Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120118

An Exponential Ratio Type Estimator of the Population Mean in The Presence of Non-Response Using Double Sampling

Adnan Khan[1], Amjad Ali[1], Muhammad Ijaz[2], Muhammad Azeem[1] and Mahmoud El-Morshedy[3,4,]*

[1]Department of Statistics, Islamia College, Peshawar, Pakistan
[2]Department of Mathematics and Statistics, University of Haripur, Haripur, Pakistan
[3] Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[4] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, Egypt

Received: 22 Nov. 2021, Revised: 13 Feb. 2022, Accepted: 15 May 2022
Published online: 1 Jan. 2023

Abstract: In modern era, proper and effective planning can be only possible using statistical techniques to estimate different characteristics of population under studies. An appropriate sample design based on efficient estimation technique is desirable to extract maximum information from sample data. It is a well-known phenomenon to use auxiliary information and to reduce the negative impact of non-response using Hansen & Hurwitz approach that further increase the efficiency of an estimator. Information on one or more auxiliary variables correlated with study variable in several ways to get more reliable estimate. The current paper presents a novel Exponential ratio type estimator to estimate the population mean under the problem of non- response. The proposed estimator further reduces the mean square error in the case of double sampling scheme. Approximate algebraic expressions of the mean square error are discussed; in addition, two real applications are also presented. Several ratio and regression type estimators were developed which perform better with several optimization constants under double sampling in the presence of non-response. However, the proposed estimator has utilized information on auxiliary variables in exponential form and place optimization constants in such positions which further increase the efficiency of the estimator even in all level of correlation coefficients among study and auxiliary variables. Real world data examples, as well as simulation study have been performed to know efficiency of the proposed method over mentioned competitors.

Keywords: Auxiliary Information, Non-response, Estimation, Mean Square Error, Double sampling.

1 Introduction

In recent years, estimating the population mean in the presence of non-response using auxiliary information has gained much more attention from the researchers. In practice, during survey sampling, some of the sampling units do not respond on the first attempt due to certain reasons such as lack of understanding, unavailability of the respondents, or other serious reasons. When non-responding units are totally ignored, the reliability of the estimate decreases. So, it is mandatory to keep the non-response into account to get more reliable estimates. To improve sample surveys, the auxiliary information is used at design or at the stage of estimation to control sampling error. Auxiliary variable(s) is chosen at the estimation stage having a significant correlation with the variable of interest. Moreover, when information is not available about population characteristics of auxiliary variables then two phase sampling approach is useful to estimate it from the first phase sample relatively cheaper than on main variable during the survey process. For detailed information and study on the use of auxiliary variables, we refer to see [1-4].

The pioneer work to deal with the problem of non-response had been done by Hansen and Hurwitz [5]. They introduced sub-sampling method to draw apart from the bulk of non-respondents in the original sample to collect information about no responding units which reduced systematic error due to non-responses. They also developed a weighted estimator of the population mean. Later on, the idea of Hansen and Hurwitz [5] has been expanded to reduce the impact of non-response in

[*]Corresponding author e-mail: mah_elmorshedy@mans.edu.eg

ratio, regression and product type estimators. For example, ratio and regression estimators to inculcate information from the non-respondents as well in single phase paradigm [6]. Singh et al [7] studied exponential estimator, Bahl and Tuteja [8] exponential estimator in the same situation, and confirmed its performance in case of double sampling and under non-response. In similar fashion, Khare and Srivastava [9] suggested transform ratio type estimator for population mean for the said purpose. Furthermore, a number of ratio, product, regression and exponential ratio type estimators have been developed to estimate the population mean. Considering information on a single auxiliary variable, many researchers developed different estimators for estimating the population mean [10-12]. Singh et.al [13] and Shabbir and Saeed [14] use two auxiliary variables for further extensions of the existing estimator. The estimation method introduced by Shabbir and Saeed [14] in the presence of non-response under sampling using auxiliary information with mixture idea of ratio and regression estimators compared and confirmed their efficiency with Singh et.al [13] in four different possible situations. Following the work given in [14-15] proposed a class of estimators in same situations with different strategies show equal efficiency in all situations with Shabbir and Saeed [14]. Muneer et al. [16-17] developed classed of ratio estimators coping non-response in single phase sampling. Unal and Kadilar [18] introduced a new ratio cum exponential estimator using single auxiliary variable for population mean in the presence non-response. Ensuing the idea of [13-15] and [15] several efficient types of estimators of population mean in the presence of non-response in different situations, Boushun and Pandey [3] a modified regression estimator with three optimization constants dealing the problem of non-response and compared with mentioned competitors in terms of efficiency in double sampling. . In practice, several ratio and regression type estimators were developed in literature including [13-15], and [3] utilized auxiliary information strongly correlated with study variable and perform better with several optimization constants under double sampling in the presence of non-response. However, the proposed estimator has utilized information on auxiliary variables in exponential form and place optimization constants in such positions which further increase the efficiency of the estimator even in all level of correlation coefficients among study and auxiliary variables.

The novelty of the proposed work is to make the estimation results better than others by reducing the mean square error. Secondly, the proposed estimator performs better for all possible levels of the correlation among the study and auxiliary variables in the case of ratio and regression type estimator. Thirdly, the additional parameters k_1, k_2 are used to control abnormal variation happened in the study variable in both situations.

In the remaining parts of the manuscript, in section 2, the mathematical expression of the proposed estimators with optimization constants and minimum mean square errors in both situations are mentioned. In section 3, we placed pioneer estimators and some recently developed estimators studied in the same situations and properly cited for the purpose of comparison. In section 4, a Monte Carlo simulation and a real data set have been considered for comparison purposes, and in last section 5, the concluding remarks are presented. The mathematical derivation of the mean square error with proper notations are given in the appendix.

2 Proposed Estimator

The motivation behind the proposed exponential type estimator of the population mean handling non-response under two phase sampling are mentioned in introduction section. Moreover, the proposed estimator has utilized information on auxiliary variables in exponential form and place optimization constants in such positions which further increase the efficiency of the estimator even in all level of correlation coefficients among study and auxiliary variables. The proposed estimator is presented under different situations in the following ways:

2.1 Proposed Estimator under condition-I: (When $\bar{X}$ and $\bar{Z}$ are unknown and non- response occurs on Y only)

The proposed estimator under condition-I is defined by

$$T_{n(1)} = k_1 \bar{y}^* \exp\left(\frac{\bar{x}_1 - \bar{x}}{\bar{x}_1 + (\alpha_1 - 1)\bar{x}} + \frac{\bar{z}_1 - \bar{z}}{\bar{z}_1 + (\beta_1 - 1)\bar{z}} \right) \tag{1}$$

where k_1, α_1 and β_1 are optimimzation constants

The Mean square error of (1) is given by

$$MSE\left(T_{n(1)}\right) \approx k_1^2 \lambda_3 \bar{Y}^2 \left(C_y^2 + \frac{1}{\alpha_1^2} C_x^2 + \frac{1}{\beta_1^2} C_z^2 - \frac{2\rho_{yx} C_y C_x}{\alpha_1} - \frac{2\rho_{yz} C_y C_z}{\beta_1} + \frac{2\rho_{xz} C_z C_x}{\alpha_1 \beta_1} \right) + (k_1 - 1)^2 \bar{Y}^2 + k_1^2 \lambda_1 \bar{Y}^2 C_y^2 + k_1^2 \theta \bar{Y}^2 C_{y(2)}^2 \quad (2)$$

Differentiating (2) w.r.t. k1, α1 and β1, the optimum values of k1, α1 and β1 are given by

$$\alpha_{1opt} = \frac{1 - H_{zx} H_{xz}}{H_{yx} - H_{yz} H_{zx}}, \qquad \beta_{1opt} = \frac{1}{H_{yz} - \frac{1}{\alpha_{1opt}} H_{xz}}, \quad k_{1opt} = \frac{1}{\left(1 + \lambda_3 A^* + \lambda_1 C_y^2 + \theta C_{y(2)}^2\right)}.$$

where

$$H_{yx} = \rho_{yx} \frac{C_y}{C_x}, \quad H_{yz} = \rho_{yz} \frac{C_y}{C_z}, \quad H_{xz} = \rho_{xz} \frac{C_x}{C_z}, \quad H_{zx} = \rho_{zx} \frac{C_z}{C_x},$$

$$A^* = C_y^2 + \frac{1}{\alpha_{opt}^2} C_x^2 + \frac{1}{\beta_{opt}^2} C_z^2 - \frac{2\rho_{yx} C_y C_x}{\alpha_{opt}} - \frac{2\rho_{yz} C_y C_z}{\beta_{opt}} + \frac{2\rho_{xz} C_z C_x}{\alpha_{opt} \beta_{opt}}.$$

2.2 Proposed Estimator under condition-II (When $\bar{X}, \bar{Z}$ are unknown and non-response occurs on X, Z and Y)

The proposed estimator under condition-II is defined by

$$T_{n(2)} = k_2 \bar{y}^* \exp\left(\frac{\bar{x}_1 - \bar{x}^*}{\bar{x}_1 + (\alpha_2 - 1)\bar{x}^*} + \frac{\bar{z}_1 - \bar{z}^*}{\bar{z}_1 + (\beta_2 - 1)\bar{z}^*} \right) \quad (3)$$

.

where k_2, α_2 and β_2 are optimization constants

The Mean square error of (3) is given by

$$MSE\left(T_{n(2)}\right) \approx k_2^2 \lambda_3 \bar{Y}^2 \left(C_y^2 + \frac{1}{\alpha_2^2} C_x^2 + \frac{1}{\beta_2^2} C_z^2 - \frac{2\rho_{yx} C_y C_x}{\alpha_2} - \frac{2\rho_{yz} C_y C_z}{\beta_2} + \frac{2\rho_{xz} C_z C_x}{\alpha_2 \beta_2} \right)$$
$$+ k_2^2 \theta \bar{Y}^2 \left(C_{y(2)}^2 + \frac{1}{\alpha_2^2} C_{x(2)}^2 - \frac{2\rho_{yx(2)} C_{y(2)} C_{x(2)}}{\alpha_2} - \frac{2\rho_{yz(2)} C_{y(2)} C_{z(2)}}{\beta_2} + \frac{2\rho_{xz(2)} C_{x(2)} C_{z(2)}}{\alpha_2 \beta_2} \right)$$
$$+ (k_2 - 1)^2 \bar{Y}^2 + k_2^2 \lambda_1 \bar{Y}^2 C_y^2. \quad (4)$$

where,

$$\alpha_{2opt} = \frac{FL - E^2}{GL - EM}, \qquad \beta_{2opt} = \frac{L}{M - \frac{1}{\alpha_{2opt}} E}, \quad k_{2opt} = \frac{1}{1 + \lambda_3 A^* + \theta C^* + \lambda_1 C_y^2},$$

where

$$F = \lambda_3 C_x^2 + \theta C_{x(2)}^2, \qquad G = \lambda_3 \rho_{yx} C_y C_x + \theta \rho_{yx(2)} C_{y(2)} C_{x(2)},$$

$$E = \lambda_3 \rho_{xz} C_x C_z + \theta \rho_{xz(2)} C_{x(2)} C_{z(2)}, \quad L = \lambda_3 C_z^2 + \theta C_{z(2)}^2,$$

$$M=\lambda_3\rho_{yz}C_yC_z+\theta\rho_{yz(2)}C_{y(2)}C_{z(2)},$$

$$C^*=C^2_{y(2)}+\frac{1}{\alpha^2_{opt}}C^2_{x(2)}+\frac{1}{\beta^2_{opt}}C^2_{z(2)}-\frac{2\rho_{yx(2)}C_{y(2)}C_{x(2)(}}{\alpha_{opt}}-\frac{2\rho_{yz(2)}C_{y(2)}C_{z(2)}}{\beta_{opt}}+\frac{2\rho_{xz(2)}C_{z(2)}C_{x(2)}}{\alpha_{opt}\beta_{opt}}$$

3 Comparisons with other Estimators

The proposed estimator is compared with the following estimators under condition I and II. The detail properties with proper notations of the proposed estimator are given in Appendix.

T_0 Hansen & Hurwitz [5]

$T_{1(1)}$, $T_{1(2)}$ Cochran's [6] Ratio and regression estimators under Situation-I

$T_{2(1)}$, $T_{2(2)}$ Cochran's [6] Ratio and regression estimators under situation-II

$T_{3(1)}$, $T_{3(2)}$ Singh &Kumar [7] Estimator under Situation-I &II

$T_{4(1)}$, $T_{4(2)}$ Shabbir & Saeed [14] under Situation-I &II

$T_{5(1)}$, $T_{5(2)}$ Boushun and Naqvi [15] under Situation-I &II

$T_{6(1)}$, $T_{6(2)}$ Boushun and Pandey [3] under Situation-I &II

$T_{n(1)}$, $T_{n(2)}$ Proposed Estimators under situation -I & II

4 Monte Carlo Simulations

A Simulation study is conducted to study the performance of the proposed estimators against other competitors as the expressions of MSEs are very complicated to compare the relative efficiencies analytically. To validate the performance of proposed estimators, we use the computational of power of modern computing package R. For this purpose, we have generated data from the multivariate normal distribution having three positively correlated variables Y, X, and Z with different parameters and sample sizes. The notations of the proposed estimator are $T_{n(1)}$, and $T_{n(2)}$ under condition -I and II respectively. The population data of size N=1000 on three correlated variables is taken from a multivariate normal distribution which consists of a study variable and two supplementary or auxiliary variables.

The Relative Efficiency of each estimator has been calculated with respect to Hansen and Hurwitz [1] estimator. The general form of the percent relative efficiency (PRE) is defined by

$$PRE=\frac{MSE(T_0)}{MSE(T.)}\times 100$$

Table 1 describes various a sets of parameter values. Table 2-6 reflects the average MSE and the percent relative efficiency averaged out over 1000 simulations.

Table 1: Various sets of Parameter values.

Parameters	Set 1	set 2	set 3	set 4	set 5
n_1	100	100	100	100	100
n_2	40	20	20	40	20

ρ_{yx}	0.45	0.45	0.75	0.92	0.92
ρ_{yz}	0.50	0.50	0.75	0.88	0.88
ρ_{xz}	0.35	0.35	0.75	0.80	0.80
C_y	1.35	1.35	1.35	1.35	1.35
C_x	1.0	1.0	1.0	1.0	1.0
C_z	1.9	1.9	1.9	1.9	1.9

Table 2: MSE and Percent Relative Efficiencies for set 1.

Estimators	$\left(1/h\right)$			
	(1/5)	(1/4)	(1/3)	(1/2)
T_0	1.634329 100.0000	1.483181 100.000	1.219081 100.0000	1.004967 100.0000
$T_{1(1)}$	1.634233 100.0059	1.482757 100.0286	1.219015 100.0055	1.003272 100.1690
$T_{2(1)}$	1.634233 104.5127	1.411284 105.0945	1.148108 106.1817	0.933821 107.6188
$T_{3(1)}$	1.503126 108.7287	1.348766 109.9658	1.086992 112.1518	0.870359 115.4658
$T_{4(1)}$	1.494941 109.3240	1.340186 110.6699	1.078694 113.0146	0.861615 116.6375
$T_{5(1)}$	1.494941 109.3240	1.340186 110.6699	1.078694 113.0146	0.861615 116.6375
$T_{6(1)}$	1.40435 116.3758	1.309866 113.2315	1.004436 121.3697	0.830003 121.1755
$T_{n(1)}$	**1.301963** **125.5280**	**1.186910** **124.9616**	**0.980592** **124.3210**	**0.801034** **125.4587**
$T_{1(2)}$	1.600430 102.1181	1.450891 102.2255	1.215516 100.2933	0.995926 100.9077
$T_{2(2)}$	1.423808 114.786	1.291105 114.8769	1.094117 111.4215	0.899795 111.6885
$T_{3(2)}$	0.977480 167.1982	0.968753 153.1021	0.876200 139.1327	0.810320 124.0210
$T_{4(2)}$	0.972697 168.0203	0.988059 150.1105	0.891680 136.7174	0.787975 130.9930

$T_{5(2)}$	0.972697 168.0203	0.988059 150.1105	0.891680 136.7174	0.787975 130.9930
$T_{6(2)}$	0.932366 175.2884	0.956734 155.0254	0.856789 142.2849	0.776455 129.4302
$T_{n(2)}$	**0.904341** **180.7204**	**0.893936** **182.8239**	**0.824948** **147.7768**	**0.736555** **136.4416**

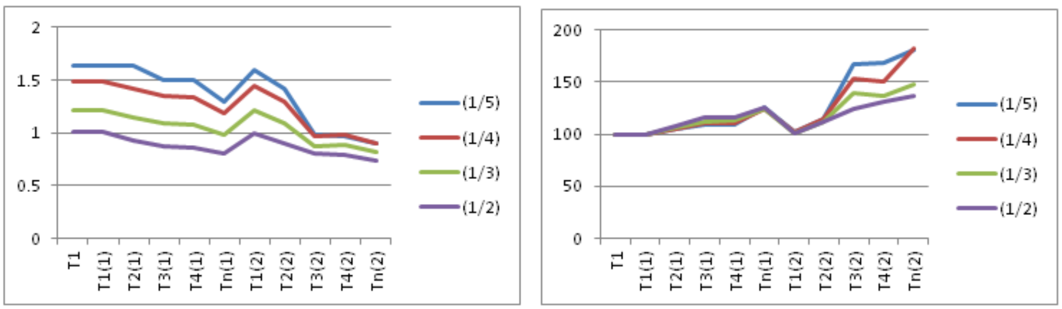

Fig. 1: MSE and PRE for a data set 1.

Table 3: MSE and Percent Relative Efficiencies for Set 2.

Estimators	$\left(\frac{1}{h}\right)$			
	(1/5)	(1/4)	(1/3)	(1/2)
T_0	2.500784 100.0000	2.150689 100.0000	1.800594 100.0000	1.450499 100.0000
$T_{1(1)}$	2.494205 100.2638	2.144110 100.3069	1.794015 100.3667	1.443920 100.4557
$T_{1(2)}$	2.381696 105.0001	2.031601 105.8618	1.681506 107.0822	1.331411 108.9445
$T_{3(1)}$	2.280354 109.6665	1.930259 111.4197	1.580164 113.9498	1.230069 117.9201
$T_{4(1)}$	2.266653 110.3294	1.916558 112.2162	1.566463 114.9465	1.216368 119.2484
$T_{5(1)}$	2.266653 110.3294	1.916558 112.2162	1.566463 114.9465	1.216368 119.2484
$T_{6(1)}$	2.097543 119.224	1.789666 120.172	1.497666 120.226	1.19876 120.999
$T_{n(1)}$	**1.815391** **137.7546**	**1.592595** **135.0431**	**1.353122** **133.0696**	**1.092920** **132.7177**
$T_{1(2)}$	2.378845 105.1260	2.057590 104.5247	1.736335 103.7009	1.415080 102.5030
$T_{2(2)}$	2.117582 118.0962	1.833516 117.2986	1.549449 116.2087	1.265383 114.6293
$T_{3(2)}$	1.425446 175.4387	1.318801 163.0791	1.198985 150.1766	1.057357 137.1816

$T_{4(2)}$	1.299034 192.5111	1.212813 177.3306	1.111711 161.9660	1.003516 144.5417
$T_{5(2)}$	1.299034 192.5111	1.212813 177.3306	1.111711 161.9660	1.003516 144.5417
$T_{6(2)}$	1.268776 197.102	1.197659 179.574	1.099865 163.710	0.997544 145.407
$T_{n(2)}$	**1.256729** **198.9915**	**1.176607** **182.7874**	**1.083170** **166.2337**	**0.968047** **149.8376**

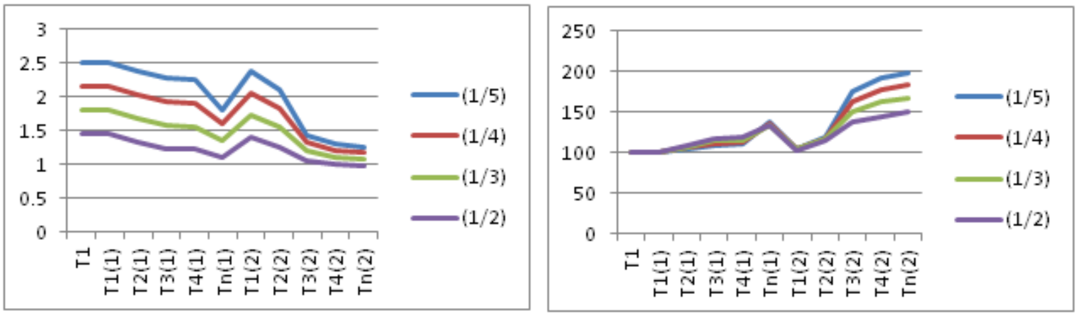

Fig. 2: MSE and PRE for a data set 2.

Table 4: MSE and Percent Relative Efficiencies for set 3.

Estimators	$\left(\frac{1}{h}\right)$			
	(1/5)	(1/4)	(1/3)	(1/2)
T_0	17.44664 100.0000	13.77122 100.0000	10.09581 100.0000	6.420394 100.0000
$T_{1(1)}$	16.58908 105.1694	12.91367 106.6407	9.238250 109.2827	5.562834 115.4159
$T_{1(2)}$	16.04594 108.7293	12.37053 111.3229	8.695110 116.1090	5.019695 127.9041
$T_{3(1)}$	15.83601 110.1707	12.16059 113.2447	8.485175 118.9817	4.809759 133.4868
$T_{4(1)}$	15.73187 110.9000	12.05645 114.2229	8.381037 120.4601	4.705621 136.4409
$T_{5(1)}$	15.73187 110.9000	12.05645 114.2229	8.381037 120.4601	4.705621 136.4409
$T_{6(1)}$	5.982344 291.6355	3.965344 347.2894	2.90568 347.4509	2.002444 320.6279
$T_{n(1)}$	**2.297296** **759.4425**	**1.977292** **696.4691**	**1.622300** **622.3146**	**1.218076** **527.0930**
$T_{1(2)}$	15.46558 112.8095	12.07104 114.0849	8.676498 116.3581	5.281958 121.5533
$T_{2(2)}$	14.48700 120.4296	11.20132 122.9429	7.915639 127.5426	4.629959 138.6706

$T_{3(2)}$	3.790468 460.2767	3.179638 433.1067	2.552019 395.6009	1.891580 339.4197
$T_{4(2)}$	7.400583 235.7468	6.438146 213.9005	5.257791 192.0162	3.661943 175.3275
$T_{5(2)}$	7.400583 235.7468	6.438146 213.9005	5.257791 192.0162	3.661943 175.3275
$T_{6(2)}$	3.446343 506.2363	2.998756 459.2311	1.987222 508.0363	1.009563 635.9577
$T_{n(2)}$	**1.136067** **1535.705**	**1.062612** **1295.979**	**0.978234** **1032.044**	**0.874580** **734.1109**

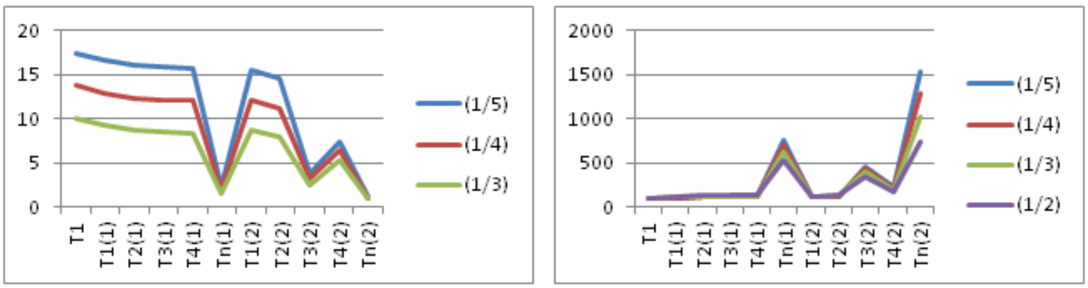

Fig. 3: MSE and PRE for Data Set 3.

Table 5: MSE and Percent Relative Efficiencies for set 4

Estimators	$\left(\frac{1}{h}\right)$			
	(1/5)	(1/4)	(1/3)	(1/2)
T_0	1.754905 100.0000	1.514154 100.0000	1.273404 100.0000	1.032653 100.0000
$T_{1(1)}$	1.359976 129.0394	1.119226 135.2859	0.878475 144.9562	0.637724 161.9278
$T_{1(2)}$	1.320490 132.8980	1.079739 140.2334	0.838988 151.7785	0.598238 172.6158
$T_{3(1)}$	1.309213 134.0428	1.068462 141.7134	0.827711 153.8463	0.586961 175.9322
$T_{4(1)}$	1.289961 136.0432	1.049211 144.3137	0.808460 157.5098	0.567709 181.8982
$T_{5(1)}$	1.289961 136.0432	1.049211 144.3137	0.808460 157.5098	0.567709 181.8982
$T_{6(1)}$	1.209431 145.101	0.976511 155.057	0.789776 161.236	0.557734 185.151
$T_{n(1)}$	**1.120211** **156.6585**	**0.935976** **161.7727**	**0.741965** **171.6257**	**0.536575** **192.4526**
$T_{1(2)}$	0.375025 467.9434	0.568287 266.4419	0.511182 249.1094	0.454078 227.4175

$T_{2(2)}$	0.519088 338.0745	0.478688 316.3134	0.438287 290.5406	0.397887 259.5339
$T_{3(2)}$	0.418160 419.6731	0.401023 377.5729	0.383551 332.0030	0.365514 282.5202
$T_{4(2)}$	0.388695 451.4858	0.376477 402.1904	0.363049 350.7523	0.347512 297.1561
$T_{5(2)}$	0.388695 451.4858	0.376477 402.1904	0.363049 350.7523	0.347512 297.1561
$T_{6(2)}$	0.384323 456.622	0.370932 408.202	0.361234 161.236	0.344562 299.700
$T_{n(2)}$	**0.375025** **467.9434**	**0.363704** **416.3145**	**0.351224** **362.5609**	**0.336736** **306.6655**

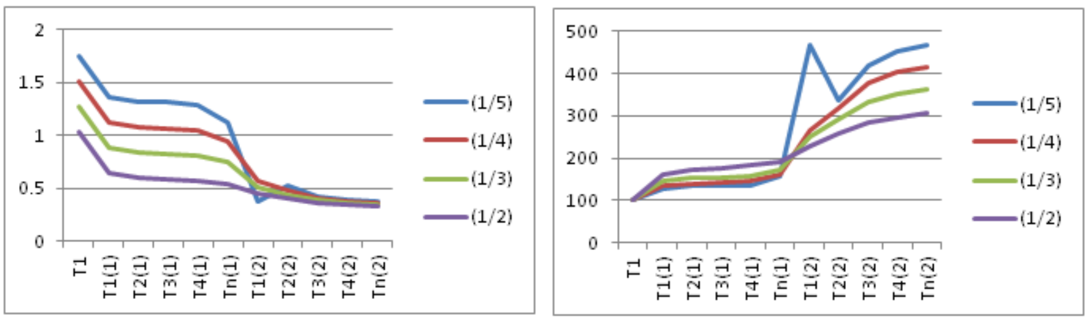

Fig. 4. MSE and PRE for a data set 4

Table 6: MSE and Percent Relative Efficiencies for set 5

Estimators	$\left(1/h\right)$			
	(1/5)	(1/4)	(1/3)	(1/2)
T_0	7.189085 100.0000	6.09891 100.000	5.00874 100.000	3.91856 100.000
$T_{1(1)}$	6.004701 119.7243	4.91453 124.099	3.82435 130.969	2.73418 143.317
$T_{1(2)}$	5.196541 138.3436	4.10637 148.523	3.01619 166.062	1.92602 203.454
$T_{3(1)}$	5.148907 139.6235	4.05873 150.266	2.96856 168.726	1.87838 208.613
$T_{4(1)}$	5.063383 141.9819	3.97321 153.501	2.88303 173.731	1.79286 218.565
$T_{5(1)}$	5.063383 141.9819	3.97321 153.501	2.88303 173.731	1.79286 218.565
$T_{6(1)}$	3.075342 233.765	2.69991 225.89	1.87231 267.51	1.00883 388.42
$T_{n(1)}$	**1.922057** **374.0309**	**1.62707** **374.840**	**1.30278** **384.464**	**0.93888** **417.365**

$T_{1(2)}$	4.482800 160.3704	3.77310 161.642	3.06340 163.502	2.35370 166.485
$T_{2(2)}$	2.053222 350.1367	1.74888 348.733	1.44453 346.738	1.14018 343.677
$T_{3(2)}$	1.405338 511.5556	1.28379 475.069	1.14377 437.916	0.96823 404.714
$T_{4(2)}$	1.158345 620.6343	1.06884 570.608	0.97809 512.089	0.88527 442.638
$T_{5(2)}$	1.158345 620.6343	1.06884 570.608	0.97809 512.089	0.88527 442.638
$T_{6(2)}$	0.858345 837.55	0.82063 743.19	0.77644 645.09	0.67527 580.29
$T_{n(2)}$	**0.715302** **1005.042**	**0.67843** **898.972**	**0.63625** **787.234**	**0.58470** **670.182**

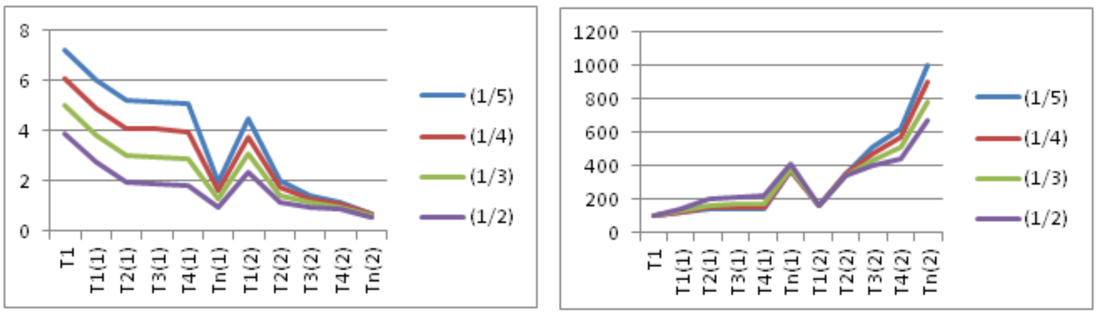

Fig. 5: MSE and PRE for a data set 5.

Table 2-6 reflects the comparison of the average MSE and PRE of the proposed and other estimators for various parameter values and response rates $\left(1/h\right)$. It is clear from table 2-6 that the proposed estimator under condition-I ($T_{n(1)}$) and condition-II ($T_{n(2)}$) have the less MSE and a large PRE than all other existing estimators. Hence, it is evident that the proposed estimator plays an important role in further reeducation of the MSE and leads to the best estimate of the population mean. Fig 1-5 describes the MSE and PRE respectively for the proposed estimators and other existing estimators. The figures clearly define the extreme decline in MSE for the proposed estimators $T_{n(1)}$, and $T_{n(2)}$ when the response rate $\left(1/h\right)$ is increases. Moreover, the line of PRE in increases and reached to it extreme end for the proposed estimators $T_{n(1)}$, and $T_{n(2)}$ when the response rate $\left(1/h\right)$ is increases.

5 Real Data Example

To confirm the performance of the proposed estimator, the following real data set was taken which is earlier used by Khare and Sinha [12].

DATA SET: The data were recorded on three different variables relating to the physical growth of high socio economic class of 95 children under the Indian council for medical research (ICMR) study, pediatrics department, at Banaras Hindu University. Among them, 25% non-responding units were found which were contacted by an expensive method. The necessary descriptive statistics are as under.

y : Children Mass in (Kg),
x : Chest Size in (cm),

z : Arm Circumference in (cm).
Population parameters are given below

$$N = 95, \quad n_1 = 45, \quad n_2 = 35, \quad W_2 = 0.25, \quad \bar{Y} = 19.4968, \quad \bar{X} = 55.8611, \quad \bar{Z} = 16.7968$$

$$C_y = 1.1562, \quad C_{y(2)} = 1.121, \quad C_x = 0.0586, \quad C_{x(2)} = 0.0541, \quad C_z = 0.0865,$$

$$C_{z(2)} = 0.071251, \quad \rho_{yx} = 0.846, \quad \rho_{yx(2)} = 0.729, \quad \rho_{yz} = 0.797, \quad \rho_{yz(2)} = 0.757, \quad \rho_{xz} = 0.725,$$

$$\rho_{xz(2)} = 0.641.$$

Table 7: MSE and Percent Relative Efficiencies of a Real Data

Estimators	$\left(\frac{1}{h}\right)$			
	(1/5)	(1/4)	(1/3)	(1/2)
T_0	22.81579 100.0000	19.40407 100.0000	15.99234 100.0000	12.58061 100.0000
$T_{1(1)}$	22.31260 102.2552	18.90087 102.6623	15.48915 103.2487	12.07742 104.1664
$T_{2(1)}$	18.48649 123.4188	15.07476 128.7189	11.66303 137.1199	8.251305 152.4681
$T_{3(1)}$	18.28247 124.7960	14.87074 130.4848	11.45902 139.5612	8.047291 156.3335
$T_{4(1)}$	18.05642 126.3584	14.64469 132.4990	11.23296 142.3697	7.821237 160.8520
$T_{5(1)}$	18.05642 126.3584	14.64469 132.4990	11.23296 142.3697	7.821237 160.8520
$T_{6(1)}$	17.05632 133.7674	13.76432 140.9737	11.00322 145.3424	7.699222 163.4011
$T_{n(1)}$	**16.45787** **138.6315**	**13.57960** **142.8913**	**10.59816** **150.8973**	**7.509435** **167.5307**
$T_{1(2)}$	21.38414 106.6949	18.20452 106.5892	15.02491 106.4388	11.84530 106.2076
$T_{2(2)}$	11.31395 201.6608	9.695356 200.1377	8.076763 198.0043	6.458170 194.8015
$T_{3(2)}$	9.977321 228.6765	8.654636 224.2043	7.327265 218.2580	5.991139 209.9870
$T_{4(2)}$	8.930169 255.4912	7.810671 248.4302	6.687695 239.1308	5.557700 226.3636
$T_{5(2)}$	8.930169 255.4912	7.810671 248.4302	6.687695 239.1308	5.557700 226.3636
$T_{6(2)}$	8.899444 256.3732	7.696754 252.1072	6.598997 242.345	5.528754 227.5487
$T_{n(2)}$	**8.725174** **261.4938**	**7.653399** **253.5353**	**6.572061** **243.3383**	**5.477607** **229.6735**

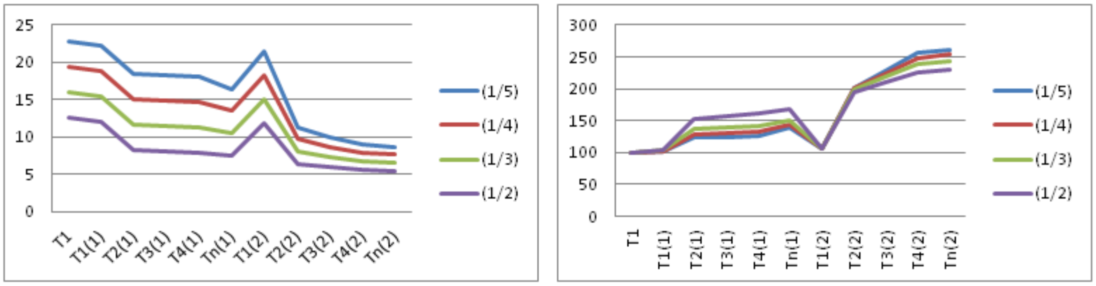

Fig. 6: MSE and PRE for a data set 6.

Table 7 presents the comparison of the average MSE and PRE of the proposed and other estimators with various values of response rates $\left(\frac{1}{h}\right)$ using real data. It is clear from table 7 that the proposed estimator under condition-I ($T_{n(1)}$) and condition-II ($T_{n(2)}$) have less MSE and a large PRE than all other existing estimators. Fig 6 demonstrates that the MSE for the proposed estimators is less whiles the PRE is high than other existing estimators.

5 Conclusions

The performance of the ratio, regression and exponential type estimator of the population mean depends upon auxiliary information, sample size, coefficient of variations of the study variable as well as of the auxiliary variable(s) and the strength of the correlation coefficients.

In this paper, we proposed an optimum estimator of the population mean with two phase sampling in the presence of non-response and its performance is measured over other estimators using simulation and a real data. In a simulation study, we considered 5% and 3% sample sizes of the population under study with different levels of correlation coefficients and coefficient of variations of the study and auxiliary variables. It is concluded that the efficiency of the proposed estimator in all situations is better than its competitors especially, in the small sample size, the performance is far better than their counterparts. The performance of the proposed estimator is compared with Hansen & Hurwitz [5], classical Cochran ratio and regression estimators [6], Singh & Kumar [7] , Shabbir and Saeed [14], Boushun and Naqvi [15] and Boushun and Pandey [3] estimators where non-response was handled under two phase sampling. It is concluded that both the simulation and real data analysis favored the proposed estimator when the auxiliary variable(s) are positively correlated with a study variable. Moreover, in future, one can study the properties of the proposed estimator in other sampling design such as stratified random sampling, systematic, cluster sampling and even unequal probability sampling design with proper adjustment of the auxiliary information.

Acknowledgement

This project was supported by the deanship of scientific research at Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia.

References

[1] Shabbir, Javid, and Nasir Saeed Khan. On estimating the finite population mean using two auxiliary variables in two phase sampling in the presence of non-response. *Communications in Statistics-Theory and Methods*,42(22)., 4127-4145. 2013.

[2] Bhushan, Shashi, and Nazia Naqvi. Generalized efficient classes of estimators in presence of non-response using two auxiliary variables. *Journal of Statistics and Management Systems*, 18(6)., 573-602, 2015.

[3] Bhushan, Shashi, and Abhay Pratap Pandey. An improved estimation procedure of population mean using bivariate auxiliary information under non-response. *Communications for Statistical Applications and Methods*, 26(4)., 347-357, 2019.

[4] Singh, Housila P., Sunil Kumar, and Marcin Kozak. 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, 2010.

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

[6] Cochran, William G. *Sampling techniques*. John Wiley & Sons, 1977.

[7] Singh, Housila P., and Sunil Kumar. Combination of regression and ratio estimate in presence of nonresponse. *Brazilian journal of Probability and Statistics*, 25(2)., 205-217, 2011.

[8] Bahl, Shashi, and RK11082970727 Tuteja. Ratio and product type exponential estimators. *Journal of information and optimization sciences*, 12(1).,159-164, 1991.

[9] Khare, B. B., and Srivastava Srivastava. Transformed ratio type estimators for the population mean in the presence of nonresponse. *Communications in Statistics-Theory and Methods*, 26(7)., 779-1791, 1997.

[10] Särndal, Carl-Erik, and Sixten Lundström. *Estimation in surveys with nonresponse*. John Wiley & Sons, 2005.

[11] Tabasum, R., and I. A. Khan. Double sampling ratio estimator for the population mean in presence of non-response. *Assam Statistical Review*, 20(1).,73-83, 2006.

[12] Khare, B. B., and R. R. Sinha. Estimation of the ratio of the two population means using multi auxiliary characters in the presence of non-response. *Statistical Techniques in Life Testing, Reliability, Sampling Theory and Quality Control* ,1.,63-171, 2007.

[13] Singh, Housila P., Sunil Kumar, and Sandeep Bhougal. Estimation of population mean in successive sampling by sub-sampling non-respondents. *Journal of Modern Applied Statistical Methods*, 10(1)., 6, 2011.

[14] Shabbir, Javid, and Nasir Saeed Khan. On estimating the finite population mean using two auxiliary variables in two phase sampling in the presence of non-response. *Communications in Statistics-Theory and Methods*, 42(22)., 4127-4145, 2013.

[15] Bhushan, Shashi, and Nazia Naqvi. Generalized efficient classes of estimators in presence of non-response using two auxiliary variables. *Journal of Statistics and Management Systems*, 18(6)., 573-602, 2015.

[16] Muneer, Siraj, Javid Shabbir, and Alamgir Khalil. Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables. *Communications in Statistics-Theory and Methods*, 46(5)., 2181-2192, 2017.

[17] Muneer, Siraj, Javid Shabbir, and Alamgir Khalil. A Generalized exponential type estimator of population mean in the presence of non-response. *Statistics in Transition New Series*, 19(2)., 2018.

[18] Ünal, Ceren, and Cem Kadilar. Exponential type estimator for the population mean in the presence of non-response. *Journal of Statistics and Management Systems*, 23(3)., 603-615, 2020.

APPENDIX

NOTATIONS:

N: Size of population, N_1: Responding population size on first attempt

$N_2 = N - N_1$ (Non-responding population size on first attempt)

n_1: First phase sample size, n_2: Second phase sample size, r_1: No. of responding units in second phase sample, r_2: Non-respondents size,

$k = r_2/h$, where $h \geq 2$, where k is the size of the subsample from non-respondings units,

$$\bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2,\ W_1 = \frac{N_1}{N},\ W_2 = \frac{N_2}{N},\ \bar{Y}_1 = \frac{\sum_{i=1}^{N_1} Y_i}{N_1},\ \bar{Y}_2 = \frac{\sum_{i=1}^{N_2} Y_i}{N_2},\ \bar{y}_{r_1} = \frac{\sum_{i=1}^{r_1} y_i}{r_1},\ \bar{y}_k = \frac{\sum_{i=1}^{k} y_i}{k}$$

$$\bar{x}_{r_1} = \frac{\sum_{i=1}^{r_1} x_i}{r_1},\ \bar{x}_k = \frac{\sum_{i=1}^{k} x_i}{k}$$

$$\bar{Z}_1 = \frac{\sum_{i=1}^{N_1} Z_i}{N_1},\ \bar{Z}_2 = \frac{\sum_{i=1}^{N_2} Z_i}{N_2},\ \bar{X}_1 = \frac{\sum_{i=1}^{N_1} X_i}{N_1},\ \bar{X}_2 = \frac{\sum_{i=1}^{N_2} X_i}{N_2},\ \theta = \frac{W_2(h-1)}{n_2},\ \lambda_1 = \frac{1}{n_1} - \frac{1}{N},\ \lambda_2 = \frac{1}{n_2} - \frac{1}{N}$$

$$\lambda_3 = \frac{1}{n_2} - \frac{1}{n_1},\quad S_{y_2}^2 = \frac{\sum_{i=1}^{N_2}(y_i - \bar{Y}_2)^2}{N_2 - 1},\ S_{yx} = \frac{\sum_{i=1}^{N_1}(X_i - \bar{X})(Y_i - \bar{Y})}{N_1 - 1},\ S_{yz} = \frac{\sum_{i=1}^{N_1}(Z_i - \bar{Z})(Y_i - \bar{Y})}{N_1 - 1}$$

$$S_{zx} = \frac{\sum_{i=1}^{N_1}(Z_i - \bar{Z})(X_i - \bar{X})}{N_1 - 1},\beta_{yx} = \frac{S_{yx}}{S_x^2},\beta_{yx(2)} = \frac{S_{yx(2)}}{S_{x(2)}^2},C_y = \frac{S_y}{\bar{Y}},C_{y(2)} = \frac{S_{y(2)}}{\bar{Y}},C_x = \frac{S_x}{\bar{X}},C_{x(2)} = \frac{S_{x(2)}}{\bar{X}},$$

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

$$\left\{\begin{array}{l} \bar{e}_y^* = \bar{y}^* - \bar{Y},\ \bar{e}_{x1} = \bar{x}_1 - \bar{X},\ \bar{e}_x = \bar{x} - \bar{X},\ \bar{e}_{z1} = \bar{z}_1 - \bar{Z},\ \bar{e}_z^* = \bar{z}^* - \bar{Z},\ \bar{e}_z = \bar{z} - \bar{Z}, \bar{e}_x^* = \bar{x}^* - \bar{X}, \\ E(\bar{e}_y^*)E(\bar{e}_{x1})E(\bar{e}_x) = E(\bar{e}_{z1}) = E(\bar{e}_z) = E(\bar{e}_x^*) = E(\bar{e}_z^*) = 0 \end{array}\right\}$$

$$\left\{\begin{array}{l} E(\bar{e}_y^{*2}) = \lambda_2 S_y^2 + \theta S_{y(2)}^2,\ E(\bar{e}_x^{*2}) = \lambda_2 S_x^2 + \theta S_{x(2)}^2,\ E(\bar{e}_z^{*2}) = \lambda_2 S_z^2 + \theta S_{z(2)}^2,\ E(\bar{e}_{x1}^2) = \lambda_1 S_{x1}^2 \\ E(\bar{e}_x^2) = \lambda_2 S_x^2,\ E(\bar{e}_z^2) = \lambda_2 S_z^2,\ E(\bar{e}_y^* \bar{e}_{x1}) = \lambda_1 S_{yx}^2,\ E(\bar{e}_y^* \bar{e}_{z1}) = \lambda_1 S_{yz}^2,\ E(\bar{e}_y^* \bar{e}_x) = \lambda_2 S_{yx}^2, \\ E(\bar{e}_y^* \bar{e}_z) = \lambda_2 S_{yz}^2,\ E(\bar{e}_y^* \bar{e}_x^*) = \lambda_2 S_{yx}^2 + \theta S_{yx(2)}^2,\ E(\bar{e}_y^* \bar{e}_z^*) = \lambda_2 S_{yz}^2 + \theta S_{yz(2)}^2. \end{array}\right\}$$

The mean Square Error of the proposed estimator under Situation-I can be obtained as:

$$T_{n(1)} = k_1(\bar{e}_y^* + \bar{Y})\exp\left(\frac{\bar{e}_{x1} + \bar{X} - \bar{e}_x - \bar{X}}{\bar{e}_{x1} + \bar{X} + (\alpha_1 - 1)(\bar{e}_x + \bar{X})} + \frac{\bar{e}_{z1} + \bar{Z} - \bar{e}_z - \bar{Z}}{\bar{e}_{z1} + \bar{Z} + (\beta_1 - 1)(\bar{e}_z + \bar{Z})}\right),$$

$$T_{n(1)} = k_1(\bar{e}_y^* + \bar{Y})\exp\left(\frac{\bar{e}_{x1} - \bar{e}_x}{\alpha_1 \bar{X}}\left(1 + \frac{\bar{e}_{x1} + (\alpha_1 - 1)\bar{e}_x}{\alpha_1 \bar{X}}\right)^{-1} + \frac{\bar{e}_{z1} - \bar{e}_z}{\beta_1 \bar{Z}}\left(1 + \frac{\bar{e}_{z1} + (\beta_1 - 1)\bar{e}_z}{\beta_1 \bar{Z}}\right)^{-1}\right),$$

$$T_{n(1)} \approx k_1(\bar{e}_y^* + \bar{Y})\left(1 + \frac{\bar{e}_{x1} - \bar{e}_x}{\alpha_1 \bar{X}} - \left(\frac{\bar{e}_{x1} - \bar{e}_x}{\alpha_1 \bar{X}}\right)\frac{\bar{e}_{x1} + (\alpha_1 - 1)\bar{e}_x}{\alpha_1 \bar{X}} + \frac{\bar{e}_{z1} - \bar{e}_z}{\beta_1 \bar{Z}} - \left(\frac{\bar{e}_{z1} - \bar{e}_z}{\beta_1 \bar{Z}}\right)\frac{\bar{e}_{z1} + (\beta_1 - 1)\bar{e}_z}{\beta_1 \bar{Z}} + \frac{1}{2}\left(\frac{\bar{e}_{x1} - \bar{e}_x}{\alpha_1 \bar{X}}\right)^2 + \frac{1}{2}\left(\frac{\bar{e}_{z1} - \bar{e}_z}{\beta_1 \bar{Z}}\right)^2\right),$$

$$E(T_{n(1)} - \bar{Y})^2 » k^2 E(\bar{e}_y^*)^2 + (k-1)^2\bar{Y}^2 + (k\bar{Y})^2\frac{E(\bar{e}_{x1} - \bar{e}_x)^2}{(\alpha\bar{X})^2} + (k\bar{Y})^2\frac{E(\bar{e}_{z1} - \bar{e}_z)^2}{(\beta\bar{Z})^2} + 2k^2\bar{Y}\frac{E\bar{e}_y^*(\bar{e}_{x1} - \bar{e}_x)}{\alpha\bar{X}}$$

$$+2k^2\bar{Y}\frac{E\bar{e}_y^*(\bar{e}_{z1} - \bar{e}_z)}{\beta\bar{Z}} + 2\frac{k^2\bar{Y}^2}{\alpha\beta\bar{X}\bar{Z}}E(\bar{e}_{x1} - \bar{e}_x)(\bar{e}_{z1} - \bar{e}_z),$$

$$MSE(T_{n(1)}) \approx k^2\lambda_2 S_y^2 + k^2\theta S_{y(2)}^2 + (k-1)^2\bar{Y}^2 + \frac{(k\bar{Y})^2}{(\alpha\bar{X})^2}\lambda_1 S_x^2 + \frac{(k\bar{Y})^2}{(\alpha\bar{X})^2}\lambda_2 S_x^2 - 2\frac{(k\bar{Y})^2}{\alpha\bar{X})^2}\lambda_1 S_x^2 + \frac{(k\bar{Y})^2}{(\beta\bar{Z})^2}\lambda_1 S_z^2$$

$$+\frac{(k\bar{Y})^2}{(\beta\bar{Z})^2}\lambda_2 S_z^2 - 2\frac{(k\bar{Y})^2}{(\beta\bar{Z})^2}\lambda_1 S_z^2 + 2\frac{k^2\bar{Y}}{\alpha\bar{X}}\lambda_1 S_{yx} - 2\frac{k^2\bar{Y}}{\alpha\bar{X}}\lambda_2 S_{yx} + 2\frac{k^2\bar{Y}}{\beta\bar{Z}}\lambda_1 S_{yz} - 2\frac{k^2\bar{Y}}{\beta\bar{Z}}\lambda_2 S_{yz} + 2\frac{k^2\bar{Y}^2}{\alpha\beta\bar{X}\bar{Z}}\lambda_1 S_{xz}$$

$$-2\frac{k^2\bar{Y}^2}{\alpha\beta\bar{X}\bar{Z}}\lambda_1 S_{xz} - 2\frac{k^2\bar{Y}^2}{\alpha\beta\bar{X}\bar{Z}}\lambda_1 S_{xz} + 2\frac{k^2\bar{Y}^2}{\alpha\beta\bar{X}\bar{Z}}\lambda_2 S_{xz}.$$

On simplification, we get

$$MSE(T_{n(1)}) \approx k_1^2\lambda_3\bar{Y}^2\left(C_y^2 + \frac{1}{\alpha_1^2}C_x^2 + \frac{1}{\beta_1^2}C_z^2 - \frac{2\rho_{yx}C_yC_x}{\alpha_1^2} - \frac{2\rho_{yz}C_yC_z}{\beta_1^2} + \frac{2\rho_{xz}C_zC_x}{\alpha_1^2\beta_1^2}\right)$$

$$+(k_1^2-1)^2\bar{Y}^2+k_1^2\lambda_1\bar{Y}^2C_y^2+k_1^2\theta\bar{Y}^2C_{y(2)}^2,$$

where $\lambda_3=\frac{1}{n_2}-\frac{1}{n_1}$.

Similarly, in Situation-II, the bias can be derived as:

$$T_{n(2)}-\bar{Y}\approx k_2\bar{e}_y^*+(k_2-1)\bar{Y}+k_2\bar{e}_y^*\frac{(\bar{e}_{x1}-\bar{e}_x^*)}{\alpha_2\bar{X}}+k_2\bar{e}_y^*\frac{(\bar{e}_{z1}-\bar{e}_z^*)}{\beta_2\bar{Z}}+k_2\bar{Y}\frac{(\bar{e}_{x1}-\bar{e}_x^*)}{\alpha\bar{X}}+k_2\bar{Y}\frac{(\bar{e}_{z1}-\bar{e}_z^*)}{\beta_2\bar{Z}},$$

Also,

$$E\left(T_{n(2)}-\bar{Y}\right)^2\approx k_2^2E\left(\bar{e}_y^*\right)^2+(k_2-1)^2\bar{Y}^2+(k_2\bar{Y})^2\frac{E(\bar{e}_{x1}-\bar{e}_x^*)^2}{(\alpha_2\bar{X})^2}+(k_2\bar{Y})^2\frac{E(\bar{e}_{z1}-\bar{e}_z^*)^2}{(\beta_2\bar{Z})^2}+2k_2^2\bar{Y}\frac{E\bar{e}_y^*(\bar{e}_{x1}-\bar{e}_x^*)}{\alpha_2\bar{X}}$$

$$+2k_2^2\bar{Y}\frac{E\bar{e}_y^*(\bar{e}_{z1}-\bar{e}_z^*)}{\beta_2\bar{Z}}+2\frac{k_2^2\bar{Y}^2}{\alpha_2\beta_2\bar{X}\bar{Z}}E(\bar{e}_{x1}-\bar{e}_x^*)(\bar{e}_{z1}-\bar{e}_z^*).$$