New generalized class of estimators for estimation of finite population mean based on probability proportional to size sampling using two auxiliary variables: A simulation study

Sohaib Ahmad[1] ⓘ, Javid Shabbir[2,3], Erum Zahid[4], Muhammad Aamir[1] and Mohammed Alqawba[5]

[1]Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan
[2]Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan
[3]Department of Statistics, University of Wah, Wah Cantt, Pakistan
[4]Department of Applied Mathematics and Statistics, Institute of Space Technology, Islamabad, Pakistan
[5]Department of Mathematics, College of Science and Arts, Qassim University, Ar Rass, Saudi Arabia

Abstract

This article aims to suggest a new generalized class of estimators based on probability proportional to size sampling using two auxiliary variables. The numerical expressions for the bias and mean squared error (MSE) are derived up to the first order of approximation. Four actual data sets are used to examine the performances of a new improved generalized class of estimators. From the results of real data sets, it is examined that the suggested estimator gives the minimum MSE and the percentage relative efficiency is higher than all existing estimators, which shows the importance of the new generalized class of estimators. To check the strength and generalizability of our proposed class of estimators, a simulation study is also accompanied. The consequence of the simulation study shows the worth of newly found proposed class estimators. Overall, we get to the conclusion that the proposed estimator outperforms as compared to all other estimators taken into account in this study.

Corresponding author:
Muhammad Aamir, Department of Statistics, Abdul Wali Khan University Mardan, KPK, Khyber Pakhtunkhwa, 23200, Pakistan.
Email: aamirkhan@awkum.edu.pk

Introduction

In the survey sampling approach, estimating the finite population mean is a common issue, and many efforts have been made to improve the precision of the estimators. A comprehensive range of approaches for incorporating the auxiliary variables by using ratio, product, and regression-type estimates are defined in the literature. Mainly when there are multiple auxiliary variables, a wide range of estimators have been presented, each one combining ratio, product, or regression estimators. Researchers have previously attempted to use the best statistical features to estimate population parameters including variance, coefficient of variation, and kurtosis. A representative sample of the population is required for this set-up. If the population of interest is similar, then selecting units can be done using simple random sampling with or without replacement. The population parameters of the auxiliary variable should also be previously known when using the ratio, product, and regression estimation methods. By suitably adapting the auxiliary variables, many authors have suggested several estimators. The researcher can investigate these research findings by looking the Kadilar and Cingi[1] who recommended improvement in estimating the population mean in simple random sampling. Al-Omari[2] suggested ratio estimation of the population mean using the auxiliary information in simple random sampling and median ranked set sampling. Ozturk[3] proposed estimation of population mean and total in a finite population setting using multiple auxiliary variables. Yadav et al.[4] recommended the use of the auxiliary variables in searching efficient estimators of a population mean. Bhushan and Pandey[5] discussed the optimality of ratio-type imputation methods for the estimation of population mean using the higher order moment of an auxiliary variable. Zaman et al.[6] recommended robust regression-ratio-type estimators of the mean utilizing two auxiliary variables. Kumar and Saini[7] discussed a predictive approach for the finite population mean when auxiliary variables are attributes. Singh and Nigam[8] recommended a generalized class of estimators for finite population mean using two auxiliary variables in sample surveys. Bhushan et al.[9] proposed some improved classes of estimators in stratified sampling using bivariate auxiliary information. Shahzad et al.[10] discussed mean estimation using robust quantile regression with two auxiliary variables. Zaman et al.[6] recommended robust regression-ratio-type estimators of the mean utilizing two auxiliary variables. Mahdizadeh and Zamanzade[11] proposed an interval estimation of the population mean in ranked set sampling. Ahmad et al.[12] recommended a new improved generalized class of estimators for population distribution function using the auxiliary variable under simple random sampling. Muhammad et al.[13] suggested an enhanced ratio-type estimator for finite population mean using the auxiliary variable in simple random sampling. Ahmad et al.[14] discussed an improved generalized class of estimators in estimating the finite population mean using two auxiliary variables under two-stage sampling. Shahzad et al.[15] proposed a three-fold utilization of supplementary information for mean estimation under median-ranked set sampling scheme. Shahzad et al.[16] discussed the estimation of the population mean by successive use of an auxiliary variable in median ranked set sampling.

Yasmeen et al.[17] proposed generalized exponential estimators of finite population mean using transformed auxiliary variables. Singh et al.[18] discussed an alternative efficient class of estimators for finite population mean using information on an auxiliary attribute in sample surveys. Singh et al.[19] recommended the estimation of finite population variance using scrambled responses in the presence of auxiliary information.

In many conditions, the population differs considerably in size, for example, in a medical study, the number of patients having a specific disease, and the size of health units may differ. Likewise, in a survey related to the income of the household, the household may have a different number of siblings, and then in such circumstances, the probability of units may change. For dealing with such unequal probability, we use probability proportional to size (PPS) sampling. PPS is an unequal random sampling in which the chance of gathering information is proportional to an auxiliary variable, for each sampling unit in the population. Consider the case where we need to assess the population in a province within a country; we take the auxiliary variable that has an association with the study variable. For example (i) Population of all provinces within the country (correlated with study variable = 0.95). (ii) Number of households in all communities within the province (correlated with the study variable = 0.99). Based on these facts (ii) may be more useful at the estimation stage. Many researchers have suggested several estimators by efficiently adjusting the auxiliary variables under PPS. The researcher can examine this research by Akpanta[20] who proposed the problems of PPS sampling in multicharacter surveys. Agarwal and Mannai[21] recommended a linear combination of estimators in PPS sampling to estimate the population mean and its robustness to optimum value. Abdulla et al.[22] suggested the selection of samples in PPS sampling using the cumulative relative frequency method. Andersen et al.[23] discussed optimal PPS sampling by vanishing the auxiliary variables with applications in microscopy. Alam et al.[24] discussed the selection of the samples with PPS. Patel and Bhatt[25] recommended the estimation of finite population total under PPS sampling in the presence of extra auxiliary information. Singh et al.[26] discussed an improved estimator of population total in PPS sampling. Makela et al.[27] suggested Bayesian inference under cluster sampling with PPS. Ahmad and Shabbir[28] discussed the use of extreme values to estimate the finite population mean under the PPS sampling scheme. Ozturk[29] proposed poststratified PPS sampling from stratified populations. Latpate et al.[30] discussed the scheme of PPS sampling. Sohil et al.[31] recommended optimum second call imputation in PPS sampling. Sinha and Khanna[32] discussed the estimation of population mean under PPS sampling with and without measurement errors. Zangeneh and Little[33] discussed Bayesian inference for the finite population total from a heteroscedastic PPS. Hentschel et al.[34] recommended exact PPS sampling with a bounded sample size. Barbiero et al.[35] proposed bootstrapping PPS samples via a calibrated empirical population. Gupt and Ahamed[36] discussed optimum stratification for a generalized auxiliary variable proportional to allocation under a super-population model. Ponkaew and Lawson[37] recommended new estimators for estimating the population total with an application to water demand in Thailand under unequal probability sampling without replacement for missing data. Al-Jararha[38] discussed a class of estimators using two units with PPS. Al-Marzouki et al.[39] proposed an estimation of finite population mean under PPS in the presence of maximum and minimum values. Zheng and Little[40] suggested penalized

spline model-based estimation of the finite population total. Zheng and Little[41] recommended inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. Amab[42] proposed the optimum estimation of a finite population total in PPS sampling with a replacement for multicharacter surveys. Olayiwolla et al.[43] suggested the PPS method to enhance the efficiency of the estimator in two-stage sampling.

(i)   In this article, the primary aim of the current work is to propose a new improved generalized class of estimators for the estimation of finite population mean using two auxiliary variables under PPS.

(ii) The bias and mean squared error (MSE) of the proposed estimator is derived up to the first order of approximation.

(iii) Through use of the real data sets from various domains and a simulation study, the application of the proposed estimator is highlighted.

All notations and symbols are given in the Appendix.

Review of existing estimators

In this section, we have studied some well-known existing estimators under PPS sampling.

1. The usual estimator under PPS, is given by:

$$\widehat{\bar{Y}}_u = \bar{u} \tag{1}$$

The variance of $\widehat{\bar{Y}}_u$ is given by:

$$\mathrm{Var}(\widehat{\bar{Y}}_u) = \lambda \bar{Y}^2 C_u^2 \tag{2}$$

2. The ratio estimator under PPS, is given by:

$$\widehat{\bar{Y}}_{R,PPS} = \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right) \tag{3}$$

The bias and MSE of $\widehat{\bar{Y}}_{R,PPS}$, are given by:

$$\mathrm{Bias}(\widehat{\bar{Y}}_{R,PPS}) \cong \bar{Y}\lambda[C_u^2 - \rho_{uv}C_uC_v],$$

and

$$\mathrm{MSE}(\widehat{\bar{Y}}_{R,PPS}) \cong \bar{Y}^2\lambda[C_u^2 + C_v^2 - 2\rho_{uv}C_uC_v]. \tag{4}$$

3. Murthy,[44] suggested a product estimator, given by:

$$\widehat{\bar{Y}}_{R,PPS} = \bar{u}\left(\frac{\bar{v}}{\bar{X}}\right) \tag{5}$$

The bias and MSE of $\widehat{\bar{Y}}_{P,PPS}$, are given by:

$$\text{Bias}(\widehat{\bar{Y}}_{P,PPS}) \cong \bar{Y}\lambda\rho_{uv}C_uC_v,$$

and

$$\text{MSE}(\widehat{\bar{Y}}_{R,PPS}) \cong \bar{Y}^2\lambda[C_u^2 + C_v^2 - 2\rho_{uv}C_uC_v]. \quad (6)$$

4. The regression estimator, is given by:

$$\widehat{\bar{Y}}_{(Reg,PPS)} = \bar{u} + \Omega_1(\bar{X} - \bar{v}), \quad (7)$$

where Ω_1 is constant. The optimum values of Ω_1 is given by:

$$\Omega_{1(opt)} = \frac{\rho_{uv}S_u}{S_v}$$

The minimum variance of $\widehat{\bar{Y}}_{(Reg,PPS)}$, is given by:

$$\text{Var}(\widehat{\bar{Y}}_{(Reg,PPS)})_{min} \cong \lambda\bar{Y}^2C_u^2(1 - \rho_{uv}^2) = MSE(\widehat{\bar{Y}}_{(Reg,PPS)}) \quad (8)$$

5. Bai et al.[45] proposed the following estimator, is given by:

$$\widehat{\bar{Y}}_{(Rao,PPS)} = \Omega_2\bar{u} + \Omega_3(\bar{X} - \bar{v}), \quad (9)$$

where Ω_2 and Ω_3 are the unknown constants, the optimum values are given by:

$$\Omega_{2(opt)} = \frac{1}{1 + \lambda C_u^2(1 - \rho_{uv}^2)},$$

and

$$\Omega_{3(opt)} = \frac{\bar{Y}C_u\rho_{uv}}{XC_v\{1 + \lambda C_u^2(1 - \rho_{uv}^2)\}}$$

The minimum MSE of $\widehat{\bar{Y}}_{(Rao,PPS)}$ is given by:

$$\text{MSE}(\widehat{\bar{Y}}_{(Rao,PPS)})_{min} = \frac{\lambda\,\bar{Y}^2C_u^2(1 - \rho_{uv}^2)}{\{1 + \lambda C_u^2(1 - \rho_{uv}^2)\}} \quad (10)$$

6. Bahl and Tuteja[46] suggested the following ratio and product exponential type estimators, are given by:

$$\widehat{\bar{Y}}_{(BR,PPS)} = \bar{u}\exp\left(\frac{\bar{X} - \bar{v}}{\bar{X} + \bar{v}}\right), \quad (11)$$

and

$$\widehat{\overline{Y}}_{(BR,PPS)} = \bar{u}\exp\left(\frac{\bar{v}-\bar{X}}{\bar{v}+\bar{X}}\right). \tag{12}$$

The biases and MSEs of $\widehat{\overline{Y}}_{(BR,PPS)}$, $\widehat{\overline{Y}}_{(BP,PPS)}$ are given by:

$$\text{Bias}(\widehat{\overline{Y}}_{(BR,PPS)}) \cong \lambda\bar{Y}\left(\frac{3}{8}\,C_v^2 - \frac{1}{2}\rho_{uv}C_uC_v\right)$$

$$\text{MSE}(\widehat{\overline{Y}}_{(BR,PPS)}) \cong \lambda\bar{Y}^2\left(C_u^2 + \frac{1}{4}C_v^2 - \rho_{uv}C_uC_v\right), \tag{13}$$

$$\text{Bias}(\widehat{\overline{Y}}_{(BR,PPS)}) \cong \lambda\bar{Y}\left(\rho_{uv}C_uC_v - \frac{1}{4}C_v^2\right),$$

$$\text{MSE}(\widehat{\overline{Y}}_{(BR,PPS)}) \cong \lambda\bar{Y}^2\left[C_u^2 + \frac{1}{4}C_v^2 + \rho_{uv}C_uC_v\right]. \tag{14}$$

7. Haq and Shabbir[47] suggested the following exponential-type estimators, which are given by:

$$\widehat{\overline{Y}}_{(H1,PPS)} = \left\{\frac{\Omega_4}{2}\bar{u}\left(\frac{\bar{X}}{\bar{v}} + \frac{\bar{v}}{\bar{X}}\right) + \Omega_5(\bar{X} - \bar{v})\right\}\exp\left(\frac{\bar{X}-\bar{v}}{\bar{X}+\bar{v}}\right) \tag{15}$$

where Ω_4 and Ω_5 are constants. The bias and MSE of $\widehat{\overline{Y}}_{(H1,PPS)}$ are given by:

$$\text{Bias}(\widehat{\overline{Y}}_{(H1,PPS)}) \cong \bar{Y}\left[(\Omega_4 - 1) + \Omega_4\lambda\left(\frac{7}{8}C_v^2 - \frac{1}{2}\rho_{uv}C_uC_v\right) + \Omega_5R\frac{C_v^2}{2}\right],$$

where $= \frac{\bar{X}}{\bar{Y}}$,
    The optimum values are

$$\Omega_{4(opt)} = \frac{B_{h1}C_{h1} - D_{h1}E_{h1}}{A_{h1}B_{h1-E_{h1}^2}}$$

and

$$\Omega_{5(opt)} = \frac{A_{h1}D_{h1} - C_{h1}E_{h1}}{B_{h1}B_{h1} - E_{h1}^2}$$

The minimum MSE of $\widehat{\overline{Y}}_{(H1,PPS)}$ at the optimum values, is given by:

$$\text{MSE}(\widehat{\overline{Y}}_{(H1,PPS)})_{min} \cong \bar{Y}^2\left[1 - \left(\frac{A_{h1}D_{h1}^2 + B_{h1}C_{h1}^2 - C_{h1}D_{h1}E_{h1}}{A_{h1}B_{h1} - E_{h1}^2}\right)\right], \tag{16}$$

where

$$A_{h1} = 1 + \lambda[C_u^2 + 2C_v^2 - 2\rho_{uv}C_uC_v],$$

$$B_{h1} = R^2 \lambda C_v^2,$$

$$C_{h1} = 1 + \lambda\left[\frac{7}{8}C_v^2 - \frac{1}{2}\rho_{uv}C_uC_v\right],$$

$$D_{h1} = R\lambda\frac{C_v^2}{2},$$

$$E_{h1} = R\lambda[C_v^2 - \rho_{uv}C_uC_v].$$

The second proposed estimator of $\widehat{\bar{Y}}_{(H2,PPS)}$, is given by:

$$\widehat{\bar{Y}}_{(H2,PPS)} = \{\Omega_6\widehat{\bar{Y}}_{BTA,PPS} + \Omega_7(\bar{X} - \bar{v})\} \exp\left(\frac{\bar{X} - \bar{v}}{\bar{X} + \bar{v}}\right) \tag{17}$$

where Ω_6 and Ω_7 are constants.

$$\widehat{\bar{Y}}_{BTA,PPS} = \frac{\bar{u}}{2}\left[\exp\left(\frac{\bar{X} - \bar{v}}{\bar{X} + \bar{v}}\right) + \exp\left(\frac{\bar{v} - \bar{X}}{\bar{v} + \bar{X}}\right)\right],$$

The bias of $\widehat{\bar{Y}}_{(H1,PPS)}$ is given by:

$$\text{Bias}(\widehat{\bar{Y}}_{(H2,PPS)}) \cong \bar{Y}\left[(\Omega_6 - 1) + \Omega_6\lambda\left(\frac{C_v^2}{2} - \frac{1}{2}\rho_{uv}C_uC_v\right) + \Omega_7R\frac{C_v^2}{2}\right]$$

The optimum values of Ω_6 and Ω_7 are given by:

$$\Omega_{6(opt)} = \frac{B_{h2}C_{h2} - D_{h2}E_{h2}}{A_{h2}B_{h2} - E_{h2}^2}$$

and

$$\Omega_{7(opt)} = \frac{A_{h2}D_{h2} - C_{h2}E_{h2}}{B_{h2}B_{h2} - E_{h2}^2},$$

The minimum MSE of $\widehat{\bar{Y}}_{(H2,PPS)}$ at the optimum values, is given by:

$$\text{MSE}(\widehat{\bar{Y}}_{(H2,PPS)})_{min} \cong \bar{Y}^2\left[1 - \left(\frac{A_{h2}D_{h2}^2 + B_{h2}C_{h2}^2 - 2C_{h2}D_{h2}E_{h2}}{A_{h2}B_{h2} - E_{h2}^2}\right)\right], \tag{18}$$

where

$$A_{h2} = 1 + \lambda \left[C_u^2 + \frac{5}{4} C_v^2 - 2\rho_{uv} C_u C_v \right],$$

$$B_{h2} = R^2 \, \lambda C_v^2,$$

$$C_{h2} = 1 + \lambda \left[\frac{1}{2} C_v^2 - \frac{1}{2} \rho_{uv} C_u C_v \right],$$

$$D_{h2} = R\lambda \frac{C_v^2}{2},$$

$$E_{h2} = R\lambda [C_v^2 - \rho_{uv} C_u C_v]$$

8. Ekpenyong and Enang[48] suggested the following estimator:

$$\widehat{\bar{Y}}_{(EE,PPS)} = \Omega_8 \bar{u} + \Omega_9 (\bar{X} - \bar{v}) \exp\left(\frac{\bar{X} - \bar{v}}{\bar{X} + \bar{v}} \right) \qquad (19)$$

where Ω_8 and Ω_9 are constants.

The bias of $\widehat{\bar{Y}}_{(EE,PPS)}$, is given by:

$$\text{Bias}(\widehat{\bar{Y}}_{(EE,PPS)}) \cong \bar{Y} \left\{ (\Omega_8 - 1) + \Omega_9 R\lambda \frac{C_v^2}{2} \right\}.$$

The optimum values of Ω_8 and Ω_9 are given by:

$$\Omega_{8(opt)} = \frac{B_e C_e - D_e E_e}{A_e B_e - E_e^2}$$

and

$$\Omega_{9(opt)} = \frac{A_e D_e - C_e E_e}{B_e B_e - E_e^2}.$$

The minimum MSE of $\widehat{\bar{Y}}_{(EE,PPS)}$, at the optimal values, is given by:

$$\text{MSE}(\widehat{\bar{Y}}_{(EE,PPS)})_{min} \cong \bar{Y}^2 \left[1 - \left(\frac{A_e D_e^2 + B_e C_e^2 - 2C_e D_e E_e}{A_e B_e - E_e^2} \right) \right], \qquad (20)$$

where

$$A_e = 1 + \lambda C_u^2,$$

$$B_e = \lambda R^2 C_v^2,$$

$$C_e = 1,$$

$$D_e = R\lambda \frac{\theta^2 C_v^2}{2},$$

$$E_e = \lambda R\left[\frac{C_v^2}{2} - \rho_{uv} C_u C_v\right]$$

9. Singh et al.[49] suggested the following class of estimators, is given by:

$$\widehat{\bar{Y}}^*_{(S,PPS)} = \bar{u}\exp\left(\frac{a(\bar{X} - \bar{v})}{a(\bar{X} + \bar{v}) + 2b}\right), \tag{21}$$

where a and b are known constants.

The bias and MSE of $\widehat{\bar{Y}}^*_{(S,PPS)}$, to the first order approximation are given by:

$$\text{Bias}(\widehat{\bar{Y}}^*_{(S,PPS)}) = \bar{Y}\lambda\left(\frac{3}{8} \theta^2 C_v^2 - \frac{1}{2} \theta\rho_{uv} C_u C_v\right) \tag{22}$$

and

$$\text{MSE}(\widehat{\bar{Y}}^*_{(S,PPS)}) = \frac{\lambda \bar{Y}^2}{4}(4C_u^2 + \theta^2 C_v^2 - 4\theta\rho_{uv} C_u C_v) \tag{23}$$

where $\theta = \frac{a\bar{X}}{a\bar{X}+b}$.

10. Grover and Kaur[50] suggested the following estimators and is given by:

$$\widehat{\bar{Y}}^*_{(GK,PPS)} = [\Omega_{10}\bar{u} + \Omega_{11}(\bar{X} - \bar{v})] \exp\left(\frac{a(\bar{X} - \bar{v})}{a(\bar{X} - \bar{v}) + 2b}\right) \tag{24}$$

where Ω_{10} and Ω_{11} are constants.

The bias and MSE of $\widehat{\bar{Y}}^*_{(GK,PPS)}$, are given by:

$$\text{Bais}(\widehat{\bar{Y}}^*_{(GK,PPS)}) \cong \bar{Y}\left\{(\Omega_{10} - 1) + \Omega_{10}\left\{\frac{3}{8}\theta^2 C_v^2 - \frac{1}{2}\theta\rho_{uv} C_u C_v\right\} + \Omega_{11}R\lambda\frac{\theta^2 C_v^2}{2}\right\} \tag{25}$$

The optimum values of Ω_{10} and Ω_{11} are:

$$\Omega_{10(opt)} = \frac{B_g C_g - D_g E_g}{A_g B_g - E_g^2}$$

and

$$\Omega_{11(opt)} = \frac{A_g D_g - C_g E_g}{B_g B_g - E_g^2},$$

where

$$\theta = \frac{a\bar{X}}{a\bar{X} + b} \tag{26}$$

The minimum MSE of $\widehat{\bar{Y}}^*_{(GK,PPS)}$, is given by:

$$\text{MSE}(\widehat{\bar{Y}}^*_{(GK,PPS)})_{min} \cong \bar{Y}^2 \left[1 - \left(\frac{A_g D_g^2 + B_g C_g^2 - 2C_g D_g E_g}{A_g B_g - E_g^2} \right) \right], \quad (27)$$

where

$$A_g = 1 + \lambda[C_u^2 + \theta^2 C_v^2 - 2\theta\rho_{uv}C_u C_v],$$

$$B_g = R^2 \lambda C_v^2,$$

$$C_g = 1 + \lambda\left[\frac{3}{8}\theta^2 C_v^2 - \frac{1}{2}\theta\rho_{uv}C_u C_v \right],$$

$$D_g = R\lambda\frac{C_v^2}{2},$$

$$E_g = \lambda[\theta^2 C_v^2 - \theta\rho_{uv}C_u C_v].$$

Proposed estimator

An estimator's performance can be improved by using appropriate use of the auxiliary variables at the design or estimation stage. Based on these ideas, we examine to use one auxiliary varible (Z) under PPS and the other auxiliary variable (X) at the estimation stage. The proposed estimator is more robust as compared to ratio, product and regression estimators as it can take any type of data that exists in literature. Taking motivation from Ahmad et al.,[51,52] we propose a new class of estimators using two auxiliary variables under PPS sampling.

$$\widehat{\bar{Y}}^*_{(Prop,PPS)} = [\Psi_{18}\widehat{\bar{Y}}_A + \Psi_{19}(\bar{X} - \bar{v})] \left[\exp\left(\frac{\alpha(\bar{X} - \bar{v})}{\alpha(\bar{X} + \bar{v}) + 2b} \right) \right], \quad (28)$$

where

$$\widehat{\bar{Y}}_A = \bar{u}\left\{ \frac{1}{4}\left(\frac{\bar{X}}{\bar{v}} + \frac{\bar{v}}{\bar{X}} \right)\left(\exp\left(\frac{\bar{X} - \bar{v}}{\bar{X} + \bar{v}} \right) + \exp\left(\frac{\bar{v} - \bar{X}}{\bar{v} + \bar{X}} \right) \right) \right\}$$

where Ψ_{18} and Ψ_{19} are the unknown constants, a and b are described earlier. Some family members of estimators are given in Table 1.

After simplification of $\widehat{\bar{Y}}^*_{(Prop,PPS)}$, we have

$$\widehat{\bar{Y}}^*_{(Prop,PPS)} = \left[\Psi_{18}\bar{Y}(1 + \xi_0)\left(1 + \frac{5}{8}\xi_1^2 \right) - \Psi_{19}\bar{X}\xi_1 \right]\left[1 - \frac{1}{2}\theta\xi_1 + \frac{3\theta^2}{8}\xi_1^2 \right] \quad (29)$$

Table 1. Family members of the suggested generalized class of estimators.

A	b	$\widehat{\overline{Y}}^*_{(S,PPS)}$	$\widehat{\overline{Y}}^*_{(GK,PPS)}$	$\widehat{\overline{Y}}^*_{(Prop,PPS)}$
I	C_v	$\widehat{\overline{Y}}^{*(1)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(1)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(1)}_{(Prop,PPS)}$
I	$\beta_{2(v)}$	$\widehat{\overline{Y}}^{*(2)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(2)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(2)}_{(Prop,PPS)}$
$\beta_{2(v)}$	C_v	$\widehat{\overline{Y}}^{*(3)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(3)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(3)}_{(Prop,PPS)}$
C_v	$\beta_{2(v)}$	$\widehat{\overline{Y}}^{*(4)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(4)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(4)}_{(Prop,PPS)}$
I	ρ_{uv}	$\widehat{\overline{Y}}^{*(5)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(5)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(5)}_{(Prop,PPS)}$
C_v	ρ_{uv}	$\widehat{\overline{Y}}^{*(6)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(6)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(6)}_{(Prop,PPS)}$
ρ_{uv}	C_v	$\widehat{\overline{Y}}^{*(7)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(7)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(7)}_{(Prop,PPS)}$
$\beta_{2(v)}$	ρ_{uv}	$\widehat{\overline{Y}}^{*(8)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(8)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(8)}_{(Prop,PPS)}$
ρ_{uv}	$\beta_{2(v)}$	$\widehat{\overline{Y}}^{*(9)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(9)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(9)}_{(Prop,PPS)}$
I	$N\bar{X}$	$\widehat{\overline{Y}}^{*(10)}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*(10)}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*(10)}_{(Prop,PPS)}$

Expanding (29), we get

$$\widehat{\overline{Y}}^*_{(Prop,PPS)} - \bar{Y} = \bar{Y} + \bar{Y}$$
$$\left[(\Psi_{18} - 1) + \Psi_{18} \left\{ \xi_0 - \frac{1}{2}\theta\xi_1 - \frac{1}{2}\theta\xi_0\xi_1 + \frac{1}{8}(5 + 3\theta^2)\xi_1^2 - \Psi_{19}R\left\{ \xi_1 - \frac{1}{2}\theta\xi_1^2 \right\} \right\} \right]$$
(30)

From (30), the bias of $\widehat{\overline{Y}}^*_{(Prop,PPS)}$ is given by:

$$\text{Bias } (\widehat{\overline{Y}}^*_{(Prop,PPS)}) = \bar{Y}\left[(\Psi_{18} - 1) + \lambda\Psi_{18}\left\{ \frac{1}{8}(5 + 3\theta^2)C_v^2 - \frac{1}{2}\theta\rho_{uv}C_uC_v \right\} + \Psi_{19}R\lambda\frac{1}{2}\theta C_v^2 \right]$$
(31)

Squaring (31) and taking expectations, we obtain the MSE of $\widehat{\overline{Y}}^*_{(Prop,PPS)}$ as given by:

$$\text{MSE } (\widehat{\overline{Y}}^*_{(Prop,PPS)}) = \bar{Y}^2[1 + \Psi_{18}^2 A + \Psi_{19}^2 B - 2\Psi_{18}C - 2\Psi_{19}D + 2\Psi_{18}\Psi_{19}E], \quad (32)$$

where

$$A = 1 + \lambda\left\{ C_u^2 + \left(\frac{5}{4} + \theta^2 \right)C_v^2 - 2\theta\rho_{uv}C_uC_v \right\},$$

$$B = R^2\lambda C_v^2,$$

$$C = 1 + \lambda\left\{ \left(\frac{(5 + 3\theta^2)}{8} \right)C_v^2 - \frac{1}{2}\theta\rho_{uv}C_uC_v \right\},$$

$$D = R\lambda \frac{\theta C_v^2}{2},$$

$$E = R\lambda \{\theta C_v^2 - \rho_{uv} C_u C_v\}.$$

Differentiate (32) w.r.t Ψ_{18} and Ψ_{19}, we get the optimum values of Ψ_{18} and Ψ_{19} as given by:

$$\Psi_{18(opt)} = \left[\frac{BC - DE}{AB - E^2}\right],$$

and

$$\Psi_{19(opt)} = \left[\frac{AD - CE}{AB - E^2}\right].$$

Putting the optimum values of Ψ_{18} and Ψ_{19} in (32), we get the minimum MSE of $\widehat{\bar{Y}}^*_{(Prop,PPS)}$ as given by:

$$\text{MSE}(\widehat{\bar{Y}}^*_{(Prop,PPS)})_{min} \cong \bar{Y}^2 \left[1 - \frac{AD^2 + BC^2 - 2CDE}{AB - E^2}\right] \tag{33}$$

Numerical study

We carry out a numerical study to evaluate the performances of estimators. The following numerical expression is used to compute the percentage relative efficiency (*PRE*).

$$PRE = \frac{Var(\widehat{\bar{Y}}_u)}{MSE(i)} \times 100$$

where

$$(i = \widehat{\bar{Y}}_{R,PPS}, \widehat{\bar{Y}}_{P,PPS}, \widehat{\bar{Y}}_{Reg,PPS}, \widehat{\bar{Y}}_{Rao,PPS}, \widehat{\bar{Y}}_{BR,PPS}, \widehat{\bar{Y}}_{BP,PPS}, , \widehat{\bar{Y}}_{H1,PPS} \widehat{\bar{Y}}_{H2,PPS}, \widehat{\bar{Y}}_{EE,PPS},$$

$$\widehat{\bar{Y}}^*_{(GK,PPS)}, \widehat{\bar{Y}}^*_{(S,PPS)}, \widehat{\bar{Y}}^*_{(Prop,PPS)})$$

Population-I: (Source: Singh[53])
   Y = Expected total of fish in the year 1995,
   X = Expected total of fish in the year 1994,
   Z = Expected total of fish in the year 1993.
   Population-II: (Source: Punjab Bureau of Statistics[54])
   Y = Total number of beds on the 30th June 2021,
   X = Total allocated beds for COVID-19, 2021,
   Z = Beds used by COVID-19, 2021.
   Population-III: (Source: Punjab Bureau of Statistics[54])
   Y = Kids under age 5 whose childbirths are described listed with a public consultant,
   X = Kids aged 5–17 years who are involved in child labor during the last week,
   Z = Women aged 20–24 years who were first married before age 16.

Population-IV: (Source: Singh[53])
Y = Expected total of fish in the year 1995,
X = Expected total of fish in the year 1994,
Z = Expected total of fish in the year 1992.
    The summary statistics is given in Table 2 and results based on four populations are given in Tables 3–10. The simulation results are given in Tables 11–18.

Simulation analysis

We have produced four populations of size 5000 from a bivariate normal distribution with unlike covariance matrices. The population means and covariance matrices are given below:

Population-I:

$$\mu_1 = \begin{bmatrix} 500 \\ 500 \\ 500 \end{bmatrix}$$

and

$$\sum = \begin{bmatrix} 1000 & 800 & 810 \\ 800 & 850 & 820 \\ 810 & 820 & 840 \end{bmatrix}$$

Population-II:

$$\mu_1 = \begin{bmatrix} 500 \\ 500 \\ 500 \end{bmatrix}$$

and

$$\sum = \begin{bmatrix} 1500 & 870 & 820 \\ 870 & 900 & 800 \\ 820 & 800 & 740 \end{bmatrix}$$

Table 2. Summary statistics using real data sets.

Parameters	Population-I	Population-II	Population-III	Population-IV
N	69	36	36	69
n	15	7	6	14
$\bar{Y}$	4514.899	76.22889	660.1389	4514.899
$\bar{X}$	4954.435	14.77222	215.6389	4954.435
C_u	0.4720461	1.568599	0.7089214	0.8523346
C_v	0.5049075	0.7550356	0.7816491	0.8925777
ρ_{uv}	0.2660536	0.3004429	0.1852395	0.1542462
C_{uv}	0.06341113	0.3558289	0.1026463	0.1173466
$\beta_{2(v)}$	9.985055	2.567602	19.72718	9.985055

Table 3. MSE using Population-I.

Estimators		$\widehat{\overline{Y}}^*_{(S,PPS)}$	$\widehat{\overline{Y}}^*_{(GK,PPS)}$	$\widehat{\overline{Y}}^*_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	349398.9	349874.6	317890.6	306281.5
$\widehat{\overline{Y}}_{R,PPS}$	5502774	349701.8	317896.8	306291.6
$\widehat{\overline{Y}}_{P,PPS}$	947998	349900.5	317889.6	306280
$\widehat{\overline{Y}}_{Reg,PPS}$	324666.8	349761.7	317894.6	306288.1
$\widehat{\overline{Y}}_{Rao,PPS}$	319576.8	349898	317889.7	306280.1
$\widehat{\overline{Y}}_{BR,PPS}$	349903.4	349899.6	317889.7	306280
$\widehat{\overline{Y}}_{BP,PPS}$	548763.7	349795.1	317893.4	306286.2
$\widehat{\overline{Y}}_{H1,PPS}$	309030.9	349902.9	317889.6	306279.8
$\widehat{\overline{Y}}_{H2,PPS}$	315997.70	349154	317916.6	306323.9
$\widehat{\overline{Y}}_{EE,PPS}$	317893.26	347998.8	319576.5	309104

MSE: mean squared error.

Table 4. PRE using Population-I.

Estimators		$\widehat{\overline{Y}}^*_{(S,PPS)}$	$\widehat{\overline{Y}}^*_{(GK,PPS)}$	$\widehat{\overline{Y}}^*_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	100	99.86404	109.9117	114.0777
$\widehat{\overline{Y}}_{R,PPS}$	63.49505	99.91337	109.9095	114.0739
$\widehat{\overline{Y}}_{P,PPS}$	36.8565	99.85663	109.912	114.0783
$\widehat{\overline{Y}}_{Reg,PPS}$	107.6177	99.89625	109.9103	114.0752
$\widehat{\overline{Y}}_{Rao,PPS}$	109.3317	99.85734	109.912	114.0782
$\widehat{\overline{Y}}_{BR,PPS}$	99.8558	99.85688	109.912	114.0782
$\widehat{\overline{Y}}_{BP,PPS}$	63.67018	99.88673	109.9107	114.076
$\widehat{\overline{Y}}_{H1,PPS}$	113.0628	99.85596	109.912	114.0783
$\widehat{\overline{Y}}_{H2,PPS}$	110.568	100.0701	109.9027	114.0619
$\widehat{\overline{Y}}_{EE,PPS}$	109.91	100.4023	109.3318	113.036

PRE: percentage relative efficiency.

Table 5. MSE using Population-II.

Estimators		$\widehat{\overline{Y}}^*_{(S,PPS)}$	$\widehat{\overline{Y}}^*_{(GK,PPS)}$	$\widehat{\overline{Y}}^*_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	2042.513	1867.158	1380.42	1266.972
$\widehat{\overline{Y}}_{R,PPS}$	1924,985	1876.736	1386.883	1273.903
$\widehat{\overline{Y}}_{P,PPS}$	3106.509	1866.099	1379.451	1265.934
$\widehat{\overline{Y}}_{Reg,PPS}$	1858.144	1892.744	1393.463	1280.965
$\widehat{\overline{Y}}_{Rao,PPS}$	1407.928	1866.659	1379.973	1266.493
$\widehat{\overline{Y}}_{BR,PPS}$	1865.441	1868.404	1381.469	1268.096
$\widehat{\overline{Y}}_{BP,PPS}$	2456.203	1871.413	1383.692	1270.48
$\widehat{\overline{Y}}_{H1,PPS}$	1288.355	1865.91	1379.269	1265.739
$\widehat{\overline{Y}}_{H2,PPS}$	1357.061	1902.869	1396.35	1284.067
$\widehat{\overline{Y}}_{EE,PPS}$	1348.015	2034.616	1407.907	1296.493

MSE: mean squared error.

Table 6. PRE using Population-II.

Estimators		$\widehat{\overline{Y}}^{*}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*}_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	100	109.3916	147.9632	161.2122
$\widehat{\overline{Y}}_{R,PPS}$	106.1054	108.8333	147.2736	160.3351
$\widehat{\overline{Y}}_{P,PPS}$	65.74947	109.4536	148.0671	161.3444
$\widehat{\overline{Y}}_{Reg,PPS}$	109.9222	107.9128	146.5782	159.4511
$\widehat{\overline{Y}}_{Rao,PPS}$	109.9222	109.4208	148.0111	161.2731
$\widehat{\overline{Y}}_{BR,PPS}$	145.0723	109.3186	147.8508	161.0692
$\widehat{\overline{Y}}_{BP,PPS}$	109.4923	109.1428	147.6132	160.767
$\widehat{\overline{Y}}_{H1,PPS}$	83.15736	109.4647	148.0866	161.3692
$\widehat{\overline{Y}}_{H2,PPS}$	150.511	107.3386	146.2751	159.066
$\widehat{\overline{Y}}_{EE,PPS}$	151.520	100.3881	145.0744	157.5414

PRE: percentage relative efficiency.

Table 7. MSE using Population-III.

Estimators		$\widehat{\overline{Y}}^{*}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*}_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	31287.35	40038.97	31729.26	26349.13
$\widehat{\overline{Y}}_{R,PPS}$	56543.03	38983.65	31868.64	26559.61
$\widehat{\overline{Y}}_{P,PPS}$	82103.92	40135.31	31716.69	26330.22
$\widehat{\overline{Y}}_{Reg,PPS}$	30213.76	39332.51	31822.19	26489.27
$\widehat{\overline{Y}}_{Rao,PPS}$	28254.8	40127.88	31717.65	26331.68
$\widehat{\overline{Y}}_{BR,PPS}$	34406.16	40132.11	31717.1	26330.85
$\widehat{\overline{Y}}_{BP,PPS}$	47186.6	39620.37	31784.16	26431.83
$\widehat{\overline{Y}}_{H1,PPS}$	24431.36	40139.87	31716.09	26329.33
$\widehat{\overline{Y}}_{H2,PPS}$	26895.349	36482.52	32226.4	27108.34
$\widehat{\overline{Y}}_{EE,PPS}$	27802.337	36308.52	32610.92	27713.8

MSE: mean squared error.

Table 8. PRE using Population-III.

Estimators		$\widehat{\overline{Y}}^{*}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*}_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	100	91.16594	114.3605	138.5317
$\widehat{\overline{Y}}_{R,PPS}$	55.3337	93.63389	114.5386	137.4339
$\widehat{\overline{Y}}_{P,PPS}$	38.10701	90.94712	115.0874	138.6312
$\widehat{\overline{Y}}_{Reg,PPS}$	103.5533	92.80341	114.7058	137.7988
$\widehat{\overline{Y}}_{Rao,PPS}$	103.5533	90.96396	115.0839	138.6235
$\widehat{\overline{Y}}_{BR,PPS}$	111.9295	90.95436	115.0859	138.6279
$\widehat{\overline{Y}}_{BP,PPS}$	90.93532	92.12913	114.8431	138.0983
$\widehat{\overline{Y}}_{H1,PPS}$	66.30558	90.93677	115.0896	138.6359
$\widehat{\overline{Y}}_{H2,PPS}$	116.330	100.0531	113.2671	134.6519
$\widehat{\overline{Y}}_{EE,PPS}$	112.535	100.5326	111.9315	131.7102

PRE: percentage relative efficiency.

Table 9. MSE using Population-IV.

Estimators		$\widehat{Y}^*_{(S,PPS)}$	$\widehat{Y}^*_{(GK,PPS)}$	$\widehat{Y}^*_{(Prop,PPS)}$
$\widehat{Y}_{u,PPS}$	1057763	1176787	967856.1	874307.9
$\widehat{Y}_{R,PPS}$	1876050	1176083	967911	874391
$\widehat{Y}_{P,PPS}$	2559487	1176893	967847.9	874295.4
$\widehat{Y}_{Reg,PPS}$	1032596	1176327	967891.9	874362.1
$\widehat{Y}_{Rao,PPS}$	982810.8	1176892	967847.9	874295.5
$\widehat{Y}_{BR,PPS}$	1176905	1176896	967847.6	874295.1
$\widehat{Y}_{BP,PPS}$	1518623	1176144	967906.2	874383.7
$\widehat{Y}_{H1,PPS}$	895768.3	1176903	967847	874294.2
$\widehat{Y}_{H2,PPS}$	952013.35	1171676	968255	874911.8
$\widehat{Y}_{EE,PPS}$	973909.40	1055381	982807.9	897499

MSE: mean squared error.

Table 10. PRE using Population-IV.

Estimators		$\widehat{Y}^*_{(S,PPS)}$	$\widehat{Y}^*_{(GK,PPS)}$	$\widehat{Y}^*_{(Prop,PPS)}$
$\widehat{Y}_{u,PPS}$	100	89.88564	109.2892	120.9829
$\widehat{Y}_{R,PPS}$	56.38245	89.93946	109.2831	120.9714
$\widehat{Y}_{P,PPS}$	41.32714	89.87756	109.2902	120.9846
$\widehat{Y}_{Reg,PPS}$	102.4372	89.92077	109.2852	120.9754
$\widehat{Y}_{Rao,PPS}$	107.6263	89.87763	109.2902	120.9846
$\widehat{Y}_{BR,PPS}$	89.87666	89.87734	109.2902	120.9846
$\widehat{Y}_{BP,PPS}$	69.65273	89.93477	109.2836	120.9724
$\widehat{Y}_{H1,PPS}$	118.0844	89.87676	109.2903	120.9848
$\widehat{Y}_{H2,PPS}$	100.5374	90.27774	109.2442	120.8993
$\widehat{Y}_{EE,PPS}$	105.366739	100.2257	107.626627	117.8567

PRE: percentage relative efficiency.

Table 11. MSE using Population-I based on the simulation study.

Estimators		$\widehat{Y}^*_{(S,PPS)}$	$\widehat{Y}^*_{(GK,PPS)}$	$\widehat{Y}^*_{(Prop,PPS)}$
$\widehat{Y}_{u,PPS}$	0.4223217	0.1634819	0.08864383	0.05332498
$\widehat{Y}_{R,PPS}$	0.0899548	0.1634819	0.08864764	0.05332499
$\widehat{Y}_{P,PPS}$	1.49594	0.1636143	0.08864378	0.05332498
$\widehat{Y}_{Reg,PPS}$	0.0886746	0.1636143	0.08870236	0.05332507
$\widehat{Y}_{Rao,PPS}$	0.08896743	0.1634819	0.08864491	0.05332507
$\widehat{Y}_{BR,PPS}$	0.1634819	0.1636143	0.08866297	0.05332498
$\widehat{Y}_{BP,PPS}$	0.8664743	0.1636143	0.08864384	0.05332501
$\widehat{Y}_{H1,PPS}$	0.08896705	0.1634819	0.08864414	0.05332498
$\widehat{Y}_{H2,PPS}$	0.0889674	0.1636143	0.08864813	0.05332498
$\widehat{Y}_{EE,PPS}$	0.4223205	0.4059884	0.08896743	0.05332547

MSE: mean squared error.

Table 12. PRE using Population-I based on the simulation study.

Estimators		$\widehat{\overline{Y}}^*_{(S,PPS)}$	$\widehat{\overline{Y}}^*_{(GK,PPS)}$	$\widehat{\overline{Y}}^*_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	100	258.3294	476.4254	491.0944
$\overline{Y}_{R,PPS}$	469.4821	258.3294	476.4049	491.0942
$\overline{Y}_{P,PPS}$	28.2312	258.1202	476.4257	491.0948
$\overline{Y}_{Reg,PPS}$	474.6923	258.1202	476.111	491.0936
$\overline{Y}_{Rao,PPS}$	474.6925	258.3294	476.4196	491.0944
$\overline{Y}_{BR,PPS}$	258.3294	258.1202	476.3225	491.0942
$\overline{Y}_{BP,PPS}$	48.74025	258.1202	476.4253	491.0944
$\overline{Y}_{H1,PPS}$	474.6945	258.3294	476.4237	491.0945
$\overline{Y}_{H2,PPS}$	474.6927	258.1202	476.4023	491.0944
$\overline{Y}_{EE,PPS}$	100.0003	104.0231	474.6925	491.0899

PRE: percentage relative efficiency.

Table 13. MSE using Population-II based on the simulation study.

Estimators		$\widehat{\overline{Y}}^*_{(S,PPS)}$	$\widehat{\overline{Y}}^*_{(GK,PPS)}$	$\widehat{\overline{Y}}^*_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	1.097883	0.351711	0.1064713	0.1060176
$\overline{Y}_{R,PPS}$	0.1083106	0.351711	0.1064953	0.1060176
$\overline{Y}_{P,PPS}$	4.098543	0.3517714	0.106471	0.1060517
$\overline{Y}_{Reg,PPS}$	0.1082472	0.3517714	0.1068609	0.1060517
$\overline{Y}_{Rao,PPS}$	0.1082472	0.351711	0.1064787	0.1060176
$\overline{Y}_{BR,PPS}$	0.351711	0.3517714	0.106609	0.1060517
$\overline{Y}_{BP,PPS}$	2.346827	0.3517714	0.1064713	0.1060517
$\overline{Y}_{H1,PPS}$	0.108245	0.351711	0.1064735	0.1060176
$\overline{Y}_{H2,PPS}$	0.1082468	0.3517714	0.1064966	0.1060517
$\overline{Y}_{EE,PPS}$	1.097874	1.061622	0.1081982	0.1064735

MSE: mean squared error.

Table 14. PRE using Population-II based on the simulation study.

Estimators		$\widehat{\overline{Y}}^*_{(S,PPS)}$	$\widehat{\overline{Y}}^*_{(GK,PPS)}$	$\widehat{\overline{Y}}^*_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	100	312.1549	1031.154	1035.567
$\overline{Y}_{R,PPS}$	1013.643	312.1549	1030.921	1035.567
$\overline{Y}_{P,PPS}$	26.78716	312.1013	1031.157	1035.234
$\overline{Y}_{Reg,PPS}$	1014.237	312.1013	1027.395	1035.234
$\overline{Y}_{Rao,PPS}$	1014.237	312.1549	1031.083	1035.567
$\overline{Y}_{BR,PPS}$	312.1549	312.1013	1029.822	1035.234
$\overline{Y}_{BP,PPS}$	46.78159	312.1013	1031.154	1035.234
$\overline{Y}_{H1,PPS}$	1014.257	312.1549	1031.133	1035.567
$\overline{Y}_{H2,PPS}$	1014.241	312.1013	1030.697	1035.234
$\overline{Y}_{EE,PPS}$	100.0008	103.4115	1014.697	1031.133

PRE: percentage relative efficiency.

Table 15. MSE using Population-III based on the simulation study.

Estimators		$\widehat{\overline{Y}}^{*}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*}_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	0.60094	0.51537	0.482104	0.478010
$\widehat{\overline{Y}}_{R,PPS}$	0.9517461	0.51537	0.48214	0.478010
$\widehat{\overline{Y}}_{P,PPS}$	2.33747	0.51542	0.4821036	0.478013
$\widehat{\overline{Y}}_{Reg,PPS}$	0.4859471	0.51542	0.4827192	0.478013
$\widehat{\overline{Y}}_{Rao,PPS}$	0.4859461	0.51537	0.4821088	0.478010
$\widehat{\overline{Y}}_{BR,PPS}$	0.515426	0.51542	0.4822075	0.478013
$\widehat{\overline{Y}}_{BP,PPS}$	1.208288	0.51542	0.4821048	0.478013
$\widehat{\overline{Y}}_{H1,PPS}$	0.4859419	0.51537	0.4821052	0.478010
$\widehat{\overline{Y}}_{H2,PPS}$	0.4859467	0.51542	0.4821863	0.478013
$\widehat{\overline{Y}}_{EE,PPS}$	0.6009381	0.58768	0.4858253	0.48186

MSE: mean squared error.

Table 16. PRE using Population-III based on the simulation study.

Estimators		$\widehat{\overline{Y}}^{*}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*}_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	100	116.6020	124.6495	125.717
$\widehat{\overline{Y}}_{R,PPS}$	63.14079	116.6020	124.6401	125.715
$\widehat{\overline{Y}}_{P,PPS}$	25.70899	116.5909	124.6496	125.716
$\widehat{\overline{Y}}_{Reg,PPS}$	123.6637	116.5909	124.4906	125.716
$\widehat{\overline{Y}}_{Rao,PPS}$	123.6639	116.6020	124.6482	125.717
$\widehat{\overline{Y}}_{BR,PPS}$	116.5909	116.5909	124.6227	125.716
$\widehat{\overline{Y}}_{BP,PPS}$	49.73483	116.5909	124.6492	125.716
$\widehat{\overline{Y}}_{H1,PPS}$	123.665	116.6020	124.6491	125.717
$\widehat{\overline{Y}}_{H2,PPS}$	123.6638	116.5909	124.6282	125.716
$\widehat{\overline{Y}}_{EE,PPS}$	100.0003	102.2563	123.6947	124.710

PRE: percentage relative efficiency.

Table 17. MSE using Population-IV based on the simulation study.

Estimators		$\widehat{\overline{Y}}^{*}_{(S,PPS)}$	$\widehat{\overline{Y}}^{*}_{(GK,PPS)}$	$\widehat{\overline{Y}}^{*}_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	3.976607	1.535259	1.123713	1.118145
$\widehat{\overline{Y}}_{R,PPS}$	1.397566	1.535259	1.124288	1.118145
$\widehat{\overline{Y}}_{P,PPS}$	15.77027	1.535259	1.123705	1.118147
$\widehat{\overline{Y}}_{Reg,PPS}$	1.17434	1.535259	1.13242	1.118147
$\widehat{\overline{Y}}_{Rao,PPS}$	1.174335	1.535259	1.123868	1.118145
$\widehat{\overline{Y}}_{BR,PPS}$	1.535259	1.535259	1.126424	1.118147
$\widehat{\overline{Y}}_{BP,PPS}$	8.174274	1.535259	1.123715	1.118147
$\widehat{\overline{Y}}_{H1,PPS}$	1.174274	1.535259	1.123757	1.118145
$\widehat{\overline{Y}}_{H2,PPS}$	1.17433	1.535259	1.124399	1.118147
$\widehat{\overline{Y}}_{EE,PPS}$	3.976492	1.55843	1.174339	1.121434

MSE: mean squared error.

Table 18. PRE using Population-IV based on the simulation study.

Estimators		$\widehat{\overline{Y}}^*_{(S,PPS)}$	$\widehat{\overline{Y}}^*_{(GK,PPS)}$	$\widehat{\overline{Y}}^*_{(Prop,PPS)}$
$\widehat{\overline{Y}}_{u,PPS}$	100	259.1356	353.8811	355.6423
$\overline{Y}_{R,PPS}$	284.5381	259.1356	353.7001	355.6423
$\overline{Y}_{P,PPS}$	25.21584	259.0187	353.8835	355.6423
$\overline{Y}_{Reg,PPS}$	338.6248	259.0187	351.1601	355.6423
$\overline{Y}_{Rao,PPS}$	338.6264	259.1356	353.8322	355.6423
$\overline{Y}_{BR,PPS}$	259.0187	259.0187	353.0294	355.6423
$\overline{Y}_{BP,PPS}$	45.59487	259.0187	353.8804	355.6423
$\overline{Y}_{H1,PPS}$	338.6438	259.1356	353.8671	355.6423
$\overline{Y}_{H2,PPS}$	338.6279	259.0187	353.6651	355.6423
$\overline{Y}_{EE,PPS}$	100.0029	255.1670	338.3200	354.6000

PRE: percentage relative efficiency.

Population-III:

$$\mu_1 = \begin{bmatrix} 500 \\ 500 \\ 500 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 400 & 270 & 220 \\ 270 & 500 & 300 \\ 220 & 300 & 645 \end{bmatrix}$$

Population-IV:

$$\mu_1 = \begin{bmatrix} 50 \\ 50 \\ 50 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 70 & -50 & -52 \\ -57 & 50 & 30 \\ -52 & 50 & 85 \end{bmatrix}$$

Discussion

To calculate the achievability of the proposed estimators in comparison to the existing estimators, four data sets and a simulation analysis were performed. Four natural data sets were used in the empirical study. We also performed the simulation study, to check the reliability and generalizability of the new improved class of estimators. The consistency findings demonstrated that the proposed estimators were more accurate

and less biased than conventional and other well-known existing estimators. Table 2 provides summary statistics for the available datasets. Tables 3–10 contain the MSE and PRE results based on the real data sets. The numerical findings based on real data sets show that our suggested estimators are the best among all existing estimators. Tables 11–18 include the MSE and PRE results utilizing simulated data sets. The results of the simulation analysis also clearly show that the PRE of the proposed estimator is higher than the existing estimators, which are considered in this study. Therefore it observed from the numerical results that our proposed estimators are the best among all the existing counterparts.

From the numerical results, presented in Tables 3–18, we would like to remind that the MSE and percentage relative efficiency of all the proposed classes of estimators changeover according to different choices of a and b. Based on both real data sets and a simulation analysis, if we used ($a = 1$ and $b = \rho_{uv}$), ($C_v = 1$ and $b = \rho_{uv}$), ($a = \beta_{2(v)}$ and $b = \rho_{uv}$) we get the largest values of percentage relative efficiencies of all families among different classes. In this way, choosing a and b as the coefficient of variation, kurtosis, and association coefficient in the families of estimators give the best results. While from the numerical results the percentage relative efficiencies of our suggested family are declining across the values of ($a = 1$ and $b = N \bar{X}_1$). Greater improvements in efficiency are observed by using the proposed estimator over some existing estimators under probability proportional to size sampling. The results incorporated in this study are very sound and quite enlightening. Therefore, it is recommended that the proposed estimator is useful in practice.

Concluding remarks

In this article, we proposed an improved generalized class of estimators using two auxiliary information based on probability proportional to size sampling. Ten new estimators are generated from the proposed class of estimators, which are presented in Table 1. The proposed generalized class of estimators is compared with several existing estimators to judge their uniqueness and superiority using four real data sets. Moreover, a simulation study is also conducted to check the robustness and generalizability of the proposed estimator. The MSE of the proposed and existing estimators are derived up to the first order of approximation. The proposed class of estimators performs well as compared to its existing estimators, as shown by the results of four real data sets and a simulation study. It has been validated through empirical efficiency comparisons that our proposed class of estimators performs more effectively than the traditional estimators. The current work can be extended easily to an estimation of population means using the auxiliary variables based on measurement error, non-response, and stratified random sampling.

ORCID iD

Sohaib Ahmad https://orcid.org/0000-0003-2582-2265

References

1. Kadilar C and Cingi H. Improvement in estimating the population mean in simple random sampling. *Appl Math Lett* 2006; 19: 75–79.
2. Al-Omari AI. Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling. *Stat Probab Lett* 2012; 82: 1883–1890.
3. Ozturk O. Estimation of population mean and total in a finite population setting using multiple auxiliary variables. *J Agric Biol Environ Stat* 2014; 19: 161–184.
4. Yadav SK, Sharma DK, Mishra SS, et al. Use of auxiliary variables in searching efficient estimator of population mean. *Int J Multivar Data Anal* 2018; 1: 230–244.
5. Bhushan S and Pandey AP. Optimality of ratio-type imputation methods for estimation of population mean using higher order moment of an auxiliary variable. *J Stat Theory Pract* 2021; 15: 1–35.
6. Zaman T, Dünder E, Audu A, et al. Robust regression-ratio-type estimators of the mean utilizing two auxiliary variables: a simulation study. *Math Probl Eng* 2021; 2021: 1–9.
7. Kumar A and Saini M. A predictive approach for finite population mean when auxiliary variables are attributes. *Thailand Stat* 2022; 20: 575–584.
8. Singh HP and Nigam P. A generalized class of estimators for finite population mean using two auxiliary variables in sample surveys. *J Reliab StatStud* 2022; 15: 61–104.
9. Bhushan S, Kumar A, Onyango R, et al. Some improved classes of estimators in stratified sampling using bivariate auxiliary information. *J Probab Stat* 2022; 2022: 1–23.
10. Shahzad U, Ahmad I, Almanjahie IM, et al. Mean estimation using robust quantile regression with two auxiliary variables. *Sci Iran* 2022; 30: 1245–1254.
11. Mahdizadeh M and Zamanzade E. On interval estimation of the population mean in ranked set sampling. *Commun Stat - Simul Comput* 2022; 51: 2747–2768.
12. Ahmad S, Ullah K, Zahid E, et al. A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling. *Sci Rep* 2023; 13: 5415.
13. Muhammad I, Zakari Y, Abdu M, et al. Enhanced ratio-type estimator for finite population mean using auxiliary variable in simple random sampling. *Ratio (Oxf)* 2023; 5: 242–252.
14. Ahmad S, Hussain S, Shabbir J, et al. Improved generalized class of estimators in estimating the finite population mean using two auxiliary variables under two-stage sampling. *AIMS Mathematics* 2022; 7: 10609–10624.
15. Shahzad U, Ahmad I, Almanjahie IM, et al. Three-fold utilization of supplementary information for mean estimation under median ranked set sampling scheme. *Plos One* 2022; 17: e0276514.
16. Shahzad U, Ahmad I, Oral E, et al. Estimation of the population mean by successive use of an auxiliary variable in median ranked set sampling. *Math Popul Stud* 2021; 28: 176–199.
17. Yasmeen U, Noor ul Amin M and Hanif M. Generalized exponential estimators of finite population mean using transformed auxiliary variables. *Int J Appl Comput Math* 2015; 1: 589–598.
18. Singh HP, Malviya P and Tailor R. An alternative efficient class of estimators for finite population mean using information on an auxiliary attribute in sample surveys. *J Stat Theory Pract* 2023; 17: 2.
19. Singh S, Sedory SA and Arnab R. Estimation of finite population variance using scrambled responses in the presence of auxiliary information. *Commun Stat - Simul Comput* 2015; 44: 1050–1065.

20. Akpanta AC. On the problems of PPS sampling in multi-character surveys. *Global J Math Sci* 2009; 8: 31–42.
21. Agarwal SK and Al Mannai M. Linear combination of estimators in probability proportional to sizes sampling to estimate the population mean and its robustness to optimum value. *Statistica* 2009; 69.
22. Abdulla F, Hossain M and Rahman M. On the selection of samples in probability proportional to size sampling: cumulative relative frequency method. *Math Theor Model* 2014; 4: 102Á7.
23. Andersen IT, Hahn U and Vedel Jensen EB. Optimal PPS sampling with vanishing auxiliary variables–with applications in microscopy. *Scand J Stat* 2015; 42: 1136–1148.
24. Alam M, Sumy SA and Parh YA. Selection of the samples with probability proportional to size. *Sci J Appl Math Stat* 2015; 3: 230–233.
25. Patel PA and Bhatt S. Estimation of finite population total under PPS sampling in presence of extra auxiliary information. *Int J Stat Anal* 2016; 6: 9–16.
26. Singh HP, Mishra AC and Pal SK. Improved estimator of population total in PPS sampling. *Commun Stat - Theory Methods* 2018; 47: 912–934.
27. Makela S, Si Y and Gelman A. Bayesian Inference under cluster sampling with probability proportional to size. *Stat Med* 2018; 37: 3849–3868.
28. Ahmad S and Shabbir J. Use of extreme values to estimate finite population mean under PPS sampling scheme. *J Reliab Stat Stud* 2018; 11: 99–112.
29. Ozturk O. Post-stratified probability-proportional-to-size sampling from stratified populations. *J Agric Biol Environ Stat* 2019; 24: 693–718.
30. Latpate R, Kshirsagar J, Kumar Gupta V, et al. Probability proportional to size sampling. In: *Advanced sampling methods*. Singapore: Springer, 2021, pp.85–98.
31. Sohil F, Sohail MU and Shabbir J. Optimum second call imputation in PPS sampling. *PLoS One* 2022; 17: e0261834.
32. Sinha RR and Khanna B. Estimation of population mean under probability proportional to size sampling with and without measurement errors. *Concurrency Comput Pract Exper* 2022; 34: e7023.
33. Zangeneh SZ and Little RJ. Bayesian Inference for the finite population total from a heteroscedastic probability proportional to size sample. *J Surv Stat Methodol* 2015; 3: 162–192.
34. Hentschel B, Haas PJ and Tian Y. Exact PPS sampling with bounded sample size. *Inf Process Lett* 2023; 182: 106382.
35. Barbiero A, Manzi G and Mecatti F. Bootstrapping probability-proportional-to-size samples via calibrated empirical population. *J Stat Comput Simul* 2015; 85: 608–620.
36. Gupt BK and Ahamed MI. Optimum stratification for a generalized auxiliary variable proportional allocation under a superpopulation model. *Commun Stat-Theory Methods* 2022; 51: 3269–3284.
37. Ponkaew C and Lawson N. New estimators for estimating population total: an application to water demand in Thailand under unequal probability sampling without replacement for missing data. *PeerJ* 2022; 10: e14551.
38. Al-Jararha J. A class of sampling two units with probability proportional to size. *Commun Stat – Simul Comput* 2013; 42: 1906–1916.
39. Al-Marzouki S, Chesneau C, Akhtar S, et al. Estimation of finite population mean under PPS in presence of maximum and minimum values. *AIMS Math* 2021; 6: 5397–5409.
40. Zheng H and Little RJ. Penalized spline model-based estimation of the finite populations total from probability-proportional-to-size samples. *J Off Stat* 2003; 19: 99.
41. Zheng H and Little JA. Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. *J Off Stat* 2005; 21: 1.
42. Amab R. Optimum estimation of a finite population total in PPS sampling with replacement for multi-character surveys. *J Ind Soc Agril Statist* 2004; 58: 231–243.

43. Olayiwolla OM, Apantaku FS, Wale-Orojo OA, et al. Probability proportional to size (PPS) method to enhance efficiency of estimator in two stage sampling. *Ann Comput Sci Ser* 2019; 17: 311–315.
44. Murthy MN. Product method of estimation. *Sankhya: Indian J Stat, Ser A* 1964; 26: 69–74.
45. Bai ZD, Miao BQ and Rao CR, Estimation of direction of arrival of signals: Asymptotic results. In: Haykin S (ed) *Advances in spectrum analysis and array processing*, vol. II, Chapter 9, Englewood, Cliffs, NJ: Prentice Hall, 1991.
46. Bahl S and Tuteja R. Ratio and product type exponential estimators. *J Inf Optim Sci* 1991; 12: 159–164.
47. Abdul HAQ and Shabbir J. Improved exponential type estimators of finite population mean under complete and partial auxiliary information. *Hacettepe J Math Stat* 2014; 43: 1079–1093.
48. Ekpenyong EJ and Enang EI. Efficient exponential ratio estimator for estimating the population mean in simple random sampling. *Hacettepe J Math Stat* 2015; 44: 689–705.
49. Singh R, Chauhan P, Sawan N, et al. Improvement in estimating the population mean using exponential estimator in simple random sampling. *Int J Stat Econ* 2009; 3: 13–18.
50. Grover LK and Kaur P. A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. *Commun Stat - Simul Comput* 2014; 43: 1552–1574.
51. Ahmad S, Hussain S, Zahid E, et al. A simulation study: Population distribution function estimation using dual auxiliary information under stratified sampling scheme. *Math Probl Eng* 2022; 2022: 1–13.
52. Ahmad S, Hussain S, Aamir M, et al. Dual use of auxiliary information for estimating the finite population mean under the stratified random sampling scheme. *J of Math* 2021; 2021: 1–12.
53. Singh S. *Advanced sampling theory with applications: How Michael "selected" Amy*. Berlin, Germany: Springer Science & Business Media, 2003.
54. Punjab Bureau of Statistics, Punjab, Pakistan, 2021–2022. https://bos.punjab.gov.pk

Author biographies

Sohaib Ahmad is a PhD scholar at Abdul Wali Khan University Mardan. His research interests include survey sampling, randomized response, and Data analysis. He published several research articles in the same field.

Javid Shabbir is a professor in the Department of Statistics, University of Wah, Pakistan. His research direction is advanced survey sampling and randomized response.

Erum Zahid is working in the Department of Applied Mathematics and Statistics, Institute of Space Technology Islamabad, Pakistan. Her research direction includes survey sampling, spatial statistics and data analysis.

Muhammad Aamir working as an assistant professor at Abdul Wali Khan University, Mardan, Pakistan. His research direction is survey sampling, time series analysis, machine learning, and he has deep insights on the accuracy of forecasting models.

Mohammed Alqawba is working in the Department of Mathematics, College of Science and Arts, Qassim University, Ar Rass, Saudi Arabia. Hir research direction includes time series analysis, survey sampling, distribution theory and stochastic processes.

Appendix

Notations and symbols

Consider a finite population $\Omega = \{\Omega_1, \Omega_2, \ldots, \Omega_N\}$. Let y_i and $\{x_i, z_i\}$ be the characteristics of the study variable (Y) and the auxiliary variable (X, Z) respectively. We draw a sample of size n by using PPS sampling taking Z as size of the unity, i.e., $H_i = \frac{z_i}{\sum_{i=1}^{n} z_i}$, be the PPS sampling for obtaining the units. We draw a sample of size n by adopting the PPS sampling with replacement.

Define

$$u_i = \frac{y_i}{NH_i} = v_i = \frac{x_i}{NH_i},$$

$\bar{u} = \frac{\sum_{i=1}^{n} u_i}{n} = \bar{y}_{pps}, \bar{v} = \frac{\sum_{i=1}^{n} v_i}{n} = \bar{x}_{pps}$, be the sample means corresponding to population means $\bar{Y}$ and $\bar{X}$.

We consider the following error terms for obtaining the properties, i.e., bias and MSE of the estimators is given by:

$$\text{Let } \xi_0 = \frac{\bar{u} - \bar{Y}}{\bar{Y}}, \ \xi_1 = \frac{\bar{v} - \bar{X}}{\bar{X}}, \ E(\xi_i) = 0, i = 0, 1, .$$

$$E(\xi_0^2) = \frac{C_u^2}{n}, \ E(\xi_1^2) = \frac{C_v^2}{n}, \ E(\xi_0 \xi_1) = \lambda \rho_{uv} C_u C_v,$$

$C_u = \frac{S_u}{\bar{Y}}, \ C_v = \frac{S_v}{\bar{X}}$, be the population coefficient of variation such that

$$S_u = \sqrt{\sum_{i=1}^{N} H_i (u_i - \bar{Y})^2}$$

and

$$S_u = \sqrt{\sum_{i=1}^{N} H_i (v_i - \bar{X})^2}$$

Let $\rho_{uv} = \frac{\sum_{i=1}^{N} H_i (u_i - \bar{Y}) (v_i - \bar{X})}{S_u S_v}$, be the correlation coefficient of u on v.