Using Auxiliary Information and Their Ranks for Estimation of Finite Population Mean

By

Sughra

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2023

بسم الله الرحمن الرحيم

In the Name of ALLAH The Most Merciful and The Most Beneficent

Using Auxiliary Information and Their Ranks for Estimation of Finite Population Mean

By

Sughra

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PHILOSOPHY IN STATISTICS

Supervised By

Dr. Manzoor Khan

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2022

CERTIFICATE

Using Auxiliary Information and It's Ranks for Estimation of Finite Population Mean

By

SUGHRA

(Reg. No. 02222113013)

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF M.PHIL. IN

STATISTICS

We accept this thesis as conforming to the required standards

1. ______________________
Dr. Manzoor Khan
(Supervisor)

2. ______________________
Dr. Saadia Masood
(External Examiner)

3. ______________________
Prof. Dr. Ijaz Hussain
(Chairman)

DEPARTMENT OF STATISTICS
QUAID-I-AZAM UNIVERSITY
ISLAMABAD, PAKISTAN
2023

Declaration

I "Sughra" hereby solemnly declare that this thesis titled, "Using Auxiliary Information and Their Ranks for Estimation of Finite Population Mean".

- This work was done wholly in candidature for a degree of M.Phil Statistics at this University.
- Where I got help from the published work of others, this is always clearly stated.
- Where I have quoted from the work of others, the source is always mentioned. Except of such quotations, this thesis is entirely my own research work.
- Where the thesis is based on work done by myself jointly with my supervisor, I have made clear exactly what was done by others and what I have suggested.

Dated:____________________ Signature:____________________

Dedication

I dedicate this research to my Baba who always support me, brought me up to be an independent women. Also my siblings, who were a good support towards me. I dedicate this to my sister who is my motivation to be successful in life, who taught me how to forgive and forget and how to be an womanly man in the man's world. Special thanks to my respected teachers, my friends, who supports me and always there for me.

Contents

List of Tables

Abstract

Survey statistician's purpose has always been to produce the best estimates of unknown population parameters. In survey sampling, the selected sample might contain tied values. The idea of ranking with tied values takes the variation of the auxiliary variable into account. We have propose a two-fold application of auxiliary information in which auxiliary information is supported by original and ranked auxiliary variables. Simple and stratified random sampling schemes are used to establish mathematical developments. The results align perfectly with the idea of using auxiliary information to help estimate the required characteristics.We observe that using relevant information more thoroughly enhances the efficiency of the estimation process. Up to the first level of approximation, we formulate expressions for the bias and mean square errors of the proposed estimators. We conduct a theoretical assessment of the suggested estimators. We employed real-world data sets to demonstrate, through numerical analysis, how the proposed estimators perform, thus corroborating our theoretical conclusions. The analysis demonstrated that the proposed estimator performed better than the existing estimators.

Chapter 1

Introduction

1.1 Background of the study

Many researcher's primary focus has always been on determining the many traits and Population statistics, including measures like average, variability, and proportion, but examining the entire population is challenging. Due to limited resources such as time, labour or money, having access to every member of the population is frequently neither feasible nor practical. To address this issue, we choose a manageable subset of a population that is Reflective of the whole population.

In statistical analysis, the act of choosing a certain number of observations from a larger population is known as sampling. Various sampling designs, like simple random sampling, stratified random sampling, systematic sampling, and cluster sampling, are employed to select samples from a larger population, and their choice depends on the specific type of study being conducted, which spans virtually every facet of human activity.

Sampling is the structured method of choosing a sample that accurately represents a population in research, enabling us to draw conclusions about the entire population. The advantages of sampling is the possibility of saving money and human resources. Within a given sample size, the goal of sample selection is to optimise estimation accuracy while avoiding bias. This is critical since bias can undermine information integrity and affect a researcher's findings.

1.2 Survey Sampling

Sample survey uses the sampling method to survey only a part of the population rather than the complete population in order to estimate the Characteristics of a population (such as average, median, variance, mode, total, ratio, distribution function). A sample survey is a method of determining features of a population or universe by analysing a subset of it. Researchers, organisations, and government bodies all around the world use these methods to investigate, make taxes, price fixing, and minimum wage decisions, as well as plan and project future economic structure. Sample surveys have several advantages, including the fact that they are simple, inexpensive, and require a smaller scale of operation, which decreases the time required to collect and process data. It enables the measuring of demographic features that would otherwise be impossible to quantify. Importantly, in order to make exact inferences, sample surveys must acquire a population-representative sample.

1.3 Auxiliary information

Estimators can become more precise when auxiliary information is employed, particularly when there is a correlation between the study variable and the auxiliary variables. Auxiliary variables are information available for all units of the population before data collection on the study variable. The researchers utilize these information in developing efficient estimators. Using auxiliary data to estimate the population parameter, Neyman (1938) improved his results. In the literature of survey sampling, various methods exist to enhance the efficiency of population mean estimations through the incorporation of auxiliary information. These methods encompass ratio, product, and regression approaches. By effectively utilizing auxiliary information, we can decrease the mean square error (MSE) of the estimator, resulting in a more efficient estimation.When there's a positive correlation between the study variable (Y) and the auxiliary variable (X), the ratio estimator is the optimal choice for estimating the population mean. Conversely, when they exhibit negative correlation, the product estimator is the more efficient option. The ratio estimator is most effective when there is a linear relationship between the study variable and the auxiliary variable, and the regression line intersects the origin. However, if the regression line doesn't pass through the origin, the ratio estimator loses its effectiveness, and the preference shifts to the regression estimator. A large number of researchers work with supplementary data in the form of variables. Cochran (1977) centered his work on employing them during the stage of estimation aimed at enhancing the efficiency of the estimators.

In the field of survey sampling, the use of auxiliary information has a history of being applied. Hansen and Hurwitz (1943) suggest incorporating auxiliary data when selecting samples with varying probabilities. Koyuncu and Kadilar (2009a) introduces a set of estimators that utilize auxiliary information within the context of stratified random sampling. Numerous authors have developed different estimators to enhance ratio estimations in simple random sampling by incorporating auxiliary variables, including Upadhyaya and Singh (1999), Sisodia and Dwivedi (1981), Singh and Tailor (2003), Shabbir and Gupta (2007), and Khoshnevisan et al. (2007). Modifications were made to ratio estimators in the context of stratified random sampling by Kadilar and Cingi (2003), Shabbir and Gupta (2005), Kadilar and Cingi (2005),Koyuncu and Kadilar (2009a), Koyuncu and Kadilar (2009b) to enhance the effectiveness of the estimators.

When there is a correlation between the values of the study variable and the auxiliary variable, the study variable values are also correlated with the rankings of the auxiliary variable. Utilizing the idea, Haq et al. (2017) introduced a novel estimator for the mean of a finite population, which utilizes dual sources of auxiliary information, including auxiliary variables and their rankings. The bias and mean squared error (MSE) of the proposed estimator were calculated using a first-order approximation. Both theoretical analysis and empirical studies indicate that the suggested estimator performs better than all other commonly used estimators.

Yaqub et al. (2017) addressed the problem of estimating the finite population mean that utilizes the auxiliary information and ranks of the auxiliary variable in the presence of non-response. Up to the first order of approximation, expressions for bias and mean squared error of considered estimators are generated. Numerical study suggests that the proposed estimator is more efficient than the usual estimators.

Shabbir and Gupta (2019) presented an extensive range of estimators for finite population variance within the framework of stratified sampling, which includes the utilization of auxiliary variables and their respective rankings. The bias and mean square estimator were calculated using a first-order approximation. The outcomes illustrate that the proposed generalized class of estimators is superior in efficiency compared to the traditional sample variance estimator..

Javed and Irfan (2020) expanded upon the concept introduced by Haq et al. (2017) to develop novel exponential-type estimators optimized for estimating the mean of a finite population within a stratified random sampling framework. The bias and mean squared error were calculated for the suggested estimators. The numerical findings indicate that these estimators outperform the competition in terms of efficiency

Ahmad et al. (2021) introduced an enhanced set of estimators for the mean of a finite population within the context of stratified random sampling. Bias and mean square error estimates were calculated for both the proposed and existing estimators using a first-order approximation. The results from the simulation study led to the conclusion that the suggested family of estimators performed superiorly.

When information regarding the study variable is lacking, the utility of supplemental knowledge becomes vital. Ullah et al. (2021) introduces a novel exponential-type estimator for estimating the mean of a finite population within a simple random sampling framework. Mathematical expressions for the bias and mean squared error were derived of the proposed estimator using a first-order approximation. Both theoretical analysis and empirical studies consistently showed that the proposed estimator surpasses existing estimators in terms of percentage relative efficiency.

The utilization of dual auxiliary information is instrumental in enhancing estimator efficiency. In pursuit of this objective, Irfan et al. (2022) introduced exponential estimators of the difference type, leveraging dual auxiliary information, for estimating the population mean in the context of simple random sampling. Mathematically, the bias and mean squared error of the proposed estimators were defined. Empirical and simulation outcomes, considering mean square errors and percentage relative efficiencies, consistently demonstrated that the suggested estimators outperformed their counterparts.

In their work, Shabbir and Onyango (2022) introduced an enhanced unbiased estimator for estimating the mean of a finite population. This was achieved by utilizing a single auxiliary variable along with its ranking, incorporating the Hartley-Ross procedure, particularly when certain parameters of the auxiliary variable are known. Mathematical expressions for the bias and mean square error (or variance) of the estimator are derived up to the first-order approximation. The proposed unbiased estimator outperformed all other considered estimators.

1.4 Sampling Methods

The primary goal of sampling theory is to gauge characteristics of the population, specifically the population mean. In the survey sampling literature, there are several approaches and sample procedures that are extensively used for estimating characteristics of the population. To improve the accuracy of the mean estimators, auxiliary information is typically used throughout the sampling and/or estimating stages. When the study variable strongly correlates with the auxiliary variable, we anticipate that the ratio between the study and auxiliary variables will exhibit less variability compared to the study variable alone. Conversely, when there is a negative correlation between the study and auxiliary variables, the product method of estimation becomes the preferable choice. When contrasted with the conventional mean-per-unit estimator under simple random sampling (SRS), these estimators are known to have lower variances under certain conditions. To properly begin statistical research, the sampling technique and estimators to be used must be clearly determined.

Numerous sampling techniques have been developed to obtain effective estimations of attributes, with simple random sampling and stratified random sampling being the most prevalent methods.

1.4.1 Simple random sampling

Simple random sampling (SRS) is the most commonly used data collection strategy when we need to draw conclusions about a population through the analysis of only a portion of it. When the population units are homogeneous, the results are more precise. In SRS, each individual unit of a population is chosen at random, so that each sample has the same chance of being chosen at any point of the process. There are two methods for drawing a sample from SRS: one is with replacement scheme and the other is without replacement. In the context of simple random sampling with replacement (SRSWR), every unit within the population has an identical likelihood of being chosen and replaced before drawing the next sample unit. In sampling without replacement (SRSWOR), a subset of the observations is chosen at random, and once chosen, an observation cannot be chosen again.

1.4.2 Stratified random sampling

When the population units are heterogeneous, SRS is not successful because the selected sample may not be a good representative of the population, and the estimator's precision falls as a result. To address this issue, stratified random sampling is used. The sampling approach of stratified random sampling splits the entire population of interest into strata, which are homogeneous groups from which samples are taken separately and independently. These classifications must be mutually exclusive or non-overlapping. The stratum should contain uniform units, or variability within subgroups should be minimal, but diversity between subgroups should be maximal. Many allocation methods are used to calculate sample size, including proportional allocation, equal allocation, optimum allocation, and Neyman's allocation.

1.5 Tied Ranks

A simple rank is a number assigned to a single sample item based on its position in the sorted list. The first item is given a rank of one, the second a rank of two, and so on. There could be two or more observations with the same rank. These are known as tied values. When two or more values are tied, the mean of their ranks is assigned to each. There would be no change in the ranks of the other values. For instance, if we have 1,2,3,4,4,4,5,6,6 as the dataset then the 4's and 6's are tied values. Tied data is a concern because identical numbers must now be transformed into rank. Sometimes ranks are assigned at random, while other times an average rank is used. Ranking is a strategy that is often used in Non-parametric Statistics. Most significantly, a methodology for breaking tied ranks must be provided in order for the result to be consistent.

1.6 Motivation for this research

Tied ranks are used in research when there are two or more observations that have the same value. It can happen for various reasons, such as measurement error, rounding, or limitations of the measurement instrument. In some cases, tied ranks can be problematic when computing statistical tests or measures that assume unique ranks, such as the Spearman correlation coefficient or the Mann-Whitney U-test. Simple ranks are equally spaced and does not take the variation into account, while tied rank are not equally spaced and better represents the scattered the data set. So, it is expected that the estimator using tied ranks would perform better than those using simple ranks as it takes the variation of the auxiliary variable into account.

1.7 Thesis outline

The remainder of this thesis is organized as follows: Chapter 2 provides a review of relevant literature. In Chapter 3, we develop an estimator for estimating the mean of a finite population under simple random sampling, utilizing auxiliary information through tied ranks. Chapter 4 extends this concept to stratified random sampling. Finally, chapter 5 presents the conclusions drawn from the research conducted in this thesis.

Chapter 2

Literature Review

The purpose of this research is to derive an estimate for the mean of the finite population by incorporating auxiliary data. As many authors have contributed to this with an aim and have suggested improved estimators of the population parameters. Some of them are listed here.

Laplace (1820) was the first to use supplementary data to estimate the population. The work of Neyman (1938) is regarded as preliminary work in which supplementary data is used to improve an estimator. Numerous statisticians specializing in surveys have contributed thereafter to estimate the finite population mean and other population characteristics using auxiliary data to enhance efficiency of estimators. The traditional unbiased estimator for the mean per unit is represented as $\bar{y} = \sum_{i=1}^{n} y_i/n$, and its associated variance is provided as

$$Var(\bar{y}) = \lambda \bar{Y}^2 C_y^2 \tag{2.1}$$

where $\lambda = (1-f)/n$ and $C_y = S_y^2/\bar{Y}$ which is the coefficient of variation.

Cochran (1940) introduced the incorporation of an auxiliary variable in the estimation process and introduced the ratio estimator for the population mean. It's widely recognized that ratio-type estimators excel over the sample mean per unit estimator when there's a strong positive correlation between the study variable and an auxiliary variable, and the regression line goes through the origin. The usual ratio estimator is

$$\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}} \bar{X} \tag{2.2}$$

where $\bar{y}$ represents the sample mean of the study variable, $\bar{x}$ signifies the sample mean of the auxiliary variable, and $\bar{X}$ denotes the population mean of the auxiliary variable.

The Bias and MSE of usual estimator are

$$Bias(\hat{\bar{Y}}_R) = \lambda \bar{Y}[C_x^2 - \rho_{yx} \cdot Cx \cdot Cy] \tag{2.3}$$

$$MSE(\hat{\bar{Y}}_R) = \lambda \bar{Y}^2 (C_y^2 + C_x^2 - 2\rho_{yx} \cdot Cx \cdot Cy) \tag{2.4}$$

where $C_y = S_y^2/\bar{Y}$ and $C_x = S_x^2/\bar{X}$ are the population coefficient of variation and $\rho_{yx} = S_{yx}/S_y S_x$ be the correlation coefficient of Y with X.

The proposal for the linear regression estimator came from Hansen et al. (1953), and is defined as

$$\bar{Y}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) \tag{2.5}$$

$$V(\hat{\bar{Y}}_{lr}) = \lambda S_y^2(1 - \rho_{yx}^2) \tag{2.6}$$

where $b = S_{yx}/S_x^2$ is the regression coefficient.

The goal of survey statisticians has been to search for the best estimates of unknown population parameters. Khoshnevisan et al. (2007) proposed a comprehensive class of estimators denoted as $\hat{\bar{Y}}_K$, encompassing various adapted ratio-type estimators proposed by different authors as specific instances. This class is defined as

$$\hat{\bar{Y}}_K = \bar{y}\left[\frac{a\bar{X} + b}{\alpha(a\bar{x} + b) + (1 - \alpha)(a\bar{X} + b)}\right]^g, \tag{2.7}$$

where $a(\neq 0)$ and 'b' can be functions or known constants that depend on any population parameters known, including coefficients of skewness such as $(\beta_1(x), \beta_2(x))$, correlation coefficient ρ_{yx}, coefficient of variation C_x, etc.

Please observe that the mean per unit, ratio, and product estimators are instances of $\hat{\bar{Y}}_K$, given that $\alpha = a = 1$ and $b = 0$, when $g = 0, 1$ and -1, respectively.

In the framework of Simple Random Sampling Without Replacement (SRSWOR), we obtain expressions up to the first-order approximation.

The Bias and MSE of $\hat{\bar{Y}}_K$, where $\theta = a\bar{X}/(a\bar{X} + b)$, are given by

$$Bias(\hat{\bar{Y}}_K) = g\bar{Y}\alpha\theta\lambda\Big(gC_x^2\lambda\theta - \rho_{xy}C_xC_y\Big), \tag{2.8}$$

$$MSE(\hat{\bar{Y}}_K) = \bar{Y}^2\lambda C_y^2\Big(1 + g^2\alpha^2\theta^2\frac{C_x^2}{C_Y^2} - 2g\alpha\theta\rho_{xy}\frac{C_x}{C_y}\Big). \tag{2.9}$$

The optimal values of ($\alpha\theta$g) yield the minimum Mean Squared Error (MSE) for ($\hat{\bar{Y}}_K$), and this minimum MSE is expressed as

$$MSE_{min}\Big(\hat{\bar{Y}}_K\Big) \cong \lambda\bar{Y}^2C_y^2\Big(1 - \rho_{yx}^2\Big). \tag{2.10}$$

The minimum Mean Squared Error (MSE) coincides with the approximate variance of the conventional linear regression estimator.The empirical findings demonstrate that the suggested group of estimators exhibits greater efficiency compared to the currently available estimators.

Rao (1991) proposed an enhanced difference-type estimator, which is formulated as follows

$$\hat{\bar{Y}}_{R,D} = t_1\bar{y} + t_2\Big(\bar{X} - \bar{x}\Big) \tag{2.11}$$

where t_1 and t_2 are appropriately selected constants.

The Bias and MSE of $\hat{\bar{Y}}_{R,D}$ are

$$Bias(\hat{\bar{Y}}_{R,D}) \cong \bar{Y}\Big(t_1 - 1\Big) \tag{2.12}$$

$$MSE(\hat{\hat{Y}}_{R,D}) \cong t_1^2\bar{Y}^2 + t_1^2\bar{Y}^2\lambda C_y^2 + t_2^2\bar{X}^2\lambda C_x^2 + \bar{Y}^2 - 2t_1\bar{Y}^2 - 2t_1t_2\bar{Y}\bar{X}\lambda\rho_{xy}C_xC_y \tag{2.13}$$

Differentiating (2.12) with respect to t_1 and t_2, we get

$$t_{1(opt)} = \frac{1}{1 + \lambda C_y^2(1 - \rho_{yx}^2)}$$

$$t_{2(opt)} = \frac{\bar{Y}C_y\rho_{yx}}{\bar{X}C_x\left[1 + \lambda C_y^2(1 - \rho_{yx}^2)\right]}$$

The optimal values yield the minimum Mean Squared Error (MSE) for $\hat{\hat{Y}}_{R,D}$, and this minimum MSE is expressed as

$$MSE_{min}\left(\hat{\hat{Y}}_{R,D}\right) \cong \frac{\lambda\bar{Y}^2C_y^2\left(1 - \rho_{yx}^2\right)}{1 + \lambda C_y^2\left(1 - \rho_{yx}^2\right)} \tag{2.14}$$

Bahl and Tuteja (1991) was a trailblazer in the development of exponential estimators in simple random sampling. These estimators utilize an exponential function to estimate the mean of a finite population, making use of information from a single auxiliary variable. The suggested estimators are given by

$$\hat{\hat{Y}}_{BT,R} = \bar{y}\exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \tag{2.15}$$

$$\hat{\hat{Y}}_{BT,P} = \bar{y}\exp\left(\frac{\bar{x} - \bar{X}}{\bar{X} + \bar{x}}\right) \tag{2.16}$$

The Biases and MSE's of $\hat{\hat{Y}}_{BT,R}$ and $\hat{\hat{Y}}_{BT,P}$, respectively, are given by

$$Bias(\hat{\hat{Y}}_{BT,R}) = \bar{Y}\lambda\left[\frac{3C_x^2}{8} - \frac{\rho_{xy}C_xC_y}{2}\right] \tag{2.17}$$

$$Bias(\hat{\hat{Y}}_{BT,P}) = \bar{Y}\lambda\left[\frac{C_x^2}{8} - \frac{\rho_{xy}C_xC_y}{2}\right] \tag{2.18}$$

and

$$MSE(\hat{\hat{Y}}_{BT,R}) \cong \frac{\lambda\bar{Y}^2}{4}\left(4C_y^2 + C_x^2 - 4\rho_{xy}C_yC_x\right) \tag{2.19}$$

$$MSE(\hat{\hat{Y}}_{BT,P}) \cong \frac{\lambda\bar{Y}^2}{4}\left(4C_y^2 + C_x^2 + 4\rho_{xy}C_yC_x\right) \tag{2.20}$$

In numerous practical scenarios, these estimators have demonstrated superior efficiency and have been assessed for their precision in comparison to conventional mean per unit, ratio, and product estimators.

Building upon the research conducted by Bahl and Tuteja (1991) as a foundation, Singh et al. (2009) introduces a ratio-type exponential estimator for the estimation of the population mean of the variable of interest. The Mean Squared Error (MSE) equations for all the proposed estimators are derived and subsequently compared in the context of simple random sampling without replacement (SRSWOR). The proposed estimator is

$$\hat{\bar{Y}}_S = \bar{y}\exp\left(\frac{a(\bar{X}-\bar{x})}{a(\bar{X}+\bar{x})+2b}\right) \tag{2.21}$$

Note that both the estimators of Bahl and Tuteja (1991) are special cases of $\hat{\bar{Y}}_S$ when (a= 1, -1 and b= 0), respectively.

The bias and Mean Squared Error (MSE) of the proposed estimator are provided as follows.

$$Bias(\hat{\bar{Y}}_S) = \bar{Y}\theta\lambda\left[\frac{3\theta C_x}{2} - \rho_{xy}C_y\right] \tag{2.22}$$

$$MSE(\hat{\bar{Y}}_S) = \bar{Y}^2\lambda\left[\theta^2C_x^2 + C_y^2 - 2\theta\rho_{xy}C_xC_y\right] \tag{2.23}$$

The minimum Mean Squared Error (MSE) for $\hat{\bar{Y}}_S$ occurs at the optimal values, which are expressed as

$$MSE_{min}(\hat{Y}_S) \cong \lambda\bar{Y}^2C_Y^2(1-\rho_{yx}^2) \tag{2.24}$$

Grover and Kaur (2014) proposed a versatile category of exponential estimators based on ratios by combining the estimators of Rao (1991) and Singh et al. (2009), given as

$$\hat{\bar{Y}}_{GK,G} = \left[\omega_1\bar{y} + \omega_2(\bar{X}-\bar{x})\right]\exp\left(\frac{a\left(\bar{X}-\bar{x}\right)}{a(\bar{X}+\bar{x})+2b}\right) \tag{2.25}$$

where ω_1 and ω_2 are appropriately selected constants.

The minimum MSE of $\hat{\bar{Y}}_{GK,G}$ at the optimum values,

$$\omega_{1(opt)} = \frac{8-\lambda\theta^2C_x^2}{8[1+\lambda C_y^2(1-\rho_{yx}^2)]}$$

and

$$\omega_{2(opt)} = \frac{\bar{Y}[\lambda\theta^3C_x^3 + 8C_y\rho_{yx} - \lambda\theta^2C_x^2C_y\rho_{yx} - 4\theta C_x(1-\lambda C_y^2(1-\rho_{yx}^2))]}{8\bar{X}C_x[1+\lambda C_y^2(1-\rho_{yx}^2)]}$$

is given by

$$MSE_{min}\left(\hat{\bar{Y}}_{GK,G}\right) \cong \frac{\lambda\bar{Y}^2[64C_y^2(1-\rho_{yx}^2) - \lambda\theta^4C_x^4 - 16\lambda\theta^2C_x^2C_y^2(1-\rho_{yx}^2)]}{64[1+\lambda C_y^2(1-\rho_{yx}^2)]} \tag{2.26}$$

A simulation study was conducted to assess the comparative performance of the proposed estimator in relation to several existing estimators. The findings from the study demonstrate that the proposed estimator exhibits higher efficiency when compared to the existing ones.

In various methodologies and practical applications, traditional ratio, product, and classical regression estimators are commonly employed to estimate unknown population parameters, under the condition of a substantial correlation between the study variable and the auxiliary variable.

When there is a correlation between the study variable and the auxiliary variable, the ranks of the auxiliary variable also exhibit a correlation with the values of the study variable. Consequently, the ranked auxiliary variable can be regarded as a novel auxiliary variable. Based on these ideas, Haq et al. (2017) Suggested an enhanced estimator for estimating the mean of a finite population. The author suggested the following estimator.

$$\hat{\hat{Y}}^{*}_{Pr} = \left[\omega_1 \bar{y} + \omega_2(\bar{X} - \bar{x}) + \omega_3(\bar{R}_x - \bar{r}_x)\right] \exp\left(\frac{a(\bar{X} - \bar{x})}{a(\bar{X} + \bar{x}) + 2b}\right) \tag{2.27}$$

where a and b are constants. The first-order approximation yields the bias and Mean Squared Error (MSE) of the proposed estimator as

$$Bias(\hat{\hat{Y}}^{*}_{Pr}) = \frac{1}{8}\left[-8\bar{Y} + 4\lambda\theta C_x(\bar{X}C_x\omega_2 + \bar{R}_x C_r \omega_3 \rho_{xr_x}) + \bar{Y}\omega_1[8 + \lambda\theta C_x(3\theta C_x - 4C_y\rho_{yx})]\right] \tag{2.28}$$

$$\begin{aligned} MSE(\hat{\hat{Y}}^{*}_{Pr}) = {} & \bar{Y}^2 + \lambda\bar{X}C_x^2\omega_2(-\bar{Y}\theta + \bar{X}\omega_2) + \lambda\bar{R}_x^2 C_r^2 \omega_3^2 + \lambda\bar{R}_x C_x C_r(-\bar{Y}\theta + 2\bar{X}\omega_2)\omega_3\rho_{xr_x} \\ & + \bar{Y}^2\omega_1^2\left[1 + \lambda\left(C_y^2 + \theta C_x(\theta C_x - 2C_y\rho_{yx})\right)\right] \\ & + \frac{1}{4}\bar{Y}\omega_1\Big[-8\bar{Y} + \lambda C_x\Big(\theta C_x(-3\bar{Y}\theta + 8\bar{X}\omega_2) \\ & + 8\bar{R}_x\theta C_r\omega_3\rho_{xr_x} + 4C_y(\bar{Y}\theta - 2\bar{X}\omega_2)\rho_{yx}\Big) - 8\bar{R}_x\lambda C_y C_r \omega_3 \rho_{yr_x}\Big] \end{aligned} \tag{2.29}$$

The values of ω_1, ω_2, and ω_3 that optimize the expression (2.29) are determined by minimizing it, resulting in the following values.

$$\omega_{1(opt)} = \frac{8 - \lambda\theta^2 C_x^2}{8\left\{1 + \lambda C_y^2\left(1 - Q^2_{y.xr_x}\right)\right\}}$$

$$\omega_{2(opt)} = \frac{\bar{Y}\left[\begin{array}{c} \lambda\theta^3 C_x^3\left(-1 + \rho^2_{xr_x}\right) + (-8C_y + \lambda\theta^2 C_x^2 C_y)\left(\rho_{yx} - \rho_{xr_x}\rho_{yr_x}\right) \\ +4\theta C_x\left(-1 + \rho^2_{xr_x}\right)\left\{-1 + \lambda C_y^2\left(1 - Q^2_{y\cdot xr_x}\right)\right\} \end{array}\right]}{8\bar{X}C_x\left(-1 + \rho^2_{xr_x}\right)\left\{1 + \lambda C_y^2\left(1 - Q^2_{y\cdot xr_x}\right)\right\}}$$

and

$$\omega_{3(opt)} = \frac{\bar{Y}\left(8 - \lambda\theta^2 C_x^2\right)C_y\left(\rho_{xr_x}\rho_{yx} - \rho_{yr_x}\right)}{8\bar{R}_x C_r\left(-1 + \rho^2_{xr_x}\right)\left\{1 + \lambda C_y^2\left(1 - y \cdot xr_x{}^2\right)\right\}}$$

Upon substituting the optimal values of ω_1, ω_2, and ω_3 into equation (2.29) and performing some simplifications, we arrive at the minimum Mean Squared Error (MSE) for $\hat{\hat{Y}}^{*}_{Pr}$, which is expressed as

$$\text{MSE}_{\min}\left(\hat{\hat{Y}}^{*}_{P_r}\right) \cong \frac{\lambda\bar{Y}^2\left\{64C_y^2\left(1 - Q^2_{y.xr_x}\right) - \lambda\theta^4 C_x^4 - 16\lambda\theta^2 C_x^2 C_y^2\left(1 - Q^2_{y.xr_x}\right)\right\}}{64\left\{1 + \lambda C_y^2\left(1 - Q^2_{y.xx_x}\right)\right\}}. \tag{2.30}$$

Through a combination of theoretical analysis and numerical results, it is evident that the proposed estimator exhibits greater efficiency when compared to conventional estimators such as the mean, ratio, product, exponential-ratio, exponential-product, and classical regression estimators.

Javed and Irfan (2020) extended the idea of Haq et al. (2017) to the newly optimized exponential-type estimators for estimating the mean of a finite population, utilizing an auxiliary variable in a stratified random sampling framework. The dual utilization of the auxiliary variable encompasses both the original auxiliary information and the ranked auxiliary information. The initial proposed estimator is formulated as follows

$$\hat{\hat{Y}}_{P1} = \frac{u_{11}}{2}\left(\frac{\bar{X}}{\bar{x}_{st}} + \frac{\bar{x}_{st}}{\bar{X}}\right)\hat{\bar{Y}}_{BT_{st},AVG} + u_{12}(\bar{Z} - \bar{z}_{st}) + u_{13}(\bar{X} - \bar{x}_{st})\exp\left(\frac{\bar{X} - \bar{x}_{st}}{\bar{X} + \bar{x}_{st}}\right) \tag{2.31}$$

where u_{11}, u_{12} and u_{13} represent appropriately selected weights.

The Bias and MSE of $\hat{\hat{Y}}_{P1}$ are as follows

$$Bias(\hat{\hat{Y}}_{P1}) \cong (u_{11} - 1)\bar{Y} + \frac{V_{002}}{2}\left(u_{13}\bar{X} + \frac{5u_{11}\bar{Y}}{4}\right) \tag{2.32}$$

and

$$\begin{aligned} MSE(\hat{\hat{Y}}_{P1}) \cong \bar{Y}^2\Big[1 &+ u_{11}^2\Big(1 + V_{200} + \frac{5}{4}V_{002}\Big) + u_{12}^2\tilde{R}^2V_{020} + u_{13}^2R'^2V_{002} - u_{11}\Big(2 + \frac{5}{4}V_{002}\Big) \\ &+ 2u_{12}\tilde{R}\Big(u_{13}R'V_{011} - u_{11}V_{110}\Big) - u_{13}R'V_{002}\Big(2k'u_{11} - u_{11} + 1\Big)\Big] \end{aligned} \tag{2.33}$$

The optimum weights u_{11}, u_{12} and u_{13} were obtained by minimizing (2.33) and are given as follows

$$u_{11(opt)} = \frac{E_1E_2 - 2V_{002}V_{020}E_3}{E_2E_4 - 2E_3^2}$$

$$u_{12(opt)} = \frac{2V_{110}(E_1E_2 - 2V_{002}V_{020}E_3) + V_{011}(E_1E_3 - V_{002}V_{020}E_4)}{2\tilde{R}V_{020}(E_2E_4 - 2E_3^2)}$$

and

$$u_{13(opt)} = \frac{V_{002}V_{020}E_4 - E_1E_3}{2R'(E_2E_4 - 2E_3^2)}$$

Upon replacing the optimal values of u_{11}, u_{12}, and u_{13} into equation (2.33) and simplifying further, we obtain the minimum Mean Squared Error (MSE) for $\hat{\hat{Y}}_{P1}$, given by

$$\begin{aligned} MSE_{min(\hat{\hat{Y}}_{P1(st)})} \cong \frac{\bar{Y}^2}{4V_{020}F_1^2}\Big[& 4V_{020}F_1^2 + (4V_{020} + 4V_{020}V_{200} - 4V_{110}^2 + 5V_{002}V_{020})F_2^2 \\ & + (V_{002}V_{020} - V_{011}^2)F_3^2 - V_{020}(8 + 5V_{002})F_1F_2 \\ & + 2(2V_{110}V_{011} - V_{020}V_{002}\big(2\acute{k} - 1\big))F_2F_3 - 2V_{002}V_{020}F_1F_3\Big] \end{aligned} \tag{2.34}$$

The second Proposed estimator is given as

$$\hat{\hat{Y}}_{P2} = u_{14}\bar{y}_{st} + u_{15}(\bar{Z} - \bar{z}_{st}) + u_{16}(\bar{X} - \bar{x}_{st}) \exp\left(\frac{2(\bar{X} - \bar{x}_{st})}{\bar{X} + \bar{x}_{st}}\right) \tag{2.35}$$

The bias and MSE of $\hat{\hat{Y}}_{P2}$ are

$$Bias(\hat{\hat{Y}}_{P2}) \cong \bar{Y}\Big[(u_{14} - 1) + u_{16}V_{002}\acute{R}\Big] \tag{2.36}$$

$$\begin{aligned} MSE(\hat{\hat{Y}}_{P2}) \cong \bar{Y}^2\Big[& 1 + u_{14}^2(1 + V_{200}) + u_{15}^2V_{020}\tilde{R}^2 + u_{16}^2V_{002}\acute{R}^2 - 2u_{14} - 2u_{16}V_{002}\acute{R} \\ & - 2u_{14}u_{16}V_{002}\acute{R}(\acute{k} - 1) + 2u_{15}u_{16}\acute{R}\tilde{R}V_{011} - 2u_{14}u_{15}\tilde{R}V_{110}\Big]. \end{aligned} \tag{2.37}$$

The minimum Mean Squared Error (MSE) for $\hat{\hat{Y}}_{P2}$ is achieved at the optimal values, and it is expressed as

$$\begin{aligned} MSE_{min}\Big(\hat{\hat{Y}}_{P2(st)}\Big) \cong \frac{\bar{Y}^2}{(1 + V_{200})F_4^2}\Big[& (1 + V_{200})F_4^2 + V_{002}(1 + V_{200})F_5^2 + V_{020}(1 + V_{200})F_6^2 \\ & + F_7^2 - 2(V_{002}F_4 - V_{011}F_6)(1 + V_{200}F_5) \\ & - 2(V_{110}F_6 + V_{002}(\acute{k} - 1)F_5 + F_4)F_7)\Big] \end{aligned} \tag{2.38}$$

The proposed estimators are underscored through empirical analysis using real-world data. The numerical outcomes from these proposed estimators demonstrate their superior efficiency compared to competing methods.

Chapter 3

Estimator of Finite Population Mean Using Tied Ranks

The accuracy of an estimator can be enhanced through appropriate utilization of auxiliary information, both during the estimation process and during the design phase. Usually, a sampling frame is available along with the complete information of auxiliary variables in the socio-economic and natural surveys. The symbols and notations used in subsequent derivations are given in the following section.

3.1 Notations and symbols

Let N be the finite size of population consisting of units $U = \{U_1, U_2, ..., U_N\}$. Let a simple random sample of size n be drawn from a population of size U without replacement. It is assumed that the complete auxiliary variable information e.g. ranks of the auxiliary variable and X is available and also the population parameters of the auxiliary X including mean $\bar{X}$, coefficient of variation, coefficient of kurtosis $B_2(x)$ are also known. Let (y_i, x_i) be the values of the study variable Y and auxiliary variable X for the ith unit of the population, respectively. Let $\bar{Y}=\sum_{i=1}^{N} y_i/N$ and $\bar{X}=\sum_{i=1}^{N} x_i/N$ be the population means of study variable and auxiliary variables. Let $S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/(N-1)$ and $S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2/(N-1)$ be the corresponding population variances of the study variable and auxiliary variable, respectively. Let $\bar{y}=\sum_{i=1}^{n} y_i/n$ and $\bar{x}=\sum_{i=1}^{n} x_i/n$ be the sample means corresponding to population means. In the same way, let $s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$ and $s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$ be the corresponding sample variances. Also, let $\hat{C}_y = s_y^2/\bar{y}$ and $\hat{C}_x = s_x^2/\bar{x}$ be the respective sample coefficient of variations corresponding of the population coefficient of variations $C_y = S_y^2/\bar{Y}$ and $C_x = S_x^2/\bar{X}$. Let $\widehat{\rho}_{yx}$ be the sample correlation coefficient corresponding to the population correlation coefficient ρ_{xy}.

3.2 Proposed estimator

When a substantial correlation exists between the study variable and the auxiliary variable, it is reasonable to anticipate a correlation between the values of the study variable and the ranks assigned to the auxiliary variable. Consequently, the variable containing these ranks can be denoted as a novel auxiliary variable, referred to as the ranked auxiliary variable. The efficiency of the estimator improves when it incorporates information from the ranked auxiliary variable. Building on this concept, Haq et al. (2017) introduced a novel estimator for estimating the mean of a finite population.

The estimator given by Haq et al. (2017) uses the idea of ranking which are equally spaced values. Equally spaced ranks ignore the variation that exists within the population of auxiliary variables. For example, consider a small population of patients in a hospital whose ages are $12, 23, 23, 23, 31, 39, 39, 41, 45$, and 50. The usual ranks of these values are $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ which are equally spaced and ignore the variation of the auxiliary variable. We proposed a technique that deals with tied ranks. The tied ranks of the above population are $1, 3, 3, 3, 5, 6.5, 6.5, 8, 9, 10$. This ranking scheme incorporates the variation of the auxiliary variable. The suggested estimator integrates auxiliary information in the form of both a ranked auxiliary variable and another auxiliary variable. Furthermore, the proposed estimator makes advantage of auxiliary data to improve efficiency and also decrease the variation of the estimator.

Recalling that in the underlying finite population of U, x_1, x_2,...., x_N are the N values of X. Let r_{x1}, r_{x2}, ..., r_{xN} be the tied ranks of the X denoted by R_x. Let S^2_{rx}, $\bar{R}_x$ and $\bar{r}_x$ be the population variance, population mean and sample mean of R_x, respectively where $S^2_{rx} = \sum_{i=1}^{N}(r_{x,i} - \bar{R}_x)^2/(N-1)$, $\bar{R}_x = \sum_{i=1}^{N} r_{x,i}/N = (N+1)/2$ and $\bar{r}_x = \sum_{i=1}^{n} r_{x,i}/n$ where $r_{x,i}$ denotes the ith value of the R_x in the population U. Let the correlation coefficient between the Z and R_x be the $\rho_{zr_x} = S_{zr_x}/S_z S_{r_x}$ where $S_{zr_x} = \sum_{i=1}^{N}(Z_i - \bar{Z})(r_{x,i} - \bar{R}_x)/(N-1)$ is the population covariance between Z and R_x for $Z = Y, X$. Let the coefficient of variation of R_x be $C_r = S_{r_x}/\bar{R}_x$.

Table 3.1 lists specific members of the classes of existing estimator and the proposed estimator. Using the values of a and b, the sub cases of $\hat{\bar{Y}}^*_{Haq}$ and $\hat{\bar{Y}}^*_{Pr}$ can be developed.

The new proposed estimator based on tied ranks of the auxiliary variable is

$$\hat{\bar{Y}}^*_{Pr} = [\omega_1 \bar{y} + \omega_2(\bar{X} - \bar{x}) + \omega_3(\bar{R}_x - \bar{r_x})] \exp\left(\frac{a(\bar{X} - \bar{x})}{a(\bar{X} + \bar{x}) + 2b}\right), \tag{3.1}$$

where a and b are suitable constants.

3.3 Properties of proposed estimator

In order to derive the properties, namely bias and Mean Squared Error (MSE), of the proposed estimator, we define the following relative error terms based on the simple random sample.

Table 3.1: Some members of the existing and proposed estimator.

a	b	$\hat{\bar{Y}}^{*}_{Haq}$	$\hat{\bar{Y}}^{*}_{Pr}$
1	C_X	$\hat{\bar{Y}}^{(1)}_{Haq}$	$\hat{\bar{Y}}^{*(1)}_{Pr}$
1	$\beta_2(X)$	$\hat{\bar{Y}}^{(2)}_{Haq}$	$\hat{\bar{Y}}^{*(2)}_{Pr}$
$\beta_2(X)$	C_X	$\hat{\bar{Y}}^{(3)}_{Haq}$	$\hat{\bar{Y}}^{*(3)}_{Pr}$
C_X	$\beta_2(X)$	$\hat{\bar{Y}}^{(4)}_{Haq}$	$\hat{\bar{Y}}^{*(4)}_{Pr}$
1	ρ_{YX}	$\hat{\bar{Y}}^{(5)}_{Haq}$	$\hat{\bar{Y}}^{*(5)}_{Pr}$
C_X	ρ_{YX}	$\hat{\bar{Y}}^{(6)}_{Haq}$	$\hat{\bar{Y}}^{*(6)}_{Pr}$
ρ_{YX}	C_X	$\hat{\bar{Y}}^{(7)}_{Haq}$	$\hat{\bar{Y}}^{*(7)}_{Pr}$
$\beta_2(X)$	ρ_{YX}	$\hat{\bar{Y}}^{(8)}_{Haq}$	$\hat{\bar{Y}}^{*(8)}_{Pr}$
ρ_{YX}	$\beta_2(X)$	$\hat{\bar{Y}}^{(9)}_{Haq}$	$\hat{\bar{Y}}^{*(9)}_{Pr}$
1	$N\bar{X}$	$\hat{\bar{Y}}^{(10)}_{Haq}$	$\hat{\bar{Y}}^{*(10)}_{Pr}$

$e_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$, $e_1 = \frac{\bar{x}-\bar{X}}{\bar{X}}$ and $e_2 = \frac{\bar{r_x}-\bar{R_x}}{\bar{R_x}}$, such that $E(e_i) = 0$ for $i = 0, 1, 2$, i.e,

$$E(e_0) = E(e_1) = E(e_2) = 0 \tag{3.2}$$

It is easy to compute the following relative error terms as:

$$\begin{aligned}
E(e_0^2) &= E\left(\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right)^2 = \frac{1}{\bar{Y}^2}E(\bar{y}-\bar{Y})^2 = \frac{\lambda var(\bar{Y})}{\bar{Y}^2} = \lambda C_y^2,\\
E(e_1^2) &= E\left(\frac{\bar{x}-\bar{X}}{\bar{X}}\right)^2 = \frac{1}{\bar{X}^2}E(\bar{x}-\bar{X})^2 = \frac{\lambda var(\bar{X})}{\bar{X}^2} = \lambda C_x^2,\\
E(e_2^2) &= E\left(\frac{\bar{r_x}-\bar{R_x}}{\bar{R_x}}\right)^2 = \frac{1}{\bar{R_x}^2}E(\bar{r_x}-\bar{R_x})^2 = \frac{\lambda var(\bar{R_x})}{\bar{R_x}^2} = \lambda C_r^2,\\
E(e_0e_1) &= E\left(\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{x}-\bar{X}}{\bar{X}}\right) = \frac{1}{\bar{X}\bar{Y}}E(\bar{y}-\bar{Y})(\bar{x}-\bar{X}) = \frac{1}{\bar{Y}\bar{X}}Cov(\bar{y},\bar{x})\\
&= \frac{1}{\bar{Y}\bar{X}}\lambda\rho_{xy}S_xS_y = \lambda\rho_{xy}C_xC_y,\\
E(e_0e_2) &= E\left(\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{r}_x-\bar{R}_x}{\bar{R}_x}\right) = \frac{1}{\bar{R}_x\bar{Y}}E(\bar{y}-\bar{Y})(\bar{r}_x-\bar{R}_x) = \frac{1}{\bar{Y}\bar{R}_x}Cov(\bar{y},\bar{r}_x)\\
&= \frac{1}{\bar{Y}\bar{R}_x}\lambda\rho_{r_xy}S_{r_x}S_y = \lambda\rho_{yr_x}C_yC_{r_x},\\
E(e_1e_2) &= E\left(\frac{\bar{x}-\bar{X}}{\bar{X}}\right)\left(\frac{\bar{r}_x-\bar{R}_x}{\bar{R}_x}\right) = \frac{1}{\bar{R}_x\bar{X}}E(\bar{x}-\bar{X})(\bar{r}_x-\bar{R}_x) = \frac{1}{\bar{X}\bar{R}_x}Cov(\bar{x},\bar{r_x})\\
&= \frac{1}{\bar{X}\bar{R}_x}\lambda\rho_{xr_x}S_xS_{r_x} = \lambda\rho_{xr_x}C_xC_{r_x}.
\end{aligned}$$

To determine the Bias and Mean Squared Error (MSE) of the suggested estimator, we can express the proposed estimator $\hat{\hat{Y}}^*_{Pr}$ in terms of the relative error terms, yielding

$$\hat{\hat{Y}}^*_{Pr} = \Big[\omega_1\bar{y} + \omega_2(\bar{X} - \bar{x}) + \omega_3(\bar{R}_x - \bar{r}_x)\Big] \exp\left(\frac{a(\bar{X} - \bar{x})}{a(\bar{X} + \bar{x}) + 2b}\right)$$

$$\hat{\hat{Y}}^*_{Pr} = \Big[\omega_1\bar{Y}(1 + e_0) + \omega_2(\bar{X} - \bar{X}(1 + e_1)) + \omega_3(\bar{R}_x - \bar{R}_x(1 + e_2))\Big]$$
$$\exp\left(\frac{a(\bar{X} - \bar{X}(1 + e_1))}{a(\bar{X} + \bar{X}(1 + e_1)) + 2b}\right)$$

$$\hat{\hat{Y}}^*_{Pr} = \Big[\omega_1\bar{Y}(1 + e_0) - \omega_2\bar{X}e_1 - \omega_3\bar{R}_x e_2\Big] \exp\left(\frac{-a\bar{X}e_1}{2a\bar{X} + a\bar{X}e_1 + 2b}\right) \tag{3.3}$$

Now consider the exponential term

$$\begin{aligned}
\exp\left(\frac{-a\bar{X}e_1}{2a\bar{X} + a\bar{X}e_1 + 2b}\right) &= \exp\left(\frac{-a\bar{X}e_1}{(2a\bar{X} + 2b)\Big(1 + \frac{a\bar{X}e_1}{2a\bar{X}+2b}\Big)}\right) \\
&= \exp\left(\frac{-\theta e_1}{2(1 + \frac{\theta e_1}{2})}\right) \\
&= \exp\left(\frac{-\theta e_1}{2}\Big(\frac{1 + \theta e_1}{2}\Big)^{-1}\right) \\
&= \exp\left(\frac{-\theta e_1}{2}\Big(1 - \frac{\theta e_1}{2} + \frac{\theta^2 e_1^2}{4} - \frac{\theta^3 e_1^3}{8} + ...\Big)\right) \\
&\approx \exp\left(\frac{-\theta e_1}{2} + \frac{\theta^2 e_1^2}{4}\right) \\
&\approx 1 - \Big(\frac{\theta e_1}{2} + \frac{\theta^2 e_1^2}{4}\Big) + \frac{(\frac{-\theta e_1}{2} + \frac{\theta^2 e_1^2}{4})^2}{2} \\
&\approx 1 - \frac{\theta e_1}{2} + \frac{\theta^2 e_1^2}{4} + \frac{\theta^2 e_1^2}{8} \\
&\approx 1 - \frac{\theta e_1}{2} + \frac{3\theta^2 e_1^2}{8}.
\end{aligned}$$

We consider the relative error terms up to the first-order approximation. Also $\theta = -a\bar{X}/(a\bar{X} + b)$.

Substituting the exponential term in (3.3), we get

$$\hat{\hat{Y}}^*_{Pr} = \Big[\omega_1\bar{Y}(1 + e_0) - \omega_2\bar{X}e_1 - \omega_3\bar{R}_x e_2\Big]\Big[1 - \frac{\theta e_1}{2} + \frac{3\theta^2 e_1^2}{8}\Big]. \tag{3.4}$$

Expanding (3.4) and keeping the terms up to the order two in e_is, we can write

$$\begin{aligned}(\hat{\hat{Y}}^*_{Pr} - \bar{Y}) = &-\bar{Y} + \bar{Y}\omega_1 + \bar{Y}\omega_1 e_0 - \bar{X}\omega_2 e_1 - \bar{R}_x\omega_3 e_2 - \frac{\bar{Y}\omega_1\theta e_1}{2} - \frac{\bar{Y}\omega_1\theta e_0 e_1}{2} \\ &+ \frac{\bar{X}\omega_2\theta e_1^2}{2} + \frac{\bar{R}_x\omega_3\theta e_1 e_2}{2} + \frac{3\bar{Y}\omega_1\theta^2 e_1^2}{8}.\end{aligned} \tag{3.5}$$

Now taking expectations on both sides of equation (3.5), We get

$$\begin{aligned}E(\hat{\hat{Y}}^*_{Pr} - \bar{Y}) &= Bias(\hat{\hat{Y}}^*_{Pr}) \\ &\cong -\bar{Y} + \bar{Y}\omega_1 + \bar{Y}\omega_1 E(e_0) - \bar{X}\omega_2 E(e_1) - \bar{R}_x\omega_3 E(e_2) - \frac{\bar{Y}\omega_1\theta E(e_1)}{2} \\ &- \frac{\bar{Y}\omega_1\theta E(e_0 e_1)}{2} + \frac{\bar{X}\omega_2\theta E(e_1)^2}{2} + \frac{\bar{R}_x\omega_3\theta E(e_1 e_2)}{2} + \frac{3\bar{Y}\omega_1\theta^2 E(e_1)^2}{8}\end{aligned}$$

Substituting the results of error terms computed earlier in the above expression

$$\begin{aligned}Bias(\hat{\hat{Y}}^*_{Pr}) &\cong -\bar{Y} + \bar{Y}\omega_1 - \frac{\bar{Y}\omega_1\theta\lambda\rho_{xy}C_xC_y}{2} + \frac{\bar{X}\omega_2\theta\lambda C_x^2}{2} + \frac{\bar{R}_x\omega_3\theta\lambda\rho_{xr_x}C_xC_{r_x}}{2} + \frac{3\bar{Y}\omega_1\theta^2\lambda C_x^2}{8} \\ &\cong -\bar{Y} + \frac{\bar{X}\omega_2\theta\lambda C_x^2}{2} + \frac{\bar{R}_x\omega_3\theta\lambda\rho_{xr_x}C_xC_{r_x}}{2} + \bar{Y}\omega_1 - \frac{\bar{Y}\omega_1\theta\lambda\rho_{xy}C_xC_y}{2} + \frac{3\bar{Y}\omega_1\theta^2\lambda C_x^2}{8} \\ &\cong \frac{1}{8}\Big[-8\bar{Y} + 4\theta\lambda C_x(\omega_2\bar{X}C_x + \omega_3\bar{R}_x\rho_{xr_x}C_{r_x}) + \bar{Y}\omega_1(8 - 4\theta\lambda\rho_{xy}C_xC_y + 3\lambda\theta^2C_x^2)\Big]\end{aligned}$$

The Bias of the suggested estimator can be expressed as follows.

$$\begin{aligned}Bias\Big(\hat{\hat{Y}}^*_{Pr}\Big) \cong \frac{1}{8}\Big[&-8\bar{Y} + 4\lambda\theta C_x\Big(\bar{X}C_x\omega_2 + \bar{R}_xC_{r_x}\omega_3\rho_{xr_x}\Big) \\ &+ \bar{Y}\omega_1\Big[8 + \lambda\theta C_x\Big(3\theta C_x - 4C_y\rho_{yx}\Big)\Big]\Big].\end{aligned} \tag{3.6}$$

Taking the square of (3.5) to find out its MSE,

$$\begin{aligned}(\hat{\hat{Y}}^*_{Pr} - \bar{Y})^2 \cong \Big(&-\bar{Y} + \bar{Y}\omega_1 + \bar{Y}\omega_1 e_0 - \bar{X}\omega_2 e_1 - \bar{R}_x\omega_3 e_2 - \frac{\bar{Y}\omega_1\theta e_1}{2} - \frac{\bar{Y}\omega_1\theta e_0 e_1}{2} \\ &+ \frac{\bar{X}\omega_2\theta e_1^2}{2} + \frac{\bar{R}_x\omega_3\theta e_1 e_2}{2} + \frac{3\bar{Y}\omega_1\theta^2 e_1^2}{8}\Big)^2.\end{aligned}$$

Expanding and keeping the terms up to order two in e_is, we get

$$\begin{aligned}(\hat{\hat{Y}}^*_{Pr} - \bar{Y})^2 \cong\ &\bar{Y}^2 + \bar{Y}^2\omega_1^2 + \bar{Y}^2\omega_1^2 e_0^2 + \bar{X}^2\omega_2^2 e_1^2 + \bar{R_x}^2\omega_3^2 e_2^2 - \frac{3\bar{Y}^2\omega_1\theta^2 e_1^2}{4} - 2\bar{Y}^2\omega_1 \\ &- 2\omega_1\omega_2\bar{X}\bar{Y}e_0e_1 - 2\omega_1\omega_3\bar{Y}\bar{R}_xe_0e_2 - 2\omega_1^2\bar{Y}^2\theta e_0e_1 + 2\omega_2\omega_3\bar{X}\bar{R}_xe_1e_2 \\ &+ 2\omega_1\omega_2\bar{X}\bar{Y}e_1^2\theta + \omega_1\bar{Y}^2\theta e_0e_1 - \bar{Y}\bar{X}\omega_2\theta e_1^2 - \omega_3\bar{Y}\bar{R}_x\theta e_1e_2 \\ &+ \omega_1^2\bar{Y}^2\theta^2e_1^2 + 2\omega_1\omega_3\bar{Y}\bar{R}_x\theta e_1e_2\end{aligned}$$

After applying the expectation operator to both sides and simplifying the equation, we obtain

$$
\begin{aligned}
E(\hat{\hat{Y}}^*_{Pr}-\bar{Y})^2 &= MSE(\hat{\hat{Y}}^*_{Pr})\\
&\cong \bar{Y}^2+\lambda\bar{X}C_x^2\omega_2(-\bar{Y}\theta+\bar{X}\omega_2)+\lambda\bar{R_x}^2C_r^2\omega_3^2+\lambda\bar{R}_xC_xC_r(-\bar{Y}\theta+2\bar{X}\omega_2)\omega_3\rho_{xr_x}\\
&+\bar{Y}^2\omega_1^2\Big[1+\lambda\Big(C_y^2+\theta C_x(\theta C_x-2C_y\rho_{yx})\Big)\Big]\\
&+\frac{1}{4}\bar{Y}\omega_1\Big[-8\bar{Y}+\lambda C_x\Big(\theta C_x(-3\bar{Y}\theta+8\bar{X}\omega_2)+8\bar{R}_x\theta C_r\omega_3\rho_{xr_x}\\
&+4C_y(\bar{Y}\theta-2\bar{X}\omega_2)\rho_{yx}\Big)-8\bar{R}_x\lambda C_yC_r\omega_3\rho_{yr_x}\Big]
\end{aligned}
\tag{3.7}
$$

We calculated the Bias and Mean Squared Error (MSE) up to the first-order approximation. To find the optimal values of ω_1, ω_2, and ω_3, we minimized equation 3.7 with respect to w_i where i ranges from 1 to 3. We set the derivative of the equation to zero and solved for w_i simultaneously, resulting in the following optimal values.

$$
\omega_{1(opt)}=\frac{8-\lambda\theta^2C_x^2}{8\Big[1+\lambda C_y^2(1-Q^2_{y.xr_x})\Big]}
\tag{3.8}
$$

$$
\omega_{2(opt)}=\frac{\bar{Y}\left[\begin{array}{r}\lambda\theta^3C_x^3(-1+\rho^2_{xr_x})+(-8C_y+\lambda\theta^2C_x^2C_y)(\rho_{yx}-\rho_{xr_x}\rho_{yr_x})\\ +4\theta C_x(-1+\rho^2_{xr_x})\Big(-1+\lambda C_y^2(1-Q^2_{y.xr_x})\Big)\end{array}\right]}{8\bar{X}C_x(-1+\rho^2_{xr_x})\Big[1+\lambda C_y^2(1-Q^2_{y.xr_x})\Big]}
\tag{3.9}
$$

and

$$
\omega_{3(opt)}=\frac{\bar{Y}(8-\lambda\theta^2C_x^2)C_y(\rho_{xr_x}\rho_{yx}-\rho_{yr_x})}{8\bar{R}_xC_r(-1+\rho^2_{xr_x})\Big[1+\lambda C_y^2(1-Q^2_{y.xr_x})\Big]}
\tag{3.10}
$$

where $Q^2_{y.xr_x}$ is the coefficient of multiple determination of Y on X and R_x in simple random sampling. It is given as

$$
Q^2_{y.xr_x}=\frac{\rho^2_{yx}+\rho^2_{yr_x}-2\rho_{yx}\rho_{yr_x}\rho_{xr_x}}{1-\rho^2_{xr_x}}
$$

Now, by substituting the optimal values of ω_1, ω_2, and ω_3 into equation (3.7) and conducting some algebraic simplifications, we arrive at the minimum Mean Squared Error (MSE) for the proposed estimator $\hat{\hat{Y}}^*_{Pr}$ as

$$
MSE_{min}\Big(\hat{\hat{Y}}^*_{Pr}\Big)\cong\frac{\lambda\bar{Y}^2\Big[64C_y^2\Big(1-Q^2_{y.xr_x}\Big)-\lambda\theta^4C_x^4-16\lambda\theta^2C_x^2C_y^2\Big(1-Q^2_{y.xr_x}\Big)\Big]}{64\Big[1+\lambda C_y^2\Big(1-Q^2_{y.xr_x}\Big)\Big]}.
\tag{3.11}
$$

3.4 Theoretical comparison

In this section, we have contrasted the minimum Mean Squared Error (MSE) of the proposed estimator with that of several existing estimators. The obtained expressions are given in the following section.

(a) Through a comparison between the proposed estimator and the traditional unbiased estimator, we have

$$MSE_{min}(\hat{\bar{Y}}^*_{Pr}) < MSE(\hat{\bar{Y}}_R)$$

$$MSE(\hat{\bar{Y}}_R) - MSE_{min}(\hat{\bar{Y}}^*_{Pr}) > 0$$

The condition is true if and only if

$$\frac{\lambda\bar{Y}^2\Big[\lambda\theta^4C_x^4 + 16C_y^2\Big(4Q^2_{y.xr_x} + \lambda\theta^2C_x^2(1-Q^2_{y.xr_x})\Big) + 64\lambda C_y^4(1-Q^2_{y.xr_x})\Big]}{64\Big[1+\lambda C_y^2(1-Q^2_{y.xr_x})\Big]} > 0 \tag{3.12}$$

(b) By comparing the proposed estimator $(\hat{\bar{Y}}^*_{Pr})$ and the usual ratio estimator $(\hat{\bar{Y}}_R)$, we have

$$MSE_{min}(\hat{\bar{Y}}^*_{Pr}) < MSE(\hat{\bar{Y}}_R).$$

$$MSE(\hat{\bar{Y}}_R) - MSE_{min}(\hat{\bar{Y}}^*_{Pr}) > 0.$$

The condition is true if and only if

$$\begin{aligned}\lambda\bar{Y}^2\left(C_x - C_y\rho_{yx}\right)^2 &+ \frac{\lambda^2\bar{Y}^2\left\{\theta^2C_x^2 + 8C_y^2\left(1-\rho^2_{yx}\right)\right\}^2}{64\left\{1+\lambda C_y^2\left(1-\rho^2_{yx}\right)\right\}} \\ &+ \frac{\lambda\bar{Y}^2C_y^2\left(\rho_{yr_x}-\rho_{yx}\rho_{xr_x}\right)^2\left(-8+\lambda\theta^2C_x^2\right)^2}{64\left(1-\rho^2_{xr_x}\right)\left\{1+\lambda C_y^2\left(1-\rho^2_{yx}\right)\right\}\left\{1+\lambda C_y^2\left(1-Q^2_{y.xr_x}\right)\right\}} > 0.\end{aligned} \tag{3.13}$$

(c) By comparing the proposed estimator $(\hat{\bar{Y}}^*_{Pr})$ and the Bahl and Tuteja (1991) ratio type exponential estimator $(\hat{\bar{Y}}_{BT,R})$, we have

$$MSE_{min}(\hat{\bar{Y}}^*_{Pr}) < MSE(\hat{\bar{Y}}_{BT,R})$$

$$MSE(\hat{\bar{Y}}_{BT,R}) - MSE_{min}(\hat{\bar{Y}}^*_{Pr}) > 0$$

The condition is true if and only if

$$\begin{aligned}\frac{\lambda\bar{Y}^2}{4}\left(C_x - 2C_y\rho_{yx}\right)^2 &+ \frac{\lambda^2\bar{Y}^2\left\{\theta^2C_x^2 + 8C_y^2\left(1-\rho^2_{yx}\right)\right\}^2}{64\left\{1+\lambda C_y^2\left(1-\rho^2_{yx}\right)\right\}} \\ &+ \frac{\lambda\bar{Y}^2C_y^2\left(\rho_{yr_x}-\rho_{yx}\rho_{xr_x}\right)^2\left(-8+\lambda\theta^2C_x^2\right)^2}{64\left(1-\rho^2_{xr_x}\right)\left\{1+\lambda C_y^2\left(1-\rho^2_{yx}\right)\right\}\left\{1+\lambda C_y^2\left(1-Q^2_{y.xr_x}\right)\right\}} > 0.\end{aligned} \tag{3.14}$$

(d) By comparing the proposed estimator $(\hat{\bar{Y}}^*_{Pr})$ and Rao (1991) difference type estimator $(\hat{\bar{Y}}_{R,D})$, we have

$$MSE_{min}(\hat{\bar{Y}}^*_{Pr}) < MSE_{min}(\hat{\bar{Y}}_{R,D})$$

$$MSE_{min}(\hat{\bar{Y}}_{R,D}) - MSE_{min}(\hat{\bar{Y}}^*_{Pr}) > 0$$

The condition is true if and only if

$$\frac{\lambda^2\theta^2\bar{Y}^2C_x^2\left\{\theta^2C_x^2+16C_y^2\left(1-\rho_{yx}^2\right)\right\}}{64\left\{1+\lambda C_y^2\left(1-\rho_{yx}^2\right)\right\}} + \frac{\lambda\bar{Y}^2C_y^2\left(\rho_{yr_x}-\rho_{yx}\rho_{xr_x}\right)^2\left(-8+\lambda\theta^2C_x^2\right)^2}{64\left(1-\rho_{xr_x}^2\right)\left\{1+\lambda C_y^2\left(1-\rho_{yx}^2\right)\right\}\left\{1+\lambda C_y^2\left(1-Q_{y.xr_x}^2\right)\right\}} > 0. \tag{3.15}$$

(e) By comparing the proposed estimator ($\hat{\hat{Y}}_{Pr}^*$) and the ratio type exponential estimator ($\hat{\hat{Y}}_{GK,G}$), we have

$$MSE_{min}(\hat{\hat{Y}}_{Pr}^*) < MSE_{min}(\hat{\hat{Y}}_{GK,G})$$

$$MSE_{min}(\hat{\hat{Y}}_{GK,G}) - MSE_{min}(\hat{\hat{Y}}_{Pr}^*) > 0$$

The condition is true if and only if

$$\frac{\lambda\bar{Y}^2C_y^2\left(\rho_{yr_x}-\rho_{yx}\rho_{xr_x}\right)^2\left(-8+\lambda\theta^2C_x^2\right)^2}{64\left(1-\rho_{xr_x}^2\right)\left\{1+\lambda C_y^2\left(1-\rho_{yx}^2\right)\right\}\left\{1+\lambda C_y^2\left(1-Q_{y.xr_x}^2\right)\right\}} > 0 \tag{3.16}$$

It's important to note that all the conditions derived above, from (a) to (e), consistently hold true. Consequently, the suggested estimator consistently outperforms and exhibits greater efficiency than all the existing estimators.

3.5 Numerical comparison

In this section, we analyze various population datasets to conduct numerical comparisons between the existing estimators and the proposed estimators. We assess the performance of these estimators by examining the percentage of relative efficiencies (PREs). Below, we provide a description of the population along with key data statistics.

Dataset 1. (source: Montgomery et al. (2021) , p.159)
y: Product Viscosity
x: Temperature
The summary statistics of the population are given as

$$N=17,\ n=10,\ \bar{Y}=22.55,\ \bar{X}=4.382353,\ \bar{R}_x=9,$$
$$\rho_{yx}=0.6978707\ ,\ \rho_{yr_x}=0.5473379\ ,\ \rho_{xr_x}=0.9743457,$$
$$C_y=0.244811\ ,\ C_x=0.412654\ ,\ C_r=0.557289\ ,\ \beta_2(x)=2.344069$$

Dataset 2. (source: Montgomery et al. (2021), p.565)
y: PITCH (Results of the pitch carbon analysis test)
x: SOAKTIME (Duration of the carburizing cycle)
The summary statistics of the population are given as

$$N=32,\ n=10,\ \bar{Y}=0.02628125\ ,\ \bar{X}=2.75375\ ,\ \bar{R}_x=16.5\ ,$$

$$\rho_{yx} = 0.9371165 \ , \ \rho_{yr_x} = 0.8575938 \ , \ \rho_{xr_x} = 0.6659805 \ ,$$
$$C_y = 0.4495106 \ , \ C_x = 1.304772 \ , \ C_r = 0.5574865, \ \beta_2(x) = 14.12413$$

Dataset 3. (source: Montgomery et al. (2021), p.557)

y: Sale price of the house

x: Number of garage stalls

The summary statistics of the population are given as

$$N = 24, \ n = 10, \ \bar{Y} = 34.6125, \ \bar{X} = 1.3125, \ \bar{R}_x = 12.5,$$
$$\rho_{yx} = 0.4588588, \ \rho_{yr_x} = 0.6484538, \ \rho_{xr_x} = 0.2514236,$$
$$C_y = 0.1734573, \ C_x = 0.4606148, \ C_r = 0.5263162, \ \beta_2(x) = 2.659623$$

Dataset 4. (source:Montgomery et al. (2021), p.169)

y: Measure of the whiteness of rayon

x: Amount of chlorine bleach

The summary statistics of the population are given as

$$N = 26, \ n = 10, \ \bar{Y} = 92.54615, \ \bar{X} = 45, \ \bar{R}_x = 13.5,$$
$$\rho_{yx} = 0.4437157, \ \rho_{yr_x} = -0.0551814, \ \rho_{xr_x} = 0.4271411,$$
$$C_y = 0.2673327, \ C_x = 0.2177324, \ C_r = 0.5416026, \ \beta_2(x) = 2.166667$$

Table 3.2: PREs of estimators for various choices of a and b.

Estimator		Population 1		Population 2	
		$\hat{\bar{Y}}_{Haq}$	$\hat{\bar{Y}}^{*}_{Pr}$	$\hat{\bar{Y}}_{Haq}$	$\hat{\bar{Y}}^{*}_{Pr}$
$\hat{\bar{Y}}^{(1)}_{Haq}$	$\hat{\bar{Y}}^{*(1)}_{Pr}$	354.2649	604.0788	4050.129	4246.627
$\hat{\bar{Y}}^{(2)}_{Haq}$	$\hat{\bar{Y}}^{*(2)}_{Pr}$	354.1885	603.9333	4004.295	4196.906
$\hat{\bar{Y}}^{(3)}_{Haq}$	$\hat{\bar{Y}}^{*(3)}_{Pr}$	354.2826	604.1134	4172.921	4381.058
$\hat{\bar{Y}}^{(4)}_{Haq}$	$\hat{\bar{Y}}^{*(4)}_{Pr}$	354.1485	603.8609	4004.872	4197.518
$\hat{\bar{Y}}^{(5)}_{Haq}$	$\hat{\bar{Y}}^{*(5)}_{Pr}$	354.2472	604.0444	4068.562	4266.751
$\hat{\bar{Y}}^{(6)}_{Haq}$	$\hat{\bar{Y}}^{*(6)}_{Pr}$	354.2056	603.9652	4084.495	4284.166
$\hat{\bar{Y}}^{(7)}_{Haq}$	$\hat{\bar{Y}}^{*(7)}_{Pr}$	354.2534	604.0564	4046.816	4243.015
$\hat{\bar{Y}}^{(8)}_{Haq}$	$\hat{\bar{Y}}^{*(8)}_{Pr}$	354.2731	604.0948	4178.982	4387.711
$\hat{\bar{Y}}^{(9)}_{Haq}$	$\hat{\bar{Y}}^{*(9)}_{Pr}$	354.1705	603.9002	4004.193	4196.798
$\hat{\bar{Y}}^{(10)}_{Haq}$	$\hat{\bar{Y}}^{*(10)}_{Pr}$	354.1187	603.809	4003.433	4195.997

We conduct a numerical comparison between the proposed estimator and existing estimators using the Percentage Relative Efficiencies (PREs), as defined by

$$PRE(\cdot) = (Var(\bar{y})/MSE(\cdot)) \times 100$$

Table 3.3: PREs of estimators for various choices of a and b.

Estimator		Population 3		Population 4	
		$\hat{\hat{Y}}_{Haq}$	$\hat{\hat{Y}}^*_{Pr}$	$\hat{\hat{Y}}_{Haq}$	$\hat{\hat{Y}}^*_{Pr}$
$\hat{\hat{Y}}^{(1)}_{Haq}$	$\hat{\hat{Y}}^{*(1)}_{Pr}$	198.2112	205.9961	130.1007	137.4516
$\hat{\hat{Y}}^{(2)}_{Haq}$	$\hat{\hat{Y}}^{*(2)}_{Pr}$	198.1343	205.9158	130.0989	137.4496
$\hat{\hat{Y}}^{(3)}_{Haq}$	$\hat{\hat{Y}}^{*(3)}_{Pr}$	198.2572	206.0444	130.1009	137.4517
$\hat{\hat{Y}}^{(4)}_{Haq}$	$\hat{\hat{Y}}^{*(4)}_{Pr}$	198.1225	205.9035	130.0934	137.4438
$\hat{\hat{Y}}^{(5)}_{Haq}$	$\hat{\hat{Y}}^{*(5)}_{Pr}$	198.2114	205.9964	130.1005	137.4514
$\hat{\hat{Y}}^{(6)}_{Haq}$	$\hat{\hat{Y}}^{*(6)}_{Pr}$	198.1702	205.9533	130.099	137.4498
$\hat{\hat{Y}}^{(7)}_{Haq}$	$\hat{\hat{Y}}^{*(7)}_{Pr}$	198.1698	205.9529	130.1005	137.4513
$\hat{\hat{Y}}^{(8)}_{Haq}$	$\hat{\hat{Y}}^{*(8)}_{Pr}$	198.2574	206.0445	130.1008	137.4516
$\hat{\hat{Y}}^{(9)}_{Haq}$	$\hat{\hat{Y}}^{*(9)}_{Pr}$	198.1224	205.9035	130.0967	137.4473
$\hat{\hat{Y}}^{(10)}_{Haq}$	$\hat{\hat{Y}}^{*(10)}_{Pr}$	198.1174	205.8983	130.0782	137.4278

Based on the datasets above, the proposed and existing estimator's PREs are calculated with respect to $\bar{y}$ and reported in Table 2 and Table 3. Table 2 and Table 3 clearly demonstrate that the proposed estimator consistently outperforms all the other existing estimators considered in this analysis. In datasets from 1 to 4, the proposed estimator PREs are greater than those with their counterparts.

It is to be noted that the more we have tied values in our data set, the result will be more precise.

3.6 Conclusion and discussion

In this chapter, we have proposed a technique that deals with tied ranks. The proposed estimator leverages both a ranked auxiliary variable and an auxiliary variable to incorporate auxiliary information. This approach harnesses auxiliary data to enhance efficiency and reduce variance. We have obtained the expressions for both bias and Mean Squared Error (MSE) using a first-order approximation. The efficiency of the suggested estimator has been established theoretically and the theoretical findings have been validated using population datasets. The suggested estimator outperforms the existing estimators according to the numerical results and is suggested to use for efficiently estimating finite population means when tied values are present in population datasets.

Chapter 4

Estimation of Finite Population Mean Using Tied Ranks Under Stratified Random Sampling

In contrast to simple random sampling, it is recommended to employ stratified random sampling when a population being studied is heterogeneous. In such a situation, a particular sampling design is used to select a sample at random from each stratum after dividing the population into strata. To fulfill the goal of an estimate, It is crucial to establish the count of strata and the corresponding sample sizes in each stratum. A Simple Random Sample (SRS) may not yield a representative sample of a diverse population. SRS might miss the variety that can be uncovered through stratified random sampling.

A population should be stratified based on some information. According to Cingi (1994), if there is no prior knowledge of the stratification scheme, around 10 strata are the ideal number for large population size. The three most common methods for determining the ideal stratum sample sizes are neyman allocation, proportional allocation, and equal allocation. When stratifying the data, each unit of the population should be assigned to just one of the strata, and the strata should contain all of the population's data. This stratification is also done to keep variance between strata to a maximum and variance within each stratum to a minimum. Getting the most precise number of units inside each stratum is the aim of population stratification. A few benefits of stratified random sampling are improved estimation, cheaper surveys, and simpler administration.

In survey sampling, utilizing auxiliary information effectively can improve the accuracy of an estimate when it is presented in the form of auxiliary data. Classical ratio, product, and regression estimators are examples of how auxiliary information is used. One method for increasing the precision of estimates is stratified random sampling. It is a versatile and powerful strategy that is commonly used in practice.

Using stratified random sampling and utilizing information from the auxiliary variable is likely to enhance an estimator's efficiency.

4.1 Definition and notations under stratified random sampling

Let a population comprising of N units be divided into L mutually exclusive strata of sizes N_h $(h = 1, 2, 3, ..., L)$, such that $\sum_{h=1}^{L} N_h = N$. Let a random sample of size n_h is drawn from each stratum independently by simple random sampling without replacement such that $\sum_{h=1}^{L} n_h = n$, where n is the total number of units in the sample. Let y_{hi} and x_{hi} be the sample values of the study variable and auxiliary variable respectively, corresponding to the population values of the study variable Y_{hi} and auxiliary variable X_{hi} for the i^{th} unit $(i = 1, 2, 3, ..., N_h)$ in the h^{th} stratum (h=1,2,3,..., L), respectively.

Let $\bar{y}_h = \sum_{h=1}^{n_h} y_{hi}/n_h$ and $\bar{x}_h = \sum_{h=1}^{n_h} x_{hi}/n_h$ be the sample means corresponding to the population means $\bar{Y}_h = \sum_{h=1}^{N_h} Y_{hi}/N_h$ and $\bar{X}_h = \sum_{h=1}^{N_h} X_{hi}/N_h$ in the h^{th} (h=1,2,3,...,L) stratum, respectively.

Let $s^2_{yh} = 1/(n_h - 1)\sum_{h=1}^{n_h}(y_{hi} - \bar{y}_h)^2$ and $s^2_{xh} = 1/(n_h - 1)\sum_{h=1}^{n_h}(x_{hi} - \bar{x}_h)^2$ be the sample variances corresponding to the population variances $S^2_{yh} = 1/(N_h - 1)\sum_{h=1}^{N_h}(Y_{hi} - \bar{Y}_h)^2$ and $S^2_{xh} = 1/(N_h - 1)\sum_{h=1}^{N_h}(X_{hi} - \bar{X}_h)^2$ in the h^{th} stratum, respectively.

Let $\hat{\rho}_{yxh}$ be the sample correlation coefficient corresponding to the population correlation coefficient in the h^{th} stratum. Let $\bar{y}_{st} = \sum_{h=1}^{L} W_h\bar{y}_h$ and $\bar{x}_{st} = \sum_{h=1}^{L} W_h\bar{x}_h$ be the unbiased sample means corresponding to the population means $\bar{Y} = \sum_{i=1}^{N} Y_i/N$ and $\bar{X} = \sum_{i=1}^{N} X_i/N$, respectively.

Let $S_{yx} = S_y \cdot S_x \cdot \rho_{yx}$ be the population covariance of Y and X. Let $b_h = S_{hyx}/S^2_{hx}$ be the population regression coefficient for the h^{th} stratum and $W_h = N_h/N$ be the stratum weight in the h^{th} stratum, $f_h = n_h/N_h$ be the sample fraction in the h^{th} stratum. Moreover, let $R = \bar{Y}/\bar{X}$ be the population ratio and $R_h = \bar{Y}_h/\bar{X}_h$ be the population ratio in the h^{th} stratum.

4.2 Proposed estimators

Haq et al. (2017) developed an efficient class of estimators for the population mean estimation using additional information from the auxiliary variable known as the ranked auxiliary variable with equally spaced values. These estimators were only designed to work with a simple random sampling design. A new idea for investigating more optimal estimators utilizing dual use of auxiliary information with tied ranks under stratified random sampling is presented. The challenge is satisfactorily met, and using a stratified random sampling method, new optimal estimators for finite population means are constructed. In the following section, we present an enhanced version of our suggested estimator (3.1) to efficiently handle the population's underlying heterogeneous structure.

Let $S^2_{hr_x}$, $\bar{R}_{hx}$ and $\bar{r}_{hx}$ be the population variance, population mean, and sample mean of R_{hx}, respectively where $S^2_{hr_x} = \sum_{h=1}^{N_h}(R_{x,ih} - \bar{R}_{xh})^2/(N_h - 1)$, $\bar{R}_{xh} = \sum_{h=1}^{N_h} R_{x,ih}/N_h = (N_h + 1)/2$ and $\bar{r}_{xh} = \sum_{h=1}^{n_h} r_{x,hi}/n_h$. Let the correlation coefficient between the Z and R_x

be the $\rho_{z_h r_{xh}} = S_{z_h r_{xh}}/S_{zh}S_{hr_x}$ where $S_{z_h r_{xh}} = \sum_{h=1}^{N_h}(Z_{ih} - \bar{Z}_h)(r_{x,ih} - \bar{R}_{xh})/(N_h - 1)$ is the population covariance between Z and R_x for $Z= Y,X$. Let the coefficient of variation of R_x be $C_{rh} = S_{hr_x}/\bar{R}_{xh}$.

In stratified random sampling, there are two methods for calculating estimates. These are the separate and combined estimators. The combined estimator is as follows.

$$\hat{\bar{Y}}_C^* = [\omega_1\bar{y}_{st} + \omega_2(\bar{X} - \bar{x}_{st}) + \omega_3(\bar{R}_x - \bar{r}_{x_{st}})] \exp\left(\frac{a(\bar{X} - \bar{x}_{st})}{a(\bar{X} + \bar{x}_{st}) + 2b}\right) \tag{4.1}$$

where ω_1, ω_2, and ω_3 are the unknown constants minimizing the proposed estimator's MSE.

Table 4.1: Some members of the existing estimator along with separate and combined estimator.

a	b	$\hat{\bar{Y}}_{Haq}^*$	$\hat{\bar{Y}}_S^*$	$\hat{\bar{Y}}_C^*$
1	C_X	$\hat{\bar{Y}}_{Haq}^{(1)}$	$\hat{\bar{Y}}_S^{*(1)}$	$\hat{\bar{Y}}_C^{*(1)}$
1	$\beta_2(X)$	$\hat{\bar{Y}}_{Haq}^{(2)}$	$\hat{\bar{Y}}_S^{*(2)}$	$\hat{\bar{Y}}_C^{*(2)}$
$\beta_2(X)$	C_X	$\hat{\bar{Y}}_{Haq}^{(3)}$	$\hat{\bar{Y}}_S^{*(3)}$	$\hat{\bar{Y}}_C^{*(3)}$
C_X	$\beta_2(X)$	$\hat{\bar{Y}}_{Haq}^{(4)}$	$\hat{\bar{Y}}_S^{*(4)}$	$\hat{\bar{Y}}_C^{*(4)}$
1	ρ_{YX}	$\hat{\bar{Y}}_{Haq}^{(5)}$	$\hat{\bar{Y}}_S^{*(5)}$	$\hat{\bar{Y}}_C^{*(5)}$
C_X	ρ_{YX}	$\hat{\bar{Y}}_{Haq}^{(6)}$	$\hat{\bar{Y}}_S^{*(6)}$	$\hat{\bar{Y}}_C^{*(6)}$
ρ_{YX}	C_X	$\hat{\bar{Y}}_{Haq}^{(7)}$	$\hat{\bar{Y}}_S^{*(7)}$	$\hat{\bar{Y}}_C^{*(7)}$
$\beta_2(X)$	ρ_{YX}	$\hat{\bar{Y}}_{Haq}^{(8)}$	$\hat{\bar{Y}}_S^{*(8)}$	$\hat{\bar{Y}}_C^{*(8)}$
ρ_{YX}	$\beta_2(X)$	$\hat{\bar{Y}}_{Haq}^{(9)}$	$\hat{\bar{Y}}_S^{*(9)}$	$\hat{\bar{Y}}_C^{*(9)}$
1	$N\bar{X}$	$\hat{\bar{Y}}_{Haq}^{(10)}$	$\hat{\bar{Y}}_S^{*(10)}$	$\hat{\bar{Y}}_C^{*(10)}$

Table 4.1 lists specific members of the estimators of existing, separate and combined. Using the values of a and b, we can develop the sub-cases of the $\hat{\bar{Y}}_{Haq}^*$, $\hat{\bar{Y}}_S^*$ and $\hat{\bar{Y}}_C^*$ under stratified random sampling.

The separate estimator is given as

$$\hat{\bar{Y}}_S^* = \sum_{h=1}^{L} W_h^2\left[[\omega_1\bar{y}_h + \omega_2(\bar{X} - \bar{x}_h) + \omega_3(\bar{R}_x - \bar{r}_{x_h})] \exp\left(\frac{a(\bar{X} - \bar{x}_h)}{a(\bar{X} + \bar{x}_h) + 2b}\right)\right] \tag{4.2}$$

The separate regression estimator is the sub-case of the separate estimator i.e if we have a= 0, $\omega_1 = 1$, and $\omega_3 = 0$ then the separate estimator reduces to separate regression estimator given as

$$\hat{\bar{Y}}_{lrS}^* = \sum_{h=1}^{L} W_h^2\Big(\bar{y}_h + \omega_2(\bar{X} - \bar{x}_h)\Big) \tag{4.3}$$

4.3 Statistical properties of proposed estimators

In many practices and approaches, the ratio and classical linear regression estimators are widely used for the estimation of unknown population parameters provided there exists a sufficient correlation between the response variable and the auxiliary variable.

The difference between an estimator's expected value and the true value of the parameter being estimated is the bias of the estimator. i.e. $Bias(\hat{\bar{Y}}) = E(\hat{\bar{Y}} - \bar{Y})$ and MSE is defined as the deviation of estimator values from the true parameter value i.e. $MSE(\hat{\bar{Y}}) = E(\hat{\bar{Y}} - \bar{Y})^2$. It is known that when the first degree of approximation is used in obtaining the Mean Square Error (MSE) of the ratio estimate, it is equal to the variance. To obtain the properties, i.e., the bias and MSE of the proposed estimator, the following relative error terms based on the stratified random sample are defined:

$e_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$, $e_1 = \frac{\bar{x}-\bar{X}}{\bar{X}}$ and $e_2 = \frac{\bar{r}_x-\bar{R}_x}{\bar{R}_x}$, such that $E(e_i) = 0$ for $i = 0, 1, 2$. i-e

$$E(e_0) = E(e_1) = E(e_2) = 0. \tag{4.4}$$

The expected values of the relative error terms can be shown to be given by

$$\begin{aligned}
E(e_0^2) &= E\left(\frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}}\right)^2 = \frac{1}{\bar{Y}^2}E(\bar{y}_{st} - \bar{Y})^2 = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{Y_h}^2}{\bar{Y}^2} = \sum_{h=1}^{L} W_h^2 \lambda_h C_{Y_h}^2 \\
&= V_{200}, \\
E(e_1^2) &= E\left(\frac{\bar{x}_{st} - \bar{X}}{\bar{X}}\right)^2 = \frac{1}{\bar{X}^2}E(\bar{x}_{st} - \bar{X})^2 = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{X_h}^2}{\bar{X}^2} = \sum_{h=1}^{L} W_h^2 \lambda_h C_{X_h}^2 \\
&= V_{020}, \\
E(e_2^2) &= E\left(\frac{\bar{r}_{x_{st}} - \bar{R}_x}{\bar{R}_x}\right)^2 = \frac{1}{\bar{R}_x^2}E(\bar{r}_{x_{st}} - \bar{R}_x)^2 = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{R_{xh}}^2}{\bar{R}_x^2} = \sum_{h=1}^{L} W_h^2 \lambda_h C_{R_{xh}}^2 \\
&= V_{002}, \\
E(e_0 e_1) &= E\left(\frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{x}_{st} - \bar{X}}{\bar{X}}\right) = \frac{1}{\bar{X}\bar{Y}}E(\bar{y}_{st} - \bar{Y})(\bar{x}_{st} - \bar{X}) = \frac{1}{\bar{Y}\bar{X}}Cov(\bar{y}_{st}, \bar{x}_{st}) \\
&= \frac{1}{\bar{Y}\bar{X}}\sum_{h=1}^{L} W_h^2 \lambda_h S_{Y_h X_h} = \sum_{h=1}^{L} W_h^2 \lambda_h \rho_{Y_h X_h} C_{Y_h} C_{X_h} \\
&= V_{110}, \\
E(e_0 e_2) &= E\left(\frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{r}_{x_{st}} - \bar{R}_x}{\bar{R}_x}\right) = \frac{1}{\bar{R}_x\bar{Y}}E(\bar{y}_{st} - \bar{Y})(\bar{r}_{x_{st}} - \bar{R}_x) = \frac{1}{\bar{Y}\bar{R}_x}Cov(\bar{y}_{st}, \bar{r}_{x_{st}}) \\
&= \frac{1}{\bar{Y}\bar{R}_x}\sum_{h=1}^{L} W_h^2 \lambda_h S_{Y_h R_{xh}} = \sum_{h=1}^{L} W_h^2 \lambda_h \rho_{Y_h R_h} C_{Y_h} C_{R_{xh}} \\
&= V_{101}, \\
E(e_1 e_2) &= E\left(\frac{\bar{x}_{st} - \bar{X}}{\bar{X}}\right)\left(\frac{\bar{r}_{x_{st}} - \bar{R}_x}{\bar{R}_x}\right) = \frac{1}{\bar{R}_x\bar{X}}E(\bar{x}_{st} - \bar{X})(\bar{r}_{x_{st}} - \bar{R}_x) = \frac{1}{\bar{X}\bar{R}_x}Cov(\bar{x}_{st}, \bar{r}_{x_{st}}) \\
&= \frac{1}{\bar{X}\bar{R}_x}\sum_{h=1}^{L} W_h^2 \lambda_h S_{X_h R_{xh}} = \sum_{h=1}^{L} W_h^2 \lambda_h \rho_{X_h R_h} C_{X_h} C_{R_{xh}} \\
&= V_{011},
\end{aligned}$$

where the notations V_{iii} have been introduced to save space in the subsequent derivations, for $i = 0, 1, 2$.

Utilizing the error terms, the MSE of the separate estimator is given as

$$MSE_{min}\left(\hat{\hat{Y}}_S^*\right) \cong \sum_{h=1}^{L} W_h^2 \left[\frac{\lambda_h \bar{Y}^2 \Big[64 C_{Y_h}^2 \Big(1 - Q_{Y_h.X_h R_{x_h}}^2\Big) - \lambda_h \theta^4 C_{X_h}^4 - 16 \lambda_h \theta^2 C_{X_h}^2 C_{Y_h}^2 \Big(1 - Q_{Y_h.X_h R_{x_h}}^2\Big)\Big]}{64\Big[1 + \lambda_h C_{Y_h}^2 \Big(1 - Q_{Y_H.X_h R_{x_h}}^2\Big)\Big]} \right] \tag{4.5}$$

where

$$Q_{Y_h.X_h R_{xh}}^2 = \frac{\rho_{Y_h X_h}^2 + \rho_{Y_h R_{xh}}^2 - 2\rho_{Y_h X_h} \rho_{Y_h R_{xh}} \rho_{X_h R_{xh}}}{1 - \rho_{X_h R_{xh}}^2}$$

and the optimum values of separate estimator is defined as

$$\omega_{1(opt)} = \sum_{h=1}^{L} W_h^2 \left[\frac{8 - \lambda_h \theta^2 C_{x_h}^2}{8\Big[1 + \lambda_h C_{y_h}^2 (1 - Q_{y.xr_{xh}}^2)\Big]} \right] \tag{4.6}$$

$$\omega_{2(opt)} = \sum_{h=1}^{L} \frac{W_h^2 \bar{Y}_h \begin{bmatrix} \lambda_h \theta^3 C_{x_h}^3 (-1 + \rho_{xr_{xh}}^2) + (-8C_{y_h} + \lambda_h \theta^2 C_{x_h}^2 C_{y_h})(\rho_{yx_h} - \rho_{xr_{xh}} \rho_{yr_{xh}}) \\ + 4\theta C_{x_h} (-1 + \rho_{xr_{xh}}^2)\Big(-1 + \lambda_h C_{y_h}^2 (1 - Q_{y.xr_{xh}}^2)\Big) \end{bmatrix}}{8 \bar{X}_h C_{x_h} (-1 + \rho_{xr_{xh}}^2)\Big[1 + \lambda_h C_{y_h}^2 (1 - Q_{y.xr_{xh}}^2)\Big]} \tag{4.7}$$

and

$$\omega_{3(opt)} = \sum_{h=1}^{L} W_h^2 \left[\frac{\bar{Y}_h (8 - \lambda_h \theta^2 C_{x_h}^2) C_{y_h} (\rho_{xr_{xh}} \rho_{yx_h} - \rho_{yr_{xh}})}{8 \bar{R}_{xh} C_{r_h} (-1 + \rho_{xr_{xh}}^2)\Big[1 + \lambda_h C_{y_h}^2 (1 - Q_{y.xr_{xh}}^2)\Big]} \right] \tag{4.8}$$

Now, to find the bias and MSE of the combined estimator, rewriting $\hat{\hat{Y}}_C^*$ in terms of relative error terms, we have

$$\hat{\hat{Y}}_C^* = \Big[\omega_1 \bar{Y}(1 + e_0) + \omega_2 (\bar{X} - \bar{X}(1 + e_1)) + \omega_3 (\bar{R}_x - \bar{R}_x (1 + e_2))\Big]$$
$$\exp\left(\frac{a(\bar{X} - \bar{X}(1 + e_1))}{a(\bar{X} + \bar{X}(1 + e_1)) + 2b} \right)$$

$$\hat{\hat{Y}}_C^* = \Big[\omega_1 \bar{Y}(1 + e_0) - \omega_2 \bar{X} e_1 - \omega_3 \bar{R}_x e_2\Big] \exp\left(\frac{-a\bar{X} e_1}{2a\bar{X} + a\bar{X} e_1 + 2b} \right) \tag{4.9}$$

Now consider the exponential term

$$
\begin{aligned}
\exp\left(\frac{-a\bar{X}e_1}{2a\bar{X}+a\bar{X}e_1+2b}\right) &= \exp\left(\frac{-a\bar{X}e_1}{(2a\bar{X}+2b)\left(1+\frac{a\bar{X}e_1}{2a\bar{X}+2b}\right)}\right)\\
&= \exp\left(\frac{-\theta e_1}{2(1+\frac{\theta e_1}{2})}\right)\\
&= \exp\left(\frac{-\theta e_1}{2}\left(\frac{1+\theta e_1}{2}\right)^{-1}\right)\\
&= \exp\left(\frac{-\theta e_1}{2}\left(1-\frac{\theta e_1}{2}+\frac{\theta^2 e_1^2}{4}-\frac{\theta^3 e_1^3}{8}+....\right)\right)\\
&\approx \exp\left(\frac{-\theta e_1}{2}+\frac{\theta^2 e_1^2}{4}\right)\\
&\approx 1-\left(\frac{\theta e_1}{2}+\frac{\theta^2 e_1^2}{4}\right)+\frac{(\frac{-\theta e_1}{2}+\frac{\theta^2 e_1^2}{4})^2}{2}\\
&\approx 1-\frac{\theta e_1}{2}+\frac{\theta^2 e_1^2}{4}+\frac{\theta^2 e_1^2}{8}\\
&\approx 1-\frac{\theta e_1}{2}+\frac{3\theta^2 e_1^2}{8},
\end{aligned}
$$

the last expression is the relative error terms taken up to the first order of approximation, and $\theta = \frac{-a\bar{X}}{a\bar{X}+b}$.

Substituting the exponential term in (4.9), we get

$$
\hat{\hat{Y}}_C^* = \left[\omega_1\bar{Y}(1+e_0)-\omega_2\bar{X}e_1-\omega_3\bar{R}_x e_2\right]\left[1-\frac{\theta e_1}{2}+\frac{3\theta^2 e_1^2}{8}\right] \tag{4.10}
$$

Expanding (4.10) and keeping the terms up to the order two in e_is, we can write as

$$
\begin{aligned}
(\hat{\hat{Y}}_C^*-\bar{Y}) &= -\bar{Y}+\bar{Y}\omega_1+\bar{Y}\omega_1 e_0-\bar{X}\omega_2 e_1-\bar{R}_x\omega_3 e_2-\frac{\bar{Y}\omega_1\theta e_1}{2}-\frac{\bar{Y}\omega_1\theta e_0 e_1}{2}\\
&+\frac{\bar{X}\omega_2\theta e_1^2}{2}+\frac{\bar{R}_x\omega_3\theta e_1 e_2}{2}+\frac{3\bar{Y}\omega_1\theta^2 e_1^2}{8}.
\end{aligned} \tag{4.11}
$$

Now taking expectations on both sides, we get

$$
\begin{aligned}
E(\hat{\hat{Y}}_C^*-\bar{Y}) &= Bias(\hat{\hat{Y}}_C^*)\\
&\cong -\bar{Y}+\bar{Y}\omega_1+\bar{Y}\omega_1 E(e_0)-\bar{X}\omega_2 E(e_1)-\bar{R}_x\omega_3 E(e_2)-\frac{\bar{Y}\omega_1\theta E(e_1)}{2}\\
&-\frac{\bar{Y}\omega_1\theta E(e_0e_1)}{2}+\frac{\bar{X}\omega_2\theta E(e_1)^2}{2}+\frac{\bar{R}_x\omega_3\theta E(e_1e_2)}{2}+\frac{3\bar{Y}\omega_1\theta^2 E(e_1)^2}{8}
\end{aligned}
$$

Using the results of error terms computed earlier under a stratified random sample, we get the bias of the combined estimator given as

$$
Bias(\hat{\hat{Y}}_C^*) \cong \omega_1\bar{Y}-\bar{Y}+\frac{3}{8}\theta^2\bar{Y}\omega_1 V_{020}-\frac{1}{2}\omega_1\bar{Y}\theta V_{110}+\frac{1}{2}\omega_2\bar{X}\theta V_{020}+\frac{1}{2}\omega_3\bar{R}_x\theta V_{011}
$$

Taking square of (4.11) to find out MSE.

$$(\hat{\hat{Y}}^{*}_{Pr} - \bar{Y})^2 \cong \Big(-\bar{Y} + \bar{Y}\omega_1 + \bar{Y}\omega_1 e_0 - \bar{X}\omega_2 e_1 - \bar{R}_x\omega_3 e_2 - \frac{\bar{Y}\omega_1\theta e_1}{2} - \frac{\bar{Y}\omega_1\theta e_0 e_1}{2} + \frac{\bar{X}\omega_2\theta e_1^2}{2} + \frac{\bar{R}_x\omega_3\theta e_1 e_2}{2} + \frac{3\bar{Y}\omega_1\theta^2 e_1^2}{8}\Big)^2$$

Simplifying and Keeping the terms up to order two in e_is, we get

$$\begin{aligned}(\hat{\hat{Y}}^{*}_{Pr} - \bar{Y})^2 \cong{}& \bar{Y}^2 + \bar{Y}^2\omega_1^2 + \bar{Y}^2\omega_1^2 e_0^2 + \bar{X}^2\omega_2^2 e_1^2 + \bar{R}_x^2\omega_3^2 e_2^2 - \frac{3\bar{Y}^2\omega_1\theta^2 e_1^2}{4} - 2\bar{Y}^2\omega_1 \\ &- 2\omega_1\omega_2\bar{X}\bar{Y}e_0e_1 - 2\omega_1\omega_3\bar{Y}\bar{R}_x e_0 e_2 - 2\omega_1^2\bar{Y}^2\theta e_0 e_1 + 2\omega_2\omega_3\bar{X}\bar{R}_x e_1 e_2 \\ &+ 2\omega_1\omega_2\bar{X}\bar{Y}e_1^2\theta + \omega_1\bar{Y}^2\theta e_0 e_1 - \bar{Y}\bar{X}\omega_2\theta e_1^2 - \omega_3\bar{Y}\bar{R}_x\theta e_1 e_2 \\ &+ \omega_1^2\bar{Y}^2\theta^2 e_1^2 + 2\omega_1\omega_3\bar{Y}\bar{R}_x\theta e_1 e_2.\end{aligned}$$

Taking expectations on both sides and simplifying algebraically, we get

$$\begin{aligned}E(\hat{\hat{Y}}^{*}_{C} - \bar{Y})^2 ={}& MSE(\hat{\hat{Y}}^{*}_{C}) \\ \cong{}& -2\omega_1\bar{Y}\omega_2\bar{X}V_{110} - 2\omega_1\bar{Y}\omega_3\bar{R}_x V_{101} + 2\omega_2\bar{X}\omega_3\bar{R}_x V_{011} + \omega_1^2\bar{Y}^2\theta^2 V_{020} \\ &+ \omega_3\bar{R}_x^2 V_{002} + \omega_1^2\bar{Y}^2 V_{200} + \omega_2^2\bar{X}^2 V_{020} + \omega_1\bar{Y}^2\theta V_{110} - \frac{3}{4}\theta^2\bar{Y}^2\omega_1 V_{020} \\ &- 2\omega_1^2\bar{Y}^2\theta V_{110} + \omega_1^2\bar{Y}^2 - 2\omega_1\bar{Y}^2 + \bar{Y}^2 - \bar{Y}\omega_3\bar{R}_x\theta V_{011} - \bar{Y}\omega_2\bar{X}\theta V_{020} \\ &+ 2\omega_1\bar{Y}\omega_2\bar{X}\theta V_{020} + 2\omega_1\bar{Y}\omega_3\bar{R}_x\theta V_{011}.\end{aligned} \tag{4.12}$$

The bias and MSE were obtained up to the first order of approximation. The optimum values of ω_1, ω_2 and ω_3 were obtained by minimizing 4.12 and are given as

$$\omega_{1(opt)} = \frac{-1}{8T}\left[(J_1 - V_{011}^2)(V_{020}\theta^2 - 8)\right] \tag{4.13}$$

$$\begin{aligned}\omega_{2(opt)} ={}& \frac{\bar{Y}}{8\bar{X}T}\Big(J_1V_{020}\theta^3 - J_7\theta^3 - J_8\theta^2 + J_9\theta^2 + 4J_2\theta - 4J_3\theta - 4\theta J_4 + 8\theta J_5 - 4\theta J_6 V_{101} \\ &-4J_1\theta + 4\theta V_{011}^2 + 8V_{002}V_{110} - 8V_{011}V_{101}\Big),\end{aligned} \tag{4.14}$$

and

$$\omega_{3(opt)} = \frac{\bar{Y}}{8\bar{R}_x}\left(\frac{(V_{011}V_{110} - J_6)(V_{020}\theta^2 - 8)}{T}\right). \tag{4.15}$$

Now substituting the optimum values of ω_1, ω_2 and ω_3 in equation (4.12) and after doing some simplifications, we achieve the minimum MSE of the proposed combined estimator $\hat{\hat{Y}}^{*}_{C}$

given as

$$MSE_{min}(\hat{\bar{Y}}_C^*) \cong \frac{\bar{Y}^2}{64T}\Big(-J_1V_{020}^2\theta^4 + J_7V_{020}\theta^4 - 16J_2\theta^2V_{020} + 16J_8\theta^2V_{110} + 16J_{10}\theta^2$$
$$-32\theta^2J_9V_{110} + 16J_6^2\theta^2 + 64J_2 - 64J_3 - 64J_4 + 128J_5 - 64J_6V_{101}\Big) \quad (4.16)$$

where

$$J_1 = V_{002}V_{020}\ ,\ J_2 = V_{002}V_{020}V_{200},\ J_3 = V_{002}V_{110}^2,\ J_4 = V_{011}^2V_{200}$$
$$J_5 = V_{011}V_{101}V_{110},\ J_6 = V_{020}V_{101},\ J_7 = V_{011}^2V_{020},\ J_8 = V_{002}V_{020}V_{110},$$
$$J_9 = V_{011}V_{020}V_{101},\ J_{10} = V_{011}^2V_{020}V_{200}$$

and

$$T = J_2 - J_3 - J_4 + 2J_5 - J_6V_{101} + J_1 - V_{011}^2$$

4.4 Theoretical comparison of separate estimator

Under this section, we have compared the separate estimator with some of the existing estimator under stratified random sampling. The obtained expressions are given in the following section.

(a) By comparing the proposed separate estimator and the conventional unbiased estimator, we have

$$MSE_{min}(\hat{\bar{Y}}_S^*) < MSE(\hat{\bar{Y}}_R)$$

$$MSE(\hat{\bar{Y}}_R) - MSE_{min}(\hat{\bar{Y}}_S^*) > 0$$

The condition is true if and only if

$$\sum_{h=1}^{L} W_h^2\left[\frac{\lambda_h\bar{Y}_h^{\,2}\Big[\lambda_h\theta^4C_{x_h}^4 + 16C_{y_h}^2\Big(4Q_{y.xr_xh}^2 + \lambda_h\theta^2C_{x_h}^2(1-Q_{y.xr_xh}^2)\Big) + 64\lambda_hC_{y_h}^4(1-Q_{y.xr_xh}^2)\Big]}{64\Big[1+\lambda_hC_{y_h}^2(1-Q_{y.xr_xh}^2)\Big]}\right] > 0 \quad (4.17)$$

(b) By comparing the proposed separate estimator $(\hat{\bar{Y}}_S^*)$ and the usual ratio estimator $(\hat{\bar{Y}}_R)$, we have

$$MSE_{min}(\hat{\bar{Y}}_S^*) < MSE(\hat{\bar{Y}}_R).$$

$$MSE(\hat{\bar{Y}}_R) - MSE_{min}(\hat{\bar{Y}}_S^*) > 0.$$

The condition is true if and only if

$$\sum_{h=1}^{L} W_h^2 \Bigg[\lambda_h \bar{Y_h}^2 \left(C_{x_h} - C_{y_h}\rho_{yx_h}\right)^2 + \frac{\lambda_h^2 \bar{Y_h}^2 \left\{\theta^2 C_{x_h}^2 + 8C_{y_h}^2 \left(1-\rho_{yx_h}^2\right)\right\}^2}{64\left\{1+\lambda_h C_{y_h}^2\left(1-\rho_{yx_h}^2\right)\right\}} + \frac{\lambda_h \bar{Y_h}^2 C_{y_h}^2 \left(\rho_{yr_{xh}} - \rho_{yx_h}\rho_{xr_{xh}}\right)^2 \left(-8+\lambda_h\theta^2 C_{x_h}^2\right)^2}{64\left(1-\rho_{xr_{xh}}^2\right)\left\{1+\lambda_h C_{y_h}^2\left(1-\rho_{yx_h}^2\right)\right\}\left\{1+\lambda_h C_{y_h}^2\left(1-Q_{y.xr_{xh}}^2\right)\right\}} \Bigg] > 0. \tag{4.18}$$

(C) By comparing the proposed separate estimator ($\hat{\hat{Y}}_S^*$) and the Bahl and Tuteja (1991) ratio type exponential estimator ($\hat{\hat{Y}}_{BT,R}$), we have

$$MSE_{min}(\hat{\hat{Y}}_S^*) < MSE(\hat{\hat{Y}}_{BT,R})$$

$$MSE(\hat{\hat{Y}}_{BT,R}) - MSE_{min}(\hat{\hat{Y}}_S^*) > 0$$

The condition is true if and only if

$$\sum_{h=1}^{L} W_h^2 \Bigg[\frac{\lambda_h \bar{Y_h}^2}{4} \left(C_{x_h} - 2C_{y_h}\rho_{yx_h}\right)^2 + \frac{\lambda_h^2 \bar{Y_h}^2 \left\{\theta^2 C_{x_h}^2 + 8C_{y_h}^2 \left(1-\rho_{yx_h}^2\right)\right\}^2}{64\left\{1+\lambda_h C_{y_h}^2\left(1-\rho_{yx_h}^2\right)\right\}} + \frac{\lambda_h \bar{Y_h}^2 C_{y_h}^2 \left(\rho_{yr_{xh}} - \rho_{yx_h}\rho_{xr_{xh}}\right)^2 \left(-8+\lambda_h\theta^2 C_{x_h}^2\right)^2}{64\left(1-\rho_{xr_{xh}}^2\right)\left\{1+\lambda_h C_{y_h}^2\left(1-\rho_{yx_h}^2\right)\right\}\left\{1+\lambda_h C_{y_h}^2\left(1-Q_{y.xr_{xh}}^2\right)\right\}} \Bigg] > 0. \tag{4.19}$$

(d) By comparing the proposed separate estimator ($\hat{\hat{Y}}_S^*$) and Rao (1991) difference type estimator ($\hat{\hat{Y}}_{R,D}$) , we have

$$MSE_{min}(\hat{\hat{Y}}_S^*) < MSE_{min}(\hat{\hat{Y}}_{R,D})$$

$$MSE_{min}(\hat{\hat{Y}}_{R,D}) - MSE_{min}(\hat{\hat{Y}}_S^*) > 0$$

The condition is true if and only if

$$\sum_{h=1}^{L} W_h^2 \Bigg[\frac{\lambda_h^2 \theta^2 \bar{Y_h}^2 C_{x_h}^2 \left\{\theta^2 C_{x_h}^2 + 16C_{y_h}^2 \left(1-\rho_{yx_h}^2\right)\right\}}{64\left\{1+\lambda_h C_{y_h}^2\left(1-\rho_{yx_h}^2\right)\right\}} + \frac{\lambda_h \bar{Y_h}^2 C_{y_h}^2 \left(\rho_{yr_{xh}} - \rho_{yx_h}\rho_{xr_{xh}}\right)^2 \left(-8+\lambda_h\theta^2 C_{x_h}^2\right)^2}{64\left(1-\rho_{xr_{xh}}^2\right)\left\{1+\lambda_h C_{y_h}^2\left(1-\rho_{yx_h}^2\right)\right\}\left\{1+\lambda_h C_{y_h}^2\left(1-Q_{y.xr_{xh}}^2\right)\right\}} \Bigg] > 0. \tag{4.20}$$

(e) By comparing the proposed estimator ($\hat{\hat{Y}}_S^*$) and the ratio type exponential estimator ($\hat{\hat{Y}}_{GK,G}$) , we have

$$MSE_{min}(\hat{\hat{Y}}_S^*) < MSE_{min}(\hat{\hat{Y}}_{GK,G})$$

$$MSE_{min}(\hat{\hat{\bar{Y}}}_{GK,G}) - MSE_{min}(\hat{\hat{\bar{Y}}}^*_S) > 0$$

The condition is true if and only if

$$\sum_{h=1}^{L} W_h^2 \left[\frac{\lambda_h \bar{Y_h}^2 C_{y_h}^2 \left(\rho_{yr_{xh}} - \rho_{yx_h}\rho_{xr_{xh}}\right)^2 \left(-8 + \lambda_h \theta^2 C_{x_h}^2\right)^2}{64\left(1-\rho_{xr_{xh}}^2\right)\left\{1+\lambda_h C_{yh}^2\left(1-\rho_{yx_h}^2\right)\right\}\left\{1+\lambda_h C_{y_h}^2\left(1-Q_{y.xr_{xh}}^2\right)\right\}} \right] > 0 \qquad (4.21)$$

Note that all the conditions derived above from (a) to (e) always hold true. Thus, the suggested estimator always performs better and is more efficient than all the existing estimators.

4.5 Theoretical comparison of combined estimator

The proposed combine estimator is given as

$$MSE_{min}(\hat{\hat{\bar{Y}}}^*_C) \cong \frac{\bar{Y}^2}{64T}\Big(-J_1 V_{020}^2\theta^4 + J_7 V_{020}\theta^4 - 16J_2\theta^2 V_{020} + 16J_8\theta^2 V_{110} + 16J_{10}\theta^2$$
$$-32\theta^2 J_9 V_{110} + 16J_6^2\theta^2 + 64J_2 - 64J_3 - 64J_4 + 128J_5 - 64J_6 V_{101}\Big) \qquad (4.22)$$

The proposed combine estimator holds always true when

$$MSE_{min}(\hat{\hat{\bar{Y}}}^*_C) < MSE_{min}(\hat{\hat{\bar{Y}}}^*_{Haq})$$

$$MSE_{min}(\hat{\hat{\bar{Y}}}^*_{Haq}) - MSE_{min}(\hat{\hat{\bar{Y}}}^*_C) > 0$$

The theoretical comparison of proposed combine estimator and the existing estimator makes the analytical comparison complicated. However, we have done the numerical comparison that proves that our proposed combine estimator always perform better than the existing estimator.

4.6 Numerical comparison

In this section, we examined three real datasets with the intention of conducting a numerical comparison between the estimators that are already in use and the new ones we're proposing. The summary statistics for Data 1, Data 2, and Data 3 can be found in Tables 4.2, 4.3, and 4.4, respectively.

Data 1: (Source: Singh (2003), p.1116)

y: Estimated numbers of fish caught by marine recreational fishermen in 1996

x: Estimated number of fish caught in 1994

Data 2: (Source: Särndal et al. (2003), p.652)

y: Population in 1985

Table 4.2: Summary statistics for Data 1

h	N_h	n_h	W_h	λ_h	$\bar{Y}_h$	$\bar{X}_h$	$\bar{R}_{x_h}$
1	9	5	0.1304	0.0889	5942.22	7065.11	5
2	12	5	0.1739	0.1167	4672.5	4201.91	6.5
3	8	5	0.1159	0.075	2184.5	2576.25	4.5
4	14	5	0.2029	0.1285	5276.42	5094.92	7.5
5	10	5	0.1449	0.1	5790.4	7300.9	5.5
6	16	5	0.2318	0.1375	3295.5	3931.18	8.5
h	S_{Y_h}	S_{X_h}	$S_{R_{x_h}}$	$\rho_{Y_h X_h}$	$\rho_{Y_h R_{x_h}}$	$\rho_{X_h R_{x_h}}$	
1	9956.77	12379.5	2	0.9966448	0.1174785	0.09585862	
2	5656.82	4535.59	3.44	0.9535943	0.5181903	0.593571	
3	1513.19	1908.78	2.41	0.742415	0.05958569	-0.04888645	
4	7009.47	6407.96	3.97	0.9937896	0.4590239	0.4391694	
5	5459.11	8565.1	2.89	0.9727229	-0.2217078	-0.2693162	
6	4839.05	6067.90	4.3050	0.9830913	0.2357735	0.317363	

x: Municipal tax revenue in 1985

Table 4.3: Summary statistics for Data 2

h	N_h	n_h	W_h	λ_h	$\bar{Y}_h$	$\bar{X}_h$	$\bar{R}_{x_h}$
1	48	20	0.169	0.0291	49.5833	421.7291	24.5
2	35	15	0.1232	0.0381	24.1428	176.3714	18
3	68	25	0.2394	0.0252	33.1176	314.3823	34.5
4	39	10	0.1373	0.0743	20.2564	145.3846	20
5	45	18	0.1584	0.0333	20.8444	163.667	23
6	49	17	0.1725	0.0384	23.1428	179.102	25
h	S_{Y_h}	S_{X_h}	$S_{R_{x_h}}$	$\rho_{X_h Y_h}$	$\rho_{Y_h R_{x_h}}$	$\rho_{X_h R_{x_h}}$	
1	94.5859	902.6175	13.5218	0.9986	-0.0875	-0.0959	
2	20.5471	174.0182	10.1691	0.9735	0.0081	0.0678	
3	57.4691	896.1789	17.7333	0.9836	0.0705	0.0915	
4	19.5592	153.0143	11.2921	0.9962	-0.1387	-0.1068	
5	23.313	199.8747	13.0148	0.9979	-0.0635	-0.0643	
6	24.0381	202.6107	14.1888	0.9957	-0.2046	-0.1760	

Data 3: (Source: Särndal et al. (2003), p.654)
y: Population in 1985
x: Number of municipal employees in 1984

Table 4.5 and 4.6 shows the percentage relative efficiencies of the proposed and existing estimator. Table 4.5 contains the results of separate estimator while table 4.6 shows results of combine estimator.Three real world data sets have been considered for this purpose. Results from these data sets show that the proposed estimators outperform their respective competitor.

Table 4.4: Summary statistics for Data 3

h	N_h	n_h	W_h	λ_h	$\bar{Y}_h$	$\bar{X}_h$	$\bar{R}_{x_h}$
1	48	20	0.169	0.0291	49.5833	3062.125	24.5
2	35	15	0.1232	0.0381	24.1428	1320.77	18
3	68	25	0.2394	0.0252	33.1176	2244.32	34.5
4	39	10	0.1373	0.0743	20.2564	1069.46	20
5	45	18	0.1584	0.0333	20.8444	1154.2	23
6	49	17	0.1725	0.0384	23.1428	1341.93	25
h	S_{Y_h}	S_{X_h}	$S_{R_{x_h}}$	$\rho_{X_hY_h}$	$\rho_{Y_hR_{x_h}}$	$\rho_{X_hR_{x_h}}$	
1	94.5859	6519.65	13.4741	0.9975	-0.0875	-0.1041	
2	20.5471	1369.35	10.1431	0.9604	0.0081	0.0828	
3	57.4691	6309.55	17.5654	0.9862	0.0705	0.0872	
4	19.5592	1142.83	11.2816	0.9974	-0.1387	-0.1223	
5	23.313	1459.78	13.0043	0.9966	-0.0635	-0.0786	
6	24.0381	1439.1	14.1983	0.9946	-0.2046	-0.1531	

4.7 Conclusion and discussion

In this chapter, we have extended the idea of the estimator proposed in Chapter 3 and a proposed estimator of finite population mean utilizing tied ranks under stratified random sampling. Up to the first degree of approximation, the bias and MSE of suggested estimators were calculated using stratified random sampling. It has been theoretically and quantitatively proved that the extended use of auxiliary information improves the performance of the estimators of finite population mean. Thus the suggested estimator outperforms all other considered estimators.

Table 4.5: PREs of separate estimator with existing estimator for various choices of a and b.

		Population 1		Population 2		Population 3	
Estimators		$\hat{\hat{Y}}^*_{Haq}$	$\hat{\hat{Y}}^*_{S}$	$\hat{\hat{Y}}^*_{Haq}$	$\hat{\hat{Y}}^*_{S}$	$\hat{\hat{Y}}^*_{Haq}$	$\hat{\hat{Y}}^*_{S}$
$\hat{\hat{Y}}^{*(1)}_{Haq}$	$\hat{\hat{Y}}^{*(1)}_{S}$	3439.108	3667.428	7476.129	7527.666	7582.873	7671.653
$\hat{\hat{Y}}^{*(2)}_{Haq}$	$\hat{\hat{Y}}^{*(2)}_{S}$	3438.971	3667.277	7409.084	7459.847	7571.559	7660.118
$\hat{\hat{Y}}^{*(3)}_{Haq}$	$\hat{\hat{Y}}^{*(3)}_{S}$	3439.151	3667.476	7482.186	7533.794	7583.689	7672.484
$\hat{\hat{Y}}^{*(4)}_{Haq}$	$\hat{\hat{Y}}^{*(4)}_{S}$	3439.028	3667.34	7449.572	7500.792	7578.91	7667.559
$\hat{\hat{Y}}^{*(5)}_{Haq}$	$\hat{\hat{Y}}^{*(5)}_{S}$	3439.124	3667.447	7479.96	7531.542	7583.385	7672.175
$\hat{\hat{Y}}^{*(6)}_{Haq}$	$\hat{\hat{Y}}^{*(6)}_{S}$	3439.134	3667.458	7481.369	7532.967	7583.575	7672.368
$\hat{\hat{Y}}^{*(7)}_{Haq}$	$\hat{\hat{Y}}^{*(7)}_{S}$	3439.106	3667.427	7476.045	7527.581	7582.863	7671.643
$\hat{\hat{Y}}^{*(8)}_{Haq}$	$\hat{\hat{Y}}^{*(8)}_{S}$	3439.154	3667.48	7482.283	7533.892	7583.702	7672.497
$\hat{\hat{Y}}^{*(9)}_{Haq}$	$\hat{\hat{Y}}^{*(9)}_{S}$	3438.968	3667.273	7408.354	7459.109	7571.427	7659.983
$\hat{\hat{Y}}^{*(10)}_{Haq}$	$\hat{\hat{Y}}^{*(10)}_{S}$	3372.418	3594.414	7246.961	7296.044	7356.441	7441.204

Table 4.6: PREs of combine estimator with existing estimator for various choices of a and b.

		Population 1		Population 2		Population 3	
Estimators		$\hat{\hat{Y}}^*_{Haq}$	$\hat{\hat{Y}}^*_{C}$	$\hat{\hat{Y}}^*_{Haq}$	$\hat{\hat{Y}}^*_{C}$	$\hat{\hat{Y}}^*_{Haq}$	$\hat{\hat{Y}}^*_{C}$
$\hat{\hat{Y}}^{*(1)}_{Haq}$	$\hat{\hat{Y}}^{*(1)}_{C}$	2161.021	2170.533	1803.772	1814.426	1899.414	1903.984
$\hat{\hat{Y}}^{*(2)}_{Haq}$	$\hat{\hat{Y}}^{*(2)}_{C}$	2160.995	2170.506	1802.016	1812.654	1898.996	1903.564
$\hat{\hat{Y}}^{*(3)}_{Haq}$	$\hat{\hat{Y}}^{*(3)}_{C}$	2161.027	2170.539	1803.903	1814.558	1899.433	1904.003
$\hat{\hat{Y}}^{*(4)}_{Haq}$	$\hat{\hat{Y}}^{*(4)}_{C}$	2161.025	2170.537	1803.574	1814.226	1899.383	1903.953
$\hat{\hat{Y}}^{*(5)}_{Haq}$	$\hat{\hat{Y}}^{*(5)}_{C}$	2161.01	2170.521	1803.693	1814.346	1899.402	1903.971
$\hat{\hat{Y}}^{*(6)}_{Haq}$	$\hat{\hat{Y}}^{*(6)}_{C}$	2161.027	2170.539	1803.883	1814.538	1899.431	1904
$\hat{\hat{Y}}^{*(7)}_{Haq}$	$\hat{\hat{Y}}^{*(7)}_{C}$	2161.025	2170.537	1803.848	1814.502	1899.425	1903.995
$\hat{\hat{Y}}^{*(8)}_{Haq}$	$\hat{\hat{Y}}^{*(8)}_{C}$	2161.028	2170.539	1803.904	1814.559	1899.434	1904.004
$\hat{\hat{Y}}^{*(9)}_{Haq}$	$\hat{\hat{Y}}^{*(9)}_{C}$	2161.024	2170.535	1803.403	1814.053	1899.354	1903.924
$\hat{\hat{Y}}^{*(10)}_{Haq}$	$\hat{\hat{Y}}^{*(10)}_{C}$	2154.087	2163.56	1800.444	1811.072	1895.799	1900.357

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Survey sampling is used by academics to accurately estimate the population mean. This can be accomplished with efficient and unbiased estimators. Effective utilization of the auxiliary information may increase an estimator's efficiency and accuracy. In this research, we concentrated on utilizing the auxiliary variable and the ranked auxiliary variable to construct new estimators of the finite population mean under simple and stratified random sampling. We propose a two-fold application of auxiliary information in which auxiliary information is complemented by ranks of the auxiliary variable.

Suppose that we want to check the performance of graduated students of specific university and their level of attaining a good job in the society. Consider the number of graduated students of a university as a study variable(Y) and the number of teachers as an auxiliary variable(X). There are 26 departments in the university and is possible that the number of teachers can be same in some of the departments. We assign tied ranks to the teachers that is the departments having same number of teachers are assigned tied rank that is the average of the tied value. This takes the variation into account. So in this way, the attained information will be less scattered. Consider an other example. We want to check the production of crops in a village. The farmers have land of 50 acres. The production depends on the amount of fertilizer farmers are giving to the field. Some farmers may be giving same amount of fertilizer to the fields. Here production of crops is study variable(Y) and amount of fertilizer giving to the land is auxiliary variable(X). Giving tied ranks to the same amount of fertilizer the farmers are giving. In this way, the scattered data will be less scattered and give as more precise result.

Simple random sampling and stratified random sampling are used to establish mathematical developments of the estimator of a population mean. The findings are completely consistent with the concept of employing auxiliary information to aid in the estimation of required attributes. We see that more rigorous usage of relevant information improves the efficiency of the estimators.

Past research has used the idea of ranking with equally spaced values to develop efficient estimators of the finite population mean. This ignores the variation in the auxiliary variable. The current research has extended the idea of ranking with tied values as it takes the variation of the auxiliary variable into account.

Chapter 3 introduced a new estimator under simple random sampling for the estimation of finite population mean utilizing original and ranked auxiliary information. Bias and MSE have been expressed up to the first order of approximation. The derived estimator has been theoretically compared against competing estimators. Conditions that establish the superiority of the generated estimator over the existing estimators have been derived. Various real-world data sets were used to show the performance of the proposed estimators.

Chapter 4 extended the idea of developed estimator into stratified random sampling which is used to reduce the heterogeneity of population for improved estimation. Upto the first degree of approximation, mathematical equations for the suggested estimator's biases and MSEs have been derived. It has been demonstrated that the proposed estimator outperforms the existing estimators both theoretically and numerically. Some real-world data sets were used for numerical illustration.

The usage of supplementary information improves an estimator's efficiency. It is recommended to use the newly developed estimators for population mean estimation because they are more efficient.

5.2 Future work

- The current work could also be extended to other sampling schemes like probability proportional to size etc.
- A similar technique can be used to estimate other population parameters i-e median etc.
- The current work can be extended to the other probability sampling techniques like systematic sampling and cluster sampling.

References

Ahmad, S., Hussain, S., Aamir, M., Yasmeen, U., Shabbir, J., and Ahmad, Z. (2021). Dual use of auxiliary information for estimating the finite population mean under the stratified random sampling scheme. *Journal of Mathematics*, 2021:1–12.

Bahl, S. and Tuteja, R. (1991). Ratio and product type exponential estimators. *Journal of information and optimization sciences*, 12(1):159–164.

Cingi, H. (1994). Sampling theory. *Beytepe, Ankara, Turkey: Hacettepe University Arts and Science School Publication*.

Cochran, W. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science*, 30(2):262–275.

Cochran, W. G. (1977). *Sampling Techniques*. John Wiley and Sons Inc., New York.

Grover, L. K. and Kaur, P. (2014). A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. *Communications in Statistics-Simulation and Computation*, 43(7):1552–1574.

Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *The Annals of Mathematical Statistics*, 14(4):333–362.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). Sample survey methods and theory, volume 1: Methods and applications. 1:1–664.

Haq, A., Khan, M., and Hussain, Z. (2017). A new estimator of finite population mean based on the dual use of the auxiliary information. *Communications in Statistics-Theory and Methods*, 46(9):4425–4436.

Irfan, M., Javeda, M., and Bhatti, S. H. (2022). Difference-type-exponential estimators based on dual auxiliary information under simple random sampling. *Scientia Iranica. Transaction E, Industrial Engineering*, 29(1):343–354.

Javed, M. and Irfan, M. (2020). A simulation study: new optimal estimators for population mean by using dual auxiliary information in stratified random sampling. *Journal of Taibah University for Science*, 14(1):557–568.

Kadilar, C. and Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 45(2):218–225.

Kadilar, C. and Cingi, H. (2005). A new ratio estimator in stratified random sampling. *Communications in Statistics—Theory and Methods*, 34(3):597–602.

Khoshnevisan, M., Singh, R., Chauhan, P., and Sawan, N. (2007). A general family of estimators for estimating population mean using known value of some population parameter (s). 22(2):181–191.

Koyuncu, N. and Kadilar, C. (2009a). Family of estimators of population mean using two auxiliary variables in stratified random sampling. *Communications in Statistics-Theory and Methods*, 38(14):2398–2417.

Koyuncu, N. and Kadilar, C. (2009b). Ratio and product estimators in stratified random sampling. *Journal of Statistical Planning and Inference*, 139(8):2552–2558.

Laplace, P. S. (1820). *Théorie analytique des probabilités*. Courcier.

Montgomery, D. C., Peck, E. A., and vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association*, 33(201):101–116.

Rao, T. (1991). On certail methods of improving ration and regression estimators. *Communications in Statistics-Theory and Methods*, 20(10):3325–3340.

Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.

Shabbir, J. and Gupta, S. (2005). Improved ratio estimators in stratified sampling. *American Journal of Mathematical and Management Sciences*, 25(3-4):293–311.

Shabbir, J. and Gupta, S. (2007). On improvement in variance estimation using auxiliary information. *Communications in Statistics—Theory and Methods*, 36(12):2177–2185.

Shabbir, J. and Gupta, S. (2019). Using rank of the auxiliary variable in estimating variance of the stratified sample mean. *International Journal of Computational and Theoretical Statistics*, 6(2).

Shabbir, J. and Onyango, R. (2022). Use of an efficient unbiased estimator for finite population mean. *Plos one*, 17(7):e0270277.

Singh, H. and Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, 6(4):555–560.

Singh, R., Chauhan, P., Sawan, N., and Smarandache, F. (2009). Improvement in estimating the population mean using exponential estimator in simple random sampling. *Auxiliary Information and a priori Values in Construction of Improved Estimators*, 33.

Singh, S. (2003). *Advanced Sampling Theory With Applications: How Michael"" Selected"" Amy*, volume 2. Springer Science & Business Media.

Sisodia, B. and Dwivedi, V. (1981). Modified ratio estimator using coefficient of variation of auxiliary variable. *Journal-Indian Society of Agricultural Statistics*, 33(2):13–18.

Ullah, K., Hudson, I., Cheema, S., Khan, A., Rahman, A., and Hussian, Z. M. (2021). Use of auxiliary information in estimation of the finite population mean: An exponential type estimator. In *Proceedings od the 24th International Congress on Modelling and Simulation (MODSIM2021)*, pages 1–7. Modelling and Simulation Society of Australia and New Zealand.

Upadhyaya, L. N. and Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 41(5):627–636.

Yaqub, M., Shabbir, J., and Gupta, S. N. (2017). Estimation of population mean based on dual use of auxiliary information in non response. *Communications in Statistics-Theory and Methods*, 46(24):12130–12151.

Document Viewer

Turnitin Originality Report

Processed on: 23-Aug-2023 11:56 PKT
ID: 2149838857
Word Count: 13588
Submitted: 1

Using Auxiliary Information and Their Ranks f... By Sughra Sarwar

Similarity Index	Similarity by Source	
18%	Internet Sources:	17%
	Publications:	16%
	Student Papers:	6%

include quoted | include bibliography | excluding matches < 20 words | mode: quickview (classic) report | print | refresh | download

1% match (Internet from 30-Oct-2022)
http://prr.hec.gov.pk

1% match (Internet from 08-Nov-2022)
http://www.hjms.hacettepe.edu.tr

1% match (Maria Javed, Muhammad Irfan. "A simulation study: new optimal estimators for population mean by using dual auxiliary information in stratified random sampling", Journal of Taibah University for Science, 2020)
Maria Javed, Muhammad Irfan. "A simulation study: new optimal estimators for population mean by using dual auxiliary information in stratified random sampling", Journal of Taibah University for Science, 2020

1% match ()
Muhammad Tahir, Bu Yude, Saima Bashir, Sardar Hussain, Tahir Munir. "A new improved estimator for the population mean using twofold auxiliary information under simple random sampling", Management Science Letters, 2023

1% match (Javid Shabbir, Sat Gupta. "Estimation of finite population mean using two auxiliary variables in stratified two-phase sampling", Communications in Statistics - Simulation and Computation, 2015)
Javid Shabbir, Sat Gupta. "Estimation of finite population mean using two auxiliary variables in stratified two-phase sampling", Communications in Statistics - Simulation and Computation, 2015

1% match ()
Sousa, Rita Cristina Pinto de. "Parameter estimation in the presence of auxiliary information", Faculdade de Ciências e Tecnologia, 2013

1% match (Internet from 24-Aug-2022)
https://www.hindawi.com/journals/mpe/2022/2545517/

1% match (student papers from 11-Mar-2013)
Submitted to Staffordshire University on 2013-03-11

<1% match (Internet from 16-Dec-2022)
http://prr.hec.gov.pk

<1% match (Internet from 31-Oct-2022)
http://prr.hec.gov.pk

<1% match (Internet from 08-Nov-2022)
http://www.hjms.hacettepe.edu.tr

<1% match (Maria Javed, Muhammad Irfan. "A simulation study: new optimal estimators for population mean by using dual auxiliary information in stratified random sampling", Journal of Taibah University for Science, 2020)
Maria Javed, Muhammad Irfan. "A simulation study: new optimal estimators for population mean by using dual auxiliary information in stratified random sampling", Journal of Taibah University for Science, 2020

<1% match (Internet from 28-Feb-2021)
https://doaj.org/article/cf3013d3de2e438b99a8d87bb3c94c8e

<1% match ()
Editors, JMASM. "Vol. 13, No. 1 (Full Issue)", 'Wayne State University Library System', 2014

<1% match ()
Editors, JMASM. "Vol. 12, No. 2 (Full Issue)", 'Wayne State University Library System', 2013

<1% match (student papers from 19-Apr-2021)
Submitted to Higher Education Commission Pakistan on 2021-04-19

<1% match (student papers from 18-May-2017)
Submitted to Higher Education Commission Pakistan on 2017-05-18

<1% match (student papers from 24-Aug-2016)
Submitted to Higher Education Commission Pakistan on 2016-08-24

<1% match (Sarjinder Singh. "Advanced Sampling Theory with Applications", Springer Science and Business Media LLC, 2003)
Sarjinder Singh. "Advanced Sampling Theory with Applications", Springer Science and Business Media LLC, 2003

<1% match (Internet from 02-Nov-2022)
https://libres.uncg.edu/ir/uncg/f/Zatezalo_uncg_0154D_12085.pdf

<1% match (Internet from 02-Jan-2023)