Heliyon 10 (2024) e28272

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

An improved class of estimators for estimation of population distribution functions under stratified random sampling

Sohaib Ahmad, PhD [a,*], Javid Shabbir [b], Walid Emam [c], Erum Zahid [d], Muhammad Aamir [a], Mohd Khalid [e], Malik Muhammad Anas [f]

[a] Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan
[b] Department of Statistics, University of Wah, Wah Cantt, Pakistan
[c] Department of Statistics and Operations Research, Faculty of Science, King Saud University, P.O. Box 2455, Riyadh, 11451, Saudi Arabia
[d] Department of Applied Mathematics and Statistics, Institute of Space Technology, Islamabad, Pakistan
[e] Department of Statistics, Aligarh Muslim University, Aligarh, India
[f] Department of Economics and Statistics, University of Salerno, Fisciano (Salerno) 84084, Italy

ARTICLE INFO

Keywords:
Simulation study
Stratified random sampling
Population DF
Auxiliary variables
Bias
MSE
PRE

ABSTRACT

The main objective of the current study is to suggest an enhanced family of log ratio-exponential type estimators for population distribution function (DF) using auxiliary information under stratified random sampling. Putting different choices in our suggested generalized class of estimators, we found some Specific estimators. The bias and MSE expressions of the estimators have been approximated up to the first order. By using the actual and simulated data sets, we measured the performance of estimators. Based on the results, the suggested estimators for DF show better performance as compared to the preliminary estimators considered here. The suggested estimators have a advanced efficiency than the other estimators examined with the estimators $\hat{\hat{F}}^{2}_{logPR(st)}$, and $\hat{\hat{F}}^{4}_{logPR(st)}$ for both the actual and simulated data sets. The magnitude of the improvement in efficiency is noteworthy, indicating the superiority of the proposed estimators in terms of MSE.

1. Introduction

In general, it is broadly recognized that the estimator precision increases when the auxiliary information is employed properly in survey sampling. In the literature, there are numerous estimators for assessing numerous population parameters, comprising mean, variance and total. A widespread variety of techniques for comprising the supporting information using ratio, product, and regression type methods of estimate are covered in the literature on survey sampling. Several estimators have been recommended by numerous researchers by suitably modifying the auxiliary variable. Some notable work by these authors under stratified random sampling includes [17] discussed improved estimator under simple random sampling and stratified random sampling [18]. recommended a general class of estimators under stratified random sampling [19,20]. discussed estimation of finite population mean under Stratified Random Sampling. The authors in Refs. [21–24] suggested improved estimators under stratified sampling schemes. The authors in

* Corresponding author.
E-mail addresses: sohaib_ahmad@awkum.edu.pk (S. Ahmad), javidshabbir@gmail.com (J. Shabbir), wemam.c@ksu.edu.pk (W. Emam), erumzahid22@gmail.com (E. Zahid), aamirkhan@awkum.edu.pk (M. Aamir), mohdkhalid4870@gmail.com (M. Khalid), anas.uaar@gmail.com (M. Muhammad Anas).

https://doi.org/10.1016/j.heliyon.2024.e28272
Received 5 November 2023; Received in revised form 6 March 2024; Accepted 14 March 2024
Available online 22 March 2024

Refs. [25–28] recommended some novel estimators for estimation of population parameters under stratified random sampling.

Stratification is used to increase the efficiency of estimators. Stratified planning divides the entire population into smaller, more manageable subsets, or "strata." This classification was made based on relevant characteristics or factors. The basic objective of stratification is to achieve internal uniformity. When a population is divided into subgroups with comparable values, there is less room for variation within each stratum. The estimations are more reliable because of the independent sampling within each strata. Simple random sampling is typically used for selecting samples within each stratum. As a result, each individual or unit in a given stratum can be regarded of as a microcosm of that stratum's total population. In order to more effectively capture the variance present within the subgroups of a population, researchers conducting surveys can increase the precision of their estimates by applying stratification and selecting samples using simple random sampling within each stratum. Sometimes we want to examine the proportion or percentage of a population that exhibits a specific attribute or falls within a particular range, we essential to evaluate the population DF. When applied to a finite population, the finite population DF can be used to learn the likelihood of each value. This function is what connects the population values with the associated probabilities. The DF is implemented in the computation of total, coefficient of variation, and standard deviation in populations. Statistical inference techniques are commonly used in survey sampling to extrapolate results from a subset of the population to the entire population, yielding an estimate of the DF for the entire population. Specific estimators that are good choices for the proposed enhanced class of estimators have been found. Factors comprising sample size, sampling technique, and population fluctuation all affect how well the finite population DF works. Greater precision can be achieved by using a larger sample size and a random sampling technique. In conclusion, when we need to estimate population characteristics like the proportion or percentage of the population possessing a given trait, estimation of finite population DF is an essential part of survey sampling. Some notable work in the DF field comprises: The authors in Ref. [1] recommended estimation population distribution function using auxiliary information. The authors in Refs. [2–7] discussed improved estimators for population DF under different sampling schemes. The [8–10] suggested efficient estimators for population DF using ranked set sampling [11–14]. recommend some new estimators for population DF utilizing auxiliary information [29–31]. also discussed improved class of estimators for population DF under stratified random sampling. The authors in Refs. [32–36] discuss enhanced estimator using auxiliary variable and attribute under stratified random sampling.

The organization of the article is given by:

Introduction of the article is discussed in Section 1. Notations and methods are discussed in Section 2. In Section 3, we discussed some well-known existing estimators. The proposed class f estimators are given in Section 4. Data description using actual data sets are given in Section 5. The simulation analysis is assumed in Section 6. Discussion of the article is given in Section 7. Finally the conclusion of the article is given in section 8.

2. Methods and materials

Consider $\Omega = (1,2,\ldots,\Omega_N)$ consist of N distinctive elements, which is alienated into L similar strata, where the size of h^{th} stratum is N_h for $h = 1,2,\ldots,L$, such that $\sum_{h=}^{L} N$. Let Y and X be the values of the study variable and the auxiliary variable which takes values y_i and x_i, where $i = 1,2,\ldots,\Omega_N$. A sample of size $\sum_{h=1}^{L} n_h = n$, is chosen from the h^{th} stratum using simple random sampling without replacement (SRSWOR).

Let $[F_{st}(y) = F_y = \sum_{h=1}^{L} W_h F_{h(y)}$ and $F_{st}(x) = F_x = \sum_{h=1}^{L} W_h F_{h(x)}, \widehat{F}_{st}(y) = \widehat{F}_y = \sum_{h=1}^{L} W_h \widehat{F}_h(y)$ and $\widehat{F}_{st}(x) = \widehat{F}_x = \sum_{h=1}^{L} W_h \widehat{F}_h(x)\,]$ be the DF of the study variable Y and the auxiliary variabe X, where $W_h = N_h/N$,

$$\left[F_{h(y)} = \sum_{i=1}^{N_h} I(Y_{ih} \le Y) \Big/ N_h, \widehat{F}_{h(y)} = \sum_{i=1}^{n_h} I(Y_{ih} \le Y) \Big/ n_h, F_{h(x)} = \sum_{i=1}^{N_h} I(X_{ih} \le X) \Big/ N_h, \widehat{F}_{h(x)} = \sum_{i=1}^{n_h} I(X_{ih} \le X) \Big/ n_h\right]$$

Let, $e_0 = \frac{\widehat{F}_y - F_y}{F_y}$, $e_1 = \frac{\widehat{F}_x - F_x}{F_x}$ a, such that

$E(e_i) = 0$ for $i = 1,2$

Let $\varphi_{rs} = E[e_1^r\ e_2^s\,]$, where $r,s = 1,2$. Here,

$$E(e_0^2) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 C_{F_{yh}}^2 = \varphi_{20}, E(e_1^2) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 C_{F_{xh}}^2 = \varphi_{02},$$

$$E(e_0 e_1) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 \rho_{(F_y F_x)_h} C_{F_{yh}} C_{F_{xh}} = \varphi_{11},$$

$$\text{Let } \rho_{F_{yh}}^2 = \sum_{i=1}^{N_h} \left(I(Y_i \le Y) - F_y\right)^2 \Big/ (N-1),$$

$$\rho_{F_{xh}}^2 = \sum_{i=1}^{N_h} \left(I(X_i \le X) - F_x\right)^2 \Big/ (N-1),$$

$$\rho_{(F_yF_x)_h} = \sigma_{(F_yF_x)_h} \Big/ \left(\sigma_{F_{yh}}\sigma_{F_{xh}}\right), C_{F_y} = \frac{\rho_{F_{yh}}}{F_y}, C_{F_x} = \frac{\rho_{F_{xh}}}{F_x},$$

where $\lambda_h = (1/n_h - 1/N_h)$.

3. Review of existing estimators

In this section, we have discussed some exisiting estimators for population DF under stratified random sampling:

1. The conventional estimator of $\widehat{\widehat{F}}_u$, is given in equation (1):

$$\widehat{\widehat{F}}_1 = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le \text{Y}). \quad (1)$$

The variance of $\widehat{\widehat{F}}_1$, is given in equation (2):

$$\text{Var}(\widehat{\widehat{F}}_1) = F_y^2\varphi_{20}. \quad (2)$$

2. The usual ratio estimator is given in equation (3):

$$\widehat{\widehat{F}}_2 = \widehat{F}_{st}(y)\left(\frac{F_x}{\widehat{F}_{st}(x)}\right) \quad (3)$$

The bias of $\widehat{\widehat{F}}_2$, is given by:

$$\text{Bias}(\widehat{\widehat{F}}_2) \cong F_y(\varphi_{02} - \varphi_{11}),$$

The MSE of $\widehat{\widehat{F}}_2$, is given in equation (4):

$$\text{MSE}(\widehat{\widehat{F}}_2) \cong F_y^2(\varphi_{20} + \varphi_{02} - 2\varphi_{11}). \quad (4)$$

5. The regression type estimator of $\widehat{\widehat{F}}_3$ is given in equation (5):

$$\widehat{\widehat{F}}_3 = [\widehat{F}_{st}(y) + w(F_x - \widehat{F}_{st}(x))], \quad (5)$$

where w is constant.

$$w_{(\text{opt})} = \left(F_y\varphi_{11}\right) / \left(F_x\varphi_{02}\right)$$

The variance of $\widehat{\widehat{F}}_3$ at the optimum value is given in equation (6):

$$\text{Var}_{\min}(\widehat{\widehat{F}}_3) = \frac{F_y^2\left(\varphi_{20}\varphi_{02} - \varphi_{11}^2\right)}{\varphi_{02}} \quad (6)$$

Equation (6) is also inscribed, a simplified form given in equation (7):

$$\text{Var}_{\min}(\widehat{\widehat{F}}_3) = F_y^2\varphi_{20}\left(1 - \rho^2_{(F_yF_x)_h}\right). \quad (7)$$

4. The exponential ratio and produc ttype estimators are given in equation (8) and in equation (9):

$$\widehat{\widehat{F}}_4 = \widehat{F}_{st}(y)\exp\left(\frac{F_x - \widehat{F}_{st}(x)}{\widehat{F}_{st}(x) + F_x}\right), \quad (8)$$

$$\widehat{\widehat{F}}_5 = \widehat{F}_{st}(y)\exp\left(\frac{\widehat{F}_{st}(x) - F_x}{\widehat{F}_{st}(x) + F_x}\right). \quad (9)$$

The bias of $\widehat{\widehat{F}}_4$, is given by:

$$\text{Bias}(\widehat{\widehat{F}}_4) \cong F_y\left(\frac{3}{8}\varphi_{02} - \frac{1}{2}\varphi_{11}\right),$$

The MSE of $\widehat{\widehat{F}}_4$, is given in equation (10):

$$\text{MSE}(\widehat{\widehat{F}}_4) \cong \frac{F_y^2}{4}(4\varphi_{20} + \varphi_{02} - 4\varphi_{11}) \quad (10)$$

The bias of $\widehat{\widehat{F}}_5$, is given by:

$$\text{Bias}(\widehat{\widehat{F}}_5) \cong F_y\left(\frac{1}{2}\varphi_{11} - \frac{1}{8}\varphi_{02}\right),$$

The MSE of $\widehat{\widehat{F}}_5$, is given in equation (11):

$$\text{MSE}(\widehat{\widehat{F}}_5) \cong \frac{F_y^2}{4}(4\varphi_{20} + \varphi_{02} + 4\varphi_{11}), \tag{11}$$

4. Proposed estimator

The main purpose of survey researchers is to increase the efficiency of an estimator. The utilization of auxiliary information can boost estimator efficiency either at the design or at the estimation stage. In this article, we develop the enhanced estimator using the DF of *Y* and *X*. The developed estimator is the combination of regression cum log exponential type estimator. The logarithmic function is indeed a fundamental mathematical concept that appears in many areas of science and non-science including mathematics, economics, and engineering among others. We suggested a new class of estimators for population DF under stratified sampling design, which is given in equation (12):

$$\widehat{\widehat{F}}^G_{logPR(st)} = (K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x)))\left[1 + log\left(\frac{F_x}{\widehat{F}_{st}(x)}\right)\right]^{\alpha_{1(st)}}\left[\exp\left(\frac{F_x - \widehat{F}_{st}(x)}{F_x + \widehat{F}_{st}(x)}\right)\right]^{\alpha_{2(st)}}, \tag{12}$$

where K_3 and K_4 are constants.

$$\widehat{\widehat{F}}^G_{logPR(st)} = \left[K_3F_y(1+e_0) - K_4F_x\, e_1\right]\left[1 - \alpha_{1(st)}e_1 - \alpha_2\frac{e_1}{2}\right]\left[1 - \alpha_{2(st)}\frac{e_1}{2} + \frac{\alpha_{2(st)}\left(\alpha_{2(st)}+2\right)}{8}e_1^2\right]$$

Simplify the above equation we got equation (13):

$$\widehat{\widehat{F}}^G_{logPR(st)} - \widehat{\widehat{F}}_y = (K_3-1)F_y + K_3F_y\left[e_0 - \delta_{1(st)}e_1 - \delta_{1(st)}e_0e_1 + \delta_{2(st)}e_1^2\right] - K_4F_x\left[e_1 - \delta_{1(st)}e_1^2\right] \tag{13}$$

where $\delta_{1(st)} = \alpha_{1(st)} + \frac{\alpha_{2(st)}}{2}$, and

$$\delta_{2(st)} = \frac{1}{2}\left[\alpha_{1(st)}\alpha_{2(st)} + \alpha^2_{1(st)} + \alpha_{2(st)}\left(\alpha_{2(st)}+2\right)\right].$$

$$\text{Bias}\left(\widehat{\widehat{F}}^G_{logPR(st)}\right) = F_y\left[(K_3-1) + K_3\lambda\left\{\delta_{2(st)}C^2_{F_x} - \delta_{1(st)}\rho_{F_yF_x}C_{F_y}C_{F_x}\right\} + K_4R\delta_{1(st)}C^2_{F_x}\right],$$

where $R = \frac{F_y}{F_x}$.

Squaring equation (13) and then taking expectations, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^G_{logPR(st)}\right) = F_y^2\ \text{E}\ \left[(K_3-1) + K_3\left\{e_0 - \delta_{1(st)}e_1 - \delta_{1(st)}e_0e_1 + \delta_{2(st)}e_1^2\right\} + K_4\left\{e_1 - \delta_{1(st)}e_1^2\right\}\right]^2$$

Applying the values we got equation (14):

$$\text{MSE}\left(\widehat{\widehat{F}}^G_{logPR(st)}\right) = F_y^2\left[1 + K_3^2A + K_4^2B - 2K_3C - 2K_4D + 2K_3K_4E\right], \tag{14}$$

where

$$A = 1 + \lambda\left\{C^2_{F_y} + \left(\delta^2_{1(st)} + 2\delta_{2(st)}\right)C^2_{F_x} - 4\delta_{1(st)}\rho_{F_yF_x}C_{F_y}C_{F_x}\right\},$$

$$B = R^2\lambda C^2_{F_x},$$

$$C = 1 + \lambda\left\{\delta_{2(st)}C^2_{F_x} - \delta_{1(st)}\rho_{F_yF_x}C_{F_y}C_{F_x}\right\},$$

$$D = R\delta_{1(st)}\lambda C^2_{F_x},$$

$$E = R\ \lambda\left(2\delta_{1(st)}C^2_{F_x} - \rho_{F_yF_x}C_{F_y}C_{F_x}\right)$$

The ideal values of K_3 and K_4 are

$$K_{3(opt)} = \frac{BC - DE}{AB - E^2}, K_{3(opt)} = \frac{AD - CE}{AB - E^2},$$

Putting the values of $K_{3(opt)}$ and $K_{4(opt)}$ in equation (13), we got equation (15):

$$\text{MSE}\left(\widehat{\widehat{F}}^{G}_{logPR(st)}\right)_{min} = F_y^2\left[1 - \frac{AD^2 + BC^2 - 2CDE}{AB - E^2}\right] \tag{15}$$

Some Special Cases.

Case 1. When $\alpha_{1(st)} = 1$, $\alpha_{2(st)} = 1$, $\delta_{1(st)} = 1$, $\delta_{2(st)} = \frac{1}{2}$, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^{1}_{logPR(st)}\right) = [K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x))]\left[1 + log\left(\frac{F_x}{\widehat{F}_{st}(y)}\right)\right]$$

Case 2. When $\alpha_{1(st)} = 0$, $\alpha_{2(st)} = 1$, $\delta_{1(st)} = \frac{1}{2}$, $\delta_{2(st)} = \frac{3}{2}$, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^{2}_{logPR(st)}\right) = [K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x))]\left[\exp\left(\frac{F_x - \widehat{F}_{st}(x)}{F_x + \widehat{F}_{st}(x)}\right)\right]$$

Case 3. When $\alpha_{1(st)} = 1$, $\alpha_{2(st)} = 1$, $\delta_{1(st)} = \frac{3}{2}$, $\delta_{2(st)} = \frac{5}{2}$, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^{3}_{logPR(st)}\right) = [K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x))]\left[1 + log\left(\frac{F_x}{\widehat{F}_{st}(y)}\right)\right]\left[\exp\left(\frac{F_x - \widehat{F}_{st}(x)}{F_x + \widehat{F}_{st}(x)}\right)\right]$$

Case 4. When $\alpha_{1(st)} = -1$, $\alpha_{2(st)} = 1$, $\delta_{1(st)} = \frac{-1}{2}$, $\delta_{2(st)} = \frac{3}{2}$, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^{4}_{logPR(st)}\right) = [K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x))]\left[1 + log\left(\frac{F_x}{\widehat{F}_{st}(y)}\right)\right]^{-1}\left[\exp\left(\frac{F_x - \widehat{F}_{st}(x)}{F_x + \widehat{F}_{st}(x)}\right)\right]$$

Case 5. When $\alpha_{1(st)} = 1$, $\alpha_{2(st)} = -1$, $\delta_{1(st)} = \frac{1}{2}$, $\delta_{2(st)} = \frac{-1}{2}$, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^{5}_{logPR(st)}\right) = [K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x))]\left[1 + log\left(\frac{F_x}{\widehat{F}_{st}(y)}\right)\right]\left[\exp\left(\frac{\widehat{F}_{st}(x) - F_x}{\widehat{F}_{st}(x) + F_x}\right)\right]$$

Case 6. When $\alpha_{1(st)} = -1$, $\alpha_{2(st)} = 0$, $\delta_{1(st)} = -1$, $\delta_{2(st)} = \frac{1}{2}$, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^{6}_{logPR(st)}\right) = [K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x))]\left[1 + log\left(\frac{F_x}{\widehat{F}_{st}(y)}\right)\right]^{-1}$$

Case 7. When $\alpha_{1(st)} = 0$, $\alpha_{2(st)} = -1$, $\delta_{1(st)} = \frac{1}{2}$, $\delta_{2(st)} = \frac{-1}{2}$, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^{7}_{logPR(st)}\right) = [K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x))]\left[\exp\left(\frac{\widehat{F}_{st}(x) - F_x}{\widehat{F}_{st}(x) + F_x}\right)\right]$$

Case 8. When $\alpha_{1(st)} = -1$, $\alpha_{2(st)} = -1$, $\delta_{1(st)} = \frac{1}{2}$, $\delta_{2(st)} = \frac{-1}{2}$, we get

$$\text{MSE}\left(\widehat{\widehat{F}}^{8}_{logPR(st)}\right) = [K_3\widehat{F}_{st}(y) + K_4(F_x - \widehat{F}_{st}(x))]\left[1 + log\left(\frac{F_x}{\widehat{F}_{st}(y)}\right)\right]^{-2}\left[\exp\left(\frac{\widehat{F}_{st}(x) - F_x}{\widehat{F}_{st}(x) + F_x}\right)\right]$$

5. Data description and numerical results

Three actual data are used to empirically measure the suggested generalized class of estimators which is compared to the pre-

Table 1
Description using Population-I.

N_h	n_h	W_h	λ_h
127	31	0.1375	0.0244
117	21	0.1267	0.0390
103	29	0.1115	0.0248
170	38	0.1841	0.0204
205	22	0.2221	0.0406
201	39	0.2177	0.0207
F_{yh}	S_{yh}	F_{xh}	S_{xh}
0.3543	0.4802	0.3779	0.4868
0.4188	0.4955	0.4872	0.5019
0.4272	0.4970	0.4660	0.5013
0.5765	0.4956	0.6118	0.4888
0.6146	0.4879	0.6537	0.4769
0.5025	0.5012	0.3532	0.4792

liminary estimators. To check the performances of estimators in terms of PRE we use the following relation, which is given by:

$$PRE(.) = \frac{Var(\widehat{\overline{F}}_1)}{MSE(\widehat{\overline{F}}_{i(st)(min)})} \times 100,$$

where $i = (\widehat{\overline{F}}_1, \widehat{\overline{F}}_2, \widehat{\overline{F}}_3, \widehat{\overline{F}}_4, \widehat{\overline{F}}_5\ \widehat{\overline{F}}^1_{logPR(st)}, \widehat{\overline{F}}^2_{logPR(st)}, \widehat{\overline{F}}^3_{logPR(st)}, \widehat{\overline{F}}^4_{logPR(st)}, \widehat{\overline{F}}^5_{logPR(st)}, \widehat{\overline{F}}^6_{logPR(st)}, \widehat{\overline{F}}^7_{logPR(st)}, \widehat{\overline{F}}^8_{logPR(st)})$.

Population-I [Source: [15]]:
Y: The number of teachers.
X: the number of pupils attending primary and intermediate schools.
Population-II [Source: [15]]:
Y: The number of teachers,
X: the number of students enrolled in basic and secondary education in 923 districts throughout six regions in 2007.
Population-III [Source: [16]]:
Y: The yield of apples in 1999,
X: the number of apples timber in 1999.

6. Simulation study

Three populations, each with 1000 observations, were generated from a multivariate normal distribution. Population-I, exhibits negative correlation. Population-II, exhibits positive association. Population-III, is highly positive correlation.

Population-I:

$$\mu =$$

and

$$\sum = \begin{bmatrix} 4 & -9.0 \\ -9.0 & 64 \end{bmatrix}$$

$N_1 = 500$ and $N_2 = 500, \rho_{XY} = -0.590220$
Population-II:

$$\mu =$$

and

$$\sum = \begin{bmatrix} 4 & 9.5 \\ 9.5 & 63 \end{bmatrix}$$

$N_1 = 500$ and $N_2 = 500$, $\rho_{XY} = 0.612254$.
Population-III:

$$\mu = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

and

Table 2
Description using population-II.

N_h	n_h	W_h	λ_h
127	31	0.1375	0.0244
117	21	0.1267	0.0391
103	29	0.1115	0.0248
170	38	0.1841	0.0204
205	22	0.2221	0.0406
201	39	0.2177	0.0207
F_{yh}	S_{yh}	F_{xh}	S_{xh}
0.3543	0.4802	0.3700	0.4847
0.4188	0.4955	0.4700	0.5013
0.4272	0.4970	0.4272	0.4970
0.5765	0.4956	0.5882	0.4936
0.6146	0.4879	0.6146	0.4879
0.5025	0.5012	0.4527	0.4990

Table 3
Description using population-III.

N_h	n_h	W_h	λ_h	$\overline{Y}_h$
106	9	0.1241	0.1017	1536.774
106	17	0.1241	0.0494	2212.594
94	38	0.1100	0.0157	9384.309
171	67	0.2002	0.0090	5588.012
204	7	0.2389	0.1379	966.955
173	2	0.2026	0.4942	404.398
S^2_{yh}	C^2_{yh}	$\overline{X}_h$	S^2_{xh}	C^2_{xh}
41281746	17.479	24711.81	2414224935	3.95
133437791	27.256	26840.04	2913701588	4.04
894457433	10.156	72723.76	25956279019	4.90
820445636	26.274	73191.2	68903936687	12.86
5710999	6.107	26833.75	2040714047	2.83
894440.3	5.469	9903.30	360137210	3.67

Table 4
MSE using actual Populations.

Estimators	Populations		
	I	II	III
$\widehat{F}_1$	0.0009240016	0.0009240016	0.00519513
$\widehat{F}_2$	0.0001780699	0.000221005	0.002569816
$\widehat{F}_3$	0.0001690221	0.0002206926	0.002565836
$\widehat{F}_4$	0.0003187042	0.0003891876	0.003275305
$\widehat{F}_5$	0.001993962	0.0018255447	0.008329292
$\widehat{F}^1_{logPR(st)}$	0.0001673105	0.0002202792	0.002364296
$\widehat{F}^2_{logPR(st)}$	0.0001368365	0.0001970156	0.0009471063
$\widehat{F}^3_{logPR(st)}$	0.0001658819	0.0002175351	0.001524696
$\widehat{F}^4_{logPR(st)}$	0.0001368365	0.0001970156	0.009471063
$\widehat{F}^5_{logPR(st)}$	0.0001635881	0.0002185955	0.002508543
$\widehat{F}^6_{logPR(st)}$	0.0001673105	0.0002202792	0.002364296
$\widehat{F}^7_{logPR(st)}$	0.0001635881	0.0002185955	0.002508543
$\widehat{F}^8_{logPR(st)}$	0.0001635881	0.0002185955	0.002508543

Table 5
PRE using actual Populations.

Estimators	Populations		
	I	II	III
$\widehat{F}_1$	100	100	100
$\widehat{F}_2$	518.8982	418.0909	202.1596
$\widehat{F}_3$	546.675	418.6825	202.4732
$\widehat{F}_4$	289.9245	237.418	158.6152
$\widehat{F}_5$	46.33998	50.61783	62.37181
$\widehat{F}^1_{logPR(st)}$	552.2675	419.4684	219.7327
$\widehat{F}^2_{logPR(st)}$	675.2595	468.9992	548.52266
$\widehat{F}^3_{logPR(st)}$	557.0236	424.7597	340.7321
$\widehat{F}^4_{logPR(st)}$	675.2595	468.9992	548.5266
$\widehat{F}^5_{logPR(st)}$	564.8343	422.6993	207.0975
$\widehat{F}^6_{logPR(st)}$	552.2675	419.4684	219.7327
$\widehat{F}^7_{logPR(st)}$	564.8343	422.6993	207.0975
$\widehat{F}^8_{logPR(st)}$	564.8343	422.6993	207.0975

Table 6
Mean squared error using simulation analysis.

Estimators	Populations		
	I	II	III
$\widehat{F}_1$	0.001001643	0.001000641	0.001001323
$\widehat{F}_2$	0.0028426528	0.0011419570	0.0005401607
$\widehat{F}_3$	0.0008131456	0.0008314480	0.0004650720
$\widehat{F}_4$	0.0016769029	0.0008367699	0.0005175744
$\widehat{F}_5$	0.0008168738	0.0016335711	0.0019914054
$\widehat{F}^1_{logPR(st)}$	0.0008127221	0.0008308942	0.0004650653
$\widehat{F}^2_{logPR(st)}$	0.0007965114	0.0008147885	0.0004534033
$\widehat{F}^3_{logPR(st)}$	0.0008085939	0.0008266602	0.0004630307
$\widehat{F}^4_{logPR(st)}$	0.0007965114	0.0008147885	0.0004534033
$\widehat{F}^5_{logPR(st)}$	0.0008131212	0.0008313792	0.0004647290
$\widehat{F}^6_{logPR(st)}$	0.0008127221	0.0008308942	0.0004650653
$\widehat{F}^7_{logPR(st)}$	0.0008131212	0.0008313792	0.0004647290
$\widehat{F}^8_{logPR(st)}$	0.0008131212	0.0008313792	0.0004647290

$$\sum = \begin{bmatrix} 2 & 4 \\ 6 & 10 \end{bmatrix}$$

$N_1 = 500$ and $N_2 = 500$, $\rho_{XY} = 0.902645$.

7. Discussion

We evaluated the effectiveness of our suggested novel improved class of estimators using simulation and three real populations. We also took into account various population sample sizes. Tables 1–3 contain descriptions of actual data sets. Table 4 and Table 5 provide the numerical outcomes of MSE and PRE developed on actual data sets. It is also emphasized that the suggested estimators are more effective than the existing estimators based on the numerical illustration. It has been noted that the proposed estimator outperforms its previous counterparts in terms of less MSE and higher PRE. Table 6 and Table 7 provide the MSE and PRE results based on simulation. We visualize actual population 1–3 for the result of mean square error and PRE in Figs. 1 and 2. Similarly, the simulated population is also presented visually on Figs. 3 and 4. As the sample size increases, the suggested estimators are closely to the true value. This property is known as consistency, which means that as the sample size grows, the estimated values become more accurate and coverage

Table 7
Percentage relative efficiency using simulation study.

Estimators	Populations		
	I	II	III
$\widehat{F}_1$	100	100	100
$\widehat{F}_2$	35.23622	87.62513	185.37494
$\widehat{F}_3$	123.1813	120.3492	215.3049
$\widehat{F}_4$	59.73174	119.58380	193.46447
$\widehat{F}_5$	122.61910	61.25483	50.28221
$\widehat{F}^1_{logPR(st)}$	123.2455	120.4295	215.3080
$\widehat{F}^2_{logPR(st)}$	125.7538	122.8099	220.8459
$\widehat{F}^3_{logPR(st)}$	123.8747	121.0463	216.2541
$\widehat{F}^4_{logPR(st)}$	125.7538	122.8099	220.8459
$\widehat{F}^5_{logPR(st)}$	123.1850	120.3592	215.4638
$\widehat{F}^6_{logPR(st)}$	123.2455	120.4295	215.3080
$\widehat{F}^7_{logPR(st)}$	123.1850	120.3592	215.4638
$\widehat{F}^8_{logPR(st)}$	123.1850	120.3592	215.4638

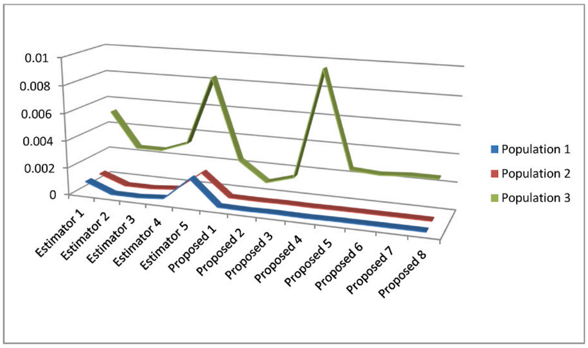

Fig. 1. Show the mean square error using population 1-3.

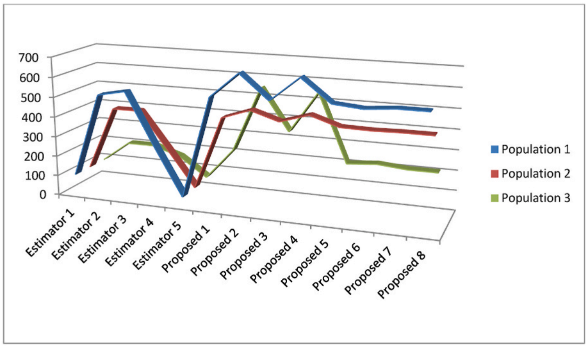

Fig. 2. Show percentage relative efficiency using populations 1-3.

to the true value of the population parameters being estimated. Consistency is a desirable property for estimators because it indicates that as more data is collected, the estimators improve and provide more reliable results. It suggests that the proposed estimators are effective in estimating the finite population DF accurately.

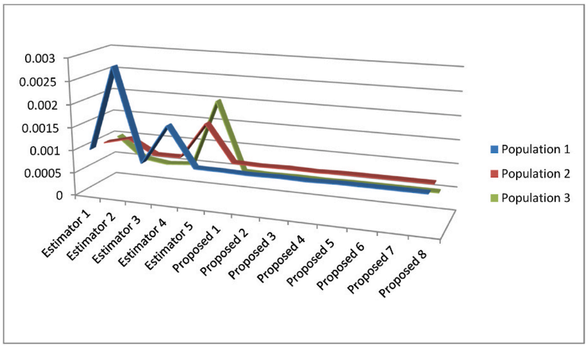

Fig. 3. MSE using simulated populations 1–3.

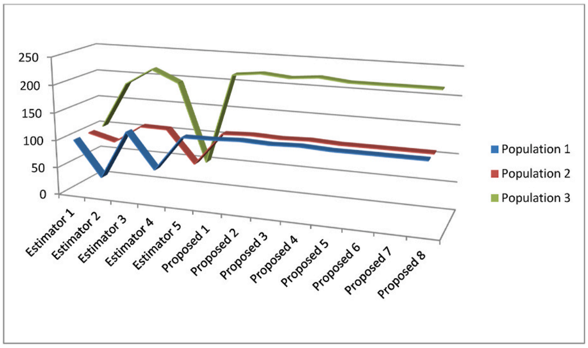

Fig. 4. PRE using simulated Population 1-3.

8. Conclusion

In this article, we propose an improved generalized class of estimators. These estimators are developed by combining regression type, log ratio and exponential ratio type estimators. We utilize transformation of population DF under stratified random sampling to construct these estimators. These expressions allow for a quantitative evaluation of the performance of the estimator. To validate the effectiveness of the proposed estimators, we conduct numerical investigations using actual data sets. We also perform a simulation analysis on an artificially created population. The simulation outcomes are accessible in Tables 6 and 7, which validate the dominance of the suggested improved class of estimators equated to usual estimators. Furthermore, the authors identify 8 sub-classes within the recommended class of estimators. These sub-classes are formed by employing various combinations of the proposed estimators, potentially offering different estimation approaches for different estimation approaches for different scenarios. Overall, the article introduces a novel approach to estimation by combining different types of estimators and provides empirical evidence supporting the helpfulness of the recommended generalized class of estimators. Based on the actual data and a simulation study the proposed subclass i.e. T^5_{logPR}, T^7_{logPR} and T^8_{logPR} provide similar results. The gain in efficiency of T^2_{logPR} and T^4_{logPR} are the best among all the entire suggested sub class of estimators. Based on the numerical results, we can see that the suggested sub-classes of improved estimators are out-performing as compared to existing counterparts. Therefore, the suggested improved class of estimators is preferable in further study.

Data availability

Data will be made available on request.

CRediT authorship contribution statement

Sohaib Ahmad: Writing – original draft, Visualization, Validation, Software, Data curation, Conceptualization. **Javid Shabbir:** Supervision. **Walid Emam:** Investigation, Funding acquisition. **Erum Zahid:** Software. **Muhammad Aamir:** Visualization, Supervision. **Mohd Khalid:** Methodology. **Malik Muhammad Anas:** Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationship that could have appeared to influence the work reported in this paper.

Acknowledgements

The study was funded by Researchers Supporting Project number (RSPD2024R749), King Saud University, Riyadh, Saudi Arabia.

References

[1] M.S. Ahmed, W. Abu-Dayyeh, Estimation of finite-population distribution function using multivariate auxiliary information, Statistics in Transition 5 (3) (2001) 501–507.
[2] R.L. Chambers, R. Dunstan, Estimating distribution functions from survey data, Biometrika 73 (3) (1986) 597–604.
[3] R.L. Chambers, A.H. Dorfman, P. Hall, Properties of estimators of the finite population distribution function, Biometrika 79 (3) (1992) 577–582.
[4] A.H. Dorfman, A comparison of design-based and model-based estimators of the finite population distribution function, Aust. J. Stat. 35 (1) (1993) 29–41.
[5] H.P. Singh, S. Singh, M. Kozak, A family of estimators of finite-population distribution function using auxiliary information, Acta applicandae mathematicae 104 (2) (2008) 115–130.
[6] S. Ahmad, S. Hussain, E. Zahid, A. Iftikhar, S. Hussain, J. Shabbir, M. Aamir, A simulation study: population distribution function estimation using dual auxiliary information under stratified sampling scheme, Math. Probl Eng. 2022 (2022) 3263022.
[7] A. Iftikhar, S. Hongbo, S. Hussain, S. Khan, B. Alnssyan, J. Shabbir, M. Aamir, A new exponential factor-type estimator for population distribution function using dual auxiliary variables under stratified random sampling, Math. Probl Eng. 2022 (2022) 7673945.
[8] E. Zamanzade, M. Mahdizadeh, H.M. Samawi, Efficient estimation of cumulative distribution function using moving extreme ranked set sampling with application to reliability, AStA Advances in Statistical Analysis 104 (3) (2020) 485–502.
[9] E. Zamanzade, M. Mahdizadeh, Distribution function estimation using concomitant-based ranked set sampling, Hacettepe Journal of Mathematics and Statistics 47 (3) (2018) 755–761.
[10] M. Mahdizadeh, E. Zamanzade, Estimation of a symmetric distribution function in multistage ranked set sampling, Stat. Pap. 61 (2) (2020) 851–867.
[11] S. Ahmad, K. Ullah, E. Zahid, J. Shabbir, M. Aamir, H.M. Alshanbari, A.A.A.H. El-Bagoury, A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling, Sci. Rep. 13 (1) (2023) 5415.
[12] M. Yaqub, J. Shabbir, Estimation of population distribution function involving measurement error in the presence of non-response, Commun. Stat. Theor. Methods 49 (10) (2020) 2540–2559.
[13] M. Yaqub, F. Sohil, J. Shabbir, M.U. Sohail, Estimation of population distribution function in the presence of non-response using stratified random sampling, Commun. Stat. Simulat. Comput. (2022) 1–29.
[14] S. Hussain, S. Ahmad, M. Saleem, S. Akhtar, Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling, PLoS One 15 (9) (2020) e0239098.
[15] N. Koyuncu, C. Kadilar, Ratio and product estimators in stratified random sampling, J. Stat. Plann. Inference 139 (8) (2009) 2552–2558.
[16] C. Kadilar, H. Cingi, Ratio estimators in stratified random sampling, Biom. J. 45 (2) (2003) 218–225.
[17] S. Aladag, H. Cingi, Improvement in estimating the population median in simple random sampling and stratified random sampling using auxiliary information, Commun. Stat. Theor. Methods 44 (5) (2015) 1013–1032.
[18] G.N. Singh, D. Bhattacharyya, A. Bandyopadhyay, A general class of calibration estimators under stratified random sampling in presence of various kinds of non-sampling errors, Commun. Stat. Simulat. Comput. 52 (2) (2023) 320–333.
[19] K. Ullah, Z. Hussain, I. Hussain, S.A. Cheema, Z. Almaspoor, M. El-Morshedy, Estimation of Finite Population Mean in Simple and Stratified Random Sampling by Utilizing the Auxiliary, Ranks, and Square of the Auxiliary Information, Mathematical Problems in Engineering, 2022, 2022.
[20] S. Ahmad, S. Hussain, J. Shabbir, E. Zahid, M. Aamir, R. Onyango, Improved Estimation of Finite Population Variance Using Dual Supplementary Information under Stratified Random Sampling, Mathematical Problems in Engineering, 2022, 2022.
[21] R. Tailor, S. Chouhan, Ratio-cum-product type exponential estimator of finite population mean in stratified random sampling, Commun. Stat. Theor. Methods 43 (2) (2014) 343–354.
[22] J.O. Muili, R.V.K. Singh, G.I. Onwuka, A. Audu, New calibration of finite population mean of combined ratio estimators in stratified random sampling, Fudma Journal of Sciences 6 (4) (2022) 37–44.
[23] C. Kadilar, H. Cingi, A new ratio estimator in stratified random sampling, Commun. Stat. Theor. Methods 34 (3) (2005) 597–602.
[24] M. Mishra, S. Singh, B.B. Khare, Separate ratio estimator using calibration approach for the population mean in stratified random sampling, Asian Journal of Probability and Statistics 20 (3) (2022) 64–73.
[25] H.A. Lone, R. Tailor, H.P. Singh, Generalized ratio-cum-product type exponential estimator in stratified random sampling, Commun. Stat. Theor. Methods 45 (11) (2016) 3302–3309.
[26] H.A. Lone, R. Tailor, M.R. Verma, Efficient separate class of estimators of population mean in stratified random sampling, Commun. Stat. Theor. Methods 46 (2) (2017) 554–573.
[27] S. Kumar, M. Trehan, J.S. Joorel, A simulation study: estimation of population mean using two auxiliary variables in stratified random sampling, J. Stat. Comput. Simulat. 88 (18) (2018) 3694–3707.
[28] U. Shahzad, M. Hanif, N. Koyuncu, A new estimator for mean under stratified random sampling, Mathematical Sciences 12 (2018) 163–169.
[29] S. Hussain, S. Ahmad, S. Akhtar, A. Javed, U. Yasmeen, Estimation of finite population distribution function with dual use of auxiliary information under non-response, PLoS One 15 (12) (2020) e0243584.
[30] S. Ahmad, S. Hussain, A. Al Mutairi, M. Kamal, M.U. Rehman, M.S. Mustafa, Improved estimation of population distribution function using twofold auxiliary information under simple random sampling, Heliyon 10 (2) (2024) 12345.
[31] L. Gu, S. Wang, L. Yang, Simultaneous confidence bands for the distribution function of a finite population in stratified sampling, Ann. Inst. Stat. Math. 71 (4) (2019) 983–1005.
[32] T. Zaman, C. Kadilar, On estimating the population mean using auxiliary character in stratified random sampling, J. Stat. Manag. Syst. 23 (8) (2020) 1415–1426.
[33] T. Zaman, Efficient estimators of population mean using auxiliary attribute in stratified random sampling, Adv. Appl. Stat. 56 (2) (2019) 153–171.
[34] T. Zaman, An efficient exponential estimator of the mean under stratified random sampling, Math. Popul. Stud. 28 (2) (2021) 104–121.
[35] S. Bhushan, A. Kumar, S.A. Lone, S. Anwar, N.M. Gunaime, An efficient class of estimators in stratified random sampling with an application to real data, Axioms 12 (6) (2023) 576.
[36] S. Bhushan, A. Kumar, S. Singh, Some efficient classes of estimators under stratified sampling, Commun. Stat. Theor. Methods 52 (6) (2023) 1767–1796.