^{*2}}{E_1^*}\right)$$

with equality holding if

$$M_1 = \frac{E_4^*}{E_1^*}.$$

2.2 Efficiency Comparison

- From (8) and (13) we have that

$MSE_{\min}(t_{(M_1,M_2)}) < MSE_{\min}(t_M)$ if

$$\frac{(E_2E_4^2 - 2E_3E_4E_5 + E_1E_5^2)}{(E_1E_2 - E_3^2)} > 2E_5 - E_2 + \frac{(E_2 - E_3 + E_4 - E_5)^2}{(E_1 + E_2 - 2E_3)}. \tag{23}$$

This always met in survey situations. Thus the proposed class of estimators $t_{(M_1,M_2)}$ is more efficient than the class of estimators t_M.

- From (8) and (18) we note that

$$MSE_{\min}(t_{M_1}) - MSE_{\min}(t_{(M_1,M_2)}) = \frac{\bar{Y}^2(E_1E_5 - E_3E_4)^2}{E_1(E_1E_2 - E_3^2)} > 0, \tag{24}$$

which follows that the proposed class of estimators $t_{(M_1,M_2)}$ is better than t_{M_1}-family of estimators and hence it is more efficient than $t_{M_{1(1)}}$-family of estimators.

If we set $M_1 = 1$ in (19) we get a class of estimators due to Chaudhury et al. (2009):

$$t_C = \bar{y}_{st}^*\left\{\frac{a_x\bar{X} + b_x}{\alpha_x(a_x\bar{x}_{st} + b_x) + (1 - \alpha_x)(a_x\bar{X} + b_x)}\right\}^{g_x} \tag{25}$$

To the first degree of approximation the bias and *MSE* of t_C are respectively given as

$$B(t_C) = \bar{Y}(E_4^* - 1), \tag{26}$$

$$MSE(t_C) = \bar{Y}^2(1 + E_1^* - 2E_4^*). \tag{27}$$

We have from (22) and (27) we have

$$MSE(t_C) - MSE_{\min}(t_{M_{1(1)}}) = \frac{\bar{Y}^2(E_1^* - E_4^*)^2}{E_1^*} > 0 \tag{28}$$

which follows that the proposed $t_{M_{1(1)}}$-family of estimators is more efficient than Chaudhury et al. (2009) class of estimators t_C.

Finally it follows from (24) and (28) that the proposed $t_{(M_1,M_2)}$-family of estimators is better than $t_{M_{1(1)}}$ and t_C-families of estimators.

3 Numerical Illustration

For numerical illustration we consider a data set [Source: Koyuncu and Kadilar (2009)], in which *y*: Number of teachers; *x*: number of students and *z*: number of classes in primary and secondary schools for 923 districts and 6 regions in Turkey in 2007.

Stratum (*h*)		1	2	3	4	5	6
Stratified mean,	N_h	127	117	103	170	205	201
Standard deviations	n_h	31	21	29	38	22	39
and Correlation	n'_h	70	50	75	95	70	90
coefficients	S_{yh}	883.84	644.92	1033.40	810.58	403.65	711.72
	S_{xh}	30486.70	15180.77	27549.69	18218.93	8497.77	23094.14
	S_{zh}	555.58	365.46	612.95	458.03	260.85	397.05
	$\bar{Y}_h$	703.74	413.00	573.17	424.66	267.03	393.84
	$\bar{X}_h$	20804.59	9211.79	14309.30	9478.85	5569.95	12997.59
	$\bar{Z}_h$	498.28	318.33	431.36	311.32	227.20	313.71
	ρ_{yxh}	0.94	1.00	0.99	0.98	0.99	0.97
	ρ_{xzh}	0.94	0.97	0.98	0.96	0.97	1.00
	ρ_{yzh}	0.98	0.98	0.98	0.98	0.96	0.98
W_h=10%	$S_{yh(2)}$	510.57	386.77	1872.88	1603.30	264.19	497.84
Non-response	$S_{xh(2)}$	9446.93	9198.29	52429.99	34794.90	4972.56	12485.10

Stratum (*h*)		1	2	3	4	5	6
	$S_{zh(2)}$	303.92	278.51	960.71	821.29	190.85	287.99
	$\rho_{yxh(2)}$	1.00	1.00	1.00	0.97	1.00	0.93
	$\rho_{xzh(2)}$	0.99	0.99	1.00	0.96	0.99	0.98
	$\rho_{yzh(2)}$	0.99	0.99	1.00	0.99	0.99	0.96
W_h=20%	$S_{yh(2)}$	396.77	406.15	1654.40	1333.35	335.83	903.91
Non-response	$S_{xh(2)}$	7439.16	8880.46	45784.78	29219.30	6540.43	28411.44
	$S_{zh(2)}$	244.56	274.42	965.42	680.28	214.49	469.86
	$\rho_{yxh(2)}$	1.00	0.99	1.00	0.98	1.00	0.99
	$\rho_{xzh(2)}$	0.99	0.99	0.98	0.96	0.98	0.98
	$\rho_{yzh(2)}$	0.99	0.98	0.98	0.99	0.98	0.99
W_h=30%	$S_{yh(2)}$	500.26	356.95	1383.70	1193.47	289.41	825.24
Non-response	$S_{xh(2)}$	14017.99	7812.00	38379.77	26090.60	5611.32	24571.95
	$S_{zh(2)}$	284.44	247.63	811.21	631.28	188.30	437.90
	$\rho_{yxh(2)}$	0.96	0.99	1.00	0.98	1.00	0.97
	$\rho_{xzh(2)}$	0.91	0.98	0.98	0.97	0.98	0.96
	$\rho_{yzh(2)}$	0.97	0.98	0.98	0.99	0.98	0.98

Table 1 *PRE* of t_C when $W_h = 10\%, 20\%$ and 30% non-response for different values of the constants $(g_x, \alpha_x, h_x, a_x, b_x)$

g_x	α_x	h_x	a_x	b_x	10%	20%	30%
−0.5	−0.20	0.75	1	1	118.84	115.69	113.44
−0.5	−0.25	0.75	1	1	124.26	145.26	166.26
−0.5	−0.30	0.75	1	1	130.01	151.98	173.95
−0.5	−0.40	0.75	1	1	142.56	166.65	190.73
−0.5	−0.50	0.75	1	1	156.64	183.11	209.58
−0.5	−0.60	0.75	1	1	172.43	201.56	230.7

Table 1 gives the *PRE* of the Chaudhury et al. (2009) class of estimators t_C when $W_h = 10\%, 20\%$ and 30% non-response respectively for different values of the constants $(g_x, \alpha_x, h_x, a_x, b_x)$.

Table 2 gives the *PRE* of estimator t_M when $W_h = 10\%, 20\%$ and 30% non-response respectively for various values of the constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$.

Table 2 *PRE* of t_M when $W_h = 10\%, 20\%$ and 30% non-response for different values of the constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$

g_x	g_z	α_x	α_z	h_x	h_z	a_x	a_z	b_x	b_z	10%	20%	30%
−0.5	−0.5	−0.2	−0.2	0.75	1.25	1	1	1	1	841	406.07	292.87
−0.5	−0.5	−0.3	−0.3	0.75	1.25	1	1	1	1	652.71	764.80	875.35
−0.5	−0.5	−0.25	−0.25	0.75	1.25	1	1	1	1	715.58	838.56	959.76
−0.5	−0.5	−0.4	−0.3	0.75	1.25	1	1	1	1	841.53	986.46	1129.05
−0.5	−0.5	−0.3	−0.25	0.75	1.25	1	1	1	1	928.01	1088.16	1245.45
−0.5	−0.5	−0.5	−0.3	0.75	1.25	1	1	1	1	936.55	1097.88	1256.58
−0.5	−0.5	−0.4	−0.25	0.75	1.25	1	1	1	1	1304.33	1530.74	1751.99
−0.5	−0.5	−0.5	−0.25	0.75	1.25	1	1	1	1	1409.92	1654.48	1893.63
−0.5	−0.5	−0.25	−0.2	0.75	1.25	1	1	1	1	1515.89	1780.35	2037.68
−0.5	−0.5	−0.6	−0.2	0.75	1.25	1	1	1	1	2619.87	3079.66	3524.80
−0.5	−0.5	−0.3	−0.2	0.75	1.25	1	1	1	1	2854.18	3364.82	3851.19
−0.5	−0.5	−0.5	−0.2	0.75	1.25	1	1	1	1	4795.72	5670.80	6490.48

Table 3 *PRE* of t_{M_1} when $W_h = 10\%, 20\%$ and 30% non-response for different values of the constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$

g_x	g_z	α_x	α_z	h_x	h_z	a_x	a_z	b_x	b_z	10%	20%	30%
−0.5	−0.5	−0.2	−0.2	0.75	1.25	1	1	1	1	137.09	130.58	126.19
−0.5	−0.5	−0.25	−0.2	0.75	1.25	1	1	1	1	143.51	167.78	192.03
−0.5	−0.5	−0.25	−0.25	0.75	1.25	1	1	1	1	148.61	173.73	198.85
−0.5	−0.5	−0.3	−0.2	0.75	1.25	1	1	1	1	150.33	175.75	201.15
−0.5	−0.5	−0.3	−0.25	0.75	1.25	1	1	1	1	155.75	182.09	208.41
−0.5	−0.5	−0.3	−0.3	0.75	1.25	1	1	1	1	161.42	188.73	216.01
−0.5	−0.5	−0.4	−0.25	0.75	1.25	1	1	1	1	171.36	200.35	229.31
−0.5	−0.5	−0.4	−0.3	0.75	1.25	1	1	1	1	177.79	207.87	237.92
−0.5	−0.5	−0.5	−0.2	0.75	1.25	1	1	1	1	181.97	212.74	243.49
−0.5	−0.5	−0.5	−0.25	0.75	1.25	1	1	1	1	188.91	220.86	252.79
−0.5	−0.5	−0.5	−0.3	0.75	1.25	1	1	1	1	196.17	229.37	262.52
−0.5	−0.5	−0.6	−0.2	0.75	1.25	1	1	1	1	200.74	234.69	268.61

Table 3 gives the *PRE* of estimator t_{M_1} when $W_h = 10\%, 20\%$ and 30% non-response respectively for multiple values of the constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$.

Table 4 shows the *PRE* of the proposed estimator $t_{(M_1,M_2)}$ when W_h is 10%, 20% and 30% respectively at varying constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$.

Table 4 *PRE* of the proposed estimator $t_{(M_1,M_2)}$ when $W_h = 10\%, 20\%$ and 30% non-response for different values of the constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$

g_x	g_z	α_x	α_z	h_x	h_z	a_x	a_z	b_x	b_z	10%	20%	30%
−0.5	−0.5	−0.2	−0.2	0.75	1.25	1	1	1	1	919.84	415.39	295.18
−0.5	−0.5	−0.3	−0.3	0.75	1.25	1	1	1	1	693.78	813.14	930.68
−0.5	−0.5	−0.25	−0.25	0.75	1.25	1	1	1	1	767.55	899.74	1029.80
−0.5	−0.5	−0.4	−0.3	0.75	1.25	1	1	1	1	931.43	1092.35	1250.25
−0.5	−0.5	−0.3	−0.25	0.75	1.25	1	1	1	1	1038.48	1218.37	1394.48
−0.5	−0.5	−0.5	−0.3	0.75	1.25	1	1	1	1	1065.42	1249.68	1430.31
−0.5	−0.5	−0.4	−0.25	0.75	1.25	1	1	1	1	1598.60	1878.31	2149.81
−0.5	−0.5	−0.5	−0.25	0.75	1.25	1	1	1	1	1802.76	2118.4	2424.60
−0.5	−0.5	−0.25	−0.2	0.75	1.25	1	1	1	1	1906.51	2242.58	2566.73
−0.5	−0.5	−0.3	−0.2	0.75	1.25	1	1	1	1	5238.85	6226.55	7126.56
−0.5	−0.5	−0.6	−0.2	0.75	1.25	1	1	1	1	5327.30	6303.86	7215.05
−0.5	−0.5	−0.5	−0.2	0.75	1.25	1	1	1	1	68342.8	96766.1	110753

It is observed from Tables 2–4 that for the constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$ considered in these tables, the *PRE*s of the estimators t_C, t_M, t_{M_1} and $t_{(M_1,M_2)}$ are larger than 100%. So the estimators t_C, t_M, t_{M_1} and $t_{(M_1,M_2)}$ are more efficient than the usual unbiased estimator $\bar{y}^*_{st}$ which does not utilize auxiliary information. It shows that the use of auxiliary variable(s) at the estimation stage is advantageous. For all the choices of constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$ the *PRE*s increase for increasing values of W_h expect for the values of constants given in first row of the Tables 2–4, where the values of *PRE*s decrease with increasing values of W_h. Larger gain in efficiency is observed by using the proposed class of estimators $t_{(M_1,M_2)}$ over $\bar{y}^*_{st}$ as compared to the estimators t_C, t_M and t_{M_1}. From the results of the Table 4 it is clear that there is enough scope of selecting the values of the constants $(g_x, g_z, \alpha_x, \alpha_z, h_x, h_z, a_x, a_z, b_x, b_z)$ in obtaining estimators from the suggested class of estimators $t_{(M_1,M_2)}$ better than the estimators $\bar{y}^*_{st}$, t_C, t_M and t_{M_1}. Thus the proposal of the class of estimators $t_{(M_1,M_2)}$ is justified.

4 Conclusion

In this article we have developed the generalized version of Chaudhury et al. (2009) estimator using information on two auxiliary variables in presence of non-response under stratified sampling. In addition to Chaudhury et al.

(2009) estimator, a large number of estimators can be identified as a member of the suggested class of estimators. We have obtained the bias and *MSE* of the envisaged class of estimators $t_{(M_1,M_2)}$ up to first order of approximation. The conditions are obtained at which the class of estimators $t_{(M_1,M_2)}$ has the minimum *MSE*. Thus this study unifies several results at one place. So it is advantageous to the researchers engaged in this area. It has been demonstrated both theoretically and numerically that proposed class of estimators $t_{(M_1,M_2)}$ is more efficient than the Chaudhury et al. (2009) estimator. Thus we recommend the proposed class of estimators $t_{(M_1,M_2)}$ for its use in practice.

Acknowledgement

We are grateful to the two learned referees, associate editor and editor-in-chief for their valuable suggestions that helped to improved our paper.

References

[1] Chaudhury, M. K., Singh, R., Shukla, R. K., Kumar, M. and Smarandache, F. (2009). A family of estimators for estimating population mean in stratified sampling under non-response, *Pakistan Journal of Statistics and Operation Research*, 5(1), 47–54.

[2] Chaudhury, M. K., Singh, V. K., Singh, R., and Smarandache, F. (2011). Estimating the population mean in stratified population using auxiliary information under non-response. *Studies in Sampling Techniques and Time Series Analysis*, Zip Publishing, USA.

[3] Cochran, W. G. (1977). Sampling techniques, *3rd Edition, John Wiley: New York.*

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

[5] Haq, A. and Shabbir, J. (2013). Improved family of ratio estimators in simple and stratified random sampling, *Communications in Statistics-Theory and Methods*, 42(5), 782–799.

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

[7] Khare, B. B. and Srivastava, S. (1997). Transformed ratio-type estimators for the population mean in the presence of non-response, *Communications in Statistics-Theory and Methods*, 26(7), 1779–1791.

[8] Khare, B. B., Srivastava, U. and Kamlesh, K. (2013). Generalized chain type estimators for ratio of two population means using two auxiliary characters in the presence of non-response, *International Journal of Statistics and Economics*, 10(1).

[9] Koyuncu, N. and Kadilar, C. (2009). Ratio and product estimators in stratified random sampling, *Journal of Statistical Planning and Inference*, 139, 2552–2558.

[10] Okafor F. C. and Lee, H. (2000). Double sampling for ratio and regression estimation with sub-sampling the non-respondent, *Survey Methodology*, 26(2), 183–188.

[11] Rao, P. S. R. S. (1986). Ratio estimation with sub-sampling the non-respondents, *Survey Methodology*, 12(2), 217–230.

[12] Saleem, I., Sanaullah, A. and Hanif, M. (2018). A generalized class of estimators for estimating population mean in the presence of non-response, *Journal of Statistical Theory and Applications*, 17(4), 616–626.

[13] Sanaullah, A., Noor-ul-Amin, M. and Hanif, M. (2015). Generalized exponential-type ratio-cum-ratio and product-cum-product estimators for population mean in the presence of non-response under stratified two-phase random sampling, *Pakistan Journal of Statistics and Operation Research*, 31(1), 71–94.

[14] Singh, G. N. and Khalid, M. (2015). Exponential chain dual to ratio and regression type estimators of population mean in two-phase sampling, *Statistica*, 75(4), 379–389.

[15] Singh, H. P. and Kumar, S. (2008). A general family of estimators of finite population ratio, product and mean using two-phase sampling scheme in presence of non-response, *Journal of Statistical Theory and Practice*, 2(4), 677–692.

[16] Singh, H. P. and Kumar, S. (2009). A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information, *SORT*, 33(1), 71–84.

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

[18] Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asok , C. (1984). "Sampling theory and surveys with applications", *Towa State University Press, Ames.*

[19] Tabasum, R. and Khan, I. A. (2004). Double sampling for ratio estimation with non-response, *Journal of Indian Society of Agricultural Statistics*, 53(3), 300–306.
[20] Tabasum, R. and Khan, I. A. (2006). Double sampling ratio estimator for population mean in presence of non-response. *Assam Statistical Review*, 20, 73–83.
[21] Tripathi, T. P. and Khare, B. B. (1997). Estimation of mean vector in presence of non-response, *Communications and Statistics-Theory and Methods*, 26(9), 2255–2269.

Biographies

Housila P. Singh is Professor of Statistics at School of Studies in Statistics, Vikram University, Ujjain. He has guided 22 Ph.D. scholars and has published more than 500 research papers in national and international journals of repute.

Pragati Nigam is Assistant Professor of Statistics at Mandsaur University, Mandsaur. She completed her Ph.D. in 2021 from Vikram University, Ujjain under the guidance of Prof. H. P. Singh. Dr. Pragati published 4 research papers in national and international journals of repute.