



A new estimator for mean under stratified random sampling

Usman Shahzad¹ · Muhammad Hanif¹ · Nursel Koyuncu²

Received: 1 March 2018 / Accepted: 14 July 2018 / Published online: 23 July 2018
© The Author(s) 2018

Abstract

In this paper, we have proposed an estimator of finite population mean in stratified random sampling. The expressions for the bias and mean square error of the proposed estimator are obtained up to the first order of approximation. It is found that the proposed estimator is more efficient than the traditional mean, ratio, exponential, regression, Shabbir and Gupta (in *Commun Stat Theory Method* 40:199–212, 2011) and Khan et al. (in *Pak J Stat* 31:353–362, 2015) estimators. We have utilized four natural and four artificial data sets under stratified random sampling scheme for assessing the performance of all the estimators considered here.

Keywords Bias · Efficiency · Mean square error · Stratified random sampling

Mathematics Subject Classification 62D05 · 62D99

Introduction

Stratification is a designing tool that is used in modern surveys for improving the precision of estimates. In stratified design, the whole population is divided into number of strata for getting homogeneity within each stratum, and samples are selected within each stratum mostly through simple random sampling. There are numerous authors who have suggested different estimators by utilizing some known population parameters of an auxiliary variable. Sisodia and Dwivedi [11] presented a proportion estimator by utilizing coefficient of variety of an auxiliary variable. Singh and Kakran [9] proposed another proportion estimator by utilizing known coefficient of kurtosis of an auxiliary variable. Upadhyaya and Singh [12] likewise examined a ratio-type estimator by using the linear

combination of coefficient of variation and kurtosis of an auxiliary variable. Kadilar and Cingi [2] utilized the stratified forms of these specified estimators keeping in mind the end goal to improve the efficiency of the suggested estimators.

The fundamental goal of this paper is to propose an improved estimator of the finite population mean utilizing data on an auxiliary variable in stratified random sampling. The expressions for the bias and mean square error (MSE) of the proposed estimator are inferred up to the first order of approximation. On the bases of theoretical and numerical comparisons, we demonstrate that the proposed estimator is more efficient than existing estimators.

The rest of the paper is organized as follows: “[Some existing estimators](#)” section consists in the estimators which we reviewed from the literature, and also some useful preliminaries results for obtaining the properties of proposed and existing estimators are available here. “[Proposed estimator](#)” section introduces an improved estimator using stratified random sampling scheme. “[Numerical illustration](#)” section is devoted to the efficiency comparison. A numerical evaluation is presented in “[Conclusion](#)” section to highlight the contribution of the paper.

✉ Usman Shahzad
usman.stat@yahoo.com

Muhammad Hanif
mhpuno@hotmail.com

Nursel Koyuncu
nkoyuncu@hacettepe.edu.tr

¹ Department of Mathematics and Statistics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan

² Department of Statistics, Hacettepe University, Ankara, Turkey

Some existing estimators

Suppose $U = \{U'_1, U'_2, \dots, U'_N\}$ be the population containing N finite units. Let X be an auxiliary and Y be the study variable taking values y_{hi} and x_{hi} in the unit ($i = 1, 2, \dots, N$) in the h th stratum consisting of N_h units such that $\sum_{h=1}^L N_h = N$. Let n_h be the size of the sample drawn from the h th stratum by using simple random sampling without replacement scheme such that $\sum_{h=1}^L n_h = n$. Suppose $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$, where $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$. Let $\bar{x}_{st} = \sum_{h=1}^L W_h \bar{x}_h$, where $\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$ and $W_h = \frac{N_h}{N}$ is the stratum weight. The expressions for Y are also defined in similar way.

To find the MSE of the proposed and existing estimators, let us define

$$e_{0st} = \frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}}, \quad e_{1st} = \frac{\bar{x}_{st} - \bar{X}}{\bar{X}}.$$

The expectations of e terms are given below

$$E(e_{0st}) = 0, E(e_{1st}) = 0, \\ E(e_{0st}^2) = \sum_{h=1}^L W_h^2 f'_h \frac{S_{yh}^2}{\bar{Y}^2} = V_{2.0}, E(e_{1st}^2) \\ = \sum_{h=1}^L W_h^2 f'_h \frac{S_{xh}^2}{\bar{X}^2} = V_{0.2},$$

$$E(e_{0st} e_{1st}) = \sum_{h=1}^L W_h^2 f'_h \frac{S_{yxh}}{\bar{X}\bar{Y}} = V_{1.1}$$

where

$$f_h = \frac{n_h}{N_h}, f'_h = \left(\frac{1 - f_h}{n_h} \right),$$

and

$$V_{a.b} = \sum_{h=1}^L W_h^{a+b} f'_h \sum_{i=1}^{N_h} \frac{(y_{hi} - \bar{Y}_h)^a (x_{hi} - \bar{X}_h)^b}{\bar{Y}^a \bar{X}^b}.$$

The variance of the sample mean in stratified random sampling without replacement is given by:

$$\text{Var}(\bar{y}_{st}) = \bar{Y}^2 V_{2.0}.$$

The stratified version of classical ratio estimator for mean is given by

$$\hat{y}_{Rst} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X}. \tag{1}$$

The bias and MSE of classical ratio estimator given in (1) up to the first order of approximation are given below:

$$\text{Bias}(\hat{y}_{Rst}) = \bar{Y}[V_{0.2} - V_{1.1}] \\ \text{MSE}(\hat{y}_{Rst}) = \bar{Y}^2[V_{2.0} + V_{0.2} - 2V_{1.1}]. \tag{2}$$

The stratified version of Bahl and Tuteja [1] estimator is

$$\hat{y}_{BTst} = \bar{y}_{st} \exp \left[\frac{\bar{X} - \bar{x}_{st}}{\bar{X} + \bar{x}_{st}} \right]. \tag{3}$$

The bias and MSE of \hat{y}_{BTst} up to the first order of approximation are given below:

$$\text{Bias}(\hat{y}_{BTst}) = \bar{Y} \left[\frac{3}{8} V_{0.2} - \frac{1}{2} V_{1.1} \right] \\ \text{MSE}(\hat{y}_{BTst}) = \bar{Y}^2 \left[V_{2.0} + \frac{1}{4} V_{0.2} - V_{1.1} \right]. \tag{4}$$

The traditional regression estimator is given by

$$\hat{y}_{Reg(st)} = \bar{y}_{st} + b_{st}(\bar{X} - \bar{x}_{st}).$$

The MSE of $\hat{y}_{Reg(st)}$ is,

$$\text{MSE}(\hat{y}_{Reg(st)}) = \bar{Y}^2 V_{2.0} (1 - \rho_{st}^2), \tag{5}$$

where $\rho_{st} = \frac{V_{1.1}}{\sqrt{V_{2.0} V_{0.2}}}$ is the combined correlation between study and auxiliary variate.

Shabbir and Gupta [8] introduced an exponential ratio estimator as:

$$\hat{y}_{SGst} = [w_1^{SG} \bar{y}_{st} + w_2^{SG} (\bar{X} - \bar{x}_{st})] \exp \left(\frac{\bar{A} - \bar{a}_{st}}{\bar{A} + \bar{a}_{st}} \right), \tag{6}$$

where $\bar{a} = \bar{x}_{st} + N\bar{X}$, $\bar{A} = \bar{X} + N\bar{X}$.

The optimum values of w_1^{SG} and w_2^{SG} are

$$w_1^{SG(\text{opt})} = \frac{B_{SG} D_{SG} - \frac{C_{SG} E_{SG}}{2}}{A_{SG} B_{SG} - C_{SG}^2} \text{ and } w_2^{SG(\text{opt})} = \frac{-C_{SG} D_{SG} + \frac{A_{SG} E_{SG}}{2}}{A_{SG} B_{SG} - C_{SG}^2}.$$

The minimum MSE of \hat{y}_{SG} is

$$\text{MSE}_{\min}(\hat{y}_{SGst}) = \bar{Y}^2 - \frac{B_{SG} D_{SG}^2 + \frac{A_{SG} E_{SG}^2}{4} - C_{SG} D_{SG} E_{SG}}{A_{SG} B_{SG} - C_{SG}^2}, \tag{7}$$

where

$$A_{SG} = \bar{Y}^2 \left[1 + V_{2.0} + \frac{V_{0.2}}{(1+N)^2} - \frac{2V_{1.1}}{(1+N)} \right], \\ B_{SG} = \bar{X}^2 V_{0.2}, C_{SG} = \bar{X}\bar{Y} \left[\frac{V_{0.2}}{(1+N)} - V_{1.1} \right], \\ D_{SG} = \bar{Y}^2 \left[1 + \frac{3V_{0.2}}{8(1+N)^2} - \frac{V_{1.1}}{2(1+N)} \right] \text{ and } E_{SG} = \bar{X}\bar{Y} \frac{V_{0.2}}{(1+N)}.$$

Khan et al. [3] proposed the following ratio estimator in simple random sampling:

$$\hat{y}_{Kalst} = [w_1^{Kal} \bar{y} + w_2^{Kal} (\bar{X} - \bar{x})] \exp \left(\frac{\bar{A}' - \bar{a}'}{\bar{A}' + \bar{a}'} \right), \tag{8}$$

In stratified random sampling, it will be of the form:

$$\hat{y}_{Kal(st)} = [w_1^{Kal} \bar{y}_{st} + w_2^{Kal} (\bar{X} - \bar{x}_{st})] \exp\left(\frac{\bar{A}' - \bar{a}'_{st}}{\bar{A}' + \bar{a}'_{st}}\right), \quad (9)$$

where

$$\bar{a}'_{st} = \bar{X}c_{ki}, \bar{A} = \bar{x}_{st} + \bar{X}(c_{ki} - 1), \quad \text{for } i = 1, 2, 3.$$

$$c_{k1} = \rho_{st} + 1, c_{k2} = \frac{\rho_{st} + 1}{2}, c_{k3} = \frac{\rho_{st} + 1}{3}.$$

The optimum values of w_1^{Kal} and w_2^{Kal} are

$$w_1^{Kal(opt)} = \frac{B_{Kal}D_{Kal} - \frac{C_{Kal}E_{Kal}}{2}}{A_{Kal}B_{Kal} - C_{Kal}^2} \text{ and } w_2^{Kal(opt)} = \frac{-C_{Kal}D_{Kal} + \frac{A_{Kal}E_{Kal}}{2}}{A_{Kal}B_{Kal} - C_{Kal}^2}.$$

The minimum MSE of $\hat{y}_{Kal(st)}$ is

$$MSE_{\min}(\hat{y}_{Kal(st)}) = \bar{Y}^2 - \frac{B_{Kal}D_{Kal}^2 + \frac{A_{Kal}E_{Kal}^2}{4} - C_{Kal}D_{Kal}E_{Kal}}{A_{Kal}B_{Kal} - C_{Kal}^2}, \quad (10)$$

where

$$A_{Kal} = \bar{Y}^2 \left[1 + V_{2.0} + \frac{V_{0.2}}{c_{ki}^2} - \frac{2V_{1.1}}{c_{ki}} \right],$$

$$B_{Kal} = \bar{X}^2 V_{0.2}, C_{Kal} = \bar{X}\bar{Y} \left[\frac{V_{0.2}}{c_{ki}} - V_{1.1} \right],$$

$$D_{Kal} = \bar{Y}^2 \left[1 + \frac{3V_{0.2}}{8c_{ki}^2} - \frac{V_{1.1}}{2c_{ki}} \right] \text{ and } E_{Kal} = \bar{X}\bar{Y} \frac{V_{0.2}}{c_{ki}}.$$

Proposed estimator

In this section, an improved estimator of finite population mean in stratified random sampling is proposed. The properties of the proposed estimator are studied up to the first order of approximation. Development of the proposed estimator is given step-by-step below.

Rao [7] introduced the following estimator in simple random sampling

$$\hat{y}_{Rao} = [w_1(\bar{X} - \bar{x}) + w_2\bar{y}] \quad (11)$$

and its stratified version can be written as

$$\hat{y}_{Rao(st)} = [w_1(\bar{X} - \bar{x}_{st}) + w_2\bar{y}_{st}]. \quad (12)$$

As we know, the average of Bahl and Tuteja [1] estimators is

$$\hat{y}_{ABT} = \frac{1}{2} \left\{ \bar{y}_{st} \exp\left(\frac{\bar{X} - \bar{x}_{st}}{\bar{X} + \bar{x}_{st}}\right) + \bar{y}_{st} \exp\left(\frac{\bar{x}_{st} - \bar{X}}{\bar{X} + \bar{x}_{st}}\right) \right\}, \quad (13)$$

Now, by adding (12) and (13), one can propose the following estimator

$$\hat{y}_{AA} = \left[\frac{1}{2} \left\{ \bar{y}_{st} \exp\left(\frac{\bar{X} - \bar{x}_{st}}{\bar{X} + \bar{x}_{st}}\right) + \bar{y}_{st} \exp\left(\frac{\bar{x}_{st} - \bar{X}}{\bar{X} + \bar{x}_{st}}\right) \right\} + w_1(\bar{X} - \bar{x}_{st}) + w_2\bar{y}_{st} \right],$$

Hence, taking motivation from \hat{y}_{AA} & \hat{y}_{BTst} , we propose the following estimator

$$\hat{y}_{Nst} = \left[\frac{1}{2} \left\{ \bar{y}_{st} \exp\left(\frac{\bar{X} - \bar{x}_{st}}{\bar{X} + \bar{x}_{st}}\right) + \bar{y}_{st} \exp\left(\frac{\bar{x}_{st} - \bar{X}}{\bar{X} + \bar{x}_{st}}\right) \right\} + w_1(\bar{X} - \bar{x}_{st}) + w_2\bar{y}_{st} \right] \exp\left[\frac{\bar{X}'' - \bar{x}_{st}''}{\bar{X}'' + \bar{x}_{st}''}\right], \quad (14)$$

where $\bar{X}'' = \bar{X}k$ and $\bar{x}_{st}'' = \bar{x}_{st} + \bar{X}(k - 1)$, (see [3]). Further, $k = \frac{\rho_{st} + 1}{4}$, suitably chosen constant.

The bias, MSE and minimum MSE of \hat{y}_{Nst} are given by,

$$\text{Bias}(\hat{y}_{Nst}) = \bar{Y} \left\{ \frac{V_{0.2}}{8} \left(1 + \frac{3}{k^2} \right) - \frac{V_{1.1}}{2k} \right\} + \frac{\bar{X}V_{0.2}}{2k} w_1 + \bar{Y} \left\{ \frac{3V_{0.2}}{8k^2} - \frac{V_{1.1}}{2k} w_2 \right\},$$

$$MSE(\hat{y}_{Nst}) = \bar{Y}^2 L + w_1^2 \lambda_A + w_2^2 \lambda_B + 2w_1 w_2 \lambda_C - 2w_1 \lambda_D - w_2 \lambda_E,$$

$$MSE_{\min}(\hat{y}_{Nst}) = \left[\bar{Y}^2 L - \frac{\lambda_B \lambda_D^2 + \frac{\lambda_A \lambda_E^2}{4} - \lambda_C \lambda_D \lambda_E}{\lambda_A \lambda_B - \lambda_C^2} \right].$$

The detailed proofs of the above expressions are provided in Appendix.

Numerical illustration

Real data sets

To investigate the theoretical results, the following real data sets are considered as:

Population 1 [10, p. 219]

X = Amount of milky cows in the year 1990 and Y = Amount of milky cows in the year 1993.

$N_{1st} = 7$	$N_{2st} = 12$	$N_{3st} = 5$
$n_{1st} = 3$	$n_{2st} = 5$	$n_{3st} = 2$
$W_{1st} = 0.2916$	$W_{2st} = 0.5000$	$W_{3st} = 0.2083$
$\bar{X}_{1st} = 15.2857$	$\bar{X}_{2st} = 17.2500$	$\bar{X}_{3st} = 20.6000$
$\bar{Y}_{1st} = 17.4285$	$\bar{Y}_{2st} = 20.4166$	$\bar{Y}_{3st} = 17.8000$
$S_{x1st} = 4.5721$	$S_{x2st} = 5.4958$	$S_{x3st} = 3.6469$
$S_{y1st} = 4.1975$	$S_{y2st} = 4.0778$	$S_{y3st} = 3.2710$
$\rho_{st} = 0.29$	$n = 10$	$N = 24$

Population 2 [6, p. 228]

X = Amount of workers working in a factor and Y = Yield or output for factories in an area.

$N_{1st} = 25$	$N_{2st} = 23$	$N_{3st} = 16$	$N_{4st} = 16$
$n_{1st} = 14$	$n_{2st} = 13$	$n_{3st} = 9$	$n_{4st} = 9$
$W_{1st} = 0.3125$	$W_{2st} = 0.2875$	$W_{3st} = 0.2000$	$W_{4st} = 0.2000$
$\bar{X}_{1st} = 71.00$	$\bar{X}_{2st} = 140.69$	$\bar{X}_{3st} = 362.93$	$\bar{X}_{4st} = 749.50$
$\bar{Y}_{1st} = 3156.64$	$\bar{Y}_{2st} = 4766.21$	$\bar{Y}_{3st} = 6334.18$	$\bar{Y}_{4st} = 7795.31$
$S_{x1st} = 14.6116$	$S_{x2st} = 28.0364$	$S_{x3st} = 91.3823$	$S_{x4st} = 174.46$
$S_{y1st} = 740.01$	$S_{y2st} = 515.69$	$S_{y3st} = 501.39$	$S_{y4st} = 653.09$
$n = 45$	$N = 80$	$\rho_{st} = 0.67$	

Strata are formed by grouping randomly into four strata on the premise of auxiliary variate (X). The criteria of constructions are $x < 100.0$, $100.0 \geq x < 200.0$, $200.0 \geq x < 500.0$ and $x \geq 500.0$ respectively. Proportional allocation is used for selecting sample from each stratum. Note that we use $n = 45$.

Population 3 ([2])

X = Apple trees amount in $N = 854$ towns in Turkey in (1999) and Y = Level of apple production.

$N_{1st} = 106$	$N_{2st} = 106$	$N_{3st} = 94$	$N_{4st} = 171$	$N_{5st} = 204$
$N_{6st} = 173$	$n_{1st} = 9$	$n_{2st} = 17$	$n_{3st} = 38$	$n_{4st} = 67$
$n_{5st} = 7$	$n_{6st} = 2$	$W_{1st} = 0.1241$	$W_{2st} = 0.1241$	$W_{3st} = 0.1101$
$W_{4st} = 0.2002$	$W_{5st} = 0.2386$	$W_{6st} = 0.2025$	$X_{st1st} = 243.76$	$\bar{X}_{2st} = 274.22$
$\bar{X}_{3st} = 724.10$	$\bar{X}_{4st} = 773.65$	$\bar{X}_{5st} = 264.42$	$\bar{X}_{6st} = 98.44$	$\bar{Y}_{1st} = 15.37$
$\bar{Y}_{2st} = 22.13$	$\bar{Y}_{3st} = 93.84$	$\bar{Y}_{4st} = 55.88$	$\bar{Y}_{5st} = 9.67$	$\bar{Y}_{6st} = 4.04$
$C_{x1st} = 2.02$	$C_{x2st} = 2.10$	$C_{x3st} = 2.22$	$C_{x4st} = 3.84$	$C_{x5st} = 1.72$
$C_{x6st} = 1.91$	$C_{y1st} = 4.18$	$C_{y2st} = 5.22$	$C_{y3st} = 3.19$	$C_{y4st} = 5.13$
$C_{y5st} = 2.47$	$C_{y6st} = 2.31$	$\rho_{1st} = 0.82$	$\rho_{2st} = 0.86$	$\rho_{3st} = 0.90$
$\rho_{4st} = 0.99$	$\rho_{5st} = 0.71$	$\rho_{6st} = 0.89$	$\rho_{st} = 0.82$	$n = 140$

Population 4 ([4])

X = Amount of students in secondary schools plus primary consisting $N = 923$ at six different districts in Turkey in the year 2007, and Y = Number of teachers.

$N_{1st} = 127$	$N_{2st} = 117$	$N_{3st} = 103$	$N_{4st} = 170$	$N_{5st} = 205$
$N_{6st} = 201$	$n_{1st} = 31$	$n_{2st} = 21$	$n_{3st} = 29$	$n_{4st} = 38$
$n_{5st} = 22$	$n_{6st} = 39$	$W_{1st} = 0.11375$	$W_{2st} = 0.1267$	$W_{3st} = 0.1115$
$W_{4st} = 0.1841$	$W_{5st} = 0.2221$	$W_{6st} = 0.2177$	$\bar{X}_{1st} = 20804.59$	$\bar{X}_{2st} = 9211.79$
$\bar{X}_{3st} = 14309.30$	$\bar{X}_{4st} = 9478.85$	$\bar{X}_{5st} = 5569.95$	$\bar{X}_{6st} = 12997.59$	$\bar{Y}_{1st} = 703.74$
$\bar{Y}_{2st} = 413$	$\bar{Y}_{3st} = 573.17$	$\bar{Y}_{4st} = 424.66$	$\bar{Y}_{5st} = 267.03$	$\bar{Y}_{6st} = 393.84$
$C_{x1st} = 1.465$	$C_{x2st} = 1.648$	$C_{x3st} = 1.925$	$C_{x4st} = 1.922$	$C_{x5st} = 1.526$
$C_{x6st} = 1.777$	$C_{y1st} = 1.256$	$C_{y2st} = 1.562$	$C_{y3st} = 1.803$	$C_{y4st} = 1.909$
$C_{y5st} = 1.512$	$C_{y6st} = 1.807$	$\rho_{1st} = 0.936$	$\rho_{2st} = 0.996$	$\rho_{3st} = 0.994$
$\rho_{4st} = 0.983$	$\rho_{5st} = 0.989$	$\rho_{6st} = 0.965$	$\rho_{st} = 0.95$	$n = 180$

MSEs and their respective constants are calculated by using the above-mentioned data sets. After that, improved ratio estimator is compared with all of the reviewed estimators, through MSE and percent relative efficiency (PRE). In Table 1, we can observe the MSE and PRE based on populations 1, 2, 3 and 4, respectively.

Simulation study

For assessing the performance of the proposed estimator, we use four different artificial populations where x'_{hi} and y'_{hi} are from different distributions, as given in Table 2. In order to propose different level of correlations between study and auxiliary variables, some transformations are given in Table 3. Each population contains three strata having five units. We selected all possible $n_h = 2, 3, 4$ units from each stratum, respectively, and in this way, we get ${}^5C_2 {}^5C_3 {}^5C_4 = 500$ samples.

The degree of linear relationship between study and auxiliary variables is taken as 0.50, 0.70, 0.90 for each stratum respectively. For more details about these artificial stratified populations, see Koyuncu and Kadilar (2014). The MSE and PRE based on artificial populations 1,2,3,4 are available in Table 4.

Conclusion

We have developed an improved estimator for the population mean in stratified random sampling. MSE of improved estimator has been found and compared with some of the existing estimators. Also numerical illustration has been carried out using four real data sets. From Table 1, we can see that our proposed estimator is less bias and has minimum MSE value for all real data sets. We have also conducted a simulation study to see the efficiency of our proposed estimator for different artificial data sets. We have calculated both theoretical and empirical bias and MSE values for all estimators. From Table 4, we can conclude that our proposed estimator is highly efficient

Table 1 Bias, MSE and PRE based on real data sets

Est.	Pop-1			Pop-2			Pop-3			Pop-4		
	Bias	MSE	PRE	Bias	MSE	PRE	Bias	MSE	PRE	Bias	MSE	PRE
\hat{y}	0.9538	0.9098	100	61.17	3734.12	100	8.35	69.80	100	47.21	2229.6	100
\hat{y}_{Rst}	1.3571	1.8418	49.39	40.49	16402.2	22.76	4.72	22.33	312.57	14.74	217.45	1025.32
\hat{y}_{BTst}	0.9794	0.9594	94.82	60.05	3606.40	103.54	6.02	36.24	192.60	24.54	603.18	369.63
$\hat{y}_{Reg(st)}$	0.9104	0.8290	109.74	45.31	2053.81	181.81	4.67	21.90	318.63	13.97	195.27	1141.79
\hat{y}_{SGst}	0.9094	0.8271	109.99	45.31	2053.65	181.82	4.62	21.36	326.76	13.97	195.07	1142.96
$\hat{y}_{Sal1(st)}$	0.9091	0.8265	110.07	45.31	2053.43	181.84	4.61	21.28	327.92	13.95	194.87	1144.14
$\hat{y}_{Sal2(st)}$	0.9078	0.8242	110.38	45.30	2052.15	181.96	4.58	21.03	331.92	13.92	193.85	1150.14
$\hat{y}_{Sal3(st)}$	0.9047	0.8186	111.13	45.25	2048.02	182.32	4.52	20.50	340.42	13.81	190.78	1168.68
\hat{y}_{Nst}	0.8966	0.8039	113.17	45.08	2032.71	183.70	4.36	19.03	366.81	13.34	178.21	1251.10

Table 2 Parameters and distributions of study and auxiliary variables

Artificial population 1	
$f(y'_{hi}) = \frac{1}{\Gamma(1.5)} y'^{1.5-1}_{hi} e^{-y'_{hi}}$	$f(x'_{hi}) = \frac{1}{\Gamma(0.3)} x'^{0.3-1}_{hi} e^{-x'_{hi}}$
Artificial population 2	
$f(y'_{hi}) = \frac{1}{\Gamma(0.3)} y'^{0.3-1}_{hi} e^{-y'_{hi}}$	$f(x'_{hi}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x'^2_{hi}}{2}}$
Artificial population 3	
$f(y'_{hi}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y'^2_{hi}}{2}}$	$f(x'_{hi}) = \frac{1}{\Gamma(0.3)} x'^{0.3-1}_{hi} e^{-x'_{hi}}$
Artificial Population 4	
$f(y'_{hi}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y'^2_{hi}}{2}}$	$f(x'_{hi}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x'^2_{hi}}{2}}$

Table 3 Characteristics of strata

Stratum 1	
$y_{1i} = y'_{1i} + 50$	$x_{1i} = \rho_{xy1} \frac{S_{1x}}{S_{1y}} y'_{1i} + \sqrt{1 - \rho_{xy1}^2} x'_{1i} + 15$
Stratum 2	
$y_{2i} = y'_{2i} + 150$	$x_{2i} = \rho_{xy2} \frac{S_{2x}}{S_{2y}} y'_{2i} + \sqrt{1 - \rho_{xy2}^2} x'_{2i} + 100$
Stratum 3	
$y_{3i} = y'_{3i} + 100$	$x_{3i} = \rho_{xy2} \frac{S_{2x}}{S_{2y}} y'_{2i} + \sqrt{1 - \rho_{xy2}^2} x'_{2i} + 200$

than existing estimators. As a result from numerical illustration, it is derived that the new estimator is more efficient than the classical mean, ratio, exponential, regression, [8] and [3] estimators. We suggest the use of the proposed estimator for a more efficient estimation of the finite population mean in stratified random sampling.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

We can write (14) with the e terms as:

$$\hat{y}_{Nst} = \left[\bar{Y} \left\{ 1 + \frac{1}{8} e_{1st}^2 + e_{ost} \right\} - w_1 \bar{X} e_{1st} + w_2 \bar{Y} + w_2 \bar{Y} e_{ost} \right] \left[1 - \frac{e_{1st}}{2k} + \frac{3e_{1st}^2}{8k^2} + \dots \right]. \tag{15}$$

It implies that

$$\begin{aligned} \hat{y}_{Nst} - \bar{Y} = & \frac{\bar{Y}}{8} e_{1st}^2 + \bar{Y} e_{ost} - w_1 \bar{X} e_{1st} + w_2 \bar{Y} + w_2 \bar{Y} e_{ost} - \bar{Y} \frac{e_{1st}}{2k} \\ & - \bar{Y} \frac{e_{ost} e_{1st}}{2k} + w_1 \bar{X} \frac{e_{1st}^2}{2k} \\ & - w_2 \bar{Y} \frac{e_{1st}}{2k} - w_2 \bar{Y} \frac{e_{ost} e_{1st}}{2k} + \bar{Y} \frac{3e_{1st}^2}{8k^2} + w_2 \bar{Y} \frac{3e_{1st}^2}{8k^2}. \end{aligned} \tag{16}$$

Now taking expectation we obtain the bias of \hat{y}_{Nst} as,

$$\begin{aligned} \text{Bias}(\hat{y}_{Nst}) = & \bar{Y} \left\{ \frac{V_{0.2}}{8} \left(1 + \frac{3}{k^2} \right) - \frac{V_{11}}{2k} \right\} \\ & + \frac{\bar{X} V_{0.2}}{2k} w_1 + \bar{Y} \left\{ \frac{3V_{0.2}}{8k^2} - \frac{V_{1.1}}{2k} w_2 \right\}. \end{aligned}$$

We get $MSE(\hat{y}_N)$ by squaring both sides of (16), ignoring higher order terms, taking expectation as

Table 4 Bias, MSE and PRE based on simulation

Est.	Pop-1			Pop-2			Pop-3			Pop-4		
	Bias	MSE	PRE	Bias	MSE	PRE	Bias	MSE	PRE	Bias	MSE	PRE
<i>Empirical</i>												
\hat{y}	47.45	2252.19	100	55.79	3113.16	100	90.25	8146.11	100	80.24	6440.01	100
\hat{y}_{Rst}	47.03	2212	101.80	55.78	3112.19	100.03	90.19	8135.01	100.16	80.01	6400.91	100.61
\hat{y}_{BTst}	45.96	2113.12	106.58	55.67	3100.21	100.41	90.54	8199.08	99.35	802.01	6402.09	100.59
$\hat{y}_{Reg(st)}$	41.23	1700.17	132.46	46.89	2199.08	141.56	89.44	8000.29	101.82	79.37	6301.07	102.20
\hat{y}_{SGst}	41.22	1699.16	132.54	45.93	2110.01	147.54	88.94	7910.45	102.97	79.06	6250.51	103.03
$\hat{y}_{Sal1(st)}$	39.43	1555.15	144.82	42.56	1812.03	171.81	89.05	7930.04	102.72	77.46	6000.91	107.31
$\hat{y}_{Sal2(st)}$	39.44	1556.13	144.73	42.56	1812.06	171.80	89.05	7931.05	102.71	77.45	6000.02	107.33
$\hat{y}_{Sal3(st)}$	39.42	1554.13	144.91	42.53	1809.15	172.07	89.02	7925.39	102.78	77.40	5991.49	107.48
\hat{y}_{Nst}	37.41	1400.09	160.86	41.24	1701.27	182.99	70.88	5025.31	162.10	64.19	4121.01	156.27
<i>Theoretical</i>												
\hat{y}	45.14	2247.11	100	53.83	3103.23	100	92.17	8140.91	100	75.01	6442.92	100
\hat{y}_{Rst}	46.42	2210	101.22	56.62	3110.14	100.08	89.21	8130.83	100.12	74.62	6409.51	101.46
\hat{y}_{BTst}	44.72	2117.51	107.72	54.64	3102.27	102.43	85.51	8197.07	96.31	800.71	6400.02	101.33
$\hat{y}_{Reg(st)}$	42.55	1702.15	134.88	44.75	2190.75	140.22	87.02	8001.35	103.98	76.74	6304.61	103.18
\hat{y}_{SGst}	39.17	1694.32	130.52	44.21	2113.02	145.43	89.90	7913.67	104.83	76.31	6253.22	102.44
$\hat{y}_{Sal1(st)}$	37.51	1553.12	142.81	42.11	1810.09	173.23	88.01	7935.74	104.51	79.51	6001.93	105.74
$\hat{y}_{Sal2(st)}$	37.31	1553.51	142.70	44.71	1810.05	170.36	90.06	7934.66	104.53	75.43	6004.04	109.32
$\hat{y}_{Sal3(st)}$	38.35	1551.17	148.92	45.50	1803.10	170.85	87.92	7926.32	104.02	78.91	5988.32	109.71
\hat{y}_{Nst}	35.21	1403.62	163.44	40.33	1704.39	180.85	73.82	5023.30	160.01	62.86	4120.50	158.44

$$MSE(\hat{y}_{Nst}) = \bar{Y}^2 L + w_1^2 \lambda_A + w_2^2 \lambda_B + 2w_1 w_2 \lambda_C - 2w_1 \lambda_D - w_2 \lambda_E, \tag{17}$$

where

$$L = f' \left\{ V_{2.0} + \frac{V_{0.2}}{4k^2} - \frac{V_{1.1}}{k} \right\},$$

$$\lambda_A = \bar{X}^2 V_{0.2},$$

$$\lambda_B = \bar{Y}^2 \left\{ 1 + V_{2.0} + \frac{V_{0.2}}{k^2} - \frac{2V_{1.1}}{k} \right\},$$

$$\lambda_C = \bar{X}\bar{Y} \left\{ \frac{V_{0.2}}{k} - V_{1.1} \right\},$$

$$\lambda_D = \bar{X}\bar{Y} \left\{ -\frac{V_{0.2}}{2k} + V_{1.1} \right\},$$

$$\lambda_E = \bar{Y}^2 \left\{ \frac{3V_{1.1}}{k} - 2V_{2.0} - \frac{V_{0.2}}{4} - \frac{5V_{0.2}}{4k^2} \right\}.$$

Partially differentiating (17) with respect to w_1 and w_2 and equating to zero, we have the following equations

$$w_1 \lambda_A + w_2 \lambda_C = \lambda_D, \tag{17}$$

$$w_1 \lambda_C + w_2 \lambda_B = \frac{\lambda_E}{2}. \tag{18}$$

Now by solving matrix inversion method, we get the optimum values of w_1, w_2 i.e.

$$w_1^{opt} = \left[\frac{\lambda_B \lambda_D - \frac{\lambda_C \lambda_E}{2}}{\lambda_A \lambda_B - \lambda_C^2} \right], \tag{19}$$

and

$$w_2^{opt} = \left[\frac{-\lambda_C \lambda_D + \frac{\lambda_A \lambda_E}{2}}{\lambda_A \lambda_B - \lambda_C^2} \right]. \tag{20}$$

By putting w_1^{opt}, w_2^{opt} in $MSE(\hat{y}_{Nst})$, we get the minimum MSE of \hat{y}_{Nst} , i.e.,

$$MSE_{min}(\hat{y}_{Nst}) = \left[\bar{Y}^2 L - \frac{\lambda_B \lambda_D^2 + \frac{\lambda_A \lambda_E^2}{4} - \lambda_C \lambda_D \lambda_E}{\lambda_A \lambda_B - \lambda_C^2} \right]. \tag{21}$$

References

1. Bahl, S., Tuteja, R.K.: Ratio and product type exponential estimator. *J. Inf. Optim. Sci.* **12**, 159–163 (1991)

2. Kadilar, C., Cingi, H.: Ratio estimators in stratified sampling. *Biom. J.* **45**, 218–225 (2003)
3. Khan, S.A., Ali, H., Manzoor, S., Alamgir, : A class of transformed efficient ratio estimators of finite population mean. *Pak. J. Stat.* **31**, 353–362 (2015)
4. Koyuncu, N., Kadilar, C.: Ratio and product estimators in stratified random sampling. *J. Stat. Plann. Inference* **139**, 2552–2558 (2009)
5. Koyuncu, N., Kadilar, C.: Calibration weighting in stratified random sampling. *Commun. Stat. Simul. Comput.* **45**(7), 2267–2275 (2016)
6. Murthy, M.N.: *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta (1967)
7. Rao, T.: On certain methods of improving ratio and regression estimators. *Commun. Stat. Theory Methods* **20**, 3325–3340 (1991)
8. Shabbir, J., Gupta, S.: On estimating finite population mean in simple and stratified random sampling. *Commun. Stat. Theory Method* **40**, 199–212 (2011)
9. Singh, H.P., Kakran, M.S.: A Modified Ratio Estimator using Known Coefficient of Kurtosis of an Auxiliary Character.(unpublished) (1993)
10. Singh, R., Mangat, N.S.: *Elements of Survey Sampling*. Kluwer Academic Publishers, London (1996)
11. Sisodia, B.V.S., Dwivedi, V.K.: A modified ratio estimator using coefficient of variation of auxiliary variable. *J. Indian Soc. Agric. Stat.* **33**, 13–18 (1981)
12. Upadhyaya, L.N., Singh, H.P.: Use of transformed auxiliary variable in estimating the finite population mean. *Biom. J.* **41**, 627–636 (1999)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

