Improved Estimation of Finite Population Variance in Stratified Random Sampling

By

Amir Yousaf

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2024

بسم الله الرحمن الرحيم

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

Improved Estimation of Finite Population Variance in Stratified Random Sampling

By

Amir Yousaf

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

2024

Declaration

I “Amir Yousaf” hereby solemnly declare that this thesis titled, “Your thesis title is written here...”.

- 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 am feeling great honor and pleasure to dedicate this research work to

My parent and sibling

Whose endless affection, prayers and wishes have been a great source of comfort for me during my whole education period and my life

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Dr. Manzoor Khan, for his invaluable guidance, unwavering support, and mentorship throughout the course of my research. His expertise, encouragement, and constructive feedback have played a pivotal role in shaping the direction of my work and enhancing its quality.

I am also thankful for the support and insights provided by Dr. Abdul Haq. His expertise and willingness to share knowledge have been instrumental in enriching my understanding of the subject matter.

This research would not have been possible without the collective support, guidance, and encouragement from Dr. Manzoor Khan, and Dr. Abdul Haq. I am truly grateful for the opportunity to work under their mentorship and benefit from their wisdom.

Abstract

This paper focuses on the formulation of estimators for the finite population variance, leveraging information on the first and second raw moments of the study variable under stratified random sampling. Furthermore, we introduce combined and separate estimators that utilize supplementary information on a study variable along with auxiliary variable to estimate the population variance. The comparative evaluation of the proposed estimators against existing one is conducted based on absolute bias and relative efficiency. Both simulated and empirical studies indicate that the newly proposed estimators demonstrate potential superiority over their existing counterparts.

Contents

List of Tables

Chapter 1

Introduction

1.1 History

Early censuses and data gathering for resource allocation and governance demonstrate the ancient origins of the concept of sampling. But sampling became formally recognized as a scientific approach in the 17th and 18th centuries, thanks in large part to the work of John Graunt in the field of population statistics. With the important contributions of statisticians like Fisher, Neyman, and Wald, sampling theory, experimental design, confidence interval construction, and hypothesis testing made tremendous strides in the 20th century. This development, influenced by advances in mathematics and technology, has taken sampling from a crude instrument for gathering data to a sophisticated, vital methodology used in a wide range of study fields.

1.2 Sampling Designs

The techniques and procedures used to choose a sample from a population for investigation or research are known as sampling designs. These methods provide a framework for choosing a sample that is representative and able to yield reliable and accurate results. There are some common sampling methods that are briefly covered. By using probability sampling procedures, it is ensured that there is a known, non-zero chance of selection for each individual or unit in the population. These solutions rely on statistical ideas and random selection to attain representativeness. Cluster sampling(CS), stratified sampling(SS), and simple random sampling(SRS) are examples of probability sampling techniques. Random selection is not used in non-probability sampling techniques, and it is not ensured that every single unit of the population has an equivalent chance of being taken into account. Usually, these techniques are used when probability sampling is not feasible. Chain-referral sampling, convenience sampling, and selective sampling are examples of non-probability sampling techniques. The process of cluster sampling involves dividing the population into groups or clusters, usually based on how close they are to one another. The sample is drawn at random from the designated clusters and includes all individuals or items found within them. Cluster sampling is often more practical and cost-effective when the population is dispersed geographically. Systematic sampling is the process of regularly choosing individuals or groups of individuals from a community. Divide the population size by the suitable sample size to find the sampling interval. The sample interval is added to the serial number of the first selected unit to determine the subsequent choices, which are made after the first person or item is randomly selected. Snowball sampling is used when the population of interest is hard to find or identify. The original participants, sometimes referred to as "seeds," are selected using non-probability techniques. After that, participants are asked to recommend people who meet the study's inclusion requirements. This sampling method uses referrals to expand the sample size. Purposive sampling involves the deliberate selection of individuals or elements based on predefined characteristics or attributes that align with the objectives of the research project. Researchers use their best judgment to choose participants who they think are most pertinent or knowledgeable about the topic at hand. Purposive sampling is useful in qualitative research and when studying uncommon or unusual groups, though subjective approaches are unlikely to yield a representative sample.

1.3 Transformation

Data transformation is the process of modifying the original data using mathematical or statistical methods to meet predetermined assumptions or accomplish specific goals. Data transformation may be useful to enhance linear correlations between variables, normalize variables, reduce skewness, improve data distribution, or stabilize variance. We applied the antithetic variable technique in transformation, a well-liked method for lowering variance in Monte Carlo simulations. By lowering the variance of the estimates, the technique aims to improve the estimators' accuracy and efficiency. Two sets of random variables the original set and the antithetic set are used in the antithetic variable technique. To create the antithetic set, the original set is transformed while preserving the correlation structure among the variables. This transformation typically involves the random variables' values being flipped or their signs being altered. The researcher uses this data to develop effective estimators.

1.4 Auxiliary Information

Auxiliary variables play a crucial role in statistical analysis and survey sampling by providing supplementary details about the population under study. While not the main focus, these variables are linked to the study variables and can impact the study design and data analysis. The relationship between study variables (Y) and auxiliary variables (X) varies based on the research context and objectives. Including auxiliary variables enhances the precision of estimates and improves the overall design and analysis of a study.

Researchers should carefully consider the choice and inclusion of auxiliary variables, taking into account their relevance to the study and their potential to enhance analysis quality. The incorporation of auxiliary variables in sampling is particularly important for improving the effectiveness, accuracy, and representativeness of survey estimates. Wisely selecting auxiliary variables allows researchers to optimize sampling design, allocate sample units strategically, and make necessary estimation adjustments. Inclusion of supplementary data results in more accurate and reliable study outcomes. Also known as ancillary data, auxiliary information is vital to improving the accuracy and efficiency of sampling techniques. It entails the utilization of extra external data that offers clarification or helps to enhance projections regarding the intended audience. Auxiliary data may comprise, for example, demographic information, historical trends, or additional measurements pertaining to the main study. Reducing sampling errors and improving estimators are two benefits of incorporating such auxiliary data into sampling designs. Auxiliary data, when properly included, can result in more accurate conclusions and more efficient use of resources in surveys and sampling, enhancing the overall quality of statistical estimations Hansen et al. (1953).

The literature on survey sampling often overlooks the estimation of finite population variance using auxiliary data in Stratified Random Sampling (STRS), mainly due to the challenges posed by bias and complexity in the conventional variance estimator for STRS. However, some studies have begun to address this gap. For example, Kadilar and Cingi (2006) introduced a biased estimator for the finite population variance under STRS. They also improved the accuracy of this estimator by incorporating a combined ratio estimator of the population variance, which utilizes secondary info from a specific correlated auxiliary variable. This ratio estimator works well when there is a strong positive correlation (e.g., above 0.50) among the main study variable and the auxiliary variable.

Building on Kadilar and Cingi (2006)'s work, Sidelel et al. (2014) proposed an extension by introducing a combined ratio estimator that incorporates info from two auxiliary variables when estimating the population variance under STRS. Additionally, Özel et al. (2014) as well as Cekim and Kadilar (2020a) suggested separate biased ratio estimators for the population variance.

For further exploration, researchers can delve into works such as Singh and Solanki (2013), Cekim and Kadilar (2020b), Qian (2020), Mahanty and Mishra (2020), Singh et al. (2021), Shahzad et al. (2023), and other related references cited therein, which offer insights into the estimate of finite population mean and variance.

1.5 Sampling Methods

Measuring population features, particularly the population mean, is the main objective of sampling theory. The literature at review sampling defines a number of sample processes and approaches that are widely used to estimate population characteristics. Auxiliary information is usually employed throughout the sampling and/or estimating stages in instruction to improve the accuracy of the mean estimator. It is expected that in cases while here is a significant correlation among the study and auxiliary variables, the ratio of the study to auxiliary variables will be less changeable than the study variable alone. On the other hand, the product technique of estimation is the better option while here is a negative correlation among the study and auxiliary variables. In some situations, these estimators are known to exhibit smaller variances than the traditional mean-per-unit estimator when used with simple random sampling (SRS). The sampling strategy and estimators to be employed must be decided upon in detail before conducting statistical research can begin.

To provide accurate estimates, many sampling strategies have been devised; the most widely used ones are SS and basic random sampling.

1.5.1 Simple Random Sampling(SRS)

SRS stands out as the most commonly employed strategy for data collection when the goal is to make inferences about a population based on the analysis of a subset. Precision in results is particularly enhanced when population units exhibit homogeneity. In SRS, each individual unit in a population is randomly selected, ensuring that every sample has an equal chance of being chosen at any stage of the process. Two methods are employed for drawing a sample from SRS: one involves a replacement scheme, and the other operates without replacement. Within the context of SRS with replacement (SRSWR), every unit in the population has an equally likelihood of individual chose, and then replaced before the next sample unit is drawn. On the other hand, in sampling without replacement (SRSWOR), a subset of observations is randomly chosen, and once selected, an observation cannot be chosen again.

1.5.2 Stratified Random Sampling(STRS)

When dealing with heterogeneous population units, simple random sampling (SRS) may not yield satisfactory results because the selected sample might not adequately represent the population, leading to a decrease in the precision of the estimator. To address this limitation, researchers often turn to (STRS. In the STRS approach, the entire population of interest is divide into strata—homogeneous groups from which samples are independently and separately drawn. It is essential that these strata are mutually exclusive and do not overlap. The strata should consist of uniform units, with minimal variability within subgroups and maximal diversity between them. Various allocation methods are employed to determine sample size, such as proportional allocation, equal allocation, optimum allocation, and Neyman's allocation. This ensures a more accurate representation of the entire population and enhances the precision of the estimator.

1.6 Motivation for the Present Study

In simulation studies, the utilization of antithetic variables aims at enhancing estimation efficiency. Similarly, in the context of survey sampling, the primary objective is the efficient estimation of the population variation by incorporating auxiliary variables. In this scenario, the focus is on converting the study variable into the auxiliary variable to achieve a precise and effective approximation of the population variation.

1.7 Thesis Outline

We consumed the first and second raw moments of the study variables to determine the finite population variance(FPV) under (STRS) in Chapter 2. Then we operated simple and combined/separate (difference) estimators used for these moments. These estimators need data on the study variable and another auxiliary variable, which agrees us to originate biased and unbiased(B/Ub) estimators for the FPV under (STRS). We calculated the proposed estimators by doing simulation and real data and linked them with existing estimators by Kadilar and Cingi (2006), Özel et al. (2014), and Cekim and Kadilar (2020b). The assessment utilized both the Absolute Bias (AB) and Relative Efficiency (RE) as criteria, offering a comprehensive evaluation of the measured variance estimators under (STRS).

Similar to Chapter 2, in Chapter 3, we've extended our approach toward show the FPV under (STRS) while accounting for measurement errors, utilizing the first and second raw moments of the study variables. Subsequently, we applied simple and combined/separate (difference) estimators for the first and second raw moments. These estimators necessitate Details on the study variable and one or more auxiliary variables, facilitating the development of B/Ub estimators for the FPV under (STRS) within the presence of measurement errors. To assess how well the suggested estimators performed in the context of measurement errors, actual data was used. The findings revealed that the separate class of estimators demonstrates greater precision compared to the combined class of estimators.

Chapter 2

Literature Review

In the context of in-sample surveys, stratification is a widely used sampling method aimed at mitigating population heterogeneity, ultimately enhancing the precision of estimators for underlying population parameters. This sampling strategy proves valuable in practical scenarios, for example, calculating the mean salary of public employees at various ranks. In the method of (STRS), the population are initially distributed into distinct groups, referred to as strata. This division ensures homogeneity within each stratum concerning the study variable. Subsequently, random samples are selected from each stratum, typically utilizing a simple random sampling approach. The benefits of stratification encompass cost reduction in surveys and increased administrative efficiency.

Among sample surveys, improving the exactness of an estimator during the approximation phase is achievable through the judicious utilization of supplementary information. This information is often provided in the form of one or more auxiliary variables or characteristics. Usually employed estimator for population parameters, such as the quantiles, distribution function, variance, mean, and median, involve classical ratio, product, and difference/regression estimators in survey sampling. For further insights, one can refer to Sampling Techniques (1977).

The literature on survey sampling often overlooks the estimation of FPV using auxiliary information in (STRS), mainly due to the challenges posed by bias and complexity in the conventional variance estimator for STRS. However, some studies have begun to address this gap. For example, Kadilar and Cingi (2006) introduced a bias estimators for the FPV under (STRS). Additionally, they increased the accuracy of this estimate by adding a combined ratio estimator of the population variance, which makes use of additional data from a single auxiliary variable that is correlated. When the primary study variable and the auxiliary variable have a significant positive correlation (i.e., above 0.50), the ratio estimator performs well.

Building on Kadilar and Cingi (2006)'s work, Sidelel et al. (2014) proposed an extension by introducing a combined ratio estimator that incorporates information from two auxiliary variables when estimate the population variance under (STRS). Additionally, Özel et al.

(2014) as well as Cekim and Kadilar (2020a) suggested separate biassed ratio estimators for the populations variance.

For further exploration, researchers can delve into works such as Singh and Solanki (2013), Cekim and Kadilar (2020b), Qian (2020), Mahanty and Mishra (2020), Singh et al. (2021), Shahzad et al. (2023), and other related references cited therein, which provide information on how to estimate the mean and variance of a finite population.

In this investigation, we start by deriving a mathematical expression for the FPV in (STRS) using the first and second raw moments of the population. We then develop both biased and unbiased estimators for this variance, utilizing the corresponding sample raw moments. Drawing from techniques in survey sampling literature, we integrate supplementary information, often in the form of auxiliary variables, to enhance estimator precision. Building upon this, we introduce combined or separate difference estimators for the population's first and second raw moments, which in turn serve as the basis for constructing B/Ub estimators for the FPV, requiring data on one or two auxiliary variables in addition to the study variable.

To evaluate the effectiveness of both current and proposed variance estimators in (STRS), we assess them based on (ABs) and (REs) using simulated and real population data. Results suggest that the proposed estimators generally demonstrate superior precision when compared to existing methods.

2.1 Notations and Symbols

Consider a population denoted as $U = \{1, 2, 3, \ldots, N\}$ comprising N individuals. Let (Y_i, X_i) denote the study variable's values (Y) and auxiliary variable (X) for the ith unit within a finite population. Consider a population, denoted as U, which, according to a stratifying variable, is divided into L strata. For $h = 1, 2, 3, \ldots, L$, the h-th stratum has N_h units, so that $\sum_{h=1}^{L} N_h = N$.Let the weight of the h-th stratum be given by $W_h = \frac{N_h}{N}$. In the h-th stratum, where $i = 1, 2, 3, \ldots, N_h$, let Y and X stand for the study and auxiliary variables, respectively, which take values $Y_{i,h}$ and $X_{i,h}$. Let $(\bar{Y}, \bar{X})$ and (S_Y^2, S_X^2) represent the population mean and variance of (Y, X), respectively, where $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, $S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$, and $S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$. In addition, let $(\bar{Y'}_2, \bar{X'}_2)$ and $(S_{Y^2}^2, S_{X^2}^2)$ be the population mean and variance of (Y^2, X^2), respectively, where $\bar{Y'}_2 = \frac{1}{N}\sum_{i=1}^{N} Y_i^2$, $\bar{X'}_2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2$, $S_{Y^2}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i^4 - \bar{Y'})^2$ and $S_{X^2}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i^4 - \bar{X'}_2)^2$. Similarly, let $(\bar{Y'}_{2,h}, \bar{X}_{2,h})$ and $(S_{Y^2,h}^2, S_{X^2,h}^2)$ be the population mean and variance of (Y_h^2, X_h^2) for the h-th stratum, respectively, where $\bar{Y'}_{2,h} = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_i^2$, $\bar{X'}_{2,h} = \frac{1}{N_h}\sum_{i=1}^{N_h} X_i^2$, $S_{Y^2,h}^2 = \frac{1}{N_h-1}\sum_{i=1}^{Nh}(Y_{i,h}^2 - \bar{Y}_{2,h})^2$, $S_{X^2,h}^2 = \frac{1}{N_h-1}\sum_{i=1}^{Nh}(X_{i,h}^2 - \bar{X}_{2,h})^2$.To estimate the population variance, a (STRS) is taken from N units, with size n.. In this process, to choose n_h units from N_h units, use SRS with out replacement, with the constraint that $\sum_{h=1}^{L} n_h = n$. The determination of n_h values may involve employing an allocation scheme, such as proportional allocation, Neyman allocation, and so forth.

Under STRS, let $\bar{Y}_{st} = \sum_{h=1}^{L} W_h \bar{Y}_h = \bar{Y}$ and $\bar{X}_{st} = \sum_{h=1}^{L} W_h \bar{X}_h = \bar{X}$ be the population means of Y and X respectively, with their respective unbiased estimators: $\hat{\bar{Y}}_{st} = \sum_{h=1}^{L} W_h \hat{\bar{Y}}_h$ and $\hat{\bar{X}}_{st} = \sum_{h=1}^{L} W_h \hat{\bar{X}}_h$ where $\hat{\bar{Y}}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} Y_{i,h}$ and $\hat{\bar{X}}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} X_{i,h}$, along with their respective variances: $\mathrm{Var}(\hat{\bar{Y}}_{st}) = \sum_{h=1}^{L} \psi_h S_{Y,h}^2$ and $\mathrm{Var}(\hat{\bar{X}}_{st}) = \sum_{h=1}^{L} \psi_h S_{X,h}^2$, where $\psi_h = W_h^2 \lambda_h$ with $\lambda_h = \frac{1}{n_h} - \frac{1}{N_h}$. On similar lines, let $\bar{Y}_{2,st} = \sum_{h=1}^{L} W_h \bar{Y'}_{2,h}$,$\bar{X}_{2,st} = \sum_{h=1}^{L} W_h \bar{X}_{2,h}$, be the population means of Y^2 and X^2, respectively, with their respective unbiased estimators: $\hat{\bar{Y}}'_{2,st} = \sum_{h=1}^{L} W_h \hat{Y'}_{2,h}$, $\hat{\bar{X}}'_{2,st} = \sum_{h=1}^{L} W_h \hat{X'}_{2,h}$, where $\hat{\bar{Y}}'_{2,h} = \frac{1}{n_h}\sum_{i=1}^{n_h} Y_{i,h}^2$, $\hat{\bar{X}}'_{2,h} = \frac{1}{n_h}\sum_{i=1}^{n_h} X_{i,h}^2$, along with their respective variances: $\mathrm{Var}(\hat{\bar{Y}}'_{2,st}) = \sum_{h=1}^{L} \psi_h S_{Y^2,h}^2$, $\mathrm{Var}(\hat{\bar{X}}'_{2,st}) = \sum_{h=1}^{L} \psi_h S_{X^2,h}^2$.

In the context of STRS, the expression for the finite population variance S_Y^2 may be formulated as follows:

$$\begin{aligned} S_Y^2 &= \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{Nh}\left(Y_{i,h} - \bar{Y}_{st}\right)^2 \\ &= \sum_{h=1}^{L}\left(\frac{N_h - 1}{N-1}\right) S_{y,h}^2 + \sum_{h=1}^{L}\left(\frac{N_h}{N-1}\right)(\bar{Y}_h - \bar{Y}_{st})^2 \end{aligned} \tag{2.1}$$

Similarly, mathematical expression can be formulated for S_X^2.

We analyse the resulting relative error term to get the mean and variance of combined and separate estimators based on auxiliary information under (STRS) for the population variance S_Y^2. Let's consider the following expressions:

$$V_{rst} = E\left(\xi_0^r \xi_1^s \xi_2^t\right) = E\left[\left(\frac{\hat{\bar{Y}}_{st} - \bar{Y}}{\bar{Y}}\right)^r \left(\frac{\hat{\bar{X}}_{st} - \bar{X}}{\bar{X}}\right)^s \left(\frac{\hat{\bar{Z}}_{st} - \bar{Z}}{\bar{Z}}\right)^t\right] \tag{2.2}$$

$$V'_{rst} = E\left(\xi_0'^r \xi_1'^s \xi_2'^t\right) = E\left[\left(\frac{\hat{\bar{Y}}'_{2,st} - \bar{Y}'_2}{\bar{Y}'_2}\right)^r \left(\frac{\hat{\bar{X}}'_{2,st} - \bar{X}'_2}{\bar{X}'_2}\right)^s \left(\frac{\hat{\bar{Z}}'_{2,st} - \bar{Z}'_2}{\bar{Z}'_2}\right)^t\right] \tag{2.3}$$

This expression yields

$$V_{200} = E\left(\xi_0^2\right) = \frac{1}{\bar{Y}^2}\sum_{h=1}^{L}\psi_h S_{Y,h}^2, \qquad V'_{200} = E\left(\xi_0'^2\right) = \frac{1}{\bar{Y'}_2^2}\sum_{h=1}^{L}\psi_h S_{Y^2,h}^2,$$

$$V_{020} = E\left(\xi_1^2\right) = \frac{1}{\bar{X}^2}\sum_{h=1}^{L}\psi_h S_{X,h}^2, \qquad V'_{020} = E\left(\xi_1'^2\right) = \frac{1}{\bar{X'}_2^2}\sum_{h=1}^{L}\psi_h S_{X^2,h}^2,$$

$$V_{002} = E\left(\xi_2^2\right) = \frac{1}{\bar{Z}^2}\sum_{h=1}^{L}\psi_h S_{Z,h}^2, \qquad V'_{002} = E\left(\xi_2'^2\right) = \frac{1}{\bar{Z'}_2^2}\sum_{h=1}^{L}\psi_h S_{Z^2,h}^2,$$

$$V_{110}=E\left(\xi_0\xi_1\right)=\frac{1}{\bar{Y}\bar{X}}\sum_{h=1}^{L}\psi_h S_{YX,h}, V'_{110}=E\left(\xi'_0\xi'_1\right)=\frac{1}{\bar{Y'}_2\bar{X'}_2}\sum_{h=1}^{L}\psi_h S_{Y^2X^2,h},$$

$$V_{101}=E\left(\xi_0\xi_2\right)=\frac{1}{\bar{Y}\bar{Z}}\sum_{h=1}^{L}\psi_h S_{YZ,h}, V'_{101}=E\left(\xi'_0\xi'_2\right)=\frac{1}{\bar{Y'}_2\bar{Z'}_2}\sum_{h=1}^{L}\psi_h S_{Y^2Z^2,h},$$

$$V_{011}=E\left(\xi_1\xi_2\right)=\frac{1}{\bar{X}\bar{Z}}\sum_{h=1}^{L}\psi_h S_{XZ,h}, V'_{011}=E\left(\xi'_1\xi'_2\right)=\frac{1}{\bar{X'}_2\bar{Z'}_2}\sum_{h=1}^{L}\psi_h S_{X^2Z^2,h},$$

where $S_{YX,h}=\rho_{YX,h}S_{Y,h}S_{X,h}$, $S_{YZ,h}=\rho_{YZ,h}S_{Y,h}S_{Z,h}$, $S_{XZ,h}=\rho_{XZ,h}S_{X,h}S_{Z,h}$, $S_{Y^2X^2,h}=\rho_{Y^2X^2,h}S_{Y^2,h}S_{X^2,h}$, $S_{Y^2Z^2,h}=\rho_{Y^2Z^2,h}S_{Y^2,h}S_{Z^2,h}$, and
$S_{X^2Z^2,h}=\rho_{X^2Z^2,h}S_{X^2,h}S_{Z^2,h}$.

Here, $\rho_{YX,h}$ ($\rho_{Y^2X^2,h}$) represents the coefficient of correlation between Y_h and X_h (Y_h^2 and X_h^2) for the h-th stratum. Similar reasoning holds true for other correlation coefficients between (X,Z) and (Y,Z).

In a comparable manner, consider the following

$$V_{rst,h}=E\left(\xi_{0,h}^{r}\xi_{1,h}^{s}\xi_{2,h}^{t}\right)=E\left[\left(\frac{\hat{\bar{Y}}_h-\bar{Y}_h}{\bar{Y}_h}\right)^r\left(\frac{\hat{\bar{X}}_h-\bar{X}_h}{\bar{X}_h}\right)^s\left(\frac{\hat{\bar{Z}}_h-\bar{Z}_h}{\bar{Z}_h}\right)^t\right] \quad (2.4)$$

$$V'_{rst,h}=E\left(\xi_{0,h}^{\prime r}\xi_{1,h}^{\prime s}\xi_{2,h}^{\prime t}\right)=E\left[\left(\frac{\hat{\bar{Y}}'_{2,h}-\bar{Y}'_{2,h}}{\bar{Y}'_{2,h}}\right)^r\left(\frac{\hat{\bar{X}}'_{2,h}-\bar{X}'_{2,h}}{\bar{X}'_{2,h}}\right)^s\left(\frac{\hat{\bar{Z}}'_{2,h}-\bar{Z}'_{2,h}}{\bar{Z}'_{2,h}}\right)^t\right] \quad (2.5)$$

which results in

$$V_{200,h}=E\left(\xi_{0,h}^2\right)=\frac{1}{\bar{Y}_h^2}\lambda_h S_{Y,h}^2, \qquad V'_{200,h}=E\left(\xi_{0,h}^{\prime 2}\right)=\frac{1}{\bar{Y'}_{2,h}^2}\lambda_h S_{Y^2,h}^2,$$

$$V_{020,h}=E\left(\xi_{1,h}^2\right)=\frac{1}{\bar{X}_h^2}\lambda_h S_{X,h}^2, \qquad V'_{020,h}=E\left(\xi_{1,h}^{\prime 2}\right)=\frac{1}{\bar{X'}_{2,h}^2}\lambda_h S_{X^2,h}^2,$$

$$V_{002,h}=E\left(\xi_{2,h}^2\right)=\frac{1}{\bar{Z}_h^2}\lambda_h S_{Z,h}^2, \qquad V'_{002,h}=E\left(\xi_{2,h}^{\prime 2}\right)=\frac{1}{\bar{Z'}_{2,h}^2}\lambda_h S_{Z^2,h}^2,$$

$$V_{110,h}=E\left(\xi_{0,h}\xi_{1,h}\right)=\frac{1}{\bar{Y}_h\bar{X}_h}\lambda_h S_{YX,h}, V'_{110,h}=E\left(\xi'_{0,h}\xi'_{1,h}\right)=\frac{1}{\bar{Y'}_{2,h}\bar{X'}_{2,h}}\lambda_h S_{Y^2X^2,h},$$

$$V_{101,h}=E\left(\xi_{0,h}\xi_{2,h}\right)=\frac{1}{\bar{Y}_h\bar{Z}_h}\lambda_h S_{YZ,h}, V'_{101,h}=E\left(\xi'_{0,h}\xi'_{2,h}\right)=\frac{1}{\bar{Y'}_{2,h}\bar{Z'}_{2,h}}\lambda_h S_{Y^2Z^2,h},$$

$$V_{011'h}=E\left(\xi_{1,h}\xi_{2,h}\right)=\frac{1}{\bar{X}_h\bar{Z}_h}\lambda_h S_{XZ,h}, V'_{011,h}=E\left(\xi'_{1,h}\xi'_{2,h}\right)=\frac{1}{\bar{X'}_{2,h}\bar{Z'}_{2,h}}\lambda_h S_{X^2Z^2,h},$$

2.2 Existing estimators

We give a quick summary of some of the current combined and separate estimators for the population variance under (STRS) in this section.

2.2.1 Kadilar and Cingi (2006) Combined Estimators

Kadilar and Cingi (2006) proposed several estimators to assess the FPV under Stratified Random Sampling (STRS). In alignment with equation (3.1), they recommended the following estimators for S_Y^2 and S_X^2

$$\hat{S}_Y = \sum_{h=1}^{L}\left(\frac{n_h - 1}{n-1}\right)\hat{S}^2_{X,h} + \sum_{h=1}^{L}\left(\frac{n_h}{n-1}\right)(\hat{\bar{Y}}_h - \hat{\bar{Y}}_{st})^2 \tag{2.6}$$

and

$$\hat{S}_X = \sum_{h=1}^{L}\left(\frac{n_h - 1}{n-1}\right)\hat{S}^2_{X,h} + \sum_{h=1}^{L}\left(\frac{n_h}{n-1}\right)(\hat{\bar{X}}_h - \hat{\bar{X}}_{st})^2 \tag{2.7}$$

Furthermore, this one be able to demonstrated that these estimators $\hat{S}^2_{Y,st}$ and $\hat{S}^2_{X,st}$ exhibit bias in estimating S_Y^2 and S_X^2 respectively, in (STRS).

Kadilar and Cingi (2006) also suggested a combined ratio estimator for the estimate of S_Y^2, which is expressed as using data from an auxiliary variable X.

$$t_1 = \hat{S}^2_{Y,st}\left(\frac{aS_X^2 + b}{a\hat{S}^2_{X,st} + b}\right) \tag{2.8}$$

In this case, a and b stand for the given parameters, usually unit-free coefficients, related from the auxiliary factor X; for example, C_X and β_{2X} indicate the auxiliary variable X's kurtosis and coefficients of variations, respectively. As expected, t_1 is a biassed estimator of S_Y^2. The mean square error (MSE) of t1 was determined mathematically by Kadilar and Cingi (2006).

2.2.2 Sidelel et al. (2014) combined estimator

When we have access to a third auxiliary variable Z in addition to Y and X, we may use data from Y and (X, Z) to extend the combined ratio estimator that was previously discussed for predicting the population variance. Sidelel et al. (2014) developed combined ratio estimators that use data from two auxiliary variables to estimate the FPV under (STRS), improving on the foundation established by Kadilar and Cingi (2006). The expression for these estimators is:

$$u_1 = \hat{S}^2_{Y,st}\left(\frac{aS_X^2 + b}{a\hat{S}^2_{X,st} + b}\right)\left(\frac{aS_Z^2 + b}{a\hat{S}^2_{Z,st} + b}\right), \tag{2.9}$$

where

$$\hat{S}_Z = \sum_{h=1}^{L}\left(\frac{n_h - 1}{n-1}\right)\hat{S}^2_{Z,h} + \sum_{h=1}^{L}\left(\frac{n_h}{n-1}\right)(\hat{\bar{Z}}_h - \hat{\bar{Z}}_{st})^2$$

The parameters c and d represent known coefficients of Z, such as C_Z and β_{2Z}As expected, the estimator u_1 exhibits bias in estimating S^2_Y.The Mean Squared Error (MSE) within this estimate has a quantitative formulation given by Sidelel et al. (2014).

2.2.3 Özel et al. (2014) separate estimator

In order to determine the population variance with (STRS), Özel et al. (2014) suggested a collection of distinct ratio-type estimators using data from a single auxiliary variable (X). These estimators are expressed as:

$$t_2 = \sum_{h=1}^{L} W_h \hat{S}^2_{Y,h}\left(\frac{a_h S^2_X + b_h}{a_h \hat{S}^2_{X,h} + b_h}\right) \tag{2.10}$$

In which the parameters or coefficients specified in relation to X for the h-th stratum are represented by a_h and b_h. It is clear that bias in estimating S^2_Y is introduced by t_2. The Mean Squared Error (MSE) used in this estimator was determined mathematically by Özel et al. (2014) over the first order of approximation.

Using data from two auxiliary variables (X, Z), a set of improved distinct ratio-type estimators for population-level variance are given as follows, in accordance with the methodology of Özel et al. (2014):

$$u_2 = \sum_{h=1}^{L} W_h \hat{S}^2_{Y,h}\left(\frac{a_h S^2_X + b_h}{a_h \hat{S}^2_{X,h} + b_h}\right)\left(\frac{c_h S^2_Z + d_h}{c_h \hat{S}^2_{Z,h} + d_h}\right) \tag{2.11}$$

whereas the parameters mentioned above or coefficients related to Z for the h-th stratum are indicated by the variables c_h and d_h.

This is essential to note that both of the estimators t_2 and u_2 only take account of differences within the strata; they do not take into consideration differences between the strata. As a result, while measuring the population variance with (STRS), such estimators do not achieve unbiasedness in estimating S^2_Y. This limitation stands as a significant drawback for these estimators.

2.2.4 Cekim and Kadilar (2020b) separate estimator

Cekim and Kadilar (2020b) expanded upon the work of Özel et al. (2014) by developing a distinct ratio-type estimator for the FPV with stratified sampling that included the conventional logarithmic modification. The estimator they suggest is stated as follows:

$$t_3 = \sum_{h=1}^{L} W_h \hat{S}^2_{Y,h} \ln\left(\frac{S^2_X}{\hat{S}^2_{X,h}} - 3\right) \tag{2.12}$$

This estimator is biased when estimating S^2_Y. Much like t_2, a limitation of t_3 lies in its reliance solely on within-strata variations, overlooking variances between stratum. Additionally, within its first order of approximating, Cekim and Kadilar (2020b) offered a mathematical formula for the Mean Squared Error (MSE) used in this estimator.

Using data collected from the two auxiliary variables (X, Z), a modified distinct ratio-type estimator that the overall variance is formulated as follows, in accordance with the methods suggested by Cekim and Kadilar (2020b).

$$u_3 = \sum_{h=1}^{L} W_h \hat{S}^2_{Y,h} \ln\left(\frac{S^2_X}{\hat{S}^2_{X,h}} - 3\right) \ln\left(\frac{S^2_Z}{\hat{S}^2_{Z,h}} - 3\right) \tag{2.13}$$

This estimator is also characterized as a biased estimator for S^2_Y.

2.3 Haq et al. (2023) Combined and Separate estimators

In this investigation, we initially explore unbiased difference estimators for both the population mean and the second raw moment in Stratified Random Sampling (STRS). These estimators incorporate data from the study variable as well as one or two auxiliary variables, either combined or separately. Following this, we utilize these estimators for development B/Ub estimators for the FPV within the framework of STRS.

The population variance S^2_Y as defined in equation (3.1) can be expressed as:

$$\begin{aligned} S^2_Y &= \frac{1}{N-1} \sum_{h=1}^{L} \sum_{i=1}^{N_h} (Y_{i,h} - \bar{Y}_{st})^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \\ &= \frac{N}{N-1} \left[\frac{1}{N} \sum_{i=1}^{N} Y_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} Y_i \right)^2 \right] = \gamma[\bar{Y}_2 - \bar{Y}^2] \\ &= \gamma[E(Y^2) - (E(Y))^2] \end{aligned} \tag{2.14}$$

In this context, where $\gamma = \frac{N}{N-1}$, the initial approach involves constructing difference estimators for $E(Y)$ and $E(Y^2)$ under Stratified Random Sampling (STRS). Subsequently, these estimators are utilized in equation (2.15) to formulate unbiased estimators for S^2_Y.

With the information on (Y, X) and (Y, X, Z), this combined difference estimators of

$E(Y_{y'})$ and $E(Y_{y'}^2)$ are provided by

$$\hat{\bar{Y}}_{c,1} = \hat{\bar{Y}}_{\text{st}} + K_{1,1}(X_{\text{s})} - \hat{\bar{X}}_{\text{st}}) \tag{2.15}$$

$$\hat{\bar{Y}}'_{c,1} = \hat{\bar{Y}}'_{2,\text{st}} + K'_{1,1}(X'_{2,\text{st}} - \hat{\bar{X}}'_{2,\text{st}}) \tag{2.16}$$

and

$$\hat{\bar{Y}}_{c,2} = \hat{\bar{Y}}_{st} + K_{1,2}(\bar{X} - \hat{\bar{X}}_{st}) + K_{2,2}(\bar{Z} - \hat{\bar{Z}}_{st}) \tag{2.17}$$

$$\hat{\bar{Y}}'_{c,2} = \hat{\bar{Y}}'_{2,st} + K'_{1,2}(\bar{X}'_2 - \hat{\bar{X}}'_{2,st}) + K'_{2,2}(\bar{Z}'_2 - \hat{\bar{Z}}'_{2,st}) \tag{2.18}$$

accordingly, where

$$K_{1,1} = \frac{\bar{Y}V_{110}}{\bar{X}V_{020}}, \qquad K'_{1,1} = \frac{\bar{Y}'_2 V'_{110}}{\bar{X}'_2 V'_{020}}$$

$$K_{1,2} = \frac{\bar{Y}(V_{011}V_{110} - V_{110}V_{020})}{\bar{X}(V_{011}^2 - V_{002}V_{020})}, \qquad K'_{1,2} = \frac{\bar{Y}'_2(V'_{011}V_{110} - V'_{110}V'_{020})}{\bar{X}'_2(V_{011'}^2 - V'_{002}V'_{020})},$$

$$K_{2,2} = \frac{\bar{Y}(V_{011}V_{110} - V_{101}V_{020})}{\bar{Z}(V_{011}^2 - V_{002}V_{020})}, \qquad K'_{2,2} = \frac{\bar{Y}'_2(V'_{011}V_{110} - V'_{101}V'_{020})}{\bar{Z}'_2(V_{011'}^2 - V'_{002}V'_{020})},$$

are established constants. The variance of $\hat{\bar{Y}}_{c,1}$ and $\hat{\bar{Y}}_{c,2}$ are provided by

$$\text{Var}(\hat{\bar{Y}}_{c,1}) = \bar{Y}^2 V_{200}\left(1 - \frac{V_{110}^2}{V_{200}V_{020}}\right) = \bar{Y}^2 V_{200}(1 - \rho_{YX,st}^2) \tag{2.19}$$

and

$$\begin{aligned}\text{Var}(\hat{\bar{Y}}_{c,2}) &= \bar{Y}^2 V_{200}\left(1 - \frac{V_{110}^2 V_{002} + V_{101}^2 V_{020} - 2V_{110}V_{101}V_{011}}{V_{200}(V_{020}V_{002} - V_{011}^2)}\right) \\ &= \bar{Y}^2 V_{200}(1 - R_{Y,XZ,st}^2)\end{aligned} \tag{2.20}$$

where the correlation coefficient between Y and X is denoted by $\rho_{YX,st}^2$ and multiple correlation coefficient between Y on X and Z is denoted by $R_{Y,XZ,st}^2$ under STRS.

Sampling Techniques (1977) has provided the mean and variance of $\hat{\bar{Y}}_{c,1}$. By following analogous steps, one can determine the mean and variance of $\hat{\bar{Y}}'_{c,1}$.
The expression for the estimator $\hat{\bar{Y}}_{c,2}$ can be formulated as:

$$\begin{aligned}\hat{Y}_{c,2} &= \bar{Y}(1 + \xi_0) - K_{1,2}\bar{X}\xi_1 - K_{2,2}\bar{X}\xi_2 \\ \left(\hat{Y}_{c,2} - \bar{Y}\right) &= \bar{Y}\xi_0) - K_{1,2}\bar{X}\xi_1 - K_{2,2}\bar{Z}\xi_2\end{aligned} \tag{2.21}$$

By squaring both sides and then calculating the expectation, we derive

$$\begin{aligned}\text{Var}(\hat{\bar{Y}}_{c,2}) =& \bar{Y}^2 V_{200} + K_{1,2}X\bar{Y}^2 V_{020} + K_{2,2}Z\bar{Y}^2 V_{002} - 2K_{1,2}\bar{Y}XV_{110} \\ &- 2K_{2,2}\bar{Y}ZV_{101} + 2K_{1,2}K_{2,2}XZV_{011}.\end{aligned} \tag{2.22}$$

The optimal values for $K_{1,2}$ and $K_{2,2}$ are determined by minimizing the variance of $\hat{\hat{Y}}_{c,2}$, i.e., by taking the derivatives of $\mathrm{Var}(\hat{\hat{Y}}_{c,2})$ with respect to $K_{1,2}$ and $K_{2,2}$ and setting The subsequent computation that leads to 0. The minimum variance of $\hat{\hat{Y}}_{c,2}$ is achieved by substituting these optimal values back into the expression for $\mathrm{Var}(\hat{\hat{Y}}_{c,2})$.
Similarly, by following a similar procedure, one can determine the optimal values for $K'_{1,2}$ and $K'_{2,2}$ and minimize the variance of $\hat{\hat{Y}}'_{c,2}$. The Appendix contains the condensed versions of those constants.

Using the data on (Y, X) and (Y, X, Z), the separate difference estimators of $E(Y)$ and $E(Y^2)$ are provided by

$$\hat{\hat{Y}}_{s,1} = \sum_{h=1}^{L} W_h \hat{\hat{Y}}_{s,1,h}, \qquad \hat{\hat{Y}}'_{s,1} = \sum_{h=1}^{L} W_h \hat{\hat{Y}}'_{s,1,h} \tag{2.23}$$

$$\hat{\hat{Y}}_{s,2} = \sum_{h=1}^{L} W_h \hat{\hat{Y}}_{s,2,h}, \qquad \hat{\hat{Y}}'_{s,2} = \sum_{h=1}^{L} W_h \hat{\hat{Y}}'_{s,2,h} \tag{2.24}$$

where

$$\begin{aligned}
\hat{\hat{Y}}_{s,1,h} &= \hat{\hat{Y}}_h + K_{1,1,h}(\bar{X}_h - \hat{\hat{X}}_h) \\
\hat{\hat{Y}}'_{s,1,h} &= \hat{\hat{Y}}'_{2,h} + K'_{1,1,h}(\bar{X}'_{2,h} - \hat{\hat{X}}'_{2,h}) \\
\hat{Y}_{s,2,h} &= \hat{\hat{Y}}_h + K_{1,2,h}(\bar{X}_h - \hat{\hat{X}}_h) + K_{2,2,h}(\bar{Z}_h - \hat{\hat{Z}}_h) \\
\hat{Y}'_{s,2,h} &= \hat{\hat{Y}}'_{2,h} + K'_{1,2,h}(\bar{X}'_{2,h} - \hat{\hat{X}}'_{2,h}) + K'_{2,2,h}(\bar{Z}'_{2,h} - \hat{\hat{Z}}'_{2,h})
\end{aligned}$$

$$K_{1,1,h} = \frac{\bar{Y}_h V_{110,h}}{\bar{X}_h V_{020,h}}, \qquad K'_{1,1,h} = \frac{\bar{Y}'_{2,h} V'_{110,h}}{\bar{X}'_{2,h} V'_{020,h}},$$

$$K_{1,2,h} = \frac{\bar{Y}_h (V_{011,h} V_{110,h} - V_{110,h} V_{020,h})}{\bar{X}_h (V_{011,h}^2 - V_{002,h} V_{020,h})}, K'_{1,2,h} = \frac{\bar{Y'}_{2,h} (V'_{011,h} V_{110,h} - V'_{110,h} V'_{020,h})}{\bar{X'}_{2,h} (V_{011,h}^2 - V'_{002,h} V'_{020,h})},$$

$$K_{2,2'h} = \frac{\bar{Y}_h (V_{011,h} V_{110,h} - V_{101,h} V_{020,h})}{\bar{Z}_h (V_{011,h}^2 - V_{002,h} V_{020,h})}, K'_{2,2,h} = \frac{\bar{Y'}_{2,h} (V'_{011,h} V_{110,h} - V'_{101,h} V'_{020,h})}{\bar{Z'}_{2,h} (V_{011,h'}^2 - V'_{002,h} V'_{020,h})},$$

The constants $K_{1,1,h}$ and $K'_{1,1,h}$ are known constants, and their simplified expressions can be found in the Appendix. The variances of $\hat{\hat{Y}}_{s,1}$ and $\hat{\hat{Y}}_{s,2}$ are given by:

$$\begin{aligned}
\mathrm{Var}(\hat{\hat{Y}}_{s,1}) &= \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 V_{200,h} \left(1 - \frac{V_{110,h}^2}{V_{200,h} V_{020,h}}\right) \\
&= \sum_{h=1}^{L} \psi S_{Y,h}^2 (1 - \rho_{YX,h}^2)
\end{aligned} \tag{2.25}$$

and

$$\begin{aligned}\mathrm{Var}(\hat{Y}_{s,2}) &= \sum_{h=1}^{L} W_{2,h}\bar{Y}_{2,h}V_{200,h}\left(1-\frac{V_{110,h}^2V_{002,h}+V_{101,h}^2V_{020,h}-2V_{110,h}V_{101,h}V_{011,h}}{V_{200,h}(V_{020,h}V_{002,h}-V_{011,h}^2)}\right)\\ &= \sum_{h=1}^{L}\psi S_{Y,h}^2(1-R_{Y,XZ,h}^2)\end{aligned} \tag{2.26}$$

The quantity $R^2_{Y.XZ,h}$ represents the multiple correlation coefficient of Y_h on X_h and Z_h for the h-th stratum.

Utilising just Y data, unbiased and biassed estimators that estimate the FPV over (STRS) are provided as
The FPV over (STRS) can be estimated with both biased and unbiased estimators, utilizing information just from variable Y.

$$t_p = \gamma\left(\hat{\bar{Y}}'_{2st)} - \hat{\bar{Y}}^2_{st}\right) \tag{2.27}$$

and

$$t_p^* = \gamma\left(\hat{\bar{Y}}'_{2st} - \hat{\bar{Y}}^2_{st} + \sum_{h=1}^{L}\psi_h\hat{S}_h^2\right) \tag{2.28}$$

In the same way, $\hat{S}_h^2$ is a traditional, unbiased estimator of S_h^2.
The expectation of t_p mathematically is

$$\begin{aligned}\mathrm{E}(t_p) &= \gamma\left[\mathrm{E}\left(\hat{\bar{Y}}_{2st}\right) - \mathrm{E}\left(\hat{\bar{Y}}^2_{st}\right)\right]\\ &= \gamma\left[\bar{Y'}_2 - \bar{Y}^2 - \mathrm{V}(\bar{Y}_{st})\right]\\ &= S_Y^2 - \gamma\sum_{h=1}^{L}\psi_h S_{Y,h}^2\end{aligned}$$

This indicates that t_p is a biased estimator of S_Y^2 since it underestimates S_Y^2. A reasonable estimate for S_Y^2 can be obtained by

$$\begin{aligned}\mathrm{E}(t_p^*) &= \mathrm{E}(t_p) + \gamma\mathrm{E}\left(\sum_{h=1}^{L}\psi_h\hat{S}_{Y,h}^2\right)\\ &= \mathrm{E}(t_p) + \gamma\sum_{h=1}^{L}\psi_h S_{Y,h}^2 = S_Y^2\end{aligned} \tag{2.29}$$

Biased combined and separate estimators of the FPV over (STRS), utilizing kwonledge about on (Y, X) and (Y, X, Z), are presented as follows:

$$t_{p,1} = \gamma\left(\hat{\hat{Y}}'_{c,1} - \hat{\hat{Y}}^2_{c,1}\right) \tag{2.30}$$

$$t_{p,2} = \gamma\left(\hat{\hat{Y}}'_{c,1} - \hat{\hat{Y}}^2_{st}\right) \tag{2.31}$$

$$t_{p,3} = \gamma\left(\hat{\hat{Y}}'_{s,1} - \hat{\hat{Y}}^2_{s,1}\right) \tag{2.32}$$

$$t_{p,4} = \gamma\left(\hat{\hat{Y}}'_{s,1} - \hat{\hat{Y}}^2_{st}\right) \tag{2.33}$$

and

$$u_{p,1} = \gamma\left(\hat{\hat{Y}}'_{c,2} - \hat{\hat{Y}}^2_{c,2}\right) \tag{2.34}$$

$$u_{p,2} = \gamma\left(\hat{\hat{Y}}'_{c,2} - \hat{\hat{Y}}^2_{st}\right) \tag{2.35}$$

$$u_{p,3} = \gamma\left(\hat{\hat{Y}}'_{s,2} - \hat{\hat{Y}}^2_{s,2}\right) \tag{2.36}$$

$$u_{p,4} = \gamma\left(\hat{\hat{Y}}'_{s,2} - \hat{\hat{Y}}^2_{st}\right) \tag{2.37}$$

The expectation of $t_{p,1}$ mathematically is

$$\begin{aligned} \mathrm{E}(t_{p,1}) &= \gamma\left[\mathrm{E}\left(\hat{\hat{Y}}'_{c,1}\right) - \mathrm{E}\left(\hat{\hat{Y}}^2_{c,1}\right)\right] \\ &= \gamma\left[\bar{Y'}_2 - \bar{Y}^2 - \mathrm{V}(\hat{\hat{Y}}_{c,1})\right] \\ &= S^2_Y - \gamma\left(\sum_{h=1}^{L}\psi_h S^2_{Y,h}\right)\left(1 - \rho^2_{YX,st}\right) \end{aligned} \tag{2.38}$$

This indicates that $t_{p,1}$ also provides a biased estimation of S^2_Y, demonstrating that the estimator is biassed. It has been shown that there is bias in various estimators of S^2_Y.

$$t^*_{p,1} = t_{p,1} + \gamma\sum_{h=1}^{L}\psi_h\hat{S}^2_{Y,h}(1 - \rho^2_{YX,st}) \tag{2.39}$$

$$t^*_{p,2} = t_{p,2} + \gamma\sum_{h=1}^{L}\psi_h\hat{S}^2_{Y,h} \tag{2.40}$$

$$t^*_{p,3} = t_{p,3} + \gamma\sum_{h=1}^{L}\psi_h\hat{S}^2_{Y,h}(1 - \rho^2_{YX,h}) \tag{2.41}$$

$$t^*_{p,4} = t_{p,4} + \gamma\sum_{h=1}^{L}\psi_h\hat{S}^2_{Y,h} \tag{2.42}$$

and

$$u^*_{p,1} = u_{p,1} + \gamma \sum_{h=1}^{L} \psi_h \hat{S}^2_{Y,h}(1 - R^2_{Y,XZ,st}) \tag{2.43}$$

$$u^*_{p,2} = u_{p,2} + \gamma \sum_{h=1}^{L} \psi_h \hat{S}^2_{Y,h} \tag{2.44}$$

$$u^*_{p,3} = u_{p,3} + \gamma \sum_{h=1}^{L} \psi_h \hat{S}^2_{Y,h}(1 - R^2_{Y,XZ,h}) \tag{2.45}$$

$$u^*_{p,4} = u_{p,4} + \gamma \sum_{h=1}^{L} \psi_h \hat{S}^2_{Y,h} \tag{2.46}$$

Demonstrating the unbiasedness of the mentioned estimators is a straightforward process.

Chapter 3

Variance of a Finite Population Estimated Utilising Transformations

In this chapter, a set of unique estimators leveraging known auxiliary variables under SRS is proposed for the estimation of the finite population variance. At the initial approximation level, statements for the bias and mean square error of both the current and proposed families generated from estimators. Subsequently, we provide a conceptual comparison between the proposed set of estimators and other existing methods.

3.1 Notations and Symbols

Let U = $U = \{1, 2, \ldots, N\}$ be a finite population made up of N units., partitioned into L strata based on a stratifying variable. Each stratum, denoted by the index h (ranging from 1 to L), contains N_h units, and the total population is distributed across these strata such that the sum of N_h for all h equals N. The stratum's weight in hth, represented by W_h, is defined as N_h/N. The study variable Y'_Y and auxiliary variable (fx) are defined, taking values $(Y_{Y'(i,h)}, fx_{i,h})$ $i = 1, 2, \ldots, N_h$ in the hth stratum.

Let $(\bar{Y}_{Y'}, \bar{fx})$ and (S'^2_Y, S^2_{fx}) be the population mean and variance of $(Y_{Y'}, fx)$, respectively, where $\bar{Y}_{Y'} = \frac{1}{N}\sum_{i=1}^{N} Y_{Y'i}$, $\bar{fx} = \frac{1}{N}\sum_{i=1}^{N} fx_i$, $S^2_{Y'} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{Y'i} - \bar{Y}_{Y'})^2$, and $S^2_{fx} = \frac{1}{N-1}\sum_{i=1}^{N}(fx_i - \bar{fx})^2$. In addition, let $(\bar{Y'}_{Y'(2)}, \bar{fx'}_2)$ and $(S^2_{Y'^2}, S^2_{fx^2})$ be the population mean and variance of (Y'^2, fx^2), respectively, where $\bar{Y'}_{Y'(2)} = \frac{1}{N}\sum_{i=1}^{N} Y^2_{Y'i}$, $F\bar{X'}_2 = \frac{1}{N}\sum_{i=1}^{N} fx_i^2$, $S^2_{Y'^2} = \frac{1}{N-1}\sum_{i=1}^{N}(Y^4_{Y'i} - \bar{Y'}_{Y'})^2$ and $S^2_{fx^2} = \frac{1}{N-1}\sum_{i=1}^{N}(fx_i^4 - \bar{fx}_2)^2$. Similarly, let $(\bar{Y'}_{Y'(2,h)}, \bar{fx}_{2,h})$ and $(S^2_{Y'^2(h)}, S^2_{fx^2(h)})$ be the population mean and variance of $(Y^2_{Y'(h)}, fx_h^2)$ for the h-th stratum, respectively, where $\bar{Y'}_{Y'(2,h)} = \frac{1}{N_h}\sum_{i=1}^{N_h} Y^2_{Y'i}$, $\bar{fx'}_{2,h} = \frac{1}{N_h}\sum_{i=1}^{N_h} fx_i^2$, $S^2_{Y'(^2,h)} = \frac{1}{N_h-1}\sum_{i=1}^{Nh}(Y^2_{Y'(i,h)} - \bar{Y}_{Y'(2,h)})^2$, $S^2_{fx^2,h} = \frac{1}{N_h-1}\sum_{i=1}^{Nh}(fx^2_{i,h} - \bar{fx}_{2,h})^2$.To estimate the population variance, a (STRS) sample of size n units is drawn from N units. In this process, n_h units are selected from N_h units using (SRSWOR), ensuring that $\sum_{h=1}^{L} n_h = n$. The determination of n_h values may involve using an allocation scheme, such as proportional allocation, Neyman allocation, etc.

Under STRS, let $\bar{Y}_{Y'(st)} = \sum_{h=1}^{L} W_h \bar{Y}_{Y'(h)} = \bar{Y}_{Y'}$ and $\bar{fx}_{st} = \sum_{h=1}^{L} W_h \bar{fx}_h = \bar{fx}$ be the population means of $Y_{Y'}$ and fx respectively, with their respective unbiased estimators: $\hat{\bar{Y}}_{Y'(st)} = \sum_{h=1}^{L} W_h \hat{\bar{Y}}_{Y'(h)}$ and $\hat{\bar{fx}}_{st} = \sum_{h=1}^{L} W_h \hat{\bar{fx}}_h$ where $\hat{\bar{Y}}_{Y'(h)} = \frac{1}{n_h}\sum_{i=1}^{n_h} Y_{Y'(i,h)}$ and $\hat{\bar{fx}}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} fx_{i,h}$, along with their respective variances: $\mathrm{Var}(\hat{\bar{Y}}_{st}) = \sum_{h=1}^{L} \psi_h S^2_{Y,h}$ and $\mathrm{Var}(\hat{\bar{fx}}_{st}) = \sum_{h=1}^{L} \psi_h S^2_{fx,h}$, where $\psi_h = W_h^2 \lambda_h$ with $\lambda_h = \frac{1}{n_h} - \frac{1}{N_h}$. On similar lines, let $\bar{Y}'_{Y'(2,st)} = \sum_{h=1}^{L} W_h \bar{Y'}_{Y'(2,h)}, \bar{fx}'_{2,st} = \sum_{h=1}^{L} W_h \bar{fx}'_{2,h}$, be the population means of $Y^2_{Y'}$ and fx^2, respectively, with their respective unbiased estimators: $\hat{\bar{Y}}'_{Y'(2,st)} = \sum_{h=1}^{L} W_h \hat{\bar{Y}'}_{Y'(2,h)}, \hat{\bar{fx}}'_{2,st} = \sum_{h=1}^{L} W_h \hat{\bar{fx}}'_{2,h}$, where $\hat{\bar{Y}}'_{Y(2,h)} = \frac{1}{n_h}\sum_{i=1}^{n_h} Y^2_{Y(i,h)}$, $\hat{\bar{fx}}'_{2,h} = \frac{1}{n_h}\sum_{i=1}^{n_h} fx^2_{i,h}$, along with their respective variances: $\mathrm{Var}(\hat{\bar{Y}}'_{Y'(2,st)}) = \sum_{h=1}^{L} \psi_h S^2_{Y'^2,h}$, $\mathrm{Var}(\hat{\bar{fx}}'_{2,st}) = \sum_{h=1}^{L} \psi_h S^2_{fx^2,h}$.

In the context of Stratified Random Sampling (StRS), the expression for the finite population variance $S^2_{Y'}$ can be formulated as follows:

$$\begin{aligned} S^2_{Y'} &= \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}(Y_{Y'(i,h)} - \bar{Y}_{Y'(st)})^2 \\ &= \sum_{h=1}^{L}\left(\frac{N_h-1}{N-1}\right) S^2_{Y',h} + \sum_{h=1}^{L}\left(\frac{N_h}{N-1}\right)\left(\bar{Y}_{Y'(h)} - \bar{Y}_{Y(st)}\right)^2 \end{aligned} \tag{3.1}$$

Similarly, mathematical expression can be formulated for S^2_{fx}.

To determine the mean and variance of auxiliary-information-based (AIB) combined and separate estimators under Stratified Random Sampling (STRS) for the population variance with measurement error, denoted as $S_{Y'^2}$, we examine the following relative error terms.

$$V_{Y'(rst)} = E\left(\xi_0^r \xi_1^s\right) = E\left[\left(\frac{\hat{\bar{Y}}_{Y'(st)} - \bar{Y}_{Y'}}{\bar{Y}_{Y'}}\right)^r \left(\frac{\hat{\bar{fx}}_{st} - \bar{fx}}{\bar{fx}}\right)^s\right] \tag{3.2}$$

$$V'_{Y'(rst)} = E\left(\xi_0'^r \xi_1'^s\right) = E\left[\left(\frac{\hat{\bar{Y}}'_{Y'(2,st)} - \bar{Y}'_{Y'(2)}}{\bar{Y}'_{Y'(2)}}\right)^r \left(\frac{\hat{\bar{fx}}'_{2,st} - \bar{fx}'_2}{\bar{fx}'_2}\right)^s\right] \tag{3.3}$$

This expression yields

$$V_{Y'(20)} = E\left(\xi_0^2\right) = \frac{1}{\bar{Y}^2_{Y'}}\sum_{h=1}^{L}\psi_h S^2_{Y',h}, \qquad V'_{Y'(20)} = E\left(\xi_0'^2\right) = \frac{1}{\bar{Y'}^2_{Y'(2)}}\sum_{h=1}^{L}\psi_h S^2_{Y'^2,h},$$

$$V_{Y(02)} = E\left(\xi_1^2\right) = \frac{1}{\bar{fx}^2}\sum_{h=1}^{L}\psi_h S^2_{fx,h}, \qquad V'_{Y'(02)} = E\left(\xi_1'^2\right) = \frac{1}{\bar{fx}'^2_2}\sum_{h=1}^{L}\psi_h S^2_{fx^2,h},$$

$$V_{Y'(11)} = E\left(\xi_0\xi_1\right) = \frac{1}{\bar{Y}_{Y'}\bar{fx}}\sum_{h=1}^{L}\psi_h S_{Y'fx,h}, V'_{Y'(11)} = E\left(\xi_0'\xi_1'\right) = \frac{1}{\bar{Y'}_{Y(2)}\bar{fx'}_2}\sum_{h=1}^{L}\psi_h S_{Y^2_{Y'},fx^2,h},$$

where $S_{Y'fx,h} = \rho_{Y'fx,h} S_{Y',h} S_{fx,h}$, $S_{Y'^2fx^2,h} = \rho_{Y'^2fx^2,h} S_{Y'^2,h} S_{fx^2,h}$. Here, $\rho_{Y'fx,h}$ ($\rho_{Y'^2fx^2,h}$) represents the correlation coefficient between Y'_h and fx_h ($Y'^2_h andfx^2_h$) for the h-th stratum.

In a comparable manner, consider the following

$$V_{Y'(rst,h)} = E\left(\xi^r_{0,h}\xi^s_{1,h}\right) = E\left[\left(\frac{\hat{\bar{Y}}_{Y'(h)} - \bar{Y}_{Y'(h)}}{\bar{Y}_{Y'(h)}}\right)^r \left(\frac{\hat{\bar{f}}x_h - \bar{f}x_h}{\bar{f}x_h}\right)^s\right] \tag{3.4}$$

$$V'_{Y'(rst,h)} = E\left(\xi'^r_{0,h}\xi'^s_{1,h}\right) = E\left[\left(\frac{\hat{\bar{Y}}'_{Y'(2,h)} - \bar{Y}'_{Y'(2,h)}}{\bar{Y}'_{Y'(2,h)}}\right)^r \left(\frac{\hat{\bar{f}}x'_{2,h} - \bar{f}x'_{2,h}}{\bar{f}x'_{2,h}}\right)^s\right] \tag{3.5}$$

This expression yields

$$V_{Y'(20,h)} = E\left(\xi^2_{0,h}\right) = \frac{1}{\bar{Y}^2_{Y',h}}\lambda_h S^2_{Y',h}, \qquad V'_{Y'(20,h)} = E\left(\xi'^2_{0,h}\right) = \frac{1}{\bar{Y'}^2_{Y'(2,h)}}\lambda_h S^2_{Y'^2,h},$$

$$V_{Y(02,h)} = E\left(\xi^2_{1,h}\right) = \frac{1}{\bar{fx}^{2,h}}\lambda_h S^2_{fx,h}, \qquad V'_{Y'(02,h)} = E\left(\xi'^2_{1,h}\right) = \frac{1}{\bar{fx'}^2_{2,h}}\lambda_h S^2_{fx^2,h},$$

$$V_{Y'(11,h)} = E\left(\xi_{0,h}\xi_{1,h}\right) = \frac{1}{\bar{Y}_{Y',h}\bar{f}x_h}\lambda_h S_{Y'fx,h}, E\left(\xi'_{0,h}\xi'_{1,h}\right) = \frac{1}{\bar{Y'}_{Y'(2,h)}\bar{fx'}_{2,h}}\lambda_h S_{Y'^2fx^2,h},$$

where $S_{Y'fx,h} = \rho_{Y'fx,h} S_{Y',h} S_{fx,h}$, $S_{Y'^2fx^2,h} = \rho_{Y'^2fx^2,h} S_{Y'^2,h} S_{fx^2,h}$. Here, $\rho_{Y'fx,h}$ ($\rho_{Y'^2fx^2,h}$) represents the correlation coefficient between Y'_h and fx_h ($Y'^2_h andfx^2_h$) for the h-th stratum.

3.2 Proposed Estimators

In this investigation, our initial focus is on unbiased (both combined and separate) difference estimators for the population mean and second-raw moment within the framework of Stratified Random Sampling (STRS). We leverage information from the study variable and auxiliary variable. Subsequently, these estimators are employed to formulate B/Ub estimators for the FPV over STRS Haq et al. (2023). Following the previously mentioned estimator, an alternative approach to reorganize the population units is outlined as follows. The population of size N is arranged based on the ascending order of the auxiliary variable magnitude, as

$$Y = Y_1, Y_2, \ldots, Y_{N-1}, Y_n$$

The auxiliary variables are first organized in ascending order based on their magnitudes, and the median is determined. Following this, fx_1 is added to the smallest value of the auxiliary variable, and the same value is subtracted from the largest one. Similarly, fx_2 is added to the second smallest value, and it is subtracted from the second largest variable. This process continues for the remaining values of the auxiliary variable, excluding the median. Mathematically, the scheme is

$$Y' = Y_{Y'1}, Y_{Y'2}, \ldots, Y_{Y'i}, \ldots, Y_{Y'N-1}, Y_{Y'N}$$

$$Y' = (y_1 + fx_1),\ (y_2 + fx_2),\ \ldots,\ y_i,\ \ldots,\ (y_{N-1} - fx_2),\ (y_N - fx_1)$$

where $Y_{Y'i} = (y_i \pm fx_i), \qquad i = 1, 2, \ldots, N$.

Therefore, $(Y_{Y'1}, Y_{Y'2}, Y_{Y'3}, \ldots, Y_{Y'N-2}, Y_{Y'N-1}, Y_{Y'N})$ represents the transformed population values. The finite population variance under STRS is then estimated using these estimators to create biased and unbiased estimators.

$$\begin{aligned} S^2_{Y'} &= \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}(Y_{Yi',h} - \bar{Y}_{Y',st})^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{Y'i} - \bar{Y}_{Y'})^2 \\ &= \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}Y^2_{Y'i} - \left(\frac{1}{N}\sum_{i=1}^{N}Y'_{Y'i}\right)^2\right] = \gamma_{Y'}[\bar{Y}_{Y',2} - \bar{Y}^2_{Y'}] \\ &= \gamma_{Y'}[E(Y^2_{Y'}) - (E(Y_{Y'}))^2] \end{aligned} \tag{3.6}$$

In this context, where $\gamma_{Y'} = \frac{N}{N-1}$, the initial approach involves constructing difference estimators for $E(Y_{Y'})$ and $E(Y^2_{Y'})$ under Stratified Random Sampling (STRS). Subsequently, these estimators are utilized in equation (3.1) to formulate unbiased estimators for $S^2_{Y'}$.

Using the data on $(Y_{Y'}, fx)$, the combined difference estimators of $E(Y_{Y'})$ and $E(Y^2_{Y'})$ are provided by

$$\hat{\bar{Y}}_{Y'(c,1)} = \hat{\bar{Y}}_{Y'(\mathrm{st})} + K_{Y'(1,1)}(fx_{\mathrm{st}} - \hat{\bar{f}}x_{\mathrm{st}}) \tag{3.7}$$

and

$$\hat{\bar{Y}}'_{Y'(c,1)} = \hat{\bar{Y}}'_{Y'(2,\mathrm{st})} + K'_{Y'(1,1)}(fx'_{2,\mathrm{st}} - \hat{\bar{f}}x'_{2,\mathrm{st}}) \tag{3.8}$$

accordingly, where

$$K_{Y'(1,1)} = \frac{\bar{Y}'_{Y'}V'_{Y'(11)}}{\bar{f}xV'_{Y'(02)}}, \qquad K'_{Y'(1,1)} = \frac{\bar{Y}'_{Y'(2)}V'_{Y'(11)}}{\bar{f}x'_2V'_{Y'(02)}}$$

are established constants. The variance of $\hat{\bar{Y}}'_{Y'(c,1)}$ are provided by

$$\mathrm{Var}(\hat{\bar{Y}}_{Y'(c,1)}) = \bar{Y}^2_{Y'}V_{Y'(20)}\left(1 - \frac{V^2_{Y'(11)}}{V_{Y'(20)}V_{Y'(02)}}\right) = \bar{Y}^2_{Y'}V_{Y'(20)}(1 - \rho^2_{Y'fx,st}) \tag{3.9}$$

where the correlation coefficient between $Y_{Y'}$ and fx is denoted by $\rho^2_{Y'fx,st}$ under STRS. Using the data on $(Y_{Y'}, fx)$, the separate difference estimators of $E(Y_{Y'})$ and $E(Y^2_{Y'})$ are provided by

$$\hat{\hat{Y}}_{Y'(s,1)} = \sum_{h=1}^{L} W_h \hat{\hat{Y}}_{Y'(s,1,h)}, \qquad \hat{\hat{Y}}'_{Y'(s,1)} = \sum_{h=1}^{L} W_h \hat{\hat{Y}}'_{Y'(s,1,h)}$$

where

$$\hat{\hat{Y}}_{Y'(s,1,h)} = \hat{\hat{Y}}_{Y'(h)} + K_{Y'(1,1,h)}(\bar{f}x_h - \hat{\bar{f}}x_h)$$
$$\hat{\hat{Y}}'_{Y'(s,1,h)} = \hat{\hat{Y}}'_{Y'(h)} + K'_{Y'(1,1,h)}(\bar{f}x'_h - \hat{\bar{f}}x'_h)$$

with

$$K_{Y'(1,1,h)} = \frac{\bar{Y'}_{Y'(h)} V'_{Y'(11,h)}}{\bar{f}x_h V'_{Y'(02,h)}}, \qquad K'_{Y'(1,1,h)} = \frac{\bar{Y'}_{Y'(2,h)} V'_{Y'(11,h)}}{\bar{f}x'_2 V'_{Y'(02,h)}}$$

are recognized constants.The following is the variance of $\hat{\hat{Y}}_{Y'(s,1)}$.

$$\begin{aligned} \mathrm{Var}(\hat{\hat{Y}}_{Y'(s,1)}) &= \sum_{h=1}^{L} W_{Y'(h)} \bar{Y}^2_{Y'(h)} V_{Y'(20,h)} \left(1 - \frac{V^2_{Y'(11,h)}}{V_{Y'(20,h)} V_{Y'(02,h)}}\right) \\ &= \sum_{h=1}^{L} \psi S^2_{Y' fx,h}(1 - \rho^2_{Y' fx,h}) \end{aligned} \tag{3.10}$$

where $Y_{Y'}$ on fx for the hth stratum is correlated with a coefficient of $\rho^2_{Y' fx,h}$.

Using only data on $Y_{Y'}$ under STRS, B/Ub estimators of the FPV are provided by

$$b_p = \gamma_{Y'}\left(\hat{\hat{Y}}'_{Y'(2,st)} - \hat{\hat{Y}}^2_{Y'(st)}\right) \tag{3.11}$$

and

$$b_p^* = \gamma_{Y'}\left(\hat{\hat{Y}}'_{Y'(2,st} - \hat{\hat{Y}}^2_{Y'(st)}\right) + \sum_{h=1}^{L} \psi_h \hat{S}^2_{Y'(h)} \tag{3.12}$$

In the same way, $\hat{S}^2_{y'(h)}$ is a traditional, Ub estimator of $S^2_{y'(h)}$.
The expectation of b_p mathematically is

$$\begin{aligned} \mathrm{E}(b_p) &= \gamma_{Y'}\left[\mathrm{E}\left(\hat{\hat{Y}}_{Y'(2,st)}\right) - \mathrm{E}\left(\hat{\hat{Y}}^2_{Y'(st)}\right)\right] \\ &= \gamma_{Y'}\left[\bar{Y'}_{Y'(2)} - \bar{Y}^2_{Y'} - \mathrm{V}(\bar{Y}_{Y'(st)})\right] \\ &= S^2_{Y'} - \gamma_{Y'} \sum_{h=1}^{L} \psi_h S^2_{Y'(h)} \end{aligned} \tag{3.13}$$

This indicates that b_p is a biased estimator of $S^2_{Y'}$ since it underestimates $S^2_{Y'}$. A reasonable estimate for $S^2_{Y'}$ can be obtained by

$$\begin{aligned} \mathrm{E}(b_p^*) &= \mathrm{E}(b_p) + \gamma_{Y'}\mathrm{E}\left(\sum_{h=1}^{L}\psi_h \hat{S}^2_{Y'(h)}\right) \\ &= \mathrm{E}(b_p) + \gamma_{Y'}\sum_{h=1}^{L}\psi_h S^2_{Y'(h)} = S^2_{Y'} \end{aligned} \tag{3.14}$$

Using the data on $(Y_{Y'}, fx)$ under STRS, B and Ub estimators of the FPV are provided by

$$b_{p,1} = \gamma_{Y'}\left(\hat{\bar{Y}}'_{Y'(c,1)} - \hat{\bar{Y}}^2_{Y'(c,1)}\right) \tag{3.15}$$

and

$$b^*_{p,1} = b_{p,1} + \gamma_{Y'}\left(\sum_{h=1}^{L}\psi_h \hat{S}^2_{Y',h}\right)\left(1 - \rho^2_{Y'fx,st}\right) \tag{3.16}$$

Similarly, an old-fashioned, Ub estimator of $S^2_{Y'(h)}$ is $\hat{S}^2_{Y'(h)}$.
The expectation of $b_{p,1}$ mathematically is

$$\begin{aligned} \mathrm{E}(b_{p,1}) &= \gamma_{Y'}\left[\mathrm{E}\left(\hat{\bar{Y}}'_{Y'(c,1)}\right) - \mathrm{E}\left(\hat{\bar{Y}}^2_{Y'(c,1)}\right)\right] \\ &= \gamma_{Y'}\left[\bar{Y}'_{Y'(2)} - \bar{Y}^2_{Y'} - \mathrm{V}(\hat{\bar{Y}}_{Y'(c,1)})\right] \\ &= S^2_{Y'} - \gamma_{Y'}\left(\sum_{h=1}^{L}\psi_h S^2_{Y',h}\right)\left(1 - \rho^2_{Y'fx,st}\right) \end{aligned} \tag{3.17}$$

This also indicates that $b_{p,1}$ is a biased estimator of $S^2_{Y'}$ since it underestimates $S^2_{Y'}$. A reasonable estimate for $S^2_{Y'}$ can be obtained by

$$\begin{aligned} \mathrm{E}(b^*_{p,1}) &= \mathrm{E}(b_{p,1}) + \gamma_{Y'}\mathrm{E}\left(\sum_{h=1}^{L}\psi_h \hat{S}^2_{Y',h}\right)\left(1 - \rho^2_{Y'fx,st}\right) \\ &= \mathrm{E}(b_{p,1}) + \gamma_{Y'}\left(\sum_{h=1}^{L}\psi_h S^2_{Y',h}\right)\left(1 - \rho^2_{Y'fx,st}\right) = S^2_{Y'} \end{aligned} \tag{3.18}$$

Using the data on $(Y_{Y'}, fx)$, biassed combined/separate estimators of the FPV under STRS are provided by

$$b_{p,2} = \gamma_{Y'}\left(\hat{\bar{Y}}'_{Y'(c,1)} - \hat{\bar{Y}}^2_{Y'(st)}\right) \tag{3.19}$$

$$b_{p,3} = \gamma_{Y'}\left(\hat{\bar{Y}}'_{Y'(s,1)} - \hat{\bar{Y}}^2_{Y'(s,1)}\right) \tag{3.20}$$

$$b_{p,4} = \gamma_{Y'}\left(\hat{\bar{Y}}'_{Y'(s,1)} - \hat{\bar{Y}}^2_{Y'(st)}\right) \tag{3.21}$$

The expectation of $b_{p,2}$ mathematically is

$$\begin{aligned} \mathrm{E}(b_{p,2}) &= \gamma_{Y'}\left[\mathrm{E}\left(\hat{\bar{Y}}'_{Y'(c,1)}\right) - \mathrm{E}\left(\hat{\bar{Y}}^2_{Y'(st)}\right)\right] \\ &= \gamma_{Y'}\left[\bar{Y'}_{Y'(2)} - \bar{Y}^2_{Y'} - \mathrm{V}(\hat{\bar{Y}}_{Y'(st)})\right] \\ &= S^2_{Y'} - \gamma_{Y'}\left(\sum_{h=1}^{L}\psi_h S^2_{Y',h}\right) \end{aligned} \tag{3.22}$$

This also indicates that $b_{p,2}$ is a biassed estimator of $S^2_{Y'}$ since it underestimates $S^2_{Y'}$. A reasonable estimate for $S^2_{Y'}$ can be obtained by

$$\begin{aligned} \mathrm{E}(b^*_{p,2}) &= \mathrm{E}(b_{p,2}) + \gamma_{Y'}\mathrm{E}\left(\sum_{h=1}^{L}\psi_h \hat{S}^2_{Y',h}\right) \\ &= \mathrm{E}(b_{p,2}) + \gamma_{Y'}\left(\sum_{h=1}^{L}\psi_h S^2_{Y',h}\right) = S^2_{Y'} \end{aligned} \tag{3.23}$$

Given the data on $(Y_{Y'}, fx)$, unbiassed combined/separate estimators of the FPV under STRS are provided by

$$b^*_{p,3} = b_{p,3} + \gamma_{Y'}\left(\sum_{h=1}^{L}\psi_h \hat{S}^2_{Y',h}\right)\left(1 - \rho^2_{Y'fx,h}\right) \tag{3.24}$$

$$b^*_{p,4} = b_{p,4} + \gamma_{Y'}\left(\sum_{h=1}^{L}\psi_h \hat{S}^2_{Y',h}\right) \tag{3.25}$$

The proof of the provided estimators is alike to the case of biased combined and separate estimators.

3.3 Simulation and Real Data

In this section, actual populations are examined, and the suggested estimators' values in numbers for both ABs and REs are computed.

3.3.1 Population

The dataset, sourced from Kadilar and Cingi (2006), focuses on apple production, measured in 100 tonnes, denoted as $Y_{Y'}$, Using the auxiliary variable fx representing the quantity of apple trees, expressed in terms of 100 trees. For the duration of 1999, it includes 854 villages spread over six regions: Marmarian, Agean, Mediterranean, Black Sea, Central Anatolia, and East and Southeast Anatolia. The Neyman allocation method is used to assign numbers of samples to each of these areas. Table 3.1 provides a summary of the dataset's statistics.

The biass and mean squared errore (MSE) of the proposed estimators are assessed over ten thousand iterations (denoted as = 10000 per simulation). Absolute biases (ABs), MSEs, and relative efficiencies (REs) of the proposed variance estimators within the framework of STRS are computed, taking into account measurement error. Formulas for computing ABs, MSEs, and REs are provided as follows:

$$AB(b_i) = \frac{1}{r}\sum_{i=1}^{r}|b_i - S_Y^2| \tag{3.26}$$

$$\text{MSE}(b_i) = \frac{1}{r}\sum_{i=1}^{r}(b_i - S_Y^2)^2 \tag{3.27}$$

$$\text{RE}(b, b_p) = \frac{\text{MSE}(b)}{\text{MSE}(b_p)} \tag{3.28}$$

In the context of the given variance estimator under Stratified Random Sampling (STRS), denoted as b, the numerical outcomes are displayed in Tables 2.4. These tables utilize information on just one auxiliary variable (either fx). Additionally, when discussing the existing estimators, they assume $b = d = 0$ (bh, dh) and $a = c = 1$ $(ah = ch = 1)$ for the purpose of simplicity.

Conversely, when incorporating information from the auxiliary variable fx, all proposed estimators (b_p, b_p^* for $i = 1, 2, 3, 4$) surpass the precision of existing estimators ($\hat{S}^2_{Y',st}$ and t_i for $i = 1, 2, 3$). Cases highlighted in bold in Table 2.4 emphasize this observation. As expected, Amongst the estimators that are suggested, the Ub ones ($b^*_{p,i}$ for $i = 1, 2, 3, 4$) exhibit fewer Absolute Biases compared to their biased counterparts ($b_{p,i}$ for $i = 1, 2, 3, 4$), with the Relative Efficiencies of the former being almost equivalent to those of the latter.

Table 3.1: Descriptive statistics for actual Population

	1	2	3	4	5	6
N_h	106	106	94	171	204	173
n_h	11	20	47	81	8	3
$\bar{Y}_h$	1536.77	2212.59	9384.31	5588.01	966.956	404.399
$\bar{X}_h$	24375.6	27421.7	72409.9	74364.7	26441.7	9843.83
$\bar{Y}'_{2,h}$	4.3254×10^7	1.37075×10^8	9.73007×10^8	8.46874×10^8	6.61801×10^6	1.05281×10^6
$\bar{X}'_{2,h}$	2.99091×10^9	4.02252×10^9	3.08112×10^{10}	8.66222×10^{10}	2.75047×10^9	4.48072×10^8
$S_{Y,h}$	6425.09	11551.5	29907.5	28643.4	2389.77	945.749
$S_{X,h}$	49189.1	57460.6	160757	285603	45402.8	18794
$S_{Y^2,h}$	3.82606×10^8	1.34521×10^9	4.97464×10^9	8.52128×10^9	4.60968×10^7	5.51211×10^6
$S_{X^2,h}$	1.45741×10^{10}	2.21695×10^{10}	1.53676×10^{11}	8.39145×10^{11}	1.31575×10^{10}	2.23427×10^9
$\rho_{YX,h}$	0.815641	0.855999	0.90112	0.985876	0.713099	0.893599
$\rho_{Y^2X^2,h}$	0.716031	0.967265	0.792559	0.996738	0.534591	0.78124

Table 3.2: A comparison between the REs and ABs for the proposed and current variance estimators in the context of STRS, utilizing simulated population and information derived from auxiliary variable.

Existing	Simulated		Proposed	Simulated	
Estimators	AB	RE	Estimators	AB*	RE
$\hat{S}^2_{Y,\mathrm{st}}$	0.0066	1.000	$\hat{S}^2_{Y',\mathrm{st}}$	0.0060	1.100
t_p	0.0008	1.010	b_p	0.0008	5.2728
t_p^*	0.0001	1.010	b_p^*	0.0002	5.2759
t_1	0.0021	2.188	b_1	2.6804	0.8774
$t_{p,1}$	0.0002	3.791	$b_{p,1}$	2.6733	0.09464
$t_{p,1}^*$	0.0000	3.791	$b_{p,1}^*$	2.6740	0.09461
$t_{p,2}$	0.0005	0.305	$b_{p,2}$	2.6696	0.09446
$t_{p,2}^*$	0.0001	0.305	$b_{p,2}^*$	2.6702	0.0944
t_2	2.8519	0.001	b_2	0.0573	1.6765
t_3	2.7441	0.001	b_3	0.0414	1.8190
$t_{p,3}$	0.0002	5.207	$b_{p,3}$	0.0665	1.8073
$t_{p,3}^*$	0.0000	5.206	$b_{p,3}^*$	0.0659	1.8079
$t_{p,4}$	0.0005	0.326	$b_{p,4}$	0.0305	5.1993
$t_{p,4}^*$	0.0001	0.326	$b_{p,4}^*$	0.0299	5.2032

Table 3.3: With population 1 from an auxiliary variable, real population data are used to analyse the ABs and REs values of the proposed and existing variance estimators under STRS.

Existing	Data		Proposed	Data	
Estimators	AB	RE	Estimators	AB*	RE
$\hat{S}^2_{Y,\mathrm{st}}$	37633.900	1.000	$\hat{S}^2_{Y',\mathrm{st}}$	29789	1.2633
t_p	71.307	11.767	b_p	63.8905	153.0285
t^*_p	19.581	11.732	b^*_p	46.9726	153.0549
t_1	6043.980	11.423	b_1	0.8779	13.2907
$t_{p,1}$	8.687	93.640	$b_{p,1}$	6.6761	173.7761
$t^*_{p,1}$	8.601	93.478	$b^*_{p,1}$	5.8643	174.6355
$t_{p,2}$	43.196	94.771	$b_{p,2}$	33.3556	84.4209
$t^*_{p,2}$	8.531	94.331	$b^*_{p,2}$	4.9800	84.5127
t_2	2396.900	53.653	b_2	2.4887	1.9903
t_3	9265.290	6.621	b_3	2.0885	2.8263
$t_{p,3}$	3.110	188.932	$b_{p,3}$	56.1821	176.2353
$t^*_{p,3}$	5.850	188.744	$b^*_{p,3}$	56.0531	177.0221
$t_{p,4}$	45.490	188.502	$b_{p,4}$	36.0000	87.1095
$t^*_{p,4}$	6.236	188.038	$b^*_{p,4}$	34.0000	87.3153

Table 3.4: Descriptive statistics for actual Population 2

	1	2	3	4	5	6
N_h	48	35	68	39	45	49
n_h	20	15	25	10	18	17
W_h	0.169	0.1232	0.2394	0.1373	0.1584	0.1725
λ_h	0.0291	0.0381	0.0252	0.0743	0.0333	0.0384
$\bar{Y}_h$	49.5833	24.1428	33.1176	20.2564	20.8444	23.1428
$\bar{X}_h$	24375.6	176.3714	314.3823	145.3846	163.667	179.102
$\bar{R}_{x_h}$	24.5	18	34.5	20	23	25
$S_{Y,h}$	94.5859	20.5471	57.4691	19.5592	23.313	945.749
$S_{X,h}$	902.6175	174.0182	896.1789	153.0143	199.8747	202.6107
$S_{R_{x_h}}$	13.5218	10.1691	17.7333	11.2921	13.0148	14.1888
$\rho_{Y_h X_h}$	0.9986	0.9735	0.9836	0.9962	0.9979	0.9957
$\rho_{Y_h R_{x_h}}$	-0.0875	0.0081	0.0705	-0.1387	-0.0635	-0.2046
$\rho_{X_h R_{x_h}}$	-0.0959	0.0678	0.0915	-0.1068	-0.0643	-0.1760

Table 3.5: With population 2 from an auxiliary variable, real population data are used to analyse the ABs and REs values of the proposed and existing variance estimators under STRS.

Existing	Data		Proposed	Data	
Estimators	AB	RE	Estimators	AB*	RE
$\hat{S}^2_{Y,\mathrm{st}}$	40923.739	1.000	$\hat{S}^2_{Y',\mathrm{st}}$	39657.160	1.032
t_p	89.307	8.845	b_p	43.564	167.194
t_p^*	35.637	34.745	b_p^*	29.640	169.269
t_1	4032.675	21.604	b_1	3.648	9.064
$t_{p,1}$	18.047	57.374	$b_{p,1}$	13.937	138.428
$t_{p,1}^*$	16.453	55.953	$b_{p,1}^*$	11.749	139.633
$t_{p,2}$	51.578	69.093	$b_{p,2}$	61.628	57.837
$t_{p,2}^*$	13.492	68.792	$b_{p,2}^*$	8.592	56.973
t_2	3952.547	42.683	b_2	8.278	1.368
t_3	18224.833	7.273	b_3	3.793	2.021
$t_{p,3}$	9.264	153.932	$b_{p,3}$	67.427	153.763
$t_{p,3}^*$	12.538	152.734	$b_{p,3}^*$	70.482	153.104
$t_{p,4}$	63.794	124.773	$b_{p,4}$	48.106	58.798
$t_{p,4}^*$	3.741	207.630	$b_{p,4}^*$	60.492	57.810

Chapter 4

Conclusion and future works

4.1 Conclusion

In survey sampling, we need accurate and precise estimates of the parameters of a finite population. To achieve this, we often use extra information during the estimation process.This chapter has demonstrated how to use the first and second raw moments of the variables under study to compute the FPV under STRS with measurement error. We have also developed biased and unbiased estimators for the FPV under STRS, taking into account measurement errors. Based on data regarding the variable under study and another variable, these estimators employ simple and combined/separate (difference) estimators for the first and second raw moments.

An empirical study was undertaken to evaluate the performance of the proposed estimators in the presence of measurement errors. The findings indicated that the separate class of estimators demonstrates superior precision compared to the combined class of estimators. In conclusion, it is advisable to use the estimators b_p,b_p^*, $(b_2, b_3, b_{p,3}, b_{p,3}^* and b_{p,4}, b_{p,4}^*)$, when information is available on $(Y_{Y'})$ $(Y_{Y'}, fx)$.

The Absolute Biases (ABs) and Relative Efficiencies (REs) of both the existing and proposed variance estimators are detailed in Table 3.2 for the simulated populations and Table 3.3 and 3.5 for the real population 1 and population 2, respectively. Analysis of Table 3.2 indicates that, in general, the proposed variance estimators (b_p, b_p^*, b_2, b_3, $b_{p,3}$, $b_{p,3}^*$, $b_{p,4}^*$, $b_{p,4}^*$), whether simple or combined/separate, exhibit greater precision compared to their existing counterparts. While the other proposed estimators are not as efficient as the existing estimators (b_1, $b_{p,1}$, $b_{p,1}^*$, $b_{p,2}$, $b_{p,2}^*$), which target the estimation of the population's first raw moment through $\bar{Y}_{st}$, they demonstrate lower efficiency than both the existing (b_2, b_3) and proposed (b_p, b_p^*, $b_{p,3}$, $b_{p,3}^*$, $b_{p,4}$, $b_{p,4}^*$) estimators.

Overall, the combined estimators (b_p, b_p^*) demonstrate greater precision than the separate estimators $(b_{p,i}, b_{p,i}^*$ for $i = 3, 4)$.

4.2 Future works

- The current study has the potential for extension to encompass other probability sampling techniques, such as systematic sampling and cluster sampling.

References

Cekim, H. O. and Kadilar, C. (2020a). In-type estimators for the population variance in stratified random sampling. *Communications in Statistics-Simulation and Computation*, 49(7):1665–1677.

Cekim, H. O. and Kadilar, C. (2020b). Ln-type variance estimators in simple random sampling. *Pakistan Journal of Statistics and Operation Research*, pages 689–696.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). Sample survey methods and theory. vol. i. methods and applications.

Haq, A., Usman, M., and Khan, M. (2023). Estimation of finite population variance under stratified random sampling. *Communications in Statistics-Simulation and Computation*, 52(12):6193–6209.

Kadilar, C. and Cingi, H. (2006). Improvement in variance estimation using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, 35(1):111–115.

Kadilar, C. and Cingi, H. (2006). Ratio estimators for the population variance in simple and stratified random sampling. *Applied Mathematics and Computation*, 173(2):1047–1059.

Mahanty, B. and Mishra, G. (2020). An unbiased estimator of finite population mean using auxiliary information. *Journal of Statistical Theory and Applications*, 19(4):534–539.

Özel, G., Çingi, H., and Oğuz, M. (2014). Separate ratio estimators for the population variance in stratified random sampling. *Communications in Statistics-Theory and Methods*, 43(22):4766–4779.

Qian, J. (2020). Variance estimation with complex data and finite population correction—a paradigm for comparing jackknife and formula-based methods for variance estimation. *ETS Research Report Series*, 2020(1):1–16.

Sampling Techniques (1977). Sampling techniques. *New York*, pages 423–457.

Shahzad, U., Ahmad, I., Almanjahie, I. M., Al-Noor, N. H., and Hanif, M. (2023). A novel family of variance estimators based on l-moments and calibration approach under stratified random sampling. *Communications in Statistics-Simulation and Computation*, 52(8):3782–3795.

Sidelel, E. B., Orwa, G. O., and Otieno, R. O. (2014). Variance estimation in stratified random sampling in the presence of two auxiliary random variables.

Singh, G., Bhattacharyya, D., and Bandyopadhyay, A. (2021). Calibration estimation of population variance under stratified successive sampling in presence of random non response. *Communications in Statistics-Theory and Methods*, 50(19):4487–4509.

Singh, H. P. and Solanki, R. S. (2013). A new procedure for variance estimation in simple random sampling using auxiliary information. *Statistical papers*, 54:479–497.