Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling

Sardar Hussain¹, Sohaib Ahmad¹, Mariyam Saleem², Sohail Akhtar*²

¹ Department of Statistics, Quaid-I-Azam University, Islamabad, Pakistan, ² Department of Statistics, Government College University, Lahore, Pakistan

* s.akhtar@gcu.edu.pk

Abstract

The main purpose of this paper is to propose two new estimators for estimating the finite population distribution function under simple and stratified random sampling schemes using supplementary information on the distribution function, mean and ranks of the auxiliary variable. The mathematical expressions for the bias and mean squared error of the proposed estimators are derived under the first order of approximation. The theoretical and empirical studies showed that the proposed estimators uniformly perform better than the existing estimators in terms of the percentage relative efficiency.

1 Introduction

In survey sampling, the use of suitable auxiliary information improves the precision of estimators of the unknown population parameter(s). Several estimators of population parameters, including the population mean, median, total, distribution function, quantiles, etc., exist in the literature, and requires, supplementary information on one or more auxiliary variables along with the information on the study variable. A number of studies have been published on the estimation of the population mean. Some important references to the population mean estimation using auxiliary information include Murthy [1], Sisodia and Dwivedi [2], Srivastava and Jhajj [3], Rao [4], Upadhyaya and Singh [5], Singh [6], Kadilar and Cingi [7], Kadilar and Cingi [8], Gupta and Shabbir [9], Grover and Kaur [10], Grover and Kaur [11], Lu [12], Muneer et al. [13], Shabbir and Gupta [14], and Gupta and Yadav [15]. In these studies, the authors have proposed improved ratio, product, and regression type estimators for estimating the finite population mean. These authors have used a single auxiliary variable for the estimation procedure. In a recent study, Haq et al. [16] suggested using ranks of the auxiliary variable as an additional auxiliary variable to increase the precision of the estimator of the population mean in simple random sampling. However, to the best of our knowledge, there is no study concerning the use of two auxiliary information variables for the estimation of the finite population distribution function.
The problem of estimating the finite population cumulative distribution function (CDF) arises when the interest lies in finding out the proportion of the values of the study variable that are less or equal to a certain value. There are situations where estimating the CDF is deemed necessary. For example, for a nutritionist, it is interesting to know the proportion of the population that consumes 25% or more of the calorie intake from saturated fat. In the literature, many authors have estimated the CDF using information about one or more auxiliary variables. Chambers and Dunstan [17] suggested an estimator for estimating the CDF that requires information both on the study and auxiliary variables. Similarly, Rao et al. [18] and Rao [19] proposed ratio and difference/regression estimators for estimating the CDF under a general sampling design. Kuk [20] suggested a kernel method for estimating the CDF using the auxiliary information. Ahmed and Abu-Dayyeh [21] estimated the CDF using the information on multiple auxiliary variables. Rueda et al. [22] used a calibration approach to develop an estimator for estimating the CDF. Singh et al. [23] considered the problem of estimating the CDF and quantiles with the use of at the estimation stage of a survey. Moreover, Yaqub and Shabbir [24] considered a generalised class of estimators for estimating the CDF in the presence of non-response. Chen and Chen [25] investigated the injury severities of truck drivers in single- and multi-vehicle accidents on rural highways while Zeng et al. [26] worked on a multi-variate random-parameter Tobit model for analysing highway crash rates by injury severity, and Yaqub and Shabbir [24] considered a generalised class of estimators for estimating the CDF in the presence of non-response. Dong et al. [27] investigated the differences of single-vehicle and multi-vehicle accident probability using a mixed logit model, Chen et al. [28] worked on an analysis of hourly crash likelihood using an unbalanced panel data mixed logit model and real-time driving environmental big data, Zeng et al. [29] suggested jointly modelling area-level crash rates by severity, and Zeng et al. [30] used spatial joint analysis for zonal daytime and night-time crash frequencies using a Bayesian bivariate conditional autoregressive model. However, these estimators only used one auxiliary variate.

In this paper, we propose two new families of estimators for estimating the CDF using the information on the distribution function, ranks, and mean of the auxiliary variable under simple random sampling and stratified random sampling. The bias and mean squared errors (MSEs) of the existing and proposed estimators of the CDF are derived under the first order of approximation. The theoretical and numerical comparisons showed that the proposed estimators are more precise than the existing adapted estimators when estimating the CDF of a finite population.

2 Notation in simple random sampling

Consider a finite population \( \Omega = \{1, 2, \ldots, N\} \) of \( N \) distinct units. In order to estimate the finite population distribution function, a sample of size \( n \) units is drawn from \( \Omega \) using simple random sampling without replacement. Suppose \( Y \) and \( X \) are the study and auxiliary variables, respectively. Let \( Z \) denote the ranks of \( X \), \( I(Y \leq y) \) is indicator variable based on \( Y \) and \( I(X \leq x) \) is indicator variable based on \( X \). Similarly, \( F(y) = \sum_{i=1}^{N} I(Y_i \leq y) / N \) and \( \hat{F}(y) = \sum_{i=1}^{n} I(Y_i \leq y) / n \) are the population and sample distribution functions of \( Y(X) \), respectively. Let \( \bar{X} = \sum_{i=1}^{N} X_i / N \) and \( \hat{X} = \sum_{i=1}^{n} X_i / n \) \( (\bar{Z} = \sum_{i=1}^{N} Z_i / N \) and \( \hat{Z} = \sum_{i=1}^{n} Z_i / n \) are the population and sample means of \( X \) (\( Z \)), respectively.

In order to obtain the biases and mean squared errors (MSEs) of the adapted and proposed estimators of \( F(y) \), we consider the following relative error terms. Let

\[
e_1 = \frac{\hat{F}(y) - F(y)}{F(y)}, e_2 = \frac{\hat{F}(x) - F(x)}{F(x)}, e_3 = \frac{\hat{X} - \bar{X}}{\bar{X}} \quad \text{and} \quad e_4 = \frac{\hat{Z} - \bar{Z}}{\bar{Z}},
\]
such that $e_i = 0$ for $i = 0, 1, 2, 3$, where $E(\cdot)$ is the mathematical expectation of $(\cdot)$. Let

$$V_{r1}\hat{e} = E[e'_1 e'_2 e'_3 e'_4],$$

$$E(e'_1) = \hat{C} V_{2001}, E(e'_2) = \hat{C} V_{2001}, E(e'_3) = \hat{C} V_{2001}, E(e'_4) = \hat{C} V_{2001},$$

$$E(e_1 e_2) = \hat{R}_{12} C_1 C_2 = V_{1000}, E(e_1 e_3) = \hat{R}_{13} C_1 C_3 = V_{1000}, E(e_1 e_4) = \hat{R}_{14} C_1 C_4 = V_{1000},$$

$$E(e_2 e_3) = \hat{R}_{23} C_2 C_3 = V_{1000}, E(e_2 e_4) = \hat{R}_{24} C_2 C_4 = V_{1000},$$

where $\lambda = (N-n)/(nN)$, $S_i^2 = \sum_{i=1}^N (I(Y_i \leq y) - F(y))^2/(N-1)$,

$S_2^2 = \sum_{i=1}^N (I(X_i \leq x) - F(x))^2/(N-1),$

$S_1^2 = \sum_{i=1}^N (X_i - \bar{X})^2/(N-1),$

$S_3^2 = \sum_{i=1}^N (Z_i - \bar{Z})^2/(N-1),$

$C_1 = S_1 F(x), C_2 = S_2 F(x), C_3 = S_3 / \bar{X}, C_4 = S_4 / \bar{Z}, R_{13} = S_{13} / (F(y) \bar{X}), R_{23} = S_{23} / (F(y) \bar{Z}), R_{14} = S_{14} / (F(x) \bar{X}), R_{24} = S_{24} / (F(x) \bar{Z}),$

$S_{12} = \sum_{i=1}^N ([I(Y_i \leq y) - F(y)](I(X_i \leq x) - F(x))) / (N-1),$

$S_{13} = \sum_{i=1}^N ([I(Y_i \leq y) - F(y)](X_i - \bar{X})) / (N-1),$

$S_{23} = \sum_{i=1}^N ([I(X_i \leq x) - F(x)](X_i - \bar{X})) / (N-1),$

$S_{34} = \sum_{i=1}^N ([I(Y_i \leq y) - F(y)](Z_i - \bar{Z})) / (N-1),$

$S_{24} = \sum_{i=1}^N ([I(X_i \leq x) - F(x)](Z_i - \bar{Z})) / (N-1),$

In addition, let $R_{23}^2 = (R_{12}^2 + R_{13}^2 - 2R_{12} R_{13} R_{23})/(1 - R_{23}^2)$ be the multiple correlation coefficient of $I(Y \leq y)$ on $I(X \leq x)$ and $X$, and let $R_{12}^2 = (R_{12}^2 + R_{13}^2 - 2R_{12} R_{13} R_{23})/(1 - R_{23}^2)$ be the multiple correlation coefficient of $I(Y \leq y)$ on $I(X \leq x)$ and $Z$, under simple random sampling.

### 3 Adapeted estimators in simple random sampling

In this section, some estimators of finite population mean are adapted for estimating the finite CDF under simple random sampling. The biases and MSEs of these adapted estimators are derived under the first order of approximation.

1. The traditional unbiased estimator of $F(y)$ is

$$\hat{F}_1(y) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq y).$$

The variance of $\hat{F}_1(y)$ is

$$\text{Var}(\hat{F}_1(y)) = F'(y) V_{2001}.$$  

2. Cochran [31] adapted ratio estimator of $F(y)$ is

$$\hat{F}_2(y) = \hat{F}(y) \left( \frac{F(x)}{F'(x)} \right).$$


The bias and MSE of \( \hat{F}_2(y) \), to the first order of approximation, are

\[
\text{Bias}(\hat{F}_2(y)) \approx F(y)(V_{020} - V_{1100}),
\]

\[
\text{MSE}(\hat{F}_2(y)) \approx F^2(y)(V_{2000} + V_{0200} - 2V_{1100}).
\]  \hspace{1cm} (4)

If \( R_{12} > C_2/(2C_1) \), then \( \hat{F}_2(y) \) is better than \( \hat{F}_1(y) \) in terms of MSE.

3. Murthy [32] adapted product estimator of \( F(y) \) is

\[
\hat{F}_3(y) = \hat{F}(y) \left( \frac{\hat{F}(x)}{F(x)} \right).
\]  \hspace{1cm} (5)

The bias and MSE of \( \hat{F}_3(y) \), to the first order of approximation, are

\[
\text{Bias}(\hat{F}_3(y)) = F(y)V_{1100},
\]

\[
\text{MSE}(\hat{F}_3(y)) \approx F^2(y)(V_{2000} + V_{0200} + 2V_{1100}).
\]  \hspace{1cm} (6)

If \( -C_2/(2C_1) > R_{12} \), then \( \hat{F}_3(y) \) is better than \( \hat{F}_1(y) \) in terms of MSE.

4. The adapted difference estimator of \( F(y) \) is

\[
\hat{F}_4(y) = \hat{F}(y) + k(F(x) - \hat{F}(x)).
\]  \hspace{1cm} (7)

where \( k \) is an unknown constant. Here, \( \hat{F}_4(y) \) is an unbiased estimator of \( F(y) \). The minimum variance of \( \hat{F}_4(y) \) at the optimum value \( k_{(opt)} = (F(y)V_{1100})/(F(x)V_{0200}) \) is

\[
\text{Var}_{\text{min}}(\hat{F}_4(y)) = \frac{F^2(y)(V_{2000}V_{0200} - V_{1100}^2)}{V_{0200}}.
\]  \hspace{1cm} (8)

Here, (8) may be written as

\[
\text{Var}_{\text{min}}(\hat{F}_4(y)) = F^2(y)V_{2000}(1 - R^2_{12}).
\]  \hspace{1cm} (9)

5. Rao [4] adapted difference-type estimator of \( F(y) \) is

\[
\hat{F}_5(y) = k_1\hat{F}(y) + k_2(F(x) - \hat{F}(x)),
\]  \hspace{1cm} (10)

where \( k_1 \) and \( k_2 \) are unknown constants. The bias and MSE of \( \hat{F}_5(y) \), to the first order of approximation, are

\[
\text{Bias}(\hat{F}_5(y)) = F(y)(k_1 - 1),
\]

\[
\text{MSE}(\hat{F}_5(y)) \approx F^2(y) - 2k_1F^2(y) + k_2^2F^2(y) + k_3^2F^2(y)V_{2000}
\]

\[
-2k_1k_3F(y)F(x)V_{1100} + k_2^2F^2(x)V_{0200}.
\]  \hspace{1cm} (11)
The optimum values of $k_1$ and $k_2$, determined by minimizing (11), are

$$
k_{1(\text{opt})} = \frac{V_{0200}}{(V_{0200}V_{1200} - V_{2000}^2 + V_{0200})}$$

$$
k_{2(\text{opt})} = \frac{F(y)V_{1100}}{F(x)(V_{2000}V_{0200} - V_{1100}^2 + V_{0200})}.
$$

The minimum MSE of $\hat{F}_5(y)$ at the optimum values of $k_1$ and $k_2$ is

$$
\text{MSE}_{\text{min}}(\hat{F}_5(y)) = -\frac{F^2(y)V_{2000}V_{0200} - V_{1100}^2}{(V_{2000}V_{0200} - V_{1100}^2 + V_{0200})}.
$$

(12)

Here, (12) may be written as

$$
\text{MSE}_{\text{min}}(\hat{F}_5(y)) = \frac{F^2(y)V_{2000}(1 - R_{12}^2)}{1 + V_{2000}(1 - R_{12}^2)}.
$$

(13)

6. Singh et al. [33] adapted generalized ratio-type exponential estimator of $F(y)$ is

$$
\hat{F}_6(y) = \hat{F}(y)\exp\left(\frac{a(F(x) - \hat{F}(x))}{a(F(x) + F(x)) + 2b}\right),
$$

(14)

where $a$ and $b$ are known constants. The bias and MSE of $\hat{F}_6(y)$, to the first order of approximation, are

$$
\text{Bias}(\hat{F}_6(y)) \approx F(y)\left(\frac{3}{8} \theta^2 V_{0200} - \frac{1}{2} \theta V_{1100}\right),
$$

$$
\text{MSE}(\hat{F}_6(y)) \approx \frac{F^2(y)}{4} \left(4V_{2000} + \theta^2 V_{0200} - 4\theta V_{1100}\right),
$$

(15)

where $\theta = aF(x)/(aF(x)+b)$.

7. Grover and Kaur [11] adapted generalized class of ratio-type exponential estimator of $F(y)$ is

$$
\hat{F}_7(y) = \{k_3\hat{F}(y) + k_4(F(x) - \hat{F}(x))\}\exp\left(\frac{a(F(x) - \hat{F}(x))}{a(F(x) + F(x)) + 2b}\right),
$$

(16)

where $k_3$ and $k_4$ are unknown constants. The bias and MSE of $\hat{F}_7(y)$, to the first order of approximation, are
approximation, are

\[
\text{Bias}(\hat{F}_k(y)) \cong F(y)(k_3 - 1) + \frac{3}{8} \theta^2 k_4 F(y) + \frac{1}{2} \theta k_3 F(x) V_{2000} - \frac{1}{2} \theta F(y) V_{1100},
\]

\[
\text{MSE}(\hat{F}_k(y)) \cong k_4^2 F^2(x) V_{2000} + k_4^2 F^2(y) V_{2000} + 2 \theta k_3 k_4 F(y) F(x) V_{2000}
\]

\[-2k_3 F(x) V_{1100} + F^2(y) - 2k_4 F^2(y) + \theta k_3^2 F^2(y)
\]

\[+ k_4 F^2(y) V_{1100} - \theta k_3 F(x) F(y) V_{2000} - 2 \theta k_4^2 F^2(y) V_{1100} - \frac{3}{4} \theta^2 k_3 F^2(y) V_{2000} + \theta^2 k_4^2 F^2(y) V_{2000}
\]

(17)

The optimum values of \(k_3\) and \(k_4\), determined by minimizing (17), are

\[
k_{3(\text{opt})} = \frac{V_{2000}(\theta^2 V_{2000} - 8)}{8(-V_{2000}^2 V_{1100} - V_{2000})},
\]

\[
k_{4(\text{opt})} = \frac{F(y)(\theta^2 V_{2000}^2 - \theta^2 V_{2000} V_{1100} + 4R_{2000} V_{2000} - 4\theta V_{1100} - 4\theta V_{2000} + 8V_{1100})}{8F(x)(V_{2000} - V_{2000} V_{1100} - V_{2000}).}
\]

The simplified minimum MSE of \(\hat{F}_i(y)\) at the optimum values of \(k_3\) and \(k_4\) is

\[
\text{MSE}_{\text{min}}(\hat{F}_i(y)) \cong \frac{E(y)(\theta^2 V_{2000}^2 - \theta^2 V_{2000} V_{1100} + 4R_{2000} V_{2000} - 4\theta V_{1100} - 4\theta V_{2000} + 8V_{1100})}{8F(x)(V_{2000} - V_{2000} V_{1100} - V_{2000}).}
\]

(18)

Here, (18) may be written as

\[
\text{MSE}_{\text{min}}(\hat{F}_i(y)) \cong \text{Var}_{\text{min}}(\hat{F}_i(y)) - \frac{E^2(y)(\theta^2 V_{2000}^2 - \theta^2 V_{2000} V_{1100} + 8V_{2000} V_{2000})^2}{64 V_{2000}^2 (1 + V_{2000}(1 - R_{12}^2))},
\]

(19)

which shows that \(\hat{F}_i(y)\) is more precise than \(\hat{F}_i(y)\).

### 4 Proposed estimators in simple random sampling

The precision of an estimator increases by using suitable auxiliary information at the estimation stage. In previous studies, the sample distribution function of the auxiliary variable was used to improve the efficiencies of the existing distribution function estimators. In a recent study, Haq et al. [16] suggested using ranks of the auxiliary variable as an additional auxiliary variable to increase the precision of an estimator of the population mean. On similar lines, we use additional auxiliary information on sample means of the auxiliary and ranked-auxiliary variables along with the sample distribution function estimators of \(F(y)\) and \(F(x)\) to estimate the finite CDF. For this purpose, we propose two families of estimators for estimating \(F(y)\).

#### 4.1 First proposed family of estimators

For the first family of estimators, the sample mean of the auxiliary variable is used as an additional auxiliary variable; whilst in the second family of estimators, the sample mean of the ranked auxiliary variable is used as an additional auxiliary variable.
On the lines of $\hat{F}_s(y)$ and $\hat{F}_i(y)$, first proposed family of estimators for estimating $F(y)$ is given by

$$
\hat{F}_s(y) = \left\{ k_s \hat{F}(y) + k_6 \left( \frac{F(x) - \hat{F}(x)}{F(x)} \right) + k_7 \left( \frac{X - \hat{X}}{X} \right) \right\} \exp \left( \frac{a(F(x) - \hat{F}(x))}{a(F(x) + \hat{F}(x)) + 2b} \right),
$$

where $k_s$, $k_6$ and $k_7$ are unknown constants, $a(\neq 0)$ and $b$ are either two real numbers or function of known population parameters of $I(X \leq x)$, like $R_{12}$, $\beta_2$ (coefficient of kurtosis), $C_2$, etc.

The estimator $\hat{F}_s(y)$ can also be written as

$$
\hat{F}_s(y) = (k_s F(y)(1 + e_1) - k_6 e_2 - k_7 e_3) \left( 1 - \frac{1}{2} \theta e_2 + \frac{3}{8} \theta^2 e_2^2 + \cdots \right).
$$

Simplifying (21) and keeping terms only up to the second power of $e_i$'s, we can write

$$
(\hat{F}_s(y) - F(y)) = -F(y) + k_s F(y) + k_6 F(y) e_1 - \frac{1}{2} \theta k_6 F(y) e_2 - k_6 e_2 - k_7 e_3
$$

$$
+ \frac{3}{8} \theta^2 k_6 F(y) e_2^2 + \frac{1}{2} \theta k_6 F(y) e_3 - \frac{1}{2} \theta k_7 F(y) e_1 e_2 + \frac{1}{2} \theta k_7 e_1 e_3.
$$

The bias and MSE of $\hat{F}_s(y)$, to the first order of approximation, are

$$
\text{Bias}(\hat{F}_s(y)) \approx F(y)(k_s - 1) + \frac{3}{8} \theta^2 k_6 F(y) V_{0200} + \frac{1}{2} \theta k_6 V_{0200} - \frac{1}{2} \theta k_6 F(y) V_{1100} + \frac{1}{2} \theta k_7 V_{0110},
$$

$$
\text{MSE}(\hat{F}_s(y)) \approx F^2(y)(k_s - 1)^2 + k_6^2 F^2(y) V_{0200} + k_6^2 V_{0200} + k_7^2 V_{0200} + \theta^2 k_6^2 F^2(y) V_{0200}
$$

$$
- \theta k_s F(y) V_{0200} + 2k_s k_6 F(y) V_{0200} - k_6^2 F^2(y) V_{0200} + \theta k_6 F(y) V_{1100}
$$

$$
- 2k_6^2 F^2(y) V_{1100} - 2k_s k_6 F(y) V_{1100} - 2k_6 k_7 F(y) V_{0110} - \theta k_7 F(y) V_{0110}
$$

$$
+ 2k_6 k_7 F(y) V_{0110} - 2k_6 k_7 V_{0110}.
$$

The optimum values of $k_s$, $k_6$ and $k_7$, determined by minimizing (23), are

$$
k_{s(\text{opt})} = \frac{8 - \theta^2 V_{0200}}{8(1 + V_{0200}(1 - R_{12}^2))},
$$

$$
k_{6(\text{opt})} = \frac{\theta^2 V_{0200} (R_{23}^2 - 1) + V_{0200} (8 + \theta^2 V_{0200}) (R_{12} - R_{23} R_{13})}{8 V_{0200} (R_{23}^2 - 1) \left\{ -1 + V_{0200} (1 - R_{12}^2) \right\} - 1},
$$

$$
k_{7(\text{opt})} = \frac{F(y) V_{0200} (8 - \theta^2 V_{0200})(R_{12} - R_{23} R_{13})}{8 V_{0200} (R_{23}^2 - 1) \left\{ -1 + V_{0200} (1 - R_{12}^2) \right\}}.
$$
On similar lines, second proposed family of estimators for estimating

4.2 Second proposed family of estimators

On similar lines, second proposed family of estimators for estimating \( F(y) \) is given by

\[
\hat{F}_3(y) = \left\{ k_8 \tilde{F}(y) + k_9 \left( \frac{F(x) - \tilde{F}(x)}{F(x)} \right) + k_{10} \left( \frac{\bar{Z} - \bar{Y}}{Z} \right) \right\} \exp \left( \frac{a(F(x) - \hat{F}(x))}{a(F(x) - \tilde{F}(x)) + 2b} \right),
\]

where \( k_8, k_9 \) and \( k_{10} \) are unknown constants, \( a(\neq 0) \) and \( b \) are either two real numbers or functions of known population parameters of \( I(X \leq x) \), like \( R_{12}, \beta_2 \) (coefficient of kurtosis), \( C_2 \), etc.

The estimator \( \hat{F}_3(y) \) can also be written as

\[
\hat{F}_3(y) = \{ k_8 F(y)(1 + e_1) - k_9 e_2 - k_{10} e_1 \} \left( 1 - \frac{1}{2} \theta e_2 + \frac{3}{8} \theta^2 e_2^2 + \cdots \right).
\]

Simplifying (27) and keeping terms only up to the second power of \( e_i \)'s, we can write

\[
(\hat{F}_3(y) - F(y)) = -F(y) + k_9 F(y) + k_9 F(y) e_1 - \frac{1}{2} \theta k_9 F(y) e_2 - k_9 e_2 - k_{10} e_1
\]

\[
+ \frac{3}{8} \theta^2 k_9 F(y) e_2^2 + \frac{1}{2} \theta k_{10} e_2^2 - \frac{1}{2} \theta k_9 F(y) e_1 e_2 + \frac{1}{2} \theta k_{10} e_2 e_4.
\]
The bias and MSE of \( \hat{F}_3(y) \), to the first order of approximation, are

\[
\text{Bias}(\hat{F}_3(y)) \approx F(y)(k_8 - 1) + \frac{3}{8} \theta^2 k_8 F(y) V_{0200} + \frac{1}{2} \theta k_9 F(y) V_{1100} + \frac{1}{2} \theta k_9 V_{0101},
\]

\[
\text{MSE}(\hat{F}_3(y)) \approx F^2(y)(k_8 - 1)^2 + k_8^2 F^2(y) V_{2000} + k_9^2 V_{0200} + k_{10}^2 V_{0002} + \theta^2 k_8^2 F^2(y) V_{0200}
\]

\[
- \theta k_9 F(y) V_{0200} + 2\theta k_9 F(y) V_{0200} - \frac{3}{4} \theta^2 k_8 F^2(y) V_{0200} + \theta k_9 F^2(y) V_{1100}
\]

\[
- 2\theta k_9^2 F^2(y) V_{1100} - 2k_9 k_{10} F(y) V_{1001} - \theta k_{10} F^2(y) V_{0101}
\]

\[
+ 2\theta k_9 k_{10} F(y) V_{0101} - 2k_9 k_{10} V_{0101}.
\]

The optimum values of \( k_9 \), \( k_9 \), and \( k_{10} \), determined by minimizing (29), are

\[
k_{9(\text{opt})} = \frac{8 - \theta^2 V_{2000}}{8 \{1 + V_{2000}(1 - R_{21}^2)\}},
\]

\[
k_{9(\text{opt})} = \frac{V^{1/2} V_{0200} (R_{24}^2 - 1) + V^{1/2} V_{0200} (-8 + \theta^2 V_{0200})(R_{12} - R_{24}R_{14})}{8 V^{1/2} V_{0200} (R_{24}^2 - 1) \{-1 + V_{2000}(1 - R_{21}^2)\}},
\]

\[
k_{10(\text{opt})} = \frac{F(y) V^{1/2} V_{2000} (8 - \theta^2 V_{0200})(R_{12} - R_{24}R_{14})}{8 V^{1/2} V_{0200} (R_{24}^2 - 1) \{-1 + V_{2000}(1 - R_{21}^2)\}}.
\]

The simplified minimum MSE of \( \hat{F}_3(y) \) at the optimum values of \( k_9 \), \( k_9 \), and \( k_{10} \) is

\[
\text{MSE}_{\text{min}}(\hat{F}_3(y)) \approx \frac{F^2(y) \{64 V_{2000} (1 - R_{21}^2) - \theta^4 V_{0200} - 16 \theta V_{0200} V_{2000} (1 - R_{21}^2)\}}{64 \{1 + V_{2000}(1 - R_{21}^2)\}},
\]

where \( R_{21}^2 = \left( \frac{V_{0002}^2 V_{0200}^2 + V_{0200}^2 V_{0200}^2 - 2 V_{0200} V_{0200} V_{0200} V_{0200}}{V_{0200} V_{0200} V_{0200}} \right) \).

Here, (30) may be written as

\[
\text{MSE}_{\text{min}}(\hat{F}_3(y)) \approx \text{Var}_{\text{min}}(\hat{F}_3(y)) - H_1 - H_3,
\]

where

\[
H_1 = \frac{F^2(y)(\theta^2 V_{0200} - 8 V_{1100} + 8 V_{0200})^2}{64 V_{0200}^2 \{1 + V_{2000}(1 - R_{12}^2)\}}
\]

and

\[
H_3 = \frac{F^2(y)(\theta^2 V_{0200} - 8)^2 (V_{0200} V_{1001} V_{1100})^2}{64 V_{0200} V_{0002} V_{0200} (1 - R_{24}^2) \{-1 + V_{2000} V_{2000} (1 - R_{21}^2)\}}.
\]

It is clear that \( \hat{F}_3(y) \) is more precise than \( \hat{F}_i(y) \).

In Table 1, we put some members of the Singh et al. [33], Grover and Kaur [11], and proposed families of estimators with selected choices of \( a \) and \( b \).
Finite population distribution function estimation with dual use of auxiliary information

Table 1. Some members of the adapted and proposed distribution function estimators.

| $a$  | $b$    | $F_{a_b}(y)$ | $F_{a_b}^{(1)}(y)$ | $F_{a_b}^{(2)}(y)$ | $F_{a_b}^{(3)}(y)$ | $F_{a_b}^{(4)}(y)$ |
|------|--------|--------------|--------------------|--------------------|--------------------|--------------------|
| 1    | $C_{C_2}$ | $F_{10}^{(1)}(y)$ | $F_{10}^{(2)}(y)$ | $F_{10}^{(3)}(y)$ | $F_{10}^{(4)}(y)$ |
| 1    | $\beta_2$ | $F_{2}^{(1)}(y)$ | $F_{2}^{(2)}(y)$ | $F_{2}^{(3)}(y)$ | $F_{2}^{(4)}(y)$ |
| $\beta_1$ | $C_2$ | $F_{C_2}^{(1)}(y)$ | $F_{C_2}^{(2)}(y)$ | $F_{C_2}^{(3)}(y)$ | $F_{C_2}^{(4)}(y)$ |
| 1    | $R_{12}$ | $F_{R_{12}}^{(1)}(y)$ | $F_{R_{12}}^{(2)}(y)$ | $F_{R_{12}}^{(3)}(y)$ | $F_{R_{12}}^{(4)}(y)$ |
| $R_{12}$ | $C_{R_{12}}$ | $F_{R_{12}}^{(1)}(y)$ | $F_{R_{12}}^{(2)}(y)$ | $F_{R_{12}}^{(3)}(y)$ | $F_{R_{12}}^{(4)}(y)$ |
| $\beta_2$ | $R_{12}$ | $F_{R_{12}}^{(1)}(y)$ | $F_{R_{12}}^{(2)}(y)$ | $F_{R_{12}}^{(3)}(y)$ | $F_{R_{12}}^{(4)}(y)$ |
| 1    | $NF(x)$ | $F_{NF(x)}^{(1)}(y)$ | $F_{NF(x)}^{(2)}(y)$ | $F_{NF(x)}^{(3)}(y)$ | $F_{NF(x)}^{(4)}(y)$ |

https://doi.org/10.1371/journal.pone.0239098.t001

5 Efficiency comparisons in simple random sampling

In this section, the adapted and proposed estimators of $F(y)$ are compared in terms of the minimum MSEs. [(i)]

1. From (2) and (25),

$\text{MSE}_{\text{min}}(\tilde{F}_0(y)) < \text{Var}(\tilde{F}_1(y))$ if

$$F^2(y)V_{2000}R_{12}^2 + H_1 + H_2 > 0.$$ 

2. From (4) and (25),

$\text{MSE}_{\text{min}}(\tilde{F}_s(y)) < \text{MSE}(\tilde{F}_s(y))$ if

$$\frac{F^2(y)}{V_{2000}}(V_{0200} - V_{1000})^2 + H_1 + H_2 > 0.$$ 

3. From (6) and (25),

$\text{MSE}_{\text{min}}(\tilde{F}_s(y)) < \text{MSE}(\tilde{F}_3(y))$ if

$$\frac{F^2(y)}{V_{2000}}(V_{0200} + V_{1100})^2 + H_1 + H_2 > 0.$$ 

4. From (9) and (25),

$\text{MSE}_{\text{min}}(\tilde{F}_s(y)) < \text{MSE}_{\text{min}}(\tilde{F}_1(y))$ if

$$H_1 + H_2 > 0.$$
5. From (13) and (25),
\[
\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}_{\min}(\hat{F}_5(y)) \text{ if } \frac{F^2(y)\theta^2 V_{0200}(\theta^2 V_{0200} + 16 V_{2000}(1 - R^2_{12}))}{64\{1 + V_{2000}(1 - R^2_{12})\}} + H_2 > 0.
\]

6. From (15) and (25),
\[
\text{MSE}_{\min}(\hat{F}_6(y)) < \text{MSE}(\hat{F}_4(y)) \text{ if } \frac{F^2(y)}{V_{2000}} \left(\frac{\theta V_{2000}}{2} - V_{1100}\right)^2 + H_1 + H_2 > 0.
\]

7. From (19) and (25),
\[
\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}_{\min}(\hat{F}_7(y)) \text{ if } H_2 > 0.
\]

8. From (2) and (31),
\[
\text{MSE}_{\min}(\hat{F}_9(y)) < \text{Var}(\hat{F}_1(y)) \text{ if } \frac{1}{V_{2000}} F^2(y) V_{1100}^2 + H_1 + H_3 > 0.
\]

9. From (4) and (31),
\[
\text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}(\hat{F}_4(y)) \text{ if } \frac{F^2(y)}{V_{2000}} (V_{0200} - V_{1100})^2 + H_1 + H_3 > 0.
\]

10. From (6) and (31),
\[
\text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}(\hat{F}_3(y)) \text{ if } \frac{F^2(y)}{V_{2000}} (V_{0200} + V_{1100})^2 + H_1 + H_3 > 0.
\]

11. From (9) and (31),
\[
\text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}_{\min}(\hat{F}_4(y)) \text{ if } H_1 + H_3 > 0.
\]
12. From (13) and (31),
\[ \text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}_{\min}(\hat{F}_5(y)) \text{if} \]
\[ \frac{F'(y)\theta^2V_{0200}(\theta^2V_{0200} + 16V_{2000}(1 - R_{12}^2))}{64\{1 + V_{2000}(1 - R_{12}^2)\}} + H_3 > 0. \]

13. From (15) and (31),
\[ \text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}_{\min}(\hat{F}_6(y)) \text{if} \]
\[ \frac{F'(y)}{V_{2000}} \left( \frac{\theta V_{0200}}{2} - V_{1100} \right)^2 + H_1 + H_3 > 0. \]

14. From (19) and (31),
\[ \text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}_{\min}(\hat{F}_7(y)) \text{if} \]
\[ H_3 > 0. \]

The proposed families of estimators are always more precise than the adapted estimators as the above conditions (i)–(xiv) are always true.

6 Empirical study in simple random sampling

In this section, we conduct a numerical study to investigate the performances of the adapted and CDF estimators. For this purpose, five populations are considered. The summary statistics of these populations are reported in Table 2. The percentage relative efficiency (PRE) of an estimator \( \hat{F}_i(y) \) with respect to \( \hat{F}_1(y) \) is
\[ \text{PRE}(\hat{F}_i(y), \hat{F}_1(y)) = \frac{\text{Var}(\hat{F}_1(y))}{\text{MSE}_{\min}(\hat{F}_i(y))} \times 100, \]
where \( i = 2,3, \ldots, 9 \).

The PREs of distribution function estimators, computed from five populations, are given in Tables 7–11.

**Population I** (Source: Singh, [6])
- \( Y \): Duration of sleep (in minutes) and
- \( X \): Age of old persons.

**Population II** (Source: Gujarati, [34])
- \( Y \): The eggs produced in 1990 (millions) and
- \( X \): The price per dozen (cents) in 1990.

**Population III** (Source: Murthy, [1])
- \( Y \): The output of the factory and
- \( X \): The number of workers.

**Population IV** (Source: Sarndal, [35])
- \( Y \): Population in 1983 (in million) and
- \( X \): Population in 1980 (in million).

**Population V** (Source: Koyuncu and Kadilar, [36])
- \( Y \): Number of teachers and
- \( X \): number of students.
Table 2. Summary statistics for Populations I to V.

| Population | I     | II    | III   | IV    | V     |
|------------|-------|-------|-------|-------|-------|
| N          | 30    | 50    | 80    | 120   | 923   |
| n          | 5     | 5     | 10    | 20    | 180   |
| λ          | 0.16667 | 0.18  | 0.0875 | 0.04167 | 0.00447 |
| X          | 67.267 | 78.29 | 285.125 | 33.353 | 11440.5 |
| C₃         | 0.13725 | 0.27229 | 0.94846 | 3.32354 | 1.86453 |
| R          | 15.5  | 25.5  | 4.05  | 60.5  | 462   |
| C₄         | 0.56695 | 0.57166 | 0.57377 | 0.57493 | 0.57703 |

https://doi.org/10.1371/journal.pone.0239098.t002

From the numerical results, presented in Tables 3–7, it is observed that the PREs of all families of estimators change with the choice of a or b. It is further noted that the proposed families of estimators are more precise than the adapted distribution function estimators of Cochran [31], Murthy [32], Rao [4], Singh et al. [33] and Grover and Kaur [11], in terms of PRE. It can be seen that, for data sets I, III, IV and V, the second proposed family of the estimators perform better than the first proposed family of estimators, and for data set-II, it is also observed that the first proposed family of the estimators perform better than the second proposed family of estimators.

Table 3. PREs of distribution function estimators using Population I.

| Estimator | Value   | Estimator | Value   | Estimator | Value   | Estimator | Value   | Estimator | Value |
|-----------|---------|-----------|---------|-----------|---------|-----------|---------|-----------|-------|
| F₁(y)     | 100     | F₁⁽¹⁾(y)  | 78.81   | F₁⁽³⁾(y)  | 234.70  | F₁⁽⁵⁾(y)  | 253.84  | F₁⁽⁷⁾(y)  | 255.59 |
| F₂(y)     | 28.84   | F₂⁽¹⁾(y)  | 78.60   | F₂⁽³⁾(y)  | 234.72  | F₂⁽⁵⁾(y)  | 253.87  | F₂⁽⁷⁾(y)  | 255.62 |
| F₃(y)     | 187.50  | F₃⁽¹⁾(y)  | 78.81   | F₃⁽³⁾(y)  | 234.70  | F₃⁽⁵⁾(y)  | 253.84  | F₃⁽⁷⁾(y)  | 255.59 |
| F₄(y)     | 216.34  | F₄⁽¹⁾(y)  | 78.39   | F₄⁽³⁾(y)  | 234.75  | F₄⁽⁵⁾(y)  | 253.90  | F₄⁽⁷⁾(y)  | 255.65 |
| F₅(y)     | 233.58  | F₅⁽¹⁾(y)  | 173.45  | F₅⁽³⁾(y)  | 343.92  | F₅⁽⁵⁾(y)  | 377.99  | F₅⁽⁷⁾(y)  | 381.16 |
| F₆⁽¹⁾(y) | 161.15  | F₆⁽³⁾(y)  | 372.72  | F₆⁽⁵⁾(y)  | 411.96  | F₆⁽⁷⁾(y)  | 415.63  | F₆⁽⁹⁾(y)  | 258.08 |
| F₆⁽³⁾(y) | 150.14  | F₆⁽⁵⁾(y)  | 236.97  | F₆⁽⁷⁾(y)  | 256.31  | F₆⁽⁹⁾(y)  | 377.99  | F₆⁽¹⁾(y)  | 381.16 |
| F₆⁽⁵⁾(y) | 173.45  | F₆⁽⁷⁾(y)  | 343.92  | F₆⁽⁹⁾(y)  | 377.99  | F₆⁽¹⁾(y)  | 381.16  | F₆⁽⁵⁾(y)  | 258.29 |
| F₆⁽⁹⁾(y) | 151.69  | F₆⁽¹⁾(y)  | 237.16  | F₆⁽³⁾(y)  | 256.52  | F₆⁽⁵⁾(y)  | 258.29  | F₆⁽⁹⁾(y)  | 254.39 |

https://doi.org/10.1371/journal.pone.0239098.t003
Here we take five data sets for numerical illustration. We selected different sample sizes from these populations and, then, we used simple random sampling. The MSEs (minimum) of the proposed families of estimators are pointed out in Eqs 25 and 31. Finally, the adapted estimators and proposed families of estimators were compared with each other with respect to their PRE values. These results are set out in Tables 3–7. In Table 2, we see the summary statistics about the populations. We can also note from the numerical results presented in Tables 3–7 that the PREs of all families of estimators change with the choice of a or b. It is further noted that the proposed families of estimators are more precise than the adapted distribution function estimators of Cochran [31], Murthy [32], Rao [4], Singh et al. [33] and Grover and Kaur [11], in terms of the PRE. It can be seen that, for data sets I, III, IV and V, the second proposed family of the estimators perform better than the first proposed family of estimators, and, for data set II, it is also observed that the first proposed family of the estimators performs better than the second proposed family of estimators.

### 7 Notation in stratified random sampling

Consider a finite population $\Omega = \{1, 2, \ldots, N\}$ of $N$ distinct units, which is divided into $L$ homogeneous strata, where the size of $h$th stratum is $N_h$, for $h = 1, 2, \ldots, L$, such that $\sum_{h=1}^{L} N_h = N$.

### Table 4. PREs of distribution function estimators using Population II.

| Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value |
|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
| $\hat{F}_1(y)$ | 100 | $F_1^{(1)}(y)$ | 93.70 | $\hat{F}_2(y)$ | 120.43 | $F_2^{(1)}(y)$ | 125.66 | $\hat{F}_3(y)$ | 123.19 |
| $\hat{F}_1(y)$ | 44.642 | $F_1^{(2)}(y)$ | 93.65 | $\hat{F}_2(y)$ | 120.44 | $F_2^{(2)}(y)$ | 125.67 | $\hat{F}_3(y)$ | 123.19 |
| $\hat{F}_1(y)$ | 56.810 | $F_1^{(3)}(y)$ | 93.70 | $\hat{F}_2(y)$ | 120.43 | $F_2^{(3)}(y)$ | 125.66 | $\hat{F}_3(y)$ | 123.19 |
| $\hat{F}_1(y)$ | 101.46 | $F_1^{(4)}(y)$ | 93.59 | $\hat{F}_2(y)$ | 120.45 | $F_2^{(4)}(y)$ | 125.67 | $\hat{F}_3(y)$ | 123.20 |
| $\hat{F}_1(y)$ | 119.82 | $F_1^{(5)}(y)$ | 62.86 | $\hat{F}_2(y)$ | 131.42 | $F_2^{(5)}(y)$ | 137.18 | $\hat{F}_3(y)$ | 134.46 |
| $\hat{F}_1(y)$ | 62.99 | $F_1^{(6)}(y)$ | 131.33 | $\hat{F}_2(y)$ | 137.09 | $F_2^{(6)}(y)$ | 134.37 | $\hat{F}_3(y)$ | 122.58 |
| $\hat{F}_1(y)$ | 100.66 | $F_1^{(7)}(y)$ | 119.85 | $\hat{F}_2(y)$ | 125.04 | $F_2^{(7)}(y)$ | 122.58 | $\hat{F}_3(y)$ | 134.46 |
| $\hat{F}_1(y)$ | 62.86 | $F_1^{(8)}(y)$ | 131.42 | $\hat{F}_2(y)$ | 137.18 | $F_2^{(8)}(y)$ | 134.37 | $\hat{F}_3(y)$ | 122.58 |
| $\hat{F}_1(y)$ | 100.66 | $F_1^{(9)}(y)$ | 119.85 | $\hat{F}_2(y)$ | 125.04 | $F_2^{(9)}(y)$ | 122.58 | $\hat{F}_3(y)$ | 122.56 |
| $\hat{F}_1(y)$ | 99.75 | $F_1^{(10)}(y)$ | 119.83 | $\hat{F}_2(y)$ | 125.02 | $F_2^{(10)}(y)$ | 122.56 | $\hat{F}_3(y)$ | 122.56 |

https://doi.org/10.1371/journal.pone.0239098.t004

### Table 5. PREs of distribution function estimators using Population III.

| Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value |
|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
| $\hat{F}_1(y)$ | 100 | $F_1^{(1)}(y)$ | 140.40 | $\hat{F}_2(y)$ | 1037.21 | $F_2^{(1)}(y)$ | 1059.61 | $\hat{F}_3(y)$ | 1087.07 |
| $\hat{F}_1(y)$ | 1000 | $F_1^{(2)}(y)$ | 140.62 | $\hat{F}_2(y)$ | 1037.23 | $F_2^{(2)}(y)$ | 1059.64 | $\hat{F}_3(y)$ | 1087.09 |
| $\hat{F}_1(y)$ | 25.64 | $F_1^{(3)}(y)$ | 140.40 | $\hat{F}_2(y)$ | 1037.21 | $F_2^{(3)}(y)$ | 1059.61 | $\hat{F}_3(y)$ | 1087.07 |
| $\hat{F}_1(y)$ | 1025.64 | $F_1^{(4)}(y)$ | 140.84 | $\hat{F}_2(y)$ | 1037.26 | $F_2^{(4)}(y)$ | 1059.66 | $\hat{F}_3(y)$ | 1087.12 |
| $\hat{F}_1(y)$ | 1034.50 | $F_1^{(5)}(y)$ | 142.42 | $\hat{F}_2(y)$ | 1037.44 | $F_2^{(5)}(y)$ | 1059.85 | $\hat{F}_3(y)$ | 1087.31 |
| $\hat{F}_1(y)$ | 142.64 | $F_1^{(6)}(y)$ | 1037.46 | $\hat{F}_2(y)$ | 1059.88 | $F_2^{(6)}(y)$ | 1087.34 |
| $\hat{F}_1(y)$ | 138.68 | $F_1^{(7)}(y)$ | 1037.02 | $\hat{F}_2(y)$ | 1059.42 | $F_2^{(7)}(y)$ | 1086.87 |
| $\hat{F}_1(y)$ | 142.42 | $F_1^{(8)}(y)$ | 1037.44 | $\hat{F}_2(y)$ | 1059.85 | $F_2^{(8)}(y)$ | 1087.31 |
| $\hat{F}_1(y)$ | 138.89 | $F_1^{(9)}(y)$ | 1037.04 | $\hat{F}_2(y)$ | 1059.4 | $F_2^{(9)}(y)$ | 1086.89 |
| $\hat{F}_1(y)$ | 101.18 | $F_1^{(10)}(y)$ | 1034.50 | $\hat{F}_2(y)$ | 1056.84 | $F_2^{(10)}(y)$ | 1084.22 |

https://doi.org/10.1371/journal.pone.0239098.t005
Let $Y$ and $X$ be the study and auxiliary variables which take values $y_h$ and $x_h$, respectively, where $i = 1, 2, ..., n_h$ and $h = 1, 2, ..., L$, for estimating finite population distribution function, assume a sample of size $n_h$ is drawn from the $h$th stratum using simple random sampling without replacement, such that $\sum_{h=1}^{L} n_h = n$, where $n$ is the sample size.

Let $\hat{F}_n(y) = F(y) = \sum_{h=1}^{L} w_y f_h(y)$ and $\hat{F}_n(x) = F(x) = \sum_{h=1}^{L} w_x f_h(x)$ ($\hat{F}_n(y) = \tilde{F}(y) = \sum_{h=1}^{L} w_{yh} \tilde{F}_h(y)$ and $\hat{F}_n(x) = \tilde{F}(x) = \sum_{h=1}^{L} w_{xh} \tilde{F}_h(x)$) be the population (sample) distribution functions of $Y$ and $X$ under stratified random sampling, respectively, where $W_h = N_h/N$, $F_h(y) = \sum_{i=1}^{n_h} I(Y_{ih} \leq y)/N_h$, $F_h(x) = \sum_{i=1}^{n_h} I(X_{ih} \leq x)/N_h$, $\hat{F}_h(y) = \sum_{i=1}^{n_h} I(Y_{ih} \leq y)/n_h$, $\hat{F}_h(x) = \sum_{i=1}^{n_h} I(X_{ih} \leq x)/n_h$, $\hat{X}_h = \sum_{i=1}^{n_h} W_{xh} \hat{X}_h$ and $\hat{Z}_h = \sum_{i=1}^{n_h} W_{yh} \hat{Z}_h$ ($\tilde{X}_h = \tilde{X} = \sum_{h=1}^{L} W_{xh} \tilde{X}_h$ and $\tilde{Z}_h = \tilde{Z} = \sum_{h=1}^{L} W_{yh} \tilde{Z}_h$) be the population (sample) means of $X$ and $Z$ under stratified random sampling, respectively, where $\hat{X}_h = \sum_{i=1}^{n_h} X_{ih}/n_h$, $\hat{Z}_h = \sum_{i=1}^{n_h} Z_{ih}/n_h$.

| Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value |
|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
| $\hat{F}_n(y)$ | 100 | $\hat{F}_n^{(1)}(y)$ | 141.58 | $\hat{F}_n^{(2)}(y)$ | 1531.59 | $\hat{F}_n^{(3)}(y)$ | 1533.54 | $\hat{F}_n^{(4)}(y)$ | 1599.81 |
| $\hat{F}_n(y)$ | 1500.00 | $\hat{F}_n^{(5)}(y)$ | 141.73 | $\hat{F}_n^{(6)}(y)$ | 1531.60 | $\hat{F}_n^{(7)}(y)$ | 1533.55 | $\hat{F}_n^{(8)}(y)$ | 1599.82 |
| $\hat{F}_n(y)$ | 25.42 | $\hat{F}_n^{(9)}(y)$ | 141.58 | $\hat{F}_n^{(10)}(y)$ | 1531.59 | $\hat{F}_n^{(11)}(y)$ | 1533.54 | $\hat{F}_n^{(12)}(y)$ | 1599.81 |
| $\hat{F}_n(y)$ | 1525.42 | $\hat{F}_n^{(13)}(y)$ | 141.88 | $\hat{F}_n^{(14)}(y)$ | 1531.61 | $\hat{F}_n^{(15)}(y)$ | 1533.56 | $\hat{F}_n^{(16)}(y)$ | 1599.84 |
| $\hat{F}_n(y)$ | 1529.62 | $\hat{F}_n^{(17)}(y)$ | 142.95 | $\hat{F}_n^{(18)}(y)$ | 1531.70 | $\hat{F}_n^{(19)}(y)$ | 1533.65 | $\hat{F}_n^{(20)}(y)$ | 1599.93 |
| $\hat{F}_n(y)$ | 1531.19 | $\hat{F}_n^{(21)}(y)$ | 143.11 | $\hat{F}_n^{(22)}(y)$ | 1531.71 | $\hat{F}_n^{(23)}(y)$ | 1533.66 | $\hat{F}_n^{(24)}(y)$ | 1599.94 |
| $\hat{F}_n(y)$ | 140.39 | $\hat{F}_n^{(25)}(y)$ | 1531.49 | $\hat{F}_n^{(26)}(y)$ | 1533.44 | $\hat{F}_n^{(27)}(y)$ | 1533.71 | $\hat{F}_n^{(28)}(y)$ | 1599.93 |
| $\hat{F}_n(y)$ | 140.53 | $\hat{F}_n^{(29)}(y)$ | 1531.50 | $\hat{F}_n^{(30)}(y)$ | 1533.45 | $\hat{F}_n^{(31)}(y)$ | 1533.72 | $\hat{F}_n^{(32)}(y)$ | 1599.72 |
| $\hat{F}_n(y)$ | 100.80 | $\hat{F}_n^{(33)}(y)$ | 1529.62 | $\hat{F}_n^{(34)}(y)$ | 1531.57 | $\hat{F}_n^{(35)}(y)$ | 1533.75 | $\hat{F}_n^{(36)}(y)$ | 1597.75 |

https://doi.org/10.1371/journal.pone.0239098.t006

Table 6. PREs of distribution function estimators using Population IV.

https://doi.org/10.1371/journal.pone.0239098.t007

Table 7. PREs of distribution function estimators using Population V.
Table 8. Some members of the proposed and adapted estimators.

| a  | b     | \( \hat{F}_{k}(y) \) | \( \hat{F}_{\text{dc}}(y) \) | \( \hat{F}_{\text{mpr}_{1}}(y) \) | \( \hat{F}_{\text{mpr}_{2}}(y) \) |
|----|-------|-----------------------|---------------------------|-----------------------|-----------------------|
| 1  | \( C_{2} \) | \( F_{[1]}^{(1)}(y) \) | \( F_{[1]}^{(2)}(y) \) | \( F_{[1]}^{(3)}(y) \) | \( F_{[1]}^{(4)}(y) \) |
| 1  | \( \beta_{2(0)} \) | \( F_{[1]}^{(1)}(y) \) | \( F_{[1]}^{(2)}(y) \) | \( F_{[1]}^{(3)}(y) \) | \( F_{[1]}^{(4)}(y) \) |
| 2  | \( C_{2} \) | \( F_{[2]}^{(1)}(y) \) | \( F_{[2]}^{(2)}(y) \) | \( F_{[2]}^{(3)}(y) \) | \( F_{[2]}^{(4)}(y) \) |
| 3  | \( R_{12} \) | \( F_{[3]}^{(1)}(y) \) | \( F_{[3]}^{(2)}(y) \) | \( F_{[3]}^{(3)}(y) \) | \( F_{[3]}^{(4)}(y) \) |
| 4  | \( \beta_{2(0)} \) | \( F_{[4]}^{(1)}(y) \) | \( F_{[4]}^{(2)}(y) \) | \( F_{[4]}^{(3)}(y) \) | \( F_{[4]}^{(4)}(y) \) |
| 5  | \( R_{12} \) | \( F_{[5]}^{(1)}(y) \) | \( F_{[5]}^{(2)}(y) \) | \( F_{[5]}^{(3)}(y) \) | \( F_{[5]}^{(4)}(y) \) |
| 6  | \( \beta_{2(0)} \) | \( F_{[6]}^{(1)}(y) \) | \( F_{[6]}^{(2)}(y) \) | \( F_{[6]}^{(3)}(y) \) | \( F_{[6]}^{(4)}(y) \) |

In order to obtain the biases and MSES of the adapted and proposed estimators of \( F(y) \) under stratified random sampling, the following relative error terms are considered. Let

\[
\begin{align*}
\hat{\zeta}_{1} &= \frac{\hat{F}_{d}(y) - F(y)}{F(y)}, \\
\hat{\zeta}_{2} &= \frac{\hat{F}_{\text{dc}}(x) - F(x)}{F(x)}, \\
\hat{\zeta}_{3} &= \frac{\hat{X}_{d} - \bar{X}}{\bar{X}} \quad \text{and} \quad \hat{\zeta}_{4} = \frac{\hat{Z}_{d} - \bar{Z}}{\bar{Z}},
\end{align*}
\]

such that \( E(\hat{\zeta}_{i}) = 0 \) for \( i = 1,2,3,4 \), where \( E(\cdot) \) is the mathematical expectation of \( \cdot \). Let

\[
V_{\text{ntr}} = E\left[\hat{\zeta}_{1}^{\prime} \hat{\zeta}_{2}^{\prime} \hat{\zeta}_{3}^{\prime} \hat{\zeta}_{4}^{\prime}\right].
\]

Table 9. Summary statistics for Population I.

| h  | \( N_{h} \) | \( n_{h} \) | \( W_{h} \) | \( \lambda_{h} \) | \( F(y_{h}) \) | \( F(x_{h}) \) | \( X_{h} \) | \( Z_{h} \) |
|----|-------------|-------------|-------------|----------------|----------------|----------------|--------|--------|
| 1  | 127         | 31          | 0.1375      | 0.0244         | 0.3543         | 0.3779         | 20805  | 64     |
| 2  | 117         | 21          | 0.1267      | 0.0390         | 0.4188         | 0.4872         | 9212   | 59     |
| 3  | 103         | 29          | 0.1115      | 0.0248         | 0.4272         | 0.4660         | 14309  | 52     |
| 4  | 170         | 38          | 0.1841      | 0.0204         | 0.5765         | 0.6118         | 9479   | 86     |
| 5  | 205         | 22          | 0.2221      | 0.0406         | 0.6146         | 0.6537         | 5570   | 103    |
| 6  | 201         | 39          | 0.2177      | 0.0207         | 0.5025         | 0.3532         | 12998  | 101    |
| \( S_{1h} \) | \( S_{2h} \) | \( S_{3h} \) | \( S_{4h} \) | \( S_{5h} \) | \( S_{6h} \) | \( R_{12h} \) | \( R_{13h} \) | \( R_{23h} \) | \( R_{14h} \) | \( R_{124h} \) |
| .4802 | 0.4868 | 30487 | 36.806 | 0.9164 | -0.4602 | -0.4832 | -0.8155 | -0.8399 |
| .4955 | 0.5019 | 15181 | 33.919 | 0.8709 | -0.4147 | -0.4600 | -0.8437 | -0.8658 |
| .4970 | 0.5013 | 27550 | 29.877 | 0.9244 | -0.3928 | -0.4198 | -0.8489 | -0.8640 |
| .4956 | 0.4888 | 18219 | 49.219 | 0.8805 | -0.5074 | -0.5396 | -0.8417 | -0.8441 |
| .4879 | 0.4769 | 8498 | 59.322 | 0.8772 | -0.5579 | -0.5909 | -0.8305 | -0.8241 |
| .5012 | 0.4792 | 23094 | 58.168 | 0.7145 | -0.4334 | -0.3554 | -0.8125 | -0.8279 |
Table 10. Summary statistics for Population II.

| h | N_h | n_h | W_h | λ_h  | F(y_h) | F(x_h) | X_0 | Z_h |
|---|-----|-----|-----|------|--------|--------|-----|-----|
| 1 | 127 | 31  | 0.1375 | 0.0244 | 0.3543 | 0.3700 | 498.276 | 64  |
| 2 | 117 | 21  | 0.1267 | 0.0391 | 0.4188 | 0.4700 | 318.333 | 59  |
| 3 | 103 | 29  | 0.1115 | 0.0248 | 0.4272 | 0.4272 | 431.359 | 52  |
| 4 | 170 | 38  | 0.1841 | 0.0204 | 0.5765 | 0.5882 | 311.324 | 86  |
| 5 | 205 | 22  | 0.2221 | 0.0406 | 0.6146 | 0.6146 | 227.195 | 103 |
| 6 | 201 | 39  | 0.2177 | 0.0207 | 0.5025 | 0.4527 | 312.706 | 101 |

S_{1h} S_{2h} S_{3h} S_{4h} R_{12h} R_{13h} R_{23h} R_{24h} 

| 0.4802 | 0.4847 | 555.58 | 36.805 | 0.8983 | -0.5398 | -0.5580 | -0.8114 | -0.8363 |
| 0.4955 | 0.5013 | 365.45 | 33.918 | 0.8666 | -0.5205 | -0.5589 | -0.8382 | -0.8645 |
| 0.4970 | 0.4970 | 613.95 | 29.877 | 0.9603 | -0.4803 | -0.4823 | -0.8519 | -0.8568 |
| 0.4956 | 0.4936 | 458.02 | 49.217 | 0.9277 | -0.5818 | -0.5934 | -0.8490 | -0.8525 |
| 0.4879 | 0.4879 | 260.85 | 59.321 | 0.8764 | -0.6395 | -0.6457 | -0.8285 | -0.8429 |
| 0.5012 | 0.4990 | 397.04 | 58.167 | 0.8450 | -0.5063 | -0.4873 | -0.8343 | -0.8622 |

Where

\[ E(\hat{\zeta}_1) = \sum_{h=1}^{L} W_h^2 \hat{\zeta}_h^2 C_{1h} = \psi_{2001}, \]

\[ E(\hat{\zeta}_2) = \sum_{h=1}^{L} W_h^2 \hat{\zeta}_h^2 C_{2h} = \psi_{0201}, \]

\[ E(\hat{\zeta}_3) = \sum_{h=1}^{L} W_h^2 \hat{\zeta}_h^2 C_{3h} = \psi_{0021}, \]

\[ E(\hat{\zeta}_4) = \sum_{h=1}^{L} W_h^2 \hat{\zeta}_h^2 C_{4h} = \psi_{0002}, \]

\[ E(\hat{\zeta}_5) = \sum_{h=1}^{L} W_h^2 \hat{\zeta}_h^2 R_{12h} C_{1h} C_{2h} = \psi_{1101}, \]

\[ E(\hat{\zeta}_6) = \sum_{h=1}^{L} W_h^2 \hat{\zeta}_h^2 R_{13h} C_{1h} C_{3h} = \psi_{1011}, \]

\[ E(\hat{\zeta}_7) = \sum_{h=1}^{L} W_h^2 \hat{\zeta}_h^2 R_{23h} C_{2h} C_{3h} = \psi_{0111}, \]

\[ E(\hat{\zeta}_8) = \sum_{h=1}^{L} W_h^2 \hat{\zeta}_h^2 R_{24h} C_{2h} C_{4h} = \psi_{0101}. \]

Table 11. Summary statistics for Population III.

| h | N_h | n_h | W_h | λ_h  | F(y_h) | F(x_h) | X_0 | Z_h |
|---|-----|-----|-----|------|--------|--------|-----|-----|
| 1 | 106 | 9   | 0.1241 | 0.1017 | 0.5849 | 0.5472 | 24376 | 54  |
| 2 | 106 | 17  | 0.1241 | 0.0494 | 0.5189 | 0.5660 | 27422 | 54  |
| 3 | 94  | 38  | 0.1100 | 0.0157 | 0.3298 | 0.3404 | 72410 | 48  |
| 4 | 171 | 67  | 0.2002 | 0.0090 | 0.3684 | 0.3801 | 74365 | 87  |
| 5 | 204 | 7   | 0.2389 | 0.1379 | 0.4657 | 0.4657 | 26442 | 103 |
| 6 | 173 | 2   | 0.2026 | 0.4942 | 0.7052 | 0.7225 | 9844  | 87  |

S_{1h} S_{2h} S_{3h} S_{4h} R_{12h} R_{13h} R_{23h} R_{24h} 

| 0.4950 | 0.5001 | 49189 | 30.743 | 0.7722 | -0.4470 | -0.4523 | -0.7665 | -0.8622 |
| 0.5020 | 0.4979 | 5746  | 30.743 | 0.8330 | -0.4370 | -0.4816 | -0.8285 | -0.8585 |
| 0.4728 | 0.4784 | 160757| 27.279 | 0.7854 | -0.2957 | -0.3087 | -0.7509 | -0.8208 |
| 0.4839 | 0.4868 | 285063| 49.507 | 0.7755 | -0.1848 | -0.1936 | -0.7535 | -0.8408 |
| 0.5000 | 0.4965 | 45403 | 59.033 | 0.6750 | -0.3929 | -0.4129 | -0.7218 | -0.8578 |
| 0.4573 | 0.4490 | 18794 | 50.084 | 0.7319 | -0.5598 | -0.6102 | -0.7290 | -0.7755 |

https://doi.org/10.1371/journal.pone.0239098.t010

https://doi.org/10.1371/journal.pone.0239098.t011
Table 12. Summary statistics for Population IV.

| $h$ | $N_h$ | $n_h$ | $W_h$ | $\lambda_h$ | $F(y_h)$ | $F(x_h)$ | $X_h$ | $Z_h$ |
|-----|-------|-------|-------|-------------|---------|---------|-------|-------|
| 0   | 106   | 9     | 0.1241| 0.1017      | 0.5849  | 0.5189  | 24712 | 54    |
| 1   | 106   | 17    | 0.1241| 0.0494      | 0.5189  | 0.5660  | 26840 | 54    |
| 2   | 94    | 38    | 0.1100| 0.0157      | 0.3298  | 0.3404  | 72722 | 48    |
| 3   | 171   | 67    | 0.2002| 0.0090      | 0.3684  | 0.3743  | 73191 | 87    |
| 4   | 204   | 7     | 0.2389| 0.1379      | 0.4657  | 0.4363  | 26834 | 103   |
| 5   | 173   | 2     | 0.2026| 0.4942      | 0.7052  | 0.7341  | 9903  | 87    |

8 Adapted estimators in stratified random sampling

In this section, some estimators of finite population mean are adapted for estimating the finite CDF under stratified random sampling. The biases and MSEs of these adapted estimators are derived under the first order of approximation.

1. The traditional unbiased estimator of $F(y)$ is

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

The variance of $\hat{F}_{SRS}(y)$ is

$$
\text{Var}(\hat{F}_{SRS}(y)) = F^2(y)\psi_{2000}.
$$

Table 13. PREs of distribution function estimators using Population I.

| Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value |
|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
| $\hat{F}_1(y)$ | 100   | $\hat{F}^{(1)}_1(y)$ | 133.03| $\hat{F}^{(2)}_1(y)$ | 364.28| $\hat{F}^{(4)}_1(y)$ | 370.30| $\hat{F}^{(6)}_1(y)$ | 428.83|
| $\hat{F}_5(y)$ | 341.19| $\hat{F}^{(2)}_5(y)$ | 128.23| $\hat{F}^{(3)}_5(y)$ | 364.27| $\hat{F}^{(5)}_5(y)$ | 370.29| $\hat{F}^{(7)}_5(y)$ | 428.82|
| $\hat{F}_6(y)$ | 27.38 | $\hat{F}^{(3)}_6(y)$ | 139.35| $\hat{F}^{(4)}_6(y)$ | 364.29| $\hat{F}^{(6)}_6(y)$ | 370.32| $\hat{F}^{(8)}_6(y)$ | 428.85|
| $\hat{F}_7(y)$ | 363.74| $\hat{F}^{(4)}_7(y)$ | 128.99| $\hat{F}^{(5)}_7(y)$ | 364.27| $\hat{F}^{(7)}_7(y)$ | 370.30| $\hat{F}^{(9)}_7(y)$ | 428.82|
| $\hat{F}_8(y)$ | 364.23| $\hat{F}^{(5)}_8(y)$ | 138.40| $\hat{F}^{(6)}_8(y)$ | 364.29| $\hat{F}^{(8)}_8(y)$ | 370.32| $\hat{F}^{(10)}_8(y)$ | 428.85|
| $\hat{F}_9(y)$ | 139.35| $\hat{F}^{(6)}_9(y)$ | 364.29| $\hat{F}^{(7)}_9(y)$ | 370.32| $\hat{F}^{(9)}_9(y)$ | 428.85|
| $\hat{F}_{10}(y)$ | 128.99| $\hat{F}^{(7)}_{10}(y)$ | 364.27| $\hat{F}^{(8)}_{10}(y)$ | 370.29| $\hat{F}^{(10)}_{10}(y)$ | 428.82|
| $\hat{F}_{11}(y)$ | 145.36| $\hat{F}^{(8)}_{11}(y)$ | 364.31| $\hat{F}^{(9)}_{11}(y)$ | 370.34| $\hat{F}^{(11)}_{11}(y)$ | 428.87|
| $\hat{F}_{12}(y)$ | 124.67| $\hat{F}^{(9)}_{12}(y)$ | 364.26| $\hat{F}^{(10)}_{12}(y)$ | 370.29| $\hat{F}^{(12)}_{12}(y)$ | 428.81|
| $\hat{F}_{13}(y)$ | 100.09| $\hat{F}^{(10)}_{13}(y)$ | 364.23| $\hat{F}^{(11)}_{13}(y)$ | 370.26| $\hat{F}^{(13)}_{13}(y)$ | 428.78|

https://doi.org/10.1371/journal.pone.0239098.t012

https://doi.org/10.1371/journal.pone.0239098.t013
Table 14. PREs of distribution function estimators using Population III.

| Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value |
|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
| $\hat{F}_1(y)$ | 100 | $\hat{F}^{(1)}_6(y)$ | 136.01 | $\hat{F}^{(3)}_6(y)$ | 454.79 | $\hat{F}^{(1)}_8(y)$ | 464.37 | $\hat{F}^{(3)}_8(y)$ | 502.91 |
| $\hat{F}_3(y)$ | 427.32 | $\hat{F}^{(3)}_6(y)$ | 132.94 | $\hat{F}^{(3)}_8(y)$ | 454.78 | $\hat{F}^{(1)}_8(y)$ | 464.36 | $\hat{F}^{(3)}_8(y)$ | 502.90 |
| $\hat{F}_5(y)$ | 26.53 | $\hat{F}^{(3)}_6(y)$ | 139.80 | $\hat{F}^{(3)}_8(y)$ | 454.80 | $\hat{F}^{(1)}_8(y)$ | 464.38 | $\hat{F}^{(3)}_8(y)$ | 502.92 |
| $\hat{F}_7(y)$ | 454.24 | $\hat{F}^{(3)}_6(y)$ | 133.38 | $\hat{F}^{(3)}_8(y)$ | 454.78 | $\hat{F}^{(1)}_8(y)$ | 464.36 | $\hat{F}^{(3)}_8(y)$ | 502.90 |
| $\hat{F}_9(y)$ | 454.73 | $\hat{F}^{(3)}_6(y)$ | 140.06 | $\hat{F}^{(3)}_8(y)$ | 454.80 | $\hat{F}^{(1)}_8(y)$ | 464.38 | $\hat{F}^{(3)}_8(y)$ | 502.90 |
| $\hat{F}^{(3)}_6(y)$ | 140.57 | $\hat{F}^{(3)}_8(y)$ | 454.81 | $\hat{F}^{(3)}_8(y)$ | 464.38 | $\hat{F}^{(3)}_8(y)$ | 502.92 |
| $\hat{F}^{(3)}_6(y)$ | 132.71 | $\hat{F}^{(3)}_8(y)$ | 454.78 | $\hat{F}^{(3)}_8(y)$ | 464.36 | $\hat{F}^{(3)}_8(y)$ | 502.92 |
| $\hat{F}^{(3)}_6(y)$ | 144.15 | $\hat{F}^{(3)}_8(y)$ | 454.82 | $\hat{F}^{(3)}_8(y)$ | 464.39 | $\hat{F}^{(3)}_8(y)$ | 502.93 |
| $\hat{F}^{(3)}_6(y)$ | 129.86 | $\hat{F}^{(3)}_8(y)$ | 454.77 | $\hat{F}^{(3)}_8(y)$ | 464.35 | $\hat{F}^{(3)}_8(y)$ | 502.89 |
| $\hat{F}^{(3)}_6(y)$ | 100.10 | $\hat{F}^{(3)}_8(y)$ | 454.73 | $\hat{F}^{(3)}_8(y)$ | 464.30 | $\hat{F}^{(3)}_8(y)$ | 502.84 |

2. Cochran [31] adapted ratio estimator of $F(y)$ is

$$\hat{F}_{R_0}(y) = \hat{F}_d(y) \left( \frac{\hat{F}(x)}{\hat{F}_d(x)} \right).$$

The bias and MSE of $\hat{F}_{R_0}(y)$, to the first order of approximation, are

$$\text{Bias}(\hat{F}_{R_0}(y)) \approx F(y)(\psi_{000} - \psi_{110}),$$

$$\text{MSE}(\hat{F}_{R_0}(y)) \approx F^2(y)(\psi_{000} + \psi_{000} - 2\psi_{110}).$$

Table 15. PREs of distribution function estimators using Population III.

| Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value | Estimator | Value |
|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
| $\hat{F}_1(y)$ | 100 | $\hat{F}^{(1)}_6(y)$ | 120.05 | $\hat{F}^{(3)}_6(y)$ | 211.54 | $\hat{F}^{(1)}_8(y)$ | 214.04 | $\hat{F}^{(3)}_8(y)$ | 241.14 |
| $\hat{F}_3(y)$ | 181.83 | $\hat{F}^{(3)}_6(y)$ | 121.46 | $\hat{F}^{(3)}_8(y)$ | 211.49 | $\hat{F}^{(1)}_8(y)$ | 213.99 | $\hat{F}^{(3)}_8(y)$ | 241.09 |
| $\hat{F}_5(y)$ | 29.41 | $\hat{F}^{(3)}_6(y)$ | 131.91 | $\hat{F}^{(3)}_8(y)$ | 211.61 | $\hat{F}^{(1)}_8(y)$ | 214.10 | $\hat{F}^{(3)}_8(y)$ | 241.22 |
| $\hat{F}_7(y)$ | 208.64 | $\hat{F}^{(3)}_6(y)$ | 122.03 | $\hat{F}^{(3)}_8(y)$ | 211.50 | $\hat{F}^{(1)}_8(y)$ | 214.00 | $\hat{F}^{(3)}_8(y)$ | 241.10 |
| $\hat{F}_9(y)$ | 211.39 | $\hat{F}^{(3)}_6(y)$ | 132.66 | $\hat{F}^{(3)}_8(y)$ | 211.62 | $\hat{F}^{(1)}_8(y)$ | 214.12 | $\hat{F}^{(3)}_8(y)$ | 241.23 |
| $\hat{F}^{(3)}_6(y)$ | 133.40 | $\hat{F}^{(3)}_8(y)$ | 211.63 | $\hat{F}^{(3)}_8(y)$ | 214.13 | $\hat{F}^{(3)}_8(y)$ | 241.24 |
| $\hat{F}^{(3)}_6(y)$ | 120.88 | $\hat{F}^{(3)}_8(y)$ | 211.49 | $\hat{F}^{(3)}_8(y)$ | 213.99 | $\hat{F}^{(3)}_8(y)$ | 241.08 |
| $\hat{F}^{(3)}_6(y)$ | 139.14 | $\hat{F}^{(3)}_8(y)$ | 211.71 | $\hat{F}^{(3)}_8(y)$ | 214.21 | $\hat{F}^{(3)}_8(y)$ | 241.34 |
| $\hat{F}^{(3)}_6(y)$ | 116.98 | $\hat{F}^{(3)}_8(y)$ | 211.46 | $\hat{F}^{(3)}_8(y)$ | 213.95 | $\hat{F}^{(3)}_8(y)$ | 241.05 |
| $\hat{F}^{(3)}_6(y)$ | 100.08 | $\hat{F}^{(3)}_8(y)$ | 211.39 | $\hat{F}^{(3)}_8(y)$ | 213.88 | $\hat{F}^{(3)}_8(y)$ | 240.97 |

https://doi.org/10.1371/journal.pone.0239098.t014

https://doi.org/10.1371/journal.pone.0239098.t015
Table 16. PREs of distribution function estimators using Population IV.

| Estimator   | Value | Estimator   | Value | Estimator   | Value | Estimator   | Value | Estimator   | Value |
|-------------|-------|-------------|-------|-------------|-------|-------------|-------|-------------|-------|
| \(\hat{F}_{1}(y)\) | 100   | \(\hat{F}^{(1)}_{6}(y)\) | 125.53 | \(\hat{F}^{(2)}_{6}(y)\) | 208.38 | \(\hat{F}^{(3)}_{6}(y)\) | 210.61 | \(\hat{F}^{(4)}_{6}(y)\) | 231.59 |
| \(\hat{F}_{2}(y)\) | 179.42 | \(\hat{F}^{(1)}_{6}(y)\) | 120.78 | \(\hat{F}^{(2)}_{6}(y)\) | 208.34 | \(\hat{F}^{(3)}_{6}(y)\) | 210.56 | \(\hat{F}^{(4)}_{6}(y)\) | 231.34 |
| \(\hat{F}_{3}(y)\) | 29.65 | \(\hat{F}^{(1)}_{6}(y)\) | 131.68 | \(\hat{F}^{(2)}_{6}(y)\) | 208.45 | \(\hat{F}^{(3)}_{6}(y)\) | 210.68 | \(\hat{F}^{(4)}_{6}(y)\) | 231.47 |
| \(\hat{F}_{4}(y)\) | 205.48 | \(\hat{F}^{(1)}_{6}(y)\) | 121.39 | \(\hat{F}^{(2)}_{6}(y)\) | 208.34 | \(\hat{F}^{(3)}_{6}(y)\) | 210.56 | \(\hat{F}^{(4)}_{6}(y)\) | 231.35 |
| \(\hat{F}_{5}(y)\) | 208.23 | \(\hat{F}^{(1)}_{6}(y)\) | 132.28 | \(\hat{F}^{(2)}_{6}(y)\) | 208.46 | \(\hat{F}^{(3)}_{6}(y)\) | 210.69 | \(\hat{F}^{(4)}_{6}(y)\) | 231.48 |
| \(\hat{F}_{6}(y)\) | 133.08 | \(\hat{F}^{(1)}_{6}(y)\) | 208.47 | \(\hat{F}^{(2)}_{6}(y)\) | 210.70 | \(\hat{F}^{(3)}_{6}(y)\) | 231.49 |
| \(\hat{F}_{7}(y)\) | 120.31 | \(\hat{F}^{(1)}_{6}(y)\) | 208.33 | \(\hat{F}^{(2)}_{6}(y)\) | 210.55 | \(\hat{F}^{(3)}_{6}(y)\) | 231.34 |
| \(\hat{F}_{8}(y)\) | 139.07 | \(\hat{F}^{(1)}_{6}(y)\) | 208.56 | \(\hat{F}^{(2)}_{6}(y)\) | 210.78 | \(\hat{F}^{(3)}_{6}(y)\) | 231.59 |
| \(\hat{F}_{9}(y)\) | 116.30 | \(\hat{F}^{(1)}_{6}(y)\) | 208.30 | \(\hat{F}^{(2)}_{6}(y)\) | 210.52 | \(\hat{F}^{(3)}_{6}(y)\) | 231.30 |
| \(\hat{F}_{10}(y)\) | 100.08 | \(\hat{F}^{(1)}_{6}(y)\) | 208.23 | \(\hat{F}^{(2)}_{6}(y)\) | 210.46 | \(\hat{F}^{(3)}_{6}(y)\) | 231.23 |

https://doi.org/10.1371/journal.pone.0239098.t016

3. Murthy [32] adapted product estimator of \(F(y)\) is

\[
\hat{F}_{\text{Pr}}(y) = \hat{F}_{\hat{u}}(y) \left( \frac{\hat{F}_{\hat{u}}(x)}{\hat{F}(x)} \right). \tag{36}
\]

The bias and MSE of \(\hat{F}_{\text{Pr}}(y)\), to the first order of approximation, are

\[
\text{Bias}(\hat{F}_{\text{Pr}}(y)) = F(y)\psi_{1100},
\]

\[
\text{MSE}(\hat{F}_{\text{Pr}}(y)) \approx F^2(y)(\psi_{2000} + \psi_{0200} + 2\psi_{1100}). \tag{37}
\]

4. The adapted difference estimator of \(F(y)\) is

\[
\hat{F}_{\text{Reg}}(y) = \hat{F}_{\hat{u}}(y) + m(F(x) - \hat{F}_{\hat{u}}(x)), \tag{38}
\]

where \(m\) is an unknown constant. Here, \(\hat{F}_{\text{Reg}}(y)\) is an unbiased estimator of \(F(y)\). The minimum variance of \(\hat{F}_{\text{Reg}}(y)\) at the optimum value \(m_{\text{opt}} = (F(y)\psi_{1100})/(F(x)\psi_{0200})\) is

\[
\text{Var}_{\min}(\hat{F}_{\text{Reg}}(y)) = F^2(y)(\psi_{2000}\psi_{0200} - \psi_{1100}^2). \tag{39}
\]

Here, (39) may be written as

\[
\text{Var}_{\min}(\hat{F}_{\text{Reg}}(y)) = F^2(y)\psi_{2000}(1 - R_{12}^2). \tag{40}
\]

5. Rao [4] adapted difference-type estimator of \(F(y)\) is

\[
\hat{F}_{\text{R.D}}(y) = m_1\hat{F}_{\hat{u}}(y) + m_2(F(x) - \hat{F}_{\hat{u}}(x)), \tag{41}
\]

where \(m_1\) and \(m_2\) are unknown constants. The bias and MSE of \(\hat{F}_{\text{R.D}}(y)\), to the first order of
approximation, are

\[
\text{Bias}(\hat{F}_a(y)) = F(y)(m_1 - 1),
\]

\[
\text{MSE}(\hat{F}_{RD_a}(y)) \approx F^2(y) - 2m_1F^2(y) + m_1^2F^2(y)\psi_{1100} + m_2^2F^2(y)\psi_{0200}.
\]

(42)

The optimum values of \(m_1\) and \(m_2\), determined by minimizing (42), are

\[
m_{1(\text{opt})} = \frac{\psi_{0200}}{(\psi_{0200}\psi_{2000} - \psi^2_{1100} + \psi_{0200})},
\]

\[
m_{2(\text{opt})} = \frac{F(y)\psi_{1100}}{F(x)(\psi_{2000}\psi_{0200} - \psi^2_{1100} + \psi_{0200})}.
\]

The minimum MSE of \(\hat{F}_{RD_a}(y)\) at the optimum values of \(m_1\) and \(m_2\) is

\[
\text{MSE}_{\text{min}}(\hat{F}_{RD_a}(y)) = \frac{F^2(y)(\psi_{0200}\psi_{2000} - \psi^2_{1100})}{(\psi_{2000}\psi_{0200} - \psi^2_{1100} + \psi_{0200})}.
\]

Here, (43) may be written as

\[
\text{MSE}_{\text{min}}(\hat{F}_{RD_a}(y)) = \frac{F^2(y)\psi_{2000}(1 - R_{12}^2)}{1 + \psi_{2000}(1 - R_{12}^2)}.
\]

6. Singh et al. [33] adapted generalized ratio-type exponential estimator of \(F(y)\) is

\[
\hat{F}_{sa}(y) = \hat{F}_a(y)\exp\left(\frac{a_4(F(x) - \hat{F}_a(x))}{a_4(F(x) + \hat{F}_a(x)) + 2b_a}\right),
\]

where \(a_4\) and \(b_a\) are known constants. The bias and MSE of \(\hat{F}_{sa}(y)\), to the first order of approximation, are

\[
\text{Bias}(\hat{F}_{sa}(y)) \approx F(y)\left(\frac{3}{8}\Theta^2\psi_{0200} - \frac{1}{2}\Theta\psi_{1100}\right),
\]

\[
\text{MSE}(\hat{F}_{sa}(y)) \approx \frac{F^2(y)}{4}(4\psi_{0200}\Theta^2\psi_{2000} - 4\Theta\psi_{1100}),
\]  

(46)

where \(\Theta = a_4F(x)/(a_4F(x) + b_a)\).

7. Grover and Kaur [11] adapted generalized class of ratio-type exponential estimator of \(F\) (y) is

\[
\hat{F}_{g,K_a}(y) = \left\{m_3\hat{F}_a(y) + m_4(F(x) - \hat{F}_a(x))\right\}\exp\left(\frac{a_4(F(x) - \hat{F}_a(x))}{a_4(F(x) + \hat{F}_a(x)) + 2b_a}\right).
\]

(47)

where \(m_3\) and \(m_4\) are unknown constants. The bias and MSE of \(\hat{F}_{g,K_a}(y)\), to the first order of
approximation, are

$$\text{Bias}\left( \hat{F}_{\text{G,K}_a}(y) \right) \equiv F(y)(m_3 - 1) + \frac{3}{8} \Theta^2 m_4 F(y) + \frac{1}{2} \Theta m_4 F(x) \psi_{0200} - \frac{1}{2} \Theta F(y) \psi_{1100},$$

$$\text{MSE}(\hat{F}_{\text{G,K}_a}(y)) \equiv m_3^2 F^2(y) \psi_{0200} + m_4^2 F^2(y) \psi_{0200} + 2 \Theta m_4 m_3 F(y) F(x) \psi_{0200}$$

$$-2m_3 m_4 F(y) F(x) \psi_{1100} + F^2(y) - 2m_3 F^2(y) + \Theta m_4^2 F^2(y)$$

$$+m_3 F^2(y) \psi_{1100} - \Theta m_4 F(y) F(x) \psi_{0200} - 2 \Theta m_4^2 F^2(y) \psi_{1100}$$

$$- \frac{3}{4} \Theta^2 m_3^2 F^2(y) \psi_{0200} + \Theta^2 m_4^2 F^2(y) \psi_{0200}.$$  (48)

The optimum values of $m_3$ and $m_4$, determined by minimizing (48), are

$$m_{3(\text{opt})} = \frac{\psi_{0200}(\Theta^2 \psi_{0200} - 8)}{8(-\psi_{0200} \psi_{0200} + \psi_{1100}^2 - \psi_{0200}^2)},$$

$$m_{4(\text{opt})} = \frac{F(y)(\Theta^2 \psi_{0200} - \Theta^2 \psi_{0200} \psi_{1100} + 4 \Theta \psi_{0200} \psi_{1100}^2 - 4 \Theta \psi_{0200}^2 + 8 \psi_{1100}^2)}{8F(x)(\psi_{0200} \psi_{0200} - \psi_{1100}^2)^2}.$$

The simplified minimum MSE of $\hat{F}_{\text{G,K}_a}(y)$ at the optimum values of $m_3$ and $m_4$ is

$$\text{MSE}_{\text{min}}\left( \hat{F}_{\text{G,K}_a}(y) \right) \approx \frac{F^2(y)}{64} \left( 64 - 16 \Theta^2 \psi_{0200} - \frac{\psi_{0200}(-8 + \Theta^2 \psi_{0200})^2}{\psi_{0200}^2}\right).$$  (49)

Here, (49) may be written as

$$\text{MSE}_{\text{min}}\left( \hat{F}_{\text{G,K}_a}(y) \right) \approx \text{Var}_{\text{min}}\left( \hat{F}_{\text{Reg}_a}(y) \right) - \frac{F^2(y)(\Theta^2 \psi_{0200}^2 - 8 \psi_{1100}^2 + 8 \psi_{0200} \psi_{0200}^2)}{64 \psi_{0200}^2 \left(1 + \psi_{0200}^2(1 - R_{12}^2)\right)},$$  (50)

which shows that $\hat{F}_{\text{G,K}_a}(y)$ is more precise than $\hat{F}_{\text{Reg}_a}(y)$.

9 Proposed estimators in stratified random sampling

9.1 First proposed family of estimators

On the lines of $\hat{F}_{\text{K,D}_a}(y)$ and $\hat{F}_{\text{K}_a}(y)$, first proposed family of estimators for estimating $F(y)$ is given by

$$\hat{F}_{\text{prop}_g}(y) = \left\{ m_5 \hat{F}_g(y) + m_6 \left( \frac{F(x) - \hat{F}_g(x)}{F(x)} \right) + m_7 \left( \frac{\hat{X} - \hat{X}_g}{X} \right) \right\} \exp \left( \frac{a_g(F(x) - \hat{F}_g(x))}{a_g(F(x) + \hat{F}_g(x)) + 2b_g} \right),$$  (51)

where $m_5$, $m_6$ and $m_7$ are unknown constants, $a_g(\neq 0)$ and $b_g$ are either two real numbers or functions of known population parameters of $I(X \leq x)$, like $R_{12}$, $\beta_{2(\alpha)} = \sum_{i=1}^{L} W_i \beta_{2(\alpha)}$ (coefficient of kurtosis), $C_2 = \sum_{i=1}^{L} W_i C_{2i}$, etc.
The estimator $\hat{F}_{prop1a}(y)$ can also be written as

$$\hat{F}_{prop1a}(y) = \{m_y F(y)(1 + \zeta_1) - m_{\zeta_2} - m_{\zeta_3}\} \left(1 - \frac{1}{2} \Theta \zeta_2 + \frac{3}{8} \Theta^2 \zeta^2_2 + \cdots \right).$$

(52)

Simplifying (52) and keeping terms only up to the second power of $\zeta_i$'s, we can write

$$\left(\hat{F}_{prop1a}(y) - F(y)\right) = -F(y) + m_y F(y) + m_y F(y)\zeta_1 - \frac{3}{8} \Theta m_y F(y)\zeta^2_2 + \frac{1}{2} \Theta m_y F(y)\zeta\zeta_2 + \frac{1}{2} \Theta m_y F(y)\zeta_3.$$

(53)

The bias and MSE of $\hat{F}_{prop1a}(y)$, to the first order of approximation, are

$$\text{Bias}\left(\hat{F}_{prop1a}(y)\right) \cong F(y)(m_y - 1) + \frac{3}{8} \Theta m_y F(y)\psi_{0200} + \frac{1}{2} \Theta m_y F(y)\psi_{1100} + \frac{1}{2} \Theta m_y F(y)\psi_{0110},$$

$$\text{MSE}\left(\hat{F}_{prop1a}(y)\right) \cong F^2(y)(m_y - 1)^2 + m_y^2 F^2(y)\psi_{0200} + m_y^2 F^2(y)\psi_{0202} + \Theta^2 m_y^2 F^2(y)\psi_{0200}$$

$$\quad - \Theta m_y F(y)\psi_{0200} - 2 \Theta m_y F(y)\psi_{0110} - 2 \Theta m_y F(y)\psi_{0110} - 2 \Theta m_y F(y)\psi_{1100} - 2 \Theta m_y F(y)\psi_{1100}$$

$$+ 2 \Theta m_y F(y)\psi_{0202} - 2 m_y m_y F(y)\psi_{0110} - 2 m_y m_y F(y)\psi_{0110}.$$  

(54)

The optimum values of $m_y$, $m_6$ and $m_7$, determined by minimizing (54), are

$$m_y^{(opt)} = \frac{8 - \Theta^2 \psi_{0200}}{8\{1 + \psi_{2000}(1 - R^2_{1,23})\}},$$

$$m_6^{(opt)} = \frac{\Theta \psi_{0200}(R_{23}^2 - 1) + \psi_{2000}(8 + \Theta^2 \psi_{0200})(R_{12} - R_{23}R_{13})}{8\psi_{0200}(R_{23}^2 - 1)\{1 + \psi_{2000}(1 - R^2_{1,23})\}},$$

$$m_7^{(opt)} = \frac{F(y)\psi_{2000}(R_{12} - R_{23}R_{13})}{8\psi_{0200}(R_{23}^2 - 1)\{1 + \psi_{2000}(1 - R^2_{1,23})\}}.$$  

The simplified minimum MSE of $\hat{F}_{prop1a}(y)$ at the optimum values of $m_y$, $m_6$ and $m_7$ is

$$\text{MSE}_{\min}\left(\hat{F}_{prop1a}(y)\right) \cong \frac{F^2(y)(8\psi_{2000}(1 - R^2_{1,23}) - \Theta^2 \psi_{0200})}{64\{1 + \psi_{2000}(1 - R^2_{1,23})\}}.$$  

(55)

where $R^2_{1,23} = \left(\frac{\psi_{0110}\psi_{0200} + \psi_{1010}\psi_{0200} - \psi_{1000}\psi_{1010} - \psi_{0110} - \psi_{1100}}{\psi_{0010}\psi_{0200} - \psi_{0110}}\right).$

Here, (55) may be written as

$$\text{MSE}_{\min}\left(\hat{F}_{prop1a}(y)\right) \cong \text{Var}_{\min}\left(\hat{F}_{Rega}(y)\right) - T_1 - T_2.$$  

(56)
where
\[
T_1 = \frac{F^2(y)(\Theta^2\psi_{200}^2 - 8\psi_{100}^2 + 8\psi_{020}\psi_{200})^2}{64\psi_{020}(1 + \psi_{200}(1 - R_{12}^2))}
\]
and
\[
T_2 = \frac{F^2(y)(\Theta^2\psi_{020}^2 - 8)^2(\psi_{020}\psi_{100} - \psi_{010}\psi_{100})^2}{64\psi_{020}(1 - R_{12}^2)(1 + \psi_{200}(1 - R_{12}^2))}. 
\]

It can be seen that \( \hat{F}_{prop1a}(y) \) is more precise than \( \hat{F}_{prop1b}(y) \).

### 9.2 Second proposed family of estimators

On similar lines, second proposed family of estimators for estimating \( F(y) \) is given by

\[
\hat{F}_{prop2a}(y) = \left\{ m_8\hat{F}_a(y) + m_9\left( \frac{F(x) - \hat{F}_a(x)}{F(x)} \right) + m_{10}\left( \frac{Z - \hat{Z}_a}{Z} \right) \right\} \exp\left( \frac{a_1(F(x) - \hat{F}_a(x))}{a_1(F(x) - \hat{F}_a(x)) + 2b_1} \right), \tag{57}
\]

where \( m_8, m_9 \) and \( m_{10} \) are unknown constants, \( a_1 (\neq 0) \) and \( b_1 \) are either two real numbers or functions of known population parameters of \( I(X \leq x) \), like \( R_{12}, \beta_{2(\mu)} = \sum_{i=1}^{z} W_i \beta_{2h(i)} \) (coefficient of kurtosis), \( C_2 = \sum_{i=1}^{L} W_i C_{2i} \), etc.

The estimator \( \hat{F}_{prop2a}(y) \) can also be written as

\[
\hat{F}_{prop2a}(y) = \left( m_8F(y)(1 + \zeta_1) - m_9\zeta_2 - m_{10}\zeta_4 \right) \left( 1 - \frac{1}{2}\Theta_2^2 + \frac{3}{8}\Theta_2^2\zeta_2 + \cdots \right). \tag{58}
\]

Simplifying (58) and keeping terms only up to the second power of \( \zeta_i \)'s, we can write

\[
\left( \hat{F}_{prop2a}(y) - F(y) \right) = -F(y) + m_8F(y) + m_9F(y)\zeta_1 - \frac{1}{2}\Theta m_8F(y)\zeta_2 - m_9\zeta_2 - m_{10}\zeta_4 \\
+ \frac{3}{8}\Theta^2 m_8F(y)\zeta_2^2 + \frac{1}{2}\Theta m_9\zeta_2^2 - \frac{1}{2}\Theta m_8F(y)\zeta_1\zeta_2 + \frac{1}{2}\Theta m_{10}\zeta_2\zeta_4. \tag{59}
\]

The bias and MSE of \( \hat{F}_{prop2a}(y) \), to the first order of approximation, are

\[
\text{Bias}\left( \hat{F}_{prop2a}(y) \right) \\
\cong F(y)(m_8 - 1) + \frac{3}{8}\Theta^2 m_8F(y)\psi_{020} + \frac{1}{2}\Theta m_8F(y)\psi_{020} - \frac{1}{2}\Theta m_8F(y)\psi_{110} + \frac{1}{2}\Theta m_9\psi_{010},
\]

\[
\text{MSE}(\hat{F}_{prop2a}(y)) \cong F^2(y)(m_8 - 1)^2 + m^2_8F^2(y)\psi_{200} + m^2_9\psi_{020} + m^2_{10}\psi_{002} + m^2_{11}\psi_{020}F^2(y)\psi_{020} \\
- \Theta m_8F(y)\psi_{020} + 2\Theta m_8F(y)\psi_{020} - \frac{3}{4}\Theta^2 m_8F(y)\psi_{110} - \Theta m_9F(y)\psi_{110} \\
- 2\Theta^2 m_8F^2(y)\psi_{110} - 2m_8m_9F(y)\psi_{110} - 2m_8m_{10}F(y)\psi_{101} - \Theta m_{10}F(y)\psi_{101} \\
+ 2\Theta m_{11}F(y)\psi_{011} - 2m_{11}m_9\psi_{011}. \tag{60}
\]
The optimum values of \( m_8, m_9 \) and \( m_{10} \), determined by minimizing (60), are

\[
m_{8(\text{op})} = \frac{8 - \Theta^2 \psi_{0200}}{8(1 + \psi_{2000}(1 - R_{12})^2)},
\]

\[
m_{9(\text{op})} = F(y)[\Theta^2 \psi_{0200}^2 (R_{24}^2 - 1) + \psi_{2000}^2 (-8 + \Theta^2 \psi_{0200}) (R_{12} - R_{24} R_{14})] + 4 \Theta^2 \psi_{0200} (R_{24}^2 - 1) (-1 + \psi_{2000}(1 - R_{12}^2)) \frac{8 \psi_{0200} (R_{24}^2 - 1) (-1 + \psi_{2000}(1 - R_{12}^2))}{8 \psi_{0200} (R_{24}^2 - 1) (-1 + \psi_{2000}(1 - R_{12}^2))},
\]

\[
m_{10(\text{op})} = \frac{F(y) \psi_{2000} (8 - \Theta^2 \psi_{0200}) (R_{12} - R_{24} R_{14})}{8 \psi_{0200} (R_{24}^2 - 1) (-1 + \psi_{2000}(1 - R_{12}^2))}.
\]

The simplified minimum MSE of \( \hat{F}_{\text{prop}}(y) \) at the optimum values of \( m_8, m_9 \) and \( m_{10} \) is

\[
\text{MSE}_{\text{min}}(\hat{F}_{\text{prop}}(y)) \equiv \frac{F^2(y) \{64 \psi_{2000} (1 - R_{12}^2) - \Theta^2 \psi_{0200}^2 - 16 \Theta \psi_{0200} \psi_{2000} (1 - R_{12}^2)\}}{64 \{1 + \psi_{2000}(1 - R_{12}^2)\}};
\]

where \( R_{24}^2 = \left( \frac{\psi_{0200}^2 + \psi_{1200}^2 + \psi_{1000}^2}{\psi_{0200}^2 + \psi_{1200}^2 + \psi_{1000}^2} \right)^2 \).

Here, (61) may be written as

\[
\text{MSE}_{\text{min}}(\hat{F}_{\text{prop}}(y)) \equiv \text{Var}_{\text{min}}(\hat{F}_{\text{REG}}(y)) - T_1 - T_3,
\]

where

\[
T_1 = \frac{F^2(y) (\Theta^2 \psi_{0200} - 8 \psi_{1200}^2 + 8 \psi_{0300}^2)^2}{64 \psi_{0200}^2 (1 + \psi_{2000}(1 - R_{12}^2))}
\]

and

\[
T_3 = \frac{F^2(y) (\Theta^2 \psi_{0200} - 8)^2 (\psi_{0200} \psi_{1000} - \psi_{1200} \psi_{0100})^2}{64 \psi_{0200}^2 \psi_{0200} (1 - R_{24}^2) \{1 + \psi_{2000}(1 - R_{12}^2)\} \{1 + \psi_{2000}(1 - R_{12}^2)\}}.
\]

respectively, which shows that \( \hat{F}_{\text{prop}}(y) \) is more precise than \( \hat{F}_{\text{REG}}(y) \).

In Table 8, we put some members of the Singh et al. [33], Grover and Kaur [11], and proposed families of estimators with selected choices of \( a \) and \( b \).

### 10 Efficiency comparisons in stratified random sampling

In this section, the adapted and proposed estimators of \( F(y) \) are compared in terms of the minimum MSEs. ([60])

1. From (33) and (56),

\[
\text{MSE}_{\text{min}}(\hat{F}_{\text{prop}}(y)) < \text{Var}(\hat{F}_{\text{SRS}}(y)) \text{iff}
\]

\[
F^2(y) \psi_{2000} R_{12}^2 + T_1 + T_2 > 0.
\]
2. From (35) and (56),
\[
\text{MSE}_{\min} \left( \hat{F}_{\text{prop}_a} (y) \right) < \text{MSE} \left( \hat{F}_{R_a} (y) \right) \text{if}
\]
\[
\frac{F^2(y)}{\psi_{200}} \left( \psi_{0200} - \psi_{1100} \right)^2 + T_1 + T_2 > 0.
\]

3. From (37) and (56),
\[
\text{MSE}_{\min} \left( \hat{F}_{\text{prop}_b} (y) \right) < \text{MSE} \left( \hat{F}_{R_b} (y) \right) \text{if}
\]
\[
\frac{F^2(y)}{\psi_{200}} \left( \psi_{0200} + \psi_{1100} \right)^2 + T_1 + T_2 > 0.
\]

4. From (40) and (56),
\[
\text{MSE}_{\min} \left( \hat{F}_{\text{prop}_b} (y) \right) < \text{MSE}_{\min} \left( \hat{F}_{R_{\text{Reg}}} (y) \right) \text{if}
\]
\[
T_1 + T_2 > 0.
\]

5. From (44) and (56),
\[
\text{MSE}_{\min} \left( \hat{F}_{\text{prop}_b} (y) \right) < \text{MSE}_{\min} \left( \hat{F}_{R_{\text{D}}} (y) \right) \text{if}
\]
\[
\frac{F^2(y) \Theta^2 \psi_{0200} \left\{ \Theta^2 \psi_{0200} + 16 \psi_{2000} (1 - R_{12}^2) \right\}}{64 \left\{ 1 + \psi_{2000} (1 - R_{12}^2) \right\}} + T_2 > 0.
\]

6. From (46) and (56),
\[
\text{MSE}_{\min} \left( \hat{F}_{\text{prop}_b} (y) \right) < \text{MSE} \left( \hat{F}_{X_b} (y) \right) \text{if}
\]
\[
\frac{F^2(y)}{\psi_{200}} \left( \Theta \psi_{0200} \frac{2}{2} - \psi_{1100} \right)^2 + T_1 + T_2 > 0.
\]

7. From (50) and (56),
\[
\text{MSE}_{\min} \left( \hat{F}_{\text{prop}_b} (y) \right) < \text{MSE}_{\min} \left( \hat{F}_{G_{K_a}} (y) \right) \text{if}
\]
\[
T_2 > 0.
\]

8. From (33) and (62),
\[
\text{MSE}_{\min} \left( \hat{F}_{\text{prop}_b} (y) \right) < \text{Var} \left( \hat{F}_{\text{SRS}_a} (y) \right) \text{if}
\]
\[
\frac{1}{\psi_{0200}} F^2(y) \psi_{1100}^2 + T_1 + T_3 > 0.
\]
9. From (35) and (62),
   \[ \text{MSE}_{\text{min}}(\hat{F}_{\text{Prop}}(y)) < \text{MSE}(\hat{F}_{\text{Reg}}(y)) \text{ if} \]
   \[ \frac{F^2(y)}{\psi_{200}}(\psi_{200} - \psi_{1100})^2 + T_1 + T_3 > 0. \]

10. From (37) and (62),
    \[ \text{MSE}_{\text{min}}(\hat{F}_{\text{Prop}}(y)) < \text{MSE}(\hat{F}_{\text{Reg}}(y)) \text{ if} \]
    \[ \frac{F^2(y)}{\psi_{200}}(\psi_{200} + \psi_{1100})^2 + T_1 + T_3 > 0. \]

11. From (40) and (62),
    \[ \text{MSE}_{\text{min}}(\hat{F}_{\text{Prop}}(y)) < \text{MSE}_{\text{min}}(\hat{F}_{\text{Reg}}(y)) \text{ if} \]
    \[ T_1 + T_3 > 0. \]

12. From (44) and (62),
    \[ \text{MSE}_{\text{min}}(\hat{F}_{\text{Prop}}(y)) < \text{MSE}_{\text{min}}(\hat{F}_{\text{Reg}}(y)) \text{ if} \]
    \[ \frac{F^2(y)\Theta\psi_{200}\{\Theta\psi_{200} + 16\psi_{200}(1 - R_{12})\}}{64\{1 + \psi_{200}(1 - R_{12})\}} + T_3 > 0. \]

13. From (46) and (62),
    \[ \text{MSE}_{\text{min}}(\hat{F}_{\text{Prop}}(y)) < \text{MSE}_{\text{min}}(\hat{F}_{\text{Reg}}(y)) \text{ if} \]
    \[ \frac{F^2(y)}{\psi_{200}}\left(\frac{\Theta\psi_{200}}{2} - \psi_{1100}\right)^2 + T_1 + T_3 > 0. \]

14. From (50) and (62),
    \[ \text{MSE}_{\text{min}}(\hat{F}_{\text{Prop}}(y)) < \text{MSE}_{\text{min}}(\hat{F}_{\text{Reg}}(y)) \text{ if} \]
    \[ T_3 > 0. \]

The proposed families of estimators are always more precise than the adapted estimators as the above conditions (i)–(xiv) are always true.

### 11 Empirical study in stratified random sampling

In this section, we conduct a numerical study to investigate the performances of the adapted and proposed CDF estimators. For this purpose, five populations are considered. The summary statistics of these populations are reported in Tables 9–12. The PRE of an estimator \( \hat{F}(y) \)
with respect to \( \hat{F}_i(y) \) is

\[
\text{PRE}(\hat{F}_i(y), \hat{F}_i(y)) = \frac{\text{Var}(\hat{F}_1(y))}{\text{MSE}_{\text{min}}(\hat{F}_i(y))} \times 100
\]

where \( i = 2, 3, \ldots, 9 \).

The PREs of distribution function estimators, computed from four populations, are given in Tables 9–12.

**Population I** (Source: Koyuncu and Kadilar, [36])
- \( Y \): The number of teachers and
- \( X \): The number of students in both primary and secondary schools in Turkey in 2007 for 923 districts in six regions.

**Population II** (Source: Koyuncu and Kadilar, [36])
- \( Y \): The number of teachers and
- \( X \): The number of classes in both primary and secondary schools in Turkey in 2007 for 923 districts in six regions.

**Population III** (Source: Kadilar and Cingi, [37])
- \( Y \): Apple production amount in 1999 and
- \( X \): The number of apple trees in 1999.

**Population IV** (Source: Kadilar and Cingi, [37])
- \( Y \): Apple production amount in 1999 and
- \( X \): Apple production amount in 1998.

From the numerical results, presented in Tables 13–16, it is observed that the PREs of all families of estimators change with the choices of \( a \) and \( b \). It is further noted that the proposed families of estimators are more precise than the adapted distribution function estimators of Cochran [31], Murthy [32], Rao [4], Singh et al. [33] and Grover and Kaur [11], in terms of PRE. It can be seen that, for all data sets, the second proposed family of the estimators perform better than the first proposed family of estimators.

We computed sample size in stratum \( h \). Here we took sample sizes of 180, 180, 140 and 140 from four populations. Then we used stratified random sampling, and the MSEs (minimum) of the proposed families of estimators were computed in Eqs 56 and 62, respectively. Lastly, the adapted estimators and proposed families of estimators were compared with each other with respect to their PRE values. The PRE results are shown in Tables 13–16. In Tables 9–12, we can observe the descriptive statistics regarding the populations, strata, and sample size.

From the numerical results, presented in Tables 13–16, it is observed that the PREs of all families of estimators change with the choices of \( a \) and \( b \). It is further noted that the proposed families of estimators are more precise than the adapted distribution function estimators of Cochran [31], Murthy [32], Rao [4], Singh et al. [33] and Grover and Kaur [11], in terms of the PRE. It can be seen that, for all data sets, the second proposed family of estimators perform better than the first proposed family of estimators. We can also see a rise in the value of PREs when the value of \( a = 1 \); and \( b = R_{12} \); \( a = R_{12} \) and \( b = C_{12} \); \( a = R_{12} \) and \( b = \beta_{2(i)} \), and see a slight fall in PREs when \( a = 1 \) and \( b = NF(x) \).

**12 Conclusion**

In this paper, we have proposed two new families of estimators for estimating the finite population distribution function under simple and stratified random sampling schemes. The proposed estimators required supplementary information about the sample mean and the ranks of the auxiliary variable. The biases and MSEs of the proposed families of estimators were derived using the first order approximation. Based on theoretical and numerical comparative
studies, it can be concluded that the proposed families of estimators are more precise than their existing counterparts. Thus, we recommend using the sample mean and the ranks of the auxiliary variable with the proposed families of estimators for estimating the finite population distribution function under simple or stratified random sampling. It would be interesting to extend the suggested estimators to two-phase and stratified two-phase sampling schemes. Furthermore, the proposed estimators could also be generalised by utilising information about multi-auxiliary variables.

**Author Contributions**

**Conceptualization:** Sardar Hussain.

**Formal analysis:** Sardar Hussain.

**Funding acquisition:** Sardar Hussain.

**Investigation:** Sardar Hussain.

**Methodology:** Sardar Hussain, Sohaib Ahmad.

**Validation:** Sardar Hussain.

**Writing – original draft:** Sardar Hussain, Sohaib Ahmad.

**Writing – review & editing:** Mariyam Saleem, Sohail Akhtar.

**References**

1. Murthy MN. Sampling Theory and Methods. Statistical Pub. 1967 Society, Calcutta.
2. Sisodia BV, Dwivedi VK. Modified ratio estimator using coefficient of variation of auxiliary variable. Journal-Indian Society of Agricultural Statistics. 1981.
3. Srivastava SK, Jhajj HS. A class of estimators of the population mean using multiauxiliary information. Calcutta Statistical Association Bulletin. 1983 Mar; 32(1–2):47–56.
4. Rao TJ. On certain methods of improving ration and regression estimators. Communications in Statistics-Theory and Methods. 1991 Jan 1; 20(10):3325–40.
5. Upadhyaya LN, Singh HP. Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 1999 Sep; 41(5):627–36.
6. Singh S. Advanced Sampling Theory With Applications: How Michael™ Selected™ Amy. Springer Science & Business Media; 2003.
7. Kadilar C, Cingi H. Estimator of a population mean using two auxiliary variables in simple random sampling. International Mathematical Journal. 2004; 5:357–67.
8. Kadilar C, Cingi H. Improvement in estimating the population mean in simple random sampling. Applied Mathematics Letters. 2006 Jan 1; 19(1):75–9.
9. Gupta S, Shabbir J. On improvement in estimating the population mean in simple random sampling. Journal of Applied Statistics. 2008 May 1; 35(5):559–66.
10. Grover LK, Kaur P. Ratio type exponential estimators of population mean under linear transformation of auxiliary variable: theory and methods. South African Statistical Journal. 2011 Jan 1; 45(2):205–30.
11. Grover LK, Kaur P. A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. Communications in Statistics-Simulation and Computation. 2014 Jan 1; 43(7):1552–74.
12. Lu J. Efficient estimator of a finite population mean using two auxiliary variables and numerical application in agricultural, biomedical, and power engineering. Mathematical Problems in Engineering. 2017 Jan 1;2017. https://doi.org/10.1155/2017/2696108 PMID: 28978548
13. Muneer S, Shabbir J, Khalil A. Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables. Communications in Statistics-Theory and Methods. 2017 Mar 4; 46(5):2181–92.
14. Shabbir J, Gupta S. Estimation of finite population mean in simple and stratified random sampling using two auxiliary variables. Communications in Statistics-Theory and Methods. 2017 Oct 18; 46(20):10135–48.
15. Gupta RK, Yadav SK. Improved estimation of population mean using information on size of the sample. American Journal of Mathematics and Statistics. 2018; 8(2):27–35.

16. Haq A, Khan M, Hussain Z. A new estimator of finite population mean based on the dual use of the auxiliary information. Communications in Statistics-Theory and Methods. 2017 May 3; 46(9):4425–36.

17. Chambers RL, Dunstan R. Estimating distribution functions from survey data. Biometrika. 1986 Dec 1; 73(3):597–604.

18. Rao JN, Kovar JG, Mantel HJ. On estimating distribution functions and quantiles from survey data using auxiliary information. Biometrika. 1990 Jun 1: 365–75.

19. Rao JN. Estimating totals and distribution functions using auxiliary information at the estimation stage. Journal of Official Statistics. 1994 Jun 1; 10(2):153.

20. Kuk AY. A kernel method for estimating finite population distribution functions using auxiliary information. Biometrika. 1993 Jun 1; 80(2):385–92.

21. Ahmed MS. and Abu-Daayeh W. Estimation of finite-population distribution function using multivariate auxiliary information. Statistics in Transition. 2001 53:501–507.

22. Singh HP, Singh S, Kozak M. A family of estimators of finite-population distribution function using auxiliary information. Acta applicandae mathematicae. 2008 Nov 1; 104(2):115–30.

23. Yaqub M, Shabbir J. Estimation of population distribution function in the presence of non-response. Hacettepe Journal of Mathematics and Statistics. 2018 Apr 1; 47(2):471–511.

24. Dong B, Ma X, Chen F, Chen S. Investigating the differences of single-vehicle and multivehicle accident probability using mixed logit model. Journal of Advanced Transportation. 2018 Jan 1; 2018. https://doi.org/10.1155/2018/7905140 PMID: 32908973

25. Zeng Q, Guo Q, Wong SC, Wen H, Huang H, Guo Q, Pei X. Spatial joint analysis for zonal daytime and nighttime crash frequencies using a Bayesian bivariate conditional autoregressive model. Journal of Transportation Safety & Security. 2020 Apr 20; 12(4):566–85. https://doi.org/10.4271/2016-01-1439 PMID: 27648455

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

27. Gujarati DN. Basic econometrics. Tata McGraw-Hill Education: 2009.

28. Murthy RN. Product method of estimation. Sankhýa: The Indian Journal of Statistics, Series A. 1964 Jul 1:69–74.

29. Singh R, Chauhan P, Sawan N, Smarandache F. Improvement in estimating the population mean using exponential estimator in simple random sampling. International Journal of Statistics & Economics. 2009 3(A09):13–18

30. Särndal CE. Methods for estimating the precision of survey estimates when imputation has been used. Survey methodology. 1992; 18(2):241–52.

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

32. Kadilar C, Bing HB. Ratio estimators in stratified random sampling. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 2003 Mar; 45(2):218–25.