



OPEN

A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling

Sohaib Ahmad¹, Kalim Ullah², Erum Zahid³, Javid Shabbir^{4,5}, Muhammad Aamir¹, Huda M. Alshanbari⁶✉ & Abd Al-Aziz Hosni El-Bagoury⁷

This article aims to suggest a new improved generalized class of estimators for finite population distribution function of the study and the auxiliary variables as well as mean of the usual auxiliary variable under simple random sampling. The numerical expressions for the bias and mean squared error (MSE) are derived up to first degree of approximation. From our generalized class of estimators, we obtained two improved estimators. The gain in second proposed estimator is more as compared to first estimator. Three real data sets and a simulation are accompanied to measure the performances of our generalized class of estimators. The MSE of our proposed estimators is minimum and consequently percentage relative efficiency is higher as compared to their existing counterparts. From the numerical outcomes it has been shown that the proposed estimators perform well as compared to all considered estimators in this study.

Generally it is a well-established notion that when the auxiliary variable is used appropriately in survey sampling, precision of the estimator is increased. Numbers of estimators exist in the literature for estimating different population parameters, such as mean, variance and total etc. but little attention has been paid to study the distribution (DF). The works on survey sampling discusses a diversity of procedures for incorporating the auxiliary variable via ratio, product, and regression methods of estimation. Numerous researchers have suggested various estimators by adequately adapting the auxiliary variable. These research findings can be investigated by Grover and Kaur¹ suggested a generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. Ahmad et al.² discussed use of extreme values to estimate the finite population mean under PPS sampling scheme. Audu et al.³ suggested on the efficiency of almost unbiased mean imputation when population mean of auxiliary variable is unknown. Singh and Nigam⁴ discussed efficient method of estimating the finite population mean based on two auxiliary variables in the presence of non-response under stratified sampling. Shahzad et al.⁵ discussed estimation of the population mean by successive use of an auxiliary variable in median ranked set sampling. Singh et al.⁶ proposed some imputation methods to deal with the problems of missing data in two-occasion successive sampling. Aggarwal et al.⁷ discussed estimation of the population mean by developing a new estimator. Singh et al.⁶ suggested an exponential approach for estimating population mean using two auxiliary variables in stratified random sampling. Yadav et al.⁸ proposed new modified ratio type estimator of the population mean using the known median of the study variable. Pal and Singh⁹ discussed about estimation of finite population mean using auxiliary information in presence of non-response. Pal and Singh¹⁰ proposed an efficient new approach for estimating the general parameter using

¹Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan. ²Foundation University Medical College, Foundation University School of Health Sciences, DHA-I, Islamabad 44000, Pakistan. ³Department of Applied Mathematics and Statistics, Institute of Space Technology, Islamabad, Pakistan. ⁴Department of Statistics, Quaid-I-Azam University, Islamabad, Pakistan. ⁵Department of Statistics, University of Wah at Wah Cantt, Punjab, Pakistan. ⁶Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. ⁷Basic Sciences Department, Higher Institute of Engineering and Technology, EL-Mahala EL-Kobra, EL-Mahalla EL-Kubra, Egypt. ✉email: hmalshanbari@pnu.edu.sa

auxiliary variable in sample surveys. Shahzad et al.¹¹ suggested quantile regression-ratio-type estimators for mean estimation under complete and partial auxiliary information. Zaman et al.¹² discussed robust ratio-type estimators for finite population mean in simple random sampling. Singh et al.¹³ discussed some efficient classes of estimators of population mean in two-phase successive sampling under random non response. Singh and Khalid¹⁴ proposed some imputation methods to compensate with non-response for estimation of population mean in two-occasion successive sampling. Irfan et al.¹⁵ proposed difference-type-exponential estimators based on dual auxiliary information under simple random sampling. Riyaz et al.¹⁶ discussed generalized exponential ratio estimator of population mean using two auxiliary variables in simple random sampling with an application to agricultural data. Zaman et al.¹⁷ proposed robust regression-ratio-type estimators of the mean utilizing two auxiliary variables. For assessing the finite population mean, researchers have suggested various improved ratio, product, and regression type estimators in their work.

There are numerous estimators available in literature for estimating different finite population parameters under various sampling designs, but the study based on distribution function (DF) has received less attention. When it needs to determine the percentage of particular values that are small or equal to the threshold value, the estimation of a finite population DF becomes necessary. For example a doctor is interested to distinguish the percentage of the population who consume at least 30% of their energy from a dietary cholesterol. A soil researcher is interested to determine how many peoples in a developing nation live below the poverty level. In all these circumstances, survey sampling significantly relies on estimation of finite population distribution function. In the area of DF, some significant work includes Chambers and Dunstan¹⁸ suggested estimating distribution functions from survey data. Chambers et al.¹⁹ discussed properties of estimators of the finite population distribution function. Dorfman²⁰ discussed a comparison of design-based and model-based estimators of the finite population distribution function. Ahmad and Abu-Dayyah²¹ suggested estimation of finite-population distribution function using multivariate auxiliary information. Singh et al.²² discussed a family of estimators of finite-population distribution function using auxiliary information. Onsongo et al.²³ discussed bias reduction technique for estimating finite population distribution function under simple random sampling without replacement. Ahmad et al.²⁴ discussed a new generalized class of exponential factor-type estimators for population distribution function using two auxiliary variables.

The continuing of the article is ordered as follows. Notations and symbols of the article are given in “Notations and symbols” section. The literature evaluation of the various estimators based on simple random sampling is introduced in “Review of existing estimators” section. The suggested a generalized class of estimators is given in Section 4. In Section 5, numerical investigation and data description are provided. In Section 6, a simulation study is given. Section 7 provides discussion of the article. Finally conclusion of the article is presented in Section 8.

Notations and symbols

Let a finite population $\mathcal{U} = (\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_N)$ consist of N identified and distinct units. A sample of size n is selected from \mathcal{U} by using simple random sampling without replacement. Let Y_i and X_i be the values of the study variable (Y) and the auxiliary variable (X) respectively. Let \bar{Y} and \bar{X} be the population means respectively corresponding to the sample means \bar{y} and \bar{x} . The indicator function of the study variable and the auxiliary variable are represented by $I(Y_i \leq Y)$, and $I(X_i \leq X)$. Let the finite population distribution function for the population and sample of the study variable and the auxiliary variable are: $F(y) = \frac{\sum_{i=1}^N I(Y_i \leq Y)}{N}$, $\hat{F}(y) = \frac{\sum_{i=1}^n I(Y_i \leq Y)}{n}$, $F(x) = \frac{\sum_{i=1}^N I(X_i \leq X)}{N}$, $\hat{F}(x) = \frac{\sum_{i=1}^n I(X_i \leq X)}{n}$.

To attain the bias and mean square error we consider the following relative error terms:

$$\xi_0 = \frac{\hat{F}_y - F_y}{F_y}, \xi_1 = \frac{\hat{F}_x - F_x}{F_x}, \xi_2 = \frac{\bar{x} - \bar{X}}{\bar{X}}, \text{ such that}$$

$$E(\xi_0^2) = \lambda C_{F(y)}^2 = \Theta_{200}, E(\xi_1^2) = \lambda C_{F(x)}^2 = \Theta_{020}, E(\xi_2^2) = \lambda C_x^2 = \Theta_{002},$$

$$E(\xi_0 \xi_1) = \lambda \rho_{F(y)F(x)} C_{F(y)} C_{F(x)} = \Theta_{110}, E(\xi_0 \xi_2) = \lambda \rho_{F(y)x} C_{F(y)} C_x = \Theta_{101},$$

$$E(\xi_1 \xi_2) = \lambda \rho_{F(x)x} C_{F(x)} C_x = \Theta_{011}$$

$$S_{F(y)}^2 = \frac{\sum_{i=1}^N \{I(Y_i \leq Y) - F(y)\}^2}{N-1}, S_{F(x)}^2 = \frac{\sum_{i=1}^N \{I(X_i \leq X) - F(x)\}^2}{N-1}, S_x^2 = \frac{\sum_{i=1}^N \{X_i - \bar{X}\}^2}{N-1}$$

$$C_{F(y)}^2 = \frac{S_{F(y)}^2}{F^2(y)}, C_{F(x)}^2 = \frac{S_{F(x)}^2}{F^2(x)}, C_x^2 = \frac{S_x^2}{\bar{X}^2}$$

$$\rho_{F(y)F(x)} = \frac{\sum_{i=1}^N [\{I(Y_i \leq Y) - F(y)\} \{I(X_i \leq X) - F(x)\}]}{\sum_{i=1}^N \{I(Y_i \leq Y) - F(y)\} \{I(X_i \leq X) - F(x)\}},$$

$$\rho_{F(y)x} = \sum_{i=1}^N [\{I(Y_i \leq Y) - F(y)\}\{X_i - \bar{X}\}],$$

$$\rho_{F(x)x} = \sum_{i=1}^N [\{I(X_i \leq X) - F(x)\}\{X_i - \bar{X}\}],$$

$$\lambda = \left(\frac{1}{n} - \frac{1}{N}\right).$$

Review of existing estimators

In this section, we present some existing estimators of finite population DF.

1. The usual estimator of $F(y)$, is given by:

$$\hat{F}_{usual} = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq Y). \tag{1}$$

The variance of \hat{F}_{usual} , is given by:

$$\text{Var}(\hat{F}_{usual}) = F_{(y)}^2 \Theta_{200}. \tag{2}$$

2. Cochran²⁵ presented the following estimator:

$$\hat{F}(R) = \hat{F}(y) \left(\frac{F(x)}{\hat{F}(x)}\right). \tag{3}$$

The bias and MSE of $\hat{F}(R)$, are given by

$$\text{Bias}(\hat{F}(R)) \cong F(y)(\Theta_{020} - \Theta_{110}),$$

$$\text{MSE}(\hat{F}(R)) \cong F^2(y)(\Theta_{200} + \Theta_{020} - 2\Theta_{110}). \tag{4}$$

3. Murthy²⁶ suggested the following estimator of $F(y)$ as:

$$\hat{F}(P) = \hat{F}(y) \left(\frac{\hat{F}(x)}{F(x)}\right). \tag{5}$$

The bias and MSE of $\hat{F}(P)$, are given by

$$\text{Bias}(\hat{F}(P)) \cong F(y)\Theta_{110},$$

$$\text{MSE}(\hat{F}(P)) \cong F^2(y)(\Theta_{200} + \Theta_{020} + 2\Theta_{110}). \tag{6}$$

4. Haq and Shabbir²⁷ suggested the following two estimators, which are given by:

$$\hat{F}_{HS,1} = [\psi_1 \hat{F}_{BT,A} + \psi_2 \{F(x) - \hat{F}(x)\}] \exp\left(\frac{F(x) - \hat{F}(x)}{\hat{F}(x) + F(x)}\right) \tag{7}$$

where $\psi_i (i = 1, 2, 3, 4)$ are constants and

$$\hat{F}_{BT,A} = \frac{\hat{F}(y)}{2} \left[\exp\left(\frac{F(x) - \hat{F}(x)}{\hat{F}(x) + F(x)}\right) + \exp\left(\frac{\hat{F}(x) - F(x)}{\hat{F}(x) + F(x)}\right) \right].$$

The bias of $\hat{F}_{HS,1}$ is given by:

$$\text{Bias}(\hat{F}_{HS,1}) \cong \frac{1}{2} \left[-2F(y) + F(y) \left\{ 2 + \lambda C_{F(x)} (2C_{F(x)} - \rho_{F(y)F(x)}) \right\} \psi_1 + F(x) \lambda C_{F(x)}^2 \psi_2 \right]$$

where ψ_1 and ψ_2 are constants. The optimum value of ψ_1 and ψ_2 are given by:

$$\psi_{1(\text{opt})} = \frac{4}{\left[4 + \lambda C_{F(x)}^2 - 4\lambda C_{F(y)}^2 (-1 + \rho_{F(y)F(x)}^2) \right]},$$

$$\psi_{2(\text{opt})} = \frac{F(y)}{2F(x)} \left[1 + \left(\frac{-8C_{F(x)} + 8C_{F(y)}\rho_{F(y)F(x)}}{C_{F(x)} \{4 + \lambda C_{F(x)}^2 - 4\lambda C_{F(y)}^2 (-1 + \rho_{F(y)F(x)}^2)\}} \right) \right]$$

The minimum MSE of $\widehat{F}_{HS,1}$ is given by:

$$\text{MSE}(\widehat{F}_{HS,1})_{\min} \cong \frac{\lambda F^2(y) [-\lambda C_{F(x)}^4 + 4(1 - \rho_{F(y)F(x)}^2)(-4 + \lambda C_{F(x)}^4) F^2(y)]}{4 [4 + \lambda C_{F(x)}^4 - 4\lambda C_{F(y)}^4 (-1 - \rho_{F(y)F(x)}^2)]} \tag{8}$$

and

$$\widehat{F}_{HS,2} = \frac{\psi_3}{2} \widehat{F}_{BT,A} \left\{ \frac{F(x)}{\widehat{F}(x)} + \frac{\widehat{F}(x)}{F(x)} \right\} + \psi_4 \{F(x) - \widehat{F}(x)\} \exp\left(\frac{F(x) - \widehat{F}(x)}{\widehat{F}(x) + F(x)}\right) \tag{9}$$

The bias of $\widehat{F}_{HS,2}$ are given by:

$$\text{Bias}(\widehat{F}_{HS,2}) \cong \frac{1}{2} [-2F(y) + F(y)] \left\{ 2 + \lambda C_{F(x)} (2C_{F(x)} - \rho_{F(y)F(x)} C_{F(y)}) \right\} \psi_3 + F(x) \lambda C_{F(x)}^2 \psi_4$$

where ψ_3 and ψ_4 are constants. The optimum value of ψ_3 and ψ_4 are given by:

$$\psi_{3(\text{opt})} = \frac{4 + 2\lambda C_{F(y)}^2}{[4 + 5\lambda C_{F(x)}^2 - 4\lambda C_{F(y)}^2 (-1 + \rho_{F(y)F(x)}^2)]}$$

$$\psi_{4(\text{opt})} = \left[\frac{F(y) \left\{ 8C_{F(y)}\rho_{F(y)F(x)} + C_{F(x)} \left\{ -4 + \lambda (C_{F(x)}^2 + 4\rho_{F(y)F(x)} - C_{F(y)} C_{F(x)} (-1 + \rho_{F(y)F(x)}^2)) \right\} \right\}}{2F(x) C_{F(x)} [4 + 5\lambda C_{F(x)}^2 - 4\lambda C_{F(y)}^2 (-1 + \rho_{F(y)F(x)}^2)]} \right]$$

The minimum MSE of $\widehat{F}_{HS,1}$ at the optimum values of ψ_3 and ψ_4 , is given by:

$$\text{MSE}(\widehat{F}_{HS,2})_{\min} \cong \frac{\lambda F^2(y) [-9\lambda C_{F(x)}^4 + 4(1 + \rho_{F(y)F(x)}^2)(-4 + \lambda C_{F(x)}^2) F^2(y)]}{4 [4 + 5\lambda C_{F(x)}^4 - 4\lambda C_{F(y)}^4 (-1 - \rho_{F(y)F(x)}^2)]} \tag{10}$$

5. The regression estimator of $F(y)$, is given by:

$$\widehat{F}_{Reg} = \widehat{F}(y) + \psi_5 [F(x) - \widehat{F}(x)], \tag{11}$$

where ψ_5 is constant. The optimum value of ψ_5 is

$$\psi_{5(\text{opt})} = \frac{F(y)\Theta_{110}}{F(y)\Theta_{020}} \tag{12}$$

The minimum variance of \widehat{F}_{Reg} , is given by:

$$\text{Var}(\widehat{F}_{Reg})_{\min} = F^2(y)\Theta_{200} (1 - \rho_{F(y)F(x)}^2). \tag{13}$$

where $\rho_{F(y)F(x)} = \frac{\Theta_{110}}{\sqrt{\Theta_{110}\Theta_{020}}}$

6. Following Bahl and Tuteja²⁸, exponential estimators of $F(y)$, are given by:

$$\widehat{F}_{BTR} = \widehat{F}(y) \exp\left(\frac{F(x) - \widehat{F}(x)}{\widehat{F}(x) + F(x)}\right) \tag{14}$$

and

$$\widehat{F}_{BTP} = \widehat{F}(y) \exp\left(\frac{\widehat{F}(x) - F(x)}{\widehat{F}(x) + F(x)}\right). \tag{15}$$

The bias and MSE of \widehat{F}_{BTR} and \widehat{F}_{BTP} , are given by

$$\text{Bias}(\widehat{F}_{BTR}) \cong F(y) \left(\frac{3}{8}\Theta_{020} - \frac{1}{2}\Theta_{110} \right),$$

$$MSE(\widehat{F}_{BTR}) \cong \frac{F^2(y)}{4}(4\Theta_{200} + \Theta_{020} - 4\Theta_{110}) \tag{16}$$

and

$$\begin{aligned} \text{Bias}(\widehat{F}_{BTP}) &\cong F(y) \left(\frac{1}{2}\Theta_{110} - \frac{1}{8}\Theta_{020} \right), \\ MSE(\widehat{F}_{BTP}) &\cong \frac{F^2(y)}{4}(4\Theta_{200} + \Theta_{020} + 4\Theta_{110}). \end{aligned} \tag{17}$$

7. Shabbir and Gupta²⁹ suggested estimator for $F(y)$:

$$\widehat{F}_{SG} = \psi_6 \widehat{F}(y) + \psi_7 [F(x) - \widehat{F}(x)], \tag{18}$$

where ψ_6 and ψ_7 are unknown constants. The optimum values are

$$\psi_6 = \frac{\Theta_{020}}{8} \left[\frac{8 - \Theta_{020}}{\Theta_{020} + \Theta_{200}\Theta_{020} - \Theta_{220}^2} \right],$$

and

$$\psi_7 = \frac{\widehat{F}(y)}{8F(x)} \left[\frac{-4\Theta_{200} + \Theta_{020} + 8\Theta_{220} - \Theta_{220}\Theta_{020} + 4\Theta_{200} - 4\Theta_{220}^2}{\Theta_{200} + \Theta_{200}\Theta_{020} - \Theta_{220}^2} \right].$$

The bias and minimum MSE of \widehat{F}_{SG} , are given as:

$$\begin{aligned} \text{Bias}(\widehat{F}_{SG}) &\cong F(y)(\psi_6 - 1) \\ MSE(\widehat{F}_{SG})_{min} &\cong \frac{F^2(y)}{64} \left[\frac{-4\Theta_{020}^2 - 16\Theta_{020}(1 - \rho_{F(y)F(x)}^2)(\Theta_{020} - 4)}{1 + \Theta_{200}(1 - \rho_{F(y)F(x)}^2)} \right]. \end{aligned} \tag{19}$$

8. Swain³⁰ suggested the following estimator for $F(y)$ and is, given by:

$$\widehat{F}_{SW} = \widehat{F}(y) \left[\alpha \left(\frac{F(x)}{\widehat{F}(x)} \right)^G (1 - \alpha) \left(\frac{\widehat{F}(x)}{F(x)} \right)^Q \right]^\sigma \tag{20}$$

where α is the unknown constant; and G, Q and σ are scaler quantity. The bias of \widehat{F}_{SW} is given by:

$$\text{Bias}(\widehat{F}_{SW}) = F(y)[A_1\Theta_{110} + A_2\Theta_{020}],$$

$$A_1 = \alpha\{(G + Q) - Q\} \delta$$

$$A_2 = [\alpha\{G(G + 1) - Q(Q - 1)\} + Q(Q - 1)] \delta$$

The optimum value of α i.e. α_{opt} is given by:

$$\alpha_{opt} = \frac{1}{Q - G} \left[Q(Q - 1) + \frac{\Theta_{110}}{\Theta_{020}} \right]$$

The minimum MSE of \widehat{F}_{SW} is given by

$$MSE(\widehat{F}_{SW}) \cong F^2(y)\Theta_{200} \left(1 - \rho_{F(y)F(x)}^2 \right) = \text{Var}(\widehat{F}_{Reg})_{min} \tag{21}$$

9. Chami et al.³¹ suggested the following estimator is given by:

$$\hat{F}_C = \hat{F}(y) \left\{ \psi_9 \frac{(1 - \beta_1)\hat{F}(x) + \beta_1 F(x)}{\beta_1 \hat{F}(x) + (1 - \beta_1)F(x)} \right\} + (1 - \psi_9) \left\{ \frac{\beta_1 \hat{F}(x) + (1 - \beta_1)F(x)}{\alpha(1 - \beta_1) + \beta_1 F(x)} \right\} \tag{22}$$

where ψ_9 and β_1 are constants.
The bias of \hat{F}_C is given by:

$$\text{Bias}(\hat{F}_C) = F(y) [(1 - 2\beta_1)(2\psi_9 - 1)\Theta_{110} + (1 - 2\beta_1)(1 - \beta_1 - \psi_9)\Theta_{020}].$$

Putting the values of ψ_9 and β_1 we get the minimum variance of \hat{F}_C , and is equal to $\text{Var}(\hat{F}_{Reg})_{min}$.
10. Yadav et al.³² suggested the following estimator is given by:

$$\hat{F}_{YG} = [\psi_{11}\hat{F}(y) + \psi_{12}\{F(x) - \hat{F}(x)\}] \left[\gamma \left(\frac{\alpha F(x) + \beta}{\alpha \hat{F}(x) + \beta} \right) + (1 - \gamma) \left(\frac{\alpha(F(x) - \hat{F}(x))}{\alpha(F(x) - \hat{F}(x)) + \beta} \right) \right]. \tag{23}$$

where ψ_{11} and ψ_{12} are constants, and γ is the scaler quantity.
The optimum values are:

$$\psi_{11} = \left[\frac{1 - \frac{1}{8}G^2(1 + 3\gamma + 4\gamma^2)\theta_{020}}{1 - \frac{1}{8}G^2(1 + 3\gamma)\theta_{020}^2 + \theta_{020}(1 - \rho_{F(y)F(x)}^2)} \right],$$

and

$$\psi_{12} = \frac{F(y)}{F(x)} \left[\frac{1}{2}G(1 - \gamma) + \psi_{11} \left(\frac{\theta_{110}}{\theta_{020}} - G(1 - \gamma) \right) \right],$$

$$G = \frac{\alpha}{F(x) + \beta}.$$

The bias of \hat{F}_{YG} is given by

$$\text{Bias}(\hat{F}_{YG}) = (\psi_{11} - 1)F(y) + \psi_{11}F(y) \left[\frac{1}{8}G^2(3 + 5\lambda)\theta_{020} - \frac{1}{2}G(1 + \lambda)\theta_{110} \right] + \frac{1}{2}\psi_{12}F(x)G(1 + \lambda)\theta_{020}$$

The minimum MSE of \hat{F}_{YG} is

$$\text{MSE}(\hat{F}_{YG})_{min} \cong F^2(y) \left[\frac{\frac{1}{F^2(y)} \text{Var}_{min}(\hat{F}_{Reg})(1 - \Theta_{020})}{1 - \Theta_{020} + \frac{1}{F^2(y)} \text{Var}_{min}(\hat{F}_{Reg})} \right] \tag{24}$$

11. Yaqub and Shabbir³³ suggested the following estimator, given by:

$$\hat{F}_{YS} = F(y) \left[\left\{ \psi_{14}\hat{F}(y) + \psi_{15}(F(x) - \hat{F}(x)) \right\} \left(\frac{\alpha F(x) + \beta}{\alpha \hat{F}(x) + 2\beta} \right) \right] \left[\frac{1}{2} \exp \left(\frac{\alpha(F(x) - \hat{F}(x))}{\alpha(F(x) - \hat{F}(x)) + 2\beta} \right) + \frac{1}{2} \exp \left(\frac{\alpha(\hat{F}(x) - F(x))}{\alpha(\hat{F}(x) - F(x)) + \beta} \right) \right] \tag{25}$$

where ψ_{14} and ψ_{15} are constants.
The bias of \hat{F}_{YS} is given by:

$$\text{Bias} = (\psi_{14} - 1)F(y) (\hat{F}_{YS}) + \frac{9}{8}\psi_{14}F(y)\theta^2\Theta_{020} - \psi_{14}F(y)\theta\Theta_{110} + \psi_{15}F(x)\theta\Theta_{020}$$

The optimum values are ψ_{14} and ψ_{15} are given by:

$$\psi_{14(\text{opt})} = \frac{\Theta_{020}}{2} \left[\frac{1 + 7(1 - \Theta_{020})}{\Theta_{020}^2 + 4\{\Theta_{020}(1 - \rho_{F(y)F(x)}^2)\} - \Theta_{110}^2} \right]$$

$$\psi_{15(\text{opt})} \frac{F(y)}{F(x)} \left[\frac{\Theta_{110} + (7\Theta_{110} - 8\Theta_{020})(1 - \Theta_{020}) + 8\Theta_{200}\Theta_{020} - 8\Theta_{110}^2}{\Theta_{020}^2 + 4\Theta_{020}(1 - \Theta_{020} + \Theta_{200}) - \Theta_{110}^2} \right]$$

The minimum MSE of \widehat{F}_{YS} is given by:

$$\text{MSE}(\widehat{F}_{YS}) \cong \frac{F^2(y)}{16} \left[\frac{\frac{64}{F^2(y)}(1 - \Theta_{020})\text{MSE}(\widehat{F}_{Reg}) - \Theta_{020}^2}{\Theta_{020} + 4(1 - \Theta_{200}) + 2\Theta_{110} + \frac{4}{F^2(y)}\text{MSE}(\widehat{F}_{Reg})} \right] \tag{26}$$

12. Muneer et al.³⁴ suggested the following estimator, given by:

$$\widehat{F}_{SM} = \widehat{F}(y) \left[\psi_{16} \left(\frac{F(x)}{\widehat{F}(x)} \right) + \psi_{17} \left(\frac{\widehat{F}(x)}{F(x)} \right) \right] \exp \left(\frac{F(x) - \widehat{F}(x)}{F(x) + \widehat{F}(x)} \right), \tag{27}$$

where ψ_{16} and ψ_{17} are constants. The optimum values are:

$$\psi_{16(\text{opt})} = \frac{1}{8} \left[\frac{16\Theta_{110}^2 + 6\Theta_{110}\Theta_{020} - 24\Theta_{110}\Theta_{020} - 16\Theta_{200}\Theta_{020} - \Theta_{020}^2 - 16\Theta_{110} - 8\Theta_{020}}{16\Theta_{110}^2 - 16\Theta_{110}\Theta_{020} - 4\Theta_{200}\Theta_{020} + \Theta_{020}^2 - 4\Theta_{020}} \right],$$

$$\psi_{17(\text{opt})} = \frac{1}{8} \left[\frac{48\Theta_{110}^2 + 16\Theta_{110}\Theta_{020} - 72\Theta_{110}\Theta_{020} + 16\Theta_{200}\Theta_{020} + 21\Theta_{020}^2 + 16\Theta_{110} - 24\Theta_{020}}{16\Theta_{110}^2 - 16\Theta_{110}\Theta_{020} - 4\Theta_{200}\Theta_{020} + \Theta_{020}^2 - 4\Theta_{020}} \right]$$

The bias of \widehat{F}_{SM} is given by:

$$\text{Bias}(\widehat{F}_{SM}) = F(y) \left[(\psi_{16} + \psi_{17} - 1) - \left(\frac{3}{2}\psi_{16} - \frac{1}{2}\psi_{17} \right) \Theta_{110} + \left(\frac{15}{8}\Theta_{110} - \frac{1}{8}\psi_{17} \right) \Theta_{020} \right]$$

The minimum MSE of \widehat{F}_{SM} are given by:

$$\text{MSE}(\widehat{F}_{SM}) \cong \frac{F^2(y)}{16} \left[\frac{64\Theta_{110}^2\Theta_{200} - 48\Theta_{110}\Theta_{020} - 128\Theta_{110}\Theta_{200}\Theta_{020} + 48\Theta_{110}\Theta_{020}^2 + 64\Theta_{200}\Theta_{020}^2 + 64\Theta_{200}\Theta_{020}^3 + 9\Theta_{020}^3 + 64\Theta_{110}^2 - 64\Theta_{200}\Theta_{020}}{16\Theta_{110}^2 - 16\Theta_{110}\Theta_{020} - 4\Theta_{200}\Theta_{020} + \Theta_{020}^2 - 4\Theta_{020}} \right]. \tag{28}$$

Proposed estimator

The performance of the estimator can be is increased by utilizing the proper use of the auxiliary variable at estimation or designing stages or both stages. Ahmad et al.²⁴ proposed an improved estimator based on cumulative distribution function (DF) under stratified random sampling using the dual auxiliary information. Taking motivation from Ahmad et al.²⁴, we propose a novel class of estimators using DF of the study and auxiliary variables as well as mean of the usual auxiliary variable. The main benefit of our proposed generalised class of estimators is that it is more versatile and productive in terms of efficiency than the existing estimator. The proposed estimator is given by

$$\widehat{F}_{Prop} = [\psi_{18}\widehat{F}(y)\mathbf{T} + \psi_{19}(\bar{X} - \bar{x})] \exp \left(\frac{F(x) - \widehat{F}(x)}{F(x) + \widehat{F}(x)} \right), \tag{29}$$

where $\mathbf{T} = \frac{1}{4} \left(\frac{F(x)}{\widehat{F}(x)} + \frac{\widehat{F}(x)}{F(x)} \right) \left(\exp \left(\frac{F(x) - \widehat{F}(x)}{F(x) + \widehat{F}(x)} \right) + \exp \left(\frac{\widehat{F}(x) - F(x)}{\widehat{F}(x) + F(x)} \right) \right)^\Upsilon$ where ψ_{18} and ψ_{19} are the unknown constants, Υ is a suitably chosen constant.

After simplification of \widehat{F}_{Prop} , we have

$$\widehat{F}_{Prop} = \left[\psi_{18}F(y)(1 + \xi_0) \left\{ \frac{1}{4} \left\{ (1 + \xi_1)^{-1} + 1 + \xi_1 \right\} \left(\exp \left\{ -\frac{1}{2}\xi_1 \left(1 + \frac{1}{2}\xi_1 \right)^{-1} \right\} + \exp \left\{ \frac{1}{2}\xi_1 \left(1 + \frac{1}{2}\xi_1 \right)^{-1} \right\} \right)^\Upsilon - \psi_{19}\bar{X}\xi_1 \right\} \left\{ \exp \left(-\frac{1}{2}\xi_1 \left(1 + \frac{1}{2}\xi_1 \right)^{-1} \right) \right\} \right] \tag{30}$$

Expanding (30), we get

$$\widehat{F}_{Prop} = \left[\psi_{18}F(y)(1 + \xi_0) \left\{ \frac{1}{4} \left((2 + \xi_1^2) \left(2 + \frac{1}{4}\xi_1^2 \right) \right)^\Upsilon - \psi_{19}\bar{X}\xi_1 \right\} \exp \left(1 - \frac{1}{2}\xi_1 + \frac{3}{8}\xi_1^2 \right) \right]$$

$$\widehat{F}_{Prop} - F(y) = -F(y) + \psi_{18}F(y) \left[1 + \xi_0 - \frac{1}{2}R\xi_1 - \frac{1}{2}\xi_0\xi_1 + \Phi\xi_1^2 + \psi_{19}R\xi_1^2 \right], \tag{31}$$

where $\Phi = \frac{5}{8}\Upsilon + \frac{3}{8}$, $R = \frac{S_{F(y)}^2}{X^2}$

By taking expectations of (31) we get the bias of \widehat{F}_{Prop} :

$$\text{Bias}(\widehat{F}_{Prop}) = -F(y) + \psi_{18}F(y) \left[1 - \frac{1}{2}\Theta_{110} + \Phi\Theta_{020} + \psi_{19}R\Theta_{020} \right] \tag{32}$$

Taking expectations of (31), we obtain the required MSE:

$$\begin{aligned} \text{MSE}(\widehat{F}_{Prop}) &= \psi_{18}^2\psi_{19}^2F^2(y)F^2(x)\lambda C_x^2 - 2\psi_{18}F^2(y)\lambda C_x^2 + \frac{9}{4}\psi_{18}^2F^2(y)\lambda C_x^2 \\ &+ \psi_{18}^2F^2(y)\lambda C_{F(y)}^2 + \psi_{18}F^2(y)\lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} \\ &- 2\psi_{18}\psi_{19}\bar{X}\lambda C_{F(y)}C_x\rho_{F(y)x} \\ &- \psi_{18}^2F^2(y)\psi_{19}\bar{X}\lambda C_{F(y)}C_x\rho_{F(y)x} - F^2(y) - 2\psi_{18}F^2(y) + \psi_{18}^2F^2(y) \end{aligned} \tag{33}$$

The optimal values of ψ_{18} and ψ_{19} are obtained from (33), is given by:

$$\psi_{18} = \left[\frac{2\left\{ \lambda C_{F(y)}^2\rho_{F(y)x}^2 + \rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 2\lambda C_x^2 \right\}}{4\lambda C_{F(y)}^2\rho_{F(y)x}^2 + 8\lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 4\lambda C_{F(y)}^2 - 9C_x^2 - 4} \right],$$

and

$$\psi_{19} = -\frac{1}{4} \left[\frac{\left\{ 4\lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 4\lambda C_{F(y)}^2 - \lambda C_x^2 \right\} C_{F(y)}\rho_{F(y)x}}{\left(\lambda C_{F(y)}^2\rho_{F(y)x}^2 + \lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 2\lambda C_x^2 - 2 \right) \bar{X}C_x} \right]$$

Putting the optimal values of ψ_{18} and ψ_{19} in (33) we get the minimum MSE as given by:

$$\text{MSE}(\widehat{F}_{Prop})_{min} \cong F^2(y) \left[1 - \frac{\Theta_{020}}{4} - \frac{\left\{ 1 + \Phi\Theta_{020} - \frac{1}{2}\Theta_{020} \right\}^2}{1 + (2\Phi - \frac{3}{4})\Theta_{020} + \Theta_{200} - \frac{\Theta_{110}^2}{\Theta_{020}}} \right] \tag{34}$$

Some special cases of our proposed estimator. Case 1: When $\Upsilon = 1$ in (29), we have:

$$\begin{aligned} \widehat{F}_{Prop1} &= \left[\psi_{18}\widehat{F}(y) \left\{ \frac{1}{4} \left(\frac{F(x)}{\widehat{F}(x)} + \frac{\widehat{F}(x)}{F(x)} \right) \left(\exp \left(\frac{F(x) - \widehat{F}(x)}{F(x) + \widehat{F}(x)} \right) \right. \right. \right. \\ &\left. \left. \left. + \exp \left(\frac{\widehat{F}(x) - F(x)}{\widehat{F}(x) + F(x)} \right) \right) + \psi_{19}(\bar{X} - \bar{x}) \right\} \right] \exp \left(\frac{F(x) - \widehat{F}(x)}{F(x) + \widehat{F}(x)} \right) \end{aligned} \tag{35}$$

The optimum values of ψ_{18} and ψ_{19} , are given by:

$$\begin{aligned} \psi_{18} &= \left[\frac{2\left\{ \lambda C_{F(y)}^2\rho_{F(y)x}^2 + \lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 2\lambda C_x^2 - 2 \right\}}{4\lambda C_{F(y)}^2\rho_{F(y)x}^2 + 8\lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 4\lambda C_{F(y)}^2 - 9\lambda C_x^2 - 4} \right] \\ \psi_{19} &= -\frac{1}{4} \left[\frac{\left\{ 4\lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 4\lambda C_{F(y)}^2 - \lambda C_x^2 + 4 \right\} C_{F(y)}\rho_{F(y)x}}{\left(\lambda C_{F(y)}^2\rho_{F(y)x}^2 + \lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 2\lambda C_x^2 - 2 \right) \bar{X}C_x} \right] \end{aligned}$$

Putting values of ψ_{18} and ψ_{19} , we have:

$$\begin{aligned} \text{MSE}(\widehat{F}_{Prop1})_{min} &\cong F^2(y) \\ &\left[\frac{\lambda F^2(y) \left(\lambda C_{F(y)}^2\rho_{F(y)x}^2 \left(\left(C_{F(y)}^2\rho_{F(y)F(x)}^2 - \frac{7}{4}C_x^2\rho_{F(y)x}^2 \right) \lambda + \rho_{F(y)x}^2 - 4 \right) C_{F(y)}^2 - 4\rho_{F(y)F(x)}C_{F(y)}C_{F(x)}(\lambda C_x^2 - 1)C_{F(y)} + 4\lambda C_{F(x)}^2 - C_x^2 \right)}{4 \left(\left(4\rho_{F(y)x}^2 - 4 \right) \lambda C_{F(y)}^2 + 8\lambda\rho_{F(y)F(x)}C_{F(y)}C_{F(x)} - 9\lambda C_x^2 - 4 \right)} \right] \end{aligned} \tag{36}$$

Case 2: When $\Upsilon = 2$ in (29), we have:

$$\begin{aligned} \widehat{F}_{Prop2} &= \left[\psi_{18}\widehat{F}(y) \left\{ \frac{1}{4} \left(\frac{F(x)}{\widehat{F}(x)} + \frac{\widehat{F}(x)}{F(x)} \right) \left(\exp \left(\frac{F(x) - \widehat{F}(x)}{F(x) + \widehat{F}(x)} \right) \right. \right. \right. \\ &\left. \left. \left. + \exp \left(\frac{\widehat{F}(x) - F(x)}{\widehat{F}(x) + F(x)} \right) \right) \right\}^2 + \psi_{19}(\bar{X} - \bar{x}) \right] \exp \left(\frac{F(x) - \widehat{F}(x)}{F(x) + \widehat{F}(x)} \right) \end{aligned} \tag{37}$$

The values of ψ_{18} and ψ_{19} , are given by:

$$\psi_{18} = \frac{1}{4} \left[\frac{\{-4\lambda C_F^2(y) \rho_{F(y)x}^2 - 4\lambda \rho_{F(y)F(x)} C_F(y) C_F(x) + 13\lambda C_F^2(x) + 8\}}{-2\lambda C_F^2(y) \rho_{F(y)x}^2 - 4\lambda \rho_{F(y)F(x)} C_F(y) C_F(x) + 7\lambda C_F^2(x) + 2\lambda C_F^2(y) + 2\lambda C_F^2(x) + 2} \right]$$

$$\psi_{19} = - \left[\frac{\{-4\lambda \rho_{F(y)F(x)} C_F(y) C_F(x) + \lambda C_F^2(x) + 4\lambda C_F^2(y) - 4\} C_F(y) \rho_{F(y)x}}{(-4\lambda C_F^2(y) \rho_{F(y)x}^2 - 4\lambda \rho_{F(y)F(x)} C_F(y) C_F(x) + 13\lambda C_F^2(x) + 8) \bar{X} C_x} \right]$$

Putting values of ψ_{18} and ψ_{19} , we have:

$$MSE(\widehat{F}_{Prop2})_{min} \cong 169F^2(y) \left[\frac{\left\{ \frac{16}{169} - \lambda C_F^4(y) \rho_{F(y)x}^2 + (T_{11}) + T_{12} \right\} - T_{13} - \frac{16}{169} \rho_{F(x)}^2}{-\lambda C_F^2(y) (-64 \rho_{F(y)x}^2 + 64) - T_{14} + 22 \lambda C_F^2(x) + 64} \right] \quad (38)$$

where $T_{11} = \frac{16}{169} \lambda C_F^2(y) (\rho_{F(y)F(x)}^2 - 3 \rho_{F(y)}^2 \bar{X})$

$$T_{11} = \frac{16}{169} \rho_{F(y)}^2 \bar{X} - \frac{64}{169} \lambda C_F^2(y)$$

$$T_{13} = -\frac{8}{13} (\lambda C_F^2(x) - \frac{8}{13} \rho_{F(y)F(x)} C_F(y) C_F(x))$$

$$T_{14} = 128 \lambda \rho_{F(y)F(x)} C_F(y) C_F(x)$$

Numerical study

We carry out a numerical analysis utilizing three real data sets to found the efficiency of our proposed estimator. In terms of percentage relative efficiency, we assess how well our proposed estimator performs in comparison to its existing estimators. The following expression was used to calculate the percentage relative efficiency (PRE).

$$PRE = \frac{Var(\widehat{F}_{usual})}{MSE(\widehat{F}_i)} \times 100$$

where $(i = R, P, \dots, Prop1, Prop2)$.

Population-I: [Source: Punjab Bureau of Statistics (2021–2022)].

Y = Covid-19 test performed in Punjab district during 2021.

X = Covid-19 confirmed cases in Punjab districts during 2021.

N = 228, n = 40.

Population-II: [Source: Source: Punjab Bureau of statistics (2021–2022)].

Y = Total number of beds in 30th June 2021.

X = Total allocated beds for Covid.

N = 36, n = 8.

Population-III: [Source: Herbert (2009)].

<http://archibe.ics.uci.edu/ml/datasets/wine>

Y = Aspartame.

X = Leucine.

Simulation study

We have generated population of size 5000 from a bivariate normal distribution with different covariance matrices. The population means and covariance matrices, are given below:

Population-I.

$$\mu = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 3 & 3 \\ 3 & 6 \end{bmatrix}$$

$$\rho_{F(y)F(x)} = 0.5704609, \rho_{F(x)x} = 0.56466576 \text{ and } \rho_{F(y)y} = 0.8000432$$

Population-II.

$$\mu = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 3 & 3.5 \\ 3.5 & 8 \end{bmatrix}$$

$$\rho_{F(y)F(x)} = 0.5687875, \rho_{F(x)x} = 0.5015351 \text{ and } \rho_{F(y)y} = 0.7986237.$$

Population-III.

$$\mu = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\rho_{F(y)F(x)} = 0.7955929, \rho_{F(x)x} = 0.7623945 \text{ and } \rho_{F(y)y} = 0.8042652.$$

Discussion

We used three real populations and simulation to assess the efficiency of our proposed novel generalized class of estimators. We also considered different sample size from the populations. In our study we used a variety of data sets where we have a good mix of correlations between the study variable and the auxiliary variable. The data descriptions of real data sets are given in Table 1. The numerical results of MSE and PRE established on real data sets are presented in Tables 2 and 3. Additionally, it is emphasized that based on the numerical illustration the proposed estimators are more efficient than the existing estimators. It is observed that the proposed estimator is appreciable in terms of smallest MSE and greater PRE as compared to existing counterparts. The mean square error and PRE results based on simulated data sets are given in Tables 4 and 5. Thus it is recommended that the proposed estimators are useful in practice.

Moreover, we have also checked that the proposed generalized class of estimator is consistent when estimating the finite population DF; that is n increases; the values of proposed estimators get closer and closer to the true value. The researchers are thus recommended that the proposed estimators may be preferred over the existing estimators for the use of practical applications.

Parameters	Population-I	Population-II	Population-III
N	228	36	67
n	40	8	12
Λ	0.02061	0.09722	0.06840796
$F(y)$	0.50000	0.50000	0.4925373
$F(x)$	0.50000	0.50000	0.4925373
\bar{X}	882.93420	215.6389	20.59851
$\rho_{F(y)F(x)}$	0.5789474	0.3333333	0.6417112
$\rho_{F(y)x}$	0.2369336	0.3830227	0.6654633
$\rho_{F(x)x}$	0.2552666	0.4224452	0.7896493
$C_F(y)$	1.002200	1.0141850	1.022699
$C_F(x)$	1.002200	1.0141850	1.0226990
C_x	3.482949	1.5045280	0.6279163

Table 1. Summary statistics using real data sets.

Estimators	Population-I	Population-II	Population-III
\widehat{F}_{usual}	0.005176211	0.02500000	0.01735724
\widehat{F}_R	0.01634593	0.03333333	0.01243700
\widehat{F}_P	0.01634593	0.06666667	0.05699116
$\widehat{F}_{HS,1}$	0.00345985	0.023312331	0.01035522
$\widehat{F}_{HS,2}$	0.003442017	0.02226877	0.01024069
\widehat{F}_{Reg}	0.003441249	0.02222220	0.01020965
\widehat{F}_{BTR}	0.00347351	0.02291667	0.01055822
\widehat{F}_{BTP}	0.009467018	0.03958333	0.03283489
\widehat{F}_{SG}	0.003370345	0.01975446	0.009547592
\widehat{F}_{SW}	0.003441249	0.02222220	0.01020965
\widehat{F}_C	0.003441249	0.02222220	0.01020965
\widehat{F}_{YG}	0.003393549	0.02022472	0.009766925
\widehat{F}_{YS}	0.003354136	0.01936910	0.009353307
\widehat{F}_{SM}	0.003124916	0.01585976	0.006853708
$\widehat{F}_{Prop1(\Upsilon=1)}$	0.001843958	0.01282890	0.005831504
$\widehat{F}_{Prop2(\Upsilon=2)}$	0.0006506061	0.00157237	0.00558592

Table 2. Mean square error using real data sets.

Estimators	Population-I	Population-II	Population-III
\widehat{F}_{usual}	100	100	100
\widehat{F}_R	118.75	75.0000	139.5522
\widehat{F}_P	31.66667	37.5000	30.45603
$\widehat{F}_{HS,1}$	149.608	107.2395	167.6183
$\widehat{F}_{HS,2}$	150.3831	112.2649	169.4930
\widehat{F}_{Reg}	150.4167	112.5000	170.0083
\widehat{F}_{BTR}	149.0196	109.0909	164.3956
\widehat{F}_{BTP}	54.67626	63.15789	52.86219
\widehat{F}_{SG}	153.5811	126.5537	181.7971
\widehat{F}_{SW}	150.4167	112.5000	170.0083
\widehat{F}_C	150.4167	112.5000	170.0083
\widehat{F}_{YG}	152.5309	123.6111	177.7145
\widehat{F}_{YS}	154.3232	129.0715	185.5733
\widehat{F}_{SM}	165.6432	157.6317	253.2533
$\widehat{F}_{Prop1(\Upsilon=1)}$	280.712	194.8573	297.6461
$\widehat{F}_{Prop2(\Upsilon=2)}$	795.5984	1589.956	312.2597

Table 3. PRE using real data sets.

Estimators	Population-I	Population-II	Population-III
\widehat{F}_{usual}	0.006201073	0.006199319	0.0003666761
\widehat{F}_R	0.006429300	0.004080697	0.0004166983
\widehat{F}_P	0.018840000	0.020826200	0.0010823830
$\widehat{F}_{HS,1}$	0.004744011	0.003403407	0.0002944911
$\widehat{F}_{HS,2}$	0.004704825	0.003323658	0.0002943374
\widehat{F}_{Reg}	0.004704767	0.003397053	0.0002943372
\widehat{F}_{BTR}	0.004706791	0.003576476	0.0002959711
\widehat{F}_{BTP}	0.010912140	0.011949230	0.0006288133
\widehat{F}_{SG}	0.004577778	0.003351978	0.0002938594
\widehat{F}_{SW}	0.004704767	0.003397053	0.0002943372
\widehat{F}_C	0.004704767	0.003397053	0.0002943372
\widehat{F}_{YG}	0.00461467	0.003302443	0.0002939996
\widehat{F}_{YS}	0.004553616	0.002966457	0.0002937858
\widehat{F}_{SM}	0.004233409	0.003339358	0.0002926822
$\widehat{F}_{Prop1}(\Upsilon=1)$	0.003646903	0.002966456	0.0002492984
$\widehat{F}_{Prop2}(\Upsilon=2)$	0.000665874	0.001949947	3.170472e-06

Table 4. MSE using simulated data sets.

Estimators	Population-I	Population-II	Population-III
\widehat{F}_{usual}	100	100	100
\widehat{F}_R	96.4502	151.9181	87.99558
\widehat{F}_P	32.91439	29.76692	33.87675
$\widehat{F}_{HS,1}$	130.7137	181.7912	124.5118
$\widehat{F}_{HS,2}$	131.8024	182.1504	124.5768
\widehat{F}_{Reg}	131.8004	182.4911	124.5769
\widehat{F}_{BTR}	131.7473	173.3359	123.8892
\widehat{F}_{BTP}	56.82727	51.88051	58.3124
\widehat{F}_{SG}	135.4603	186.521	124.7794
\widehat{F}_{SW}	131.804	182.4911	124.5769
\widehat{F}_C	131.8024	182.4911	124.5769
\widehat{F}_{YG}	134.3774	184.9451	124.72
\widehat{F}_{YS}	136.1791	187.719	124.8107
\widehat{F}_{SM}	146.4794	185.644	125.2813
$\widehat{F}_{Prop1}(\Upsilon=1)$	170.0367	208.9806	147.0832
$\widehat{F}_{Prop2}(\Upsilon=2)$	902.6828	317.9225	11,565.35

Table 5. PRE using simulated data sets.

Conclusion

In this article, we suggested a new class of estimators for estimation of the finite population DF based on simple random sampling, which utilizes DF of the study and the auxiliary variables as well as also mean of the auxiliary variable. Numerical expressions of the bias and mean squared errors are derived up to first order of approximation. We used three real data sets, and have been determined from numerical results, that the proposed estimators perform well as compared to existing estimators. The results of the simulation study also confirmed the efficiency of the proposed estimators. We generated two new estimators from our suggested class of estimators. The gain in efficiency of \widehat{F}_{Prop2} is more as compared to \widehat{F}_{Prop1} . The current work can be extended to generate a better family of estimators for simple and stratified random sampling for estimating the DF based on measurement errors, two-stage and three-stage sampling designs.

Data availability

All data generated or analysed during this study are included in this published article.

Received: 2 December 2022; Accepted: 16 February 2023

Published online: 03 April 2023

References

- Grover, L. K. & Kaur, P. A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. *Commun. Stat.-Simul. Comput.* **43**(7), 1552–1574 (2014).
- Ahmad, S. & Shabbir, J. Use of extreme values to estimate the finite population mean under PPS sampling scheme. *J. Reliab. Stat. Stud.*, 99–112, 2018.
- Audu, A. *et al.* On The Efficiency of Almost Unbiased Mean Imputation When Population Mean of Auxiliary Variable is Unknown. *Asian J. Probab. Stat.* **15**(4), 235–250 (2021).
- Singh, H. P. & Nigam, P. Efficient method of estimating the finite population mean based on two auxiliary variables in the presence of non-response under stratified sampling. *J. Reliab. Stat. Stud.*, 223–242 (2021)
- Shahzad, U., Ahmad, I., Oral, E., Hanif, M. & Almanjahie, I. M. Estimation of the population mean by successive use of an auxiliary variable in median ranked set sampling. *Math. Popul. Stud.* **28**(3), 176–199 (2021).
- Singh, H. P., Yadav, A. & Pal, S. K. An exponential approach for estimating population mean using two auxiliary variables in stratified random sampling. *Revista Invest. Oper.* **42**(4), 456–468 (2021).
- Aggarwal, S., Soni, S. S. & Chaubey, R. S. Estimation of the population mean by developing a new estimator. *Bull. Pure Appl. Sci.-Math. Stat.* **41**(1), 11–15 (2022).
- Yadav, S. K., Sharma, D. K. & Mishra, S. S. New modified ratio type estimator of the population mean using the known median of the study variable. *Int. J. Oper. Res.* **41**(2), 151–167 (2021).
- Pal, S. K. & Singh, H. P. Estimation of finite population mean using auxiliary information in presence of non-response. *Commun. Stat.-Simul. Comput.* **47**(1), 143–165 (2018).
- Pal, S. K. & Singh, H. P. An efficient new approach for estimating the general parameter using auxiliary variable in sample surveys. *Afr. Mat.* **33**(3), 70 (2022).
- Shahzad, U., Hanif, M., Sajjad, I., & Anas, M. M. Quantile regression-ratio-type estimators for mean estimation under complete and partial auxiliary information. *Sci. Iran.* **29**(3), 1705–1715 (2022).
- Zaman, T., Bulut, H. & Yadav, S. K. Robust ratio-type estimators for finite population mean in simple random sampling: A simulation study. *Concurr. Comput. Pract. Exp.* **34**, e7273 (2022).
- Singh, G. N., Khalid, M. & Sharma, A. K. Some efficient classes of estimators of population mean in two-phase successive sampling under random non response. *Commun. Stat.-Theory Methods* **46**(24), 12194–12209 (2017).
- Singh, G. N. & Khalid, M. Some imputation methods to compensate with non-response for estimation of population mean in two-occasion successive sampling. *Commun. Stat.-Theory Methods* **49**(14), 3329–3351 (2020).
- Irfan, M., Javeda, M., & Bhatti, S. H. Difference-type-exponential estimators based on dual auxiliary information under simple random sampling. *Scientia Iranica. Transaction E, Industrial Engineering* **29**(1), 343–354 (2022).
- Riyaz, S., Jan, R., Maqbool, S., Jan, T. R. & Rather, K. U. I. Generalized exponential ratio estimator of population mean using two auxiliary variables in simple random sampling with an application to agricultural data. *J. Stat. Appl. Probab. Lett.* **9**(2), 89–99 (2022).
- Zaman, T., Dünder, E., Audu, A., Alilah, D. A., Shahzad, U. & Hanif, M. Robust regression-ratio-type estimators of the mean utilizing two auxiliary variables: A simulation study. *Math. Probl. Eng.*, 2021 (2021).
- Chambers, R. L. & Dunstan, R. Estimating distribution functions from survey data. *Biometrika* **73**(3), 597–604 (1986).
- Chambers, R. L., Dorfman, A. H. & Hall, P. Properties of estimators of the finite population distribution function. *Biometrika* **73**(3), 597–604 (1986).
- Dorfman, A. H. A comparison of design-based and model-based estimators of the finite population distribution function. *Aust. J. Stat.* **35**(1), 29–41 (1993).
- Ahmed, M. S. & Abu-Dayyeh, W. Estimation of finite-population distribution function using multivariate auxiliary information. *Stat. Trans.* **5**(3), 501–507 (2001).
- Singh, H. P., Singh, S. & Kozak, M. A family of estimators of finite-population distribution function using auxiliary information. *Acta Appl. Math.* **104**(2), 115–130 (2008).
- Onsongo, W. M., Otieno, R. O. & Orwa, G. O. (2018). Bias reduction technique for estimating finite population distribution function under simple random sampling without replacement.
- Ahmad, S., Aamir, M., Hussain, S., Shabbir, J., Zahid, E., Subkrajang, K. & Jirawattanapanit, A. A new generalized class of exponential factor-type estimators for population distribution function using two auxiliary variables. *Math. Probl. Eng.*, 2022 (2022).
- Cochran, W. The estimation of yields of cereal experiments by sampling for the ratio of grain to total produce. *J. Agricult. Sci.* **30**(2), 262–275 (1940).
- Murthy, M. N. Product method of estimation. *Sankhya Indian J. Stat. Ser. A* **26**(1), 69–74 (1964).
- Haq, A., & Shabbir, J. An improved estimator of finite population mean when using two auxiliary attributes. *Appl. Math. Comput.* **241**, 14–24 (2014).
- Bahl, S. & Tuteja, R. Ratio and product type exponential estimators. *J. Inf. Optim. Sci.* **12**(1), 159–164 (1991).
- Shabbir, J., & Gupta, S. A. T. On improvement in variance estimation using auxiliary information. *Commun. Stat. Theory Methods* **36**(12), 2177–2185 (2007).
- Swain, A. K. P. C. Generalized estimator of finite population variance. *J. Stat. Theory Appl.* **14**(1), 45–51 (2015).
- Chami, P. S., Sing, B. & Thomas, D.: A two-parameter ratio-product-ratio estimator using auxiliary information. *Int. Sch. Res. Notices*, 2012 (2012).
- Yadav, S. K., Kadilar, C., Shabbir, J. & Gupta, S. Improved family of estimators of population variance in simple random sampling. *J. Stat. Theory Pract.* **9**(2), 219–226 (2015).
- Yaqub, M., & Shabbir, J. An improved class of estimators for finite population variance. *Hacet. J. Math. Stat.* **45**(5), 1641–1660 (2016).
- Muneer, S., Khalil, A., Shabbir, J. & Narjis, G. A new improved ratio-product type exponential estimator of finite population variance using auxiliary information. *J. Stat. Comput. Simul.* **88**(16), 3179–3192 (2018).

Acknowledgements

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R 299), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions

S.A., K.U. and M.A. wrote the original overall manuscript, J.S., H.M.A., A.al-A.H.El-B. and M.A. give conceptualization. E.Z. and J.S. help us in data curation. Project administration done by S.A. and M.A. Formal analysis done by S.A. Validation by K.U. Visualization, Revision and Final approval by E.Z., H.M.A and A.al-A.H.El-B.

Funding

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R 299), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.M.A.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023