Journal of Prime Research in Mathematics Vol. **16(1)***(2020), 52 - 63*

ON THE GENERALIZED CLASS OF ESTIMATORS FOR ESTIMATION OF FINITE POPULATION MEAN IN THE PRESENCE OF NON-RESPONSE PROBLEM

SABA RIAZ[1,*], AMNA NAZEER[2], JAVERIA ABBASI[2], SADIA QAMAR[3]

Abstract. This work considers a generalized class of biased estimators for the estimation of the unknown population mean of the variable of interest accompanying the issue of non-response in the study and in the auxiliary variables. The asymptotic bias and the asymptotic variance of the suggested class are acquired, up to the first degree of approximation and, compared with the linear regression estimator. The efficiency of the suggested estimators while comparing with the linear regression estimator and some other existing estimators are studied regarding percent relative efficiency (PRE). Furthermore, a simulation study also affirms the excellence of the considered class of estimators.

Key words: Biased estimators, incomplete information, linear regression estimator, simulation, efficacy.

1. Introduction

The problem of incomplete information is very common in surveys, specially in socio economic surveys of households, in which individual data are collected. The reasons of missing information may be migration, refusal to respond, not availability at time of survey performed etc. The estimates attained from such insufficient information perhaps ambiguous and tendentious. To evade this bias, it is essential to consider those non-respondents again and contact them either by personal interviews or any other approach to obtain complete information.

[1]Department of Computer Science, SZABIST, Islamabad, Pakistan.
[2]Department of Mathematics, COMSATS University Islamabad, Pakistan.
[3]Department of Statistics, University of Sargodha, Pakistan.
[*] Corresponding author: sabaqau@gmail.com.

Initially [1] deliberated the issue of non-response and suggested a technique for the incomplete information recovery. The key objective of their technique is to sub-sample the non-respondents with the assumption of complete response on second call. By adopting this technique, numerous authors suggested the estimators for mean estimation in the presence of non-response utilizing supplementary information. For instance, see [2, 3, 4, 5, 6] and the references therein. Motivated by the previous work, we propose a generalized class of biased estimators for mean estimation of the study variable using known mean of auxiliary variable along with the non-response issue.

In Section 2, we discuss the theoretical background and notations used in different estimation methods. We briefly discuss [1] technique and it's extensions. The underlying reasons motivating this extension are also presented. A new class of proposed estimators for population mean is presented in Section 3. Section 4 discusses the efficacy comparison of the proposed estimators. Results of simulation studies are presented in Section 5. Finally, a short overview and discussion is presented in the conclusion section.

2. Background and Notations

Let, P be a set of population comprises N different units along with the study and the auxiliary variables, denoted by Y and X, where y_i and x_i, $i = 1, \ldots, N$. Also assume that X and Y are correlated and the unknown mean $\bar{Y}$ is estimated by using known mean $\bar{X}$. Consider n $(n < N)$ size of a sample s is proceeded by simple random sampling without replacement (SRSWOR) sampling to accumulate data of Y and X. Assuming only n_1 can provide information on Y and X and remaining units $n_2 = n - n_1$, considered as non-respondents. Now considering [1] technique, n_2 units are sub-sampled and $r = n_2/k$ units among them are re-contacted, $k > 1$. Assuming all r units provide complete information on the second call. It is necessary that r should be an integer otherwise essential to round it. Hence whole population can be separated into two strata, P_1 be the first stratum of size N_1 units in which respondents give response for first call and second stratum P_2 of size N_2 units consists on persons will answer on the second call. The strata sizes N_1 and N_2 are usually unknown [7]. Authors in [1] suggested this technique for the estimator $\bar{y}^*$ when non response exists in Y.

When non-response is present in X, the similar technique can be used by adopting a two-phase sampling scheme. To estimate $\bar{X}$, the first phase is concerned for selection of a large sample s' of size n' $(n' < N)$ and the second phase is devoted to estimate $\bar{Y}$. For this, a smaller sub-sample s of size n is selected from n' units where $(n < n')$.

To represent estimators, let define a dummy variable $v = (y, x)$

$$\bar{v}^* = d_1\bar{v}_1 + d_2\bar{v}_{2r},$$

where

$$\bar{v} = \frac{\sum_{i=1}^{n} v_i}{n}, \quad \bar{v}_1 = \frac{\sum_{i=1}^{n_1} v_i}{n_1}, \quad \bar{v}_{2r} = \frac{\sum_{i=1}^{r} v_i}{r}, \quad d_1 = \frac{n_1}{n}, \quad d_2 = \frac{n_2}{n}.$$

Also we can have

$$\bar{V} = D_1 \bar{V}_1 + D_2 \bar{V}_2,$$

where

$$\bar{V}_1 = \frac{\sum_{i=1}^{N_1} v_i}{N_1}, \quad \bar{V}_2 = \frac{\sum_{i=1}^{N_2} v_i}{N_2}, \quad D_1 = \frac{N_1}{N}, \quad D_2 = \frac{N_2}{N}.$$

The variance of $\bar{v}^*$ can be written

$$\mathrm{Var}(\bar{v}^*) = \lambda_1 S_v^2 + \lambda_2 S_{v(2)}^2 = \widetilde{S}_v^2 = \bar{V}^2 \widetilde{C}_v^2,, \tag{1}$$

where

$$S_v^2 = \frac{\sum_i^N (v_i - \bar{V})^2}{N-1}, \quad S_{v(2)}^2 = \frac{\sum_i^{N_2} (v_i - \bar{V}_2)^2}{N_2 - 1},$$

$$\lambda_1 = \left(\frac{1}{n} - \frac{1}{N}\right), \quad \lambda_2 = \frac{N_2(k-1)}{nN}.$$

One can define the covariance as

$$\mathrm{Cov}(\bar{y}^*, \bar{x}^*) = \lambda_1 S_{yx} + \lambda_2 S_{yx(2)} = \widetilde{S}_{yx} = \bar{Y}\bar{X}\widetilde{C}_{yx}, \tag{2}$$

where

$$S_{yx} = \frac{\sum_i^N (y_i - \bar{Y})(x_i - \bar{X})}{N-1}, \quad S_{yx(2)} = \frac{\sum_i^{N_2} (y_i - \bar{Y}_2)(x_i - \bar{X}_2)}{N_2 - 1}.$$

Based on the above notations, [1] estimator can be determined as

$$\bar{y}_{\mathrm{HH}}^* = d_1 \bar{y}_1 + d_2 \bar{y}_{2r}. \tag{3}$$

[2] suggested an estimator with known $\bar{X}$, which can be written as,

$$\bar{y}_{\mathrm{KS}}^* = \bar{y}^* + \hat{\beta}_{yx}^* (\bar{X} - \bar{x}^*), \tag{4}$$

where $\hat{\beta}_{yx}^* = \dfrac{s_{yx}^*}{s_x^{*2}}$ is an estimator of $\beta_{yx} = \dfrac{S_{yx}}{S_x^2}$ of y on x,
$s_{yx}^* = \dfrac{\sum_{i=1}^{n_1} y_i x_i + k \sum_{i=1}^{r} y_i x_i - n\bar{y}^*\bar{x}^*}{n-1}$ and $s_x^{*2} = \dfrac{\sum_{i=1}^{n_1} x_i^2 + k \sum_{i=1}^{r} x_i^2 - n\bar{x}^{*2}}{n-1}$.

Recently, [5] proposed a class of estimators with known $\bar{X}$ and non response problem. Their proposed estimators have the form,

$$\bar{y}_{\mathrm{RDS}}^* = \left[w_1 \bar{y}^* + w_2(\bar{X} - \bar{x}^*)\right] \exp\left(\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right), \tag{5}$$

where w_1 and w_2 are constants and can be chosen suitably.

The minimum asymptotic variance of $\bar{y}^*_{\text{RDS}}$ can be written as

$$\text{minAV}(\bar{y}^*_{\text{RDS}}) = \frac{\bar{Y}^2\left[\widetilde{C}_x^6 + 16\widetilde{C}_x^4\widetilde{C}_y^2 - 16\widetilde{C}_x^2(4\widetilde{C}_y^2 + \widetilde{C}_{yx}^2) + 64\widetilde{C}_{yx}^2\right]}{64\left[\widetilde{C}_{yx}^2 - \widetilde{C}_x^2(1+\widetilde{C}_y^2)\right]}. \quad (6)$$

Now, a regression estimator for estimation of $\bar{Y}$ can be considered with assumption of non-response occurrence in both variables

$$\bar{y}^*_{\text{reg}} = \bar{y}^* + w(\bar{X} - \bar{x}^*), \quad (7)$$

where w is a suitable constant to be selected.

It is clear that, the asymptotic variance of $\bar{y}^*_{\text{reg}}$ can be minimum if

$$w = \frac{\widetilde{S}_{yx}}{\widetilde{S}_x^2} = w^o(\text{say}).$$

Therefore, The minimum asymptotic variance of $\bar{y}^*_{\text{reg}}$ can be written as

$$\text{minAV}(\bar{y}^*_{\text{reg}}) = \text{minAV}(\bar{y}^*_{\text{KS}}) = \widetilde{S}_y^2(1 - \widetilde{\rho}_{yx}^2), \quad (8)$$

where $\widetilde{\rho}_{yx}^2 = \dfrac{\widetilde{S}_{yx}^2}{\widetilde{S}_y^2\widetilde{S}_x^2}$.

[8] suggested the following exponential estimator considering the problem of non-response

$$\bar{y}^*_{\text{YK}} = \bar{y}^*\left[\alpha\text{exp}\left(\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right) + (1-\alpha)\text{exp}\left(\frac{\bar{x}^* - \bar{X}}{\bar{x}^* + \bar{X}}\right)\right], \quad (9)$$

where α is a suitably chosen constant. The minimum asymptotic variance of $\bar{y}^*_{\text{YK}}$ can be expressed as

$$\text{minAV}(\bar{y}^*_{\text{YK}}) = \widetilde{S}_y^2(1 - \widetilde{\rho}_{yx}^2) = \text{minAV}(\bar{y}^*_{\text{reg}}). \quad (10)$$

Recently, [9] proposed a class of estimators in presence of non-response

$$\bar{y}^*_{\text{PS}} = \bar{y}^*\left(\frac{\bar{X}}{\bar{x}^*}\right)^\alpha \text{exp}\left(\frac{\eta(\bar{X} - \bar{x}^*)}{\bar{X} + \bar{x}^*}\right), \quad (11)$$

where $(\alpha,\ \eta)$ are suitably chosen constants.

The asymptotic variance of $\bar{y}^*_{\text{PS}}$ can be written as

$$\text{AV}(\bar{y}^*_{\text{PS}}) = \bar{Y}^2\left[\widetilde{C}_y^2 + \widetilde{C}_x^2\left\{\frac{(\eta + 2\alpha)^2}{4} - (\eta + 2\alpha)R\right\}\right], \quad (12)$$

which is minimum when $(\eta + 2\alpha) = 2R$, where $R = \dfrac{\widetilde{C}_{yx}}{\widetilde{C}_x^2}$. Thus,

$$\text{minAV}(\bar{y}^*_{\text{PS}}) = \bar{Y}^2\left(\widetilde{C}_y^2 - \frac{\widetilde{C}_{yx}^2}{\widetilde{C}_x^2}\right) = \widetilde{S}_y^2(1 - \widetilde{\rho}_{yx}^2) = \text{minAV}(\bar{y}^*_{\text{reg}}). \quad (13)$$

3. Proposed Class of Estimators

Taking motivation from literature, one can establish a more generalized class of estimators to estimate the unknown population mean $\bar{Y}$ of the study variable using the transformation of the auxiliary variable X. The proposed class can be given by

$$\bar{y}_{\mathrm{g}}^{*} = \bar{y}^{*}\left[w_1 + w_2(\bar{Z} - \bar{z}^{*})\right]\exp\left(\frac{\eta(\bar{Z} - \bar{z}^{*})}{\bar{Z} + \bar{z}^{*}}\right), \tag{14}$$

where

$$\bar{z}^{*} = a\bar{x}^{*} + b \qquad \text{and} \qquad \bar{Z} = a\bar{X} + b,$$

$a(\neq 0)$ and b are either real numbers or functions of the known parameters of the auxiliary variable X such as $\{S_x, C_x, \beta_2(x), \rho_{yx}\}$ etc, η being a constant takes values (0, 1, -1) for designing different estimators and, (w_1, w_2) are suitably chosen constants.

The class $\bar{y}_{\mathrm{g}}^{*}$ can be expressed in terms of δ's upto first order of approximation

$$\bar{y}_{\mathrm{g}}^{*} \cong \bar{Y}\left[w_1 + w_1\delta_0^{*} - (w_1k_2 + w_2k_1)\delta_1^{*} + (w_1k_3 + w_2k_1k_2)\delta_1^{*2} - (w_1k_2 + w_2k_1)\delta_0^{*}\delta_1^{*}\right], \tag{15}$$

where

$$k_0 = \frac{a\bar{X}}{a\bar{X} + b}, \quad k_1 = a\bar{X}, \quad k_2 = \frac{k_0\eta}{2}, \quad k_3 = \frac{k_0^2\eta}{4} + \frac{k_0^2\eta^2}{8}.$$

Let us define the following error terms

$$\delta_0^{*} = \frac{(\bar{y}^{*} - \bar{Y})}{\bar{Y}}, \quad \delta_1^{*} = \frac{(\bar{x}^{*} - \bar{X})}{\bar{X}},$$

$$E(\delta_0^{*}) = E(\delta_1^{*}) = 0,$$

$$E(\delta_0^{*2}) = \widetilde{C}_y^2, \quad E(\delta_1^{*2}) = \widetilde{C}_x^2 \quad \text{and} \quad E(\delta_0^{*}\delta_1^{*}) = \widetilde{C}_{yx}.$$

The asymptotic bias (AB) and the asymptotic variance (AV) of $\bar{y}_{\mathrm{g}}^{*}$ to the first degree of approximation can be written as

$$\mathrm{AB}(\bar{y}_{\mathrm{g}}^{*}) = \bar{Y}\left[w_1\left(1 + k_3\widetilde{C}_x^2 - k_2\widetilde{C}_{yx}\right) + w_2\left(k_1k_2\widetilde{C}_x^2 - k_1\widetilde{C}_{yx}\right) - 1\right] \tag{16}$$

and

$$\mathrm{AV}(\bar{y}_{\mathrm{g}}^{*}) = \bar{Y}\left[1 + w_1^2A_1 + w_2^2A_2 - 2w_1A_3 - 2w_2A_4 + 2w_1w_2A_5\right], \tag{17}$$

where

$$A_1 = 1 + \widetilde{C}_y^2 + (k_2^2 + 2k_3)\widetilde{C}_x^2 - 4k_2\widetilde{C}_{yx},$$

$$A_2 = k_1^2\widetilde{C}_x^2,$$

$$A_3 = 1 + k_3\widetilde{C}_x^2 - k_2\widetilde{C}_{yx},$$

$$A_4 = k_1k_2\widetilde{C}_x^2 - k_1\widetilde{C}_{yx}$$

Table 3.1. Some members of the class of estimators $\bar{y}^*_{\text{g}}$

2*Estimators	a	b	η
$\bar{y}^*_{\text{g}(1)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]$	1	0	0
$\bar{y}^*_{\text{g}(2)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]\exp\left(\dfrac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right)$	1	0	1
$\bar{y}^*_{\text{g}(3)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]\exp\left(\dfrac{\bar{x}^* - \bar{X}}{\bar{x}^* + \bar{X}}\right)$	1	0	-1
$\bar{y}^*_{\text{g}(4)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]$	1	ρ	0
$\bar{y}^*_{\text{g}(5)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]\exp\left(\dfrac{(\bar{X} - \bar{x}^*)}{(\bar{X} + \bar{x}^*) + 2\rho}\right)$	1	ρ	1
$\bar{y}^*_{\text{g}(6)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]\exp\left(\dfrac{(\bar{x}^* - \bar{X})}{(\bar{x}^* + \bar{X}) + 2\rho}\right)$	1	ρ	-1
$\bar{y}^*_{\text{g}(7)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]$	1	S_x	0
$\bar{y}^*_{\text{g}(8)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]\exp\left(\dfrac{(\bar{X} - \bar{x}^*)}{(\bar{X} + \bar{x}^*) + 2S_x}\right)$	1	S_x	1
$\bar{y}^*_{\text{g}(9)} = \bar{y}^*\left[w_1 + w_2(\bar{X} - \bar{x}^*)\right]\exp\left(\dfrac{(\bar{x}^* - \bar{X})}{(\bar{x}^* + \bar{X}) + 2S_x}\right)$	1	S_x	-1
$\bar{y}^*_{\text{g}(10)} = \bar{y}^*\left[w_1 + w_2\rho(\bar{X} - \bar{x}^*)\right]$	ρ	C_x	0
$\bar{y}^*_{\text{g}(11)} = \bar{y}^*\left[w_1 + w_2\rho(\bar{X} - \bar{x}^*)\right]\exp\left(\dfrac{\rho(\bar{X} - \bar{x}^*)}{\rho(\bar{X} + \bar{x}^*) + 2C_x}\right)$	ρ	C_x	1
$\bar{y}^*_{\text{g}(12)} = \bar{y}^*\left[w_1 + w_2\rho(\bar{X} - \bar{x}^*)\right]\exp\left(\dfrac{\rho(\bar{x}^* - \bar{X})}{\rho(\bar{x}^* + \bar{X}) + 2C_x}\right)$	ρ	C_x	-1

and

$$A_5 = 2k_1k_2\widetilde{C}_x^2 - 2k_1\widetilde{C}_{yx}.$$

Now minimizing $\text{AV}(\bar{y}^*_{\text{g}})$ to achieve the optimum values of w_1 and w_2

$$w_1 = \frac{A_2A_3 - A_4A_5}{A_1A_2 - A_5^2} = w_1^o(\text{say})$$

and

$$w_2 = \frac{A_1A_4 - A_3A_5}{A_1A_2 - A_5^2} = w_2^o(\text{say})$$

$$\text{minAV}(\bar{y}^*_{\text{g}}) \cong \bar{Y}^2\left[1 - \left(\frac{A_1A_4^2 + A_2A_3^2 - 2A_3A_4A_5}{A_1A_2 - A_5^2}\right)\right]. \tag{18}$$

Here, Table 3.1 is showing some possible estimators that are members of the suggested class $\bar{y}^*_{\text{g}}$

4. Efficiency Comparison

Generally analytical and numerical collations are examined to affirm the superiority of the suggested class of estimators over the classical estimators. From Eq.(10) and Eq.(18), it seems a complicated chore to display the analytic comparison between the linear regression estimator and the proposed class of estimators. To get numerical results, two population data sets are studied, earlier considered by [5, 10] and the references therein.

Population-I: Source [5]
A data structure with 96 observations are as follows

$$Y = \text{the number of cultivators of villages,}$$
$$X = \text{the population of villages,}$$

The numerical results are provided for 25% weight of missing values and consider last 24 values as non-respondents as earlier considered by [5].
Population-II: Source [10]
The data comprises the primary and secondary schools of 923 districts of Turkey in 2007. The detail of variables is given below

$$Y = \text{the number of teachers at primary and secondary level,}$$
$$X = \text{the number of students at primary and secondary level,}$$

The numerical results are yielded for 10% weight of missing values and examine last 92 values as non-responsive as earlier considered by [10].

It is important to notice that numerous options can be shown for weights of the missing values e.g (10%, 20%, 30%, 40%) etc. All these possibilities are considered and it is examined that the relative efficiency of the suggested estimators is not influenced by different weights of missing values. Though numerical outcomes are distinct for variant weights, however the response of results is similar in all cases.

The Percent Relative Efficiency (PRE) are used to make the comparison of the suggested estimators w.r.t the regression estimator. The results for different values of k are shown in Tables 4.2 and 4.3

$$\text{PRE}(\bar{y}^*_\bullet) = \frac{\text{minAV}(\bar{y}^*_{\text{reg}})}{\text{minAV}(\bar{y}^*_\bullet)} \times 100$$

where $\bar{y}^*_\bullet = (\bar{y}^*_{\text{HH}},\ \bar{y}^*_{\text{RDS}},\ \bar{y}^*_{\text{YK}},\ \bar{y}^*_{\text{PS}},\ \bar{y}^*_{\text{g}(1)}, \cdots\ \bar{y}^*_{\text{g}(12)})$.

It can be seen in Tables 4.2 and 4.3 that the PREs of the suggested estimators are more efficient than the PREs of the estimators $\bar{y}^*_{\text{HH}}$, $\bar{y}^*_{\text{RDS}}$, $\bar{y}^*_{\text{YK}}$, $\bar{y}^*_{\text{PS}}$ and $\bar{y}^*_{\text{reg}}$. It is also observed that the estimator $\bar{y}^*_{\text{g}(3)}$ is more efficient than all other considered estimators. It is important to note that the use of known population parameters e.g (ρ, S_x, C_x) etc. helps to shape the different suggested estimators but it does not contribute much in increasing the efficiency of the suggested estimators.

TABLE 4.1. Summery statistics for Population-I and Population-II

Parameters	Pop-I	Pop-II
N	96	923
n	25	180
$\bar{Y}$	185.22	436.43
$\bar{X}$	1807.23	11440.5
S_x	1921.77	21331.13
S_y	195.03	749.94
S_{yx}	338835.88	15266040
ρ_{yx}	0.904	0.95
N_2	24	92
$S_{x(2)}$	1068.44	30647.36
$S_{y(2)}$	97.82	876.42
S_{yx}	93560.01	26480062
$\rho_{yx(2)}$	0.895	0.99

5. SIMULATION STUDY

In previous Section, it can be seen that the minimum asymptotic variance of the proposed class contains the population parameters. In Section of efficiency comparisons, we assumed that all these population parameters are known. But generally in actual circumstances, these parameters are not known and can not be predicted on the base of prior information or a preliminary survey. Thus, it is required to estimate them. In such situations, additional sources of discrepancy are imported in the estimates that could be disable the analytical collations. Thus now we focus to the efficiency comparisons when unknown population parameters are estimated from the chosen sample. The empirical performance of the estimators is analyzed by using a Monte Carlo simulation. The following simulation design is earlier considered by [11]: a numeral investigation is carried out by taking a population of size N=100,000 observations, comprising 40% missing values. Let gamma distribution is used to generate the values for X variable as $X \sim G(a, b)$ along parameters (a=2.2, b=3.5) and Y variable correlated with X is determined as $y_i = Rx_i + \varepsilon x_i^g$ where $\varepsilon \sim N(0, 1)$, R=(1.0, 2.0) and g=1.5. Let the sample size $n = 500$ is obtained by without replacement sampling scheme. The sampling has been simulated $B = 1,000$ times. Only the following estimators $\left(\bar{y}^*_{\mathrm{g}(1)},\ \bar{y}^*_{\mathrm{g}(2)},\ \bar{y}^*_{\mathrm{g}(3)},\ \bar{y}^*_{\mathrm{reg}}\right)$ are taken into account to investigate the behavior for different values of ρ_{yx} and k

TABLE 4.2. PREs of the estimators w.r.t $\bar{y}^*_{\text{reg}}$ for different values of k for Pop-I

Estimators	Values of parameters				k			
	a	b	α	η	2	3	4	5
$\bar{y}^*_{\text{HH}}$	-	-	-	-	18.48	18.64	18.77	18.88
$\bar{y}^*_{\text{reg}}=\bar{y}^*_{\text{KS}}=\bar{y}^*_{\text{PS}}$	-	-	-	-	100	100	100	100
$\bar{y}^*_{\text{RDS}}$	-	-	-	-	101.93	102.11	102.30	102.48
$\bar{y}^*_{\text{g(1)}}$	1	0	-	0	119.02	129.48	139.87	150.21
$\bar{y}^*_{\text{g(2)}}$	1	0	-	1	109.78	119.43	129.01	138.55
$\bar{y}^*_{\text{g(3)}}$	1	0	-	-1	**206.39***	**224.52***	**242.54***	**260.47***
$\bar{y}^*_{\text{g(4)}}$	1	ρ	-	0	119.02	129.48	139.87	150.21
$\bar{y}^*_{\text{g(5)}}$	1	ρ	-	1	109.78	119.43	129.01	138.55
$\bar{y}^*_{\text{g(6)}}$	1	ρ	-	-1	**206.22**	**224.35**	**242.35**	**260.26**
$\bar{y}^*_{\text{g(7)}}$	1	S_x	-	0	119.02	129.48	139.87	150.21
$\bar{y}^*_{\text{g(8)}}$	1	S_x	-	1	112.85	122.76	132.62	142.42
$\bar{y}^*_{\text{g(9)}}$	1	S_x	-	-1	133.92	145.69	157.38	169.01
$\bar{y}^*_{\text{g(10)}}$	ρ	C_x	-	0	119.02	129.48	139.87	150.21
$\bar{y}^*_{\text{g(11)}}$	ρ	C_x	-	1	113.10	123.04	132.91	142.74
$\bar{y}^*_{\text{g(12)}}$	ρ	C_x	-	-1	132.59	144.24	155.82	167.33

because these estimators are already proved good ones in terms of efficiency. Moreover it is already discussed in previous section that the use of known population parameters only build the different shapes of the estimators and contribute less in efficiency manners.

The simulated mean square error and the simulated bias for each studied estimator, are computed as

$$\widehat{\text{Bias}(\bar{y}^*_{\text{reg}})} = \frac{\sum_{i=1}^{B}\left(\bar{y}^{(i)}_{\text{reg}} - \bar{Y}\right)}{B},$$

$$\widehat{\text{Bias}(\bar{y}^*_{\text{g(.)}})} = \frac{\sum_{i=1}^{B}\left(\bar{y}^{(i)}_{\text{g(.)}} - \bar{Y}\right)}{B},$$

TABLE 4.3. PREs of the estimators w.r.t $\bar{y}^*_{\text{reg}}$ for different values of k for Pop-II

Estimators	Values of parameters				k			
	a	b	α	η	2	3	4	5
$\bar{y}^*_{\text{HH}}$	-	-	-	-	8.45	7.97	7.54	7.16
$\bar{y}^*_{\text{reg}}$=$\bar{y}^*_{\text{KS}}$=$\bar{y}^*_{\text{YK}}$=$\bar{y}^*_{\text{PS}}$	-	-		-	100	100	100	100
$\bar{y}^*_{\text{RDS}}$	-	-	-	-	101.09	101.36	101.65	101.95
$\bar{y}^*_{\text{g}(1)}$	1	0	-	0	123.36	133.20	141.91	149.91
$\bar{y}^*_{\text{g}(2)}$	1	0	-	1	111.02	119.87	127.71	134.91
$\bar{y}^*_{\text{g}(3)}$	1	0	-	-1	**224.99***	**242.93***	**258.82***	**273.41***
$\bar{y}^*_{\text{g}(4)}$	1	ρ	-	0	123.36	133.20	141.91	149.91
$\bar{y}^*_{\text{g}(5)}$	1	ρ	-	1	111.02	119.87	127.71	134.91
$\bar{y}^*_{\text{g}(6)}$	1	ρ	-	-1	**224.96**	**242.90**	**258.78**	**273.37**
$\bar{y}^*_{\text{g}(7)}$	1	S_x	-	0	123.36	133.20	141.91	149.91
$\bar{y}^*_{\text{g}(8)}$	1	S_x	-	1	117.53	126.90	135.19	142.82
$\bar{y}^*_{\text{g}(9)}$	1	S_x	-	-1	133.73	144.39	153.83	162.51
$\bar{y}^*_{\text{g}(10)}$	ρ	C_x	-	0	123.36	133.20	141.91	149.91
$\bar{y}^*_{\text{g}(11)}$	ρ	C_x	-	1	117.68	127.07	135.37	143.01
$\bar{y}^*_{\text{g}(12)}$	ρ	C_x	-	-1	133.26	143.89	153.30	161.94

$$\widehat{\text{MSE}(\bar{y}^*_{\text{reg}})} = \frac{\sum_{i=1}^{B}\left(\bar{y}^{(i)}_{\text{reg}} - \bar{Y}\right)^2}{B},$$

$$\widehat{\text{MSE}(\bar{y}^*_{\text{g}(.)})} = \frac{\sum_{i=1}^{B}\left(\bar{y}^{(i)}_{\text{g}(.)} - \bar{Y}\right)^2}{B}.$$

and

$$\widehat{\text{PRE}(\bar{y}^*_{\text{g}(.)})} = \frac{\widehat{\text{MSE}(\bar{y}^*_{\text{reg}})}}{\widehat{\text{MSE}(\bar{y}^*_{\text{g}(.)})}} \times 100.$$

Table 5.1. Simulated results of estimated population parameters

Estimators	$k=2$		$k=3$		$k=4$		$k=5$	
	$\rho=0.70$	$\rho=0.43$	$\rho=0.70$	$\rho=0.43$	$\rho=0.70$	$\rho=0.43$	$\rho=0.70$	$\rho=0.43$
Simulated PRE								
$\bar{y}^*_{\text{reg}}$	100	100	100	100	100	100	100	100
$\bar{y}^*_{\text{g}(1)}$	147.34	215.80*	163.26	245.10*	167.70*	238.13*	164.52	**246.29***
$\bar{y}^*_{\text{g}(2)}$	150.69*	211.92	165.13*	232.52	166.64	225.09	171.02*	234.25
$\bar{y}^*_{\text{g}(3)}$	97.84	199.41	104.79	217.96	112.53	212.00	106.54	213.46
Simulated Bias								
$\bar{y}^*_{\text{reg}}$	1.24	0.61	1.24	0.61	1.24	0.61	1.24	0.61
$\bar{y}^*_{\text{g}(1)}$	0.67	0.19	0.67	0.19	0.67	0.19	0.67	0.19
$\bar{y}^*_{\text{g}(2)}$	0.63	0.18	0.63	0.18	0.63	0.18	0.63	0.18
$\bar{y}^*_{\text{g}(3)}$	0.63	0.16	0.63	0.16	0.63	0.16	0.63	0.16

The results are shown in Table 5.1. To highlight the performance of the considered estimators, "*" sign is used to illustrate the more efficient estimators than other estimators and "bold" sign is showing the best one among others. Table 5.1 is showing that all the suggested estimators are more efficient than the regression estimator.

6. Conclusion

The current paper is concentrating on a generalized class of estimators for the estimation of $\bar{Y}$ using the auxiliary information. The non-response problem is deliberated on both variables. We establish the asymptotic bias and the asymptotic variance for the proposed class. Regression estimator is used to compare the efficiency of the proposed class because it is observed that regression estimator is always more efficient than [1]. Numerical results of efficiency comparison are shown in Tables 4.2 and 4.3. It is noted that the performance of the proposed estimators is better than the other considered estimators in terms of efficiency. Furthermore, we have analyzed the numerical comparison on the real population by a Monte Carlo study with the intention to comprehend the validness of certain results when extra estimates are needed.

References

[1] Hansen, M. H., and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association, 41*(236), 517-529.

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

[3] Singh, H. P., and Kumar, S. (2010). Improved estimation of population mean under two-phase sampling with subsampling the non-respondents. *Journal of Statistical Planning and Inference, 140*(9), 2536-2550.

[4] Riaz, S., Diana, G., and Shabbir, J. (2014). A general class of estimators for the population mean using multi-phase sampling with the non-respondents. *Hacettepe Journal of Mathematics and Statistics, 43*(3), 511-527.

[5] Saba Riaz, G. D., and Shabbir, J. (2014). Improved classes of estimators for population mean in presence of non-response. *Pak. J. Statist, 30*(1), 83-100.

[6] Pal, S. K., and Singh, H. P. (2016). Finite Population Mean Estimation through a Two-Parameter Ratio Estimator Using Auxiliary Information in Presence of Non-Response. *Journal of Applied Mathematics, Statistics and Informatics, 12*(2), 5-39.

[7] Tripathi, T. P., and Khare, B. B. (1997). Estimation of mean vector in presence of non-response. *Communications in statistics-theory and methods, 26*(9), 2255-2269.

[8] Yunusa, O., and Kumar, S. (2014). Ratio-cum-product estimator using exponential estimator in the presence of non-response. *Journal of Advanced Computing, 3*(1), 1-11.

[9] Pal, S. K., and Singh, H. P. (2016). Finite Population Mean Estimation through a Two-Parameter Ratio Estimator Using Auxiliary Information in Presence of Non-Response. *Journal of Applied Mathematics, Statistics and Informatics, 12*(2), 5-39.

[10] Riaz, S., Diana, G., and Shabbir, J. (2017). Modi fied classes of estimators in circular systematic sampling. *Hacettepe Journal of Mathematics and Statistics, 46*(4), 743-765.

[11] Diana, G., Riaz, S., and Shabbir, J. (2014). Hansen and Hurwitz estimator with scrambled response on the second call. *Journal of Applied Statistics, 41*(3), 596-611.