RESEARCH ARTICLE

A simulation study: Using dual ancillary variable to estimate population mean under stratified random sampling

Sohaib Ahmad[1], **Sardar Hussain**[2]*, **Uzma Yasmeen**[3], **Muhammad Aamir**[1], **Javid Shabbir**[2], **M. El-Morshedy**[4,5], **Afrah Al-Bossly**[4], **Zubair Ahmad**[6]

1 Department of Statistics Abdul Wali Khan University Mardan Pakistan, **2** Department of Statistics, Quaid-i-azam University, Islamabad, Pakistan, **3** Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore, Pakistan, **4** Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia, **5** Faculty of Science, Department of Mathematics, Mansoura University, Mansoura, Egypt, **6** Department of Statistics, Yazd University, Yazd, Iran

* shussain@stat.qau.edu.pk
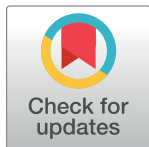
Abstract

In this paper, we propose an improved ratio-in-regression type estimator for the finite population mean under stratified random sampling, by using the ancillary varaible as well as rank of the ancillary varaible. Expressions of the bias and mean square error of the estimators are derived up to the first order of approximation. The present work focused on proper use of the ancillary variable, and it was discussed how ancillary variable can improve the precision of the estimates. Two real data sets as well as simulation study are carried out to observe the performances of the estimators. We demonstrate theoretically and numerically that proposed estimator performs well as compared to all existing estimators.

1. Introduction

In sampling theory, appropriate use of the ancillary information may increase precision of the estimators. Numerous authors employed the ancillary information at the designing stage and at the estimation stage. One purpose of sample survey theory is to estimate unknown population parameters of the variable of intrest being studied, such as mean, median, proportion and variance etc. It is preferable to employ stratified random sampling scheme rather than simple random sampling when data is based on hetrogeneous population.

In stratified random sampling, we divide the diverse population into strata or groups which are non-overlapping, and sampling is carried out from each stratum separetely. Zaman and Kadilar [1] when the difference in variance across strata is significantly greater than the difference in variance within strata, stratification enhances efficiency of the estimates. Zaman and Kadilar [2] and Zaman [3] provided an efficient estimators of population mean using the auxiliary varaible in stratified random sampling. Zaman [4] proposed an efficient exponential type estimator for the population mean under stratified random sampling. Rather and Kadilar [5] introduced dual to ratio cum product type of exponential estimator under stratified random sampling. Mradula et al. [6] obtained an efficient estimation of population mean under

stratified random smapling with linear cost function. Javed et al. [7] proposed a simulation based study for progressive estimation of population mean through traditional and non-traditional measures in stratified random sampling. Javed and Irfan [8] obtained a simulation based on new optimal estimators for population mean by using the dual auxiliary information in stratified random sampling. Yadav and Tailor [9] estimated a finite population mean using two auxiliary varaibles under stratified random sampling. Zaman and Bulut [10] proposed a modified regression estimator using robust regression methods and covaraince matrics in stratified sampling scheme. Kumar and Vishwakarma [11] proposed a generalized classes of regression-cum-ratio type estimators for population mean under stratified random sampling. Zahid et al. [12] provided a generalized class of estimators for sensitive varaible in the presence of measurment error and non-response under stratified random sampling. Some important references to the population mean under stratified random sampling using the auxiliary information include Aladag and Cingi [13], Grover and Kaur [14], Shabbir and Gupta [15], Khalid [16,17], Kadilar and Cingi [18,19], Koyuncu and Kadilar [20,21], Singh and Vishwakarma [22]. Recnetly Hussain et al. [23] proposed estimation of finite population distribution function with dual use of the anillary infromation under simple and stratified random sampling.

A unique idea for investigating more optimum estimators involving dual use of the ancillary information to deal with the stratified random sampling method has emerged recently. In this paper, we develop a new efficient estimator for finite population mean using dual ancillary variable under stratified random sampling.

Consider a finite population of distinct and identifiable units, $v = \{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_N\}$ of size N, are separated into L homogeneous strata of size $N_h (h = 1, 2,\ldots, L)$, such that $\sum_{h=1}^{L} N_h = N$. A simple random sample of size n_h is drawn without replacement from the hth stratum such that $\sum_{h=1}^{L} n_h = n$.

Let y, x and r_x be the study, auxiliary and rank of the auxiliary variables respectively, assuming values y_{h_i}, x_{h_i} and r_{xh_i} for the ith unit in the hth stratum. Let the stratum means be $\bar{Y}_h = \sum_{i=1}^{N_h} \frac{y_{h_i}}{N_h}$, $\bar{X}_h = \sum_{i=1}^{N_h} \frac{x_{h_i}}{N_h}$, $\bar{R}_{xh} = \sum_{i=1}^{N_h} \frac{r_{xh_i}}{N_h}$ respectively. Let $\bar{y}_{st} = \sum_{h=1}^{L} \omega_h \bar{y}_h$, $\bar{x}_{st} = \sum_{h=1}^{L} \omega_h \bar{x}_h$ and $\bar{r}_{xst} = \sum_{h=1}^{L} \omega_h \bar{r}_{xh}$ be the sample means of y, x and r_x respectively across the strata, where $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}, \bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$ and $\bar{r}_{xh} = \frac{1}{n_h} \sum_{i=1}^{n_h} r_{xhi}$ are the sample stratum means, $\omega_h = \frac{N_h}{N}$ are the known stratum weights. Let $\bar{Y} = \sum_{h=1}^{L} \omega_h \bar{Y}_h$, $\bar{X} = \sum_{h=1}^{L} \omega_h \bar{X}_h$ and $\bar{R}_x = \sum_{h=1}^{L} \omega_h \bar{R}_x$ are the population means of y, x and r_x respectively.

We also define the following error terms.

Let $\varepsilon_{0st} = \frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}}$, $\varepsilon_{1st} = \frac{\bar{x}_{st} - \bar{X}}{\bar{X}}$, $\varepsilon_{2st} = \frac{\bar{r}_{xst} - \bar{R}_x}{\bar{R}_x}$ such that $E(\varepsilon_i) = 0$, for (i = 0,1,2),

$$\text{where, } \nu_{rst} = \sum_{h=1}^{L} \omega_h^{r+s+t} \frac{E[(\bar{y}_h - \bar{Y})^r (\bar{x}_h - \bar{X})^s (\bar{r}_{xh} - \bar{R}_x)^t]}{\bar{Y}^r \bar{X}^s \bar{R}_x^t}$$

Using these notations, we have

$$E(\varepsilon_{0st}^2) = \frac{\sum_{h=1}^{l} \omega_h^2 \lambda_h C_{yh}^2}{\bar{Y}^2} = \nu_{200}, \quad E(\varepsilon_{1st}^2) = \frac{\sum_{h=1}^{l} \omega_h^2 \lambda_h C_{xh}^2}{\bar{X}^2} = \nu_{020}, \quad E(\varepsilon_{2st}^2) = \frac{\sum_{h=1}^{l} \omega_h^2 \lambda_h C_{rxh}^2}{\bar{R}_x^2}$$
$$= \nu_{002},$$

$$E(\varepsilon_{0st} \varepsilon_{1st}) = \frac{\sum_{h=1}^{l} \omega_h^2 \lambda_h S_{yxh}}{\bar{Y}\bar{X}} = \nu_{110}, \quad E(\varepsilon_{0st} \varepsilon_{2st}) = \frac{\sum_{h=1}^{l} \omega_h^2 \lambda_h S_{yrxh}}{\bar{Y}\bar{R}_x} = \nu_{101},$$

$$\mathrm{E}(\varepsilon_{1st}\varepsilon_{2st}) = \frac{\sum_{h=1}^{l} \omega_h^2 \lambda_h S_{xrxh}}{\bar{X}\bar{R}_x} = \boldsymbol{\nu}_{011}, \ \lambda_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right).$$

Let $C_{yh} = \frac{S_{yh}}{\bar{Y}}$, $C_{xh} = \frac{S_{xh}}{\bar{X}}$, and $C_{rxh} = \frac{S_{rxh}}{\bar{R}_x}$, be the coefficients of variation of y, x and r_x and $\rho_{y_h x_h}, \rho_{y_h rx_h}$, and $\rho_{x_h rx_h}$ be the population correlation coefficients between (y_h, x_h), (y_h, r_{xh}), and (x_h, r_{xh}) respectively in the hth stratum. Let $s_{yh}^2 = \frac{\sum_{i=1}^{N_h}(y_{hi}-\bar{y}_h)^2}{N_h-1}$, $s_{xh}^2 = \frac{\sum_{i=1}^{N_h}(x_{hi}-\bar{x}_h)^2}{N_h-1}$, and $s_{rxh}^2 = \frac{\sum_{i=1}^{N_h}(rx_{hi}-\bar{r}x_h)^2}{N_h-1}$, are the sample variances in the hth stratum, where $s_{yh} = \sqrt{\frac{\sum_{i=1}^{N_h}(y_{hi}-\bar{y}_h)^2}{N_h-1}}$, $s_{xh} = \sqrt{\frac{\sum_{i=1}^{N_h}(x_{hi}-\bar{x}_h)^2}{N_h-1}}$, and $s_{rxh} = \sqrt{\frac{\sum_{i=1}^{N_h}(rx_{hi}-\bar{r}x_h)^2}{N_h-1}}$ are the sample standard deviation in the hth stratum. Let $s_{yxh} = \frac{\sum_{i=1}^{N_h}(y_{hi}-\bar{y}_h)(x_{hi}-\bar{x}_h)}{N_h-1}$, $s_{yrxh} = \frac{\sum_{i=1}^{N_h}(y_{hi}-\bar{y}_h)(r_{xhi}-\bar{r}_{xh})}{N_h-1}$, and $s_{xrxh} = \frac{\sum_{i=1}^{N_h}(x_{hi}-\bar{x}_h)(r_{xhi}-\bar{r}_{xh})}{N_h-1}$, are the covariance between their respective subscripts.

The rest of the paper is arranged as follows: In Section 2, literature review of various estimators under stratified random sampling is introduced. In Section 3, we proposed estimator for estimating finite population mean under stratified random sampling using dual ancillary variables are defined. In Section 4, theoretical comparison is conducted to assess the performances of the estimators. In Section 5 and 6, numerical investigation and data description are given. Monte Carlo simulation is concluded in Section 7. Discussions of the numerical results are given in Section 8. Finally, concluding remarks are discussed in Section 9.

2. Literature review

This section studies various stratified estimators that are available in the literature:

(i) The traditional mean estimator in stratified random sampling is:

$$\hat{\bar{Y}}_{(st)} = \sum_{h=1}^{L} \omega_h \bar{y}_h.$$

The variance of $\hat{\bar{Y}}_{st}$, is given by

$$\mathrm{Var}(\hat{\bar{Y}}_{(st)}) = \bar{Y}^2 \boldsymbol{\nu}_{200} \tag{1}$$

(ii) Cochran [24], suggested the traditional ratio $\hat{\bar{Y}}_{R(st)}$ estimator:

$$\hat{\bar{Y}}_{R(st)} = \bar{y}_{(st)} \left(\frac{\bar{X}}{\bar{x}_{(st)}}\right) \tag{2}$$

The bias and MSE of $\hat{\bar{Y}}_R$, are given by:

$$\mathrm{Bias}(\hat{\bar{Y}}_{R(st)}) = \bar{Y}(\boldsymbol{\nu}_{020} - \boldsymbol{\nu}_{110}),$$

and

$$\text{MSE}(\hat{\bar{Y}}_{R(st)}) \cong \bar{Y}^2(\boldsymbol{\nu}_{200} + \boldsymbol{\nu}_{020} - 2\boldsymbol{\nu}_{110}) \tag{3}$$

(iii) Murthy [25] suggested the usual product estimator $\hat{\bar{Y}}_{P(st)}$, which is given by:

$$\hat{\bar{Y}}_{P(st)} = \bar{y}_{(st)}\left(\frac{\bar{x}_{(st)}}{\bar{X}}\right) \tag{4}$$

The bias and MSE of $\hat{\bar{Y}}_{P(st)}$, are given by:

$$\text{Bias}(\hat{\bar{Y}}_{P(st)}) = \bar{Y}\boldsymbol{\nu}_{110},$$

and

$$\text{MSE}(\hat{\bar{Y}}_{P(st)}) \cong \bar{Y}^2(\boldsymbol{\nu}_{200} + \boldsymbol{\nu}_{020} + 2\boldsymbol{\nu}_{110}) \tag{5}$$

(iv) Bahl and Tuteja [26] suggested the following estimators:

$$\hat{\bar{Y}}_{BT,R(st)} = \bar{y}_{(st)}\exp\left(\frac{\bar{X} - \bar{x}_{(st)}}{\bar{X} + \bar{x}_{(st)}}\right), \tag{6}$$

and

$$\hat{\bar{Y}}_{BT,P(st)} = \bar{y}_{(st)}\exp\left(\frac{\bar{x}_{(st)} - \bar{X}}{\bar{x}_{(st)} + \bar{X}}\right) \tag{7}$$

The biases and MSEs of $\hat{\bar{Y}}_{BT,R(st)}$, and $\hat{\bar{Y}}_{BT,P(st)}$, are given by:

$$\text{Bias}(\hat{\bar{Y}}_{BT,R(st)}) \cong \bar{Y}\left(\frac{3}{8}\boldsymbol{\nu}_{020} - \frac{1}{2}\boldsymbol{\nu}_{110}\right),$$

$$\text{MSE}(\hat{\bar{Y}}_{BT,R(st)}) = \frac{\bar{Y}^2}{4}\left(4\boldsymbol{\nu}_{200} + \boldsymbol{\nu}_{020} - 4\boldsymbol{\nu}_{110}\right). \tag{8}$$

and

$$\text{Bias}(\hat{\bar{Y}}_{BT,P(st)}) = \bar{Y}\left(\frac{1}{2}\boldsymbol{\nu}_{110} - \frac{1}{8}\boldsymbol{\nu}_{020}\right),$$

$$\text{MSE}(\hat{\bar{Y}}_{BT,P(st)}) = \frac{\bar{Y}^2}{4}\left(4\boldsymbol{\nu}_{200} + \boldsymbol{\nu}_{020} + 4\boldsymbol{\nu}_{110}\right). \tag{9}$$

(v) The difference estimator $\hat{\bar{Y}}_{dif}$, is given by:

$$\hat{\bar{Y}}_{dif(st)} = \bar{y}_{(st)} + d(\bar{X} - \bar{x}_{(st)}), \tag{10}$$

where d is an appropriate constant. The minimum variance of $\hat{\bar{Y}}_{dif(st)}$ at the optimal worth $d_{opt} = \frac{\bar{Y}\boldsymbol{v}_{110}}{\bar{X}\boldsymbol{v}_{020}}$, is given as:

$$\text{Var}(\hat{\bar{Y}}_{dif(st)})_{min} = \frac{\bar{Y}^2(\boldsymbol{v}_{200}\boldsymbol{v}_{020} - \boldsymbol{v}_{110})}{\boldsymbol{v}_{020}}, \tag{11}$$

(vi) Rao [27], proposed the following estimator:

$$\hat{\bar{Y}}_{R,D(st)} = Q_1\bar{y}_{(st)} + Q_2(\bar{X} - \bar{x}_{(st)}), \tag{12}$$

where Q_1 and Q_2 are constants.

The bias and MSE of $\hat{\bar{Y}}_{R,D(st)}$, given by:

$$\text{Bias}(\hat{\bar{Y}}_{R,D(st)}) = \bar{Y}(Q_1 - 1),$$

and

$$\text{MSE}(\hat{\bar{Y}}_{R,D(st)}) = \bar{Y}^2 - 2Q_1\bar{Y}^2 + Q_1^2\bar{Y}^2 + Q_1^2\bar{Y}^2\boldsymbol{v}_{200} - 2Q_1Q_2\bar{Y}\bar{X}\boldsymbol{v}_{110} + Q_2^2\bar{X}^2\boldsymbol{v}_{020}.$$

The optimum values of Q_1 and Q_2 are given by:

$$Q_{1opt} = \frac{\boldsymbol{v}_{020}}{(\boldsymbol{v}_{020}\boldsymbol{v}_{200} - \boldsymbol{v}_{110}{}^2 + \boldsymbol{v}_{020})},$$

and

$$Q_{2opt} = \frac{\bar{Y}\boldsymbol{v}_{110}}{\bar{X}(\boldsymbol{v}_{200}\boldsymbol{v}_{020} - \boldsymbol{v}_{110}{}^2 + \boldsymbol{v}_{020})}.$$

The minimum MSE of $\hat{\bar{Y}}_{R,D(st)}$ is given by:

$$\text{MSE}(\hat{\bar{Y}}_{R,D(st)})_{min} = \frac{\bar{Y}^2(\boldsymbol{v}_{200}\boldsymbol{v}_{020} - \boldsymbol{v}_{101}^2)}{(\boldsymbol{v}_{200}\boldsymbol{v}_{020} - \boldsymbol{v}_{110}{}^2 + \boldsymbol{v}_{020})}. \tag{13}$$

(vii) The suggested estimator by Singh et al. [28]:

$$\hat{\bar{Y}}_{Singh(st)} = \bar{y}_{(st)}\exp\left(\frac{a(\bar{X} - \bar{x}_{(st)})}{a(\bar{X} + \bar{x}_{(st)}) + 2b}\right), \tag{14}$$

For a = 1, and b = 0,

The bias and MSE of $\hat{\bar{Y}}_{Singh}$, is given by:

$$\text{Bias}(\hat{\bar{Y}}_{Singh(st)}) = \bar{Y}\left(\frac{3}{8}\boldsymbol{v}_{020} - \frac{1}{2}\boldsymbol{v}_{110}\right),$$

and

$$\text{MSE}\left(\hat{\bar{Y}}_{S(st)}\right) \cong \frac{\bar{Y}^2}{4}\left(4\boldsymbol{\nu}_{200} + \boldsymbol{\nu}_{020} - 4\boldsymbol{\nu}_{110}\right). \tag{15}$$

(viii) The suggested estimator of Grover and Kaur [29], is given as:

$$\hat{\bar{Y}}_{Gk(st)} = \{Z_1\bar{y}_{(st)} + Z_2(\bar{X} - \bar{x}_{(st)})\}\exp\left(\frac{a(\bar{X} - \bar{x})}{a(\bar{X} + \bar{x}) + 2b}\right). \tag{16}$$

Where Z_1 and Z_2 are unknown constants. For a = 1 and b = 0
The bias and MSE of $\hat{\bar{Y}}_{Gk(st)}$ are given by:

$$\text{Bias}\left(\hat{\bar{Y}}_{Gk(st)}\right) = \bar{Y}(Z_1 - 1) + \frac{3}{8}Z_1\bar{Y} + \frac{1}{2}Z_2\bar{X}\boldsymbol{\nu}_{020} - \frac{1}{2}\bar{Y}\boldsymbol{\nu}_{110}Z_1,$$

And

$$\begin{aligned}
\text{MSE}&(\hat{\bar{Y}}_{Gk(st)})_{min} \\
&= (Z_1 - 1)\bar{Y}^2 + Z_1^2\bar{Y}^2(\boldsymbol{\nu}_{200} + \boldsymbol{\nu}_{020} - 2\boldsymbol{\nu}_{110}) + Z_2^2\bar{X}^2\boldsymbol{\nu}_{020} - 2Z_1\bar{Y}^2\left(\frac{3}{8}\boldsymbol{\nu}_{020} - \frac{1}{2}\boldsymbol{\nu}_{110}\right) \\
&\quad - 2Z_2\bar{Y}\bar{X}\left(\frac{1}{2}\boldsymbol{\nu}_{020}\right) - 2Z_1Z_2\bar{Y}\bar{X}(\boldsymbol{\nu}_{110} - \boldsymbol{\nu}_{020})
\end{aligned}$$

$$\begin{aligned}
\text{MSE}&(\hat{\bar{Y}}_{Gk(st)})_{min} \\
&\cong Z_2^2\bar{X}^2\boldsymbol{\nu}_{020} + Z_1^2\bar{Y}^2\boldsymbol{\nu}_{200} + 2Z_1Z_2\bar{Y}\bar{X}\boldsymbol{\nu}_{020} - 2Z_1Z_2\bar{Y}\bar{X}\boldsymbol{\nu}_{110} + \bar{Y}^2 - 2Z_1\bar{Y}^2 + Z_1^2\bar{Y}^2 \\
&\quad - 2Z_1\bar{Y}^2\boldsymbol{\nu}_{110} - Z_2\bar{Y}\bar{X}\boldsymbol{\nu}_{020} - 2Z_1^2\bar{Y}^2\boldsymbol{\nu}_{110} - \frac{3}{4}Z_1\bar{Y}^2\boldsymbol{\nu}_{020} + Z_1^2\bar{Y}^2\boldsymbol{\nu}_{020}. \tag{17}
\end{aligned}$$

The optimum values of Z_1 and Z_2 are given as:

$$Z_{1(opt)} = \frac{\boldsymbol{\nu}_{020}(\boldsymbol{\nu}_{020} - 8)}{8(-\boldsymbol{\nu}_{200}\boldsymbol{\nu}_{020} - \boldsymbol{\nu}_{110}{}^2 - \boldsymbol{\nu}_{020})},$$

$$Z_{2(opt)} = \frac{\bar{Y}(\boldsymbol{\nu}_{020}^2 - 9_{020}\boldsymbol{\nu}_{020} + 4\boldsymbol{\nu}_{200}\boldsymbol{\nu}_{020} - 4\boldsymbol{\nu}_{110}{}^2 - 4\boldsymbol{\nu}_{020} + 8\boldsymbol{\nu}_{110})}{8\bar{X}(\boldsymbol{\nu}_{200}\boldsymbol{\nu}_{020} - \boldsymbol{\nu}_{110}{}^2 + \boldsymbol{\nu}_{020})},$$

The minimal MSE of $\hat{\bar{Y}}_{Gk(st)}$, are:

$$\hat{\bar{Y}}_{GK(st)min} = \frac{\bar{Y}^2}{64}\left(64 - 16\boldsymbol{\nu}_{020} - \frac{\boldsymbol{\nu}_{020}(-8 + \boldsymbol{\nu}_{020})^2}{\boldsymbol{\nu}_{020(1+\boldsymbol{\nu}_{200})} - \boldsymbol{\nu}_{110}^2}\right) \tag{18}$$

(ix) Ahmad et al. [30] proposed an improved estimator $\hat{\tilde{Y}}_{Pr_{st}}$, is given by:

$$\hat{\tilde{Y}}_{Pr_{st}} = \left\{ Q_5\hat{\tilde{Y}}_{st} + Q_6\left(\frac{\tilde{X} - \hat{\tilde{X}}_{st}}{\tilde{X}}\right) + Q_7\left(\frac{\tilde{Z} - \hat{\tilde{Z}}_{st}}{\tilde{Z}}\right) \right\} \exp\left(\frac{a(\tilde{X} - \hat{\tilde{X}}_{st})}{a(\tilde{X} + \hat{\tilde{X}}_{st}) + 2b}\right). \tag{19}$$

Where $Q_i(\mathrm{i} = 5,6,7)$ are constants.

The bias and MSE of $\hat{\tilde{Y}}_{Pr_{st}}$, are given by:

$$\mathrm{B}\left(\hat{\tilde{Y}}_{Pr_{st}}\right) \cong \tilde{Y}(Q_5 - 1) + \frac{3}{8}\theta^2 Q_5 \tilde{Y} V_{020} + \frac{1}{2}\theta Q_6 V_{020} - \frac{1}{2}\theta Q_5 \tilde{Y} V_{110} + \frac{1}{2}\theta V_{011},$$

and

$$\begin{aligned}\mathrm{MSE}\left(\hat{\tilde{Y}}_{Pr_{st}}\right) \cong{}& \tilde{Y}^2(Q_5 - 1)^2 + Q_5^2\tilde{Y}^2 V_{200} + Q_6 V_{020} + Q_7 V_{002} + \theta^2 Q_5^2\tilde{Y}^2 V_{020} - \theta Q_6\tilde{Y}V_{020} \\ &+ 2\theta\tilde{Y}V_{020} - \frac{3}{4}\theta^2 Q_5\tilde{Y}^2 V_{020} + \theta Q_5\tilde{Y}^2 V_{110} - 2Q_5^2\tilde{Y}^2 V_{110} - 2Q_5 Q_6\tilde{Y}V_{110} \\ &- 2Q_5 Q_7\tilde{Y}V_{101} - \theta Q_7\tilde{Y}V_{011} + 2\theta Q_5 Q_7\tilde{Y}V_{011} - 2Q_6 Q_7 V_{011}.\end{aligned} \tag{20}$$

The optimum values of Q_5, Q_6 and Q_7 are given by:

$$Q_{5(\mathrm{opt})} = \frac{8 - \theta^2 V_{020}}{8\{1 + V_{200}(1 - R_{y.xz}^2)\}},$$

$$Q_{6(\mathrm{opt})} = \frac{\tilde{Y}\left[\begin{array}{l}\theta^3 V_{020}^{3/2}(R_{xz}^2 - 1) + V_{200}^{1/2}(-8 + \theta^2 V_{020})(R_{yx} - R_{xz}R_{yz}) \\ +4\theta V_{020}^{1/2}(R_{xz}^2 - 1)\{-1 + V_{200}(1 - R_{y.xz}^2)\}\end{array}\right]}{8V_{020}^{1/2}(R_{xz}^2 - 1)\{-1 + V_{200}(1 - R_{y.xz}^2)\}},$$

$$Q_{7(\mathrm{opt})} = \frac{\tilde{Y}V_{200}^{1/2}(8 - \theta^2 V_{020})(R_{yx} - R_{xz}R_{yz})}{8V_{020}^{1/2}(R_{xz}^2 - 1)\{-1 + V_{200}(1 - R_{y.xz}^2)\}}.$$

The minimum MSE of $\hat{\tilde{Y}}_{Pr_{st}}$ at optimum values of Q_5, Q_6, and Q_7 are given by:

$$\mathrm{MSE}\left(\hat{\tilde{Y}}_{Pr_{st}}\right) \cong \frac{\tilde{Y}^2\{64V_{200}(1 - R_{y.xz}^2) - \theta^4 V_{020}^2 - 16\theta^2 V_{020}V_{200}(1 - R_{y.xz}^2)\}}{64\{1 + V_{200}(1 - R_{y.xz}^2)\}}, \tag{21}$$

where $R_{y.xz}^2 = \left(\frac{V_{110}^2 V_{002} + V_{101}^2 V_{020} - 2V_{101}V_{110}V_{011}}{V_{200}(V_{020}V_{002} - V_{011}^2)}\right)$.

3. Proposed estimator

Suitable use of the ancillary information may improve the precision of estimators both at the design and estimation stage. The rank of the ancillary variable is correlated with the study variable when the correlation among the study and ancillary variable is strong. In literature, dual use of ancillary variable has been rarely attempted, therefore we motivated towards it. The principal advantage of our proposed ratio-in-regression type estimator under stratified random sampling is that it is more flexible, efficient than the existing estimators. Taking motivation from Ahmad et al. [30], we propose ratio-in-regression type exponential estimator for

estimating the population mean under stratified random sampling.

$$\hat{\bar{Y}}_{ss(st)} = Q_{11}\bar{y}_{(st)} + Q_{12}\left(\bar{X} - \bar{x}_{(st)}\right)\exp\left(\frac{\bar{X} - \bar{x}_{(st)}}{\bar{X} + \bar{x}_{(st)}}\right) + Q_{13}\left(\bar{R}_x - \bar{r}_{x(st)}\right)\exp\left(\frac{\bar{R}_x - \bar{r}_{x(st)}}{\bar{R}_x + \bar{r}_{x(st)}}\right) \quad (22)$$

Where Q_{11}, Q_{12} and Q_{13} are unknown constants.

Solving $\hat{\bar{Y}}_{ss(st)}$ given in Eq (22),

$$\hat{\bar{Y}}_{ss(st)} = Q_{11}\bar{Y}(1 + \boldsymbol{\varepsilon}_{0st}) - Q_{12}\bar{X}\boldsymbol{\varepsilon}_{1st}\left(1 - \frac{1}{2}\boldsymbol{\varepsilon}_{1st} + \frac{3}{8}\epsilon_{1st}^2\right) - Q_{13}\bar{R}_x\boldsymbol{\varepsilon}_{2st}\left(1 - \frac{1}{2}\boldsymbol{\varepsilon}_{2st} + \frac{3}{8}\boldsymbol{\varepsilon}_{2st}^2\right)$$

$$\hat{\bar{Y}}_{ss(st)} - \bar{Y} = (Q_{11} - 1)\bar{Y} + Q_{11}\bar{Y}\boldsymbol{\varepsilon}_{0st} - Q_{12}\bar{X}\left(\boldsymbol{\varepsilon}_{1st} - \frac{1}{2}\boldsymbol{\varepsilon}_{1st}^2\right) - Q_{13}\bar{R}_x\left(\boldsymbol{\varepsilon}_{2st} - \frac{1}{2}\boldsymbol{\varepsilon}_{2st}^2\right) \quad (i)$$

$$\text{Bias}(\hat{\bar{Y}}_{ss(st)}) = (Q_{11} - 1)\bar{Y} + \frac{1}{2}\bar{X}Q_{12}\boldsymbol{\nu}_{020} + \frac{1}{2}Q_{13}\bar{R}_x\boldsymbol{\nu}_{002}$$

Simplify Eq (i), we have

$$\text{MSE}\left(\hat{\bar{Y}}_{ss(st)}\right) = (Q_{11} - 1)^2\bar{Y}^2 + Q_{11}^2\bar{Y}^2\epsilon_{0st}^2 + Q_{12}^2\bar{X}^2\epsilon_{1st}^2 + Q_{13}^2\bar{R}x^2\epsilon_{2st}^2 + 2\,Q_{11}(Q_{11} - 1)\bar{Y}^2\boldsymbol{\varepsilon}_{0st}$$

$$- 2\,(Q_{11} - 1)Q_{12}\bar{Y}\bar{X}\left(\boldsymbol{\varepsilon}_{1st} - \frac{1}{2}\boldsymbol{\varepsilon}_{1st}^2\right) - 2\,(Q_{11} - 1)Q_{13}\bar{Y}\bar{R}_x\left(\boldsymbol{\varepsilon}_{2st} - \frac{1}{2}\boldsymbol{\varepsilon}_{2st}^2\right)$$

$$- 2\,Q_{11}Q_{12}\bar{Y}\bar{X}(\boldsymbol{\varepsilon}_{0st}\boldsymbol{\varepsilon}_{1st}) - 2\,Q_{11}Q_{13}\bar{Y}\bar{R}_x(\boldsymbol{\varepsilon}_{0st}\boldsymbol{\varepsilon}_{2st}) - 2\,Q_{12}Q_{13}\bar{Y}\bar{R}_x\boldsymbol{\varepsilon}_{1st}\boldsymbol{\varepsilon}_{2st}$$

$$= (Q_{11} - 1)^2\bar{Y}^2 + Q_{11}^2\bar{Y}^2\boldsymbol{\nu}_{200} + Q_{12}^2\bar{X}^2\boldsymbol{\nu}_{020} + Q_{13}^2\bar{R}x^2\boldsymbol{\nu}_{002} + 2(Q_{11}Q_{12} - Q_{13})\bar{Y}\bar{X}\frac{\boldsymbol{\nu}_{020}}{2}$$

$$+ 22(Q_{11}Q_{13} - Q_{13})\bar{Y}\bar{R}_x\frac{\boldsymbol{\nu}_{002}}{2} - 2Q_{11}Q_{12}\bar{Y}\bar{X}\boldsymbol{\nu}_{110} - 2\,Q_{11}Q_{13}\bar{Y}\bar{R}_x\boldsymbol{\nu}_{101} + 2\,Q_{12}\,Q_{13}\bar{X}\bar{R}_x\boldsymbol{\nu}_{011}$$

$$\text{MSE}\left(\hat{\bar{Y}}_{ss(st)}\right) = (Q_{11} - 1)^2\bar{Y}^2 + Q_{11}^2\bar{Y}^2\boldsymbol{\nu}_{200} + Q_{12}^2\bar{X}^2\boldsymbol{\nu}_{020} + Q_{13}^2\bar{R}x^2\boldsymbol{\nu}_{002} - 2Q_{12}\bar{Y}\bar{X}\frac{\boldsymbol{\nu}_{020}}{2}$$

$$- 2Q_{13}\bar{Y}\bar{R}_x\frac{\boldsymbol{\nu}_{002}}{2} + 2Q_{11}Q_{12}\bar{Y}\bar{X}\left(\frac{\boldsymbol{\nu}_{020}}{2} - \boldsymbol{\nu}_{110}\right) + 2Q_{11}Q_{13}\bar{Y}\bar{R}_x\left(\frac{\boldsymbol{\nu}_{002}}{2} - \boldsymbol{\nu}_{101}\right)$$

$$+ 2Q_{12}Q_{13}\bar{X}\bar{R}x^2\boldsymbol{\nu}_{011}(23)$$

The optimum values of Q_{11}, Q_{12} and Q_{13} are given by:

$$Q_{11(opt)} = \frac{\left(2\lambda\frac{\boldsymbol{\nu}_{200}^{1/2}}{\lambda^{1/2}}\right)(S_{11} + B_{11}) + \lambda\left(T + \frac{\boldsymbol{\nu}_{020}}{\lambda}\right) - 4}{4\boldsymbol{\nu}_{200}\bar{Y}\left(R_{yx}^2 - 1\right) + 4\boldsymbol{\nu}_{200}^{1/2}(S_{11} + P_{11}) + \lambda\left(T + \frac{\boldsymbol{\nu}_{020}}{\lambda}\right) - 4},$$

$$Q_{12(opt)} = \frac{\frac{\boldsymbol{\nu}_{200}^{1/2}}{\lambda^{1/2}}\bar{Y}\left[2\boldsymbol{\nu}_{200}^{1/2}\lambda^{1/2}\left\{\frac{\boldsymbol{\nu}_{002}^{1/2}}{\lambda^{1/2}}S_{11+} \frac{\boldsymbol{\nu}_{020}^{1/2}}{\lambda^{1/2}}\left(\frac{\boldsymbol{\nu}_{101}^2}{\boldsymbol{\nu}_{200}\boldsymbol{\nu}_{002}}\right)\right\}\boldsymbol{\nu}_{002}^{1/2}\boldsymbol{\nu}_{020}^{1/2}D_{11} + 4S_{11}\right]}{\left[\frac{\boldsymbol{\nu}_{020}^{1/2}}{\lambda^{1/2}}\bar{X}\left(1 - \frac{\boldsymbol{\nu}_{101}^2}{\boldsymbol{\nu}_{200}\boldsymbol{\nu}_{002}}\right)\left\{4\boldsymbol{\nu}_{200}\left(R_{yx}^2 - 1\right) + \lambda\boldsymbol{\nu}_{200}^{1/2}(S_{11} + B_{11}) + \lambda\left(T + \frac{\boldsymbol{\nu}_{020}}{\lambda}\right) - 4\right\}\right]},$$

$$Q_{13(opt)} = \frac{\frac{\boldsymbol{\nu}_{200}^{1/2}}{\lambda^{1/2}}\bar{Y}\left[2\boldsymbol{\nu}_{200}^{1/2}\lambda^{1/2}\left\{\frac{\boldsymbol{\nu}_{002}^{1/2}}{\lambda^{1/2}}S_{11+} \frac{\boldsymbol{\nu}_{020}^{1/2}}{\lambda^{1/2}}\left(\frac{\vartheta_{101}^2}{\boldsymbol{\nu}_{200}\boldsymbol{\nu}_{002}}\right)\right\} + 4S_{11}\right]}{\left[\frac{\boldsymbol{\nu}_{020}^{1/2}}{\lambda^{1/2}}\bar{R}_x\left(1 - \frac{\boldsymbol{\nu}_{101}^2}{\boldsymbol{\nu}_{200}\boldsymbol{\nu}_{002}}\right)\left\{4\boldsymbol{\nu}_{200}\left(R_{yx}^2 - 1\right) + \lambda\boldsymbol{\nu}_{200}^{1/2}(S_{11} + B_{11}) + \lambda\left(T + \frac{\boldsymbol{\nu}_{020}}{\lambda}\right) - 4\right\}\right]},$$

where

$$D_{11} = \frac{\nu_{110}}{\sqrt{\nu_{200}}\sqrt{\nu_{020}}} - \frac{\nu_{101}}{\sqrt{\nu_{200}}\sqrt{\nu_{002}}}, \ R^2_{y.xz} = \left(\frac{\nu^2_{110}\nu_{002} + \nu^2_{101}\nu_{020} - 2\nu_{101}\nu_{110}\nu_{011}}{\nu_{200}(\nu_{020}\nu_{002} - \nu^2_{110})} \right),$$

$$T = \frac{\left\{ \frac{\nu^{1/2}_{002}}{\lambda^{1/2}} - \frac{\nu^{1/2}_{020}}{\lambda^{1/2}}\left(\frac{\nu_{011}}{\sqrt{\nu_{020}}\sqrt{\nu_{002}}} \right) \right\}^2}{1 - \frac{\nu^2_{011}}{\nu_{020}\nu_{002}}}, \ S_{11} = \frac{\frac{\nu^{1/2}_{020}}{\lambda^{1/2}}\left(\frac{\nu_{101}}{\sqrt{\nu_{200}}\sqrt{\nu_{002}}} - \frac{\nu_{011}}{\sqrt{\nu_{020}}\sqrt{\nu_{002}}} - \frac{\nu_{110}}{\sqrt{\nu_{200}}\sqrt{\nu_{020}}} \right)}{1 - \frac{\nu^2_{011}}{\nu_{020}\nu_{002}}},$$

$$P_{11} = \frac{\left\{ \frac{\nu_{101}}{\sqrt{\nu_{200}}\sqrt{\nu_{002}}}\left(\frac{\nu^{1/2}_{200}}{\lambda^{1/2}} \right) - \frac{\nu_{110}}{\sqrt{\nu_{020}}\sqrt{\nu_{002}}}\left(\frac{\nu^{1/2}_{002}}{\lambda^{1/2}} \right) \right\}^2}{1 - \frac{\nu^2_{011}}{\nu_{020}\nu_{002}}}, \ B_{11} = \frac{\frac{\nu^{1/2}_{002}}{\lambda^{1/2}}\left(\frac{\nu_{110}}{\sqrt{\nu_{200}}\sqrt{\nu_{002}}} - \frac{\nu_{010}}{\sqrt{\nu_{020}}\sqrt{\nu_{002}}} - \frac{\nu_{101}}{\sqrt{\nu_{200}}\sqrt{\nu_{002}}} \right)}{1 - \frac{\nu^2_{011}}{\nu_{020}\nu_{002}}}.$$

Putting the optimum values of Q_{11}, Q_{12} and Q_{13} in (23), we get the minimal mean square error of $\hat{\bar{Y}}_{ss(st)}$, given by:

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} = \frac{\bar{Y}^2\nu_{200}\{T(\lambda) - B_{11}(\lambda) + \nu_{020} + 4(R^2_{yx} - 1)\}}{\left\{ \lambda\left(4\frac{\nu_{200}}{\lambda}\left(R^2_{yx} - 1 \right) + 4\frac{\nu^{1/2}_{020}}{\lambda^{1/2}}(S_{11} + B_{11}) + T + \frac{\nu_{020}}{\lambda} - 4 \right) \right\}}. \quad (24)$$

4. Theoretical comparision

In this section, the theoretical comparison of the suggested estimator with existing estimators is considered:

(i) By taking (1) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \text{Var}(\hat{\bar{Y}}_{(st)}) \ \text{if}$$

$$\text{Var}(\hat{\bar{Y}}_{(st)}) - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0$$

$$\frac{\bar{Y}^2\nu_{200}\left[\frac{N_{11}}{\sqrt{\lambda\left(\frac{1}{2}\right)}}[N_{13}] \right]}{\lambda\left[\frac{N_{12}}{\sqrt{\lambda\left(\frac{1}{2}\right)}} \right]} > 0$$

where

$$N_{11} = (T - 4)\lambda + R^2_{yx}\nu_{200} + \nu_{020} - 4\nu_{200}\sqrt{\lambda\left(\frac{1}{2}\right)} + 4\sqrt{\nu_{020}}\lambda(S_{11} + B_{11}),$$

$$N_{12} = \left((T - 4)\lambda + 4R^2_{yx}\nu_{200} + \nu_{020} - 4\nu_{200} \right)\sqrt{\lambda\left(\frac{1}{2}\right)} + 4\sqrt{\nu_{020}}\lambda(S_{11} + B_{11}),$$

$$N_{13} = T(\lambda) - B_{11}\lambda + \nu_{020} + 4R^2_{yx} - 4$$

(ii) By taking (3) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \text{MSE}(\hat{\bar{Y}}_{R(st)}) \text{ if}$$

$$\text{MSE}(\hat{\bar{Y}}_{R(st)}) - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0$$

$$\frac{\bar{Y}^2 \left[\frac{(-\boldsymbol{\nu}_{020} + 2\boldsymbol{\nu}_{110} - \boldsymbol{\nu}_{200})\lambda(N_{11})}{\sqrt{\lambda\left(\frac{1}{2}\right)}} + \boldsymbol{\nu}_{200}[N_{13}] \right]}{\lambda \left[\frac{N_{12}}{\sqrt{\lambda\left(\frac{1}{2}\right)}} \right]} > 0$$

(iii) By taking (5) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \text{MSE}(\hat{\bar{Y}}_{P(st)}) \text{ if}$$

$$\text{MSE}(\hat{\bar{Y}}_{P(st)}) - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0.$$

$$\frac{\bar{Y}^2 \left[\frac{(-\boldsymbol{\nu}_{020} + 2\boldsymbol{\nu}_{110} - \boldsymbol{\nu}_{200})\lambda(N_{14})}{\sqrt{\lambda\left(\frac{1}{2}\right)}} + \boldsymbol{\nu}_{200}[N_{13}] \right]}{\lambda \left[\frac{N_{14}}{\sqrt{\lambda\left(\frac{1}{2}\right)}} \right]} > 0$$

Where

$$N_{14} = (T - 4) + \boldsymbol{\nu}_{020} - 4\boldsymbol{\nu}_{200}\sqrt{\lambda\left(\frac{1}{2}\right)} + 4\sqrt{\boldsymbol{\nu}_{020}}\lambda(S_{11} + B_{11}).$$

(iv) By taking (8) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \text{MSE}(\hat{\bar{Y}}_{BT,R(st)}) \text{if}$$

$$\text{MSE}(\hat{\bar{Y}}_{BT,R(st)}) - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0$$

$$\frac{\bar{Y}^2 \left[\frac{(\frac{1}{4}\boldsymbol{\nu}_{020} + \boldsymbol{\nu}_{110} - \boldsymbol{\nu}_{200})\lambda(N_{12})}{\sqrt{\lambda\left(\frac{1}{2}\right)}} - \boldsymbol{\nu}_{200}[N_{13}] \right]}{\lambda \left[\frac{N_{12}}{\bar{y}_{(st)}\sqrt{\lambda\left(\frac{1}{2}\right)\lambda}} \right]} > 0$$

(v) By taking (9) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \text{MSE}(\hat{\bar{Y}}_{BT,P(st)}) \text{if}$$

$$\text{MSE}(\hat{\bar{Y}}_{BT,P(st)}) - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0$$

$$\frac{\bar{Y}^2 \left[\frac{\left(-\frac{1}{4}\boldsymbol{\nu}_{020} - \boldsymbol{\nu}_{110} - \boldsymbol{\nu}_{200}\right)\lambda(N_{12})}{\sqrt{\lambda\left(\frac{1}{2}\right)}} + \boldsymbol{\nu}_{200}[N_{13}] \right]}{\lambda \left[\frac{N_{12}}{\sqrt{\lambda\left(\frac{1}{2}\right)\lambda}} \right]} > 0$$

(vi) By taking (11) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \text{Var}(\hat{\bar{Y}}_{dif(st)})_{min} \text{ if}$$

$$\text{Var}(\hat{\bar{Y}}_{dif(st)})_{min} - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0$$

$$\frac{\bar{Y}^2 \left[\frac{(-\boldsymbol{\nu}_{020}\boldsymbol{\nu}_{110} + \boldsymbol{\nu}_{110})\lambda(N_{12})}{\sqrt{\lambda\left(\frac{1}{2}\right)}} + \boldsymbol{\nu}_{200}[N_{13}] \right]}{\left[\frac{N_{12}}{\sqrt{\lambda\left(\frac{1}{2}\right)}} \right]} > 0$$

(vii) By taking (13) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \text{MSE}(\hat{\bar{Y}}_{R,D(st)})_{min} \text{ if}$$

$$\text{MSE}(\hat{\bar{Y}}_{R,D(st)})_{min} - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0$$

$$\frac{\bar{X}^2 \left[\frac{(\boldsymbol{\nu}_{020}\boldsymbol{\nu}_{200} - \boldsymbol{\nu}_{110}{}^2)\lambda(N_{12})}{\sqrt{\lambda\left(\frac{1}{2}\right)}} - (\boldsymbol{\nu}_{200} + 1)\boldsymbol{\nu}_{020} - \boldsymbol{\nu}_{110}{}^2[N_{13}] \right]}{((\boldsymbol{\nu}_{200} + 1)\boldsymbol{\nu}_{020} - \boldsymbol{\nu}_{110}{}^2) \left[\frac{N_{12}}{\sqrt{\lambda\left(\frac{1}{2}\right)}} \right]} > 0$$

(viii) By taking (15) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \hat{\bar{Y}}_{Singh(st)} \text{ if}$$

$$\hat{\bar{Y}}_{Singh(st)} - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0.$$

$$\frac{\bar{Y}^2 \left[\frac{(-\frac{1}{4}\boldsymbol{v}_{020} + \boldsymbol{v}_{110} - \boldsymbol{v}_{200})\lambda(N_{12})}{\sqrt{\lambda(\frac{1}{2})\lambda}} + \boldsymbol{v}_{200}[N_{13}] \right]}{\lambda \left[\frac{N_{12}}{\sqrt{\lambda(\frac{1}{2})\lambda}} \right]} > 0$$

(ix) By taking (18) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < \hat{\bar{Y}}_{GK(st)_{min}} \text{ if}$$

$$\hat{\bar{Y}}_{GK(st)_{min}} - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0$$

$$\frac{\frac{1}{64} \left[\frac{(-X^2(16\boldsymbol{v}_{020} - 64)(-\boldsymbol{v}_{110}^2 + \boldsymbol{v}_{200} + 1)\lambda(N_{12})}{\sqrt{\lambda(\frac{1}{2})}} - \boldsymbol{v}_{200}[N_{13}] \right]}{[\boldsymbol{v}_{020}(-\boldsymbol{v}_{110}^2 + \boldsymbol{v}_{200} + 1)]\lambda \left[\frac{N_{12}}{\sqrt{\lambda(\frac{1}{2})}} \right]} > 0$$

(x) By taking (21) and (24),

$$MSE(\hat{\bar{Y}}_{ss(st)})_{min} < MSE(\hat{\bar{Y}}_{Pr_{st}})_{min} \text{ if}$$

$$MSE(\hat{\bar{Y}}_{Pr_{st}})_{min} - MSE(\hat{\bar{Y}}_{ss(st)})_{min} > 0$$

$$\frac{\bar{Y}^2 \{ 64\boldsymbol{v}_{200}(1 - R_{y.xrx}^2) - \boldsymbol{v}_{020}^2 - 16\boldsymbol{v}_{020}\boldsymbol{v}_{200}(1 - R_{y.xrx}^2) \}}{64\{1 + \boldsymbol{v}_{200}(1 - R_{y.xrx}^2)\}}$$

$$- \frac{\bar{Y}^2 \boldsymbol{v}_{200} \{ T(\lambda) - B_{11}(\lambda) + \boldsymbol{v}_{020} + 4(R_{yx}^2 - 1) \}}{\left\{ \lambda \left(4\frac{\boldsymbol{v}_{200}}{\lambda}(R_{yx}^2 - 1) + 4\frac{\boldsymbol{v}_{020}^{1/2}}{\lambda^{1/2}}(S_{11} + B_{11}) + T + \frac{\boldsymbol{v}_{020}}{\lambda} - 4 \right) \right\}} > 0.$$

5. Numerical investigation

In this portion, we run a numerical test to see how well the existing and suggested estimator performed; two populations are considered for this purpose. To get the PREs, we utilize the

Table 1. Summary statistics of Population-I.

H	N_h	n_h	ω_h	λ_h	$\bar{Y}_h$	S^2_{yh}	C^2_{yh}	β_{yh}
C 1	106	9	0.1241	0.1017	1536.774	41281746	17.479	79.340
2	106	17	0.1241	0.0494	2212.594	133437791	27.256	96.075
3	94	38	0.1100	0.0157	9384.309	894457433	10.156	25.809
4	171	67	0.2002	0.0090	5588.012	820445636	26.274	101.979
5	204	7	0.2389	0.1379	966.955	5710999	6.107	55.051
6	173	2	0.2026	0.4942	404.398	894440.3	5.469	30.119
$\bar{X}_h$	S^2_{xh}	C^2_{xh}	β_{xh}	S_{xh}	$\bar{R}_{xh}$	S^2_{rxh}	C^2_{rxh}	β_{rxh}
24711.81	2414224935	3.95	27.46	49134.7	53.5	945.15	0.33	27.46
26840.04	2913701588	4.04	31.98	53978.7	53.5	945.04	0.33	31.98
72723.76	25956279019	4.90	27.40	161109.5	47.5	744.13	0.32	27.40
73191.2	68903936687	12.86	86.30	262495.6	86	2450.96	0.33	86.30
26833.75	2040714047	2.83	29.72	45174.2	102.5	3484.90	0.33	29.72
9903.30	360137210	3.67	31.87	18977.2	87	2508.36	0.33	31.87

below expression:

$$\text{PRE}(.) = \frac{Var(\hat{\bar{Y}}_{(st)})}{MSE(\hat{\bar{Y}}_{i(st)})} * 100,$$

where i $= (\hat{\bar{Y}}_R, \hat{\bar{Y}}_P, \hat{\bar{Y}}_{BT,R}, \hat{\bar{Y}}_{BT,P}, \hat{\bar{Y}}_{dif}, \hat{\bar{Y}}_{R,D}, \hat{\bar{Y}}_{Singh}, \hat{\bar{Y}}_{Gk}, \hat{\bar{Y}}_{Pr_{st}}, \hat{\bar{Y}}_{ss})$.

6. Data discription

To show the efficiancy of proposed estimator over the existing estimators, we conduct a numerical study to investigate the performances of the propose and existing estimator. For this purpose, we used two real data sets of Kadilar and Cingi (2003), summary statistics given in Tables 1 and 2 for the population-I while MSE and PRE are presented in Table 3, similarly summary statistics for population-II given in Tables 4 and 5 while MSE and PRE are presented in Table 6. For population-I study variable is apple production in 1999, and the auxiliary variable is the number of apple trees in 1999. Similarly for population-II the study variable is apple production in 1999, and the auxiliary variable is the number of apple timber in 1998. (Source: Institute of Statistics, Republic of Turkey). We have stratified the data by regions of Turkey

Table 2. Correlation and covariance's using Population I.

$\rho_{y_h x_h}$	$S_{y_h x_h}$	$\rho_{y_h r x_h}$	$S_{y_h r x_h}$	$\rho_{x_h r x_h}$	$S_{x_h r x_h}$
0.815688	257508714	0.3346206	66097.3	0.5956644	899792.9
0.835927	521230984	0.2814818	99957.52	0.6246736	1036577
0.897193	4323018145	0.4626493	377448.3	0.5893881	2590289
0.981495	7379640297	0.2979733	422543.1	0.3885082	5048833
0.710732	76727855	0.4541102	64063.84	0.6317101	1684630
0.869703	15609193	0.5366774	25420.49	0.6283556	597220.2

Table 3. MSE and PRE using Population I.

Estimators	MSE	PRE
$\hat{\bar{Y}}_{(st)}$	697789.6	100
$\hat{\bar{Y}}_{R(st)}$	243316.9	286.78
$\hat{\bar{Y}}_{P(st)}$	1878241	37.15
$\hat{\bar{Y}}_{BT,R(st)}$	379805.9	183.72
$\hat{\bar{Y}}_{BT,P(st)}$	1197268	58.28
$\hat{\bar{Y}}_{dif(st)}$	237552.8	293.74
$\hat{\bar{Y}}_{R,D(st)}$	231157	301.87
$\hat{\bar{Y}}_{Singh(st)}$	379805.9	183.72
$\hat{\bar{Y}}_{GK(st)}$	228480.4	305.4
$\hat{\bar{Y}}_{Pr_{st}}$	207876.3	335.68
$\hat{\bar{Y}}_{ss(st)}$	194803.2	358.2

https://doi.org/10.1371/journal.pone.0275875.t003

such as (1: Marmara, 2: Agean, 3: Miditerranean, 4: Central Anatolia, 5: Black sea, 6: East and Southeast Anatolia) and from each stratum, we have randomly selected the samples whose sizes are computed by using the Neyman allocation method.

Population I: (Source: Kadilar and Cingi [18])
Y is the crop of apples in 1999, and X is the number of apples timber in 1999.
Population 2: (Source: Kadilar and Cingi [18])
Y is the crop of apples in 1999, and X is the crop of apples trees in 1998.

Table 4. Summary statistics of Population II.

H	$\mathbb{N}_h$	n_h	ω_h	λ_h	$\bar{Y}_h$	S^2_{yh}	C^2_{yh}	β_{yh}
1	106	9	0.1241	0.1017	1536.774	41281746	17.479	79.340
2	106	17	0.1241	0.0494	2212.594	133437791	27.256	96.075
3	94	38	0.1100	0.0157	9384.309	894457433	10.156	25.809
4	171	67	0.2002	0.0090	5588.012	820445636	26.274	101.979
5	204	7	0.2389	0.1379	966.955	5710999	6.107	55.051
6	173	2	0.2026	0.4942	404.398	894440.3	5.469	30.119
$\bar{X}_h$	S^2_{xh}	C^2_{xh}	β_{xh}	S_{xh}	$\bar{R}_{xh}$	S^2_{rxh}	C^2_{rxh}	β_{rxh}
24375.59	2419565835	4.0	27.45	49189.0	53.5	945.142	0.3302	27.45607
27421.70	3301722268	4.3	35.90	57460.6	53.5	945.057	0.3301	35.90683
72409.95	25842911825	4.9	27.70	160757.3	47.5	744.139	0.3298	27.7032
74364.68	81569146488	14.7	97.73	285603.1	86	2450.935	0.3313	97.7351
26441.72	2061412416	2.9	29.77	45402.7	102.5	3484.892	0.3316	29.77505
9843.82	353212773	3.6	30.26	18793.9	87	2508.387	0.3314	30.26428

https://doi.org/10.1371/journal.pone.0275875.t004

Table 5. Correlation and covariance's using Population II.

$\rho_{y_h x_h}$	$S_{y_h x_h}$	$\rho_{y_h r x_h}$	$S_{y_h r x_h}$	$\rho_{x_h r x_h}$	$S_{x_h r x_h}$
0.8156414	257778692	0.3349042	66152.81	0.593033	896801.5
0.8559991	568176176	0.281644	100015.9	0.6031022	1065345
0.9011201	4332446622	0.4637681	378362.4	0.5873879	2575864
0.9858761	8065108356	0.2981778	422830.3	0.3654157	5166734
0.7130988	77372777	0.4547748	64157.44	0.6206547	1663517
0.8935989	15883145	0.5435647	25746.85	0.6262896	589509.3

7. Simulation study

The efficiency of proposed estimators over competing estimators was demonstrated clearly in the preceding section. A Monte Carlo simulation analysis with R software is also used to assess the efficiency of the proposed estimator using dual ancillary variable under stratified random sampling. The assessment of proposed estimator with existing estimators is illustrated using the percentage relative efficiency (PRE) formula. Yet again, the real population of Kadilar and Cingi [18] is used. The following steps are used in R-Language software to conduct the simulation study.

1. We considered different sample sizes (64, 80, 96, 128, 144, 160, 176, 192, 208, and 240) through the proportional allocation method.

2. With stratified sampling, the technique is repeated 100,000 times and the population is divided into six strata to calculate the numerous values of proposed and existing estimators.

3. The 100,000 values of existing estimators and proposed estimator, as well as their corresponding variances, are computed using the samples obtained.

4. The values of percentage relative efficiency (PRE) are derived using the values of variances of all existing and proposed estimator and provided in Table 7.

The consequence of the above results, the performance of the proposed estimator is the best among all the existing estimators under consideration.

Table 6. MSE and PRE using Population II.

Estimators	MSE	PRE
$\hat{\bar{Y}}_{(st)}$	697789.6	100
$\hat{\bar{Y}}_{R(st)}$	196709.7	354.73
$\hat{\bar{Y}}_{P(st)}$	1924848	36.25
$\hat{\bar{Y}}_{BT,R(st)}$	356502.2	195.73
$\hat{\bar{Y}}_{BT,P(st)}$	1220572	57.17
$\hat{\bar{Y}}_{dif(st)}$	183576.4	380.11
$\hat{\bar{Y}}_{R,D(st)}$	179733.4	388.24
$\hat{\bar{Y}}_{Singh(st)}$	356502.2	195.73
$\hat{\bar{Y}}_{GK(st)}$	177598.9	392.9
$\hat{\bar{Y}}_{Pr_{st}}$	152565.6	457.37
$\hat{\bar{Y}}_{ss(st)}$	140498.5	496.65

Table 7. Simulation results for the PRE of recommended estimator w.r.t the existing estimators by different sample sizes.

Estimator	Sample Size									
	64	80	96	128	144	160	176	192	208	240
$\hat{\bar{Y}}_{(st)}$	100	100	100	100	100	100	100	100	100	100
$\hat{\bar{Y}}_{R(st)}$	375.87	380.11	387.46	389.17	385.078	380.028	390.00	410.00	425.00	542.001
$\hat{\bar{Y}}_{P(st)}$	45.87	48.50	49.73	51.43	59.56	69.164	60.850	66.24	77.53	88.34
$\hat{\bar{Y}}_{BT,R(st)}$	199.87	210.98	215.02	216.15	223.026	225.65	226.25	228.11	230.024	234.371
$\hat{\bar{Y}}_{BT,P(st)}$	60.87	62.15	63.44	66.74	67.026	68.65	69.25	69.11	72.024	75.371
$\hat{\bar{Y}}_{dif(st)}$	382.87	398.11	399.12	400.46	410.028	416.00	419.00	437.001	467.00	489.00
$\hat{\bar{Y}}_{R,D(st)}$	390.87	395.50	399.73	415.850	416.24	419.53	426.34	428.15,	460.29	499.43
$\hat{\bar{Y}}_{Singh(st)}$	199.87	210.98	215.02	216.15	223.026	225.65	226.25	228.11	230.024	234.371
$\hat{\bar{Y}}_{GK(st)}$	399.49	411.76	423.76	445.56	456.78	459.12	467.23	489.78	490.34	499.89
$\hat{\bar{Y}}_{Pr_{st}}$	470.87	479.11	,489.12	498.46	499.17	494.078	512.028	517.00	519.00	520.001
$\hat{\bar{Y}}_{ss(st)}$	518.87	520.50	523.73	565.43	586.56	596.16	598.850	599.24	605.53	647.34

https://doi.org/10.1371/journal.pone.0275875.t007

Advantages of Monte Carlo simulation

The basis of a Monte Carlo simulation is that the probability of varying outcomes cannot be determined because of random variable interference. Therefore, a Monte Carlo simulation focuses on constantly repeating random samples to achieve certain results.

A Monte Carlo simulation takes variable that has uncertainty and assigns it a random value. The model is then run, and a result is provided. The process is repeated again and again while assigning the variable is question with many different values. Once the simulation is complete, the results are averaged together to provide an estimate.

8. Discussion

To evaluate the advantage of our propose estimator under stratified random sampling, we use two real data sets for numerical comparision. On the basis of numerical results, which are presented in Tables 3 and 6, it is observed that the proposed ratio-in-regression type estimator are more efficient than the usual sample mean estimator, Cochran [24], Murthy [25], Bahl and Tuteja [26], difference estimator, Rao [27], Singh et al. [28], Grover and Kaur [29], Ahmad et al. [30].

Table 7 gives simulation results for the percentage relative efficiency of proposed estimator w.r.t the existing estimators by using different sample sizes i.e: 64, 80, 96, 128, 144, 160, 176, 192, 208, and 240. The value of percentage relative efficiency differs depending on the sample size. From the simulation results, it is also observed that the proposed estimator is more efficient than the existing counterparts, in terms of percentage relative efficiency. As we increase the sample size, the efficiency of our proposed estimator is also increased. Overall, the gain in efficiency of our proposed estimator is the best as compared to all existing counterparts.

For visualization, the comparison of proposed estimator with existing estimators in terms of percentage relative efficiency are presented in Fig 1. The length of a line graph is directly associated with the efficiency of an estimator. More specifically, the higher the length of a line graph, efficient the estimator. In general, we recommend using our proposed estimate for the new survey instead of the existing estimator examined in this paper for estimating the finite population mean under stratified random sampling.
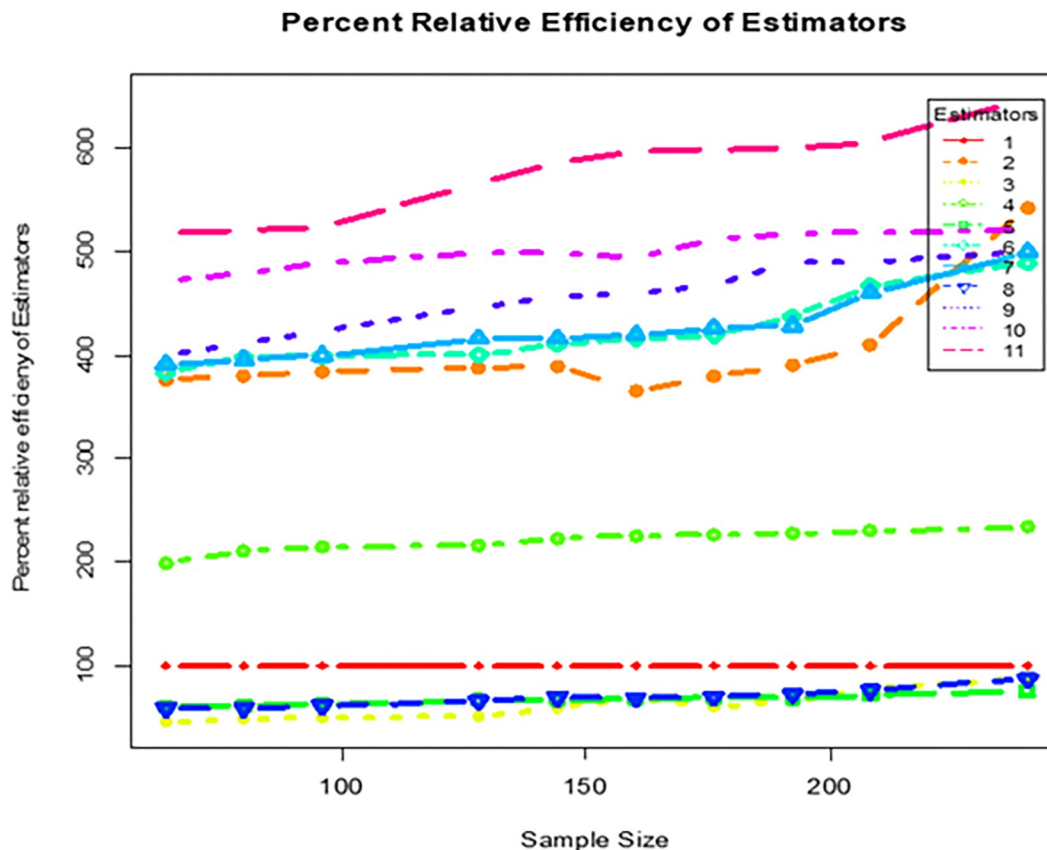
Fig 1. Percentage relative efficiency of different estimators.

9. Conclusion

In this article, we propose ratio-in-regression type exponential estimator for the finite population mean under stratified random sampling, which required an ancillary variable on the sample mean and rank of the ancillary varaible. Expressions for mean square error of the proposed estimator are derived up to first order of approximation and comparison is made with the estimators mentioned herein. According to results of real data sets, it is perceived that the proposed estimator performs well as compared to its existing counterpart. A simulation analysis is also carried out to assess the robustness and generalizability of the propose estimator. The simulation study's findings also confirm the utility of the proposed estimator. A numerical study is carried out to support the theoretical results. Therefore we recommend the use of proposed estimators for efficiently estimating the finite population mean under stratified random sampling. The current work can be extended to develop an improved class of estimators under two-phase, non-response, two-stage, and cumulative distribution function sampling scheme using information on ancillary variable for estimating the population mean under simple and stratified random sampling.

Author Contributions

Conceptualization: Sohaib Ahmad, Muhammad Aamir.

Formal analysis: Sardar Hussain, Muhammad Aamir.

References

1. Zaman T, Kadilar C. Exponential ratio and product type estimators of the mean in stratified two-phase sampling. AIMS Mathematics. 2021 Jan 1; 6(5):4265–79.

2. Zaman T, Kadilar C. On estimating the population mean using auxiliary character in stratified random sampling. Journal of Statistics and Management Systems. 2020 Nov 16; 23(8):1415–26.

3. Zaman T. Efficient estimators of population mean using auxiliary attribute in stratified random sampling. Advances and Applications in Statistics. 2019; 56(2):153–71.

4. Zaman T. An efficient exponential estimator of the mean under stratified random sampling. Mathematical Population Studies. 2021 Apr 3; 28(2):104–21.

5. Rather KU, Kadilar C. Dual to Ratio cum Product Type of Exponential Estimator for Population Mean in Stratified Random Sampling.

6. Mradula Yadav SK, Varshney R, Dube M. Efficient estimation of population mean under stratified random sampling with linear cost function. Communications in Statistics-Simulation and Computation. 2021 Dec 2; 50(12):4364–87.

7. Javed M, Irfan M, Bhatti SH, Onyango R. A Simulation-Based Study for Progressive Estimation of Population Mean through Traditional and Nontraditional Measures in Stratified Random Sampling. Journal of Mathematics. 2021 Dec 20; 2021.

8. Javed M, Irfan M. A simulation study: new optimal estimato rs for population mean by using dual auxiliary information in stratified random sampling. Journal of Taibah University for Science. 2020 Jan 1; 14 (1):557–68.

9. Yadav R, Tailor R. Estimation of finite population mean using two auxiliary variables under stratified random sampling. Statistics in Transition New Series. 2020 Mar 18; 21(1).

10. Zaman T, Bulut H. Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling. Communications in Statistics-Theory and Methods. 2020 Jul 17; 49(14):3407–20.

11. Kumar M, Vishwakarma GK. Generalized Classes of Regression-Cum-Ratio Estimators of Population Mean in Stratified Random Sampling. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences. 2020 Dec; 90(5):933–9.

12. Zahid E, Shabbir J, Alamri OA. A generalized class of estimators for sensitive variable in the presence of measurement error and non-response under stratified random sampling. Journal of King Saud University-Science. 2022 Feb 1; 34(2):101741.

13. Aladag S, Cingi H. Improvement in estimating the population median in simple random sampling and stratified random sampling using auxiliary information. Communications in Statistics-Theory and Methods. 2015 Mar 4; 44(5):1013–32.

14. Grover LK, Kaur P. A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. Communications in Statistics-Simulation and Computation. 2014 Jan 1; 43(7):1552–74.

15. Shabbir J, Gupta S. Estimation of finite population mean in simple and stratified random sampling using two auxiliary variables. Communications in Statistics-Theory and Methods. 2017 Oct 18; 46 (20):10135–48.

16. Singh GN, Khalid M. Exponential chain dual to ratio and regression type estimators of population mean in two-phase sampling. Statistica. 2015 Dec 30; 75(4):379–89.

17. Singh GN, Khalid M. Efficient class of estimators for finite population mean using auxiliary information in two-occasion successive sampling. Journal of Modern Applied Statistical Methods. 2019; 17(2):14.

18. Kadilar C., & Cingi H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 45(2), 218–225.

19. Kadilar C, Cingi H. A new ratio estimator in stratified random sampling. Communications in Statistics—Theory and Methods. 2005 Mar 1; 34(3):597–602.

20. Koyuncu N, Kadilar C. Ratio and product estimators in stratified random sampling. Journal of statistical planning and inference. 2009 Aug 1; 139(8):2552–8.

21. Koyuncu N, Kadilar C. On improvement in estimating population mean in stratified random sampling. Journal of Applied Statistics. 2010 Jun 1; 37(6):999–1013.

22. Singh HP, Vishwakarma GK. A family of estimators of population mean using auxiliary information in stratified sampling. Communications in Statistics—Theory and Methods. 2008 Feb 13; 37(7):1038–50.

23. Hussain S, Ahmad S, Saleem M, Akhtar S. Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. Plos one. 2020 Sep 28; 15(9): e0239098. https://doi.org/10.1371/journal.pone.0239098 PMID: 32986764

24. Cochran WG. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. The journal of agricultural science. 1940 Apr; 30(2):262–75.

25. Murthy MN. Product method of estimation. Sankhyā: The Indian Journal of Statistics, Series A. 1964 Jul 1:69–74.

26. Bahl S, Tuteja R. Ratio and product type exponential estimators. Journal of information and optimization sciences. 1991 Jan 1; 12(1):159–64.

27. Rao TJ. On certait methods of improving ration and regression estimators. Communications in Statistics-Theory and Methods. 1991 Jan 1; 20(10):3325–40.

28. Singh R, Chauhan P, Sawan N, Smarandache F. Improvement in estimating the population mean using exponential estimator in simple random sampling. Auxiliary Information and a priori Values in Construction of Improved Estimators. 2007; 33.

29. Grover LK, Kaur P. Ratio type exponential estimators of population mean under linear transformation of auxiliary variable: theory and methods. South African Statistical Journal. 2011 Jan 1; 45(2):205–30.

30. Ahmad S, Hussain S, Aamir M, Yasmeen U, Shabbir J, Ahmad Z. Dual use of auxiliary information for estimating the finite population mean under the stratified random sampling scheme. Journal of Mathematics. 2021 Nov 16;2021.