Scientia Iranica (2024) 31(18), 1625–1637

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
https://scientiairanica.sharif.edu

On use of subsampling of the non-respondents for estimation of distribution function

Javid Shabbir[a,*] **and Sat Gupta**[b]

a. *Department of Statistics, Quaid-i-Azam University, Islamabad, 45320, Pakistan.*
b. *Department of Mathematics and Statistics, The University of North Carolina at Greensboro, NC 27412, USA.*

Received 12 April 2020; received in revised form 10 August 2021; accepted 8 November 2021

KEYWORDS
Distribution Function (DF);
Nonresponse;
Bias;
MSE;
Efficiency.

Abstract. In this study, we propose a general class of estimators of the finite population Distribution Function (DF) using two auxiliary variables under subsampling of non-respondents. We use the Hansen and Hurwitz pioneered model in our subsampling technique. Layout of response and non-response classes are discussed in various tables in detail. Expressions for the biases and Mean Square Errors (MSEs) of the estimators are obtained up to first order of approximation. We also obtain the conditions by comparing the proposed estimator with existing estimators. Three real data sets are used to support the theoretical findings. In our findings, it is observed that the proposed class of estimators is more efficient as compared to all other existing estimators including the usual mean estimator, ratio estimator, exponential-ratio estimator, traditional difference estimator, and many well-known difference type estimators by using the criterion of MSE.

1. Introduction

The problem of non-response is common in sample survey due to many reasons such as non-availability at home or unwilling to respond due to social desirability concerns or the fear of catching some contagious disease such as Covid-19 virus by having a contact with interviewer. Hansen and Hurwitz [1] were the first who floated the indigenous idea of nonresponse. Much work has been done since to deal the non-response by constructing composite types of estimators. The ratio, product, exponential-ratio and regression type estimators are commonly in this context (see Rao [2] and Kumar et al. [3]). Some related work is credit to Gupta and Shabbir [4], Khan and Shabbir [5], Verma et al. [6], Bhushan and Kumar [7], Kumar and Bhoughal [8], Saleem et al. [9], Ahmed et al. [10], Waseem et al. [11] and Yaqub and Shabbir [12,13].

Most of this work is based on estimation of finite population mean, total and variance but very little attention has been paid to estimating the Distribution Function (DF). Some works on estimating the DF can be found in Ahmad and Abu-Dayyah [14], Wang and Dorfman [15], Singh et al. [16] and Munoz et al. [17]. Some other useful references are, Irfan et al. [18], Abid et al. [19], Abid et al. [20], Javed et al. [21], Naz et al. [22], Younis and Shabbir [23], Ahmed and Shabbir [24] and Nazir et al. [25].

In our study, we propose a new class of estimators

*. *Corresponding author. Tel.: +92 0300 5273086*
E-mail address: javid.shabbir@uow.edu.pk (J. Shabbir)

To cite this article:
J. Shabbir and S. Gupta "On use of subsampling of the non-respondents for estimation of distribution function", *Scientia Iranica*, (2024) **31**(18), pp. 1625–1637
https://doi.org/10.24200/sci.2021.55790.4406

for estimating the DF under subsampling of non-respondents when nonresponse exists on the study variable as well as on the auxiliary variables.

Consider a finite population $U=\{U_1,U_2,...,U_N\}$ of N units portioned into two classes i.e., (i) response class with size N_1 and (ii) nonresponse class with size N_2. Using Hansen and Hurwitz [1] technique, a sample of size n is drawn from U by using Simple Random Sampling Without Replacement (SRSWOR). We assume that n_1 of the sampled units respond and n_2 do not. Let a sub-sample of r units be drawn from the n_2 non-responding units by SRSWOR and we collect the information on these r units by the interviewing method as $r=\frac{n_2}{K}$, $(K>1)$. Let y_i and (x_i,z_i) $(i=1,2,\cdots,n)$ be the values of the study variable (Y) and the auxiliary variables (X,Z) respectively. We are interested in estimating the DF defined as $F_{Y(t_y)}=\frac{1}{N}\sum_{i=1}^{N}I(y_i\leq t_y)$,$-\infty<t_y<\infty$, where $I(.)$ is the indicator function such that $I=(1,0)$. Similarly, we can define:

$$F_{X(t_x)}=\frac{1}{N}\sum_{i=1}^{N}I(x_i\leq t_x),\quad \text{and}$$

$$F_{Z(t_z)}=\frac{1}{N}\sum_{i=1}^{N}I(z_i\leq t_z).$$

The DF under stratification is:

$$F_{Y(t_y)}=G_1F^{(1)}_{Y(t_y)}+G_2F^{(2)}_{Y(t_y)}\quad \text{where}$$

$$G_i=\frac{N_i}{N}\quad (i=1,2),\quad F^{(1)}_{Y(t_y)}=\frac{1}{N_1}\sum_{i=1}^{N_1}I(y_i\leq t_y),$$

$$\text{and}\quad F^{(2)}_{Y(t_y)}=\frac{1}{N_2}\sum_{i=1}^{N_2}I(y_i\leq t_y).$$

Hansen and Hurwitz [1] estimator of DF under nonresponse is defined as:

$$\hat{F}^*_{F_Y(t_y)}=g_1\hat{F}^{(1)}_{Y(t_y)}+g_2F^{(2r)}_{Y(t_y)}\quad \text{where}$$

$$g_i=\frac{n_i}{n}\quad (i=1,2),$$

$$\hat{F}^{(1)}_{Y(t_y)}=\frac{1}{n_1}\sum_{i=1}^{n_1}I(y_i\leq t_y),$$

and

$$\hat{F}^{(2r)}_{Y(t_y)}=\frac{1}{r}\sum_{i=1}^{r}I(y_i\leq t_y).$$

Similarly, we can define:

$$\hat{F}^*_{F_X(t_x)}=g_1\hat{F}^{(1)}_{X(t_x)}+g_2\hat{F}^{(2r)}_{X(t_x)},$$

and

$$\hat{F}^*_{F_Z(t_z)}=g_1\hat{F}^{(1)}_{Z(t_z)}+g_2F^{(2r)}_{Z(t_z)}.$$

Let:

$$S^2_{F_Y(t_y)}=F_{Y(t_y)}\left(1-F_{Y(t_y)}\right),$$

$$S^2_{F_X(t_x)}=F_{X(t_x)}\left(1-F_{X(t_x)}\right),$$

and

$$S^2_{F_Z(t_z)}=F_{Z(t_z)}\left(1-F_{Z(t_z)}\right),$$

be the finite population variances for Y, X, and Z respectively for the response class. Similarly, the population variances for the non-response class are defined as:

$$S^{2(2)}_{F_{Y(t_y)}}=F^{(2)}_{Y(t_y)}\left(1-F^{(2)}_{Y(t_y)}\right),$$

$$S^{2(2)}_{F_X(t_x)}=F^{(2)}_{X(t_x)}\left(1-F^{(2)}_{X(t_x)}\right),$$

and

$$S^{2(2)}_{F_{Z(t_z)}}=F^{(2)}_{Z(t_z)}\left(1-F^{(2)}_{Z(t_z)}\right).$$

Let:

$$S_{F_{YX}(t_y,t_x)}=\frac{N_{110}N_{220}-N_{120}N_{210}}{N^2},$$

$$S_{F_{YZ}(t_y,t_z)}=\frac{N_{101}N_{202}-N_{102}N_{201}}{N^2},$$

$$S_{F_{XZ}(t_x,t_z)}=\frac{N_{011}N_{022}-N_{012}N_{021}}{N^2},$$

be the population covariances for the response class in their respective subscripts and similarly the population covariances for the non-response class in their respective subscripts are:

$$S^{(2)}_{F_{YX}(t_y,t_x)}=\frac{N^{(2)}_{110}N^{(2)}_{220}-N^{(2)}_{120}N^{(2)}_{210}}{N^{(2)2}},$$

$$S^{(2)}_{F_{YZ}(t_y,t_z)}=\frac{N^{(2)}_{101}N^{(2)}_{202}-N^{(2)}_{102}N^{(2)}_{201}}{N^{(2)2}},$$

$$S^{(2)}_{F_{XZ}(t_x,t_z)}=\frac{N^{(2)}_{011}N^{(2)}_{022}-N^{(2)}_{012}N^{(2)}_{021}}{N^{(2)2}}.$$

The layout for response and non-response classes are given in Tables 1–6.

Here N_{110}, N_{120}, N_{210}, and N_{220} are the number of units in the population and similarly n_{110}, n_{120}, n_{210}, and n_{220} be the number of units in the sample in their respective cells of respondents.

Here $N^{(2)}_{110}$, $N^{(2)}_{120}$, $N^{(2)}_{210}$, and $N^{(2)}_{220}$ are the number of units in the population and similarly $n^{(2)}_{110}$, $n^{(2)}_{120}$, $n^{(2)}_{210}$,

Table 1. Layout of the response class for Y and X.

	$X \leq F_X(t_x)$	$X > F_X(t_x)$	Total
$Y \leq F_Y(t_y)$	n_{110}/N_{110}	n_{120}/N_{120}	N_{100}
$Y > F_Y(t_y)$	n_{210}/N_{210}	n_{220}/N_{220}	N_{200}
Total	N_{010}	N_{020}	N

Table 2. Layout of the non-response class for Y and X.

	$X_2 \leq F_X^{(2)}(t_x)$	$X_2 > F_X^{(2)}(t_x)$	Total
$Y_2 \leq F_Y^{(2)}(t_y)$	$n_{110}^{(2)}/N_{110}^{(2)}$	$n_{120}^{(2)}/N_{120}^{(2)}$	$N_{100}^{(2)}$
$Y_2 > F_Y^{(2)}(t_y)$	$n_{210}^{(2)}/N_{210}^{(2)}$	$n_{220}^{(2)}/N_{220}^{(2)}$	$N_{200}^{(2)}$
Total	$N_{010}^{(2)}$	$N_{020}^{(2)}$	N

Table 3. Layout of the response class for Y and Z.

	$Z \leq F_Z(t_z)$	$Z > F_Z(t_z)$	Total
$Y \leq F_Y(t_y)$	n_{101}/N_{101}	n_{102}/N_{102}	N_{100}
$Y > F_Y(t_y)$	n_{201}/N_{201}	n_{202}/N_{202}	N_{200}
Total	N_{001}	N_{002}	N

Table 4. Layout of the non-response class for Y and Z.

	$Z_2 \leq F_Z^{(2)}(t_z)$	$Z_2 > F_Z^{(2)}(t_z)$	Total
$Y_2 \leq F_Y^{(2)}(t_y)$	$n_{101}^{(2)}/N_{101}^{(2)}$	$n_{102}^{(2)}/N_{102}^{(2)}$	$N_{100}^{(2)}$
$Y_2 > F_Y^{(2)}(t_y)$	$n_{201}^{(2)}/N_{201}^{(2)}$	$n_{202}^{(2)}/N_{202}^{(2)}$	$N_{200}^{(2)}$
Total	$N_{001}^{(2)}$	$N_{002}^{(2)}$	N

Table 5. Layout of the response class for X and Z.

	$Z \leq F_Z(t_z)$	$Z > F_Z(t_z)$	Total
$X \leq F_X(t_x)$	n_{011}/N_{011}	n_{012}/N_{012}	N_{010}
$X > F_X(t_x)$	n_{021}/N_{021}	n_{022}/N_{022}	N_{020}
Total	N_{001}	N_{002}	N

Table 6. Layout of the non-response class for X and Z.

	$Z_2 \leq F_Z^{(2)}(t_z)$	$Z_2 > F_Z^{(2)}(t_z)$	Total
$X_2 \leq F_X^{(2)}(t_x)$	$n_{011}^{(2)}/N_{011}^{(2)}$	$n_{012}^{(2)}/N_{012}^{(2)}$	$N_{010}^{(2)}$
$X_2 > F_X^{(2)}(t_x)$	$n_{021}^{(2)}/N_{021}^{(2)}$	$n_{022}^{(2)}/N_{022}^{(2)}$	$N_{020}^{(2)}$
Total	$N_{001}^{(2)}$	$N_{002}^{(2)}$	N

and $n_{220}^{(2)}$ be the number of units in the sample in their respective cells of respondents.

Here N_{101}, N_{102}, N_{201}, and N_{202} are the number of units in the population and similarly n_{101}, n_{102}, n_{201}, and n_{202} be the number of units in the sample in their respective cells of respondents.

Here $N_{101}^{(2)}$, $N_{102}^{(2)}$, $N_{201}^{(2)}$, and $N_{202}^{(2)}$ are the number of units in the population and similarly $n_{101}^{(2)}$, $n_{102}^{(2)}$, $n_{201}^{(2)}$, and $n_{202}^{(2)}$ be the number of units in the sample in their respective cells of respondents.

Here N_{011}, N_{012}, N_{021}, and N_{022} are the number of units in the population and similarly n_{011}, n_{012}, n_{021}, and n_{022} be the number of units in the sample in their respective cells of respondents.

Here $N_{110}^{(2)}$, $N_{012}^{(2)}$, $N_{021}^{(2)}$, and $N_{022}^{(2)}$ are the number of units in the population and similarly $n_{011}^{(2)}$, $n_{012}^{(2)}$, $n_{021}^{(2)}$, and $n_{022}^{(2)}$be the number of units in the sample in their respective cells of respondents.

Now we define some error terms to obtain the biases and Mean Square Errors (MSEs) up to first order of approximation.

$$\Delta_0^* = \frac{F_{Y(t_y)}^* - F_{Y(t_y)}}{F_{Y(t_y)}}, \quad \Delta_1^* = \frac{F_{X(t_x)}^* - F_{X(t_x)}}{F_{X(t_x)}},$$

$$\Delta_2^* = \frac{F_{Z(t_z)}^* - F_{Z(t_z)}}{F_{Z(t_z)}},$$

such that $E(\Delta_i^*) = 0$ for $(i = 0, 1, 2)$, and

$$E(\Delta_0^{*2}) = \frac{1}{F_Y^2(t_y)} \left\{ \lambda_1 S_{F_Y(t_y)}^2 + \lambda_2 S_{F_Y(t_y)}^{(2)2} \right\} = \Lambda_{200}^*,$$

$$E(\Delta_1^{*2}) = \frac{1}{F_X^2(t_x)} \left\{ \lambda_1 S_{F_X(t_x)}^2 + \lambda_2 S_{F_X(t_x)}^{(2)2} \right\} = \Lambda_{020}^*,$$

$$E(\Delta_2^{*2}) = \frac{1}{F_Z^2(t_z)} \left\{ \lambda_1 S_{F_Z(t_z)}^2 + \lambda_2 S_{F_Z(t_z)}^{(2)2} \right\} = \Lambda_{002}^*,$$

$$E(\Delta_0^* \Delta_1^*) = \frac{1}{F_Y(t_y) F_X(t_x)} \left\{ \lambda_1 S_{F_{YX}(t_y,t_x)} + \lambda_2 S_{F_{YX}(t_y,t_x)}^{(2)} \right\} = \Lambda_{110}^*,$$

$$E(\Delta_0^* \Delta_2^*) = \frac{1}{F_Y(t_y) F_Z(t_z)} \left\{ \lambda_1 S_{F_{YZ}(t_y,t_z)} + \lambda_2 S_{F_{YZ}(t_y,t_z)}^{(2)} \right\} = \Lambda_{101}^*,$$

$$E(\Delta_1^*\Delta_2^*) = \frac{1}{F_X(t_x)F_Z(t_z)} \left\{\lambda_1 S_{F_{XZ}(t_x,t_z)} + \lambda_2 S^{(2)}_{F_{XZ}(t_x,t_z)}\right\} = \Lambda^*_{101},$$

where some equations are shown in Box I.

Now we discuss some estimators of DF using single auxiliary variable and two auxiliary variables.

2. Existing estimators

In this section, we discuss the following estimators:

(i) The variance of the usual estimator $\hat{F}^*_{F_Y(t_y)} = \hat{F}^*_0$, is given by:

$$Var(\hat{F}^*_0) = F^2_{Y(t_y)}\Lambda^*_{200}. \quad (1)$$

(ii) The traditional ratio estimator, is given by:

$$\hat{F}^*_{R_1} = \hat{F}^*_{Y(t_y)}\left(\frac{F_{X(t_x)}}{\hat{F}^*_{X(t_x)}}\right). \quad (2)$$

The bias and MSE respectively of $\hat{F}^*_{R_1}$, to first order of approximation, are given by:

$$B(\hat{F}^*_{R_1}) \cong F^2_{Y(t_y)}\left\{\Lambda^*_{020} - \Lambda^*_{110}\right\}, \quad (3)$$

and

$$MSE(\hat{F}^*_{R_1}) \cong F^2_{Y(t_y)}\left\{\Lambda^*_{200} + \Lambda^*_{020} - 2\Lambda^*_{110}\right\}. \quad (4)$$

(iii) The traditional exponential-ratio type estimator, is given by:

$$\hat{F}^*_{E_1} = \hat{F}^*_{Y(t_y)} \exp\left(\frac{F_{X(t_x)} - \hat{F}^*_{X(t_x)}}{F_{X(t_x)} + \hat{F}^*_{X(t_x)}}\right). \quad (5)$$

The bias and MSE respectively of $\hat{F}^*_{E_1}$, to first order of approximation, are given by:

$$B(\hat{F}^*_{E_1}) \cong F^2_{Y(t_y)}\left\{\frac{3\Lambda^*_{020}}{8} - \frac{\Lambda^*_{110}}{2}\right\}, \quad (6)$$

and

$$MSE(\hat{F}^*_{E_1}) \cong F^2_{Y(t_y)}\left\{\Lambda^*_{200} + \frac{\Lambda^*_{020}}{4} - \Lambda^*_{110}\right\}. \quad (7)$$

(iv) The usual difference estimator, is given by:

$$\hat{F}^*_{D_1} = \hat{F}^*_{Y(t_y)} + d_0\left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right), \quad (8)$$

where d_0 is the constant.

The minimum variance of $\hat{F}^*_{D_1}$ at the optimum value of $d_{0(opt)} = \frac{F_{Y(t_y)}\Lambda^*_{110}}{F_{X(t_x)}\Lambda^*_{020}}$, is given by:

$$Var(\hat{F}^*_{D_1})_{\min} = MSE(\hat{F}^*_{D_1})_{\min} \cong F^2_{Y(t_y)}\Lambda^*_{200}\left(1 - \rho^{*2}_{110}\right), \quad (9)$$

where $\rho^*_{110} = \frac{\Lambda^*_{110}}{\sqrt{\Lambda^*_{200}}\sqrt{\Lambda^*_{020}}}$.

(v) Rao [2] difference type estimator, is given by:

$$\hat{F}^*_{Rao} = d_1\hat{F}^*_{Y(t_y)} + d_2\left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right), \quad (10)$$

where $d_i(i = 1,2)$ are the constants.

The bias and minimum MSE respectively of $\hat{F}^*_{Rao}$ at optimum values of:

$$d_{1(opt)} = \frac{1}{1 + \Lambda^*_{020}(1 - \rho^{*2}_{110})},$$

and

$$d_{2(opt)} = \frac{F_{Y(t_y)}\Lambda^*_{110}}{F_{X(t_x)}\Lambda^*_{020}\left\{1 + \Lambda^*_{020}(1 - \rho^{*2}_{110})\right\}},$$

are given by:

$$Bias(\hat{F}^*_{Rao}) \cong (d_1 - 1)F_{Y(t_y)} \quad (11)$$

and

$$MSE(\hat{F}^*_{Rao})_{\min} \cong F^2_{Y(t_y)}\frac{\Lambda^*_{200}\left(1 - \rho^{*2}_{110}\right)}{1 + \Lambda^*_{200}\left(1 - \rho^{*2}_{110}\right)}. \quad (12)$$

(vi) Gupta and Shabbir [4] estimator using two auxiliary variables, is given by:

$$\hat{F}^*_{GS} = \left\{J_1\hat{F}^*_{Y(t_y)} + J_2\left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right)\right\}\left\{\frac{F_{X(t_x)}}{\hat{F}^*_{X(t_x)}}\right\}, \quad (13)$$

$$\lambda_1 = \left(\frac{1}{n} - \frac{1}{N}\right), \qquad \lambda_2 = \frac{N_2(K-1)}{Nn},$$

$$\Lambda^*_{def} = \frac{E\left[\left\{F^*_{Y(t_y)} - F_Y(t_y)\right\}^d\left\{F^*_{X(t_x)} - F_X(t_x)\right\}^e\left\{F^*_{Z(t_z)} - F_Z(t_z)\right\}^f\right]}{\left\{F_Y(t_y)\right\}^d\left\{F_X(t_x)\right\}^e\left\{F_Z(t_z)\right\}^f}.$$

Box I

where $J_i (i = 1, 2)$ are the constants.

The bias and minimum MSE respectively of $\hat{F}^*_{GS}$ at optimum values of

$$J_{1(opt)} = \frac{B_j C_j - D_j E_j + B_j}{A_j B_j - E_j^2 + B_j}$$

$$J_{2(opt)} = \frac{F_{Y(t_y)}(A_j D_j - C_j E_j + D_j - E_j)}{F_{X(t_x)}(A_j B_j - E_j^2 + B_j)},$$

are given by:

$$Bias(\hat{F}^*_{GS}) \cong (J_1 - 1) F_{Y(t_y)} + J_1 F_{Y(t_y)} C_j + J_2 F_{X(t_x)} D_j, \quad (14)$$

$$MSE(\hat{F}^*_{GS})_{\min} \cong F^2_{Y(t_y)} \left\{1 - \frac{A_j D_j^2 + B_j C_j^2 - 2C_j D_j E_j + 2B_j C_j - 2D_j E_j + B_j}{(A_j B_j - E_j^2 + B_j)}\right\}, \quad (15)$$

where $A_j = \Lambda^*_{200} + \Lambda^*_{020} - 2\Lambda^*_{110}$, $B_j = \Lambda^*_{020}$, $C_j = \frac{3\Lambda^*_{020}}{8} - \frac{\Lambda^*_{110}}{2}$, $D_j = \frac{\Lambda^*_{020}}{2}$, $E_j = \Lambda^*_{020} - \Lambda^*_{110}$.

(vii) The traditional ratio estimator using two auxiliary variables, is given by:

$$\hat{F}^*_{R_2} = \hat{F}^*_{Y(t_y)} \left(\frac{F_{X(t_x)}}{\hat{F}^*_{X(t_x)}}\right) \left(\frac{F_{Z(t_z)}}{\hat{F}^*_{Z(t_z)}}\right). \quad (16)$$

The bias and MSE respectively of $\hat{F}^*_{R_2}$ to first order of approximation are given by:

$$B(\hat{F}^*_{R_2}) \cong F_{Y(t_y)} \{\Lambda^*_{020} + \Lambda^*_{002} + \Lambda^*_{011} - \Lambda^*_{110} - \Lambda^*_{101}\}, \quad (17)$$

and

$$MSE(\hat{F}^*_{R_2}) \cong F^2_{Y(t_y)} \{\Lambda^*_{200} + \Lambda^*_{020} + \Lambda^*_{002} - 2\Lambda^*_{110} - 2\Lambda^*_{101} + 2\Lambda^*_{011}\}. \quad (18)$$

(viii) The traditional exponential ratio estimator using two auxiliary variables, is given by:

$$\hat{F}^*_{E_2} = \hat{F}^*_{Y(t_y)} \exp\left(\frac{F_{X(t_x)} - \hat{F}^*_{X(t_x)}}{F_{X(t_x)} + \hat{F}^*_{X(t_x)}}\right) \exp\left(\frac{F_{Z(t_z)} + \hat{F}^*_{Z(t_z)}}{F_{Z(t_z)} + \hat{F}^*_{Z(t_z)}}\right). \quad (19)$$

The bias and MSE respectively of $\hat{F}^*_{E_2}$ to first order of approximation, are given by:

$$B(\hat{F}^*_{E_2}) \cong F_{Y(t_y)} \left\{\frac{3}{8}(\Lambda^*_{020} + \Lambda^*_{002}) - \frac{1}{2}(\Lambda^*_{110} - \Lambda^*_{101}) + \frac{1}{4}\Lambda^*_{011}\right\}, \quad (20)$$

and

$$MSE(\hat{F}^*_{E_2}) \cong F^2_{Y(t_y)} \left\{\Lambda^*_{200} + \frac{1}{4}(\Lambda^*_{020} + \Lambda^*_{002}) - (\Lambda^*_{110} + \Lambda^*_{101}) + \frac{1}{2}\Lambda^*_{011}\right\}. \quad (21)$$

(ix) The usual difference estimator using two auxiliary variables, is given by:

$$\hat{F}^*_{D_2} = \hat{F}^*_{Y(t_y)} + d_1\left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right) + d_2\left(F_{Z(t_z)} - \hat{F}^*_{Z(t_z)}\right), \quad (22)$$

where $d_i (i = 1, 2)$ are constants.

The minimum variance or MSE of $\hat{F}^*_{D_2}$ at the optimum values of $d_i (i = 1, 2)$ i.e.:

$$d_{1(opt)} = \frac{F_{Y(t_y)}(\Lambda^*_{101}\Lambda^*_{011} - \Lambda^*_{002}\Lambda^*_{110})}{F_{X(t_x)}(\Lambda^{*2}_{011} - \Lambda^*_{020}\Lambda^*_{022})},$$

and

$$d_{2(opt)} = \frac{F_{Y(t_y)}(\Lambda^*_{011}\Lambda^*_{110} - \Lambda^*_{020}\Lambda^*_{101})}{F_{Z(t_x)}(\Lambda^{*2}_{011} - \Lambda^*_{020}\Lambda^*_{022})}.$$

The minimum MSE is given in Box II

or

$$\text{MSE}\ (\hat{F}^*_{D_2})_{\min} \cong F^2_{Y(t_y)}\Lambda^*_{200} \left\{1 - \frac{\rho^{*2}_{110} + \rho^{*2}_{101} - 2\rho^*_{110}\rho^*_{101}\rho^*_{011}}{1 - \rho^{*2}_{011}}\right\}, \quad (23)$$

where,

$$\rho^*_{110} = \frac{\Lambda^*_{110}}{\sqrt{\Lambda^*_{200}}\sqrt{\Lambda^*_{020}}}, \quad \rho^*_{101} = \frac{\Lambda^*_{101}}{\sqrt{\Lambda^*_{200}}\sqrt{\Lambda^*_{002}}},$$

$$\rho^*_{011} = \frac{\Lambda^*_{011}}{\sqrt{\Lambda^*_{020}}\sqrt{\Lambda^*_{002}}}.$$

$$MSE(\hat{F}^*_{D_2})_{\min} \cong F^2_{Y(t_y)} \left\{\frac{\Lambda^{*2}_{101}\Lambda^*_{020} - 2\Lambda^*_{101}\Lambda^*_{011}\Lambda^*_{110} + \Lambda^{*2}_{011}\Lambda^*_{200} - \Lambda^*_{020}\Lambda^*_{002}\Lambda^*_{200} + \Lambda^{*2}_{110}\Lambda^*_{002}}{\Lambda^{*2}_{011} - \Lambda^*_{020}\Lambda^*_{002}}\right\}.$$

Box II

(x) Kumar et al. [8] estimator using two auxiliary variables, is given by:

$$\hat{F}^*_{KU} = \hat{F}^*_{Y(t_y)} \left(\frac{F_{X(t_x)}}{\hat{F}^*_{X(t_x)}} \right) \left\{ \alpha_0 \exp\left(\frac{F_{Z(t_x)} - \hat{F}^*_{Z(t_x)}}{F_{Z(t_x)} + \hat{F}^*_{Z(t_x)}} \right) + (1-\alpha_0) \exp\left(\frac{\hat{F}^*_{Z(t_x)} - F_{Z(t_x)}}{\hat{F}^*_{Z(t_x)} + F_{Z(t_x)}} \right) \right\}, \quad (24)$$

where α_0 is the constant.

The bias and minimum MSE respectively of $\hat{F}^*_{KU}$ to first order of approximation at optimum value of:

$$\alpha_{0(opt)} = \frac{1}{2} - \frac{(\Lambda^*_{011} - \Lambda^*_{101})}{\Lambda^*_{002}},$$

are given by:

$$B(\hat{F}^*_{KU}) \cong F_{Y(t_y)} \left\{ \Lambda^*_{020} + \left(\frac{1}{2} - \alpha_0 \right) (\Lambda^*_{101} - \Lambda^*_{011}) - \left(\frac{1}{8} - \frac{1}{2}\alpha_0 \right) \Lambda^*_{002} \right\}, \quad (25)$$

and

$$MSE(\hat{F}^*_{KU})_{\min} \cong F^2_{Y(t_y)} \left\{ (\Lambda^*_{200} + \Lambda^*_{020} - 2\Lambda^*_{110}) - \frac{(\Lambda^*_{011} - \Lambda^*_{101})^2}{\Lambda^*_{002}} \right\}. \quad (26)$$

(xi) On the lines of Chami et al. [26], Guha and Chandra [27] and Singh and Usman [28] estimators using two auxiliary variables, we have:

$$\hat{F}^*_{Ch} = \hat{F}^*_{Y(t_y)} \left\{ \frac{\alpha_1 \hat{F}^*_{X(t_x)} + (1-\alpha_1) F_{X(t_x)}}{(1-\alpha_1)\hat{F}^*_{X(t_x)} + \alpha_1 F_{X(t_x)}} \right\} \left\{ \frac{\alpha_2 \hat{F}^*_{Z(t_z)} + (1-\alpha_2) F_{Z(t_z)}}{(1-\alpha_2)\hat{F}^*_{Z(t_z)} + \alpha_2 F_{Z(t_z)}} \right\}, \quad (27)$$

where $\alpha_i (i = 1, 2)$ are the constants.

The bias and minimum MSE respectively of $\hat{F}^*_{Ch}$ at the optimum values of $\alpha_i (i = 1, 2)$ i.e.:

$$\alpha_{1(opt)} = \frac{1}{2} \left\{ 1 + \frac{(\Lambda^*_{101}\Lambda^*_{011} - \Lambda^*_{002}\Lambda^*_{110})}{(\Lambda^{*2}_{011} - \Lambda^*_{020}\Lambda^*_{022})} \right\}$$

and

$$\alpha_{2(opt)} = \frac{1}{2} \left\{ 1 - \frac{(\Lambda^*_{020}\Lambda^*_{101} - \Lambda^*_{011}\Lambda^*_{110})}{(\Lambda^{*2}_{011} - \Lambda^*_{020}\Lambda^*_{022})} \right\}, \quad (28)$$

are given by:

$$Bias(\hat{F}^*_{Ch}) \cong F_{Y(t_y)} \Big\{ (2\alpha_1 - 1)\Lambda^*_{110} + (2\alpha_2 - 1) \Lambda^*_{101} + (2\alpha_1 - 1)(2\alpha_2 - 1) \Lambda^*_{011} (1-\alpha_1)(1-2\alpha_1)\Lambda^*_{020} + (1-\alpha_2) (1-2\alpha_2)\Lambda^*_{002} \Big\}, \quad (29)$$

and Eq. (30) is shown in Box III. The minimum MSE of $\hat{F}^*_{Ch}$ is equal to minimum MSE of the difference estimator $\hat{F}^*_{D_2}$.

(xii) Singh and Usman [28] estimator using two auxiliary variables, is given by:

$$\hat{F}^*_{SU} = \left\{ \hat{F}^*_{Y(t_y)} + \hat{\beta}^*_{110} \left(F_{X(t_x)} - \hat{F}^*_{X(t_x)} \right) \right\} \left\{ \frac{\gamma_1 \hat{F}^*_{X(t_x)} + (1-\gamma_1) F_{X(t_x)}}{(1-\gamma_1)\hat{F}^*_{X(t_x)} + \gamma_1 F_{X(t_x)}} \right\} \left\{ \frac{\gamma_2 \hat{F}^*_{Z(t_z)} + (1-\gamma_2) F_{Z(t_z)}}{(1-\gamma_2)\hat{F}^*_{Z(t_z)} + \gamma_2 F_{Z(t_z)}} \right\}, \quad (31)$$

where $\gamma_i (i = 1, 2)$ are constants and $\hat{\beta}^*_{110} = \frac{\hat{F}_{Y(t_y)}\hat{\Lambda}^*_{110}}{\hat{F}_{X(t_x)}\hat{\Lambda}^*_{020}}$ is the sample regression coefficient with the corresponding population regression coefficient $\beta^*_{110} = \frac{F_{Y(t_y)}\Lambda^*_{110}}{F_{X(t_x)}\Lambda^*_{020}}$. It is observed that:

$$MSE(\hat{F}^*_{SU})_{\min} = MSE(\hat{F}^*_{D_2})_{\min} = MSE(\hat{F}^*_{Ch})_{\min}.$$

$$MSE(\hat{F}^*_{Ch})_{\min} \cong F^2_{Y(t_y)} \left\{ \frac{\Lambda^{*2}_{101}\Lambda^*_{020} - 2\Lambda^*_{101}\Lambda^*_{011}\Lambda^*_{110} + \Lambda^{*2}_{011}\Lambda^*_{200} - \Lambda^*_{020}\Lambda^*_{002}\Lambda^*_{200} + \Lambda^{*2}_{110}\Lambda^*_{002}}{\Lambda^{*2}_{011} - \Lambda^*_{020}\Lambda^*_{002}} \right\}. \quad (30)$$

Box III

3. Proposed estimator

We propose the following general class of difference type estimators of DF using two auxiliary variables. This estimator is constructed by using the ratio and exponential-ratio type estimators with the difference type estimator as:

$$\hat{F}^*_{P(\delta_1,\delta_2)} = \left\{\omega_1 \hat{F}^*_{Y(t_y)} + \omega_2\left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right) + \omega_3\left(F_{Z(t_z)} - \hat{F}^*_{Z(t_z)}\right)\right\} \times\left\{\left(\frac{F_{X(t_x)}}{\hat{F}^*_{X(t_x)}}\right)^{\delta_1} \exp\ \delta_2\left(\frac{F_{X(t_x)}-\hat{F}^*_{X(t_x)}}{F_{X(t_x)}+\hat{F}^*_{X(t_x)}}\right)\right\}, \quad (32)$$

where $\omega_i (i{=}1,2,3)$ are the constants and $(0 \le \delta_i \le 1)$ $(i = 1,2)$ are known scaler values.

Rewriting $\hat{F}^*_{P(\delta_1,\delta_2)}$ in terms of errors terms, we have:

$$\hat{F}^*_{P(\delta_1,\delta_2)} - F_{Y(t_y)} \cong (\omega_1 - 1)F_{Y(t_y)} + \omega_1 F_{Y(t_y)} \left[\Delta^*_0 - \delta^*_1\Delta^*_1 + \delta^*_2\Delta^{*2}_1 - \delta^*_1\Delta^*_0\Delta^*_1\right] - \omega_2 F_{X(t_x)}\left[\Delta^*_1 - \delta^*_1\Delta^{*2}_1\right] - \omega_3 F_{Z(t_z)}\left[\Delta^*_2 - \delta^*_1\Delta^*_1\Delta^*_2\right], \quad (33)$$

where $\delta^*_1 = \left(\delta_1 + \frac{\delta_2}{2}\right)$ and:

$$\delta^*_2 = \left\{\frac{\delta_1\delta_2}{2} + \frac{\delta_1(\delta_1+1)}{2} + \frac{\delta_2(\delta_2+2)}{8}\right\}.$$

From Eq. (33), the bias of $\hat{F}^*_{P(\delta_1,\delta_2)}$, is given by:

$$Bias(\hat{F}^*_{P(\delta_1,\delta_2)}) \cong (\omega_1 - 1)F_{Y(t_y)} + \omega_1 F_{Y(t_y)}\left\{\delta^*_2\Lambda^*_{020} - \delta^*_1\Lambda^*_{110}\right\} + F_{Z(t_z)}\delta^*_1\left(\omega_2\Lambda^*_{020} + \omega_3\Lambda^*_{011}\right). \quad (34)$$

Squaring and then taking expectation on Eq. (33), we get MSE of $\hat{F}^*_{P(\delta_1,\delta_2)}$, which is given by:

$$MSE(\hat{F}^*_{P(\delta_1,\delta_2)}) \cong (\omega_1 - 1)^2 F^2_{Y(t_y)} + \omega^2_1 F^2_{Y(t_y)} A + \omega^2_2 F^2_{X(t_x)} B + \omega^2_3 F^2_{Z(t_z)} C - 2\omega_1 F^2_{Y(t_y)} D - 2\omega_2 F_{Y(t_y)} F_{X(t_x)} E - 2\omega_3 F_{Y(t_y)} F_{Z(t_z)} F + 2\omega_1\omega_2 F_{Y(t_y)} F_{X(t_x)} G + 2\omega_1\omega_3 F_{Y(t_y)} F_{Z(t_z)} H + 2\omega_2\omega_3 F_{X(t_x)} F_{Z(t_z)} I,$$

where,

$$A = \Lambda^*_{200} + \left(\delta^{*2}_1 + 2\delta^*_2\right)\Lambda^*_{020} - 4\delta^*_1\Lambda^*_{110},$$

$$B = \Lambda^*_{020}, \qquad C = \Lambda^*_{002}, \qquad D = \delta^*_2\Lambda^*_{020} - \delta^*_1\Lambda^*_{110},$$

$$E = \delta^*_1\Lambda^*_{020}, \qquad F = \delta^*_1\Lambda^*_{011}, \qquad G = 2\delta^*_1\Lambda^*_{020} - \Lambda^*_{110},$$

$$H = 2\delta^*_1\Lambda^*_{011} - \Lambda^*_{101}, \qquad I = \Lambda^*_{011}.$$

The minimum MSE of $\hat{F}^*_{P(\delta_1,\delta_2)}$ at optimum values of $\omega_i (i = 1,2,3)$ i.e., $\omega_{1(opt)} = \frac{l_5}{l_1}$ $\omega_{2(opt)} = \frac{F_{Y(t_y)} l_6}{F_{X(t_x)} l_1}$ and $\omega_{3(opt)} = \frac{F_{Y(t_y)} l_7}{F_{Z(t_z)} l_1}$, is given by:

$$MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} \cong F^2_{Y(t_y)}\left(\frac{l_2 + l_3 + l_4}{l_1}\right), \quad (35)$$

where,

$$l_1 = ABC - BH^2 - AI^2 + BC - CG^2 + 2GHI - I^2,$$

$$l_2 = -ABF^2 - BCD^2 + 2BDFH + ABC - ACE^2 + 2AEFI - 2BCD - BF^2 + 2BFH - BH^2,$$

$$l_3 = 2CDEG + D^2I^2 - 2DEHI - 2DFGI + E^2H^2 - 2EFGH + F^2G^2 - AI^2 - CE^2,$$

$$l_4 = 2CEG - CG^2 + 2DI^2 + 2EFI - 2EHI - 2FGI + 2GHI,$$

$$l_5 = BCD - BFH + BC - CEG - DI^2 + EHJ + FGI - I^2,$$

$$l_6 = ACE - AFI - CDG + DHI - EH^2 + FGH + CE - CG - FI + HI,$$

$$l_7 = ABF - BDH - AEI + BF - BH + DGI + EGH - FG^2 - EI + GI.$$

We can generate many estimators from this proposed class of estimators as follows:

(i) Putting $\delta_1 = 0$ and $\delta_2 = 0$ in Eq. (32), we get:

$$\hat{F}^*_{P(0,0)} = \omega_1\hat{F}^*_{Y(t_y)} + \omega_2\left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right) + \omega_3\left(F_{Z(t_z)} - \hat{F}^*_{Z(t_z)}\right). \quad (36)$$

(ii) Putting $\delta_1 = 1$ and $\delta_2 = 0$ in Eq. (32), we get:

$$\hat{F}^*_{P(1,0)} = \left\{\omega_1 \hat{F}^*_{Y(t_y)} + \omega_2 \left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right) + \omega_3 \left(F_{Z(t_z)} - \hat{F}^*_{Z(t_z)}\right)\right\} \left(\frac{F_{X(t_x)}}{\hat{F}^*_{X(t_x)}}\right). \quad (37)$$

(iii) Putting $\delta_1 = 0$ and $\delta_2 = 1$ in Eq. (32), we get:

$$\hat{F}^*_{P(1,1)} = \left\{\omega_1 \hat{F}^*_{Y(t_y)} + \omega_2 \left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right) + \omega_3 \left(F_{Z(t_z)} - \hat{F}^*_{Z(t_z)}\right)\right\} \left\{\left(\frac{F_{X(t_x)}}{\hat{F}^*_{X(t_x)}}\right) \exp\left(\frac{F_{X(t_x)} - \hat{F}^*_{X(t_x)}}{F_{X(t_x)} + \hat{F}^*_{X(t_x)}}\right)\right\}. \quad (38)$$

(iv) Putting $\delta_1 = 0.5$ and $\delta_2 = 0.5$ in Eq. (32), we get:

$$\hat{F}^*_{P(0.5,0.5)} = \left\{\omega_1 \hat{F}^*_{Y(t_y)} + \omega_2 \left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right) + \omega_3 \left(F_{Z(t_z)} - \hat{F}^*_{Z(t_z)}\right)\right\} \left\{\left(\frac{F_{X(t_x)}}{\hat{F}^*_{X(t_x)}}\right)^{0.5} \exp\left(0.5\frac{F_{X(t_x)} - \hat{F}^*_{X(t_x)}}{F_{X(t_x)} + \hat{F}^*_{X(t_x)}}\right)\right\}. \quad (39)$$

(v) Putting $\delta_1 = 0$ and $\delta_2 = 1$ in Eq. (32), we get:

$$\hat{F}^*_{P(0,1)} = \left\{\omega_1 \hat{F}^*_{Y(t_y)} + \omega_2 \left(F_{X(t_x)} - \hat{F}^*_{X(t_x)}\right) + \omega_3 \left(F_{Z(t_z)} - \hat{F}^*_{Z(t_z)}\right)\right\} \left\{\exp\left(\frac{F_{X(t_x)} - \hat{F}^*_{X(t_x)}}{F_{X(t_x)} + \hat{F}^*_{X(t_x)}}\right)\right\}. \quad (40)$$

The biases and minimum MSEs of above estimators can be obtained by substituting the different values $\delta_i (i = 1, 2)$ in Eqs. (34) and (35). Also, we can generate many more estimators by substituting the different values of δ_i and $\omega_i (i = 1, 2)$ in Eq. (32).

4. Comparison of estimators

We compare the proposed generalized class of estimators with some other competing estimators.

(i) By Eqs. (1) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < Var(\hat{F}^*_0)$ if:

$$\left[\Lambda^*_{200} - \left(\frac{l_2 + l_3 + l_4}{l_1}\right)\right] > 0.$$

(ii) By Eqs. (4) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{R_1})$ if:

$$\left[\{\Lambda^*_{200} + \Lambda^*_{020} - 2\Lambda^*_{110}\} - \left(\frac{l_2 + l_3 + l_4}{l_1}\right)\right] > 0.$$

(ii) By Eqs. (7) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{E_1})$ if:

$$\left[\left\{\Lambda^*_{200} + \frac{\Lambda^*_{020}}{4} - \Lambda^*_{110}\right\} - \left(\frac{l_2 + l_3 + l_4}{l_1}\right)\right] > 0.$$

(iii) By Eqs. (9) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{D_1})_{\min}$ if:

$$\left[\Lambda^*_{200} \left(1 - \rho^{*2}_{110}\right) - \left(\frac{l_2 + l_3 + l_4}{l_1}\right)\right] > 0$$

(iv) By Eqs. (12) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{Rao})_{\min}$ if:

$$\left[\frac{\Lambda^*_{200} \left(1 - \rho^{*2}_{110}\right)}{1 + \Lambda^*_{200} \left(1 - \rho^{*2}_{110}\right)} - \left(\frac{l_2 + l_3 + l_4}{l_1}\right)\right] > 0.$$

(v) By Eqs. (15) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{GS})_{\min}$ if:

$$\left[\left\{1 - \frac{A_j D_j^2 + B_j C_j^2 - 2C_j D_j E_j + 2B_j C_j - 2D_j E_j + B_j}{(A_j B_j - E_j^2 + B_j)}\right\} - \left(\frac{l_2 + l_3 + l_4}{l_1}\right)\right] > 0.$$

(vi) By Eqs. (18) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{R_2})$ if:

$$\left[\left\{\Lambda^*_{200} + \Lambda^*_{020} + \Lambda^*_{002} - 2\left(\Lambda^*_{110} + \Lambda^*_{101} - \Lambda^*_{011}\right)\right\} - \left(\frac{l_2 + l_3 + l_4}{l_1}\right)\right] > 0.$$

(vii) By Eqs. (21) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{E_2})$ if:

$$\left[\left\{\Lambda^*_{200} + \frac{1}{4}\left(\Lambda^*_{020} + \Lambda^*_{002}\right) - \left(\Lambda^*_{110} + \Lambda^*_{101}\right) + \frac{1}{2}\Lambda^*_{011}\right\} - \left(\frac{l_2 + l_3 + l_4}{l_1}\right)\right] > 0.$$

(vii) By Eqs. (23) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{D_2})_{\min}$ if:

$$\left[\Lambda^*_{200}\left\{1-\frac{\rho^{*2}_{110}+\rho^{*2}_{101}-2\rho^*_{110}\rho^*_{101}\rho^*_{011}}{1-\rho^{*2}_{011}}\right\}\right.$$
$$\left.-\left(\frac{l_2+l_3+l_4}{l_1}\right)\right]>0.$$

(viii) By Eqs. (26) and (35), $MSE(\hat{F}^*_{P(\delta_1,\delta_2)})_{\min} < MSE(\hat{F}^*_{KU})_{\min}$ if:

$$\left[\left\{(\Lambda^*_{200}+\Lambda^*_{020}-2\Lambda^*_{110})-\frac{(\Lambda^*_{011}-\Lambda^*_{101})^2}{\Lambda^*_{002}}\right\}\right.$$
$$\left.-\left(\frac{l_2+l_3+l_4}{l_1}\right)\right]>0.$$

5. Numerical study

We use the following three data sets for numerical study.

Population 1. Source: Singh [29]
Let Y, X, and Z be the number of immigrants admitted in the USA during 1996, 1995, and 1994 respectively. Let $I(y_i \le t_y) = 1$ for $t_y = 17702.76$ and $I(y_i > t_y) = 0$, otherwise; $I(x_i \le t_x) = 1$ for $t_x = 13903.24$ and $I(x_i > t_x) = 0$, otherwise; $I(z_i \le t_z) = 1$ for $t_z = 15483.67$ and $I(z_i > t_z) = 0$, otherwise. Last 25% observations i.e., 13 units are considered as non-responding units. $N = 51$, $n = 20$, $F_{Y(t_y)} = 0.8039$, $F_{X(t_x)} = 0.7647$, $F_{Z(t_z)} = 0.8039$, $S^2_{F_{Y(t_y)}} = 0.1576$, $S^2_{F_{X(t_x)}} = 0.1799$, $S^2_{F_{Z(t_z)}} = 0.1576$, $N_{110} = 39$, $N_{120} = 02$, $N_{210} = 00$, $N_{220} = 10$, $N_{101} = 40$, $N_{102} = 01$, $N_{201} = 01$, $N_{202} = 09$, $N_{011} = 39$, $N_{012} = 00$, $N_{021} = 02$, $N_{022} = 10$.

For non-response, we have: $N_2^{(2)} = 13$, $F^{(2)}_{Y(t_y)} = 0.7692$, $F^{(2)}_{X(t_x)} = 0.6923$, $F^{(2)}_{Z(t_z)} = 0.7692$, $S^{2(2)}_{F_{Y(t_y)}} = 0.1775$, $S^{2(2)}_{F_{X(t_x)}} = 0.2130$, $S^{2(2)}_{F_{Z(t_z)}} = 0.1775$, $N^{(2)}_{110} = 09$, $N^{(2)}_{120} = 01$, $N^{(2)}_{210} = 00$, $N^{(2)}_{220} = 03$, $N^{(2)}_{101} = 09$, $N^{(2)}_{102} = 01$, $N^{(2)}_{201} = 01$, $N^{(2)}_{202} = 02$, $N^{(2)}_{011} = 09$, $N^{(2)}_{012} = 00$, $N^{(2)}_{021} = 01$, $N^{(2)}_{022} = 03$.

Population 2. Source: Gujarati and Porter [30]

Let Y, X, and, Z be the production of eggs in USA during 1992, 1991, and 1990 respectively.

Let $I(y_i \le t_y) = 1$ for $t_y = 1377.854$ and $I(y_i > t_y) = 0$, otherwise; $I(x_i \le t_x) = 1$ for $t_x = 75.872$ and $I(x_i > t_x) = 0$, otherwise; $I(z_i \le t_z) = 1$ for $t_z = 78.276$ and $I(z_i > t_z) = 0$, otherwise. Last 25% observations i.e., 13 units are considered as non-responding units. $N = 50$, $n = 18$, $F_{Y(t_y)} = 0.6600$, $F_{X(t_x)} = 0.5800$, $F_{Z(t_z)} = 0.5800$, $S^2_{F_{Y(t_y)}} = 0.2244$, $S^2_{F_{X(t_x)}} = 0.2436$, $S^2_{F_{Z(t_z)}} = 0.2436$, $N_{110} = 17$, $N_{120} = 16$, $N_{210} = 12$, $N_{220} = 05$, $N_{101} = 17$, $N_{102} = 16$, $N_{201} = 12$, $N_{202} = 05$, $N_{011} = 28$, $N_{012} = 01$, $N_{021} = 01$, $N_{022} = 20$.

For nonresponse, we have: $N_2^{(2)} = 13$, $F^{(2)}_{Y(t_y)} = 0.7692$, $F^{(2)}_{X(t_x)} = 0.5385$, $F^{(2)}_{Z(t_z)} = 0.6154$, $S^{2(2)}_{F_{Y(t_y)}} = 0.1775$, $S^{2(2)}_{F_{X(t_x)}} = 0.2485$, $S^{2(2)}_{F_{Z(t_z)}} = 0.2366$, $N^{(2)}_{110} = 04$, $N^{(2)}_{120} = 06$, $N^{(2)}_{210} = 03$, $N^{(2)}_{220} = 00$, $N^{(2)}_{101} = 05$, $N^{(2)}_{102} = 05$, $N^{(2)}_{201} = 03$, $N^{(2)}_{202} = 00$, $N^{(2)}_{011} = 07$, $N^{(2)}_{012} = 00$, $N^{(2)}_{021} = 01$, $N^{(2)}_{022} = 05$.

Population 3. Source: Singh [29]

Let Y, X, and Z be the estimated number of fish caught by marine recreational fisherman by species group during 1995, 1994, and 1993 respectively.

Let $I(y_i \le t_y) = 1$ for $t_y = 4514.90$ and $I(y_i > t_y) = 0$, otherwise; $I(x_i \le t_x) = 1$ for $t_x = 4954.43$ and $I(x_i > t_x) = 0$, otherwise; $I(z_i \le t_z) = 1$ for $t_z = 4591.07$ and $I(z_i > t_z) = 0$, otherwise. Last 25% observations i.e., 17 units are considered as non-responding units.

$N = 69$, $n = 23$, $F_{Y(t_y)} = 0.7246$, $F_{X(t_x)} = 0.7681$, $F_{Z(t_z)} = 0.7391$, $S^2_{F_{Y(t_y)}} = 0.1995$, $S^2_{F_{X(t_x)}} = 0.1781$, $S^2_{F_{Z(t_z)}} = 0.1928$, $N_{110} = 47$, $N_{120} = 03$, $N_{210} = 06$, $N_{220} = 13$, $N_{101} = 48$, $N_{102} = 02$, $N_{201} = 03$, $N_{202} = 16$, $N_{011} = 49$, $N_{012} = 04$, $N_{021} = 02$, $N_{022} = 14$.

For nonresponse, we have:
$N_2^{(2)} = 17$, $F^{(2)}_{Y(t_y)} = 0.8824$, $F^{(2)}_{X(t_x)} = 0.8824$, $F^{(2)}_{Z(t_z)} = 0.8824$, $S^{2(2)}_{F_{Y(t_y)}} = 0.1038$, $S^{2(2)}_{F_{X(t_x)}} = 0.1038$, $S^{2(2)}_{F_{Z(t_z)}} = 0.1038$, $N^{(2)}_{110} = 15$, $N^{(2)}_{120} = 00$, $N^{(2)}_{210} = 00$, $N^{(2)}_{220} = 02$, $N^{(2)}_{101} = 15$, $N^{(2)}_{102} = 00$, $N^{(2)}_{201} = 00$, $N^{(2)}_{202} = 02$, $N^{(2)}_{011} = 15$, $N^{(2)}_{012} = 00$, $N^{(2)}_{021} = 00$, $N^{(2)}_{022} = 02$.

The MSE values of all estimators based on three populations are given in Tables 7–9.

From Tables 7–9, we observed that the proposed general class of estimators $\hat{F}^*_{P(\delta_1,\delta_2)}$ is performing better than all considered estimators at different choices of K.

6. Conclusion

We proposed a general class of Distribution Function (DF) estimators $\hat{F}^*_{P(\delta_1,\delta_2)}$ using two auxiliary variables under non-response in simple random sampling. It is clear from Tables 7–9, that the proposed general class of estimators $\hat{F}^*_{P(\delta_1,\delta_2)}$ for different values of K, is more efficient as compared to

Table 7. MSE values of different estimators for different values of K in Population 1.

Estimator	$K = 1.5$	$K = 2.0$	$K = 2.5$	$K = 3.0$	$K = 3.5$
$\hat{F}_0^*$	0.005922	0.007054	0.008185	0.009316	0.010448
$\hat{F}_{R_1}^*$	0.001744	0.002235	0.002726	0.003217	0.003708
$\hat{F}_{E_1}^*$	0.001947	0.002383	0.002820	0.003256	0.003692
$\hat{F}_{D_1}$	0.001369	0.001742	0.002113	0.002484	0.002853
$\hat{F}_{Rao}^*$	0.001366	0.001737	0.002106	0.002474	0.002841
$\hat{F}_{GS}^*$	0.001361	0.001729	0.002095	0.002459	0.002822
$\hat{F}_{R_2}^*$	0.009717	0.012198	0.014679	0.017160	0.019640
$\hat{F}_{E_2}^*$	0.001523	0.002136	0.002749	0.003362	0.003975
$\hat{F}_{D_2}^*, \hat{F}_{CH}^*, \hat{F}_{SU}^*$	0.001311	0.001726	0.002112	0.002481	0.002840
$\hat{F}_{KU}^*$	0.001567	0.001935	0.002293	0.002643	0.002983
$\hat{F}_{P(0,0)}^*$	0.001308	0.001721	0.002105	0.002472	0.002828
$\hat{F}_{P(1,0)}^*$	0.001308	0.001721	0.002105	0.002472	0.002827
$\hat{F}_{P(1,1)}^*$	0.001307	0.001722	0.002107	0.002475	0.002832
$\hat{F}_{P(0.5,0.5)}^*$	0.001304	0.001716	0.002097	0.002461	0.002814
$\hat{F}_{P(0,1)}^*$	0.001303	0.001713	0.002094	0.002457	0.002809

Table 8. MSE values of different estimators for different values of K in Population 2.

Estimator	$K = 1.5$	$K = 2.0$	$K = 2.5$	$K = 3.0$	$K = 3.5$
$\hat{F}_0^*$	0.009261	0.010543	0.011825	0.013107	0.014390
$\hat{F}_{R_1}^*$	0.028014	0.033371	0.038728	0.044085	0.049442
$\hat{F}_{E_1}^*$	0.015253	0.017991	0.020730	0.023468	0.026207
$\hat{F}_{D_1}$	0.008759	0.009779	0.010781	0.011772	0.012756
$\hat{F}_{Rao}^*$	0.008586	0.009564	0.010521	0.011463	0.012393
$\hat{F}_{GS}^*$	0.008513	0.009468	0.010399	0.011313	0.012212
$\hat{F}_{R_2}^*$	0.070830	0.083730	0.096630	0.109540	0.122440
$\hat{F}_{E_2}^*$	0.027187	0.032177	0.037166	0.042156	0.047146
$\hat{F}_{D_2}^*, \hat{F}_{CH}^*, \hat{F}_{SU}^*$	0.008754	0.009776	0.010780	0.011772	0.012755
$\hat{F}_{KU}^*$	0.011939	0.014108	0.016271	0.018429	0.020586
$\hat{F}_{P(0,0)}^*$	0.008582	0.009562	0.010520	0.011462	0.012392
$\hat{F}_{P(1,1)}^*$	0.008576	0.009554	0.010509	0.011447	0.012373
$\hat{F}_{P(1,0)}^*$	0.008721	0.009740	0.010740	0.011729	0.012709
$\hat{F}_{P(0.5,0.5)}^*$	0.008525	0.009488	0.010425	0.011345	0.012251
$\hat{F}_{P(0,1)}^*$	0.008509	0.009488	0.010425	0.011345	0.012251

Table 9. MSE values of different estimators for different values of K in Population 3.

Estimator	$K = 1.5$	$K = 2.0$	$K = 2.5$	$K = 3.0$	$K = 3.5$
$\hat{F}_0^*$	0.006340	0.006896	0.007451	0.008007	0.008563
$\hat{F}_{R_1}^*$	0.003569	0.003570	0.003572	0.003573	0.003574
$\hat{F}_{E_1}^*$	0.003682	0.003837	0.003992	0.004147	0.004302
$\hat{F}_{D_1}$	0.003305	0.003342	0.003373	0.003399	0.003421
$\hat{F}_{Rao}^*$	0.003284	0.003321	0.003351	0.003377	0.003399
$\hat{F}_{GS}^*$	0.003275	0.003311	0.003340	0.003365	0.003386
$\hat{F}_{R_2}^*$	0.007953	0.008426	0.008900	0.009374	0.009848
$\hat{F}_{E_2}^*$	0.002231	0.002232	0.002232	0.002233	0.002233
$\hat{F}_{D_2}^*, \hat{F}_{CH}^*, \hat{F}_{SU}^*$	0.001928	0.001936	0.001943	0.001949	0.001954
$\hat{F}_{KU}^*$	0.003471	0.003473	0.003474	0.003476	0.003477
$\hat{F}_{P(0,0)}^*$	0.001921	0.001929	0.001936	0.001942	0.001947
$\hat{F}_{P(1,1)}^*$	0.001921	0.001919	0.001936	0.001942	0.001947
$\hat{F}_{P(1,0)}^*$	0.001928	0.001936	0.001943	0.001949	0.001953
$\hat{F}_{P(0.5,0.5)}^*$	0.001917	0.001925	0.001931	0.001937	0.001941
$\hat{F}_{P(0,1)}^*$	0.001916	0.001923	0.001929	0.001935	0.001939

the estimators $\hat{F}_i^*(i = 0, R_1, E_1, D_1, Rao, GS, R_2, E_2, (D_2, Ch, SU), KU)$ when non-response exists on all the study variable (Y) and the auxiliary variables (X, Z). It is also observed that the Mean Square Error (MSE) values of all estimators increase with increase in the values of K from 1.5 to 3.5 in all Populations 1–3, which are expected results. The ratio estimator $\hat{F}_{R_2}^*$ shows poor performance in Tables 7 and 9 but in Table 8, the ratio, exponential-ratio and Kumar et al. [7] estimators i.e. $\hat{F}_i^*(i = R_1, R_2, E_1, E_2, K)$ perform poorly as compared to all other estimators. The difference estimator $(\hat{F}_{D_2}^*)$, Chami et al. [27] estimator $(\hat{F}_{Ch}^*)$ and Singh and Usman [28] estimator $(\hat{F}_{SU}^*)$ give the same Mean Square Error (MSE) values. Among proposed general class of estimators $\hat{F}_{P(\delta_1, \delta_2)^*}$, the performance of the estimator $\hat{F}_{P(0,1)}^*$ is the best in terms of MSE.

Acknowledgements

Authors are thankful to the Editor and the anonymous referees for their valuable suggestions which helped improve the quality of the manuscript.

References

1. Hansen, M.H. and Hurwitz, W.N. "The problem of non-response in sample surveys", **41**(236), pp. 517–529 (1946). DOI: 10.1080/01621459.1946.10501894
2. Rao, T.J. "On certain methods of improving ratio and regression estimators", *Communications in Statistics-Theory and Methods*, **20**(10), pp. 3325–3340 (1991). DOI: 10.1080/03610929108830705
3. Kumar, S., Trehan, M., and Joorel, J.S. "A simulation study: Estimation of population mean using two auxiliary variables in stratified sampling", *Journal of Statistical Computation and Simulation*, **88**(18), pp. 3694–3707 (2018). DOI: 10.1080/00949655.2018.1532513
4. Gupta, S. and Shabbir, J. "On improvement in estimating the population mean in simple random sampling", *Journal of Applied Statistics*, **35**(5), pp. 2540–2559 (2008). DOI: 10.1080/02664760701835839
5. Khan, M. and Shabbir, J. "A general class of estimators for finite population mean using auxiliary information in the presence of nonresponse when using second raw moments", *VFAST Transactions on Mathematics*, **2**(2), pp. 19–36 (2013). DOI: 10.21015/vtm.v2i2.129
6. Verma, H.K. Sharma, P., and Singh, R. "Some ratio-cum-product type estimators for population mean under double sampling in the presence of non-response", *Journal of Statistics, Applications and Probability*, **3**(3), pp. 379–385 (2014). DOI: 10.12785/jsap/030310
7. Bhushan, S. and Kumar, A. "On cost efficient classes of estimators for population mean in presence of measurement errors and non-response simultaneously", *International Journal of Statistics and Systems*, **12**(1), pp. 93–117 (2017).
8. Kumar, S. and Bhoughal, S. "Study on nonresponse and measurement error, using double sampling scheme", *Journal of Statistics Application and Probability Letters*, **5**(1), pp. 43-52 (2018). DOI: 10.18576/jsapl/050105
9. Saleem, I., Sanaullah, A., and Hanif, M. "A generalized class of estimators for estimating population mean in the presence of nonresponse", *Journal of Statistical Theory and Application*, **17**(4), pp. 616–626 (2018). DOI: 10.2991/jsta.2018.17.4.4
10. Ahmad, S., Arslan, M., Khan, A., et al. "A generalized exponential-type estimator for population mean when

using auxiliary attribute", *Plos One*, **16**(5), e0246947, pp. 1–29 (2021). DOI: 10.1371/journal.pone.0246947

11. Waseem, Z., Khan, H., and Shabbir, J. "Generalized exponential type estimator for the mean of sensitive variable in the presence of non-sensitive auxiliary variable", *Communications in Statistics - Theory and Methods*, **50**(14), pp. 3477–3488 (2021). DOI: 10.1080/03610926.2019.1708399

12. Yaqub, M. and Shabbir, J. "Estimation of population distribution function in the presence of nonresponse", *Hacettepe Journal of Mathematics and Statistics*, **47**(2), pp. 471–511 (2018). DOI: 10.1080/03610918.2022.2078492

13. Yaqub, M. and Shabbir, J. "Estimation of population distribution function involving measurement error in the presence of nonresponse", *Communications in Statistics-Theory and Methods*, **49**(10), pp. 2540–2559 (2020). DOI: 10.1080/03610918.2022.2078492

14. Ahmed, M.S. and Abu-Dayyah, W. "Estimation of finite population distribution function using multivariate auxiliary information", *Statistics in Transition*, **5**(3), pp. 501–507 (2001). DOI: 10.1371/journal.pone.0243584

15. Wang, S. and Dorfman, A.H. "A new estimator for the population distribution function", *Biometrika*, **83**, pp. 639–652 (1996).

16. Singh, H.P., Singh, R., and Kozak, M. "A family of estimators of finite population distribution function using auxiliary information", *Acta Applied Mathematica*, **104**, pp. 115–130 (2008). DOI: 10.1007/s10440-008-9243-1

17. Munoz, J.F., Alvarez, E., and Rueda, M. "Optimum design based ratio estimators of the distribution function", *Journal of Applied Statistics*, **41**(7), pp. 1395–1407 (2013). DOI: 10.1080/02664763.2013.870983

18. Irfan, M., Javed, M., and Lin, Z. "Efficient ratio-type estimators of finite population mean based on correlation coefficient", *Scientia Iranica*, **25**(4), pp. 2361–2372 (2018). DOI: 10.24200/sci.2017.4455

19. Abid, M., Ahmad, S., Tahir, M., et al. "Improved ratio estimators of variance based on robust measures", *Scientia Iranica*, **26**(4), pp. 2484–2494 (2019). DOI: 10.24200/sci.2018.20604

20. Abid, M., Naeem, A., Hussain, Z., et al. "Investigating the impact of simple and mixture priors on estimating sensitive proportion through a general class of randomized response models", *Scientia Iranica*, **26**(2), pp. 1009–1022 (2019). DOI: 10.24200/sci.2018.20166

21. Javed, M., Irfan, M., and Pang, T. "Hartely-Ross type unbiased estimator of population mean using two auxiliary variables", *Scientia Iranica*, **26**(6), pp. 3835–3845 (2019). DOI: 10.24200/sci.2018.5648.1397

22. Naz, F., Abid, M., Nawaz, T., et al. "Enhancing efficiency of ratio-type estimators of population variance by a combination on robust location measures", *Scientia Iranica*, **27**(4), pp. 2040–2056 (2020). DOI: 10.24200/sci.2019.5633.1385

23. Younis, F. and Shabbir, J. "Estimation of general parameters under stratified adaptive cluster sampling based on dual use of auxiliary information", *Scientia Iranica*, **28**(3), pp. 1780–1801 (2021). DOI: 10.24200/sci.2019.52515.2753

24. Ahmed, S. and Shabbir, J. "On the use of ranked set sampling for estimating super population total: Gamma population model", *Scientia Iranica*, **28**(1), pp. 465–476 (2021). DOI: 10.24200/sci.2019.50976.1946

25. Nazir, H.Z., Abid, M., Akhtar, N., et al. "An efficient mixed-memory-type control chart for normal and non-normal process", *Scientia Iranica*, **28**(3), pp. 1736–1749 (2021). DOI: 10.24200/sci.2019.51437.2177

26. Chami, P., Sing, B., and Thomas, D. "A two-parameter ratio-product-ratio estimator using auxiliary information", *ISRN Probability and Statistics*, Article ID 10368, pp. 1–15 (2012). DOI: 10.5402/2012/103860

27. Guha, S. and Chandra, H. "Improved estimation of finite population mean in two-phase sampling with subsampling of the nonrespondents", *Mathematical Population Studies*, **28**(1), pp. 24–44 (2020). DOI: 10.1080/08898480.2019.1694325

28. Singh, G.N. and Usman, M. "Improved regression cum ratio estimators using information on two auxiliary variables dealing with subsampling technique of non response", *Journal Statistical Theory and Practice*, **14**(1), pp. 1–28 (2020). DOI: 10.1007/s42519-019-0082-3

29. Singh, S. "*Advanced Theory of Sampling with Applications*: How Michal selected Ammey", Kulwer Academy, London (2003). DOI: 10.1007/978-94-007-0789-4

30. Gujrati, D.H. and Porter, D.C., *Basic Econometrics*, McGraw Hill Irwin (2020).

Biographies

Javid Shabbir is working as Tenured Professor of Statistics and Dean Faculty of Natural Sciences at Quaid-i-Azam University Islamabad, Pakistan. He completed his MS degree from the University of Southampton, UK and PhD degree from the University of Kent at Canterbury, UK. His field of interests are, Survey sampling, Nonresponse, Ranked set sampling and Randomized response techniques. He published more than 200 research papers in internationally reputed journals and produced 115 MPhil and 14 PhD students. He participated in many national and international conferences. He is an associated editor of the "Journal of Statistical Theory and Practice"-Springer-Verlag.

Sat Gupta is a Professor of Statistics and Head of the Department of Mathematics and Statistics at the University of North Carolina at Greensboro. He is a Fellow of the American Statistical Association and has bagged several other awards in his career. These include the Senior University-wide Research Excellence Award, and the Senior College of Arts and Sciences Teaching Excellence Award at University of North Carolina Greensboro; Outstanding Faculty Award and the Outstanding Teacher/Scholar Award from the University of Southern Maine; the Distinguished Service Award for the Cause of Statistics given by the North Carolina Chapter of the American Statistical Association; and the Sankhyiki Bhushan Award from The Indian Society of Agricultural Statistics. He has been the founding Editor-in-Chief since 2007 of the prestigious research journal 'Journal of Statistical Theory and Practice' published by Springer. Professor Sat Gupta had his college education from Delhi University from where he earned a PhD degree in mathematics in 1977. After teaching assignments at Delhi University Professor Gupta came to Colorado State University, USA in 1982 from where he earned a PhD degree in statistics in 1987.