Modified estimators of finite population distribution function based on dual use of auxiliary information under stratified random sampling

Sardar Hussain1, Sohail Akhtar2 and Mahmoud El-Morshedy3,4
1Department of Statistics, Quaid-i-azam University, Islamabad, Pakistan
2Department of Mathematics and Statistics, The University of Haripur, Haripur, Pakistan
3Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
4Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, Egypt

Abstract
In survey sampling, information on auxiliary variables related to the main variable is often available in many practical problems. Since the mid-twentieth century, researchers have taken a keen interest in the use of auxiliary information due to its usefulness in estimation methods. The current study presents two new estimators for the distribution function of a finite population based on dual auxiliary variables. The new estimators can be used in situations where the researchers face some sort of complex data set. The mathematical equations for the bias and mean square error have been obtained for each proposed estimator. Besides, an empirical study simulation study has also been conducted to analyse the performance of estimators. It is found that the new suggested estimators of the distribution function of a finite population are more accurate than some of the existing estimators.

Keywords
Stratified random sampling, CDF, ratio in regression type exponential estimator, auxiliary variables, Bias, MSE, PRE

Corresponding author:
Sohail Akhtar, Associate Professor in Statistics, University of Haripur, Haripur, KP, Pakistan.
Email: Akhtar013@gmail.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Introduction

Many researchers have studied the use of auxiliary variables in the literature of survey sampling to increase the efficiency of their developed estimators for estimating common parameters like mean, median, variance, and standard deviation. Traditional ratio, product, and regression estimators provide efficient results for unknown parameters in such circumstances.

Out of many practices, the ratio method and product method has been widely used for estimating unknown population parameters, when there is a high positive and a high negative correlation between study variable and auxiliary variable. In the past, several authors introduced many ratio type and product type estimators by using different type of linear transformation of original auxiliary variables. The drawback of these class of estimators are that it uses a very specific linear transformation of auxiliary variable that restrict the scope of applications of this class, in practice. To overcome this drawback, we here propose a generalized class of ratio in regression type exponential estimators of population distribution function under a very general linear transformation of auxiliary variable.

Simple random sampling works quite well if the population of interest is homogeneous. When the population of interest is heterogeneous, however, it is preferable to apply stratified random sampling rather than simple random sampling. In stratified random sampling, we divide the entire aggregate into numerous non overlapping groups or subgroups called strata. These groupings are completely homogeneous, and a sample is taken from each stratum separately. The values of the Nh must be known in order to get the most out of stratification. After the strata have been determined, a sample is taken from each stratum, and the drawings are done separately. The entire technique is represented as stratified sampling if a simple random sample is collected from each stratum. To divide the sample into strata, different researchers utilized different sample allocation procedures. If the sample size in each stratum is large enough, using a distinct ratio estimate in each stratum is more precise. As a result, we’ll apply the proportional allocation strategy in this article. The population mean under stratified random sampling has received more attention. Stratification enhances efficiency, when the variance between strata is substantially greater than the variance within strata. The problem of measuring the function of finite population cumulative distribution (CDF) arises when interest lies in knowing the proportion of study variables that are below or equal to a certain value. The need for CDF in many situations is greater than ever. For example, a physician could be interested in knowing what percentage of the population consumes 35% or more of their calories from trans fats. A soil scientist, for example, could be interested in determining the clay percentage distribution in the soil. In addition, policymakers may be curious about the percentage of people living in a developing country who are poor. The CDF has been computed using information on one or more auxiliary variables in survey sample literature. Chambers and Dunstan, Rao et al., Rao, Kuk, Ahmed and Abu-Dayyeh, Rueda et al., Singh et al., Hussain et al., and Hussain et al. proposed two new estimators for estimating the finite population distribution function using supplementary information using simple and stratified random sampling schemes. In practice, there needs to be more research on the use of both auxiliary variables and the finite population distribution function.
In survey sampling literature, the authors have estimated finite population distribution function (CDF) using on one or more auxiliary variable. Dual use of auxiliary variable has been rarely attempted while estimating finite population distribution function, therefore we motivated towards it. In this article we proposed two new estimators which are competing the existing estimators and estimators proposed by Hussain et al. \(^9\)

The paper offers two new estimators for estimating finite population distribution functions under stratified random sampling that leverage dual usage of auxiliary information. The bias and mean square error of the proposed estimators have been expressed up to first order of approximation. Cochran, \(^{10}\) Murthy, \(^{11}\) Bahl and Tuteja, \(^{12}\) Rao, \(^{13}\) Singh and Kumar, \(^{14}\) Grover and Kaur, \(^{15}\) and Hussain et al. \(^{9}\) have all demonstrated that the proposed estimators are more efficient than traditional unbiased estimators, both theoretically and empirically.

The problem of estimating the finite population 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 situations where estimating the CDF is deemed necessary. For example, for a nutritionist, it is interesting to know the proportion of population that consumes 25% or more of the calorie intake from saturated fat. Similarly, a soil scientist may be interested in estimating the distribution of clay percent in the soil.

In addition, policymakers may be interested in knowing the proportion of people living in a developing country below the poverty line.

**Sampling design and notations**

When the population is heterogeneous, stratified random sampling should be used instead of simple random sample. In stratified random sampling, we split the diverse population into a number of non-overlapping groups or subgroups termed strata. These groupings are completely homogeneous, and a sample is taken from each stratum separately. To disperse the samples in the strata, surveyors employ a variety of sample allocation techniques. If the sample size in each stratum is big enough, using independent ratio estimates in each stratum is more precise. As a result, we will apply the proportional allocation method in this article. Many publications offered many ratio type estimators in stratified sampling by changing the auxiliary variable, such as Kadilar and Cingi, \(^{16}\) Kadilar and Cingi, \(^{17}\) Koyuncu and Kadilar, \(^{18}\) Shabbir and Gupta, \(^{19}\) Aladag and Cingi, \(^{20}\) Malik and Singh. \(^{21}\)

Let \(\Omega = \{1, 2, \ldots, N\}\) be a finite population of \(N\) units, which is divided into \(L\) homogeneous strata, where the siRe of \(h\)th stratum is \(N_h\), for \(h = 1, 2, \ldots, L\), in such manner \(\sum_{h=1}^{L} N_h = N\). Assume that \(Y\) and \(X\) be the study \(y_h\) and auxiliary variable \(x_h\), where \(i = 1, 2, \ldots, N_h\) and \(h = 1, 2, \ldots, L\), a sample \(n_h\) is drawn in such a manner \(\sum_{h=1}^{L} n_h = n\), where \(n\) is the sample siRe.

Let \(F_{st}(ty) = F(ty) = \sum_{h=1}^{L} W_h F_h(ty)\) and \(F_{st}(tx) = F(tx) = \sum_{h=1}^{L} W_h F_h(tx)\), \(\hat{F}_{st}(ty) = \hat{F}(ty) = \sum_{h=1}^{L} W_h \hat{F}_h(ty)\) and \(\hat{F}_{st}(tx) = \hat{F}(tx) = \sum_{h=1}^{L} W_h \hat{F}_h(tx)\) be the population and sample distribution functions of \(Y\) and \(X\) under stratified random sampling, respectively, where \(W_h = N_h / N\), \(F_h(ty) = \sum_{i=1}^{N_h} I(Y_{ih} \leq ty) / N_h\), \(\hat{F}_h(y) = \sum_{i=1}^{n_h} I(Y_{ih} \leq ty) / n_h\), \(F_h(tx) = \sum_{i=1}^{N_h} I(X_{ih} \leq tx) / N_h\), \(\hat{F}_h(tx) = \sum_{i=1}^{n_h} I(X_{ih} \leq tx) / n_h\). Let \(\bar{X}_{st} = \bar{X} = \sum_{h=1}^{L} W_h \bar{X}_h\).
\[ \sum_{h=1}^{L} W_h \hat{X}_h \] and \( \hat{R}_st = \hat{R} = \sum_{h=1}^{L} W_h \hat{R}_h \), \( \hat{X}_st = \hat{X} = \sum_{h=1}^{L} W_h \hat{X}_h \) and \( \hat{R}_st = \hat{R} = \sum_{h=1}^{L} W_h \hat{R}_h \) be the population and sample means of \( X \) and \( Z \) under stratified random sampling, respectively, where, \( \hat{X}_h = \sum_{i=1}^{N_h} X_{ih}/N_h \), \( \hat{R}_h = \sum_{i=1}^{N_h} Z_{ih}/N_h \), \( \hat{X}_st = \sum_{i=1}^{N_h} X_{ih}/n_h \), \( \hat{R}_st = \sum_{i=1}^{N_h} Z_{ih}/n_h \).

To find the properties of the existing and proposed estimators of \( F(t) \), we consider the following relative error terms under stratified random sampling. Let

\[ \nu_1 = \frac{\hat{F}_{st}(ty) - F(ty)}{F(ty)} \text{, } \nu_2 = \frac{\hat{F}_{st}(tx) - F(tx)}{F(tx)} \text{, } \nu_3 = \frac{\hat{X}_{st} - \bar{X}}{\bar{X}} \text{ and } \nu_4 = \frac{\hat{R}_{st} - \bar{R}}{\bar{R}}, \]

such that \( E(\nu_i) = 0 \) for \( i = 1, 2, 3, 4 \), where \( E(\cdot) \) is the mathematical expectation of \( (\cdot) \).

Let \( V_{rstu} = E[\nu_1 \nu_2^* \nu_3 \nu_4^u] \),

where

\[
E(\nu_1^2) = \sum_{h=1}^{L} W_h^2 \hat{X}_h^2 \hat{R}_{F_{th}} C_{F_{th}} = \mu_{2000}, \quad E(\nu_2^2) = \sum_{h=1}^{L} W_h^2 \hat{X}_h^2 \hat{R}_{F_{tx}} C_{F_{tx}} = \mu_{0200}, \\
E(\nu_3^2) = \sum_{h=1}^{L} W_h^2 \hat{X}_h^2 C_{Xh} = \mu_{0020}, \quad E(\nu_4^2) = \sum_{h=1}^{L} W_h^2 \hat{X}_h^2 \hat{R}_{R_{th}} C_{R_{th}} = \mu_{0002},
\]

\[
E(\nu_1 \nu_2) = \sum_{h=1}^{L} W_h^2 \hat{X}_h \hat{R}_{F_{th}} F_{th} C_{F_{th}} C_{F_{th}} = \mu_{1100}, \quad E(\nu_1 \nu_3)
\]

\[
= \sum_{h=1}^{L} W_h^2 \hat{X}_h \hat{R}_{F_{th}} F_{th} C_{F_{th}} C_{xh} = \mu_{1010}, \quad E(\nu_2 \nu_3)
\]

\[
= \sum_{h=1}^{L} W_h^2 \hat{X}_h \hat{R}_{F_{th}} C_{F_{th}} C_{xh} = \mu_{0110}, \quad E(\nu_1 \nu_4)
\]

\[
= \sum_{h=1}^{L} W_h^2 \hat{X}_h \hat{R}_{F_{th}} \hat{R}_{th} C_{F_{th}} C_{R_{th}} = \mu_{1001},
\]

\[
E(\nu_2 \nu_4) = \sum_{h=1}^{L} W_h^2 \hat{X}_h \hat{R}_{F_{th}} \hat{R}_{th} C_{F_{th}} C_{R_{th}} = \mu_{0101}.
\]

**Existing estimators**

Several approximations of the finite population distribution function under stratified random sampling are described in this section. In all these estimators (Existing) the authors use single auxiliary information (Variable), except Hussain et al.\(^9\) Under the
first order of approximation, the biases and MSEs of these adapted estimators are calculated.

**Mean estimator**

The traditional unbiased mean estimator $F(ty)$, is given by

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

The variance of $\hat{F}_{SRS}^*(ty)$, is given by

$$
\text{Var}(\hat{F}_{SRS}^*(ty)) = F^2(ty)\mu_{2000}.
$$

Cochran\textsuperscript{10}

In stratified random sampling, the usual ratio estimator $F(ty)$ is provided by

$$
\hat{F}_{R}^*(ty) = \hat{F}(ty)_{st} \left( \frac{F(tx)}{\hat{F}(tx)_{st}} \right).
$$

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

$$
\text{Bias}(\hat{F}_{R}^*(ty)) \cong F(ty)(\mu_{0200} - \mu_{1100}),
$$

$$
\text{MSE}(\hat{F}_{R}^*(ty)) \cong F^2(ty)(\mu_{0200} + \mu_{0200} - 2\mu_{1100}).
$$

If $R_{F,ty,F,x} > C_{F,x}/(2C_{F,ty})$, then $\hat{F}_{R}^*(ty)$ is better than $\hat{F}_{SRS}^*(ty)$ in terms of MSE.

Murthy\textsuperscript{11}

Suggested the usual product estimator $F(ty)$ in stratified random sampling, is given by

$$
\hat{F}_{P}^*(ty) = \hat{F}(ty)_{st} \left( \frac{\hat{F}(tx)_{st}}{F(tx)} \right).
$$

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

$$
\text{Bias}(\hat{F}_{P}^*(ty)) = F(ty)\mu_{1100},
$$

and

$$
\text{MSE}(\hat{F}_{P}^*(ty)) \cong F^2(ty)(\mu_{2000} + \mu_{0200} + 2\mu_{1100}).
$$

If $-C_{F,x}/(2C_{F,ty}) > \Re_{F,ty,F,x}$, then $\hat{F}_{P}^*(ty)$ is better than $\hat{F}_{SRS}^*(ty)$ in terms of MSE.

Bahl and Tuteja\textsuperscript{12}

Presented combined ratio and product-type exponential estimators of $F(ty)$, in stratified random sampling, is given by

$$
\hat{F}_{BT,R}^*(ty) = \hat{F}(ty)_{st} \exp \left( \frac{F(tx) - \hat{F}(tx)_{st}}{F(tx) + \hat{F}(tx)_{st}} \right),
$$

$$
\hat{F}_{BT,P}^*(ty) = \hat{F}(ty)_{st} \exp \left( \frac{\tilde{F}(tx)_{st} - F(tx)}{F(tx) + \hat{F}(tx)_{st}} \right).
$$
The biases and MSEs of $\hat{F}_{BT,R}^*(ty)$ and $\hat{F}_{BT,P}^*(ty)$, are given by

$$\text{Bias}(\hat{F}_{BT,R}^*(ty)) \cong F(ty) \left( \frac{3}{8} \mu_{0200} - \frac{1}{2} \mu_{1100} \right).$$

and

$$\text{MSE}(\hat{F}_{BT,R}^*(ty)) \cong \frac{F^2(ty)}{4} (4 \mu_{2000} + \mu_{0200} - 4 \mu_{1100}),$$

and

$$\text{Bias}(\hat{F}_{BT,P}^*(ty)) \cong F(ty) \left( \frac{1}{2} \mu_{1100} - \frac{1}{8} \mu_{0200} \right),$$

and

$$\text{MSE}(\hat{F}_{BT,P}^*(ty)) \cong \frac{F^2(ty)}{4} (4 \mu_{2000} + \mu_{0200} + 4 \mu_{1100}).$$

**Regression estimator**

The usual regression estimator $\hat{F}_{Reg}(ty)$, in stratified random sampling, is given by

$$\hat{F}_{Reg}^*(ty) = \hat{F}(ty)_{st} + k(F(tx) - \hat{F}(tx)_{st}),$$

(11)

where $k$ is an appropriate chosen constant. The minimum variance of $\hat{F}_{Reg}^*(ty)$ at the optimum value $k_{\text{opt}} = (F(ty) \mu_{1100})/(F(tx) \mu_{0200})$ is

$$\text{Var}_{\text{min}}(\hat{F}_{Reg}^*(ty)) = \frac{F^2(ty)(\mu_{2000} \mu_{0200} - \mu_{1100}^2)}{\mu_{0200}}.$$ 

(12)

Here (12) may be written as

$$\text{Var}_{\text{min}}(\hat{F}_{Reg}^*(ty)) = F^2(ty) \mu_{2000} (1 - R_{F,oh}^2 R_{F,ah}).$$

(13)

Rao\textsuperscript{13}

Suggested an improved difference-type estimator $F(ty)$, in stratified random sampling, is given by

$$\hat{F}_{R,D}^*(ty) = k_1 \hat{F}(ty)_{st} + k_2 (F(tx) - \hat{F}(tx)_{st}),$$

(14)

where $k_1$ and $k_2$ are constants that are unknown. To the first order of approximation, the bias and MSE of $\hat{F}_{R,D}^*(ty)$ are given by

$$\text{Bias}(\hat{F}_{R,D}^*(ty)) = F(ty)(k_1 - 1),$$

and

$$\text{MSE}(\hat{F}_{R,D}^*(ty)) \cong F^2(ty) - 2k_1 F^2(ty) + k_1^2 F^2(ty) + k_1^2 F^2(ty) \mu_{2000} - 2k_1 k_2 F(ty) F(tx) \mu_{1100} + k_2^2 F^2(tx) \mu_{0200}. $$

(15)
The ideal values of \( k_1 \) and \( k_2 \), as obtained by minimizing (15), are \( k_1 \) and \( k_2 \), respectively

\[
k_{1(\text{opt})} = \frac{\mu_{0200}}{\mu_{0200}\mu_{200} - \mu_{1100}^2 + \mu_{0200}}
\]

\[
k_{2(\text{opt})} = \frac{F(ty)\mu_{1100}}{F(tx)(\mu_{0200}\mu_{200} - \mu_{1100}^2 + \mu_{0200})}.
\]

The minimum MSE of \( \hat{F}_{R,D}^*(ty) \) at the optimum values of \( k_1 \) and \( k_2 \) is

\[
\text{MSE}_{\text{min}}(\hat{F}_{R,D}^*(ty)) = \frac{F^2(ty)(\mu_{0200}\mu_{200} - \mu_{1100}^2)}{(\mu_{0200}\mu_{200} - \mu_{1100}^2 + \mu_{0200})}.
\]

(16)

Here (16) may be written as

\[
\text{MSE}_{\text{min}}(\hat{F}_{R,D}^*(ty)) = \frac{F^2(ty)(1 - \Re^2 F_{F_{th}F})}{1 + \mu_{0200}(1 - \Re^2 F_{F_{th}F})}.
\]

(17)

Suggested generalized ratio-type exponential estimator of \( F(ty) \), in stratified random sampling, is given by

\[
\hat{F}_S^*(ty) = \hat{F}(ty)_{st} \exp \left( \frac{a(F(tx) - \hat{F}(tx)_{st})}{a(F(tx) + \hat{F}(tx)_{st}) + 2b} \right),
\]

(18)

where, \( a = 1 \) and \( b = 0 \). The bias and MSE of \( \hat{F}_S^*(ty) \), to the first order of approximation, are given by

\[
\text{Bias}(\hat{F}_S^*(ty)) \approx F(ty) \left( \frac{3}{8} \theta^2 \mu_{0200} - \frac{1}{2} \theta \mu_{1100} \right),
\]

and

\[
\text{MSE}(\hat{F}_S^*(ty)) \approx \frac{F^2(ty)}{4} (4\mu_{0200} + \theta^2 \mu_{0200} - 4\theta \mu_{1100}),
\]

(19)

where \( \theta = aF(tx)/(aF(tx) + b) \).

Grover and Kaur\(^{15}\)

Introduced ratio-type exponential estimator of \( F(ty) \), in stratified random sampling, is given by

\[
\hat{F}_{G,K}^*(ty) = \{k_3 \hat{F}(ty)_{st} + k_4(F(tx) - \hat{F}(tx)_{st})\} \exp \left( \frac{a(F(tx) - \hat{F}(tx)_{st})}{a(F(tx) + \hat{F}(tx)_{st}) + 2b} \right),
\]

(20)

where the constants \( k_3 \) and \( k_4 \) are unknown. To the first order of approximation, the bias and MSE of \( \hat{F}_{G,K}^*(ty) \) are given by

\[
\text{Bias}(\hat{F}_{G,K}^*(ty)) \approx F(ty)(k_3 - 1) + \frac{3}{8} \theta^2 k_3 F(ty) + \frac{1}{2} \theta k_4 F(tx) \mu_{0200} - \frac{1}{2} \theta F(ty) \mu_{1100},
\]
and

\[
\text{MSE}(\hat{F}_{G,K}(ty)) \cong k_3^2F^2(tx)\mu_{0200} + k_4^2F^2(ty)\mu_{2000} + 2\theta k_3k_4F(ty)F(tx)\mu_{0200} - 2k_3k_4F(ty)F(tx)\mu_{1100} + F^2(ty) - 2k_3F^2(ty) + \theta k_3^2F^2(ty) + k_3F^2(ty)\mu_{1100} - \theta k_4F(ty)F(tx)\mu_{0200} - 2\theta k_3^2F^2(ty)\mu_{1100} - \frac{3}{4}\theta^2k_3F^2(ty)\mu_{0200} + \theta^2k_3^2F^2(ty)\mu_{0200}.
\] (21)

The minimum values of \(k_3\) and \(k_4\) are,

\[
k_{3(\text{opt})} = \frac{\mu_{0200}(\theta^2\mu_{0200} - 8)}{8(-\mu_{2000}\mu_{0200} + \mu^2_{1100} - \mu_{0200})},
\]

\[
k_{4(\text{opt})} = \frac{F(ty)(\theta^2\mu_{0200}^2 - \theta^2\mu_{0200}\mu_{1100} + 4\theta\mu_{2000}\mu_{0200} - 4\theta\mu^2_{1100} - 4\theta\mu_{0200} + 8\mu_{1100})}{8F(tx)(\mu_{2000}\mu_{0200} - \mu^2_{1100} + \mu_{0200})}.
\]

The simplified minimum MSE of \(\hat{F}_{G,K}(ty)\), at the optimum values of \(k_3\) and \(k_4\), is given by

\[
\text{MSE}_{\text{min}}(\hat{F}_{G,K}(ty)) \cong \text{Var}_{\text{min}}(\hat{F}_{\text{Reg}}(ty))
- \frac{F^2(ty)(\theta^2\mu_{0200}^2 - 8\mu_{1100}^2 + 8\mu_{0200}\mu_{2000})^2}{64\mu_{0200}^2 \{1 + \mu_{2000}(1 - \Re_{F_{\text{Reg}}F_{\text{Est}}})\}},
\] (22)

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

Hussain et al.\textsuperscript{9}

The first usual family of estimators for estimating \(F(ty)\), in stratified random sampling, is given by

\[
\hat{F}_{H1}(ty) = \left\{ k_5\hat{F}(ty)_s + k_6 \left( \frac{F(tx) - \hat{F}(tx)_s}{F(tx)} \right) + k_7 \left( \frac{\hat{X} - \hat{X}_s}{X} \right) \right\}
\]

\[
\exp \left( \frac{a(F(tx) - \hat{F}(tx)_s)}{a(F(tx) + \hat{F}(tx)_s) + 2b} \right).
\] (23)

The bias and mean square error of \(\hat{F}_{H1}(ty)\), are given by

\[
\text{Bias}(\hat{F}_{H1}(ty)) \cong F(ty)(k_5 - 1) + \frac{3}{8}\theta^2k_5F(ty)\mu_{0200} + \frac{1}{2}\theta k_6\mu_{0200} - \frac{1}{2}\theta k_5F(ty)\mu_{1100} + \frac{1}{2}\theta k_7\mu_{0110}.
\]
and

\[
\text{MSE}(\hat{F}_{H1}(t)) \cong F^2(t)(k_5 - 1)^2 + k_3^2 F^2(t)\mu_{2000} + k_6^2 \mu_{0200} + k_7^2 \mu_{0020} + \theta^2 k_3^2 F^2(t)\mu_{0200} \\
\quad - \theta k_6 F(t)\mu_{0200} + 2\theta k_5 k_6 F(t)\mu_{0200} - \frac{3}{4} \theta^2 k_5 F^2(t)\mu_{0200} \\
\quad + \theta k_5 F^2(t)\mu_{1100} - 2\theta k_5 F^2(t)\mu_{1100} - 2k_5 k_6 F(t)\mu_{1100} \\
\quad - 2k_5 k_7 F(t)\mu_{0110} - \theta k_7 F(t)\mu_{0110} + 2\theta k_5 k_7 F(t)\mu_{0110} - 2k_6 k_7 \mu_{0110}.
\]

(24)

where \(k_5, k_6\) and \(k_7\), are given by

\[
k_5(\text{opt}) = \frac{8 - \theta^2 \mu_{0200}}{8(1 + \mu_{2000}(1 - R_{Fg,F_{ax}}^2))},
\]

\[
k_6(\text{opt}) = \frac{8 \theta^3 \mu_{0200}^{1/2} (R_{F_{ax}}^2 - 1) + \mu_{2000}^{1/2} (-\theta^2 \mu_{0200}^{1/2})(R_{Fg,F_{ax}} - R_{F_{ah}h,F_{ax}h})}{8 \mu_{0200}^{1/2} (R_{F_{ah}h,F_{ax}h}^2 - 1) (-1 + \mu_{2000}(1 - R_{Fg,F_{ah}h}^2))},
\]

\[
k_7(\text{opt}) = \frac{F(t)^{1/2} (8 - \theta^2 \mu_{0200})(R_{F_{ah}h,F_{ax}} - R_{F_{ah}h,F_{ah}h})}{8 \mu_{0200}^{1/2} (R_{F_{ah}h,F_{ah}h}^2 - 1) (-1 + \mu_{2000}(1 - R_{Fg,F_{ah}h}^2))}.
\]

The minimum mean square error of \(\hat{F}_{H1}(t)\) at the optimum values of \(k_5, k_6\) and \(k_7\) is

\[
\text{MSE}_{\text{min}}(\hat{F}_{H1}(t)) = \frac{F^2(t)(64\mu_{2000} - R_{F_{ah}h,F_{ah}h} - \theta^4 \mu_{0200}^2 - 16\theta^2 \mu_{0200} \mu_{2000}(1 - R_{F_{ah}h,F_{ah}h}))}{64(1 + \mu_{2000}(1 - R_{F_{ah}h,F_{ah}h}))},
\]

(25)

where \(R_{F_{ah}h,F_{ah}h} = \frac{\mu_{100}^{2} \mu_{020} + \mu_{100}^{2} \mu_{020} - 2 \mu_{100} \mu_{100} \mu_{010} \mu_{010}}{\mu_{200}(\mu_{020}^{2} - \mu_{010}^{2})}\).

Here (64) may also be written as

\[
\text{MSE}_{\text{min}}(\hat{F}_{H1}(t)) \cong \text{Var}_{\text{min}}(\hat{F}_{\text{Reg}}^*(t)) - T_1 - T_2,
\]

(26)

where

\[
T_1 = \frac{F^2(t)(\theta^2 \mu_{0200}^2 - 8 \mu_{1100}^2 + 8 \mu_{0200} \mu_{2000})^2}{64 \mu_{0200}^2 (1 + \mu_{2000}(1 - R_{F_{ah}h,F_{ah}h}))},
\]

\[
T_2 = \frac{F^2(t)(\theta^2 \mu_{0200}^2 - 8 \mu_{0200} \mu_{2000}(1 - R_{F_{ah}h,F_{ah}h}))^2}{64 \mu_{0200} \mu_{0020}(1 - R_{F_{ah}h,F_{ah}h})(1 + \mu_{2000}(1 - R_{F_{ah}h,F_{ah}h})) (1 + \mu_{2000}(1 - R_{F_{ah}h,F_{ah}h}))}.
\]

It can be seen that \(\hat{F}_{H1}(t)\) is more precise than \(\hat{F}_{\text{Reg}}^*(t)\).
Hussain et al.9

The second family of estimators for estimating \( F(ty) \) in stratified random sampling, is given by

\[
\hat{F}_{H2}^*(ty) = \left\{ k_8 \hat{F}(ty)_{st} + k_9 \left( \frac{F(tx) - \hat{F}(tx)_{st}}{F(tx)} \right) + k_{10} \left( \frac{R_x - \hat{R}_{xt}}{R_x} \right) \right\}
\]

\[ \exp \left( \frac{a(F(tx) - \hat{F}(tx)_{st})}{a(F(tx) + \hat{F}(tx)_{st}) + 2b} \right), \]

where, \((a = 1)\) and \((b = 0)\). The bias and mean square error of \( \hat{F}_{H2}^*(ty) \), to the first degree of approximation, are given by

\[
\text{Bias}(\hat{F}_{H2}^*(ty)) \cong F(ty)(k_8 - 1) + \frac{3}{8} \theta^2 k_8 F(ty) \mu_{0200} + \frac{1}{2} \theta k_9 \mu_{0200} - \frac{1}{2} \theta k_8 F(ty) \mu_{1100} + \frac{1}{2} \theta k_9 \mu_{0101},
\]

and

\[
\text{MSE}(\hat{F}_{H2}^*(ty)) \cong F^2(ty)(k_8 - 1)^2 + k_8^2 F^2(ty) \mu_{2000} + k_9^2 \mu_{0200} + k_{10}^2 \mu_{0002}
\]

\[
+ \theta^2 k_8^2 F^2(ty) \mu_{0200} - \theta k_9 F(ty) \mu_{0200} + 2 \theta k_8 k_9 F(ty) \mu_{0200}
\]

\[
- \frac{3}{4} \theta^2 k_8 F^2(ty) \mu_{0200} + \theta k_8 F^2(ty) \mu_{1100} - 2 \theta k_8^2 F^2(ty) \mu_{1100}
\]

\[
- 2 k_8 k_9 F(ty) \mu_{1100} - 2 k_8 k_{10} F(ty) \mu_{1001} - \theta k_{10} F(ty) \mu_{0101} + 2 \theta k_8 k_{10} F(ty) \mu_{0101} - 2 k_9 k_{10} \mu_{0101}.
\]

The optimum values of \( k_8, k_9 \) and \( k_{10} \), determined by minimizing (28), are given by

\[
k_{8\,(opt)} = \frac{8 - \theta^2 \mu_{0200}}{8(1 + \mu_{2000}(1 - R_{F_{th},F_{th}}R_{th}))},
\]

\[
k_{9\,(opt)} = \frac{F(ty) \left[ \theta^3 \mu_{0200}^3 (R_{F_{th},R_{th}} - 1) + \mu_{2000}^{1/2} (-8 + \theta^2 \mu_{0200})(R_{F_{th},F_{th}} - R_{F_{th},R_{th}}) \right]}{8 \mu_{0200}^2 (R_{F_{th},R_{th}} - 1)(-1 + \mu_{2000} (1 - R_{F_{th},F_{th}}R_{th}))},
\]

\[
k_{10\,(opt)} = \frac{F(ty) \mu_{2000}^{1/2}(8 - \theta^2 \mu_{0200})(R_{F_{th},F_{th}} - R_{F_{th},R_{th}} R_{F_{th},R_{th}})}{8 \mu_{0200}^2 (R_{F_{th},R_{th}} - 1)(-1 + \mu_{2000} (1 - R_{F_{th},F_{th}}R_{th}))},
\]
The minimum MSE of $\hat{F}^*_H(t_y)$ at the optimum values of $k_8$, $k_9$ and $k_{10}$ is given by

$$\text{MSE}_{\text{min}}(\hat{F}^*_H(t_y)) \approx \frac{64 \mu_{2000}^2(1 - \mathcal{R}^2_{F_{t_y}F_{t_y}R_{th}}) - \theta^4 \mu_{2000}^2 - 16 \theta \mu_{2000} \mu_{2000}(1 - \mathcal{R}^2_{F_{t_y}F_{t_y}R_{th}})}{64 \{1 + \mu_{2000}(1 - \mathcal{R}^2_{F_{t_y}F_{t_y}R_{th}})\}},$$

(29)

where $\mathcal{R}^2_{F_{t_y}F_{t_y}R_{th}} = \left(\frac{\mu^2_{1100} + \mu^2_{0002} \mu_{0200} - 2 \mu_{1001} \mu_{0010}}{\mu_{2000} (\mu_{0002} - \mu_{0101})}\right)$.

Here (29) may be written as

$$\text{MSE}_{\text{min}}(\hat{F}^*_H(t_y)) \cong \text{Var}_{\text{min}}(\hat{F}^*_\text{Reg}(t_y)) - T_1 - T_3,$$

(30)

where

$$T_1 = \frac{F^2(t_y)(\theta^2 \mu_{0200}^2 - 8 \mu_{1100}^2 + 8 \mu_{0200}^2)^2}{64 \mu_{0200}^2(1 + \mu_{2000}(1 - \mathcal{R}^2_{F_{t_y}F_{t_y}R_{th}}))} \quad \text{and}$$

$$T_3 = \frac{F^2(t_y)(\theta^2 \mu_{0200}^2 - 8)^2(\mu_{0200} \mu_{1001} - \mu_{0101} \mu_{1100})^2}{64 \mu_{0200} \mu_{0002}(1 - \mathcal{R}^2_{F_{t_y}F_{t_y}R_{th}})(1 + \mu_{2000}(1 - \mathcal{R}^2_{F_{t_y}F_{t_y}R_{th}}))(1 + \mu_{2000}(1 - \mathcal{R}^2_{F_{t_y}F_{t_y}R_{th}}))}.$$

It is clear that $\hat{F}^*_H(t_y)$ is more precise than $\hat{F}^*_\text{Reg}(t_y)$.

**Proposed estimators**

The theory of stratified random sampling deals with the characteristics of the estimates with a great choice of sample size $n_h$ to get maximum precision. When the correlation exists between the study variable and the auxiliary variable, then there is a possibility that the correlation also exists between the study variable and CDF as well as the rank of the auxiliary variable.

In the literature of survey sampling consider here, the authors are used one or more auxiliary variables (Information) for estimation of finite population distribution function. Using dual auxiliary variables in the field of estimation of finite population distribution function are rarely attempted. The principal advantages of our proposed ratio-in-regression exponential type estimators under stratified random sampling are that it is more flexible and efficient existing then the existing estimators.

Motivated by Hussain et al., we propose a new ratio in a regression type exponential estimator of the finite population distribution function in stratified random sampling, including supplementary information in the form of CDF mean and rank of the auxiliary variable.

**First proposed estimator**

We use the same idea as the first proposed family of estimators Hussain et al., and estimate finite population CDF, which concerned with CDFs of study and auxiliary variables.
along with the mean of the auxiliary variable.

\[ \hat{F}_{prop}^*(ty) = k_{11} \hat{F}(ty)_{st} + k_{12} \left( \frac{F(tx) - \hat{F}(tx)_{st}}{F(tx)} \right) \exp \left( \frac{F(tx) - \hat{F}(tx)_{st}}{F(tx) + \hat{F}(tx)_{st}} \right) + k_{13} \left( \frac{\overline{X} - \tilde{X}}{\overline{X}} \right) \left\{ \exp \left( \frac{\overline{X} - \tilde{X}}{\overline{X} + \tilde{X}} \right) \right\}, \quad (31) \]

where \( k_{11}, k_{12} \) and \( k_{13} \) are suitable chosen constants. The estimator \( \hat{F}_{prop}^*(ty) \), in terms of errors, we have

\[ \hat{F}_{prop}^*(ty) = k_{11} F(ty)(1 + \nu_1) - k_{12} \nu_2 \left( 1 - \frac{1}{2} \nu_2 + \frac{3}{8} \nu_2^2 + \cdots \right) \]

\[ - k_{13} \nu_3 \left( 1 - \frac{1}{2} \nu_3 + \frac{3}{8} \nu_3^2 + \cdots \right). \quad (32) \]

Further simplifying (32), and keeping terms up to power 2, we have

\[ (\hat{F}_{prop}^*(ty) - F(ty)) = -F(ty) + k_{11} F(ty) - k_{12} \nu_2 + \frac{1}{2} k_{12} \frac{\nu_2^2}{2} - k_{13} \nu_3 \]

\[ + k_{13} \frac{\nu_3^3}{2}. \quad (33) \]

The bias and mean square error of \( \hat{F}_{prop}^*(ty) \), to the first degree of approximation, are given by

\[ \text{Bias}(\hat{F}_{prop}^*(ty)) \cong -F(ty) + k_{11} F(ty) + \frac{1}{2} k_{12} \mu_{0200} + \frac{1}{2} k_{13} \mu_{0020}, \]

\[ \text{MSE}(\hat{F}_{prop}^*(ty)) \cong \frac{F^2(ty)}{2} + k_{12} \mu_{0200} (-F(ty) + k_{12}) + 2 k_{12} k_{13} \mu_{0110} \]

\[ + k_{13} \mu_{0020} (F(ty) + k_{13}) + k_{11} k_{12} F(ty) (k_{11} - 2 + \mu_{2000}) \]

\[ + k_{11} k_{12} F(ty) \mu_{0200} - 2 \mu_{1100} + k_{11} k_{13} F(ty) \mu_{0020} - 2 \mu_{1010}. \quad (34) \]

The optimum values of \( k_{11}, k_{12} \) and \( k_{13} \), determined by minimizing (34), are given by

\[ k_{11(\text{opt})} = \frac{\left( \frac{2 \lambda_h \mu_{2000}^{1/2}}{\lambda_h^{1/2}} \right) (A_1 + B_1) + \lambda_h \left( A + \frac{\mu_{0200}}{\lambda_h} \right) - 4 \}

\[ 4 \mu_{2000} (\mathcal{R}_{Fxy} F_{ahxh} - 1) + 4 \mu_{2000}^{1/2} (A_1 + B_1) + \lambda_h \left( A + \frac{\mu_{0200}}{\lambda_h} \right) - 4. \]
The minimum MSE of \( \hat{F}_{\text{Prop1}}(t) \), at the optimum values of \( k_{11}, k_{12} \) and \( k_{13} \), is given by

\[
\text{MSE}_{\text{min}}(\hat{F}_{\text{Prop1}}(t)) = \frac{F^2(y)\mu_{2000} \{ A(\tilde{\lambda}_h) - B(\tilde{\lambda}_h) + \mu_{0200} + 4\left( R_{\text{FohFohxh}}^2 - 1 \right) \} \tilde{\lambda}_h \left( 4\mu_{2000} \left( R_{\text{FohFohxh}}^2 - 1 \right) + 4\frac{\mu_{0200}}{\tilde{\lambda}_h} (A + B_1) + A + \frac{\mu_{0200}}{\tilde{\lambda}_h} - 4 \right)}{\lambda_{h}^{1/2} F_{0020} \mu_{0200} \left( \frac{\mu_{1010}^2}{\mu_{2000}^2} - \frac{\mu_{0110}^2}{\mu_{2000}^2} \right)}.
\]

where

\[
R_{\text{FohFohxh}}^2 = \left( \frac{\mu_{1100}^2 \mu_{0020} + \mu_{1010}^2 \mu_{0020} - 2\mu_{0110} \mu_{1100} \mu_{0110}}{\mu_{2000}^2 \mu_{0200}^2 - \mu_{0110}^2} \right).
\]
A = \left\{ \frac{\mu_{0020}^{1/2}}{\lambda_h^{1/2}} - \frac{\mu_{0020}^{1/2}}{\lambda_h^{1/2}} \left( \frac{\mu_{0110}}{\sqrt{\mu_{0200} \sqrt{\mu_{0200}}}} \right) \right\}^2,

1 - \frac{\mu_{0110}^2}{\mu_{0200} \mu_{0020}},

A_1 = \frac{\mu_{0200}^{1/2}}{\lambda_h^{1/2}} \left( \frac{\mu_{1010}}{\sqrt{\mu_{2000} \sqrt{\mu_{0020}}}} - \frac{\mu_{0110}}{\sqrt{\mu_{2000} \sqrt{\mu_{0200}}}} - \frac{\mu_{1100}}{\sqrt{\mu_{2000} \sqrt{\mu_{0200}}}} \right),

B = \frac{\mu_{0200}^{1/2}}{\lambda_h^{1/2}} \left( \frac{\mu_{1010}}{\sqrt{\mu_{2000} \sqrt{\mu_{0020}}}} - \frac{\mu_{0110}}{\sqrt{\mu_{2000} \sqrt{\mu_{0200}}}} - \frac{\mu_{1100}}{\sqrt{\mu_{2000} \sqrt{\mu_{0200}}}} \right) \left( \frac{\mu_{0200}^{1/2}}{\lambda_h^{1/2}} \right),

1 - \frac{\mu_{0110}^2}{\mu_{0200} \mu_{0020}},

B_1 = \frac{\mu_{0200}^{1/2}}{\lambda_h^{1/2}} \left( \frac{\mu_{1100}}{\sqrt{\mu_{2000} \sqrt{\mu_{0020}}}} - \frac{\mu_{0110}}{\sqrt{\mu_{2000} \sqrt{\mu_{0200}}}} - \frac{\mu_{1010}}{\sqrt{\mu_{2000} \sqrt{\mu_{0200}}}} \right),

1 - \frac{\mu_{0110}^2}{\mu_{0200} \mu_{0020}},

Second proposed estimator

Here we use the same idea of second proposed family of estimators of Hussain et al., and estimate finite population CDFs which concern CDFs of study and auxiliary variables along with ranks of the auxiliary variable.

\[ \hat{F}_{Prop2}^* (ty) = k_{14} \hat{F}(ty) + k_{15} \left( \frac{F(tx) - \hat{F}(tx)}{F(tx)} \right) \exp \left( \frac{F(tx) - \hat{F}(tx)}{\hat{F}(tx) + \hat{F}(tx)} + \hat{F}(tx) \right) \]

\[ + k_{16} \left( \frac{\hat{R}_x - \hat{R}_x}{\hat{R}_x} \right) \left\{ \exp \left( \frac{\hat{R}_x - \hat{R}_x}{\hat{R}_x + \hat{R}_x} \right) \right\}, \]  

(36)

where \( k_{14}, k_{15} \) and \( k_{16} \) are suitable chosen constants. The estimator \( \hat{F}_{Prop2}^* (ty) \), in terms of errors, we have

\[ \hat{F}_{Prop2}^* (ty) = k_{14} F(ty)(1 + \nu_1) - k_{15} \nu_2 \left( 1 - \frac{1}{2} \nu_2 + \frac{3}{8} \nu_2^2 + \cdots \right) \]

\[- k_{16} \nu_4 \left( 1 - \frac{1}{2} \nu_4 + \frac{3}{8} \nu_4^2 + \cdots \right). \]  

(37)
Further simplifying (76), and keeping error up to power 2, we have

\[
(\hat{F}_{prop}^*(ty) - F(ty)) = -F(ty) + k_{14}F(ty) + k_{14}F(ty)\nu_1 - k_{15}\nu_2 + k_{15}\frac{\nu_2^2}{2} \\
- k_{16}\nu_4 + k_{16}\frac{\nu_4^2}{2}.
\] (38)

The bias and MSE of \(\hat{F}_{prop}^*(ty)\), to the first order of approximation, are given by

\[
\text{Bias}(\hat{F}_{prop}^*(ty)) \approx -F(ty) + k_{14}F(ty) + \frac{1}{2}k_{15}\mu_{0200} + \frac{1}{2}k_{16}\mu_{0002},
\]

and

\[
\text{MSE}(\hat{F}_{prop}^*(ty)) \approx F^2(ty) + k_{15}\mu_{0200}(-F(ty) + k_{15}) + 2k_{15}k_{16}\mu_{0101} \\
+ k_{16}\mu_{0002}(F(ty) + k_{16}) + k_{14}F^2(ty)(k_{14} - 2 + \mu_{2000}) \\
+ k_{14}k_{15}F(ty)\mu_{0200} - 2\mu_{1100} + k_{14}k_{16}F(ty)\mu_{0002} - 2\mu_{1001}.
\] (39)

The optimum values of \(k_{14}, k_{15}\) and \(k_{16}\), determined by minimizing (39), are given by

\[
k_{14(\text{opt})} = \frac{\left(2\lambda_h \frac{\mu_{2000}^{1/2}}{\lambda_h^{1/2}}\right)\left(C_1 + D_1\right) + \lambda_h\left(C + \frac{\mu_{2000}}{\lambda_h}\right) - 4}{4\mu_{2000}(\Re_{F_{\text{th}}F_{\text{th}}R_{th}}^2 - 1) + 4\mu_{2000}^{1/2}(C_1 + D_1) + \lambda_h\left(C + \frac{\mu_{2000}}{\lambda_h}\right) - 4},
\]

\[
k_{15(\text{opt})} = \frac{\left(2\mu_{2000}^{1/2}\lambda_h^{1/2}\right)\left(\frac{-\mu_{0200}^{1/2}}{\lambda_h^{1/2}}\left(\frac{\mu_{1100}}{\sqrt{\mu_{2000}\sqrt{\mu_{0200}}}}\mu_{0200} + \mu_{0101}\frac{\mu_{2000}}{\sqrt{\mu_{0002}}}ight) - \mu_{0101}\frac{\mu_{2000}}{\sqrt{\mu_{0002}}}\right)\left(C_1 + D_1\right) + \lambda_h\left(C + \frac{\mu_{2000}}{\lambda_h}\right) - 4}{4\mu_{2000}(\Re_{F_{\text{th}}F_{\text{th}}R_{th}}^2 - 1) + 4\mu_{2000}^{1/2}(C_1 + D_1) + \lambda_h\left(C + \frac{\mu_{2000}}{\lambda_h}\right) - 4},
\]

\[
k_{16(\text{opt})} = \frac{\left(2\mu_{2000}^{1/2}\lambda_h^{1/2}\right)\left(\frac{-\mu_{0200}^{1/2}}{\lambda_h^{1/2}}\left(\frac{\mu_{1100}}{\sqrt{\mu_{2000}\sqrt{\mu_{0200}}}}\mu_{0200} + \mu_{0101}\frac{\mu_{2000}}{\sqrt{\mu_{0002}}}ight) - \mu_{0101}\frac{\mu_{2000}}{\sqrt{\mu_{0002}}}\right)\left(C_1 + D_1\right) + \lambda_h\left(C + \frac{\mu_{2000}}{\lambda_h}\right) - 4}{4\mu_{2000}(\Re_{F_{\text{th}}F_{\text{th}}R_{th}}^2 - 1) + 4\mu_{2000}^{1/2}(C_1 + D_1) + \lambda_h\left(C + \frac{\mu_{2000}}{\lambda_h}\right) - 4},
\]
\[
\begin{align*}
\mu_{2000}^{1/2} f(ty) & = \left[ \mu_{0002}^{1/2} \left( \frac{\mu_{2000}}{\sqrt{\mu_{0200} \mu_{0002}}} \sqrt{\mu_{2000} \sqrt{\mu_{0200} \mu_{0002}}} \frac{\mu_{0002}}{\mu_{2000} \mu_{0002}} \frac{\mu_{1001}}{\mu_{2000} \mu_{0002}} \right) \right]^{1/2} \\
& + \mu_{0002}^{1/2} \mu_{0200}^{1/2} \left( \frac{\mu_{1001}}{\sqrt{\mu_{2000} \sqrt{\mu_{0200} \mu_{0002}}} \sqrt{\mu_{2000} \sqrt{\mu_{0200} \mu_{0002}}} \frac{\mu_{1100}}{\mu_{2000} \mu_{0002}}} \right) \\
& + 4 \left( \frac{\mu_{1100}}{\sqrt{\mu_{2000} \sqrt{\mu_{0200} \mu_{0002}}} \sqrt{\mu_{2000} \sqrt{\mu_{0200} \mu_{0002}}}} \right)
\end{align*}
\]

The minimum MSE of \( \hat{F}_{\text{prop}}^{*}(ty) \), at the optimum values of \( k_{14}, k_{15} \) and \( k_{16} \) is given by

\[
\text{MSE}_{\text{min}}(\hat{F}_{\text{prop}}^{*}(ty)) = \frac{F^2(y) \mu_{2000} \left( C(\lambda h) - D(\lambda h) + \mu_{0200} + 4(\Re_{F_{2000}^{*}} \Re_{F_{2000}^{*}}^{*} - 1) \right)}{\left( \lambda h \left( 4 \mu_{2000} \left( \Re_{F_{2000}^{*}}^{*} - 1 \right) + 4 \frac{\mu_{1001}}{\lambda h} (C_1 + D_1) + C + \frac{\mu_{0200}}{\lambda h} - 4 \right) \right)}.
\]

where

\[
\Re_{F_{2000}^{*}}^{*} = \left( \frac{\mu_{1001} + \mu_{0200} - 2 \mu_{0200} \mu_{0002}}{\mu_{2000} \mu_{0200} \mu_{0002} - \mu_{0101}^2} \right),
\]

\[
C = \left( \frac{\mu_{0002}^{1/2} \mu_{0200}^{1/2} \left( \frac{\mu_{0101}}{\mu_{2000} \mu_{0002} \mu_{0002}} \frac{\mu_{1001}}{\mu_{2000} \mu_{0002} \mu_{0002}} \right)^2}{\mu_{2000} \mu_{0200} \mu_{0002} - \mu_{0101}^2} \right),
\]

\[
C_1 = \left( \mu_{2000}^{1/2} \left( \mu_{1001} - \mu_{0101} \frac{\mu_{0200}^{1/2}}{\mu_{2000} \mu_{0002} \mu_{0002}} - \frac{\mu_{1100} \mu_{0101}}{\mu_{2000} \mu_{0002} \mu_{0002}} \right) \right)^2,
\]

\[
D = \left( \frac{\mu_{1001} \mu_{0101} \left( \frac{\mu_{2000}^{1/2}}{\mu_{2000} \mu_{0002} \mu_{0002}} - \mu_{0101} \frac{\mu_{0200}^{1/2}}{\mu_{2000} \mu_{0002} \mu_{0002}} \right)^2}{\mu_{2000} \mu_{0200} \mu_{0002} - \mu_{0101}^2} \right).
\]
Table 1. Summary statistics for Population I.

| h | Nh | nh | Wh | λh | F(τh) | F(τ₀) | X̄h | R̄h |
|---|----|----|----|----|-------|-------|-----|-----|
| 1 | 127| 31 | 0.1375 | 0.0244 | 0.3543 | 0.3779 | 20805 | 64 |
| 2 | 117| 21 | 0.1267 | 0.0390 | 0.5188 | 0.4872 | 9212  | 59 |
| 3 | 103| 29 | 0.1115 | 0.0248 | 0.4272 | 0.4660 | 14309 | 52 |
| 4 | 170| 38 | 0.1841 | 0.0204 | 0.3543 | 0.3779 | 20805 | 64 |
| 5 | 205| 22 | 0.2221 | 0.0406 | 0.6146 | 0.6537 | 5570  | 103|
| 6 | 201| 39 | 0.2177 | 0.0207 | 0.5025 | 0.3532 | 12998 | 101|

Table 2. Summary statistics for Population II.

| h | Nh | nh | Wh | λh | F(τh) | F(τ₀) | X̄h | R̄h |
|---|----|----|----|----|-------|-------|-----|-----|
| 1 | 127| 31 | 0.1375 | 0.0244 | 0.3543 | 0.3700 | 498.276 | 64 |
| 2 | 117| 21 | 0.1267 | 0.0391 | 0.4188 | 0.4700 | 318.333 | 59 |
| 3 | 103| 29 | 0.1115 | 0.0248 | 0.4272 | 0.4272 | 431.359 | 52 |
| 4 | 170| 38 | 0.1841 | 0.0204 | 0.5765 | 0.5882 | 311.324 | 86 |
| 5 | 205| 22 | 0.2221 | 0.0406 | 0.6146 | 0.5909 | 227.195 | 103|
| 6 | 201| 39 | 0.2177 | 0.0207 | 0.5025 | 0.4527 | 312.706 | 101|

Empirical study in stratified random sampling

In this section, we conduct a numerical study to investigate the performances of the existing and proposed CDF estimators in stratified random sampling. For this purpose, four populations are considered. The summary statistics of these populations are reported in
Tables 1–4. The PRE of an estimator $\hat{F}_i(y)$ with respect to $\hat{F}_1(y)$ is

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

where, $i = 2, 3, \ldots, 13$.

The mean square error and PREs of distribution function estimators, computed from four populations, are given in Tables 5–6.

Population I [(Source: Koyuncu and Kadilar16)  
Y: The number of teachers and...
Table 5. MSEs using Populations I–IV stratified.

| Estimators          | Population-I | Population-II | Population-III | Population-IV |
|---------------------|--------------|---------------|----------------|---------------|
| $\hat{F}^{\text{SRs}}_{\text{ty}}$ | 0.00122971   | 0.00122971    | 0.00691254     | 0.00691254    |
| $\hat{F}^{\text{R}}_{\text{ty}}$    | 0.00036042   | 0.00028777    | 0.00380174     | 0.00385274    |
| $\hat{F}^{\text{P}}_{\text{ty}}$    | 0.0463523    | 0.02350301    | 0.02331402     |               |
| $\hat{F}^{\text{BTs}}_{\text{ty}}$  | 0.0049599    | 0.00045079    | 0.00367218     | 0.00371493    |
| $\hat{F}^{\text{BTp}}_{\text{ty}}$  | 0.00256156   | 0.00262452    | 0.01352281     | 0.01344557    |
| $\hat{F}^{\text{Reg}}_{\text{ty}}$  | 0.00033807   | 0.00027072    | 0.00313223     | 0.00336406    |
| $\hat{F}^{\text{RD}}_{\text{ty}}$   | 0.00033762   | 0.00027043    | 0.00327009     | 0.00331960    |
| $\hat{F}^{\text{GK}}_{\text{ty}}$   | 0.0003373    | 0.00027000    | 0.00324537     | 0.00325338    |
| $\hat{F}^{\text{H1}}_{\text{ty}}$   | 0.00033164   | 0.00026443    | 0.00320746     | 0.00320993    |
| $\hat{F}^{\text{H2}}_{\text{ty}}$   | 0.00033029   | 0.00026305    | 0.00316285     | 0.00320993    |
| $\hat{F}^{\text{Prop1}}_{\text{ty}}$| 0.0002867    | 0.00024416    | 0.00284663     | 0.00288654    |
| $\hat{F}^{\text{Prop2}}_{\text{ty}}$| 0.00028566   | 0.00024318    | 0.00283534     | 0.00287695    |

Table 6. PREs using Populations I–IV stratified.

| Estimators          | Population-I | Population-II | Population-III | Population-IV |
|---------------------|--------------|---------------|----------------|---------------|
| $\hat{F}^{\text{SRs}}_{\text{ty}}$ | 100          | 100           | 100            | 100           |
| $\hat{F}^{\text{R}}_{\text{ty}}$    | 341.1900     | 427.3200      | 181.8300       | 179.4200      |
| $\hat{F}^{\text{P}}_{\text{ty}}$    | 29.41000     | 29.65000      | 29.41000       | 29.65000      |
| $\hat{F}^{\text{BTs}}_{\text{ty}}$  | 247.9300     | 272.7900      | 188.2400       | 186.0700      |
| $\hat{F}^{\text{BTp}}_{\text{ty}}$  | 48.01000     | 46.85000      | 51.12000       | 51.41000      |
| $\hat{F}^{\text{Reg}}_{\text{ty}}$  | 363.7400     | 454.2400      | 208.6300       | 205.4800      |
| $\hat{F}^{\text{RD}}_{\text{ty}}$   | 364.2300     | 454.7300      | 211.3900       | 208.2300      |
| $\hat{F}^{\text{GK}}_{\text{ty}}$   | 48.01000     | 46.85000      | 51.12000       | 51.41000      |
| $\hat{F}^{\text{H1}}_{\text{ty}}$   | 370.8000     | 465.0400      | 213.0000       | 209.8000      |
| $\hat{F}^{\text{Prop1}}_{\text{ty}}$| 372.310.     | 467.4800      | 218.5500       | 215.3500      |
| $\hat{F}^{\text{H2}}_{\text{ty}}$   | 429.420.     | 503.6500      | 242.8300       | 239.4700      |
| $\hat{F}^{\text{Prop2}}_{\text{ty}}$| 430.4900     | 505.6800      | 243.8000       | 240.2700      |

$X$: The number of students in both primary and secondary schools in Turkey in 2007 for 923 districts in six regions.

Population II [(Source: Koyuncu and Kadilar\textsuperscript{17}]

$Y$: The number of teachers and

$X$: The number of classes in both primary and secondary schools in Turkey in 2007 for 923 districts in six regions.

Population III [(Source: Kadilar and Cingi\textsuperscript{18}]

$Y$: Apple production amount in 1999 and

$X$: The number of apple trees in 1999.

Population IV [(Source: Kadilar and Cingi,\textsuperscript{18}]

$Y$: Apple production amount in 1999 and
Table 7. MSEs using simulation.

| Estimators | MSEs Using Population-1 | MSEs Using Population-2 |
|------------|--------------------------|--------------------------|
|            | N=100 | 150 | 200 | N=100 | 150 | 200 |
| \( \hat{F}^*_{\text{SRS},(ty)} \) | 0.002252252 | 0.00141800 | 0.00100100 | 0.002252252 | 0.001418085 | 0.001001001 |
| \( \hat{F}^*_{\text{S},(ty)} \) | 0.001459460 | 0.000918919 | 0.000574575 | 0.002774775 | 0.001747080 | 0.001233233 |
| \( \hat{F}^*_{\text{BT},(ty)} \) | 0.007550678 | 0.004754130 | 0.003355857 | 0.006235362 | 0.003925969 | 0.002771272 |
| \( \hat{F}^*_{\text{ST},(ty)} \) | 0.001292793 | 0.000813919 | 0.000574575 | 0.001950451 | 0.001228061 | 0.000866867 |
| \( \hat{F}^*_{\text{BT},(ty)} \) | 0.004337838 | 0.002731231 | 0.001927928 | 0.003680180 | 0.002317150 | 0.001635636 |
| \( \hat{F}^*_{\text{Prop},(ty)} \) | 0.001223027 | 0.000770054 | 0.000542388 | 0.001905509 | 0.001203161 | 0.000850494 |
| \( \hat{F}^*_{\text{G},(ty)} \) | 0.001807435 | 0.001138015 | 0.000803304 | 0.002026581 | 0.001275995 | 0.000900703 |
| \( \hat{F}^*_{\text{Reg},(ty)} \) | 0.001217073 | 0.000767689 | 0.000542388 | 0.001905509 | 0.001203161 | 0.000850494 |
| \( \hat{F}^*_{\text{Prop},(ty)} \) | 0.001216763 | 0.000767683 | 0.000542386 | 0.001905470 | 0.001203152 | 0.000850491 |
| \( \hat{F}^*_{\text{H},(ty)} \) | 0.0001013070 | 0.000640904 | 0.000453482 | 0.001786342 | 0.001294042 | 0.000798882 |
From Table 5 and Table 6, in terms of mean squared error and PRE, it is clear that proposed estimators i.e, \( \hat{F} \) Prop\(_1\) (ty) and \( \hat{F} \) Prop\(_2\) (ty) performs better than the estimators \( \hat{F} \) SRS\(_u\) (ty), \( \hat{F} \) R (ty), \( \hat{F} \) P (ty), \( \hat{F} \) BT\(_R\) (ty), \( \hat{F} \) BT\(_P\) (ty), \( \hat{F} \) Reg (ty), \( \hat{F} \) R\(_D\) (ty), \( \hat{F} \) S (ty), \( \hat{F} \) G\(_K\) (ty), \( \hat{F} \) H\(_1\) (ty), and \( \hat{F} \) H\(_2\) (ty). As we increase the sample size the mean square error values decrease, and percentage relative efficiency gives the best results, which are the expected results.

### Simulation study

We have generated two populations of size 1000 from multivariate normal distribution with different covariance matrices. The results of simulation are given in Tables 7 and 8. The population means and covariance matrices, are given below:

#### Population I

\[ \mu_1 = \begin{bmatrix} 500 \\ 500 \\ 500 \end{bmatrix} \]

and

\[ \sum_1 = \begin{bmatrix} 1000 & 800 & 810 \\ 800 & 850 & 820 \\ 810 & 820 & 840 \end{bmatrix} \]
\( \rho_{XY} = 0.8820157 \)

**Population II**

\[
\mu_2 = \begin{bmatrix} 500 \\ 500 \\ 500 \end{bmatrix}
\]

and

\[
\sum_2 = \begin{bmatrix} 400 & 270 & 220 \\ 270 & 500 & 300 \\ 220 & 300 & 675 \end{bmatrix}
\]

\( \rho_{XY} = 0.5897143 \)

Covariance matrices show the distribution of Study Variable Y, the auxiliary variable X and the ranks of the auxiliary variable Rx. There is a high correlation in Population I, and weak correlation in Population II.

We estimate the MSE using \( k = 1000 \) samples of diverse sizes selected from each population. Three different sample sizes \( n = 100, 150, 200 \) are taken from both populations.

Table 7 shows that the proposed estimators \( \hat{F}^*_{\text{Prop}_1}(ty) \) and \( \hat{F}^*_{\text{Prop}_2}(ty) \) performs better as compared to all other existing estimators for both populations in terms of MSEs. We have also seen that as the sample size increases MSE of all the decreases.

Table 8 shows that the proposed estimators \( \hat{F}^*_{\text{Prop}_1}(ty) \) and \( \hat{F}^*_{\text{Prop}_2}(ty) \) performs better as compared to all other existing estimators for both populations in terms of the PREs.

**Conclusion**

In this article, we propose ratio-in-regression type exponential estimator for the finite population distribution function under stratified random sampling, which required an ancillary variable on the sample mean and rank of the auxiliary variable. Expressions for mean square error of the proposed estimator are derived up to first order of approximation and comparison is made with the estimators mentioned herein. According to results of real data sets, and simulation it is perceived that the proposed estimator of \( \hat{F}^*_{\text{Prop}_1}(ty) \), \( \hat{F}^*_{\text{Prop}_2}(ty) \) performs better in terms of percentage relative efficiency, than usual estimator of estimator of Hussain et al.,\(^9\) Cochran,\(^10\) Murthy,\(^11\) Bahl and Tuteja,\(^12\) regression estimator, Rao,\(^13\) Singh et al.,\(^14\) and, Grover and Kaur.\(^15\)

A simulation analysis is also carried out to assess the robustness and generalizability of the propose estimator. The simulation study’s findings also confirm the utility of the proposed estimator. A numerical study is carried out to support the theoretical results. Therefore, we recommend the use of proposed estimators for efficiently estimating the finite population finite population distribution function under stratified random sampling.
Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Sohail Akhtar https://orcid.org/0000-0003-4870-152X
Sardar Hussain https://orcid.org/0000-0001-9274-0396

References

1. Chambers RL and Dunstan R. Estimating distribution functions from survey data. Biometrika 1986 Dec 1; 73: 597–604.
2. Rao JN, Kovar JG and Mantel HJ. On estimating distribution functions and quantiles from survey data using auxiliary information. Biometrika 1990 Jun 1; 70: 365–375.
3. Rao JN. Estimating totals and distribution functions using auxiliary information at the estimation stage. J Off Stat 1994 Jun 1; 10: 153.
4. Kuk AY. A kernel method for estimating finite population distribution functions using auxiliary information. Biometrika 1993 Jun 1; 80: 385–392.
5. Ahmed MS and Abu-Dayyeh W. Estimation of finite-population distribution function using multivariate auxiliary information. Stat Transit 2001; 5: 501–507.
6. Rueda M, Martínez S, Martínez H, et al. Estimation of the distribution function with calibration methods. J Stat Plan Inference 2007 Feb 1; 137: 435–448.
7. Singh HP, Singh S and Kozak M. A family of estimators of finite-population distribution function using auxiliary information. Acta Appl Math 2008 Nov; 104: 115–130.
8. Hussain S, Zichuan M, Hussain S, et al. On Estimation of Distribution Function Using Dual Auxiliary Information under Nonresponse Using Simple Random Sampling, Journal of Probability and Statistics, vol. 2020, Article ID 1693612, 13 pages, 2020. https://doi.org/10.1155/2020/1693612.
9. Hussain S, Ahmad S, Saleem M, et al. Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. Plos one 2020 Sep 28; 15: e0239098.
10. Cochran WG. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. J Agric Sci 1940 Apr; 30: 262–275.
11. Murthy MN. Product method of estimation. Sankhyā: the Indian journal of statistics. Series A 1964 Jul 1; 26: 69–74.
12. Bahl S and Tuteja R. Ratio and product type exponential estimators. J Inform Optim Sci 1991 Jan 1; 12: 159–164.
13. Rao TJ. On certain methods of improving ration and regression estimators. Commun Stat-Theory Method 1991 Jan 1; 20: 3325–3340.
14. Singh HP and Kumar S. A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information. SORT-Stat Oper Res Trans 2009 Dec 10; 33: 71–84.
15. Grover LK and Kaur P. A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. *Commun Stat-Simul Comput* 2014 Jan 1; 43: 1552–1574.

16. Kadilar C and Cingi H. Ratio estimators in stratified random sampling. *Biometrical J: J Math Methods Biosci* 2003 Mar; 45: 218–225.

17. Kadilar C and Cingi H. A new ratio estimator in stratified random sampling. *Commun Stat—Theory Methods* 2005 Mar 1; 34: 597–602.

18. Koyuncu N and Kadilar C. Ratio and product estimators in stratified random sampling. *J Stat Plan Inference* 2009 Aug 1; 139: 2552–2558.

19. Shabbir J and Gupta S. On estimating finite population mean in simple and stratified random sampling. *Commun Stat-Theory Methods* 2010 Dec 6; 40: 199–212.

20. Aladag S and Cingi H. Improvement in estimating the population median in simple random sampling and stratified random sampling using auxiliary information. *Commun Stat-Theory Methods* 2015 Mar 4; 44: 1013–1032.

21. Malik S and Singh R. A new estimator for population mean using two auxiliary variables in stratified random sampling. *J Inform Optim Sci* 2017 Nov 17; 38: 1243–1252.

**Author biographies**

Mr. Sardar Hussain is a student of M.Phil in the department of statistics Quaid-i-Azam University Islamabad, Pakistan. He has published 14 research papers in the field of survey sampling. His research area includes the estimation of the distribution function and estimation of mean and median under different sampling designs.

Dr. Sohail Akhtar is working as an Associate Professor in Statistics at the Department of Mathematics and Statistics, the University of Haripur, Haripur, KP, Pakistan. He received his Ph.D. degree from the University of Salford, the UK in 2012. He has more than 18 years of teaching and research experience. His area of interest is Forecasting, Biostatistics, Statistical Modelling, and Survey sampling.

Dr. Mahmoud El-Morshedy is a Professor in the Department of Statistics at Mansoura University, Mansoura, Egypt. He is also affiliated with the Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdul-Aziz University, Al-Kharj 11942, Saudi Arabia. He has published around 117 research papers in the field of distribution theory. His research area includes Probability Distribution Theory, Statistical Modelling, Multi-variate Analysis, Hypothesis Testing, and Computational Statistics.