

RESEARCH ARTICLE

Estimation of finite population distribution function with dual use of auxiliary information under non-response

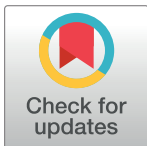
Sardar Hussain¹, Sohaib Ahmad¹, Sohail Akhtar^{2*}, Amara Javed^{2,3}, Uzma Yasmeen⁴

1 Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan, **2** Department of Statistics, Government College University, Lahore, Pakistan, **3** School of Statistics, Minhaj University, Lahore, Pakistan, **4** Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore, Pakistan

* s.akhtar@gcu.edu.pk, Akhtar013@gmail.com

Abstract

In this paper, we propose two new families of estimators for estimating the finite population distribution function in the presence of non-response under simple random sampling. The proposed estimators require information on the sample distribution functions of the study and auxiliary variables, and additional information on either sample mean or ranks of the auxiliary variable. We considered two situations of non-response (i) non-response on both study and auxiliary variables, (ii) non-response occurs only on the study variable. The performance of the proposed estimators are compared with the existing estimators available in the literature, both theoretically and numerically. It is also observed that proposed estimators are more precise than the adapted distribution function estimators in terms of the percentage relative efficiency.



OPEN ACCESS

Citation: Hussain S, Ahmad S, Akhtar S, Javed A, Yasmeen U (2020) Estimation of finite population distribution function with dual use of auxiliary information under non-response. PLoS ONE 15(12): e0243584. <https://doi.org/10.1371/journal.pone.0243584>

Editor: Feng Chen, Tongji University, CHINA

Received: September 7, 2020

Accepted: November 23, 2020

Published: December 17, 2020

Copyright: © 2020 Hussain et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

1. Introduction

One of the common problems in sample surveys is non-response. The non-response bias is serious concern in survey studies. It occurs in many ways, including linguistic problems, illness, due to response, due to non-acceptance, the process of return address misguided, captured by another person, etc. Sometime, sample survey experts use the auxiliary information to improve the precision of estimators. As expected, non-response not only reduces the precision of an estimator but also increases its bias. A number of research articles have been published on the estimation of population mean under non-response in order to control the non-response bias and to increase the efficiency of estimators. Hansen and Hurwitz [1] suggested that a sub-sample of earlier non-respondents can be re-communicated with a more expensive system. They adapted the first effort by a mail questionnaire and the second attempt by personal interview. In addition, firstly, they developed an estimator to estimate the population mean in the presence of non-response. For some related works on the non-response, we refer to Rao [2], Khare and Srivastava [3], Khare and Sinha [4, 5], Olufadi and Kumar [6], Muneer et al. [7], Pal and Singh [8], Ahmad and Shabbir [9] and the references cited therein.

The problem of estimating the finite population cumulative distribution function (CDF) arises when the interest lies in knowing the proportion of values of the study variable that are less or equal to a certain value. There are different situations where estimating the CDF is deemed necessary. For example, for an economist, it is interesting to know the proportion of the population that 27% or more Pakistanis do not have skills. Similarly, a soil scientist may be interested in estimating the distribution of clay percent in the soil. In addition, policy-makers may be interested in knowing the proportion of people living in a developing country below the poverty line.

In survey sampling literature, the authors have estimated the CDF using information on one or more auxiliary variables. Chambers and Dunstan [10] suggested an estimator for estimating the CDF that requires information both on the study and auxiliary variables. On similar lines, Rao et al. [11] and Rao [12] proposed ratio and difference/regression estimators for estimating the CDF under a general sampling design. Kuk [13] suggested a kernel method for estimating the CDF using the auxiliary information. Ahmed Abu-Dayyeh [14] estimated the CDF using information on multiple auxiliary variables. Chen and Wu [15] suggested a method for estimating the CDF and quantiles using the model-calibrated pseudo empirical likelihood. A calibration approach has been used by Rueda et al. [16] to devise an estimator for estimating the CDF. Singh et al. [17] considered the problem of estimating the CDF and quantiles with the use of auxiliary information at the estimation stage of a survey. Moreover, Chen et al. [18] investigate the injury severities of truck drivers in single-and multi-vehicle accidents on rural highways, Zeng et al. [19] worked on a multivariate random-parameters tobit model for analyzing highway crash rates by injury severity, Yaqub and Shabbir [20] considered a generalized class of estimators for estimating the CDF in the presence of non-response. Dong et al. [21] investigating the differences of single-vehicle and multi-vehicle accident probability using mixed logit model, Chen et al. [22] worked on analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data, Zeng et al. [23] suggested a jointly modeling area-level crash rates by severity and Zeng et al. [24] used spatial joint analysis for zonal daytime and night time crash frequencies using a bayesian bivariate conditional autoregressive model. Hussain et al. [25] proposed two new families of estimators using the supplementary information on auxiliary variable and exponential function for the population distribution functions in case of non-response under simple random sampling, and Hussain et al. [26] two new families of estimators for estimating the finite population distribution function are proposed under simple and stratified random sampling schemes using supplementary information on the distribution function, mean and ranks of the auxiliary variable.

In this paper, we propose two new families of estimators for estimating the CDF using information on the sample distribution function, mean and ranks of the auxiliary variable along with the information on the sample distribution function of the study variable under simple random sampling in the presence of non-response. 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 revealed that the proposed estimators are more precise than the existing adapted estimators when estimating the CDF.

The issue of non-response in survey sampling threatens to undermine the validity of inferences drawn from estimates based on those surveys. High non-response rates create opportunity or risk for bias in estimates and affect survey design, data collection, estimation and analysis Plewes et al. [27]. With these issues in mind, we propose two new families of estimators for estimating the CDF using information on the sample distribution function, mean and ranks of the auxiliary variable along with the information on the sample distribution function of the study variable under simple random sampling in the presence of non-response. 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 revealed that the proposed estimators are more precise than the existing adapted estimators when estimating the CDF.

The rest of the paper is organized as follows. In Section 2, some notations are given. In Section 3, we adapt some estimators of the finite population mean for estimating the finite CDF. The proposed estimators are given in Section 4. In Sections 5 and 6, theoretical and numerical comparisons are conducted, respectively. Finally, conclusions are drawn in Section 7.

2. Notations

Consider a finite population $\Omega = \{U_1, U_2, \dots, U_N\}$ of N distinct units, which is partitioned into N_1 respondents $\Omega_1 = \{U_1, U_2, \dots, U_{N_1}\}$ and N_2 non-respondents $\Omega_2 = \{U_1, U_2, \dots, U_{N_2}\}$, where $N = N_1 + N_2$. A sample of size n units is drawn from this population using simple random sampling without replacement. It is assumed that out of n units, n_1 units respond but $n_2 = n - n_1$ units do not respond. Clearly, n_1 and n_2 units belong to the respondent and non-respondent groups, respectively. Moreover, a sub-sample of size $r = n_2/k$ units, where $k > 1$, is drawn from n_2 units using simple random sampling without replacement, and, this time the response is obtained from r units. Let

Y : the study variable,

X : the auxiliary variable,

Z : ranks of X ,

$I(Y \leq y)$: indicator variable based on Y ,

$I(X \leq x)$: indicator variable based on X ,

$F(y) = \sum_{i=1}^N I(Y_i \leq y)/N$: the population distribution function of $I(Y \leq y)$,

$\hat{F}(y) = \sum_{i=1}^n I(Y_i \leq y)/n$: the sample distribution function of $I(Y \leq y)$,

$F(x) = \sum_{i=1}^N I(X_i \leq x)/N$: the population distribution function of $I(X \leq x)$,

$\hat{F}(x) = \sum_{i=1}^n I(X_i \leq x)/n$: the sample distribution function of $I(X \leq x)$,

$\bar{X} = \sum_{i=1}^N X_i/N$: the population mean of X ,

$\hat{\bar{X}} = \sum_{i=1}^n X_i/n$: the sample mean of X ,

$\bar{Z} = \sum_{i=1}^N Z_i/N$: the population mean of Z ,

$\hat{\bar{Z}} = \sum_{i=1}^n Z_i/n$: the sample mean of Z ,

$F_{(2)}(y) = \sum_{i=1}^{N_2} I(Y_i \leq y)/N_2$: the population distribution function of $I(Y \leq y)$ for non-response group,

$F_{(2)}(x) = \sum_{i=1}^{N_2} I(X_i \leq x)/N_2$: the population distribution function of $I(X \leq x)$ for non-response group,

$\bar{X}_{(2)} = \sum_{i=1}^{N_2} X_i/N_2$: the population mean of X for non-response group,

$\bar{Z}_{(2)} = \sum_{i=1}^{N_2} Z_i/N_2$: the population mean of Z for non-response group,

$\hat{F}_{(1)}(y) = \sum_{i=1}^{n_1} I(Y_i \leq y)/n_1$ denote the sample distribution function based on n_1 responding units out of n units,

$\hat{F}_{(1)}(x) = \sum_{i=1}^{n_1} I(X_i \leq x)/n_1$ denote the sample distribution function based on n_1 responding units out of n units,

$\bar{X}_{(1)} = \sum_{i=1}^{n_1} X_i/n_1$ denote the sample mean based on n_1 responding units out of n units,

$\bar{Z}_{(1)} = \sum_{i=1}^{n_1} Z_i/n_1$ denote the sample mean based on n_1 responding units out of n units,

$\hat{F}_{(2r)}(y) = \sum_{i=1}^r I(Y_i \leq y)/r$ denote the sample distribution function based on r responding units out of n_2 non-response units,

$\hat{F}_{(2r)}(x) = \sum_{i=1}^r I(X_i \leq x)/r$ denote the sample distribution function based on r responding units out of n_2 non-response units,

$\bar{X}_{(2r)} = \sum_{i=1}^r X_i/r$ denote the sample mean based on r responding units out of n_2 non-response units,

$\bar{Z}_{(2r)} = \sum_{i=1}^r Z_i/r$ denote the sample mean based on r responding units out of n_2 non-response units,

$S_1^2 = \sum_{i=1}^N (I(Y_i \leq y) - F(y))^2/(N - 1)$: the population variance of $I(Y \leq y)$,

$S_2^2 = \sum_{i=1}^N (I(X_i \leq x) - F(x))^2/(N - 1)$: the population variance of $I(X \leq x)$,

$S_3^2 = \sum_{i=1}^N (X_i - \bar{X})^2/(N - 1)$: the population variance of X ,

$S_4^2 = \sum_{i=1}^N (Z_i - \bar{Z})^2/(N - 1)$: the population variance of Z ,

$S_{1(2)}^2 = \sum_{i=1}^{N_2} (I(Y_i \leq y) - F_{(2)}(y))^2/(N_2 - 1)$: the population variance of $I(Y \leq y)$ for non-response group,

$S_{2(2)}^2 = \sum_{i=1}^{N_2} (I(X_i \leq x) - F_{(2)}(x))^2/(N_2 - 1)$: the population variance of $I(X \leq x)$ for non-response group,

$S_{3(2)}^2 = \sum_{i=1}^{N_2} (X_i - \bar{X}_{(2)})^2/(N_2 - 1)$: the population variance of X for non-response group,

$S_{4(2)}^2 = \sum_{i=1}^{N_2} (Z_i - \bar{Z}_{(2)})^2/(N_2 - 1)$: the population variance of Z for non-response group,

$C_1 = S_1/F(y)$: the population coefficient of variation of $I(Y \leq y)$,

$C_2 = S_2/F(x)$: the population coefficient of variation of $I(X \leq x)$,

$C_3 = S_3/\bar{X}$: the population coefficient of variation of X ,

$C_4 = S_4/\bar{Z}$: the population coefficient of variation of Z ,

$C_{1(2)} = S_{1(2)}/F_{(2)}(y)$: the population coefficient of variation of $I(Y \leq y)$ for non-response group,

$C_{2(2)} = S_{2(2)}/F_{(2)}(x)$: the population coefficient of variation of $I(X \leq x)$ for non-response group,

$C_{3(2)} = S_{3(2)}/\bar{X}_{(2)}$: the population coefficient of variation of X for non-response group,

$C_{4(2)} = S_{4(2)}/\bar{Z}_{(2)}$: the population coefficient of variation of Z for non-response group,

$S_{12} = \sum_{i=1}^N \{(I(Y_i \leq y) - F(y))(I(X_i \leq x) - F(x))\}/(N - 1)$: the population covariance

between $I(Y \leq y)$ and $I(X \leq x)$,

$$S_{13} = \sum_{i=1}^N \{(I(Y_i \leq y) - F(y))(X_i - \bar{X})\} / (N - 1): \text{the population covariance}$$

between $I(Y_i \leq y)$ and X ,

$$S_{23} = \sum_{i=1}^N \{(I(X_i \leq x) - F(x))(X_i - \bar{X})\} / (N - 1): \text{the population covariance}$$

between $I(X \leq x)$ and X ,

$$S_{14} = \sum_{i=1}^N \{(I(Y_i \leq y) - F(y))(Z_i - \bar{Z})\} / (N - 1): \text{the population covariance}$$

between $I(Y \leq y)$ and Z ,

$$S_{24} = \sum_{i=1}^N \{(I(X_i \leq x) - F(x))(Z_i - \bar{Z})\} / (N - 1): \text{the population covariance}$$

between $I(X \leq x)$ and Z ,

$$S_{12(2)} = \sum_{i=1}^{N_2} \{(I(Y_i \leq y) - F_{(2)}(y))(I(X_i \leq x) - F_{(2)}(x))\} / (N_2 - 1): \text{the population covariance between } I(Y \leq y) \text{ and } I(X \leq x) \text{ for non-response group,}$$

$$S_{13(2)} = \sum_{i=1}^{N_2} \{(I(Y_i \leq y) - F_{(2)}(y))(X_i - \bar{X}_{(2)})\} / (N_2 - 1): \text{the population covariance}$$

between $I(Y \leq y)$ and X for non-response group,

$$S_{23(2)} = \sum_{i=1}^{N_2} \{(I(X_i \leq x) - F_{(2)}(x))(X_i - \bar{X}_{(2)})\} / (N_2 - 1): \text{the population covariance}$$

between $I(X \leq x)$ and X for non-response group,

$$S_{14(2)} = \sum_{i=1}^{N_2} \{(I(Y_i \leq y) - F_{(2)}(y))(Z_i - \bar{Z}_{(2)})\} / (N_2 - 1): \text{the population covariance}$$

between $I(Y \leq y)$ and Z for non-response group,

$$S_{24(2)} = \sum_{i=1}^{N_2} \{(I(X_i \leq x) - F_{(2)}(x))(Z_i - \bar{Z}_{(2)})\} / (N_2 - 1): \text{the population covariance}$$

between $I(X \leq x)$ and Z for non-response group,

$$\Delta_{12} = S_{12} / (S_1 S_2): \text{the population correlation coefficient between } I(Y \leq y) \text{ and } I(X \leq x),$$

$$\Delta_{13} = S_{13} / (S_1 S_3): \text{the population correlation coefficient between } I(Y \leq y) \text{ and } X,$$

$$\Delta_{23} = S_{23} / (S_2 S_3): \text{the population correlation coefficient between } I(X \leq x) \text{ and } X,$$

$$\Delta_{14} = S_{14} / (S_1 S_4): \text{the population correlation coefficient between } I(Y \leq y) \text{ and } Z,$$

$$\Delta_{24} = S_{24} / (S_2 S_4): \text{the population correlation coefficient between } I(X \leq x) \text{ and } Z,$$

$$\nabla_{12(2)} = S_{12(2)} / (S_{1(2)} S_{2(2)}): \text{the population correlation coefficient between } I(Y \leq y) \text{ and } I(X \leq x) \text{ for non-response group,}$$

$$\nabla_{13(2)} = S_{13(2)} / (S_{1(2)} S_{3(2)}): \text{the population correlation coefficient between } I(Y \leq y) \text{ and } X \text{ for non-response group,}$$

$$\nabla_{23(2)} = S_{23(2)} / (S_{2(2)} S_{3(2)}): \text{the population correlation coefficient between } I(X \leq x) \text{ and } X \text{ for non-response group,}$$

$$\nabla_{14(2)} = S_{14(2)} / (S_{1(2)} S_{4(2)}): \text{the population correlation coefficient between } I(Y \leq y) \text{ and } Z \text{ for non-response group,}$$

$$\nabla_{24(2)} = S_{24(2)} / (S_{2(2)} S_{4(2)}): \text{the population correlation coefficient between } I(X \leq x) \text{ and } Z \text{ for non-response group.}$$

We can write

$$F(y) = W_1F_{(1)}(y) + W_2F_{(2)}(y) \tag{1}$$

Similarly, let

$$F(x) = W_1F_{(1)}(x) + W_2F_{(2)}(x) \tag{2}$$

Let

$$\bar{X} = W_1\bar{X}_{(1)} + W_2\bar{X}_{(2)} \tag{3}$$

Similarly let

$$\bar{Z} = W_1\bar{Z}_{(1)} + W_2\bar{Z}_{(2)} \tag{4}$$

where $W_j = N_j/N$, $F_j(y) = \sum_{i=1}^{N_j} I(Y_i \leq y)/N_j$, for $j = 1, 2$, $F_j(x) = \sum_{i=1}^{N_j} I(X_i \leq x)/N_j$, $\bar{X}_j = \sum_{i=1}^{N_j} X_i/N_j$ and $\bar{Z}_j = \sum_{i=1}^{N_j} Z_i/N_j$.

Following Hansen and Hurwitz [1], Yaqub and Shabbir [20] have suggested an unbiased estimator of $F(y)$ under non-response, which is given by

$$\hat{F}^*(y) = w_1\hat{F}_{(1)}(y) + w_2\hat{F}_{(2r)}(y)$$

with its variance

$$\text{Var}(\hat{F}^*(y)) = \lambda S_1^2 + \lambda_2 S_{1(2)}^2, \tag{5}$$

where $w_j = n_j/n$ for $j = 1, 2$, $\lambda = (1/n - 1/N)$ and $\lambda_2 = W_2(k-1)$.

Similarly, let

$$1\hat{F}^*(x) = w_1\hat{F}_{(1)}(x) + w_2\hat{F}_{(2r)}(x),$$

$$\hat{X}^* = w_1\bar{X}_{(1)} + w_2\bar{X}_{(2r)} \text{ and}$$

$$\hat{Z}^* = w_1\bar{Z}_{(1)} + w_2\bar{Z}_{(2r)},$$

are unbiased estimators of $F(x)$, \bar{X} and \bar{Z} , respectively, under non-response, with corresponding variances

$$1\text{Var}(\hat{F}^*(x)) = \lambda S_1^2 + \lambda_2 S_{1(2)}^2,$$

$$\text{Var}(\hat{X}^*) = \lambda S_2^2 + \lambda_2 S_{2(2)}^2 \text{ and}$$

$$\text{Var}(\hat{Z}^*) = \lambda S_3^2 + \lambda_2 S_{3(2)}^2,$$

respectively.

In order to obtain the biases and MSEs of the existing and proposed estimators, the following relative error terms are considered. Let

$$e_1^* = \frac{\hat{F}^*(y) - F(y)}{\frac{\frac{F(y), e_3^* = \hat{F}^*(x) - F(x)}{F(x), e_3^* = \hat{X} - \bar{X}}}{\frac{\hat{X}, e_4^* = \hat{Z} - \bar{Z}}{Z}}}$$

$$e_2 = \frac{\hat{F}(x) - F(x)}{\frac{F(x), e_3 = \hat{X} - \bar{X}}{\frac{\hat{X}, e_4 = \hat{Z} - \bar{Z}}{Z}}}$$

such that $E(e_i^*) = E(e_i) = 0$ for $i^* = 1, 2, 3, 4$ and $i = 2, 3, 4$, where $E(\cdot)$ stands for the mathematical expectation of (\cdot) . Let

$$V_{rstu} = E[e_1^r e_2^s e_3^t e_4^u] \text{ and } V_{rstu}^* = E[e_1^{*r} e_2^{*s} e_3^{*t} e_4^{*u}],$$

where $r, s, t, u = 1, 2, 3, 4$. Here,

$$E(e_1^{*2}) = \lambda C_1^2 + \lambda_2 C_{1(2)}^2 = V_{2000}^{*2}, E(e_2^{*2}) = \lambda C_2^2 + \lambda_2 C_{2(2)}^2 = V_{0200}^{*2},$$

$$E(e_3^{*2}) = \lambda C_3^2 + \lambda_2 C_{3(2)}^2 = V_{0020}^{*2}, E(e_4^{*2}) = \lambda C_4^2 + \lambda_2 C_{4(2)}^2 = V_{0002}^{*2},$$

$$E(e_2^2) = \lambda C_2^2 = V_{0200}, E(e_3^2) = \lambda C_3^2 = V_{0020}, E(e_4^2) = \lambda C_4^2 = V_{0002},$$

$$E(e_1^* e_2^*) = \lambda \Delta_{12} C_1 C_2 + \lambda_2 \nabla_{12(2)} C_{1(2)} C_{2(2)} = V_{1100}^*,$$

$$E(e_1^* e_3^*) = \lambda \Delta_{13} C_1 C_3 + \lambda_2 \nabla_{13(2)} C_{1(2)} C_{3(2)} = V_{1010}^*,$$

$$E(e_1^* e_4^*) = \lambda \Delta_{14} C_1 C_4 + \lambda_2 \nabla_{14(2)} C_{1(2)} C_{4(2)} = V_{1001}^*,$$

$$E(e_2^* e_3^*) = \lambda \Delta_{23} C_2 C_3 + \lambda_2 \nabla_{23(2)} C_{2(2)} C_{3(2)} = V_{0110}^*,$$

$$E(e_2^* e_4^*) = \lambda \Delta_{24} C_2 C_4 + \lambda_2 \nabla_{24(2)} C_{2(2)} C_{4(2)} = V_{0101}^*,$$

$$E(e_1^* e_2) = \lambda \Delta_{12} C_1 C_2 = V_{1100}, E(e_1^* e_3) = \lambda \Delta_{13} C_1 C_3 = V_{1010},$$

$$E(e_1^* e_4) = \lambda \Delta_{14} C_1 C_4 = V_{1001}, E(e_2 e_3) = \lambda \Delta_{23} C_2 C_3 = V_{0110},$$

$$E(e_2 e_4) = \lambda \Delta_{24} C_2 C_4 = V_{0101}.$$

Let $\Delta_{1.23}^2 = \left(\frac{V_{1100}^{*2} V_{0020}^* + V_{1010}^{*2} V_{0200}^* - 2V_{1010}^* V_{1100}^* V_{0110}^*}{V_{2000}^{*2} (V_{0200}^* V_{0020}^* - V_{0110}^{*2})} \right)$: the coefficient of multiple determination of $I(Y \leq y)$ on $I(X \leq x)$ and X with situation-I under non-response.

Let $\nabla_{1.23(2)}^2 = \left(\frac{V_{1100}^{*2} V_{0020} + V_{1010}^{*2} V_{0200} - 2V_{1010}^* V_{1100} V_{0110}}{V_{2000}^{*2} (V_{0200} V_{0020} - V_{0110}^2)} \right)$: the coefficient of multiple determination of $I(Y \leq y)$ on $I(X \leq x)$ and X with situation-II under non-response.

Let $\Delta_{1.24}^2 = \left(\frac{V_{1100}^{*2} V_{0002}^* + V_{1001}^{*2} V_{0200}^* - 2V_{1001}^* V_{1100}^* V_{0101}^*}{V_{2000}^{*2} (V_{0200}^* V_{0002}^* - V_{0101}^{*2})} \right)$: the coefficient of multiple determination of $I(Y \leq y)$ on $I(X \leq x)$ and Z with situation-I under non-response.

Let $\nabla_{1.24(2)}^2 = \left(\frac{V_{1100}^2 V_{0002} + V_{1001}^2 V_{0200} - 2V_{1001} V_{1100} V_{0101}}{V_{2000}^* (V_{0200} V_{0002} - V_{0101}^2)} \right)$: the coefficient of multiple determination of $I(Y \leq y)$ on $I(X \leq x)$ and Z with situation-II under non-response.

Under non-response, two situations are considered. The situation-I refers to the non-response both on study and auxiliary variables while situation-II refers to the non-response only on the study variable. For notational convenience, we follow the notations given in Table 1.

3. Adapted estimators

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

1. Cochran [28] adapted ratio estimator of $F(y)$ is

$$\hat{F}_2(y) = \hat{F}_H(y) \left(\frac{F(x)}{\hat{F}_H(x)} \right). \tag{6}$$

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

$$\text{Bias}(\hat{F}_2(y)) \cong F(y)(Q_{0200} - Q_{1100}),$$

Table 1. Estimators, variances, covariances and correlation under non-response situations.

Situation		I	II
Estimator			
$\hat{F}_H(y)$	=	$\hat{F}^*(y)$	$\hat{F}^*(y)$
$\hat{F}_H(x)$	=	$\hat{F}^*(x)$	$\hat{F}^*(x)$
\hat{X}_H	=	\hat{X}^*	\hat{X}
\hat{Z}_H	=	\hat{Z}^*	\hat{Z}
Variance/covariance			
Q_{rstu}	=	V_{rstu}^*	V_{rstu}
Q_{2000}	=	V_{2000}^*	V_{2000}
Q_{0200}	=	V_{0200}^*	V_{0200}
Q_{0020}	=	V_{0020}^*	V_{0020}
Q_{0002}	=	V_{0002}^*	V_{0002}
Q_{1100}	=	V_{1100}^*	V_{1100}
Q_{1010}	=	V_{1010}^*	V_{1010}
Q_{1001}	=	V_{1001}^*	V_{1001}
Q_{0110}	=	V_{0110}^*	V_{0110}
Q_{0101}	=	V_{0101}^*	V_{0101}
Coefficient of correlation			
ρ_{12}	=	Δ_{12}	$\nabla_{12(2)}$
ρ_{13}	=	Δ_{13}	$\nabla_{13(2)}$
ρ_{23}	=	Δ_{23}	$\nabla_{23(2)}$
ρ_{14}	=	Δ_{14}	$\nabla_{14(2)}$
ρ_{24}	=	Δ_{24}	$\nabla_{24(2)}$
Coefficient of multiple determination			
$\rho_{1.23}^2$	=	$\Delta_{1.23}^2$	$\nabla_{1.23}^2$
$\rho_{1.24}^2$	=	$\Delta_{1.24}^2$	$\nabla_{1.24}^2$

<https://doi.org/10.1371/journal.pone.0243584.t001>

$$\text{MSE}(\hat{F}_2(y)) \cong F^2(y)(Q_{2000} + Q_{0200} - 2Q_{1100}). \tag{7}$$

2. Murthy [29] adapted product estimator of $F(y)$ is

$$\hat{F}_3(y) = \hat{F}_H(y) \left(\frac{\hat{F}_H(x)}{F(x)} \right). \tag{8}$$

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

$$\text{Bias}(\hat{F}_3(y)) = F(y)Q_{1100},$$

$$\text{MSE}(\hat{F}_3(y)) \cong F^2(y)(Q_{2000} + Q_{0200} + 2Q_{1100}). \tag{9}$$

3. The adapted difference estimator of $F(y)$ is

$$\hat{F}_4(y) = \hat{F}_H(y) + k(F(x) - \hat{F}_H(x)), \tag{10}$$

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 $k_{(\text{opt})} = (F(y)Q_{1100})/(F(x)Q_{0200})$ is

$$\text{Var}_{\min}(\hat{F}_4(y)) = \frac{F^2(y)(Q_{2000}Q_{0200} - Q_{1100}^2)}{Q_{0200}}. \tag{11}$$

Here, (11) may be written as

$$\text{Var}_{\min}(\hat{F}_4(y)) = F^2(y)Q_{2000}(1 - \varrho_{12}^2). \tag{12}$$

4. Rao [30] adapted difference-type estimator of $F(y)$ is

$$\hat{F}_5(y) = k_1\hat{F}_H(y) + k_2(F(x) - \hat{F}_H(x)), \tag{13}$$

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

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

$$\begin{aligned} \text{MSE}(\hat{F}_5(y)) \cong & F^2(y) - 2k_1F^2(y) + k_1^2F^2(y) + k_1^2F^2(y)Q_{2000} \\ & - 2k_1k_2F(y)F(x)Q_{1100} + k_2^2F^2(x)Q_{0200}. \end{aligned} \tag{14}$$

The optimum values of k_1 and k_2 , determined by minimizing (??), are

$$k_{1(\text{opt})} = \frac{Q_{0200}}{(Q_{0200}Q_{2000} - Q_{1100}^2 + Q_{0200})},$$

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

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

$$MSE_{\min}(\hat{F}_5(y)) = \frac{F^2(y)(Q_{2000}Q_{0200} - Q_{1100}^2)}{(Q_{2000}Q_{0200} - Q_{1100}^2 + Q_{0200})}. \tag{15}$$

Here, (15) may be written as

$$MSE_{\min}(\hat{F}_5(y)) = \frac{F^2(y)Q_{2000}(1 - \rho_{12}^2)}{1 + Q_{2000}(1 - \rho_{12}^2)}. \tag{16}$$

5. Singh et al. [31] adapted generalized ratio-type exponential estimator of $F(y)$ is

$$\hat{F}_6(y) = \hat{F}_H(y) \exp\left(\frac{a(F(x) - \hat{F}_H(x))}{a(F(x) + \hat{F}_H(x)) + 2b}\right), \tag{17}$$

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

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

$$MSE(\hat{F}_6(y)) \cong \frac{F^2(y)}{4} (4Q_{2000} + \theta^2 Q_{0200} - 4\theta Q_{1100}), \tag{18}$$

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

6. Grover and Kaur [32] adapted generalized class of ratio-type exponential estimator of $F(y)$ is

$$\hat{F}_7(y) = \{k_3 \hat{F}_H(y) + k_4(F(x) - \hat{F}_H(x))\} \exp\left(\frac{a(F(x) - \hat{F}_H(x))}{a(F(x) + \hat{F}_H(x)) + 2b}\right), \tag{19}$$

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

$$\text{Bias}(\hat{F}_7(y)) \cong F(y)(k_3 - 1) + \frac{3}{8} \theta^2 k_3 F(y) + \frac{1}{2} \theta k_4 F(x) Q_{0200} - \frac{1}{2} \theta F(y) Q_{1100},$$

$$MSE(\hat{F}_7(y)) \cong k_1^2 F^2(x) Q_{0200} + k_3^2 F^2(y) Q_{2000} + 2\theta k_3 k_4 F(y) F(x) Q_{0200}$$

$$- 2k_3 k_4 F(y) F(x) Q_{1100} + F^2(y) - 2k_3 F^2(y) + \theta k_3^2 F^2(y)$$

$$+ k_3 F^2(y) Q_{1100} - \theta k_4 F(y) F(x) Q_{0200} - 2\theta k_3^2 F^2(y) Q_{1100}$$

$$- \frac{3}{4} \theta^2 k_3 F^2(y) Q_{0200} + \theta^2 k_3^2 F^2(y) Q_{0200}. \tag{20}$$

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

$$k_{3(\text{opt})} = \frac{Q_{0200}(\theta^2 Q_{0200} - 8)}{8(-Q_{2000} Q_{0200} + Q_{1100}^2 - Q_{0200})},$$

$$k_{4(\text{opt})} = \frac{F(y)(\theta^3 Q_{0200}^2 - \theta^2 Q_{0200} Q_{1100} + 4\theta Q_{2000} Q_{0200} - 4\theta Q_{1100}^2 - 4\theta Q_{0200} + 8Q_{1100})}{8F(x)(Q_{2000} Q_{0200} - Q_{1100}^2 + Q_{0200})}.$$

The simplified minimum MSE of $\hat{F}_7(y)$ at the optimum values of k_3 and k_4 is

$$\text{MSE}_{\min}(\hat{F}_7(y)) \cong \frac{F^2(y)}{64} \left(64 - 16\theta^2 Q_{0200} - \frac{Q_{0200}(-8 + \theta^2 Q_{0200})^2}{Q_{0200}(1 + Q_{2000}) - Q_{1100}^2} \right). \tag{21}$$

Here, (21) may be written as

$$\text{MSE}_{\min}(\hat{F}_7(y)) \cong \text{Var}_{\min}(\hat{F}_4(y)) - \frac{F^2(y)(\theta^2 Q_{0200}^2 - 8Q_{1100}^2 + 8Q_{0200} Q_{2000})^2}{64Q_{0200}^2 \{1 + Q_{2000}(1 - \rho_{12}^2)\}}, \tag{22}$$

which shows that $\hat{F}_7(y)$ is more precise than $\hat{F}_4(y)$.

4. Proposed estimators

The precision of an estimator surges by using the appropriate secondary information at the estimation stage. In previous studies, the sample distribution function of the auxiliary variable was used to expand the productivities of the prevailing distribution function estimators. In a recent study, Haq et al. [33] recommended that to use ranks of the auxiliary variable as an additional auxiliary variable to increase the precision of an estimator of the population mean. On the similar lines 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 persists, we recommend two families of estimators for estimating $F(y)$. The (first and second) families of estimators, used the (sample mean of the auxiliary variable and the sample mean of the ranked-auxiliary variable) as an additional auxiliary variable.

On the lines of $\hat{F}_5(y)$ and $\hat{F}_6(y)$, first proposed family of estimators for estimating $F(y)$ is given by

$$\hat{F}_8(y) = \left\{ k_5 \hat{F}_H(y) + k_6 \left(\frac{F(x) - \hat{F}_H(x)}{F(x)} \right) \left(\frac{\bar{X} - \hat{X}_H}{\bar{X}} \right) \right\} \exp \left(\frac{a(F(x) - \hat{F}_H(x))}{a(F(x) + \hat{F}_H(x)) + 2b} \right), \tag{23}$$

where k_5, k_6 and k_7 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}, β_2 (coefficient of kurtosis), C_2 , etc.

The estimator $\hat{F}_8(y)$ can also be written as

$$\hat{F}_8(y) = \{k_5 F(y)(1 + e_1) - k_6 e_2 - k_7 e_3\} \left(1 - \frac{1}{2} \theta e_2 + \frac{3}{8} \theta^2 e_2^2 + \dots \right). \tag{24}$$

Simplifying (24) and keeping terms only up to the second power of e_i 's, we can write

$$\begin{aligned}
 (\hat{F}_8(y) - F(y)) &= -F(y) + k_5 F(y) + k_5 F(y) e_1 - \frac{1}{2} \theta k_5 F(y) e_2 - k_6 e_2 - k_7 e_3 \\
 &+ \frac{3}{8} \theta^2 k_5 F(y) e_2^2 + \frac{1}{2} \theta k_6 e_2^2 - \frac{1}{2} \theta k_5 F(y) e_1 e_2 + \frac{1}{2} \theta k_7 e_2 e_3.
 \end{aligned}
 \tag{25}$$

The bias and MSE of $\hat{F}_8(y)$, to the first order of approximation, are

$$\text{Bias}(\hat{F}_8(y)) \cong F(y)(k_5 - 1) + \frac{3}{8} \theta^2 k_5 F(y) Q_{0200} + \frac{1}{2} \theta k_6 Q_{0200} - \frac{1}{2} \theta k_5 F(y) Q_{1100} + \frac{1}{2} \theta k_7 Q_{0110},$$

$$\begin{aligned}
 \text{MSE}(\hat{F}_8(y)) &\cong F^2(y)(k_5 - 1)^2 + k_5^2 F^2(y) Q_{2000} + k_6^2 Q_{0200} + k_7^2 Q_{0020} + \theta^2 k_5^2 F^2(y) Q_{0200} \\
 &- \theta k_6 F(y) Q_{0200} + 2\theta k_5 k_6 F(y) Q_{0200} - \frac{3}{4} \theta^2 k_5 F^2(y) Q_{0200} + \theta k_5 F^2(y) Q_{1100} \\
 &- 2\theta k_5^2 F^2(y) Q_{1100} - 2k_5 k_6 F(y) Q_{1100} - 2k_5 k_7 F(y) Q_{1010} - \theta k_7 F(y) Q_{0110} \\
 &+ 2\theta k_5 k_7 F(y) Q_{0110} - 2k_6 k_7 Q_{0110}.
 \end{aligned}
 \tag{26}$$

The optimum values of k_5 , k_6 and k_7 , determined by minimizing (26), are

$$k_{5(\text{opt})} = \frac{8 - \theta^2 Q_{0200}}{8\{1 + Q_{2000}(1 - \varrho_{1.23}^2)\}},$$

$$k_{6(\text{opt})} = \frac{F(y) \left[\begin{aligned} &\theta^3 Q_{0200}^{3/2} (\varrho_{23}^2 - 1) + Q_{2000}^{1/2} (-8 + \theta^2 Q_{0200}) (\varrho_{12} - \varrho_{23} \varrho_{13}) \\ &+ 4\theta Q_{0200}^{1/2} (\varrho_{23}^2 - 1) \{-1 + Q_{2000}(1 - \varrho_{1.23}^2)\} \end{aligned} \right]}{8Q_{0200}^{1/2} (\varrho_{23}^2 - 1) \{-1 + Q_{2000}(1 - \varrho_{1.23}^2)\}},$$

$$k_{7(\text{opt})} = \frac{F(y) Q_{2000}^{1/2} (8 - \theta^2 Q_{0200}) (\varrho_{12} - \varrho_{23} \varrho_{13})}{8Q_{0200}^{1/2} (\varrho_{23}^2 - 1) \{-1 + Q_{2000}(1 - \varrho_{1.23}^2)\}}.$$

The simplified minimum MSE of $\hat{F}_8(y)$, at the optimum values of k_5 , k_6 and k_7 is

$$\text{MSE}_{\min}(\hat{F}_8(y)) \cong \frac{F^2(y) \{64Q_{2000}(1 - \varrho_{1.23}^2) - \theta^4 Q_{0200}^2 - 16\theta^2 Q_{0200} Q_{2000}(1 - \varrho_{1.23}^2)\}}{64\{1 + Q_{2000}(1 - \varrho_{1.23}^2)\}}, \tag{27}$$

where $\varrho_{1.23}^2 = \left(\frac{Q_{1100}^2 Q_{0020} + Q_{1010}^2 Q_{0200} - 2Q_{1010} Q_{1100} Q_{0110}}{Q_{2000}(Q_{0200} Q_{0020} - Q_{0110}^2)} \right)$.

Here, (27) may be written as

$$\text{MSE}_{\min}(\hat{F}_8(y)) \cong \text{Var}_{\min}(\hat{F}_4(y)) - H_1 - H_2, \tag{28}$$

where

$$H_1 = \frac{F^2(y)(\theta^2 Q_{0200}^2 - 8Q_{1100}^2 + 8Q_{0200}Q_{2000})^2}{64Q_{0200}^2 \{1 + Q_{2000}(1 - \varrho_{12}^2)\}} \text{ and}$$

$$H_2 = \frac{F^2(y)(\theta^2 Q_{0200} - 8)^2 (Q_{0200}Q_{1010} - Q_{0110}Q_{1100})^2}{64Q_{0200}^2 Q_{0020}(1 - \varrho_{23}^2) \{1 + Q_{2000}(1 - \varrho_{12}^2)\} \{1 + Q_{2000}(1 - \varrho_{1,23}^2)\}}.$$

It is clear that $\hat{F}_8(y)$ is more precise than $\hat{F}_4(y)$.

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

$$\hat{F}_9(y) = \left\{ k_8 \hat{F}(y) + k_9 \left(\frac{F(x) - \hat{F}(x)}{F(x)} \right) + k_{10} \left(\frac{\bar{Z} - \hat{\bar{Z}}}{\bar{Z}} \right) \right\} \exp \left(\frac{a(F(x) - \hat{F}(x))}{a(F(x) - \hat{F}(x)) + 2b} \right), \quad (29)$$

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}, β_2 (coefficient of kurtosis), C_2 , etc.

The estimator $\hat{F}_9(y)$ can also be written as

$$\hat{F}_9(y) = \{k_8 F(y)(1 + e_1) - k_9 e_2 - k_{10} e_4\} \left(1 - \frac{1}{2} \theta e_2 + \frac{3}{8} \theta^2 e_2^2 + \dots \right). \quad (30)$$

Simplifying (30) and keeping terms only up to the second power of e_i 's, we can write

$$(\hat{F}_9(y) - F(y)) = -F(y) + k_8 F(y) + k_8 F(y) e_1 - \frac{1}{2} \theta k_8 F(y) e_2 - k_9 e_2 - k_{10} e_4$$

$$+ \frac{3}{8} \theta^2 k_8 F(y) e_2^2 + \frac{1}{2} \theta k_9 e_2^2 - \frac{1}{2} \theta k_8 F(y) e_1 e_2 + \frac{1}{2} \theta k_{10} e_2 e_4. \quad (31)$$

The bias and MSE of $\hat{F}_9(y)$, to the first order of approximation, are

$$\text{Bias}(\hat{F}_9(y)) \cong F(y)(k_8 - 1) + \frac{3}{8} \theta^2 k_8 F(y) Q_{0200} + \frac{1}{2} \theta k_9 Q_{0200} - \frac{1}{2} \theta k_8 F(y) Q_{1100} + \frac{1}{2} \theta k_9 Q_{0101},$$

$$\text{MSE}(\hat{F}_9(y)) \cong F^2(y)(k_8 - 1)^2 + k_8^2 F^2(y) Q_{2000} + k_9^2 Q_{0200} + k_{10}^2 Q_{0002} + \theta^2 k_8^2 F^2(y) Q_{0200}$$

$$- \theta k_9 F(y) Q_{0200} + 2\theta k_8 k_9 F(y) Q_{0200} - \frac{3}{4} \theta^2 k_8 F^2(y) Q_{0200} + \theta k_8 F^2(y) Q_{1100}$$

$$- 2\theta k_8^2 F^2(y) Q_{1100} - 2k_8 k_9 F(y) Q_{1100} - 2k_8 k_{10} F(y) Q_{1001} - \theta k_{10} F(y) Q_{0101}$$

$$+ 2\theta k_8 k_{10} F(y) Q_{0101} - 2k_9 k_{10} Q_{0101}. \quad (32)$$

The optimum values of k_8, k_9 and k_{10} , determined by minimizing (??), are

$$k_{8(\text{opt})} = \frac{8 - \theta^2 Q_{0200}}{8\{1 + Q_{2000}(1 - \varrho_{1.24}^2)\}},$$

$$k_{9(\text{opt})} = \frac{F(y) \left[\theta^3 Q_{0200}^{3/2} (\varrho_{24}^2 - 1) + Q_{2000}^{1/2} (-8 + \theta^2 Q_{0200}) (\varrho_{12} - R_{24} R_{14}) \right] + 4\theta Q_{0200}^{1/2} (\varrho_{24}^2 - 1) \{-1 + Q_{2000}(1 - \varrho_{1.24}^2)\}}{8Q_{0200}^{1/2} (\varrho_{24}^2 - 1) \{-1 + Q_{2000}(1 - \varrho_{1.24}^2)\}},$$

$$k_{10(\text{opt})} = \frac{F(y) Q_{2000}^{1/2} (8 - \theta^2 Q_{0200}) (\varrho_{12} - \varrho_{24} \varrho_{14})}{8Q_{0200}^{1/2} (\varrho_{24}^2 - 1) \{-1 + Q_{2000}(1 - \varrho_{1.24}^2)\}}.$$

The simplified minimum MSE of $\hat{F}_9(y)$ at the optimum values of k_8, k_9 and k_{10} is

$$\text{MSE}_{\min}(\hat{F}_9(y)) \cong \frac{F^2(y) \{64Q_{2000}(1 - \varrho_{1.24}^2) - \theta^4 Q_{0200}^2 - 16\theta Q_{0200} Q_{2000}(1 - \varrho_{1.24}^2)\}}{64\{1 + Q_{2000}(1 - \varrho_{1.24}^2)\}}, \tag{33}$$

where $\varrho_{1.24}^2 = \left(\frac{Q_{1100}^2 Q_{0002} + Q_{1001}^2 Q_{0200} - 2Q_{1001} Q_{1100} Q_{0101}}{Q_{2000}(Q_{0200} Q_{0002} - Q_{0101}^2)} \right)$.

Here, (33) may be written as

$$\text{MSE}_{\min}(\hat{F}_9(y)) \cong \text{Var}_{\min}(\hat{F}_4(y)) - H_1 - H_3, \tag{34}$$

where

$$H_1 = \frac{F^2(y) (\theta^2 Q_{0200}^2 - 8Q_{1100}^2 + 8Q_{0200})^2}{64Q_{0200}^2 \{1 + Q_{2000}(1 - R_{12}^2)\}} \text{ and}$$

$$H_3 = \frac{F^2(y) (\theta^2 Q_{0200} - 8)^2 (Q_{0200} Q_{1001} - Q_{0101} Q_{1100})^2}{64Q_{0200}^2 Q_{0002} (1 - R_{24}^2) \{1 + Q_{2000}(1 - R_{12}^2)\} \{1 + Q_{2000}(1 - R_{1.24}^2)\}}.$$

It is clear that $\hat{F}_9(y)$ is more precise than $\hat{F}_4(y)$.

In Table 2, we put some members of the Singh et al. [31], Grover and Kaur [32], and proposed families of estimators with selected choices of a and b .

Table 2. Some members of the adapted and proposed distribution function estimators.

a	b	$\hat{F}_6(y)$	$\hat{F}_7(y)$	$\hat{F}_8(y)$	$\hat{F}_9(y)$
1	C_2	$\hat{F}_6^{(1)}(y)$	$\hat{F}_7^{(1)}(y)$	$\hat{F}_8^{(1)}(y)$	$\hat{F}_9^{(1)}(y)$
1	β_2	$\hat{F}_6^{(2)}(y)$	$\hat{F}_7^{(2)}(y)$	$\hat{F}_8^{(2)}(y)$	$\hat{F}_9^{(2)}(y)$
β_2	C_2	$\hat{F}_6^{(3)}(y)$	$\hat{F}_7^{(3)}(y)$	$\hat{F}_8^{(3)}(y)$	$\hat{F}_9^{(3)}(y)$
C_2	β_2	$\hat{F}_6^{(4)}(y)$	$\hat{F}_7^{(4)}(y)$	$\hat{F}_8^{(4)}(y)$	$\hat{F}_9^{(4)}(y)$
1	12	$\hat{F}_6^{(5)}(y)$	$\hat{F}_7^{(5)}(y)$	$\hat{F}_8^{(5)}(y)$	$\hat{F}_9^{(5)}(y)$
C_2	R_{12}	$\hat{F}_6^{(6)}(y)$	$\hat{F}_7^{(6)}(y)$	$\hat{F}_8^{(6)}(y)$	$\hat{F}_9^{(6)}(y)$
R_{12}	C_2	$\hat{F}_6^{(7)}(y)$	$\hat{F}_7^{(7)}(y)$	$\hat{F}_8^{(7)}(y)$	$\hat{F}_9^{(7)}(y)$
β_2	R_{12}	$\hat{F}_6^{(8)}(y)$	$\hat{F}_7^{(8)}(y)$	$\hat{F}_8^{(8)}(y)$	$\hat{F}_9^{(8)}(y)$
R_{12}	β_2	$\hat{F}_6^{(9)}(y)$	$\hat{F}_7^{(9)}(y)$	$\hat{F}_8^{(9)}(y)$	$\hat{F}_9^{(9)}(y)$
1	$NF(x)$	$\hat{F}_6^{(10)}(y)$	$\hat{F}_7^{(10)}(y)$	$\hat{F}_8^{(10)}(y)$	$\hat{F}_9^{(10)}(y)$

<https://doi.org/10.1371/journal.pone.0243584.t002>

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 (5) and (28),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{Var}(\hat{F}_1(y)) \text{ if}$$

$$F^2(y)Q_{2000}\varrho_{12}^2 + H_1 + H_2 > 0.$$

2. From (7) and (28),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}(\hat{F}_2(y)) \text{ if}$$

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

3. From (9) and (28),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}(\hat{F}_3(y)) \text{ if}$$

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

4. From (12) and (28),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}_{\min}(\hat{F}_4(y)) \text{ if}$$

$$H_1 + H_2 > 0.$$

5. From (16) and (28),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}_{\min}(\hat{F}_5(y)) \text{ if}$$

$$\frac{F^2(y)\theta^2 Q_{0200} \{ \theta^2 Q_{0200} + 16Q_{2000}(1 - \varrho_{12}^2) \}}{64\{1 + Q_{2000}(1 - \varrho_{12}^2)\}} + H_2 > 0.$$

6. From (18) and (28),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}(\hat{F}_6(y)) \text{ if}$$

$$\frac{F^2(y)}{Q_{2000}} \left(\frac{\theta Q_{0200}}{2} - Q_{1100} \right)^2 + H_1 + H_2 > 0.$$

7. From (22) and (28),

$$\begin{aligned} \text{MSE}_{\min}(\hat{F}_8(y)) &< \text{MSE}_{\min}(\hat{F}_7(y)) \text{ if} \\ H_2 &> 0. \end{aligned}$$

8. From (5) and (34),

$$\begin{aligned} \text{MSE}_{\min}(\hat{F}_9(y)) &< \text{Var}(\hat{F}_1(y)) \text{ if} \\ \frac{1}{Q_{0200}} F^2(y) Q_{1100}^2 + H_1 + H_3 &> 0. \end{aligned}$$

9. From (7) and (34),

$$\begin{aligned} \text{MSE}_{\min}(\hat{F}_9(y)) &< \text{MSE}(\hat{F}_2(y)) \text{ if} \\ \frac{F^2(y)}{Q_{2000}} (Q_{0200} - Q_{1100})^2 + H_1 + H_3 &> 0. \end{aligned}$$

10. From (9) and (34),

$$\begin{aligned} \text{MSE}_{\min}(\hat{F}_9(y)) &< \text{MSE}(\hat{F}_3(y)) \text{ if} \\ \frac{F^2(y)}{Q_{2000}} (Q_{0200} + Q_{1100})^2 + H_1 + H_3 &> 0. \end{aligned}$$

11. From (12) and (34),

$$\begin{aligned} \text{MSE}_{\min}(\hat{F}_9(y)) &< \text{MSE}_{\min}(\hat{F}_4(y)) \text{ if} \\ H_1 + H_3 &> 0. \end{aligned}$$

12. From (16) and (34),

$$\begin{aligned} \text{MSE}_{\min}(\hat{F}_9(y)) &< \text{MSE}_{\min}(\hat{F}_5(y)) \text{ if} \\ \frac{F^2(y) \theta^2 Q_{0200} \{ \theta^2 Q_{0200} + 16 Q_{2000} (1 - \varrho_{12}^2) \}}{64 \{ 1 + Q_{2000} (1 - R_{12}^2) \}} + H_3 &> 0. \end{aligned}$$

13. From (18) and (34),

$$\text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}_{\min}(\hat{F}_6(y)) \text{ if}$$

$$\frac{F^2(y)}{Q_{2000}} \left(\frac{\theta Q_{0200}}{2} - Q_{1100} \right)^2 + H_1 + H_3 > 0.$$

14. From (22) and (34),

$$\begin{aligned} \text{MSE}_{\min}(\hat{F}_9(y)) &< \text{MSE}_{\min}(\hat{F}_7(y)) \text{ if} \\ H_3 &> 0. \end{aligned}$$

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 this section, we conduct a numerical study to see the performance of the adapted and proposed distribution function estimators. For this purpose, five population are considered. The summary statistics of these populations are reported in Tables 3–6. The percentage relative efficiency (PRE) of an estimator $\hat{F}_i(y)$ with respect to $\hat{F}_H(y)$ is

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

where $i = 2, 3, \dots, 9$.

The PREs of distribution function estimators, computed from five populations, are given in Tables 7–14.

Population I -Gujarati [34]

Table 3. Summary statistics for population I.

Parameter	Value	Parameter	Value
N	50	S_2	0.50508
n	15	S_3	21805
λ	0.04667	S_4	15756
\bar{X}	75.8720	R_{12}	-0.20000
\bar{R}	25.5000	R_{13}	0.25176
$F(y)$	0.50000	R_{23}	-0.79517
$F(x)$	0.50000	R_{14}	0.21345
S_1	0.50508	R_{24}	-0.86628
Non-response			
Parameter	Value	Parameter	Value
N_2	12	$S_{4(2)}$	3.60555
w_2	0.24000	$R_{12(2)}$	-0.50709
λ_2	0.01600	$R_{13(2)}$	0.35848
$S_{1(2)}$	0.51493	$R_{23(2)}$	-0.82899
$S_{2(2)}$	0.52223	$R_{14(2)}$	0.36724
$S_{3(2)}$	19.5392	$R_{24(2)}$	-0.86905

<https://doi.org/10.1371/journal.pone.0243584.t003>

Table 4. Summary statistics for population II.

Parameter	Value	Parameter	Value
N	50	S_2	0.50508
n	15	S_3	21.3175
λ	0.04667	S_4	15756
\bar{X}	78.2900	R_{12}	-0.12000
\bar{R}	25.5000	R_{13}	0.22925
$F(y)$	0.50000	R_{23}	-0.78936
$F(x)$	0.50000	R_{14}	0.18435
S_1	0.50508	R_{24}	-0.86630
Non-response			
Parameter	Value	Parameter	Value
N_2	12	$S_{4(2)}$	3.60555
w_2	0.24000	$R_{12(2)}$	-0.16903
λ_2	0.01600	$R_{13(2)}$	0.25695
$S_{1(2)}$	0.51493	$R_{23(2)}$	-0.81369
$S_{2(2)}$	0.52223	$R_{14(2)}$	0.22034
$S_{3(2)}$	18.2593	$R_{24(2)}$	-0.86905

<https://doi.org/10.1371/journal.pone.0243584.t004>

Y: The eggs produced in 1990 (millions) and

X: The price per dozen (cents) in 1991. The proportion of the non-response units in the given population is considered to be the last 25% units.

Population II -Gujarati [34]

Y: The eggs produced in 1990 (millions) and

X: The price per dozen (cents) in 1990. The proportion of the non-response units in the given population is considered to be the last 25% units.

Table 5. Summary statistics for population III.

Parameter	Value	Parameter	Value
N	30	S_2	0.50855
n	8	S_3	9.23238
λ	0.09167	S_4	8.79557
\bar{X}	67.2667	R_{12}	-0.73333
\bar{R}	15.5000	R_{13}	0.71975
$F(y)$	0.50000	R_{23}	-0.83726
$F(x)$	0.50000	R_{14}	0.73622
S_1	0.50855	R_{24}	-0.86728
Non-response			
Parameter	Value	Parameter	Value
N_2	8	$S_{4(2)}$	44949
w_2	0.26667	$R_{12(2)}$	-0.60000
λ_2	0.03333	$R_{13(2)}$	0.36387
$S_{1(2)}$	0.51755	$R_{23(2)}$	-0.81227
$S_{2(2)}$	0.51755	$R_{14(2)}$	0.28172
$S_{3(2)}$	9.38749	$R_{24(2)}$	-0.84515

<https://doi.org/10.1371/journal.pone.0243584.t005>

Table 6. Summary statistics for population IV.

Parameter	Value	Parameter	Value
N	69	S_2	0.50361
n	20	S_3	7058.99
λ	0.03550	S_{41}^2	20.0602
\bar{X}	49544	R_{12}	0.94202
\bar{R}	35.0000	R_{13}	-0.57090
$F(y)$	0.50725	R_{23}	-0.57274
$F(x)$	0.50725	R_{14}	-0.85584
S_1	0.50361	R_{24}	-0.86603
Non-response			
Parameter	Value	Parameter	Value
N_2	17	$S_{4(2)}$	5.04975
W_2	0.24638	$R_{12(2)}$	0.88273
λ_2	0.01231	$R_{13(2)}$	-0.62943
$S_{1(2)}$	0.51449	$R_{23(2)}$	-0.57830
$S_{3(2)}$	5920.86	$R_{24(2)}$	-0.85391

<https://doi.org/10.1371/journal.pone.0243584.t006>

Table 7. PREs of distribution function estimators under situation-I using population I.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	89.193	$\hat{F}_7^{(1)}(y)$	115.27	$\hat{F}_8^{(1)}(y)$	116.13	$\hat{F}_9^{(1)}(y)$	115.36
$\hat{F}_2(y)$	38.87	$\hat{F}_6^{(2)}(y)$	89.11	$\hat{F}_7^{(2)}(y)$	115.28	$\hat{F}_8^{(2)}(y)$	116.13	$\hat{F}_9^{(2)}(y)$	115.36
$\hat{F}_3(y)$	69.33	$\hat{F}_6^{(3)}(y)$	89.19	$\hat{F}_7^{(3)}(y)$	115.27	$\hat{F}_8^{(3)}(y)$	116.13	$\hat{F}_9^{(3)}(y)$	115.36
$\hat{F}_4(y)$	108.61	$\hat{F}_6^{(4)}(y)$	89.03	$\hat{F}_7^{(4)}(y)$	115.28	$\hat{F}_8^{(4)}(y)$	116.14	$\hat{F}_9^{(4)}(y)$	115.37
$\hat{F}_5(y)$	115.07	$\hat{F}_6^{(5)}(y)$	46.07	$\hat{F}_7^{(5)}(y)$	121.61	$\hat{F}_8^{(5)}(y)$	1252	$\hat{F}_9^{(5)}(y)$	121.70
		$\hat{F}_6^{(6)}(y)$	46.33	$\hat{F}_7^{(6)}(y)$	121.5	$\hat{F}_8^{(6)}(y)$	1241	$\hat{F}_9^{(6)}(y)$	121.59
		$\hat{F}_6^{(7)}(y)$	1088	$\hat{F}_7^{(7)}(y)$	115.09	$\hat{F}_8^{(7)}(y)$	115.95	$\hat{F}_9^{(7)}(y)$	115.17
		$\hat{F}_6^{(8)}(y)$	46.07	$\hat{F}_7^{(8)}(y)$	121.61	$\hat{F}_8^{(8)}(y)$	1252	$\hat{F}_9^{(8)}(y)$	121.70
		$\hat{F}_6^{(9)}(y)$	1091	$\hat{F}_7^{(9)}(y)$	115.09	$\hat{F}_8^{(9)}(y)$	115.95	$\hat{F}_9^{(9)}(y)$	115.18
		$\hat{F}_6^{(10)}(y)$	99.44	$\hat{F}_7^{(10)}(y)$	115.07	$\hat{F}_8^{(10)}(y)$	115.92	$\hat{F}_9^{(10)}(y)$	115.15

<https://doi.org/10.1371/journal.pone.0243584.t007>

Table 8. PREs of distribution function estimators under situation-II using population I.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	89.19	$\hat{F}_7^{(1)}(y)$	115.27	$\hat{F}_8^{(1)}(y)$	117.91	$\hat{F}_9^{(1)}(y)$	115.42
$\hat{F}_2(y)$	38.87	$\hat{F}_6^{(2)}(y)$	89.11	$\hat{F}_7^{(2)}(y)$	115.28	$\hat{F}_8^{(2)}(y)$	117.91	$\hat{F}_9^{(2)}(y)$	115.43
$\hat{F}_3(y)$	69.33	$\hat{F}_6^{(3)}(y)$	89.19	$\hat{F}_7^{(3)}(y)$	115.27	$\hat{F}_8^{(3)}(y)$	117.91	$\hat{F}_9^{(3)}(y)$	115.42
$\hat{F}_4(y)$	108.61	$\hat{F}_6^{(4)}(y)$	89.03	$\hat{F}_7^{(4)}(y)$	115.28	$\hat{F}_8^{(4)}(y)$	117.92	$\hat{F}_9^{(4)}(y)$	115.43
$\hat{F}_5(y)$	115.07	$\hat{F}_6^{(5)}(y)$	46.07	$\hat{F}_7^{(5)}(y)$	121.61	$\hat{F}_8^{(5)}(y)$	1241	$\hat{F}_9^{(5)}(y)$	121.76
		$\hat{F}_6^{(6)}(y)$	46.33	$\hat{F}_7^{(6)}(y)$	121.50	$\hat{F}_8^{(6)}(y)$	1231	$\hat{F}_9^{(6)}(y)$	121.66
		$\hat{F}_6^{(7)}(y)$	1088	$\hat{F}_7^{(7)}(y)$	115.09	$\hat{F}_8^{(7)}(y)$	117.72	$\hat{F}_9^{(7)}(y)$	115.24
		$\hat{F}_6^{(8)}(y)$	46.07	$\hat{F}_7^{(8)}(y)$	121.61	$\hat{F}_8^{(8)}(y)$	1241	$\hat{F}_9^{(8)}(y)$	121.76
		$\hat{F}_6^{(9)}(y)$	1091	$\hat{F}_7^{(9)}(y)$	115.09	$\hat{F}_8^{(9)}(y)$	117.72	$\hat{F}_9^{(9)}(y)$	115.24
		$\hat{F}_6^{(10)}(y)$	99.44	$\hat{F}_7^{(10)}(y)$	115.07	$\hat{F}_8^{(10)}(y)$	117.70	$\hat{F}_9^{(10)}(y)$	115.22

<https://doi.org/10.1371/journal.pone.0243584.t008>

Table 9. PREs of distribution function estimators under situation-I using population II.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	93.54	$\hat{F}_7^{(1)}(y)$	109.64	$\hat{F}_8^{(1)}(y)$	1112	$\hat{F}_9^{(1)}(y)$	111.94
$\hat{F}_2(y)$	49.21	$\hat{F}_6^{(2)}(y)$	93.49	$\hat{F}_7^{(2)}(y)$	109.64	$\hat{F}_8^{(2)}(y)$	1112	$\hat{F}_9^{(2)}(y)$	111.94
$\hat{F}_3(y)$	69.33	$\hat{F}_6^{(3)}(y)$	93.54	$\hat{F}_7^{(3)}(y)$	109.64	$\hat{F}_8^{(3)}(y)$	1112	$\hat{F}_9^{(3)}(y)$	111.94
$\hat{F}_4(y)$	103.04	$\hat{F}_6^{(4)}(y)$	93.44	$\hat{F}_7^{(4)}(y)$	109.65	$\hat{F}_8^{(4)}(y)$	1112	$\hat{F}_9^{(4)}(y)$	111.94
$\hat{F}_5(y)$	109.5	$\hat{F}_6^{(5)}(y)$	56.89	$\hat{F}_7^{(5)}(y)$	113.76	$\hat{F}_8^{(5)}(y)$	118.43	$\hat{F}_9^{(5)}(y)$	116.15
		$\hat{F}_6^{(6)}(y)$	57.17	$\hat{F}_7^{(6)}(y)$	113.69	$\hat{F}_8^{(6)}(y)$	118.36	$\hat{F}_9^{(6)}(y)$	116.08
		$\hat{F}_6^{(7)}(y)$	101.42	$\hat{F}_7^{(7)}(y)$	109.51	$\hat{F}_8^{(7)}(y)$	113.99	$\hat{F}_9^{(7)}(y)$	111.81
		$\hat{F}_6^{(8)}(y)$	56.89	$\hat{F}_7^{(8)}(y)$	113.76	$\hat{F}_8^{(8)}(y)$	118.43	$\hat{F}_9^{(8)}(y)$	116.15
		$\hat{F}_6^{(9)}(y)$	101.43	$\hat{F}_7^{(9)}(y)$	109.51	$\hat{F}_8^{(9)}(y)$	113.99	$\hat{F}_9^{(9)}(y)$	111.81
		$\hat{F}_6^{(10)}(y)$	99.70	$\hat{F}_7^{(10)}(y)$	109.50	$\hat{F}_8^{(10)}(y)$	113.97	$\hat{F}_9^{(10)}(y)$	111.79

<https://doi.org/10.1371/journal.pone.0243584.t009>

Table 10. PREs of distribution function estimators under situation-II using population II.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	95.28	$\hat{F}_7^{(1)}(y)$	107.67	$\hat{F}_8^{(1)}(y)$	111.43	$\hat{F}_9^{(1)}(y)$	109.67
$\hat{F}_2(y)$	524	$\hat{F}_6^{(2)}(y)$	95.24	$\hat{F}_7^{(2)}(y)$	107.67	$\hat{F}_8^{(2)}(y)$	111.43	$\hat{F}_9^{(2)}(y)$	109.68
$\hat{F}_3(y)$	609	$\hat{F}_6^{(3)}(y)$	95.28	$\hat{F}_7^{(3)}(y)$	107.67	$\hat{F}_8^{(3)}(y)$	111.43	$\hat{F}_9^{(3)}(y)$	109.67
$\hat{F}_4(y)$	101.07	$\hat{F}_6^{(4)}(y)$	95.20	$\hat{F}_7^{(4)}(y)$	107.68	$\hat{F}_8^{(4)}(y)$	111.43	$\hat{F}_9^{(4)}(y)$	109.67
$\hat{F}_5(y)$	107.53	$\hat{F}_6^{(5)}(y)$	69.66	$\hat{F}_7^{(5)}(y)$	109.98	$\hat{F}_8^{(5)}(y)$	113.82	$\hat{F}_9^{(5)}(y)$	1102
		$\hat{F}_6^{(6)}(y)$	69.78	$\hat{F}_7^{(6)}(y)$	109.96	$\hat{F}_8^{(6)}(y)$	113.81	$\hat{F}_9^{(6)}(y)$	1100
		$\hat{F}_6^{(7)}(y)$	100.49	$\hat{F}_7^{(7)}(y)$	107.54	$\hat{F}_8^{(7)}(y)$	111.29	$\hat{F}_9^{(7)}(y)$	109.53
		$\hat{F}_6^{(8)}(y)$	69.66	$\hat{F}_7^{(8)}(y)$	109.98	$\hat{F}_8^{(8)}(y)$	113.82	$\hat{F}_9^{(8)}(y)$	1102
		$\hat{F}_6^{(9)}(y)$	100.49	$\hat{F}_7^{(9)}(y)$	107.54	$\hat{F}_8^{(9)}(y)$	111.29	$\hat{F}_9^{(9)}(y)$	109.53
		$\hat{F}_6^{(10)}(y)$	99.82	$\hat{F}_7^{(10)}(y)$	107.53	$\hat{F}_8^{(10)}(y)$	111.28	$\hat{F}_9^{(10)}(y)$	109.52

<https://doi.org/10.1371/journal.pone.0243584.t010>

Table 11. PREs of distribution function estimators under situation-I using population III.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	79.57	$\hat{F}_7^{(1)}(y)$	208.21	$\hat{F}_8^{(1)}(y)$	210.57	$\hat{F}_9^{(1)}(y)$	216.06
$\hat{F}_2(y)$	29.47	$\hat{F}_6^{(2)}(y)$	79.36	$\hat{F}_7^{(2)}(y)$	208.22	$\hat{F}_8^{(2)}(y)$	210.59	$\hat{F}_9^{(2)}(y)$	216.07
$\hat{F}_3(y)$	1694	$\hat{F}_6^{(3)}(y)$	79.57	$\hat{F}_7^{(3)}(y)$	208.21	$\hat{F}_8^{(3)}(y)$	210.57	$\hat{F}_9^{(3)}(y)$	216.06
$\hat{F}_4(y)$	1940	$\hat{F}_6^{(4)}(y)$	79.16	$\hat{F}_7^{(4)}(y)$	208.24	$\hat{F}_8^{(4)}(y)$	210.61	$\hat{F}_9^{(4)}(y)$	216.09
$\hat{F}_5(y)$	207.46	$\hat{F}_6^{(5)}(y)$	152.74	$\hat{F}_7^{(5)}(y)$	270.64	$\hat{F}_8^{(5)}(y)$	2709	$\hat{F}_9^{(5)}(y)$	282.08
		$\hat{F}_6^{(6)}(y)$	1424	$\hat{F}_7^{(6)}(y)$	2855	$\hat{F}_8^{(6)}(y)$	288.28	$\hat{F}_9^{(6)}(y)$	296.98
		$\hat{F}_6^{(7)}(y)$	145.64	$\hat{F}_7^{(7)}(y)$	209.72	$\hat{F}_8^{(7)}(y)$	2110	$\hat{F}_9^{(7)}(y)$	217.63
		$\hat{F}_6^{(8)}(y)$	152.74	$\hat{F}_7^{(8)}(y)$	270.64	$\hat{F}_8^{(8)}(y)$	2709	$\hat{F}_9^{(8)}(y)$	282.08
		$\hat{F}_6^{(9)}(y)$	146.98	$\hat{F}_7^{(9)}(y)$	209.85	$\hat{F}_8^{(9)}(y)$	2124	$\hat{F}_9^{(9)}(y)$	217.76
		$\hat{F}_6^{(10)}(y)$	97.78	$\hat{F}_7^{(10)}(y)$	207.47	$\hat{F}_8^{(10)}(y)$	209.83	$\hat{F}_9^{(10)}(y)$	215.29

<https://doi.org/10.1371/journal.pone.0243584.t011>

Table 12. PREs of distribution function estimators under situation-II using population III.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	83.66	$\hat{F}_7^{(1)}(y)$	177.62	$\hat{F}_8^{(1)}(y)$	185.30	$\hat{F}_9^{(1)}(y)$	185.98
$\hat{F}_2(y)$	35.82	$\hat{F}_6^{(2)}(y)$	83.49	$\hat{F}_7^{(2)}(y)$	177.63	$\hat{F}_8^{(2)}(y)$	185.32	$\hat{F}_9^{(2)}(y)$	185.99
$\hat{F}_3(y)$	151.28	$\hat{F}_6^{(3)}(y)$	83.66	$\hat{F}_7^{(3)}(y)$	177.62	$\hat{F}_8^{(3)}(y)$	185.30	$\hat{F}_9^{(3)}(y)$	185.98
$\hat{F}_4(y)$	1611	$\hat{F}_6^{(4)}(y)$	83.32	$\hat{F}_7^{(4)}(y)$	177.65	$\hat{F}_8^{(4)}(y)$	185.33	$\hat{F}_9^{(4)}(y)$	186.00
$\hat{F}_5(y)$	177.16	$\hat{F}_6^{(5)}(y)$	1443	$\hat{F}_7^{(5)}(y)$	207.48	$\hat{F}_8^{(5)}(y)$	216.89	$\hat{F}_9^{(5)}(y)$	217.71
		$\hat{F}_6^{(6)}(y)$	138.06	$\hat{F}_7^{(6)}(y)$	2184	$\hat{F}_8^{(6)}(y)$	2262	$\hat{F}_9^{(6)}(y)$	223.48
		$\hat{F}_6^{(7)}(y)$	132.03	$\hat{F}_7^{(7)}(y)$	178.54	$\hat{F}_8^{(7)}(y)$	186.26	$\hat{F}_9^{(7)}(y)$	186.94
		$\hat{F}_6^{(8)}(y)$	1443	$\hat{F}_7^{(8)}(y)$	207.48	$\hat{F}_8^{(8)}(y)$	216.89	$\hat{F}_9^{(8)}(y)$	217.71
		$\hat{F}_6^{(9)}(y)$	132.90	$\hat{F}_7^{(9)}(y)$	178.62	$\hat{F}_8^{(9)}(y)$	186.34	$\hat{F}_9^{(9)}(y)$	187.02
		$\hat{F}_6^{(10)}(y)$	98.29	$\hat{F}_7^{(10)}(y)$	177.17	$\hat{F}_8^{(10)}(y)$	1883	$\hat{F}_9^{(10)}(y)$	185.50

<https://doi.org/10.1371/journal.pone.0243584.t012>

Table 13. PREs of distribution function estimators under situation-I using population IV.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	140.20	$\hat{F}_7^{(1)}(y)$	715.05	$\hat{F}_8^{(1)}(y)$	736.10	$\hat{F}_9^{(1)}(y)$	755.46
$\hat{F}_2(y)$	678.05	$\hat{F}_6^{(2)}(y)$	139.93	$\hat{F}_7^{(2)}(y)$	715.03	$\hat{F}_8^{(2)}(y)$	736.09	$\hat{F}_9^{(2)}(y)$	755.44
$\hat{F}_3(y)$	25.76	$\hat{F}_6^{(3)}(y)$	140.23	$\hat{F}_7^{(3)}(y)$	715.05	$\hat{F}_8^{(3)}(y)$	736.10	$\hat{F}_9^{(3)}(y)$	755.46
$\hat{F}_4(y)$	709.37	$\hat{F}_6^{(4)}(y)$	139.69	$\hat{F}_7^{(4)}(y)$	715.02	$\hat{F}_8^{(4)}(y)$	736.08	$\hat{F}_9^{(4)}(y)$	755.43
$\hat{F}_5(y)$	7103	$\hat{F}_6^{(5)}(y)$	141.99	$\hat{F}_7^{(5)}(y)$	715.12	$\hat{F}_8^{(5)}(y)$	736.18	$\hat{F}_9^{(5)}(y)$	755.54
		$\hat{F}_6^{(6)}(y)$	141.75	$\hat{F}_7^{(6)}(y)$	715.11	$\hat{F}_8^{(6)}(y)$	736.17	$\hat{F}_9^{(6)}(y)$	755.53
		$\hat{F}_6^{(7)}(y)$	138.24	$\hat{F}_7^{(7)}(y)$	7196	$\hat{F}_8^{(7)}(y)$	736.02	$\hat{F}_9^{(7)}(y)$	755.37
		$\hat{F}_6^{(8)}(y)$	1402	$\hat{F}_7^{(8)}(y)$	715.12	$\hat{F}_8^{(8)}(y)$	736.18	$\hat{F}_9^{(8)}(y)$	755.54
		$\hat{F}_6^{(9)}(y)$	137.98	$\hat{F}_7^{(9)}(y)$	7195	$\hat{F}_8^{(9)}(y)$	736.00	$\hat{F}_9^{(9)}(y)$	755.36
		$\hat{F}_6^{(10)}(y)$	101.35	$\hat{F}_7^{(10)}(y)$	7103	$\hat{F}_8^{(10)}(y)$	735.06	$\hat{F}_9^{(10)}(y)$	7538

<https://doi.org/10.1371/journal.pone.0243584.t013>

Table 14. PREs of distribution function estimators under situation-II using population IV.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	127.83	$\hat{F}_7^{(1)}(y)$	3060	$\hat{F}_8^{(1)}(y)$	305.59	$\hat{F}_9^{(1)}(y)$	308.98
$\hat{F}_2(y)$	297.38	$\hat{F}_6^{(2)}(y)$	127.66	$\hat{F}_7^{(2)}(y)$	3059	$\hat{F}_8^{(2)}(y)$	305.58	$\hat{F}_9^{(2)}(y)$	308.98
$\hat{F}_3(y)$	31.59	$\hat{F}_6^{(3)}(y)$	127.85	$\hat{F}_7^{(3)}(y)$	3059	$\hat{F}_8^{(3)}(y)$	305.59	$\hat{F}_9^{(3)}(y)$	308.98
$\hat{F}_4(y)$	299.62	$\hat{F}_6^{(4)}(y)$	127.50	$\hat{F}_7^{(4)}(y)$	3059	$\hat{F}_8^{(4)}(y)$	305.58	$\hat{F}_9^{(4)}(y)$	308.98
$\hat{F}_5(y)$	3029	$\hat{F}_6^{(5)}(y)$	128.96	$\hat{F}_7^{(5)}(y)$	3062	$\hat{F}_8^{(5)}(y)$	305.60	$\hat{F}_9^{(5)}(y)$	309.00
		$\hat{F}_6^{(6)}(y)$	128.80	$\hat{F}_7^{(6)}(y)$	3062	$\hat{F}_8^{(6)}(y)$	305.60	$\hat{F}_9^{(6)}(y)$	309.00
		$\hat{F}_6^{(7)}(y)$	126.58	$\hat{F}_7^{(7)}(y)$	3057	$\hat{F}_8^{(7)}(y)$	305.56	$\hat{F}_9^{(7)}(y)$	308.96
		$\hat{F}_6^{(8)}(y)$	128.98	$\hat{F}_7^{(8)}(y)$	3062	$\hat{F}_8^{(8)}(y)$	305.61	$\hat{F}_9^{(8)}(y)$	309.00
		$\hat{F}_6^{(9)}(y)$	126.41	$\hat{F}_7^{(9)}(y)$	3057	$\hat{F}_8^{(9)}(y)$	305.56	$\hat{F}_9^{(9)}(y)$	308.96
		$\hat{F}_6^{(10)}(y)$	101.02	$\hat{F}_7^{(10)}(y)$	3029	$\hat{F}_8^{(10)}(y)$	305.28	$\hat{F}_9^{(10)}(y)$	308.67

<https://doi.org/10.1371/journal.pone.0243584.t014>

Population III-Singh [35]

Y: Duration of sleep of persons with age more than 50 years and

X: The age of persons in years. The proportion of the non-response units in the given population is considered to be the last 25% units.

Population IV-Singh [35]

Y: The estimated number of fish caught by marine recreational fisherman in year 1995 and

X: The estimated number of fish caught by marine recreational fisherman in year 199 The share of the non-response units in the given population is the last 25% units.

As mention above, we used four data sets for numerical illustration. We also considered different sample sizes from these populations. We then used simple random sampling, and the MSEs (minimum) of the proposed families of estimators pointed out in Eqs 28 and 21. The proposed families of estimators and the adapted estimators were compared between each other with respect to their PREs values. The result of PREs are presented in Tables 7–14, for both Situations under non-response. In Tables 3–6, we see the summary statistics about the populations. We can also observed from numerical results shown in Tables 7–14 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 Hansen and Hurwitz [1], Cochran [28], Murthy [29], Rao [30], Singh et al. [31] and Grover and Kaur [32] in terms of the PRE. It can be seen that, for some data sets the second proposed family of the estimators perform better than the first proposed family of estimators for both situations of non-response, and for data sets, it is also observed that the first proposed family of the estimators performs better than the second proposed family of estimators under both situations of non-response.

7. Conclusion

In this paper, we have proposed two new families of estimators for estimating the finite population distribution function in the presence of non-response under simple random sampling. The proposed estimators required information on the sample distribution function of study variable, and the distribution function of auxiliary variable with additional information on either sample mean or ranks of the auxiliary variable. The bias and MSE of the proposed families of estimators were derived under the first order of approximation. Based on the theoretical and numerical (four populations considered in the numerical investigation) comparisons, it has turned out that the proposed families of estimators are more precise than their adaptive estimators under two situations of non-response. Thus, we acclaim categorically, the use of our proposed families of estimators over the existing estimators deliberate in this paper for the new survey for estimating the finite population distribution function in the presence of non-response.

Author Contributions

Conceptualization: Sardar Hussain.

Data curation: Sardar Hussain.

Formal analysis: Sardar Hussain, Sohaib Ahmad.

Investigation: Sardar Hussain.

Methodology: Sohail Akhtar.

Resources: Sohail Akhtar.

Software: Sardar Hussain.

Writing – original draft: Sardar Hussain, Sohaib Ahmad, Sohail Akhtar.

Writing – review & editing: Sardar Hussain, Sohaib Ahmad, Sohail Akhtar, Amara Javed, Uzma Yasmeen.

References

1. Hansen MH, Hurwitz WN. The problem of non-response in sample surveys. *Journal of the American Statistical Association*. 1946 Dec 1; 41(236):517–529. <https://doi.org/10.1080/01621459.1946.10501894> PMID: 20279350
2. Rao PS. Ratio estimation with sub sampling the non-respondents. *Survey methodology*. 1986 Dec; 12(2):217–230.
3. Khare BB, Srivastava S. Transformed ratio type estimators for the population mean in the presence of nonresponse. *Communications in Statistics-Theory and Methods*. 1997 Jan 1; 26(7):1779–1791.
4. Khare BB, Sinha RR. On class of estimators for population mean using multi-auxiliary characters in the presence of non-response. *Statistics in Transition*. 2009; 10(1):3–14.
5. Khare BB, Sinha RR. Estimation of population mean using multi-auxiliary characters with subsampling the nonrespondents. *Statistics in Transition new aeries*. 2011; 1(12):45–56.
6. Yunusa O, Kumar S. Ratio-cum-product estimator using exponential estimator in the presence of non-response. *Journal of Advanced Computing*. 2014; 3(1):1–11.
7. 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):2181–2192.
8. Pal SK, Singh HP. Estimation of finite population mean using auxiliary information in presence of non-response. *Communications in Statistics-Simulation and Computation*. 2018 Jan 2; 47(1):143–165.
9. Ahmad S, Shabbir J. Use of extreme values to estimate finite population mean under pps sampling scheme. *Journal of Reliability and Statistical Studies*. 2018 Oct 5:99–112.
10. Chambers RL, Dunstan R. Estimating distribution functions from survey data. *Biometrika*. 1986 Dec 1; 73(3):597–604.
11. Rao JN, Kovar JG, Mantel HJ. On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*. 1990 Jun 1:365–375.
12. Rao TJ. Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of official statistics*. 1994; 10(2):153.
13. Kuk AY. A kernel method for estimating finite population distribution functions using auxiliary information. *Biometrika*. 1993 Jun 1; 80(2):385–392.
14. Ahmed MS. and Abu-Dayyeh W. Estimation of finite-population distribution function using multivariate auxiliary information. *Statistics in Transition*. 2001 5(3):501–507
15. Chen J, Wu C. Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. *Statistica Sinica*. 2002 Oct 1:1223–1239.
16. Rueda M, Martínez S, Martínez H, Arcos A. Estimation of the distribution function with calibration methods. *Journal of statistical planning and inference*. 2007 Feb 1; 137(2):435–448.
17. 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):115–130.
18. Chen F, Chen S, Ma X. Crash frequency modeling using real-time environmental and traffic data and unbalanced panel data models. *International journal of environmental research and public health*. 2016 Jun; 13(6):609. <https://doi.org/10.3390/ijerph13060609> PMID: 27322306
19. Zeng Q, Wen H, Huang H, Pei X, Wong SC. A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. *Accident Analysis & Prevention*. 2017 Feb 1; 99:184–191. <https://doi.org/10.1016/j.aap.2016.11.018> PMID: 27914307
20. 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.

21. 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
22. Chen F, Chen S, Ma X. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. *Journal of safety research*. 2018 Jun 1; 65:153–159. <https://doi.org/10.1016/j.jsr.2018.02.010> PMID: 29776524
23. Zeng Q, Guo Q, Wong SC, Wen H, Huang H, Pei X,. Jointly modeling area-level crash rates by severity: a bayesian multivariate random-parameters spatio-temporal tobit regression. *Transportmetrica A: Transport science*. 2019 Aug 21; 15(2):1867–188.
24. Zeng Q, Wen H, Wong SC, Huang H, Guo Q, Pei X. Spatial joint analysis for zonal day time and night time crash frequencies using a Bayesian bivariate conditional autoregressive model. *Journal of Transportation Safety & Security*. 2020 Apr 20; 12(4):566–585. <https://doi.org/10.4271/2016-01-1439> PMID: 27648455
25. Hussain S, Zichuan M, Hussain S, Iftikhar A, Asif M, Akhtar S, et al. On Estimation of Distribution Function Using Dual Auxiliary Information under Nonresponse Using Simple Random Sampling. *Journal of Probability and Statistics*. 2020 Sep 26; 2020.
26. Hussain S, Ahmad S, Saleem M, Akhtar S. Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. *Plos one*. 2020 Sep 28; 15(9): e0239098. <https://doi.org/10.1371/journal.pone.0239098> PMID: 32986764
27. Plewes TJ, Tourangeau R, editors. *Nonresponse in social science surveys: A research agenda*. National Academies Press; 2013.
28. 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):262–275.
29. Murthy MN. Product method of estimation. *Sankhyā: The Indian Journal of Statistics, Series A*. 1964 Jul 1; 26(1):69–74.
30. Rao TJ. On certain methods of improving ratiion and regression estimators. *Communications in Statistics-Theory and Methods*. 1991 Jan 1; 20(10):3325–3340.
31. Singh R, Chauhan P, Sawan N, Smarandache F. Improvement in estimating the population mean using exponential estimator in simple random sampling. *Auxiliary Information and a priori Values in Construction of Improved Estimators*. 2007: 33.
32. 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):1552–1574.
33. 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):4425–4436.
34. Gujarati DN. *Basic econometrics*. Tata McGraw-Hill Education; 2009.
35. Singh S. (2003). *Advanced Sampling Theory with Applications: How Michael 'Selected' Amy*, Volume 1. Springer Science & Business Media.