A CLASS OF LOGARITHMIC-CUM-EXPONENTIAL ESTIMATORS FOR POPULATION MEAN WITH RISK ANALYSIS USING DOUBLE SAMPLING

Diwakar Shukla[1], Astha Jain[2]

[1,2] Department of Mathematics and Statistics,
Dr. Harisingh Gour Vishwavidyalaya, Sagar, M.P., 470003
[1]diwakarshukla@rediffmail.com, [2]asthjain2597@gmail.com

Abstract

In order to improve upon the efficiency of an estimate in double sampling for estimating population mean of character under study using an auxiliary variable, a part of survey resources are used to collect the information on auxiliary variable. Some authors have suggested exponential-type estimators and some others advocated for log-type estimators. But combination of such is required for specific situation. This paper presents a class of logarithmic-cum-exponential ratio estimators in double sampling setup. The expressions for the mean squared error and bias of the proposed class of estimators are derived for two different cases(sub-sample and independent sample). Sometimes the persons involved in the sample survey have to undergo for risk on life. For example, data collection in naxalites area, working in intense forest, interview during spread of epidemic or data collection in politically disturbed region. Such risk may affect the accuracy, efficiency of estimation. A linear Risk function is used for the proposed class of estimators. Two cases of double sampling are compared in terms of relative efficiency in view to risk aspect.It is found that the proposed class of estimators has a lower mean squared error than the simple mean estimator, usual ratio, usual exponential, usual log estimators in the double sampling setup. In addition, these theoretical results are supported by a numerical example. Risk function based simulated study is performed for the support of findings of the content. Optimal sample sizes under risk are derived and compared under two cases.

Keywords: Exponential estimator, Logarithmic estimator, Mean squared error, Bias, Risk function, Risk Analysis, Survey sampling, Double sampling, Simple random sampling without replacement(SRSWOR).

1. Introduction

In double sampling, some part of the resources available for the survey are used to collect data for auxiliary variable. It is because the population mean of auxiliary variable is assumed unknown. Such are collected through sample at the preliminary level and then used to estimate population mean (or population total).

In recent study on the estimators in the double sampling Sahoo et al.[9] discussed the approach of estimating the population mean using regression-type estimator. It boosted the analytical approach of estimation for dealing with double sampling scheme. Bahal and Tuteja[2] developed exponential-type ratio and product estimator for the SRSWOR setup which later extended by the many authors in verity of other sampling schemes. Shashi Bhushan et al.[5] suggested double sampling ratio type estimator using two auxiliary variables. Authors discussed asymptotic properties of the estimators with bias and mean squared error. Shabbir and Sat Gupta[14] suggested exponential ratio-type estimator for estimating the population mean in the setup of stratified sampling. Such proposal is found to perform better than the usual mean, usual ratio, usual exponential ratio, traditional regression estimators.

Zahoor et. al.[17] suggested regression estimator in double sampling using multi-auxiliary information in the presence of non-response and measurement error in the second phase sample. Such an extension of Azeem[8] who suggested ratio and ratio-cum-exponential estimators in double sampling for population mean incorporating the possibility of non-response and measurement error. The Wu and Luan[6] marked that major advantages of double sampling are the gain in high precision without much substantial increase in cost. Sanaullah et al.[10] suggested generalized exponential-type estimators for the stratified double sampling setup. Sanaullah et al.[12] developed the generalized exponential type estimators for estimating the population variance in double sampling with the help of two auxiliary variables. Zaman and Kadilar[18] proposed exponential ratio-type estimation procedures in the stratified two phase sampling setup. Shukla and Alim[1] proposed parameter estimation approach based an double sampling showing on application in big-data environment. Bhusan and Gupta[3] discussed some log-type estimators using attribute. In another useful contribution Bhushan and Kumar[4] proposed log-type estimators for population mean under the setup of ranked set sampling.

1.1. Risk in data collection

While the conduct of sample survey, using the personal interview method, some areas may be politically disturbed, some may dangerous due to being forest area, some may risky because of naxalites movement and few may under the risk of intense epidemic spread (like Covid-19). Such exposure of risk may possible on the life of field workers involved in data collection. Consider an example where area of a district exposed under risk are identified as A, B, C, D and each having different zones z_1, z_2, z_3, z_4, z_5 with percentage of risk varying over zones.

Table 1: *Risk distribution as per area and zones*

	Zones with risk (r_i)					
Area of District	z_1	z_2	z_3	z_4	z_5	Overhead Risk(r')
A	25%	10%	20%	30%	7%	8%
B	15%	13%	28%	12%	22%	10%
C	35%	14%	5%	25%	10%	11%
D	16%	11%	18%	19%	23%	13%

Risk per units (r_i) belongs to zones and overhead risk r' belong to the geographical areas of a district.

Deriving motivational idea and scientific approach from above contributions, this paper consider the development of new class of estimators under the risk of life of surveyor during data collection using double sampling.

1.2. Symbols used for population

Let a population of finite size N, D be the variable of main interest and A is an auxiliary variable correlated to D. The pair (D_i, A_i), $i = 1, 2, 3, ..., N$ represents population values such that

$$\bar{D} = \frac{1}{N}\sum_{i=1}^{N} D_i, \qquad \bar{A} = \frac{1}{N}\sum_{i=1}^{N} A_i \tag{1.1}$$

$$S_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}(D_i - \bar{D})^2, \qquad S_a^2 = \frac{1}{N-1}\sum_{i=1}^{N}(A_i - \bar{A})^2 \tag{1.2}$$

$$S_{da} = \frac{1}{N-1}\sum_{i=1}^{N}(D_i - \bar{D})(A_i - \bar{A}), \qquad C_{da} = \frac{S_{DA}}{(\bar{D}\bar{A})} \tag{1.3}$$

$$C_d = \frac{S_d}{\bar{D}}, \qquad C_a = \frac{S_a}{\bar{A}}, \qquad \rho = \frac{S_{da}}{S_d S_a}, \qquad M = \rho\frac{C_d}{C_a} \tag{1.4}$$

where C_d and C_a denote coefficient of variations, ρ correlation coefficient.

1.3. Notations in SRSWOR Setup:

Assumed that information about variable of main interest D is not available, so a simple random sampling is used, using sample of size n$(n < N)$, to predict about that. Further, in usual practice such assumes population mean of auxiliary variable $\bar{A}$ available. All possible samples are $\binom{N}{n}$.

Figure 1: *Population and Sample*

Let values of random sample by SRSWOR are $(d_i, a_i), i = 1, 2, 3, ..., n$ then one can define sample statistics as:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{N} d_i, \qquad \bar{a} = \frac{1}{n}\sum_{i=1}^{N} a_i \tag{1.5}$$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2, \qquad s_a^2 = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a})^2 \tag{1.6}$$

$$s_{da} = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})(a_i - \bar{a}), \qquad \hat{M} = \frac{s_{da}}{s_a^2} \tag{1.7}$$

1.4. Some usual estimators in SRSWOR

(a) Usual Ratio Estimator: $\hat{\bar{D}}_R = \frac{\bar{d}}{\bar{a}}\bar{A}$

(b) Usual Product Estimator: $\hat{\bar{D}}_P = \frac{\bar{d}\bar{a}}{\bar{A}}$

(c) Usual Regression Estimator: $\hat{\bar{D}}_{Re} = \bar{d} + \hat{M}(\bar{A} - \bar{a})$

(d) Usual Log Estimator:$\hat{\bar{D}}_L = \bar{d}\left[1 + \log\left(\frac{\bar{A}}{\bar{a}}\right)\right]$

(e) Usual Exponential Estimator:$\hat{\bar{D}}_{Ex} = \bar{d}\left[exp\left(\frac{\bar{A} - \bar{a}}{\bar{A} + \bar{a}}\right)\right]$

Some useful symbols are:

$$V_{qs} = \frac{E\{(\bar{d} - \bar{D})^q(\bar{a} - \bar{A})^s\}}{\bar{D}^q\bar{A}^s}, \qquad V'_{qs} = \frac{E\{(\bar{d} - \bar{D})^q(\bar{a}' - \bar{A})^s\}}{\bar{D}^q\bar{A}^s}; q, s = 0, 1, 2$$

$$V_{20} = \left(\frac{1}{n} - \frac{1}{N}\right) C_d^2, \qquad V_{02} = \left(\frac{1}{n} - \frac{1}{N}\right) C_a^2, \qquad V'_{02} = \left(\frac{1}{n'} - \frac{1}{N}\right) C_a^2$$
$$V_{11} = \left(\frac{1}{n} - \frac{1}{N}\right) \rho C_d C_a, \qquad V'_{11} = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho C_d C_a$$

Symbols have their usual meaning as adopted by the survey practitioners in the concerned literature. The Bias Bias(·) and Mean Squared Error MSE(·) of above existing estimators under SRSWOR are expressed as under:

$$Bias(\hat{\bar{D}}_R) = \bar{D}\left[V_{02} - V_{11}\right], \qquad MSE(\hat{\bar{D}}_R) = \bar{D}^2\left[V_{20} - 2V_{11} + V_{02}\right] \tag{1.8}$$
$$Bias(\hat{\bar{D}}_P) = \bar{D}\left[V_{02} + V_{11}\right], \qquad MSE(\hat{\bar{D}}_P) = \bar{D}^2\left[V_{20} + 2V_{11} + V_{02}\right] \tag{1.9}$$
$$Bias(\hat{\bar{D}}_{Re}) = \bar{D}\left[V_{02} - \hat{\beta}V_{11}\right], \qquad MSE(\hat{\bar{D}}_{Re}) = \bar{D}^2\left[V_{20} - 2\hat{\beta}V_{11} + \hat{\beta}^2 V_{02}\right] \tag{1.10}$$
$$Bias(\hat{\bar{D}}_L) = \bar{D}\left[V_{02} - V_{11}\right], \qquad MSE(\hat{\bar{D}}_L) = \bar{D}^2\left[V_{20} - 2V_{11} + V_{02}\right] \tag{1.11}$$
$$Bias(\hat{\bar{D}}_{Ex}) = \bar{D}\left[\frac{3}{8}V_{02} - \frac{1}{2}V_{11}\right], \qquad MSE(\hat{\bar{D}}_{Ex}) = \bar{D}^2\left[V_{20} + \frac{1}{4}V_{02} - V_{11}\right] \tag{1.12}$$

2. Double Sampling Approach

When the information about population mean of variable is not available then during sample survey with the extra risk and efforts, the sample could be obtained using two different strategies. Assume n' be the size of first sample with values $(a'_1, a'_2, ..., a'_{n'})$ and $\bar{a}' = \frac{1}{n'}\sum_{i=1} a'_i$

- **Case I**: When the second-phase sample of size n is a sub-sample of the first-phase sample of size n'

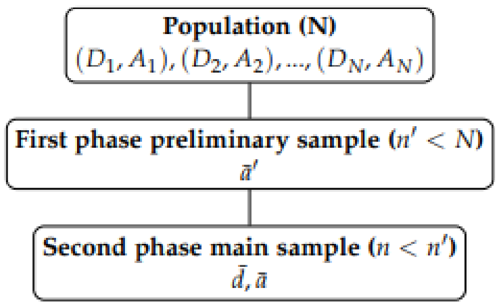


Figure 2: *Sampling strategy under case I*

- **Case II**: When the second-phase sample of size n is drawn independently of the first-phase sample of size n'.

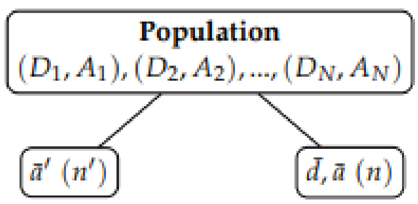


Figure 3: *Sampling strategy under case II*

2.1. Some existing estimators in double sampling

In Double sampling setup, the existing estimators with their respective bias $Bias(\cdot)_I$,$Bias(\cdot)_{II}$ and mean squared error $MSE(\cdot)_I$ & $MSE(\cdot)_{II}$ under case I and case II are as below.

(a) *Simple Random sample mean estimator*:

$$\hat{\hat{D}} = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{2.1}$$

$$V(\hat{\hat{D}}) = \bar{D}^2 V_{20} \tag{2.2}$$

where V(·) denotes variance of estimators.

(b) *Usual Ratio Estimator*:

$$\hat{\hat{D}}_{Rd} = \bar{d}\left(\frac{\bar{a}'}{\bar{a}}\right) \tag{2.3}$$

$$Bias(\hat{\hat{D}}_{Rd})_I = \bar{D}[(V_{02} - V'_{02}) - (V_{11} - V'_{11})] \tag{2.4}$$

$$Bias(\hat{\hat{D}}_{Rd})_{II} = \bar{D}[(V_{02} + V'_{02}) - V_{11}] \tag{2.5}$$

$$MSE(\hat{\hat{D}}_{Rd})_I = \bar{D}^2[V_{20} + (V_{02} - V'_{02}) - 2(V_{11} - V'_{11})] \tag{2.6}$$

$$MSE(\hat{\hat{D}}_{Rd})_{II} = \bar{D}^2[V_{20} + (V_{02} + V'_{02}) - 2V_{11}] \tag{2.7}$$

(c) *Usual Exponential Ratio Estimator*:

$$\hat{\hat{D}}_{Exd} = \bar{d}\exp\left(\frac{\bar{a}' - \bar{a}}{\bar{a}' + \bar{a}}\right) \tag{2.8}$$

$$Bias(\hat{\hat{D}}_{Exd})_I = \bar{D}[\frac{3}{8}(V_{02} - V'_{02}) - \frac{1}{2}(V_{11} - V'_{11})] \tag{2.9}$$

$$Bias(\hat{\hat{D}}_{Exd})_{II} = \bar{D}[\frac{1}{8}(3V_{02} - V'_{02}) - \frac{1}{2}V_{11}] \tag{2.10}$$

$$MSE(\hat{\hat{D}}_{Exd})_I = \bar{D}^2[V_{20} + \frac{1}{4}(V_{02} - V'_{02}) - (V_{11} - V'_{11})] \tag{2.11}$$

$$MSE(\hat{\hat{D}}_{Exd})_{II} = \bar{D}^2[V_{20} + \frac{1}{4}(V_{02} + V'_{02}) - V_{11}] \tag{2.12}$$

(d) *Usual Log Ratio Estimator*:

$$\hat{\hat{D}}_{Lod} = \bar{d}\left[1 + \log\left(\frac{\bar{a}'}{\bar{a}}\right)\right] \tag{2.13}$$

$$Bias(\hat{\hat{D}}_{Lod})_I = \bar{D}[2(V_{02} - V'_{02}) - (V_{11} - V'_{11})] \tag{2.14}$$

$$Bias(\hat{\hat{D}}_{Lod})_{II} = \bar{D}[2V_{02} + V'_{02} - V_{11}] \tag{2.15}$$

$$MSE(\hat{\hat{D}}_{Lod})_I = \bar{D}^2[V_{20} + (V_{02} - V'_{02}) - 2(V_{11} - V'_{11})] \tag{2.16}$$

$$MSE(\hat{\hat{D}}_{Lod})_{II} = \bar{D}^2[V_{20} + (V_{02} + V'_{02}) - 2V_{11}] \tag{2.17}$$

(e) *Usual Regression Estimators*:

$$\hat{\hat{D}}_{Red} = \bar{d} + \hat{M}(\bar{a}' - \bar{a}) \tag{2.18}$$

$$Bias(\hat{\hat{D}}_{Red})_I = \bar{D}[(V_{02} - V'_{02}) - (V_{11} - V'_{11})] \tag{2.19}$$

$$Bias(\hat{\hat{D}}_{Red})_{II} = \bar{D}[V_{02} + V'_{02} - V_{11}] \tag{2.20}$$

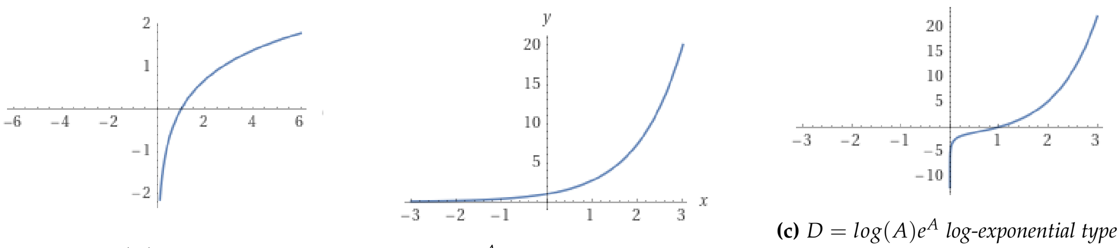
$$MSE(\hat{\hat{D}}_{Red})_I = \bar{Y}^2[V_{20} + \hat{M}^2(V_{02} - V'_{02}) - 2\hat{M}(V_{11} - V'_{11})] \tag{2.21}$$

$$MSE(\hat{\hat{D}}_{Red})_{II} = \bar{Y}^2[V_{20} + \hat{M}^2(V_{02} + V'_{02}) - 2\hat{M}V_{11}] \tag{2.22}$$

where $\hat{M}$ is the regression coefficient.

2.2. Motivation

Estimators suggested in simple random sampling, double sampling, stratified sampling may usual type or exponential type or log-type. Sometime the data may follow the pattern different that of exponential or log-type. It may be a mixture of log and exponential type (Fig4c). This motivates to look for a new combined class of log-cum-exponential type estimators. This paper considers the same in the setup of double sampling. Several authors have suggested estimators



(a) $D = log(A)$ *Log type graph*

(b) $D = e^A$ *Exponential type graph*

(c) $D = log(A)e^A$ *log-exponential type graph*

Figure 4: *Graphical pattern of relationship*

for relationship between D and A variables as shown in Fig(4a) and Fig (4b). But for relationship of type as in Fig (4c) yet needs to be explored. This paper is focused on proposing estimation methodologies with respect to mutual relation shown in fig 4c under the double sampling setup.

3. Proposed class of Logarithmic-Exponential Type Estimators

A family of estimators under the double sampling is proposed, to estimate the unknown population mean of the study variable D assuming the presence of auxiliary information A:

$$\hat{\bar{D}}_{LEd} = \bar{d}\left[\exp\left\{\left(1-\left(\frac{\bar{a}'}{\bar{a}}\right)^{\alpha}\right)\left(1+\log\left(\frac{\bar{a}'}{\bar{a}}\right)^{\beta}\right)\right\}\right] \tag{3.1}$$

assuming expo-log type relationship between D and A(fig4c), where α, β are constants may positive or negative real numbers.

Theorem 1. The bias of the proposed class of estimator for the sub-sample(Case I) and independent sample(Case II) respectively are:

$$Bias(\hat{\bar{D}}_{LEd})_I = \alpha\bar{D}((V_{11}-V'_{11})-\beta(V_{02}-V'_{02})) \tag{3.2}$$

$$Bias(\hat{\bar{D}}_{LEd})_{II} = \alpha\bar{D}(V_{11}-\beta(V_{02}+V'_{02})) \tag{3.3}$$

where $Bias(\cdot)_I$, $Bias(\cdot)_{II}$ are for case I and case II strategies respectively.

Proof. For large sample approximation, define some quantities $\epsilon_0, \epsilon_1, \epsilon_2$ with $|\epsilon_0| < 1, |\epsilon_1| < 1, |\epsilon_2| < 1$ such that

$$\bar{d} = \bar{D}(1+\epsilon_0), \qquad \bar{a} = \bar{A}(1+\epsilon_1), \qquad \bar{a}' = \bar{A}(1+\epsilon_2)$$

where $\bar{a}' = \frac{1}{n'}(\sum_{i=1}^{n'} a'_i)$ and $(a'_1, a'_2, ..., a'_n)$ is first phase sample of size n'.

$$E(\epsilon_0) = E(\epsilon_1) = E(\epsilon_2) = 0$$

Moreover,

$$E(\epsilon_0^2) = \left(\frac{1}{n}-\frac{1}{N}\right)C_d^2, \qquad E(\epsilon_1^2) = \left(\frac{1}{n}-\frac{1}{N}\right)C_a^2, \qquad E(\epsilon_2^2) = \left(\frac{1}{n'}-\frac{1}{N}\right)C_a^2$$

$$E(\epsilon_0\epsilon_1) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho C_d C_a, \quad E(\epsilon_0\epsilon_2) = \left(\frac{1}{n'} - \frac{1}{N}\right)\rho C_d C_a, \quad E(\epsilon_1\epsilon_2) = \left(\frac{1}{n'} - \frac{1}{N}\right) C_a^2$$

General expression for bias for $\hat{D}_{LEd}$ is

$$Bias(\hat{\bar{D}}_{LEd}) = [E(\hat{\bar{D}}_{LEd}) - \bar{D}]$$

Under large sampling approximation, upto first order ,

$$\hat{\bar{D}}_{LEd} = \bar{D}(1+\epsilon_0)\left[\exp\left\{(1-(1+\epsilon_2)^{\alpha}(1+\epsilon_1)^{-\alpha})(1+\beta\log(1+\epsilon_2)(1+\epsilon_1)^{-1})\right\}\right]$$

Since $|\epsilon_0| < 1, |\epsilon_1| < 1$ and $|\epsilon_2| < 1$, using Taylor series expansion upto the first order approximation, ignoring terms of higher order $(\epsilon_0^i, \epsilon_1^j, \epsilon_2^k)$ for $i > 2, j > 2, k > 2, (i+j+k) > 2$,

$$\hat{\bar{D}}_{LEd} = \bar{D}\left[1+\epsilon_0+\alpha(\epsilon_1-\epsilon_2)+\alpha\epsilon_0(\epsilon_1-\epsilon_2)+\beta\alpha(\epsilon_1^2+\epsilon_2^2-2\epsilon_1\epsilon_2)\right]$$

Using expectation E(ϵ_0)=E(ϵ_1)=E(ϵ_2)=0, which leads to bias of proposed class of estimator,

$$Bias(\hat{\bar{D}}_{LEd})_I = \alpha\bar{D}[(V_{11}-V'_{11})-\beta(V_{02}-V'_{02})] \tag{3.4}$$

$$Bias(\hat{\bar{D}}_{LEd})_{II} = \alpha\bar{D}[V_{11}-\beta(V_{02}+V'_{02})] \tag{3.5}$$

Since E($\epsilon_0\epsilon_2$)=V'_{11}=0 for case II because of sample n' being independent to n. ■

Theorem 2. The mean squared error of the proposed class of estimator for the sub-sample(Case I) and independent sample(Case II) respectively are

$$\text{MSE}(\hat{\bar{D}}_{LEd})_I = \bar{D}^2\left[V_{20}+2\alpha(V_{11}-V'_{11})+\alpha^2(V_{02}-V'_{02})\right] \tag{3.6}$$

$$\text{MSE}(\hat{\bar{D}}_{LEd})_{II} = \bar{D}^2\left[V_{20}+2\alpha V_{11}+\alpha^2(V_{02}+V'_{02})\right] \tag{3.7}$$

Proof. The proposed class in double sampling is,

$$\hat{\bar{D}}_{LEd} = \bar{d}\left[\exp\left\{\left(1-\left(\frac{\bar{a}'}{\bar{a}}\right)^{\alpha}\right)\left(1+\log\left(\frac{\bar{a}'}{\bar{a}}\right)^{\beta}\right)\right\}\right]$$

and above in terms of large sample approximation is,

$$\hat{\bar{D}}_{LEd} = \bar{d}\left[\exp\left\{(1-(1+\epsilon_2)^{\alpha}(1+\epsilon_1)^{-\alpha})(1+\beta\log(1+\epsilon_2)(1+\epsilon_1)^{-1})\right\}\right]$$

Using $|\epsilon_0| < 1, |\epsilon_1| < 1$ and $|\epsilon_2| < 1$ and Taylor series expansion upto the first order of approximation, one can get

$$\hat{\bar{D}}_{LEd} = \bar{D}\left[1+\epsilon_0+\alpha(\epsilon_1-\epsilon_2)\right]$$

by ignoring terms of higher order $(\epsilon_0^i, \epsilon_1^j, \epsilon_2^k)$ for $i > 1, j > 1, k > 1, (i+j+k) > 1$, i,j,k=0,1,2... Subtracting $\bar{D}$ and squaring both sides one can get,

$$(\hat{\bar{D}}_{LEd}-\bar{D})^2 = \bar{D}^2\left[\epsilon_0^2+2\alpha\epsilon_0(\epsilon_1-\epsilon_2)+\alpha^2(\epsilon_1-\epsilon_2)^2\right]$$

By taking expectation both sides,

$$E(\hat{\bar{D}}_{LEd}-\bar{D})^2 = \bar{D}^2 E\left[\epsilon_0^2+2\alpha\epsilon_0(\epsilon_1-\epsilon_2)+\alpha^2(\epsilon_1-\epsilon_2)^2\right]$$

So the mean squared error is for Case I and Case II are:

$$\text{MSE}(\hat{\bar{D}}_{LEd}))_I = \bar{D}^2\left[V_{20}+2\alpha(V_{11}-V'_{11})+\alpha^2(V_{02}-V'_{02})\right] \tag{3.8}$$

and

$$\text{MSE}(\hat{\hat{D}}_{LEd}))_{II} \quad = \quad \bar{D}^2\left[V_{20} + 2\alpha V_{11} + \alpha^2(V_{02} + V'_{02})\right] \tag{3.9}$$

Since E($\epsilon_0\epsilon_2$) = V'_{11} = 0 for case II. ∎

Remark 1: Gain in precision under case I and case II

$$\left[\text{MSE}(\hat{\hat{D}}_{LEd})_I - \text{MSE}(\hat{\hat{D}}_{LEd})_{II}\right] = -2\bar{D}^2(\alpha^2 V'_{02} + \alpha V'_{11}) \tag{3.10}$$

The gain in precision depends on the sign of V'_{11}. In general, case I is better, but if ($\alpha V'_{02} < V_{11}$) then case II of double sampling is better than case I. It provides range when 0< $\alpha < \left(\frac{V_{11}}{V'_{02}}\right)$ then case II is more efficient than case I.

Remark 2: Some particular estimators in the proposed class are in table6:

Table 2: *Estimators as member of proposed class.*

Estimators	α	β
$\hat{\hat{D}}_1 = \bar{d}\left[\exp\left\{\left(1-\left(\frac{\bar{a}}{\bar{a}'}\right)\right)\left(1+\log\left(\frac{\bar{a}}{\bar{a}'}\right)\right)\right\}\right]$	-1	-1
$\hat{\hat{D}}_2 = \bar{d}\left[\exp\left\{\left(1-\left(\frac{\bar{a}}{\bar{a}'}\right)\right)\right\}\right]$	-1	0
$\hat{\hat{D}}_3 = \bar{d}\left[\exp\left\{\left(1-\left(\frac{\bar{a}}{\bar{a}'}\right)\right)\left(1+\log\left(\frac{\bar{a}'}{\bar{a}}\right)\right)\right\}\right]$	-1	1
$\hat{\hat{D}}_4 = \bar{d}$	0	-1
$\hat{\hat{D}}_5 = \bar{d}$	0	0
$\hat{\hat{D}}_6 = \bar{d}$	0	1
$\hat{\hat{D}}_7 = \bar{d}\left[\exp\left\{\left(1-\left(\frac{\bar{a}'}{\bar{a}}\right)\right)\left(1+\log\left(\frac{\bar{a}}{\bar{a}'}\right)\right)\right\}\right]$	1	-1
$\hat{\hat{D}}_8 = \bar{d}\left[\exp\left\{\left(1-\left(\frac{\bar{a}'}{\bar{a}}\right)\right)\right\}\right]$	1	0
$\hat{\hat{D}}_9 = \bar{d}\left[\exp\left\{\left(1-\left(\frac{\bar{a}'}{\bar{a}}\right)\right)\left(1+\log\left(\frac{\bar{a}'}{\bar{a}}\right)\right)\right\}\right]$	1	1

Table 3: *Mean Squared Error of Estimators under case I as members of proposed class*

Mean Squared Error	α	β
$MSE(\hat{\hat{D}}_1)_I = \bar{D}^2\left[V_{20} - 2(V_{11} - V'_{11}) + (V_{02} - V'_{02})\right]$	-1	-1
$MSE(\hat{\hat{D}}_2)_I = \bar{D}^2\left[V_{20} - 2(V_{11} - V'_{11}) + (V_{02} - V'_{02})\right]$	-1	0
$MSE(\hat{\hat{D}}_3)_I = \bar{D}^2\left[V_{20} - 2(V_{11} - V'_{11}) + (V_{02} - V'_{02})\right]$	-1	1
$V(\hat{\hat{D}}_4) = \bar{D}^2 V_{20}$	0	-1
$V(\hat{\hat{D}}_5) = \bar{D}^2 V_{20}$	0	0
$V(\hat{\hat{D}}_6) = \bar{D}^2 V_{20}$	0	1
$MSE(\hat{\hat{D}}_7)_I = \bar{D}^2\left[V_{20} + 2(V_{11} - V'_{11}) + (V_{02} - V'_{02})\right]$	1	-1
$MSE(\hat{\hat{D}}_8)_I = \bar{D}^2\left[V_{20} + 2(V_{11} - V'_{11}) + (V_{02} - V'_{02})\right]$	1	0
$MSE(\hat{\hat{D}}_9)_I = \bar{D}^2\left[V_{20} + 2(V_{11} - V'_{11}) + (V_{02} - V'_{02})\right]$	1	1

3.1. Optimal sub-class of estimators

Differentiating MSE($\cdot$) with respect to α, one can obtain optimum value of α as

Case I

$$\hat{\alpha} = \frac{(V_{11} - V'_{11})}{(V_{02} - V'_{02})} = \left(-\rho\frac{C_d}{C_a}\right) = (-M) \tag{3.11}$$

Table 4: *Mean Squared Error of Estimators under case II as members of proposed class*

Mean Squared Error	α	β
$MSE(\hat{\hat{D}}_1)_{II} = \bar{D}^2\left[V_{20} - 2V_{11} + (V_{02} + V'_{02})\right]$	-1	-1
$MSE(\hat{\hat{D}}_2)_{II} = \bar{D}^2\left[V_{20} - 2V_{11} + (V_{02} + V'_{02})\right]$	-1	0
$MSE(\hat{\hat{D}}_3)_{II} = \bar{D}^2\left[V_{20} - 2V_{11} + (V_{02} + V'_{02})\right]$	-1	1
$V(\hat{\hat{D}}_4) = \bar{D}^2 V_{20}$	0	-1
$V(\hat{\hat{D}}_5) = \bar{D}^2 V_{20}$	0	0
$V(\hat{\hat{D}}_6) = \bar{D}^2 V_{20}$	0	1
$MSE(\hat{\hat{D}}_7)_{II} = \bar{D}^2\left[V_{20} + 2V_{11} + (V_{02} + V'_{02})\right]$	1	-1
$MSE(\hat{\hat{D}}_8)_{II} = \bar{D}^2\left[V_{20} + 2V_{11} + (V_{02} + V'_{02})\right]$	1	0
$MSE(\hat{\hat{D}}_9)_{II} = \bar{D}^2\left[V_{20} + 2V_{11} + (V_{02} + V'_{02})\right]$	1	1

Table 5: *Bias of Estimators under case I as members of proposed class*

Bias	α	β
$Bias(\hat{\hat{D}}_1)_I = -\bar{D}\left[(V_{11} - V'_{11}) + (V_{02} - V'_{02})\right]$	-1	-1
$Bias(\hat{\hat{D}}_2)_I = -\bar{D}(V_{11} - V'_{11})$	-1	0
$Bias(\hat{\hat{D}}_3)_I = -\bar{D}\left[(V_{11} - V'_{11}) - (V_{02} - V'_{02})\right]$	-1	1
$Bias(\hat{\hat{D}}_4) = 0$	0	-1
$Bias(\hat{\hat{D}}_5) = 0$	0	0
$Bias(\hat{\hat{D}}_6) = 0$	0	1
$Bias(\hat{\hat{D}}_7)_I = \bar{D}\left[(V_{11} - V'_{11}) - (V_{02} - V'_{02})\right]$	1	-1
$Bias(\hat{\hat{D}}_8)_I = \bar{D}(V_{11} - V'_{11})$	1	0
$Bias(\hat{\hat{D}}_9)_I = \bar{D}\left[(V_{11} - V'_{11}) + (V_{02} - V'_{02})\right]$	1	1

Table 6: *Bias of Estimators under case II as members of proposed class*

Bias	α	β
$Bias(\hat{\hat{D}}_1)_{II} = -\bar{D}\left[V_{11} + (V_{02} + V'_{02})\right]$	-1	-1
$Bias(\hat{\hat{D}}_2)_{II} = -\bar{D}\left[V_{11}\right]$	-1	0
$Bias(\hat{\hat{D}}_3)_{II} = -\bar{D}\left[V_{11} - (V_{02} + V'_{02})\right]$	-1	1
$Bias(\hat{\hat{D}}_4) = 0$	0	-1
$Bias(\hat{\hat{D}}_5) = 0$	0	0
$Bias(\hat{\hat{D}}_6) = 0$	0	1
$Bias(\hat{\hat{D}}_7)_{II} = \bar{D}\left[V_{11} - (V_{02} + V'_{02})\right]$	1	-1
$Bias(\hat{\hat{D}}_8)_{II} = \bar{D}V_{11}$	1	0
$Bias(\hat{\hat{D}}_9)_{II} = \bar{D}\left[V_{11} - (V_{02} + V'_{02})\right]$	1	1

Case II

$$\hat{\alpha} = \left[\frac{V_{11}}{V_{02} + V'_{02}}\right] = -\left[\frac{1}{(1+\delta)}\left(\rho\frac{C_d}{C_a}\right)\right] = -\left[\frac{M}{(1+\delta)}\right] \tag{3.12}$$

where, $\delta = \dfrac{\left(\frac{1}{n'} - \frac{1}{N}\right)}{\left(\frac{1}{n} - \frac{1}{N}\right)}$

The mean squared error under the optimum value of $\alpha = \hat{\alpha}$ [as per (3.8), (3.9)] are

Case I

$$[MSE(\hat{\hat{D}}_{LEd})_I]_{opt} = \bar{D}^2 C_d^2 \left\{\left(\frac{1}{n} - \frac{1}{N}\right) - \left(\frac{1}{n} - \frac{1}{n'}\right)\rho^2\right\} \tag{3.13}$$

Case II

$$[MSE(\hat{\bar{D}}_{LEd})_{II}]_{opt} = \bar{D}^2 C_d^2 \left\{ \left(\frac{1}{n} - \frac{1}{N}\right) - \left(\frac{1}{n} - \frac{1}{n'}\right)\left(\frac{\rho^2}{1+\delta}\right) \right\} \tag{3.14}$$

4. Comparison with existing estimators

The existing estimators will be less efficient to the proposed estimators for case I and case II respectively under the following conditions:

(1) ***Simple random sample mean estimator ($\bar{d}$):***

$$\textbf{Case I: } \alpha \leq \frac{-2(V_{11} - V'_{11})}{(V_{02} - V'_{02})}, \qquad \textbf{Case II: } \alpha \leq \frac{-2V_{11}}{(V_{02} + V'_{02})}$$

(2) ***Usual Ratio Estimator*** $(\hat{\bar{D}}_{Rd})$**[eq(2.3)]**

$$\textbf{Case I: } \alpha \leq \left[1 - 2\left(\rho \frac{C_d}{C_a}\right)\right], \qquad \textbf{Case II: } \alpha \leq \left[1 - \frac{2}{(1+\delta)}\left(\rho \frac{C_d}{C_a}\right)\right]$$

(3) ***Usual Exponential Ratio estimator*** $(\hat{\bar{D}}_{Ed})$**[eq(2.8)]**

$$\textbf{Case I: } \alpha \leq \frac{1}{2}\left[1 - 4\rho \frac{C_d}{C_a}\right], \qquad \textbf{Case II: } \alpha \leq \frac{1}{2}\left[1 - \frac{4}{(1+\delta)}\left(\rho \frac{C_d}{C_a}\right)\right]$$

(4) ***Usual Log Ratio Estimator*** $(\hat{\bar{D}}_{Ld})$**[eq(2.13)]**

$$\textbf{Case I: } \alpha \leq \left[1 - 2\left(\rho \frac{C_d}{C_a}\right)\right], \qquad \textbf{Case II: } \alpha \leq \left[1 - \frac{2}{(1+\delta)}\left(\rho \frac{C_d}{C_a}\right)\right]$$

(5) ***Usual Regression Estimator*** $(\hat{\bar{D}}_{Red})$**[eq(2.18)]**

$$\textbf{Case I: } \alpha \leq -2\left(\rho \frac{C_d}{C_a}\right), \qquad \textbf{Case II: } \alpha \leq \left(1 - \frac{2}{(1+\delta)}\right)\left(\rho \frac{C_d}{C_a}\right)$$

5. Risk function and the Proposed estimator

The risk in data collection for dangerous area while implementing a sampling procedure is defined as

(a) Total Risk

(b) Per unit respondent contact risk (infection, injury, life risk)

(c) General risk (area dependent risk)

Risk is associated to various ground conditions like risk in hilly area during data collection, risk of reaching to the household, risk of non-response, risk of dangerous situations, risk of attack on the life of surveyor, risk of epidemic etc.
Let us use symbols for risk as:
r': Overhead risk
r_0 : Total risk
r_1 : Risk per unit for information collection on variable D and A using second sample n.
r_2 : Risk per unit for first sample for collecting information on auxiliary variable A.

Linear risk function for collecting information is:

$$r_0 = r' + r_1 n + r_2 n'$$

It is matter of interest to determine the n and n' for a given risk r_0 at the situation when MSE of $\hat{\hat{D}}_{LEd}$ is minimum. To minimize risk function under risk constraint ϕ and optimum MSE, one can get,

Case I

$$\phi \quad = \quad [MSE(\hat{\hat{D}}_{LEd})_I]_{opt} + \lambda(r' + r_1 n + r_2 n' - r_0)$$

where λ is a Lagrange's multiplier. Differentiating with respect to n and n', equating it to zero, the optimum values of n and n' are

$$n_{opt} = \frac{(r_0 - r')\sqrt{r_1 R}}{r_1 M_1}, \qquad n'_{opt} = \frac{(r_0 - r')\sqrt{-r_2(R - C_d^2)}}{r_2 M_1} \tag{5.1}$$

where

$$M_1 = [\sqrt{r_1 R} + \sqrt{-r_2(R - C_d^2)}], \qquad R = [C_d^2 + 2\alpha C_{da} + \alpha^2 C_a^2]$$

Case II

$$\phi \quad = \quad [MSE(\hat{\hat{D}}_{LEd})_{II}]_{opt} + \lambda(r' + r_1 n + r_2 n' - r_0)$$

where λ is a Lagrange's multiplier. Now differentiating with respect to n and n', equating it to zero, the optimum values of n and n' under case II are

$$n_{opt} = \frac{(r_0 - r')\sqrt{r_1 R}}{r_1 M_2}, \qquad n'_{opt} = \frac{(r_0 - r')\alpha C_a \sqrt{r_2}}{r_2 M_2} \tag{5.2}$$

where

$$M_2 = [\sqrt{r_1 R} + \sqrt{r_2(\alpha^2 C_a^2)}], \qquad R = [C_d^2 + 2\alpha C_{da} + \alpha^2 C_a^2]$$

The ratio of optimal selection of n and n' under fixed risk c_0 is
Case I

$$\left(\frac{n_{opt}}{n'_{opt}}\right) = \frac{r_2(\sqrt{r_1 R})}{r_1(\sqrt{-r_2(R - C_d^2)})}$$

Case II

$$\left(\frac{n_{opt}}{n'_{opt}}\right) = \frac{r_2(\sqrt{r_1 R})}{r_1 \alpha C_a \sqrt{r_2}}$$

6. Empirical risk based Study

Consider a positively correlated population with two variables D and A(Data source -6th Minor Irrigation Census - Village Schedule - Assam)[19] with N=100.
The values of variable D and A are shown in Table 7, where A represents geographical area and D represents the net shown area in hectares.

Table 7: *Population Undertaken.*

D_i	152	98	75	68	60	295	72	125	16	260
A_i	165	111	80	79	78	319	86	189	26	380
D_i	62	95	210	95	175	180	100	37	87	96
A_i	74	123	220	123	185	197	120	48	105	109
D_i	80	148	85	98	38	95	200	84	18	38
A_i	110	158	121	108	40	110	350	95	28	46
D_i	53	69	30	55	29	75	78	48	81	75
A_i	71	81	45	63	45	89	110	59	95	92
D_i	103	97	82	25	76	70	57	182	55	85
A_i	113	105	96	35	94	81	70	192	65	122
D_i	70	24	190	53	190	158	80	93	176	81
A_i	75	34	200	67	232	169	100	103	186	89

Moreover, population parameters are in the Table 8.

Table 8: *Population Parameters*

$\bar{D} = 135$	$S_d^2 = 82327$	$C_d^2 = 4.534$	$S_{ad} = 96274.91$
$\bar{A} = 161$	$S_a^2 = 113076.5$	$C_a^2 = 4.356$	$C_{ad} = 4.43$

Table 9: *PREs of different estimators with respect to proposed estimator in double sampling*

	PRE	
Estimators	Case I	Case II
Simple Random sampling ($\hat{\hat{D}}$)	22.13%	56.083%
Ratio Estimator ($\hat{\hat{D}}_{Rd}$)	0.009%	41.493%
Exponential ratio estimator ($\hat{\hat{D}}_{Exd}$)	6.856%	2.011%
Log ratio Estimator ($\hat{\hat{D}}_{Lod}$)	0.009%	41.493%

where PRE is Percentage Relative Efficiency defined as:

$$(PRE)_{I,II} = \frac{MSE(T)_{I,II} - (MSE(\hat{\hat{D}}_{LEd})_{I,II})_{opt}}{MSE(T)} \times 100 \tag{6.1}$$

and T represents estiamtors like usual ratio, usual expo-ratio, usual log- ratio estimators.
It is observed that in case I, at the α_{opt}, the proposed is 22.13% efficient over sample mean estimator, 6.85% better over exponential ratio estimator and same to the usual ratio usual log ratio estimator. Moreover, in case II, at value α_{opt}, the proposed is 56% efficient to sample mean estimator, 41.4% efficient over ratio estimator, 2% efficient over to exponential estimator and 41.4% over log-ratio estimator.

In Figure 5, while general variation of α values, the case I bears lower MSE then case II. But while reaching to α_{opt}, both cases achieve the same MSE level equivalent to that of Regression estimator in double sampling.
Figure 6, reveals the variation of total risk r_0 over the optimum sample sizes (n_{opt} & n'_{opt}). It is observed that increasing fixed risk r_0 leads to larger n'_{opt} (first sample) in comparison to second sample optimum n_{opt}. Low level risk indicates for equal(but small) n and n' to be used by the survey practitioners.

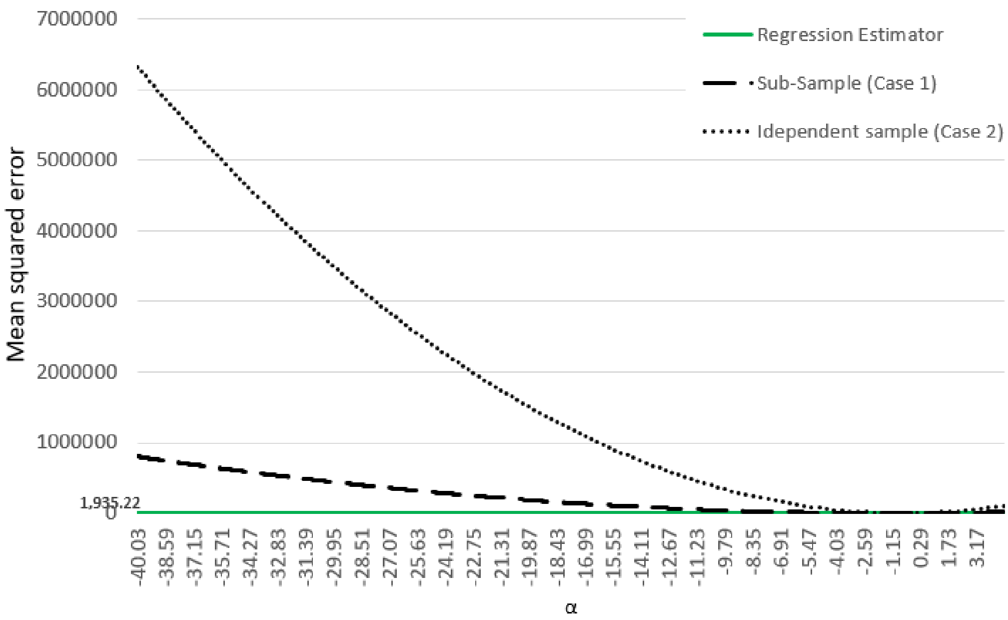


Figure 5: *Comparison between MSE's of the proposed class under case I and case II over variation of α*

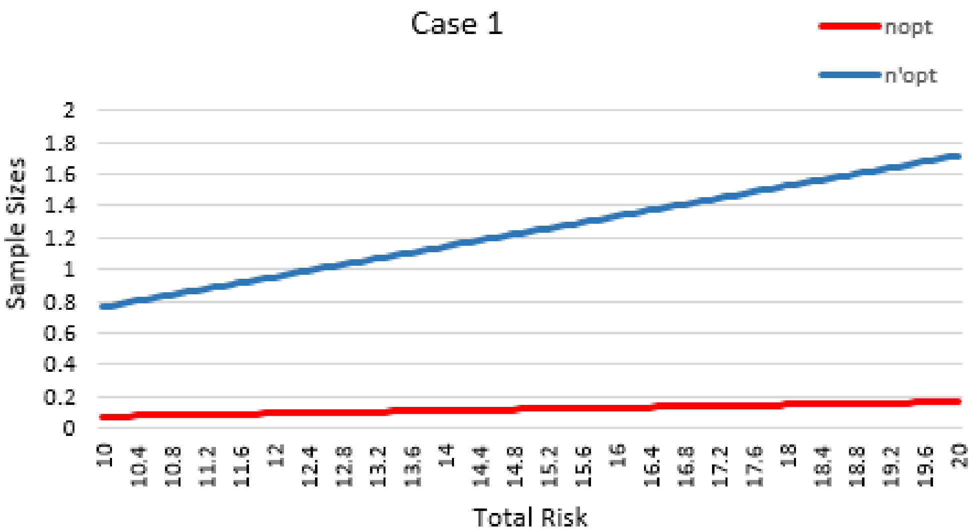


Figure 6: *n_{opt} and n'_{opt} for case I over change to total risk r_0*

Figure 7, depicts similar pattern among n'_{opt} and n_{opt} while considering variation of total risk r_0. But interesting is that with the increment in total risk r_0, the case II needs smaller optimum first phase (preliminary) sample than case I.

The Figure 8, reveals some interesting features of two cases I and II as when ratio $\left(\frac{n_{opt}}{n'_{opt}}\right)$ than case I. This feature confirms that if r_2 increases over fixed r_1 then n_{opt} increases over fixed n'_{opt}. But such increment is high in case II rather than case I.

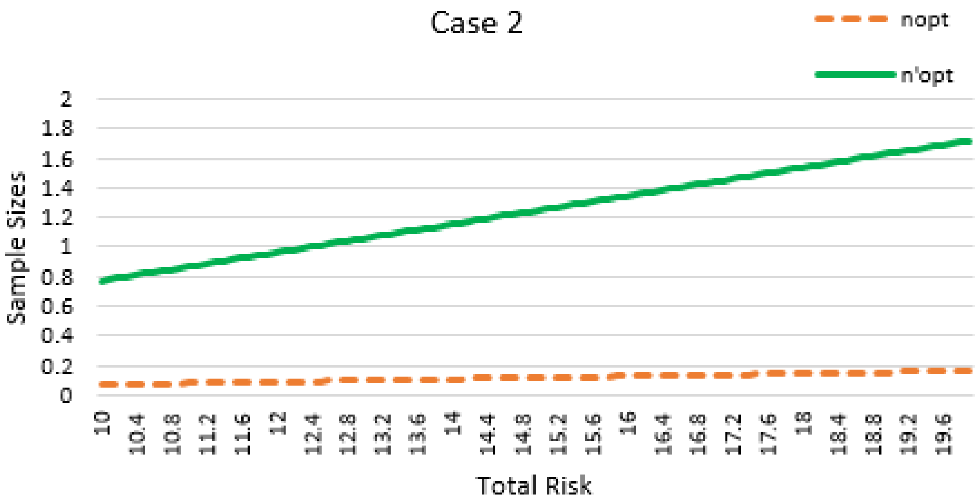


Figure 7: *Variation of n_{opt} and n'_{opt} for case II over change to total risk r_0*

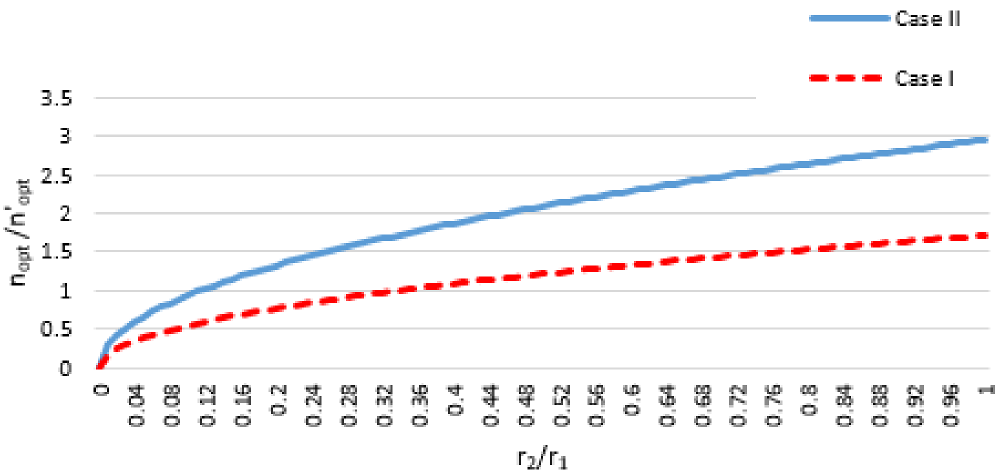


Figure 8: $\left(\frac{n_{opt}}{n'_{opt}}\right)$ *with respect to ratio of* $\left(\frac{r_2}{r_1}\right)$

7. Conclusion

On recapitulation, this paper presents a new class of estimators for estimating the unknown population mean in double sampling in the presence of auxiliary information. Some authors in literature have proposed exponential-type and some others proposed log-type estimators. The suggested estimation procedure is a combo-type class of estimators incorporating both expo and log-type structure. Its properties are discussed and compared in the set up of double sampling, under case I and case II sampling strategies. The proposed is found conditional efficient over usual expo-type and usual log-type estimators (Table 9). Moreover, a linear risk function is used in the paper with three risks parameters r_0, r_1, r_2 and expressions for optimal sample sizes n_{opt} and n'_{opt} are derived. Risk based simulation study reveals that increasing the fixed risk r_0 leads to larger n_{opt} (first sample) in comparison to equal (but small) n and n' to be used by the survey practitioner over incrementing r_0. Case II needs smaller preliminary sample size in comparison to case I. While considering variation of optimum ratio of sample sizes (n_{opt}/n'_{opt}) with respect to the risk ratio (r_2/r_1) variation, the case I graph of such ratio constantly reveals lower than the case II, graph indicating lesser need of comparative optimum sample ratio in double sampling using the suggested expo-log estimator at $\alpha = \alpha_{opt}$ choice.

References

[1] Alim A. and Shukla D. (2021). Double sampling based parameter estimation in Big Data and application in Control Charts,*Reliability(RT&M)*, 16(62):72-86.

[2] Bahl, S. and Tuteja, R.K. (1991). Ratio and product type estimators, *Journal of Information and Optimization Sciences*, 12(1):159-164.

[3] Bhushan, S. and Gupta, R (2019). Some log-type classes of estimator using auxiliary variable attribute, *Advances in Computational Science and Technology*, 12(2):99-108.

[4] Bhushan, S. and Kumar, A. (2020). Log-type estimators of population mean in rank set sampling, *Predictive analysis using statistics and Big Data: Concepts and Modeling*, 28:47-74.

[5] Bhushan, S., Pandey, A. and Shubra, K. (2008). A class of estimators in double sampling using two auxiliary variables, *Journal of Reliability and Statistical studies*, 1(1):67-73.

[6] Changbao Wu and Ying L. (2003). Optimal calibration estimators under two-phase sampling, *Quality Engineering*, 19(2):119-131.

[7] Kumari, C. and Thakur, R. K. (2021). An efficient log-type class of estimators using auxiliary information under double sampling, *Journal of Statistics Application and Probability*, 10(1):197-202.

[8] Muhammad A. and Muhammad H. (2014). On estimation of population mean in the presence of measurement error and non-response, *Pak. J. Statist.*, 31(5):657-670.

[9] Sahoo J., Sahoo L.N. and Mohanty S. (1993). A regression approach for estimation in two-phase sampling using auxiliary variables, *Current Science*, 65:73-75.

[10] Sanaullah, A., Ali, H. A., Amits, Muhammad Noor ul and Hanif, M. (2014). Generalised exponential chain ratio estimators under stratified two-phase random sampling, *Applied Mathematics and Computation*, 226: 541-547.

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

[12] Sanaullah, A., Hanif, M. and Asghar, A. (2016). Generalized exponential estimators for population variance under two-phase sampling, *International Journal of Applied Computation and Mathematics*, 2:75-84.

[13] Shabbir, J., Ahmed S., Sanaullah, A. and Onyange, R. (2021). Measuring performance of ratio-exponential-log type general class of estimators using two auxiliary variables, *Mathematical Problems in Engineering*, 2021(3):1-12.

[14] Shabbir, J. and Sat G. (2011). On estimating finite population mean in simple and stratified random sampling, *Communication in Statistics(Theory and Methods)*, 40:199-212.

[15] Shabbir, J., Sat G. and Masood, S. (2022). An improved class of estimators for finite population mean in simple random sampling, *Communication in Statistics, Theory and Method*, 51(11):3508-3520.

[16] Shukla, D. (2002). FT estimator under two phase sampling,*Metron (International Journal of Statistics)*, 60(1-2):97-106.

[17] Zahoor, A., Iqra M. and Muhammad H. (2014). Regression estimator in two phase sampling using Multi-auxiliary information in the presence of non-response and measurement error at second phase, *Journal of Applied Probability and Statistics*, 9(2):41-5.

[18] Zaman, T. and Kadilar, C. (2021). Exponential ratio and product type estimators of mean in stratified two-phase sampling, *AIMS Mathematics*, 6(5):4265–4279.

[19] https://data.gov.in/ Data source -6th Minor Irrigation Census - Village Schedule - Assam