A SUITABLE USE OF AUXILIARY INFORMATION FOR ESTIMATION OF FINITE POPULATION MEAN

By

Mehreen Shah Nawaz

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2023

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

Auxiliary Information Based Estimators of Finite Population Mean Using Transformation

By

Mehreen Shah Nawaz

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

2023

Declaration

I "Mehreen Shah Nawaz" thus sincerely proclaim that this thesis, "Auxiliary Information Based Estimators of Finite Population Mean using Transformation", is true.

- This entire project was completed as part of my application to this university for an M.Phil. in statistics.
- Where I used someone else's published work to assist me, I always make it very apparent.
- I always cite the source when I quote from someone else's writing. Except for these quotations, all of the research for this thesis was done by myself.
- Where the thesis is based on work I did with my supervisor and myself, I have made it clear exactly what was done by others and what I have advised.

Dated:________________________ Signature:________________________

Dedication

I dedicated this effort of mine to my Parents for raising me to believe that anything is possible and to my self to make everything possible along with all the respected teachers and my sisters and brothers for their love, encouragement, support and affection that make me able to get this success.

Contents

Abstract

Survey statisticians are attempting to estimate the unknown population parameters efficiently. Utilizing the auxiliary information in a meaningful way allows for the efficient estimation of the finite population mean. We have proposed some modified ratio, product, and regression type estimators for the finite population mean under simple and stratified random sampling designs using transformation. To obtain even more accurate estimates, the study variable is transformed. Similar to this, a known auxiliary variable is a source for increasing the precision of estimators. In transformation, we applied the antithetic variable technique. The antithetic variable method is a variance reduction method. The method seeks to increase the precision and effectiveness of the estimators by reducing the variance of the estimates. Up to the first level of approximation, expressions for the bias and mean squared errors of the proposed estimators are obtained. The suggested estimators are compared numerically and theoretically. The performance of the suggested estimators is numerically illustrated using actual data sets to validate theoretical conclusions. The comparison showed that the proposed estimators outperformed the usual ratio, product and regression estimators.

Chapter 1

Introduction

1.1 History

Sampling plays an important role in fields such as data analysis, quality control, epidemiology, and environmental monitoring. Social scientists and statisticians created techniques for choosing representative samples from specific groups around the middle of the 20th century. Researchers were able to draw conclusions about bigger populations from smaller samples using the techniques of randomization and stratification to assure reliable and unbiased results.To tackle complicated issues and datasets, more advanced sampling techniques including stratified sampling, cluster sampling, and adaptive sampling have been developed. Sampling became more popular in relation to national censuses. As populations swelled in the late 19th century, it became impracticable to carry out an exhaustive count of every person. In order to precisely estimate population characteristics, statisticians began to use sampling procedures. The history of sampling generally covers a wide range of fields and epochs. The essential instrument for researching and analysing populations is sampling.

1.2 Sampling Designs

Sampling designs are the methods and procedures used to choose a sample from a larger population for study or research purposes. These approaches offer a framework for selecting a sample that is representative and capable of producing accurate and trustworthy results. A few typical sampling techniques are briefly discussed as follows.

The use of probability sampling techniques guarantees that each person or unit of the population has a known, non-zero chance of being chosen for the sample. To achieve representativeness, these strategies rely on random selection and statistical concepts. The probability sampling strategies include simple random sampling, stratified sampling, and cluster sampling.

Non-probability sampling approaches do not employ random selection and do not guarantee that every member of the population has an equal chance of being considered. These methods are typically applied when probability sampling is impractical. Non-probability sampling methods include chain-referral sampling, convenience sampling, and selective sampling.

Cluster sampling entails grouping the population into clusters or groups, typically according to the proximity to one another. All persons or items found in the designated clusters are included in the sample, which is drawn at random from among the clusters. When the population is spread out geographically, cluster sampling is frequently more practicable and economical.

Selecting people or components from a community at regular intervals is known as systematic sampling. To determine the sampling interval, divide the population size by the appropriate sample size. The first person or thing is chosen at random, and then choices after that are determined by adding the sample interval to the serial number of the first selected unit.

When it is difficult to discover or identify the population of interest, snowball sampling is utilized. Using non-probability techniques, the initial participants, sometimes known as "seeds," are chosen. Then, participants are requested to suggest individuals who fit the study's inclusion criteria. Referrals are used in this sampling technique to increase the sample size.

With purposive sampling, persons or components are specifically chosen according to predetermined traits or criteria that are in line with the goals of the study. When selecting participants, researchers utilize their best judgment to select those they believe to be most pertinent or informed about the subject at hand. Although subjective approach are unlikely to provide a representative sample, purposive sampling can be helpful in qualitative research or when examining uncommon or unusual groups.

1.3 Transformation: The Antithetic Variable Technique

The process of using mathematical or statistical techniques to alter the original data in order to achieve certain objectives or satisfy particular assumptions is referred to as data transformation. In order to improve data distribution, lessen skewness, normalize variables, stabilize variance, or enhance linear correlations between variables, data transformation may be advantageous. In transformation, we used the antithetic variable technique which is a popular strategy for reducing variance in Monte Carlo simulations. The method seeks to increase the precision and effectiveness of the estimators by reducing the variance of the estimates. In the antithetic variable technique, two sets of random variables are used, the original set and the antithetic set. The original set is transformed while still maintaining the correlation structure between the variables to produce the antithetic set. Usually, this transformation entails flipping the values of the random variables or changing their sign. This information is used by the researcher to create efficient estimators.

1.4 Uses of Auxiliary Information

Auxiliary variables are additional variables that offer more details about the population of interest in statistical analysis and survey sampling. Although not the primary variables of interest, these variables may be somehow connected to the study variables. The design of a study and the data analysis may be impacted by the relationship between study variables and auxiliary variables. Depending on the specific research context and the study's objectives, there may be a difference in the relationship between the study (Y) and the auxiliary (X) variables. By adding more information and boosting the precision of estimates the design and analysis of a research can be enhanced by the inclusion of auxiliary variables. Based on their applicability to the study and potential to raise the standard of the analysis, researchers should carefully consider the choice and inclusion of auxiliary variables. In general, auxiliary data in sampling is very important for increasing the effectiveness, accuracy, and representativeness of survey estimates. Researchers can choose wisely when it comes to sampling design, sample unit allocation, and estimation adjustments by including additional variables related to the population of interest. The study's results are more accurate and reliable when supplementary data are included.

In stratified random sampling, auxiliary data are frequently used to define meaningful strata or subgroups. Researchers can make sure that the sample accurately represents each strata and enable more accurate estimates within each subgroup by stratifying the population based on auxiliary variables. Researchers can examine differences in relationships or patterns across various subgroups using subgroup analysis, which gives them a deeper understanding of the data.

Researchers can improve analysis, deal with outliers, standardize variables, and increase accuracy by using auxiliary information and its transformation in survey sampling. Researcher judgments about variable transformations can be well-informed by taking into account the correlation between the auxiliary variable (X) and the variable of interest (Y).

Auxiliary information has traditionally been used in survey sampling. Auxiliary information in sample selection with varying probabilities was suggested by Hansen and Hurwitz (1943). An assortment of estimators exploiting auxiliary information in stratified random sampling was proposed by Koyuncu and Kadilar (2009a). To improve the ratio estimators, several researchers have suggested many estimators in simple random sampling employing auxiliary variables (?). Moreover, Kadilar and Cingi (2003), Shabbir and Gupta (2005), Kadilar and Cingi (2005), Koyuncu and Kadilar (2009a), Koyuncu and Kadilar (2009b) adjusted ratio estimators in stratified random sampling to increase their efficiency.

1.5 Simple Random Sampling

Simple random sampling is a fundamental sampling technique that is commonly used in statistics and research. It comprises randomly choosing units from a population in a way that ensures each individual has an equal probability of being chosen for the sample. An accurate representation of the entire population is the goal of simple random sampling. By ensuring that every person of the population has an equal chance of being picked for the sample, simple random sampling lowers selection bias and increases the generalizability of the findings. Simple random sampling relies on the chance to choose the subjects or things. In light of this, each member of the population has an equal probability of getting selected.Simple random sampling can be used to draw a representative sample from a population. It is especially helpful when there are no recognized patterns or subgroups of interest and the population is fairly homogeneous. It is crucial to keep in mind that simple random sampling could not be practicable if the population is sizable or when the cost and time involved in sampling each element of the population are significant. In some cases, alternative sampling methods like stratified sampling or cluster sampling may be more appropriate.

1.6 Stratified Random Sampling

The stratified random sampling technique splits a population into homogenous strata based on predetermined criteria, and then a random sample is taken from each stratum.Stratified random sampling seeks to ensure that the sample is representative of the population by capturing the variability prevalent among diverse subgroups. The first step in stratified random sampling is to divide the population into strata that are exhaustive and mutually exclusive. Strata are developed depending on pertinent traits or factors that the study is interested in. Stratified random sampling has a number of benefits over simple random sampling. As the sample size is concentrated on regions with higher variability, it enables more accurate estimation within subgroups or strata. Additionally, stratification makes sure that different groups are fairly represented in the population, which is helpful when examining and contrasting traits specific to particular subgroups or when drawing conclusions about various strata separately. However, accurate knowledge of the population and the pertinent stratification variables are necessary for stratified random sampling. In comparison to simple random sampling, it also requires more preparation and resources. Despite these difficulties, stratified random sampling is a useful method for obtaining a representative and effective sample when the population is heterogeneous or when there are particular subgroups that are relevant to the analysis.

1.7 Motivation of the Study

In simulation studies, antithetic variables are used for efficient estimation. In survey sampling, the prime objective is to estimate the population mean efficiently using auxiliary variables. Here, the goal is to transform the study variable into the auxiliary variable in order to estimate the population mean in an accurate and efficient manner.

1.8 Thesis outline

The thesis' remaining sections are organised as follows. The thesis-related literature has been reviewed in Chapter 2. In Chapter 3, an estimator for calculating the finite population mean using simple random sampling is created and is founded on the utilisation of supplementary information received through transformation. The concept is extended to stratified random sampling in Chapter 4. Chapter 5 puts an end to the thesis.

Chapter 2

Literature Review

Efficient estimation of the population mean is a fundamental problem in statistics, and the literature has accorded it a lot of attention. To solve this issue, researchers have suggested several estimation techniques. There are numerous statistical methods and procedures covered in the literature to determine the population mean. To increase the population mean estimator's accuracy and effectiveness, researchers have looked into a variety of sampling strategies, estimating techniques, and statistical models. The selection of an estimation method is affected by a multitude of factors, the sampling design, the data at hand, and the required level of accuracy for the estimations.

This study aims to calculate the finite population mean with the aid of auxiliary data and transformation. As many writers have contributed to this with the intention of improving the estimators of the population parameters. The work of a few researcher is presented here in the following paragraphs.

Laplace (1820) was the first to estimate the population using additional data. It is acknowledged that the Neyman (1938) work is a draught that improves an estimate by using more data. In order to increase the effectiveness of estimators, several survey statisticians have since tried to determine the mean of the limited population and other population characteristics using auxiliary data. In addition to being the first to employ an auxiliary variable during the estimation phase, Cochran (1940) also developed the ratio estimator for the population mean. When there is a significant positive correlation between the study variable and an auxiliary variable and the regression line crosses the origin, the ratio type estimators outperform the basic mean per unit estimator. The usual Ratio Estimator, its bias and Mean Square Error (MSE) are respectively given by

$$\hat{\hat{Y}}_R = \left(\frac{\bar{y}}{\bar{x}}\bar{X}\right)$$

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

and

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

where $C_y = (S_y^2/\bar{Y}\ )$ and $C_x = (S_x^2/\bar{X}\ )$ are the population coefficient of variation.

Robson (1957) presented the product estimator for the negative correlation. It is commonly known that when the study variable (Y) and an auxiliary variable (x) have a strong negative correlation, the product type mean estimator performs better than the sample mean per unit estimator. The usual product estimator is

$$\hat{\bar{Y}}_p = \bar{y}\frac{\bar{x}}{\bar{X}}. \tag{2.1}$$

The respective bias and MSE of $\hat{\bar{Y}}_p$ are

$$Bias(\hat{\bar{Y}}_P) = \lambda\bar{Y}(\rho_{yx} \cdot Cx \cdot Cy)$$

and

$$MSE(\hat{\bar{Y}}_P) = (\lambda\bar{Y}^2[C_y^2 + C_x^2 + 2\rho_{yx} \cdot Cx \cdot Cy]$$

The simple linear regression estimator was defined by Hansen et al. (1953) and is defined as

$$\hat{\bar{Y}}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) \tag{2.2}$$

and its variance is

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

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

Survey statisticians have been trying to get precise estimate of the population parameters. However, if the selected sample contains extreme values, the variance of the mean per unit estimator will be exaggerated. Based on the inclusion of the lowest and largest values in the sample, the sample mean per unit estimator is reduced and increased by a constant.Särndal (1972) proposed a mean per unit estimator to address the problem of extreme values in the chosen sample.The following estimator was proposed by the author.

$$\bar{y}_{san} = \begin{cases} \bar{y} + c' & \text{if } y_{min'} \in \omega \text{ and } y_{max'} \notin \omega, \\ \bar{y} - c' & \text{if } y_{max'} \in \omega \text{ and } y_{min'} \notin \omega, \\ \bar{y} & \text{for all other samples} \end{cases} \tag{2.3}$$

where ω denotes the chosen sample and c denotes an adequately chosen constant which needs to be determined. The variance of $\bar{y}_{san}$ is expressed as follows.

$$V(\bar{y}_{san}) = \lambda S_y^2 - \frac{2\lambda nc(R_{y'} - nc)}{N-1}. \tag{2.4}$$

The variance of $\bar{y}_{san}$ is optimum at $c_{opt} = R_y/2n$, where $R_{y'} = y_{max} - y_{min}$. The optimal variance at c_{opt} is expressed as

$$V(\bar{y}_{san}) = V(\bar{y'}) - \frac{\lambda R_{y'}^2}{2(N-1)}, \tag{2.5}$$

where

$$V(\bar{y'}) = \lambda S_y^2. \tag{2.6}$$

$V(\bar{y}_{san})$ is always less than $V(\bar{y})$ at the optimum value of c.

By employing the extreme values existing in the sample, Khan and Shabbir (2013) expanded the concept of Särndal (1972) to include estimators that use auxiliary information to further boost the precision of the estimators. The authors introduced improved ratio, product, and regression estimators for both positive and negative situations.

The authors recommended the ratio and regression estimators as measures of positive correlation

$$\hat{\bar{Y}}_{RC} = \begin{cases} (\bar{y} + c_{1'})\dfrac{\bar{X}}{(\bar{x} + c_{2'})} & \text{if the sample includes } y_{min'} \text{ and } (x_{min'}), \\ (\bar{y} - c_{1'})\dfrac{\bar{X}}{(\bar{x} - c_{2'})} & \text{if the sample includes } y_{max'} \text{ and } (x_{max'}), \\ \bar{y}\dfrac{\bar{X}}{\bar{x}} & \text{for all other pairs,} \end{cases}$$

and

$$\bar{y}_{lrC1} = \bar{y}_{c11} + b(\bar{X} - \bar{x}_{c21}).$$

Whereas for negative correlation, the authors proposed the product and regression type estimators as

$$\hat{\bar{Y}}_{PC} = \begin{cases} (\bar{y} + c_{1'})\dfrac{(\bar{x} - c_{2'})}{\bar{X}} & \text{if sample includes } y_{min'} \text{ and } (x_{max'}), \\ (\bar{y} - c_{1'})\dfrac{(\bar{x} + c_{2'})}{\bar{X}} & \text{if sample includes } y_{min'} \text{ and } (x_{max'}), \\ \bar{y}\dfrac{\bar{x}}{\bar{X}} & \text{for all other pairs,} \end{cases}$$

and

$$\bar{y}_{lrC2} = \bar{y}_{c11} + b(\bar{X} - \bar{x}_{c22})$$

The recommended ratio, product, and regression type estimators' biases and MSE were computed and are listed below up to the first level of approximation

$$Bias(\hat{\bar{Y}}_{RC}) = \frac{\lambda}{\bar{X}}[R_{\prime}(S_x^2 - \frac{2nc_{2'}}{N-1}(x_{max'} - x_{min'} - nc_{2'})) - \{S_{yx} - \frac{n}{N-1} \times (c_{2'}(y_{max'} - y_{min'}) + c_{1'}(x_{max'} - x_{min'}) - 2nc_{1'}c_{2'})\}]$$

And optimum mean square error at optimal values of $c_{1'} = R_{y'}/2n$ and $c_{2'} = R_{x'}/2n$ is

$$MSE(\hat{\hat{Y}}_{RC})_{opt} = M(\bar{y}_R) - \frac{\lambda}{2(N-1)} \times [(y_{max'} - y_{min'}) - R_{\prime}(x_{max'} - x_{min'})]^2 \tag{2.7}$$

where $R_{\prime} = \bar{Y}/\bar{X}$.

$$Bias(\hat{\hat{Y}}_{PC}) = \frac{\lambda}{\bar{X}}[S_{yx} - \frac{n}{N-1} \times (c_{2'}(y_{max'} - y_{min'}) + c_{1'}(x_{max'} - x_{min'}) - 2nc_{1'}c_{2'})\}]$$

$$MSE(\hat{\hat{Y}}_{PC})_{opt} = M(\bar{y}_R) - \frac{\lambda}{2(N-1)} \times [(y_{max'} - y_{min'}) + R_{\prime}(x_{max'} - x_{min'})]^2 \tag{2.8}$$

$$Var(\bar{y}_{lrC})_{opt} = V(\bar{y}_{lr}) - \frac{\lambda}{2(N-1)}[(y_{max'} - y_{min'}) - |\beta|(x_{max'} - x_{min'})]^2 \tag{2.9}$$

The new estimators' superior performance over the existing estimators was demonstrated theoretically and by numerical analysis.

A family of ratio-type estimators of the population mean employing knowledge of the known population constants, i.e., coefficient of variation and coefficient of correlation of the auxiliary variable, were proposed by Mursala et al. (2015) as a further extension of Khan and Shabbir (2013).They suggested the following estimators:

$$\hat{\hat{Y}}_{P1} = \bar{y}_{c1}(\frac{\bar{X} + C_x}{\bar{x}_{c2} + C_x}),$$

$$\hat{\hat{Y}}_{P2} = \bar{y}_{c1}(\frac{\bar{X} + \rho_{yx}}{\bar{x}_{c2} + \rho_{yx}}),$$

$$\hat{\hat{Y}}_{P3} = \bar{y}_{c1}(\frac{\bar{X}C_x + \rho_{yx}}{\bar{x}_{c2}C_x + \rho_{yx}}),$$

and

$$\hat{\hat{Y}}_{P4} = \bar{y}_{c1}(\frac{\bar{X}\rho_{yx} + C_x}{\bar{x}_{c2}\rho_{yx} + C_x})$$

where $\bar{y}_{c1} = \bar{y} + c_1, \bar{x}_{c2} = \bar{x} + c_2$. Unkown constants c_1 and c_2 need to be determined. The recommended estimators' biases and MSEs are

$$Bias(\hat{\hat{Y}}_{Pi}) = \frac{\theta_{\prime} k_{Pi}}{\bar{Y}}\Bigg[k_{Pi}\Big(S_x^2 - \frac{2nc_{2'}}{N-1}(x_{max'} - x_{min'} - nc_{2'})\Big) - S_{yx} + \frac{n}{N-1}\Big(c_{2'}(y_{max'} - y_{min'}) + c_{1'}(x_{max'} - x_{min'}) - 2nc_{1'}c_{2'}\Big)\Bigg]$$

$$MSE(\hat{\hat{Y}}_{Pi}) = \theta_{\prime}[(S_y^2 + k_{Pi}^2 S_x^2 - 2k_{Pi}S_{yx}) - \frac{1}{2(N-1)}\Big((y_{max'} - y_{min'}) - k_{Pi}(x_{max'} - x_{min'})\Big)^2] \tag{2.10}$$

for $i = 1, 2, 3, 4$, where $\theta\prime = (1/n) - (1/N)$, $R\prime = \bar{Y}/\bar{X}$, and

$$k_{P1} = \frac{\bar{Y}}{\bar{X} + C_x}, \quad k_{P2} = \frac{\bar{Y}}{\bar{X} + \rho_{yx}}, \quad k_{P3} = \frac{\bar{Y}C_x}{\bar{X}C_x + \rho_{yx}}, \quad k_{P4} = \frac{\bar{Y}\rho_{yx}}{\bar{X}\rho_{yx} + C_x}.$$

The effectiveness of an estimator can also be improved by the transformation of the auxiliary variable.Using the known lowest and largest values of the auxiliary variable as a starting point, Mohanty and Sahoo (1995) presented two linear transformations and two ratio type estimators based on these two transformation. Biases and MSEs were estimated by the authors up to the first level of approximation. The author's suggested estimators' performance was quantitatively shown.

The effectiveness of estimators in probability proportional to size (PPS) sampling methods can also be affected by extreme data.To address the difficulty provided by the existence of extreme values in the chosen sample,Ahmad and Shabbir (2018) adapted Khan and Shabbir (2013) estimator to the PPS sampling scheme and proposed some ratio, product, and regression estimators. The recommended estimators outperformed the existing estimators that did not take into account extreme values, according to theoretical and numerical data.

A population might be divided into several strata, and unusual observations could occur in each stratum, increasing the mean square error of estimators of the ratio product and regression type estimators.To address this problem, Shoaib et al. (2018) introduced a distinctive family of estimators for estimation of the finite population mean based on the extreme values and fractional raw moments of a complementary variable under stratified random sampling. If stratified random sampling is used, the Särndal (1972) estimator is

$$\bar{y}_{st.c} = \sum_{h=1}^{L} W_h \bar{y}_{hc}$$

where

$$\bar{y}_{hc} = \begin{cases} \bar{y}_h + c_{h'} & \text{if a sample from the ith stratum contains } y_{hmin'} \text{ but not } y_{hmax'}, \\ \bar{y}_h - c_{h'} & \text{if a sample from the ith stratum contains } y_{hmax'} \text{ but not } y_{hmin'}, \\ \bar{y}_h, & \text{in other cases,} \end{cases}$$

where $c_{h'}(h = 1, 2, 3, .., L)$ are exogenous constants, W_h' is the stratum weight, and $\bar{y}_h'$ is the sample mean of the h^{th} stratum. Shoaib et al. (2018) suggested the respective combine ratio, product and regression type estimators under stratified random sampling design as

$$\hat{\bar{Y}}_{RC_1} = \frac{\bar{y}_{st.c_{11}}}{\bar{x}_{st.c_{21}}} \bar{X},$$

$$\hat{\bar{Y}}_{PC_1} = \frac{\bar{y}_{st.c_{12}}}{\bar{X}} \bar{x}_{st.c_{22}},$$

and

$$\hat{\bar{Y}}_{lrC_{12}} = \bar{y}_{st.c_{12}} + b_c(\bar{X} - \bar{x}_{st.c_{22}}).$$

The corresponding biases and MSEs of the combined estimators mentioned above were calculated as

$$Bias(\hat{\bar{Y}}_{RC_1}) \approx \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{\bar{X}} \left[R_{\prime} \{ S_{hx}^2 - \frac{2n_h c_{2h^\prime}}{N_h - 1}(x_{hmax^\prime} - x_{hmin^\prime} - n_{h^\prime} c_{2h^\prime}) \} - \right.$$
$$\{ S_{hyx} - \frac{n_{h^\prime}}{N_h - 1}(c_{1h^\prime}(x_{hmax^\prime} - x_{hmin^\prime}) + c_{2h^\prime}(y_{hmax^\prime} - y_{hmin^\prime})$$
$$\left. -2n_h c_{1h^\prime} c_{2h^\prime}) \} \right],$$

where $R_{\prime} = \bar{Y}/\bar{X}$, and

$$MSE(\hat{\bar{Y}}_{RC_1})_{min} = MSE(\hat{\bar{Y}}_{RC_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{ (y_{hmax^\prime} - y_{hmin^\prime}) - R(x_{hmax^\prime} - _{hmin^\prime}) \}^2, \tag{2.11}$$

where

$$MSE(\hat{\bar{Y}}_{RC_0}) \approx \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R^2 S_{hx}^2 - 2RS_{hyx}). \tag{2.12}$$

$$Bias(\hat{\bar{Y}}_{PC_1}) \approx \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{\bar{X}} [S_{hyx} - \frac{n_{h^\prime}}{N_h - 1}(c_{1h^\prime}(x_{hmax^\prime} - x_{hmin^\prime}) + c_{2h}(y_{hmax^\prime} - y_{hmin^\prime}) - 2n_h c_{1h^\prime} c_{2h^\prime})],$$

$$MSE(\hat{\bar{Y}}_{PC_1})_{min} = MSE(\hat{\bar{Y}})_{PC_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{ (y_{hmax^\prime} - y_{hmin^\prime}) + R(x_{hmax^\prime} - _{hmin^\prime}) \}^2, \tag{2.13}$$

where

$$MSE(\hat{\bar{Y}}_{PC_0}) \approx \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R^2 S_{hx}^2 + 2RS_{hyx}), \tag{2.14}$$

$$Bias(\hat{\bar{Y}}_{lrC_1}) = -cov(\bar{x}_{st.c_{21}}, b_c),$$

$$MSE(\hat{\bar{Y}}_{lrC_1})_{min} = MSE(\hat{\bar{Y}}_{lrC_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{ (y_{hmax^\prime} - y_{hmin^\prime}) - |\beta_c|(x_{hmax^\prime} - x_{hmin^\prime}) \}^2, \tag{2.15}$$

where

$$MSE(\hat{\bar{Y}}_{lrC_0}) \approx \sum W_h^2 \lambda_h S_{hy}^2 (1 - \rho_c^2), \tag{2.16}$$

where

$$\rho_c = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{hyx}}{\sqrt{\sum_{h=1}^{L} \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 \sum_{h=1}^{L} W_h^2 \lambda_h S_{hx}^2}}$$

is used to measure the population correlation coefficient between the study (Y) and the auxiliary variables (X).

The distinct ratio, product, and regression type estimators suggested by Shoaib et al.

(2018) have been given by

$$\hat{\bar{Y}}_{RS_1} = \sum_{h=1}^{L} W_h \frac{\bar{y}_{h'.c_{11'}}}{\bar{x}_{h'.c_{21'}}} \bar{X}_{h'},$$
$$\hat{\bar{Y}}_{PS_1} = \sum_{h=1}^{L} W_h \frac{\bar{y}_{h'.c_{12'}}}{\bar{X}_{h'}} \bar{x}_{h'.c_{22'}},$$
$$\hat{\bar{Y}}_{lrS_{12}} = \sum_{h=1}^{L} W_h \{\bar{y}_{h'.c_{12'}} + b_{h'}(\bar{X} - \bar{x}_{h'.c_{22'}})\}$$

respectively. The individual biases and MSEs of the distinct regression, product, and ratio type estimators are each expressed as follows:

$$Bias(\hat{\bar{Y}}_{RS_1}) \approx \sum_{i=h}^{L} W_h^2 \frac{\lambda_h}{\bar{X}_h}[R'\{S_{hx}^2 - \frac{2n_{h'}c_{h'}}{N_h - 1}(x_{hmax'} - x_{hmin'} - n_{h'}c_{2h})\} - \{S_{hyx} - \frac{n_{h'}}{N_h - 1}(c_{1h'}(x_{hmax'} - x_{hmin'}) + c_{2h'}(y_{hmax'} - y_{hmin'}) - 2n_h c_{1h'} c_{2h'})\}],$$

where $R_{h'} = \bar{Y}_h / \bar{X}_h$,

$$MSE(\hat{\bar{Y}}_{RS_1})min \approx MSE(\hat{\bar{Y}}_{RS_0}) - \sum_{i=h}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{(y_{hmax'} - y_{hmin'}) - R_h(x_{hmax'} - x_{hmin'})\}^2, \quad (2.17)$$

where

$$MSE(\hat{\bar{Y}}_{RS_0}) \approx \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R_h^2 S_{hx}^2 - 2R_h S_{hyx}), \quad (2.18)$$

$$Bias(\hat{\bar{Y}}_{PS_1}) \approx \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{\bar{X}_h}[S_{hyx-} \frac{n_{h'}}{N_h - 1} \times \{c_{1h'}(x_{hmax'} - x_{hmin'}) + c_{2h'}(y_{hmax'} - y_{hmin'}) - 2n_{h'}c_{1h'}c_{2h'}\}],$$

$$MSE(\hat{\bar{Y}}_{PS_1}) \approx MSE(\hat{\bar{Y}}_{PS_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{(y_{hmax'} - y_{hmin'}) + R_h(x_{hmax'} - x_{hmin'})\}^2, \quad (2.19)$$

where

$$MSE(\hat{\bar{Y}}_{PS_0}) \approx \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R_h^2 S_{hx}^2 + 2R_h S_{hyx}), \quad (2.20)$$

$$MSE(\hat{\bar{Y}}_{lrS_1})_{min} \approx MSE(\hat{\bar{Y}}_{lrS_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{(y_{hmax'} - y_{hmin'}) - |\beta_h|(x_{hmax'} - x_{hmin'})\}^2, \quad (2.21)$$

where

$$MSE(\hat{\bar{Y}}_{lrS_0}) \approx \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 (1 - \rho_h^2). \quad (2.22)$$

where the correlation coefficient between the study variable (Y) and the auxiliary variable (x) in the h^{th} stratum is $\rho_h = S_{hyx}/S_{hy}S_{hx}$ The performance of the recommended estimators was theoretically and quantitatively established by the authors using simulated and actual data sets.

Chapter 3

Estimators of Finite Population Mean Using Transformation

In this chapter, a family of distinctive estimators that use the known auxiliary variables under simple random sampling is suggested to estimate the finite population mean. up to the first approximation level, expressions for the bias and mean square error of the current and proposed families of estimators are derived. We next conceptually contrast the proposed family of distinctive estimators with other existing estimators. We used four real data sets to conduct a numerical investigation.

3.1 Notations and Symbols

Let $U' = \{U_{1'}, U_{2'}, ..., U_{N'}\}$ represent a population with N individuals. Let (y_i, x_i) represent the values of the study (y) and auxiliary variable (x) on the ith unit of a finite population, respectively. Let's assume that a simple random sample of size n is taken without replacement from the population U' in order to estimate the population mean. The population characteristics of the auxiliary variable is assumed to be known which include the population mean $\bar{X}$, coefficient of variation C, and coefficient of kurtosis $B_2(x)$.

Let $\bar{Y}=\sum_{i=1}^{N} y_i/N$ be the population mean of study variable y and $\bar{X}=\sum_{i=1}^{N} x_i/N$ be the population mean of auxiliary variable X. Let the respective variances of the study and auxiliary variables be $S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/(N-1)$ and $S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2/(N-1)$. Let $\bar{y}=\sum_{i=1}^{n} y_i/n$ and $\bar{x}=\sum_{i=1}^{n} x_i/n$ be the respective sample means of Y and X, and 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 sample variances of Y and X. Let $\hat{C}_y = s_y^2/\bar{y}$ be the sample coefficient of variations of study variable y. Let $\hat{C}_x = s_x^2/\bar{x}$ be the sample coefficient of variations of auxiliary variable x. Let $C_y = S_y^2/\bar{Y}$ be the population coefficient of variation of study variable Y and let $C_x = S_x^2/\bar{X}$ be the population coefficient of variations of the auxiliary variable X. Moreover, let the sample correlation coefficient be $\widehat{\rho}_{yx}$.

3.2 Proposed estimators

We propose mean per unit, ratio, product and regression type estimators of the population mean that utilize known knowledge of the auxiliary variable under SRSWOR. In order to increase the efficiency of estimator, the suggested estimator also uses auxiliary information. The methodology of the newly suggested estimator is described as follows and is based on the concept of Särndal (1972) and antithetic variable. In accordance with the estimator previously mentioned, an alternate technique to transform the population units is as follows. Let the population of size N be organized according to the magnitude of the auxiliary variable in ascending order as

$$y_1, y_2, \cdots, y_{N-1}, y_n.$$

We add fx_1 to the smallest value of the study variable and subtract the same value from the largest value of the study variable. Then, we add fx_2 to the second smallest value of the study variable and subtract it from the second largest variable. The process is continued for the remaining values of the study variable. Mathematically, the scheme is

$$y_1' = (y_1' = y_1 + fx_1), (y_2' = y_2 + fx_2), \cdots, y_j, \cdots, (y_{N-1}' = y_{N-1} - fx_2), (y_N' = y_N - fx_1).$$

Thus, $(y_1', y_2', y_3' \cdots, y_{N-2}', y_{N-1}', y_N')$ is the transformed population values. Now we define the mean per unit estimator based on the antithetic variable idea as

$$\bar{y}_T = \frac{\sum_{i=1}^{n} y_i'}{n}. \tag{3.1}$$

The variance of the suggested estimation by definition is

$$Var(\bar{y}_T) = E(\bar{y}_T)^2 - [E(\bar{y}_T)]^2. \tag{3.2}$$

First, we establish the expectation of $E(\bar{y}_T)^2$ as

$$E(\bar{y}_T)^2 = E\left(\frac{\sum_{i=1}^{n} y_i''}{n}\right)^2, \tag{3.3}$$

$$= \frac{1}{n^2} E[\sum_{i=1}^{n} {y_i''}^2 + \sum_{i \neq j} y_i'' y_j''], \tag{3.4}$$

$$= \frac{1}{n^2}[nE({y_i''}^2) + n(n-1)E(y_i'' y_j'')]. \tag{3.5}$$

Consider now $E(y_i^{''2})$

$$\begin{aligned}
E(y_i)^{'2} &= \frac{\sum_{i=1}^{N} y_i^{''2}}{N} \\
&= \frac{\sum_{i=1}^{j-1}(y_i'' + fx_i)^2 + y_j''^2 + \sum_{i=j+1}^{n}(y_i'' - fx_i)^2}{N} \\
&= \frac{\sum_{i=1}^{n} y_i''^2 + 2\sum_{i=1}^{j-1} fx_i^2 + 2\sum_{i=1}^{j-1} y_i'' fx_i - 2\sum_{i=j+1}^{N} y_i'' fx_i}{N} \\
&= E(y_i''^2) + 2\frac{\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i}{N}
\end{aligned}$$

Now consider the cross product equation. Using the expression $(\sum_{i=1}^{N} y_i'')^2 = \sum_{i=1}^{N} y_i^{''2} + \sum_{i=1}^{N}\sum_{j=1}^{N} y_i'' y_j''$,we have

$$\begin{aligned}
E(y_i'' y_j'') &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} y_i'' y_j''}{N(N-1)} \\
&= \frac{(\sum_{i=1}^{N} y_i'')^2 - (\sum_{i=1}^{N} y_i^{''2})}{N(N-1)}, \\
&= \frac{N^2\bar{Y}^2 - NE(y^2) - 2(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i)}{N(N-1)}
\end{aligned}$$

Putting the expressions $E(y_i' y_j')$ and $E(y_i^{'2})$ in (3.5) and then simplifying we get

$$\begin{aligned}
E(\bar{y}_T)^2 = \frac{1}{n^2}&\left[n[E(y''^2) + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' f(x_i) - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{N}]\right] + \\
n(n-1)\cdot&\left(\frac{N^2\bar{Y}^2 - NE(y^2) - 2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' f(x_i) - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{N(N-1)}\right) \\
= &\left[\frac{E(y^2)}{n} + \frac{(n-1)N\bar{Y}^2}{n(N-1)} - \frac{(n-1)E(y^2)}{n(N-1)} + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\right. \\
&\left. - \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' f(x_i) - \sum_{i=j+1}^{N} y_i'' fx_i\right)(n-1)}{nN(N-1)}\right]
\end{aligned}$$

Simplifying algebraically, we obtain

$$\begin{aligned}
E(\bar{y}_T)^2 = E(y)^2&\left(\frac{1}{n} - \frac{(n-1)}{n(N-1)}\right) + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\left(1 - \frac{(n-1)}{(N-1)}\right) \\
&+ \frac{(n-1)N\bar{Y}^2}{n(N-1)}
\end{aligned}$$

Note that $E(\bar{y}_T) = E(\bar{y}_i) = \overline{Y}$. The variance of $V(\bar{y}_T)$ now becomes

$$
\begin{aligned}
Var(\bar{y}_T) &= E(\bar{y}_T^2) - (E(\bar{y}_T))^2 \\
&= E(y^2)\left(\frac{1}{n} - \frac{(n-1)}{n(N-1)}\right) + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\left(1 - \frac{(n-1)}{(N-1)}\right) \\
&+ \frac{(n-1)N\bar{Y}^2}{n(N-1)} - (E(\bar{y}_T))^2 \\
&= E(y^2)\left(\frac{1}{n} - \frac{(n-1)}{n(N-1)}\right) + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\left(1 - \frac{(n-1)}{(N-1)}\right) \\
&+ \frac{(n-1)N(E(y^2))}{n(N-1)} - (E(\bar{y}))^2 \\
&= E(y^2)\left(\frac{1}{n} - \frac{(n-1)}{n(N-1)}\right) + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\left(1 - \frac{(n-1)}{(N-1)}\right) \\
&- (E(y))^2\left(1 - \frac{(n-1)N}{(N-1)n}\right) \\
&= E(y^2)\left(\frac{N-n}{n(N-1)}\right) + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\left(\frac{N-n}{N-1}\right) \\
&- (E(y))^2\left(\frac{N-n}{(N-1)n}\right) \\
&= \left(\frac{N-n}{(N-1)n}\right)\left[E(y^2) - (E(y))^2\right] + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\left(\frac{N-n}{N-1}\right) \\
&= \left(\frac{N-n}{(N-1)n}\right)\sigma^2 + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\left(\frac{N-n}{N-1}\right) \\
&= \left(\frac{N-n}{(N-1)n}\right)\left(\frac{N-1}{N}\right)S^2 + \frac{2\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right)}{Nn}\left(\frac{N-n}{N-1}\right)
\end{aligned}
$$

Now the variance of proposed estimator is as follows

$$
Var(\bar{y}_T) = \lambda S_y^2 + \frac{2\lambda}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i'' fx_i - \sum_{i=j+1}^{N} y_i'' fx_i\right) \tag{3.6}
$$

The variance of proposed estimator $\bar{y}_T$ is always less than the variance of $\bar{y}$, i.e., $V(\bar{y}_T) < Var(\bar{y})$.

The ratio estimator is seen to be ideal for estimating $\overline{Y}$ when the study and auxiliary variables are positively linked, while the product estimator is considered appropriate when a negative correlation exists between the study and the auxiliary variables. Depending on the directions of correlation between the study and the auxiliary variable, we provide novel population mean estimators that deal with each occurrences individually.

Case 1: When the study (Y) and auxiliary variables (X) are highly positively correlated, it is expected that the larger values of the study variable will be chosen when the auxiliary variable's larger values are chosen, and the smaller values of the study variable will be chosen when the auxiliary variable's smaller values are chosen. In this case, we create the following ratio type estimator under simple random sampling by using the transformation and the accompanying auxiliary data.

$$\bar{y}_{RT} = \bar{y}_T \cdot \frac{\bar{X}}{\bar{x}_t} \tag{3.7}$$

and the regression type estimator for positive correlation between the study and the auxiliary variable is defined as follows:

$$\bar{Y}_{lr(T)^+} = \bar{y}_T + b(\bar{X} - \bar{x}_T), \tag{3.8}$$

Case 2: When the study and auxiliary variable are negatively correlated, it is expected that the larger values of the study variable will be chosen when the auxiliary variable's smaller values are chosen, and the smaller values of the study variable will be chosen when the auxiliary variable's larger values are chosen. In this case, we suggest the following product type estimator under simple random sampling by using the transformation and the accompanying auxiliary data as

$$\bar{y}_{PT} = \bar{y}_T \cdot \frac{\bar{x}_T}{\bar{X}}, \tag{3.9}$$

and the regression type estimator for negative correlation between the study and the auxiliary variable is as follows:

$$\bar{Y}_{lr(T)^-} = \bar{y}_T + b(\bar{X} - \bar{x}_T), \tag{3.10}$$

A theorem is first proved for deriving the covariance between $\bar{y}_T$ and $\bar{x}_t$. Later on, the finding will be utilised to derive formulas for the biases and MSE of various suggested estimators.

Theorem 1: In the case of positively correlated variables, it can be demonstrated that the covariance between $\bar{y}_T$ and $\bar{x}_t$ for a simple random sample of size n units drawn from a population of size N units is given by

$$Cov(\bar{y}_T, \bar{x}_t) = \lambda S_{xy} + \frac{\lambda}{N-1}(\sum_{i=1}^{j-1} X_i f x_i - \sum_{i=j+1}^{n} X_i f x_i). \tag{3.11}$$

Proof. The covariance by definition is

$$Cov(\bar{y}_T, \bar{x}_T) = E(\bar{y}_T, \bar{x}_t) - E(\bar{y}_T)E(\bar{x}_T).$$

Consider,

$$
\begin{aligned}
E(\bar{y}_t, \bar{x}_t) &= E(\frac{\sum_{i=1}^{n} Y_i''}{n}, \frac{\sum_{i=1}^{n} X_j''}{n}) \\
&= \frac{1}{n^2} E[\sum_{i=j}^{n} Y_i'' X_j'' + \sum_{i \neq j}^{n} Y_i'' X_j''] \\
&= \frac{1}{n^2}[nE_{i=j} Y_i'' X_j'' + n(n-1) E_{i \neq j} Y_i'' X_j''] \qquad (3.12)
\end{aligned}
$$

Consider the expression $E_{i=j}(Y_i'' X_j'')$

$$
\begin{aligned}
E_{i=j}(Y_i'' X_j'') &= \frac{\sum_{i=1}^{n} Y_i'' X_i''}{N} \\
&= \frac{1}{N}\{(Y_1 + fx_1)(X_1) + (Y_2 + fx_2)(X_2) + ... + (Y_{j-1} + fx_{j-1})(X_{j-1}) \\
&+ Y_j + (Y_{j+1} - fx_{j+1})(X_{j+1}) + ... + (Y_{N-1} - fx_2)(X_{N-1}) + (Y_N - fx_1)(X_N)\} \\
&= \frac{1}{N}\{Y_1 X_1 + fx_1 X_1 + Y_2 X_2 + fx_2 X_2 + ... + Y_{j-1} X_{j-1} + fx_{j-1} X_{j-1} + Y_j + Y_{j+1} X_{j+1} \\
&- fx_{j+1} X_{j+1} + ... + Y_{N-1} X_{N-1} - fx_2 X_{N-1} + Y_N X_N - fx_1 X_N\} \\
&= \frac{\sum_{i=1}^{N} Y_i X_i + \sum_{i=1}^{j-1} fx_i X_i - \sum_{i=j+1}^{N} fx_i X_i}{N}
\end{aligned}
$$

This can also be described as

$$
E_{i=j}(Y_i'' X_j'') = \frac{\sum_{i=1}^{N} (Y_i X_i)}{N} + \frac{\sum_{i=1}^{j-1} fx_i X_i - \sum_{i=j+1}^{N} fx_i X_i}{N}. \qquad (3.13)
$$

Now we consider the expression

$$
\begin{aligned}
E_{i \neq j}(Y_i'' X_j'') &= \frac{\sum_{i=j=1}^{n} Y_i'' X_j''}{N(N-1)} \\
&= \frac{\sum_{i=1}^{n} Y_i'' \sum_{j=1}^{n} X_j'' - \sum_{i \neq j}^{n} Y_i'' X_j''}{N(N-1)} \\
&= \frac{\sum_{i=1}^{n} Y_i'' \sum_{j=1}^{n} X_j''}{N(N-1)} - \frac{\sum_{i \neq j}^{n} Y_i'' X_j''}{N(N-1)}.
\end{aligned}
$$

Substituting the derived expressions of $\sum_{i \neq j}^{n} Y_i'' X_j''$ in the above equation

$$
\sum_{i \neq j}^{n} Y_i'' X_j'' = \sum_{i \neq j}^{N} Y_i X_j + \sum_{i=1}^{j-1} fx_i X_i - \sum_{i=j+1}^{N} fx_i X_i
$$

$$
E_{i \neq j}(Y_i'' X_j'') = \frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j}{N(N-1)} - \frac{1}{N(N-1)} \left(\sum_{i=1}^{n} Y_i X_j + \sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{n} X_i fx_i \right)
$$

$$E_{i\neq j}(Y_i''X_j'') = \frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j}{N(N-1)} - \frac{\sum_{i=1}^{n} Y_i X_j}{N(N-1)} - \frac{\sum_{i=1}^{j-1} X_i f x_i}{N(N-1)} + \frac{\sum_{i=J+1}^{n} X_i f x_i}{N(N-1)} \tag{3.14}$$

In the expression of $Cov(\bar{y}_T, \bar{x}_t)$, equations (3.13) and (3.14) are now substituted, and the result is

$$Cov(\bar{y}_t, \bar{x}_t) = \frac{1}{n^2}\left[n\left(\frac{\sum_{i=1}^{N}(Y_iX_i)}{N} + \frac{\sum_{i=1}^{j-1} fx_iX_i - \sum_{i=j+1}^{N} fx_iX_i}{N}\right) + n(n-1)\left(\frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j}{N(N-1)} - \frac{\sum_{i=1}^{n} Y_iX_j}{N(N-1)} - \frac{\sum_{i=1}^{j-1} X_i f x_i}{N(N-1)} + \frac{\sum_{i=J+1}^{n} X_i f x_i}{N(N-1)}\right) - \bar{Y}\bar{X}\right]$$

$$= \frac{\sum_{i=1}^{n} Y_iX_i}{Nn} + \frac{\sum_{i=1}^{j-1} fx_iX_i}{Nn} - \frac{\sum_{i=j+1}^{n} fx_iX_i}{Nn} + \frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j(n-1)}{N(N-1)n} - \frac{\sum_{i=1}^{n} Y_iX_i(n-1)}{N(N-1)n} - \frac{\sum_{i=1}^{j-1} fx_iX_i(n-1)}{(N-1)Nn} + \frac{\sum_{i=j+1}^{n} X_i fx_i(n-1)}{N(N-1)n} - \bar{Y}\bar{X}$$

$$= \frac{\sum_{i=1}^{n} Y_iX_i}{Nn}\left(1 - \frac{n-1}{N-1}\right) + \frac{\sum_{i=1}^{j-1} fx_iX_i}{Nn}\left(1 - \frac{n-1}{N-1}\right) - \frac{\sum_{i=j+1}^{n} fx_iX_i}{Nn}\left(1 - \frac{n-1}{N-1}\right) + \frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j(n-1)N}{N(N-1)nN} - \bar{Y}\bar{X}$$

$$= \frac{\sum_{i=1}^{n} Y_iX_i}{Nn}\left(\frac{N-n}{N-1}\right) + \frac{\sum_{i=1}^{j-1} fx_iX_i}{Nn}\left(\frac{N-n}{N-1}\right) - \frac{\sum_{i=j+1}^{n} fx_iX_i}{Nn}\left(\frac{N-n}{N-1}\right) + \frac{\bar{Y}\bar{X}(n-1)N}{(N-1)n} - \bar{Y}\bar{X}$$

$$= \frac{\sum_{i=1}^{n} Y_iX_i}{Nn}\left(\frac{N-n}{N-1}\right) - \bar{Y}\bar{X}\left(1 - \frac{(n-1)N}{(N-1)n}\right) + \left(\frac{N-n}{N-1}\right)\frac{1}{nN}\left(\sum_{i=1}^{j-1} fx_iX_i - \sum_{i=j+1}^{n} fx_iX_i\right)$$

$$= \left(\frac{N-n}{N-1}\right)\frac{1}{n}\left[\frac{\sum_{i=1}^{n} Y_iX_i}{N} - \bar{Y}\bar{X}\right] + \left(\frac{N-n}{(N-1)Nn}\right)\left(\sum_{i=1}^{j-1} fx_iX_i - \sum_{i=j+1}^{n} fx_iX_i\right)$$

$$= \left(\frac{N-n}{N-1}\right)\frac{1}{n}.\sigma_{yx} + \left(\frac{N-n}{(N-1)Nn}\right)\left(\sum_{i=1}^{j-1} fx_iX_i - \sum_{i=j+1}^{n} fx_iX_i\right)$$

$$= \left(\frac{N-n}{N-1}\right)\frac{1}{n}.\frac{N-1}{N}S_{yx} + \left(\frac{N-n}{(N-1)Nn}\right)\left(\sum_{i=1}^{j-1} fx_iX_i - \sum_{i=j+1}^{n} fx_iX_i\right)$$

Now the covariance for the positive correlation becomes

$$Cov(\bar{y}_T, \bar{x}_T) = \lambda S_{yx} + \frac{\lambda}{N-1}\left(\sum_{i=1}^{j-1} fx_iX_i - \sum_{i=j+1}^{n} fx_iX_i\right). \tag{3.15}$$

Theorem 2: If a simple random sample of size n units is taken from a population of size N units, the covariance between $\bar{y}_T$ and $\bar{x}_T$, when there is a negative correlation between the study variable (Y) and the auxiliary variable (x), may be calculated as follows.

$$Cov(\bar{y}_T, \bar{x}_T) = \lambda S_{xy} - \frac{\lambda}{N-1}(\sum_{i=1}^{j-1} X_i f x_i - \sum_{i=j+1}^{n} X_i f x_i) \tag{3.16}$$

Proof. The covariance by definition,

$$Cov(\bar{y}_t, \bar{x}_t) = E(\bar{y}_t, \bar{x}_t) - E(\bar{y}_t)E(\bar{x}_t).$$

Consider,

$$\begin{aligned} E(\bar{y}_t, \bar{x}_t) &= E(\frac{\sum_{i=1}^{n} Y_i''}{n}, \frac{\sum_{i=1}^{n} X_j''}{n}) \\ &= \frac{1}{n^2} E[\sum_{i=j}^{n} Y_i'' X_j'' + \sum_{i \neq j}^{n} Y_i'' X_j''] \\ &= \frac{1}{n^2}[n E_{i=j} Y_i'' X_j'' + n(n-1) E_{i \neq j} Y_i'' X_j''] \end{aligned}$$

Consider the expression $E_{i=j}(Y_i'' X_j'')$

$$\begin{aligned} E_{i=j}(Y_i'' X_j'') &= \frac{\sum_{i=1}^{n} Y_i'' X_i''}{N} \\ &= \frac{1}{N}\{(Y_1 - f x_1)(X_1) + (Y_2 - f x_2)(X_2) + ... + (Y_{j-1} - f x_{j-1})(X_{j-1}) \\ &+ Y_j X_j + (Y_{j+1} + f x_{j+1})(X_{j+1}) + ... + (Y_{N-1} + f x_2)(X_{N-1}) + (Y_N + f x_1)(X_N)\} \\ &= \frac{1}{N}\{Y_1 X_1 - f x_1 X_1 + Y_2 X_2 - f x_2 X_2 - .... + Y_{j-1} X_{j-1} - f x_{j-1} X_{j-1} + Y_j X_j + Y_{j+1} X_{j+1} \\ &+ f x_{j+1} X_{j+1} + ... + Y_{N-1} X_{N-1} + f x_2 X_{N-1} + Y_N X_N + f x_1 X_N\} \\ &= \frac{\sum_{i=1}^{N} Y_i X_i - \sum_{i=1}^{j-1} f x_i X_i + \sum_{i=j+1}^{N} f x_i X_i}{N} \end{aligned}$$

This can also be reduced to

$$E_{i=j}(Y_i'' X_j'') = \frac{\sum_{i=1}^{N}(Y_i X_i)}{N} - \frac{\sum_{i=1}^{j-1} f x_i X_i + \sum_{i=j+1}^{N} f x_i X_i}{N}. \tag{3.17}$$

Now we consider the expression

$$\begin{aligned} E_{i \neq j}(Y_i'' X_j'') &= \frac{\sum_{i \neq j}^{n} Y_i'' X_j''}{N(N-1)} \\ &= \frac{\sum_{i=1}^{n} Y_i'' \sum_{j=1}^{n} X_j'' - \sum_{i=j}^{n} Y_i'' X_j''}{N(N-1)} \\ &= \frac{\sum_{i=1}^{n} Y_i'' \sum_{j=1}^{n} X_j''}{N(N-1)} - \frac{\sum_{i=j}^{n} Y_i'' X_j''}{N(N-1)} \end{aligned}$$

Putting expression of $\sum_{i=j}^{n} Y_i'' X_j''$ in the above equation, we get

$$E_{i\neq j}(Y_i'' X_j'') = \frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j}{N(N-1)} - \frac{1}{N(N-1)} \left(\sum_{i=1}^{n} Y_i X_j - \sum_{i=1}^{j-1} X_i f x_i + \sum_{i=j+1}^{n} X_i f x_i \right)$$

$$E_{i\neq j}(Y_i'' X_j'') = \frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j}{N(N-1)} - \frac{\sum_{i=1}^{n} Y_i X_j}{N(N-1)} + \frac{\sum_{i=1}^{j-1} X_i f x_i}{N(N-1)} - \frac{\sum_{i=J+1}^{n} X_i f x_i}{N(N-1)} \tag{3.18}$$

In the expression of $Cov(\bar{y}_T, \bar{x}_T)$, subsitituting equations (3.17) and (3.18), and the result can be shown to be given by

$$\begin{aligned}
Cov(\bar{y}_T, \bar{x}_T) &= \frac{1}{n^2}\left[n\left(\frac{\sum_{i=1}^{N}(Y_i X_i)}{N} - \frac{\sum_{i=1}^{j-1} f x_i X_i + \sum_{i=j+1}^{N} f x_i X_i}{N} \right) + n(n-1)\left(\frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j}{N(N-1)} \right.\right. \\
&\left.\left. - \frac{\sum_{i=1}^{n} Y_i X_j}{N(N-1)} + \frac{\sum_{i=1}^{j-1} X_i f x_i}{N(N-1)} - \frac{\sum_{i=J+1}^{n} X_i f x_i}{N(N-1)} \right) - \bar{Y}\bar{X} \right] \\
&= \frac{\sum_{i=1}^{n} Y_i X_i}{Nn} - \frac{\sum_{i=1}^{j-1} f x_i X_i}{Nn} + \frac{\sum_{i=j+1}^{n} f x_i X_i}{Nn} + \frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j (n-1)}{N(N-1)n} \\
&- \frac{\sum_{i=1}^{n} Y_i X_i (n-1)}{N(N-1)n} + \frac{\sum_{i=1}^{j-1} f x_i X_i (n-1)}{(N-1)Nn} - \frac{\sum_{i=j+1}^{n} X_i f x_i (n-1)}{N(N-1)n} - \bar{Y}\bar{X} \\
&= \frac{\sum_{i=1}^{n} Y_i X_i}{Nn}\left(1 - \frac{n-1}{N-1}\right) - \frac{\sum_{i=1}^{j-1} f x_i X_i}{Nn}\left(1 - \frac{n-1}{N-1}\right) + \frac{\sum_{i=j+1}^{n} f x_i X_i}{Nn}\left(1 - \frac{n-1}{N-1}\right) \\
&+ \frac{\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} X_j (n-1)N}{N(N-1)nN} - \bar{Y}\bar{X} \\
&= \frac{\sum_{i=1}^{n} Y_i X_i}{Nn}\left(\frac{N-n}{N-1}\right) - \frac{\sum_{i=1}^{j-1} f x_i X_i}{Nn}\left(\frac{N-n}{N-1}\right) + \frac{\sum_{i=j+1}^{n} f x_i X_i}{Nn}\left(\frac{N-n}{N-1}\right) \\
&+ \frac{\bar{Y}\bar{X}(n-1)N}{(N-1)n} - \bar{Y}\bar{X} \\
&= \frac{\sum_{i=1}^{n} Y_i X_i}{Nn}\left(\frac{N-n}{N-1}\right) - \bar{Y}\bar{X}\left(1 - \frac{(n-1)N}{(N-1)n}\right) - \left(\frac{N-n}{N-1}\right)\frac{1}{nN} \\
&\left(\sum_{i=1}^{j-1} f x_i X_i - \sum_{i=j+1}^{n} f x_i X_i \right) \\
&= \left(\frac{N-n}{N-1}\right)\frac{1}{n}\left[\frac{\sum_{i=1}^{n} Y_i X_i}{N} - \bar{Y}\bar{X} \right] - \left(\frac{N-n}{(N-1)Nn}\right)\left(\sum_{i=1}^{j-1} f x_i X_i - \sum_{i=j+1}^{n} f x_i X_i \right) \\
&= \left(\frac{N-n}{N-1}\right)\frac{1}{n}.\sigma_{yx} - \left(\frac{N-n}{(N-1)Nn}\right)\left(\sum_{i=1}^{j-1} f x_i X_i - \sum_{i=j+1}^{n} f x_i X_i \right) \\
&= \left(\frac{N-n}{N-1}\right)\frac{1}{n}.\frac{N-1}{N} S_{yx} - \left(\frac{N-n}{(N-1)Nn}\right)\left(\sum_{i=1}^{j-1} f x_i X_i - \sum_{i=j+1}^{n} f x_i X_i \right)
\end{aligned}$$

Now the covariance for the negative correlation becomes

$$Cov(\bar{y}_T, \bar{x}_T) = \lambda S_{yx} - \frac{\lambda}{N-1}\left(\sum_{i=1}^{j-1} fx_i X_i - \sum_{i=j+1}^{n} fx_i X_i\right). \tag{3.19}$$

Proposed Ratio Estimator: Based on the transformed mean per unit estimator, we propose a ratio estimator as

$$\bar{y}_{RT} = \bar{y}_t \cdot \frac{\bar{X}}{\bar{x}_t}. \tag{3.20}$$

The difference between a conventional and this ratio estimator is that here we use $\bar{y}_t$ instead of the usual mean per unit estimator $\bar{y}$. The relative error terms are defined for the purpose of deriving the bias and MSE of the proposed Ratio estimator when there is a positive correlation between study and auxiliary variable as:

$$e_0 = (\bar{y}_t - \bar{Y})/\bar{Y}$$
$$e_1 = (\bar{x}_t - \bar{X})/\bar{X}$$

Using Theorem 1, their expectation can be written as

$$E(e_0) = E(e_1) = 0, \tag{3.21}$$

$$E(e_0^2) = \frac{\lambda}{\bar{Y}^2}\left[S_y^2 + \frac{2}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right)\right] \tag{3.22}$$

$$E(e_1^2) = \frac{\lambda}{\bar{X}^2}\left[S_x^2\right] \tag{3.23}$$

$$E(e_0 e_1) = \frac{\lambda}{\bar{Y}\bar{X}}\left[S_{yx} + \frac{1}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right)\right]. \tag{3.24}$$

The suggested estimator $\bar{y}_{RT}$ cab be rewritten in terms of e_i as

$$\bar{y}_{RT} = \bar{Y}(1+e_0)(1+e_1)^{-1}.$$

Up to the first level of approximation, expanding and rearranging the right side of the equation and disregarding any terms with powers greater than two

$$(\bar{y}_{RT} - \bar{Y}) \approx \bar{Y}(e_0 - e_1 + e_1^2 - e_0 e_1). \tag{3.25}$$

To obtain the bias we apply expectation on both sides of equation (3.24) and then the bias is as follows

$$Bias(\bar{y}_{RT}) \approx \bar{Y}\left[\frac{\lambda}{\bar{X}^2}\{S_x^2\} - \frac{\lambda}{\bar{X}\bar{Y}}\{S_{xy} + \frac{1}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right)\right]. \tag{3.26}$$

In order to derive an expression for MSE, we squared both sides of (3.25), keeping just the terms up to the first level of approximation.

$$(\bar{y}_{RT} - \bar{Y})^2 \approx \bar{Y}^2[e_0^2 + e_1^2 - 2e_0e_1].$$

Applying expectations on both sides of the aforementioned equation, we obtain MSE,

$$MSE(\bar{y}_{RT}) \approx \Bigg[\lambda\bigg(S_y^2 + \delta^2 S_x^2 - 2\delta S_{xy}\bigg) + \frac{2\lambda}{N-1}\bigg(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\bigg) - \frac{2\lambda\delta}{N-1}\bigg(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\bigg)\Bigg] \tag{3.27}$$

where

$$\delta = \bar{Y}/\bar{X}$$

The Proposed Product Estimator: The suggested estimator is

$$\bar{y}_{PT} = \bar{y}_t \cdot \frac{\bar{X}}{\bar{x}_t} \tag{3.28}$$

The only difference between the suggested and conventional estimator is the transformed mean per unit estimator. The relative error terms are defined for the purpose of deriving the bias and MSE of the proposed product estimator when there is a negative correlation between study and auxiliary variable as $e_0 = (\bar{y}_t - \bar{Y})/\bar{Y}$ and $e_1 = (\bar{x}_t - \bar{X})/\bar{X}$. Using Theorem 2, their expectation can be written as

$$E(e_0) = E(e_1) = 0, \tag{3.29}$$

$$E(e_0^2) = \frac{\lambda}{\bar{Y}^2}\Bigg[S_y^2 + \frac{2}{N-1}\bigg(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\bigg)\Bigg] \tag{3.30}$$

$$E(e_1^2) = \frac{\lambda}{\bar{X}^2}\Bigg[S_x^2\Bigg] \tag{3.31}$$

and

$$E(e_0e_1) = \frac{\lambda}{\bar{Y}\bar{X}}\Bigg[S_{yx} - \frac{1}{N-1}\bigg(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\bigg)\Bigg] \tag{3.32}$$

Expressing the suggested estimator $\bar{y}_{PT}$ in terms of e_i, we obtain

$$\bar{y}_{PT} = \bar{Y}(1 + e_0)(1 + e_1).$$

Up to the first level of approximation, by enlarging and rearrange the right side of the equation and disregarding any terms with second powers of $e_i's$,

$$(\bar{y}_{PT} - \bar{Y}) = \bar{Y}(e_0 + e_1 + e_0e_1). \tag{3.33}$$

To obtain the bias we applied expectation on both sides of equation (3.32) and then the bias is as follows

$$Bias(\bar{y}_{PT}) \approx \left[\frac{\lambda}{\bar{X}}\{S_{xy} - \frac{1}{N-1}\left(\sum_{i=1}^{j-1} X_i f x_i - \sum_{i=j+1}^{N} X_i f x_i\right)\right]. \tag{3.34}$$

In order to produce MSE, we squared both sides of (3.24), keeping just the terms up to the first level of approximation.

$$(\bar{y}_{PT} - \bar{Y})^2 \approx \bar{Y}^2[e_0^2 + e_1^2 + 2e_0e_1].$$

Applying expectations on both sides of the aforementioned equation, an expression for the MSE of the proposed product estimator is

$$MSE(\bar{y}_{PT}) \approx \left[\lambda\left(S_y^2 + \delta^2 S_x^2 + 2\delta S_{xy}\right) + \frac{2\lambda}{N-1}\left(\sum_{i=1}^{j-1} f x_i^2 + \sum_{i=1}^{j-1} y_i f x_i - \sum_{i=j+1}^{N} y_i f x_i\right) - \frac{2\lambda\delta}{N-1}\left(\sum_{i=1}^{j-1} X_i f x_i - \sum_{i=j+1}^{N} X_i f x_i\right)\right] \tag{3.35}$$

Proposed Regression Estimator for positive correlation: The suggested regression estimator is

$$\bar{y}_{lr(T)} = \bar{y}_T + b(\bar{X} - \bar{x}_t)$$

The variance of the linear regression estimator, up to the first level of approximation, may now be derived as follows for positive correlation. Expressing $\bar{y}_{lr(T)}$ in terms of relative errors as

$$\bar{y}_{lr(T)} = \bar{Y}(1 + e_0) + b(\bar{X} - \bar{X}(1 + e_1)).$$

Up to first level of approximation, we expand and rearrange all the terms to obtain

$$\bar{y}_{lr(T)} - \bar{Y} = \bar{Y}e_0 - be_1\bar{X}$$

When the two sides of the aforementioned equation are squared and expectation is applied, the expression of MSE simplifies to

$$E[\bar{y}_{lr(P1)} - \bar{Y}]^2 = \lambda(S_y^2 + b^2 S_x^2 - 2bS_{xy}) + \frac{2\lambda}{N-1}\left(\sum_{i=1}^{j-1} f x_i^2 + \sum_{i=1}^{j-1} y_i f x_i - \sum_{i=j+1}^{N} y_i f x_i\right) - \frac{2b\delta}{N-1}\left(\sum_{i=1}^{j-1} X_i f x_i - \sum_{i=j+1}^{N} X_i f x_i\right)$$

where b=S_{xy}/S_x^2.

$$Var_{(\bar{y}_{lr(T)})^{+}} = \lambda S_y^2\left[1-\rho_{yx}^2\right] + \frac{2\lambda}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right) \tag{3.36}$$

$$-\frac{2b\lambda}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right). \tag{3.37}$$

Similar to positive correlation when there is a negative correlation, the variance of the regression estimate can be shown to be derived as

$$\begin{aligned} Var_{(\bar{y}_{lr(T)})^{-}} = \lambda S_y^2\left[1-\rho_{yx}^2\right] &+ \frac{2\lambda}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right) \\ &+\frac{2b\lambda}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right) \end{aligned} \tag{3.38}$$

A general expression for $Var_{(\bar{y}_{lr(T)})}$ can be described as

$$\begin{aligned} Var_{(\bar{y}_{lr(T)})} = \lambda S_y^2\left[1-\rho_{yx}^2\right] &+ \frac{2\lambda}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right) \\ &-\frac{2|b|\lambda}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right) \end{aligned} \tag{3.39}$$

3.3 Efficiences Comparison:

The circumstances in which the new estimators perform better than the existing estimators may be identified by comparing the MSE of the proposed estimators to the MSE of existing estimators. The results are described in the section below.

(1) By comparing the proposed ratio estimator and the usual mean per unit estimator we get the expression

$$Var(\hat{\bar{Y}}_{Pi}) - MSE(\bar{y}_{RT}) > 0 \tag{3.40}$$

The equation(3.39) is true if and only if

$$\begin{aligned} &-\delta S_x^2 + 2\delta S_y x - \frac{2}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right). \\ &+\frac{2\delta}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right) > 0 \end{aligned} \tag{3.41}$$

(2) By comparing the proposed ratio estimator and the usual ratio estimator we get the expression

$$MSE(\hat{\bar{Y}}_R) - MSE(\bar{y}_{RT}) > 0 \tag{3.42}$$

The equation (3.40) is true if and only if

$$-\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right) + \delta\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right) > 0 \tag{3.43}$$

(3) By comparing the proposed product estimator and the usual mean per unit estimator we get the expression

$$Var(\hat{\bar{y}}) - MSE(\bar{y}_{PT}) > 0 \tag{3.44}$$

The equation (3.42) is true if and only if

$$\begin{aligned} &-\delta S_x^2 - 2\delta S_y x - \frac{2}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right). \\ &+ \frac{2\delta}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right) > 0 \end{aligned} \tag{3.45}$$

(4) By comparing the proposed product estimator and the usual product estimator we get the expression

$$MSE(\hat{\bar{y}}_P) - MSE(\bar{y}_{PT}) > 0 \tag{3.46}$$

The equation (3.44) is true if and only if

$$-\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right) + \delta\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right) > 0 \tag{3.47}$$

(5) By comparing the proposed regression estimator and the usual mean per unit estimator for positive correlation we get the expression

$$Var(\hat{\bar{y}}) - Var(\bar{y}_{lrT}) > 0 \tag{3.48}$$

The equation (3.46) is true if and only if

$$\begin{aligned} &S_y^2\rho_{yx}^2 - \frac{2}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2 + \sum_{i=1}^{j-1} y_i fx_i - \sum_{i=j+1}^{N} y_i fx_i\right) \\ &+ \frac{2b}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i - \sum_{i=j+1}^{N} X_i fx_i\right) > 0 \end{aligned} \tag{3.49}$$

(5) By comparing the proposed regression estimator and the usual regression estimator for positive correlation we get the expression

$$Var(\bar{y}_{lr}) - Var(\bar{y}_{lrT}) > 0 \tag{3.50}$$

The equation (3.48) is true if and only if

$$-\frac{2}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2+\sum_{i=1}^{j-1} y_i fx_i-\sum_{i=j+1}^{N} y_i fx_i\right)+\frac{2b}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i-\sum_{i=j+1}^{N} X_i fx_i\right)>0 \tag{3.51}$$

(6) By comparing the proposed regression estimator and the usual mean per unit estimator for negative correlation we get the expression

$$Var(\hat{\bar{y}})-Var(\bar{y}_{lrT})>0 \tag{3.52}$$

The equation (3.50) is true if and only if

$$\begin{aligned} S_y^2\rho_{yx}^2 &- \frac{2}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2+\sum_{i=1}^{j-1} y_i fx_i-\sum_{i=j+1}^{N} y_i fx_i\right) \\ &- \frac{2b}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i-\sum_{i=j+1}^{N} X_i fx_i\right)>0 \end{aligned} \tag{3.53}$$

(7) By comparing the proposed regression estimator and the usual regression estimator for negative correlation we get the expression

$$Var(\bar{y}_{lr})-Var(\bar{y}_{lrT})>0 \tag{3.54}$$

The equation (3.52) is true if and only if

$$-\frac{2}{N-1}\left(\sum_{i=1}^{j-1} fx_i^2+\sum_{i=1}^{j-1} y_i fx_i-\sum_{i=j+1}^{N} y_i fx_i\right)-\frac{2b}{N-1}\left(\sum_{i=1}^{j-1} X_i fx_i-\sum_{i=j+1}^{N} X_i fx_i\right)>0 \tag{3.55}$$

All of the conditions from (1) to (7) that were established above are conditionally true, and the proposed estimators will work better whenever the deduced conditions are true.

3.4 Empirical Comparison of Estimators

In this part, real data sets are used to compare the performance of the recommended estimators to that of competing estimators using four real data sets. The population description and the pertinent statistical statistics are provided below.

Data 1:(Source: Montgomery et al. (2021))
Data set: Jet turbine engine thrust data [page 566]

y=Thrust

x= Primary speed of rotation

Following are the population's abridged statistics:

$N = 40$, $\lambda = 1$, $\bar{Y} = 3904$,$\bar{X} = 1809.925$, $S_y^2 = 254667.6$, $S_x^2 = 63479.05$, $\delta = 2.156995$, $\rho_{yx} = 0.9950099$ (correlation), $S_{xy} = 126511.3$, $C_y^2 = 0.01670914$, $C_x^2 = 0.01937801$, $b_{yx} = 1.992962$,

Data 2:(Source: Montgomery et al. (2021))
Data set: Solar Thermal Energy Test Data Data [page 555]

y= Total heat fl ux (kwatts)

x= Insolation (watts/m 2)

Following are the population's abridged statistics:

$N = 29$, $\lambda = 1$, $\bar{Y} = 249.6379$,$\bar{X} = 754.4741$, $S_y^2 = 524.3546$, $S_x^2 = 6367.533$, $\delta = 0.3308767$, $\rho_{yx} = 0.6276454$ (correlation), $S_{xy} = 1146.865$, $C_y^2 = 0.008414027$, $C_x^2 = 0.0111862$, $b_{yx} = 0.1801114$,

Data 3:(Source: Montgomery et al. (2021))
Data set: Belle Ayr Liquefaction Runs Data [page 558]

$y = CO_2$

$x = Spacetimemin$

Following are the population's abridged statistics:

$N = 27$, $\lambda = 1$, $\bar{Y} = 24.73037$,$\bar{X} = 20.11111$, $S_y^2 = 302.6966$, $S_x^2 = 200.961$, $\delta = 1.229687$, $\rho_{yx} = -0.7048039$ (correlation), $S_{xy} = -173.8313$, $C_y^2 = 0.4949329$, $C_x^2 = 0.4968665$, $b_{yx} = -0.8650001$,

Data 4:(Source: Montgomery et al. (2021))
Data set: Gasoline Mileage Performance for 32 Antomobiles Data [page 556]

$y = Miles/gallon$

$x = Carburetor(barrels)$

Following are the population's abridged statistics:

$N = 32$, $\lambda = 1$, $\bar{Y} = 20.22312$,$\bar{X} = 2.59375$, $S_y^2 = 39.92078$, $S_x^2 = 1.152218$, $\delta = 7.796867$, $\rho_{yx} = -0.4869972$ (correlation), $S_{xy} = -3.302883$, $C_y^2 = 0.09761183$, $C_x^2 = 0.1712688$, $b_{yx} = -2.866544$,

At this point, a numerical comparison of the proposed and current estimators is made using a percentage relative efficiency specified by

$$PRE(\cdots) = [Var(\bar{y})/MSE(\cdots)] \times 100$$

where the MSE of the competing estimators is located in the denominator. The proposed and current estimators' percentage relative efficiencies are shown in Table 3.1. The data demonstrates that the suggested estimators outperform the competing estimators.

Table 3.1: Percentage Relative Efficiencies of existing and proposed Estimators

Estimators	Data 1	Data2	Data3	Data4
$\hat{\bar{y}}$	100	100	100	100
$\hat{\bar{y}}_R$	6001.6110	113.3676	29.2717	24.7233
$\hat{\bar{y}}_{RT}$	6174.2920	125.3804	31.7711	30.1112
$\hat{\bar{y}}_P$	23.2407	26.4770	169.0480	68.2860
$\hat{\bar{y}}_{PT}$	23.2432	27.0831	309.7916	135.0076
$\bar{y}_{LR}$	10044.8900	164.9998	198.7078	131.0902
$\bar{y}_{LRT}$	10144.0500	173.2754	356.0389	281.6299

3.5 Conclusion

In this chapter, we suggested transformation-based mean per unit, ratio, product, and regression type estimators for the mean of a finite population. Additionally, under first level of approximation, expressions for the bias and MSE have been developed. The suggested estimators' effectiveness has been established theoretically, and four real data sets were utilised to validate the theoretical findings. Based on the numerical results, it have been suggested that the proposed estimator should be used for the estimation of the population mean as its relative efficiency is better than the exisitng estimators. For data set 1 and data set 2, the correlation between the study and the auxiliary variable is positive that's why the product estimators do not produce satisfactory results.

Chapter 4

Estimation of Finite Population Mean Using Transformation Under Stratified Random Sampling

A sampling method known as stratified random sampling is used in statistics and research to guarantee that the sample taken from a population accurately reflects each of its strata in proportion to their presence in the total population. In particular when the population is diverse, the objective is to create a sample that is more accurate and representative than simply random sampling.

The population is initially segmented into different, non-overlapping subgroups or strata in accordance with some traits or properties. Each component of the population only belongs to one of these strata.The proportion of the sample that should be drawn from each stratum is decided once the strata have been defined. According to the size or significance of the strata in the population as a whole, this allocation is often made.A simple random sampling technique is used to choose a random sample of items from each stratum. As a result, there is a fair possibility that each stratum member will be chosen.

In stratified random sampling, auxiliary information refers to supplementary information or variables that are known for each person or unit within the stratum. These auxiliary variables can offer insightful data on the studied research variable, or the feature of interest.When there is a significant correlation between the auxiliary variables and the research variable, the sampling procedure may produce estimates that are more precise.

When a study variable has to be transformed, it typically signifies that certain modifications are required since the variable's distribution cannot be used for sampling or analysis. Here are some common situations where transformation of study variables might be necessary in stratified random sampling:

Estimates may be skewed if the research(study) variable is extremely skewed (e.g., positively or negatively skewed).Sometimes the distribution may be made more symmetric and acceptable for sampling by transforming the study variable using mathematical procedures like the exponential function.A study variable's mean and variance can be strongly impacted by outliers, producing estimates that are biased. Outliers can have an adverse effect, but they can be reduced by transformations such swapping extreme numbers for less extreme ones or utilising rankings rather than raw values.It might be challenging to conduct a representative sample when the study variable's variation varies between strata.It could be useful to use transformations that reduce variance, such as the square root or the inverse.The correlation between the study variable and other factors may not always be linear. The relationship can be linearized and the analysis made easier by transforming the study variable or the other variables.

Shoaib et al. (2018) proposed ratio, product, and regression type estimators employing extreme data values in stratified random sampling with one auxiliary variable. In this chapter we use Antithetic variable technique in transformation to reduce the variance of the study variable to estimate the finite population mean under stratified random sampling.

4.1 Notation and symbols under stratified random sampling

Let's say that a population of N is split into L strata of sizes $(h = 1, 2, 3, ..., L)$ that are mutually exclusive and have sums of $\sum_{h=1}^{L} N_h = N$.Given that n is the total number of units in the sample, let's assume that a random sample of size n_h is taken from each stratum separately using simple random sampling without replacement.Assume that y_{hi} and x_{hi} are the sample values of the study variable and the auxiliary variable, respectively. These values correspond to the population values of the study variable Y_{hi} and the 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 $\bar{y}_h = \sum_{h=1}^{n_h} y_{hi}/n_h$ be the sample mean of study variable in the h^{th} (h=1,2,3,...,L) stratum.

Let $\bar{x}_h = \sum_{h=1}^{N_h} X_{hi}/n_h$ be the sample mean of auxiliary variable in the h^{th} (h=1,2,3,...,L) stratum.

Let $\bar{Y}_h = \sum_{h=1}^{N_h} Y_{hi}/N_h$ be the population mean of study variable in the h^{th} (h=1,2,3,...,L) stratum.

Let $\bar{X}_h = \sum_{h=1}^{N_h} X_{hi}/N_h$ be the population mean of auxiliary variable in the h^{th} (h=1,2,3,...,L) stratum.

Let $s_{yh}^2 = \sum_{h=1}^{n_h}(y_{hi} - \bar{y}_h)/(n_h - 1)$ be the sample variance of study variable in the h^{th} (h=1,2,3,...,L) stratum.

Let $s_{xh}^2 = \sum_{h=1}^{n_h}(x_{hi} - \bar{x}_h)/(n_h - 1)$ be the sample variance of auxiliary variable in the h^{th} (h=1,2,3,...,L) stratum.

Let $S_{yh}^2 = \sum_{h=1}^{N_h}(Y_{hi} - \bar{Y}_h)/(N_h - 1)$ be the population variance of study variable in the h^{th} (h=1,2,3,...,L) stratum.

Let $S_{xh}^2 = \sum_{h=1}^{N_h}(X_{hi} - \bar{X}_h)/(N_h - 1)$ be the population variance of auxiliary variable in the h^{th} (h=1,2,3,...,L) stratum.

Let $S_{hyx} = S_{hy} \cdot S_{hx} \cdot \rho_{hyx}$ be the population covariance of Y and X for the h^{th} stratum.

Let $b_h = S_{hyx}/S^2_{hx}$ be the Population Regression Coefficient for the h^{th} (h=1,2,3,...,L) stratum.

Let $W_h = N_h/N$ be the stratum weight in the h^{th} (h=1,2,3,...,L) stratum.

Let $f_h = n_h/N_h$ be the sampling fraction in the h^{th} (h=1,2,3,...,L) stratum.

Let $\delta = \bar{Y}/\bar{X}$ be the Population ratio.

Let $\delta_h = \bar{Y}_h/\bar{X}_h$ be the population ratio in the h^{th} (h=1,2,3,...,L) stratum.

Let ρ_{yxh} be the sample correlation coefficient in the h^{th} (h=1,2,3,...,L) stratum.

4.2 Proposed Estimator based on transformation

The mean per unit estimator and its variance under stratified random sampling is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{4.1}$$

$$var(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \lambda_h S^2_{hy} \tag{4.2}$$

where $\lambda_h = \dfrac{1-f_h}{n_h}$

we have proposed the following estimator under stratified random sampling using transformation.

$$\bar{y}_{st.T} = \frac{\sum_{h=1}^{L} W_h \bar{y}_h}{n_h}. \tag{4.3}$$

The proposed estimator's variance is obtained as follows by definition:

$$Var(\bar{y}_{st.T}) = E(\bar{y}^2_{st.T}) - [E(\bar{y}_{st.T})]^2 \tag{4.4}$$

where

$$\begin{aligned} E(\bar{y}_{hT})^2 &= E\left(\frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}\right)^2 \\ &= \frac{1}{n_h^2} E\left[\sum_{i=1}^{n_h} y^2_{hi} + \sum_{i=j} y_{hi} y_{hj}\right] \end{aligned}$$

$$E(\bar{y}_{hT})^2 = \frac{1}{n_h^2}[n_h E(y_{hi})^2 + n_h(n_h-1)E(y_{hi}y_{hj})] \tag{4.5}$$

Now Consider

$$E(y_{hi})^2 = \frac{1}{N_h}\sum_{i=1}^{N_h} y^2_{hi}$$

$$
\begin{aligned}
E(y_{hi}^2) &= \frac{\sum_{i=1}^{N_h} y_{hi}^2}{N_h} \\
&= \frac{1}{N_h}\{(y_{h1} + fx_1)^2 + (y_{h2} + fx_2)^2 + ... + (y_{hj-1} + fx_{hj-1})^2 \\
&+ (y_{jh})^2 + (y_{hj+1} - fx_{hj+1})^2 + ... + (y_{Nh-1} - fx_{2h})^2 + (y_{Nh} - fx_{1h})^2\} \\
&= \frac{\sum_{i=1}^{Nh} yhi^2 + 2\sum_{i=1}^{jh-1} fx_{ih}^2 + 2\sum_{i=1}^{jh-1} y_{hi}fx_{ih} - 2\sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}}{Nh} \\
&= E(y_{ih}^2) + \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{Nh}
\end{aligned}
$$

Now consider the cross product term and using the expression $(\sum_{i=1}^{N_h} y_{hi})^2 = \sum_{i=1}^{N_h} y_{hi}^2 + \sum_{i=1}^{N_h}\sum_{j=1}^{N_h} y_{hi}y_{hj}$

$$
\begin{aligned}
E(y'_{hi}y'_{hj}) &= \frac{\sum_{i=1}^{N_h}\sum_{j=1}^{N_h} y_{hi}y_{hj}}{N_h(N_h-1)} \\
&= \frac{(\sum_{i=1}^{N_h} y_{hi})^2 - (\sum_{i=1}^{N_h} y_{hi}^2)}{N_h(N_h-1)} \\
&= \frac{N_h^2\bar{Y}_h^2 - N_hE(\bar{y}_h^2) - 2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{Nh(Nh-1)}
\end{aligned}
$$

substituting the expressions of $E(y_{hi}y_{hj})$ and $E(y_{hi}^2)$ in (4.5) and simplifying we obtain

$$
\begin{aligned}
E(\bar{y}_{hT}^2) &= \frac{1}{n_h^2}\Bigg[n_h(E(y_{hi}^2) + \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{Nh}) + n_h(n_h-1) \\
&\left(\frac{N_h^2\bar{Y}_h^2 - N_hE(\bar{y}_h^2) - 2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{N_h(N_h-1)}\right)\Bigg] \\
&= \frac{E(y_h^2)}{n_h} + \frac{(n_h-1)N_h^2\bar{Y}_h^2}{n_hN_h(N_h-1)} - \frac{(n_h-1)E(\bar{y}_h^2)}{n_h(N_h-1)} \\
&+ \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{N_hn_h} \\
&- \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)(n_h-1)}{N_hn_h(N_h-1)}
\end{aligned}
$$

$$E(\bar{y}_{hT}^2) = E(y_h^2)\left[\frac{1}{n_h} - \frac{(n_h-1)}{n_h(N_h-1)}\right] + \frac{(n_h-1)N_h\bar{Y}_h^2}{n_h(N_h-1)}$$
$$+ \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{N_h n_h}\left[1 - \frac{(n_h-1)}{(N_h-1)}\right]$$

Now the variance of $Var(\bar{y}_{hT})$ becomes

$$Var(\bar{y}_{hT}) = E(\bar{y}_{hT}^2) - (E(\bar{y}_{hT}))^2$$
$$= E(y_h^2)\left[\frac{1}{n_h} - \frac{(n_h-1)}{n_h(N_h-1)}\right] + \frac{(n_h-1)N_h\bar{Y}_h^2}{n_h(N_h-1)}$$
$$+ \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{N_h n_h}\left[1 - \frac{(n_h-1)}{(N_h-1)}\right] - (E(\bar{y}_{hT}))^2$$
$$= (E(y_h^2))\left[\frac{1}{n_h} - \frac{(n_h-1)}{n_h(N_h-1)}\right] + \frac{(n_h-1)N_hE(Y_h)^2}{n_h(N_h-1)}$$
$$+ \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{N_h n_h}\left[1 - \frac{(n_h-1)}{(N_h-1)}\right] - (E(Y_h))^2$$
$$= (E(y_h^2))\left[\frac{N_h-n_h}{n_h(N_h-1)}\right] + \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{N_h n_h}\left[\frac{N_h-n_h}{N_h-1}\right]$$
$$- (E(Y_h))^2\left[\frac{N_h-n_h}{n_h(N_h-1)}\right]$$
$$= \left[\frac{N_h-n_h}{n_h(N_h-1)}\right]\left(E(y_h)^2 - (E(y_h))^2\right) + \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{N_h n_h}$$
$$\left[\frac{N_h-n_h}{N_h-1}\right]$$
$$= \left[\frac{N_h-n_h}{n_h(N_h-1)}\right]\sigma_{yh}^2 + \frac{2\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right)}{N_h n_h}\left[\frac{N_h-n_h}{N_h-1}\right]$$

$$Var(\bar{y}_{hT}) = \lambda_h S_{hy}^2 + \frac{2\lambda_h}{N_h-1}\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right) \tag{4.6}$$

Now the variance of

$Var(\bar{y}_{st.T})$=$\sum_{h=1}^{L} W_h^2 var(\bar{y}_{hT})$

$$Var(\bar{y}_{st.T}) = \sum_{h=1}^{L} W_h^2 \lambda_h \left(S_{hy}^2 + \frac{2}{N_h - 1} \left(\sum_{i=1}^{jh-1} f x_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} f x_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} f x_{ih} \right) \right) \tag{4.7}$$

The variance of$\bar{y}_{st.T}$ is always less than the variance of $\bar{y}_{st}$, i.e.,$Var(\bar{y}_{st.T}) < Var(\bar{y}_{st})$

4.3 Proposed Estimators in case of Positive Correlation

It is generally known that the ratio estimator is appropriate for population mean estimate when the study and auxiliary variable have a positive correlation.The greater value(s) of the study variable is predicted to be chosen in the sample when there is a positive correlation between the study variable and the auxiliary variable, which is achieved by choosing larger value(s) of the auxiliary variable. Likewise, it is reasonable to anticipate small values of Y if small values of X were chosen for the sample.Under the aforementioned circumstances, we define the ratio and regression estimators utilising the auxiliary variable by using transformation in stratified random sampling as

Combined Ratio Estimator,

$$\bar{y}_{CRT} = \bar{y}_{st.t} \frac{\bar{X}}{\bar{x}_{st.t}}. \tag{4.8}$$

Separate ratio estimator,

$$\bar{y}_{SRT} = \sum_{h=1}^{L} W_h \bar{y}_{h.t} \frac{\bar{X}_h}{\bar{x}_{h.t}} \tag{4.9}$$

Combined regression estimator,

$$\bar{y}_{ClrT} = \bar{y}_{st.t} + b_c(\bar{X} - \bar{x}_{st.t}) \tag{4.10}$$

Separate regression estimator,

$$\bar{y}_{SlrT} = \sum_{h=1}^{L} W_h \{\bar{y}_{h.t} + b_h(\bar{X}_h - \bar{x}_{h.t})\}. \tag{4.11}$$

4.4 Proposed estimators in case of negative correlation

However, when there is a negative correlation between the study variable and the auxiliary variable, it is predicted that the sample would have a lesser number of study variable values when the greater values of the auxiliary variable are chosen. The choice of a larger value for the study variable follows the choice of a lower value for the auxiliary variable. Under these circumstances, the suggested product and regression estimators using stratified random sampling with transformation are defined as follows:

Combined product estimator.

$$\bar{y}_{CPT} = \bar{y}_{st.t} \cdot \frac{\bar{x}_{st.t}}{\bar{X}} \tag{4.12}$$

Separate product estimator,

$$\bar{y}_{SPT} = \sum_{h=1}^{L} \bar{y}_{h.t} \cdot \frac{\bar{x}_{h.t}}{\bar{X}_h} \tag{4.13}$$

Combined regression estimator,

$$\bar{y}_{CLRT} = \bar{y}_{st.t} + b_c(\bar{X} - \bar{x}_{st.t}) \tag{4.14}$$

Separate regression estimator,

$$\bar{y}_{SLRT} = \sum_{h=1}^{L} W_h\{\bar{y}_{h.t} + b_h(\bar{X} - \bar{x}_{h.t})\}. \tag{4.15}$$

We first establish the following theorems before determining the bias and mean square errors for the proposed estimators.

Theorem 1:

For a sample of size 'n_h' (for h=1, 2, 3,..., L), units are chosen at random from a bivariate sub population of size N_h for (h=1, 2, 3,..., L), forming a stratified random sample of size $n = \sum_{h=1}^{L} n_h$ from a population of size N.If two variables are positively correlated, it can be demonstrated that the covariance between $\bar{y}_{h.t}$ and $\bar{x}_{h.t}$ is provided by;

$$COV(\bar{y}_{ht}, \bar{x}_{ht}) = \lambda_h S_{hyx} + \frac{\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} X_{ih} f x_{ih} - \sum_{i=jh+1}^{Nh} X_{hi} f x_{ih}\right) \tag{4.16}$$

Proof: Using the covariance definition,

$$Cov(\bar{y}_{hT}, \bar{x}_{hT}) = E(\bar{y}_{ht}, \bar{x}_{ht}) - E(\bar{y}_{ht})E(\bar{x}_{ht})$$

$$\begin{aligned}
E(\bar{y}_{ht}, \bar{x}_{ht}) &= E(\frac{\sum_{i=1}^{n_h} Y_{hi}}{n_h}, \frac{\sum_{i=1}^{n_h} X_{hi}}{n_h}) \\
&= \frac{1}{n_h^2} E[\sum_{i=j}^{n_h} Y_{hi} X_{hj} + \sum_{i \neq j}^{n_h} Y_{hi} X_{hj}] \\
E(\bar{y}_{ht}, \bar{x}_{ht}) &= \frac{1}{n_h^2}[n_h \cdot E_{i=j} Y_{ih} X_{jh} + n_h(n_h - 1) E_{i \neq j} Y_{hi} X_{jh}]
\end{aligned}$$

Consider the expression $E_{i=j}(Y_{hi}X_{hj})$

$$E_{i=j}(Y_{hi}X_{hj}) = \frac{\sum_{i=1}^{N_h} Y_{hi}X_{hj}}{N_h}$$
$$= \frac{1}{N_h}\{(Y_{h1} + fx_1h)X_{h1} + (Y_{h2} + fx_2h)X_{h2} + ... + (Y_{hj-1} + fx_{hj-1})X_{hj-1}$$
$$+ Y_{jh}X_{jh} + (Y_{hj+1} - fx_{hj+1})X_{hj+1} + ... + (Y_{Nh-1} - fx_{2h})X_{Nh-1} + (y_{Nh} - fx_{1h})X_{Nh}\}$$

$$= \frac{\sum_{i=1}^{Nh} yhiX_{hi} + \sum_{i=1}^{jh-1} fx_{ih}X_{ih} - \sum_{i=jh+1}^{Nh} X_{hi}fx_{ih}}{Nh}$$
$$E_{i=j}(Y_{hi}X_{hj}) = \frac{\sum_{i=1}^{N_h} Y_{hi}X_{hj}}{N_h} + \frac{\left(\sum_{i=1}^{jh-1} X_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} X_{hi}fx_{ih}\right)}{Nh}$$

Consider Now

$$E_{i\neq j}(Y_{hi}X_{hj}) = \frac{\sum_{i\neq j} Y_{hi}X_{hj}}{N_h(N_h-1)}$$
$$= \frac{\sum_{i=1}^{N_h} Y_{hi} \sum_{j=1}^{Nh} X_{hj} - \sum_{i=j} Y_{hi}X_{hj}}{N_h(N_h-1)}$$
$$= \frac{\sum_{i=1}^{Nh} Y_{hi} \sum_{j=1}^{Nh} X_{hj}}{N_h(N_h-1)} - \frac{\sum_{i=j} Y_{hi}X_{hj}}{N_h(N_h-1)}$$

In the equation above, if we replace the formula $\sum_{i=j} Y_{hi}X_{hj}$, we obtain

$$E_{i\neq j}(Y_{hi}X_{hj}) = \frac{\sum_{i=1}^{Nh} Y_{hi} \sum_{j=1}^{Nh} X_{hj}}{N_h(N_h-1)} - \frac{1}{Nh(N_h-1)}\left(\sum_{i=1}^{Nh} Y_{hi}X_{hj} + \sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{Nh} X_{hi}fx_{ih}\right)$$

$$E_{i\neq j}(Y_{hi}X_{hj}) = \frac{\sum_{i=1}^{Nh} Y_{hi} \sum_{j=1}^{Nh} X_{hj}}{N_h(N_h-1)} - \frac{\sum_{i=1}^{Nh} Y_{hi}X_{hj}}{Nh(N_h-1)} - \frac{\sum_{i=1}^{jh-1} fx_{hi}X_{ih}}{Nh(Nh-1)} + \frac{\sum_{i=jh+1}^{Nh} X_{hi}fx_{ih}}{Nh(Nh-1)}$$

Substituting expressions of the form $E_{i=j}(Y_{hi}X_{hj})$ and $E_{i\neq j}(Y_{hi}X_{hj})$ in the equation of

$Cov(\bar{y}_{ht}, \bar{x}_{ht})$ which results in

$$Cov(\bar{y}_{ht}, \bar{x}_{ht}) = \frac{1}{n_h^2}\left[n_h\left(\frac{\sum_{i=1}^{Nh} Y_{hi}X_{hj}}{N_h} + \frac{\left(\sum_{i=1}^{jh-1} X_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} X_{hi}fx_{ih}\right)}{Nh}\right) + \right.$$

$$\left. n_h(n_h-1)\left(\frac{\sum_{i=1}^{Nh} Y_{hi}\sum_{j=1}^{Nh} X_{hj}}{N_h(N_h-1)} - \frac{\sum_{i=1}^{Nh} Y_{hi}X_{hj}}{Nh(N_h-1)} - \frac{\sum_{i=1}^{jh-1} fx_{hi}X_{ih}}{N_h(N_h-1)} + \frac{\sum_{i=jh+1}^{N_h} X_{hi}fx_{ih}}{Nh(Nh-1)}\right)\right]$$

$$- \bar{Y}_h\bar{X}_h$$

$$= \frac{\sum_{i=1}^{Nh} Y_{hi}X_{hi}}{N_h n_h} + \frac{\sum_{i=1}^{jh-1} fx_{hi}X_{ih}}{n_h N_h} - \frac{\sum_{i=jh+1}^{N_h} X_{hi}fx_{ih}}{N_h n_h} + \frac{\sum_{i=1}^{Nh} Y_{hi}\sum_{j=1}^{Nh} X_{hj}(n_h-1)}{N_h n_h(N_h-1)}$$

$$- \frac{\sum_{i=1}^{Nh} Y_{hi}X_{hi}(n_h-1)}{N_h n_h(n_h-1)} - \frac{\sum_{i=1}^{jh-1} fx_{hi}X_{ih}(n_h-1)}{n_h N_h(N_h-1)} + \frac{\sum_{i=j+1}^{N_h} fx_{hi}X_{ih}(n_h-1)}{n_h N_h(N_h-1)} - \bar{Y}_h\bar{X}_h$$

$$= \frac{\sum_{i=1}^{Nh} Y_{hi}X_{hi}}{N_h n_h}\left(\frac{N_h-n_h}{N_h-1}\right) + \frac{\sum_{i=1}^{jh-1} fx_{hi}X_{ih}}{n_h N_h}\left(\frac{N_h-n_h}{N_h-1}\right) - \frac{\sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}}{n_h N_h}\left(\frac{N_h-n_h}{N_h-1}\right)$$

$$+ \frac{\bar{Y}_h\bar{X}_h(n_h-1)N_h}{n_h(N_h-1)} - \bar{Y}_h\bar{X}_h$$

$$= \frac{\sum_{i=1}^{Nh} Y_{hi}X_{hi}}{N_h n_h}\left(\frac{N_h-n_h}{N_h-1}\right) - \bar{Y}_h\bar{X}_h\left(1 - \frac{(n_h-1)N_h}{(N_h-1)n_h}\right) + \frac{1}{N_h n_h}\left(\frac{N_h-n_h}{N_h-1}\right)$$

$$\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=j+1}^{N_h} fx_{hi}X_{ih}\right)$$

$$= \left(\frac{N_h-n_h}{(N_h-1)n_h}\right)\left[\frac{\sum_{i=1}^{Nh} Y_{hi}X_{hi}}{N_h} - \bar{Y}_h\bar{X}_h\right] + \frac{1}{N_h n_h}\left(\frac{N_h-n_h}{N_h-1}\right)$$

$$\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=j+1}^{N_h} fx_{hi}X_{ih}\right)$$

$$= \left(\frac{N_h-n_h}{(N_h-1)n_h}\right)\sigma_{yhx} + \frac{1}{N_h n_h}\left(\frac{N_h-n_h}{N_h-1}\right)\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right)$$

$$= \left(\frac{N_h-n_h}{(N_h-1)n_h}\right)\left(\frac{N_h-1}{N_h}\right)S_{hyx} + \frac{1}{N_h n_h}\left(\frac{N_h-n_h}{N_h-1}\right)\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right)$$

now the covariance in case of positive correlation is

$$Cov(\bar{y}_{ht}, \bar{x}_{ht}) = \lambda_h S_{hyx} + \frac{\lambda_h}{N_h-1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right) \tag{4.17}$$

Theorem 2:

For a sample of size 'n_h' (for h=1, 2, 3,..., L), units are chosen at random from a bivariate sub population of size N_h for (h=1, 2, 3,..., L), forming a stratified random sample of size $n = \sum_{h=1}^{L} n_h$ from a population of size N.If two variables are negetively correlated, it can be demonstrated that the covariance between $\bar{y}_{h.t}$ and $\bar{x}_{h.t}$ is provided by;

$$Cov(\bar{y}_{ht}, \bar{x}_{ht}) = \lambda_h S_{hyx} - \frac{\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right) \tag{4.18}$$

Relative Errors terms:

The relative error terms are defined for the purpose of deriving the biases and Mean Square Error of the proposed estimator:

Let

$$e_{st.T_1} = \frac{\bar{y}_{st.t} - \bar{Y}}{\bar{Y}}$$

$$e_{st.T_2} = \frac{\bar{x}_{st.t} - \bar{X}}{\bar{X}}$$

$$e_{h.T_1} = \frac{\bar{y}_{h.t} - \bar{Y}_h}{\bar{Y}_h}$$

$$e_{h.T_2} = \frac{\bar{x}_{h.t} - \bar{X}_h}{\bar{X}_h}$$

Using theorems 1 and 2, expectations of relative error terms up to the first degree of approximation may be expressed as

$$E(e_{st.T_1}) = E(e_{st.T_2}) = 0$$

$$E(e_{h.T_1}) = E(e_{h.T_2}) = 0$$

$$E(e^2_{st.T_1}) = \frac{1}{\bar{Y}^2}\sum_{h=1}^{L} W_h^2 \lambda_h \left[S^2_{hy} + \frac{2}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx^2_{ih} + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih}\right)\right]$$

$$E(e^2_{st.T_2}) = \frac{1}{\bar{X}^2}\sum_{h=1}^{L} W_h^2 \lambda_h S^2_{hx}$$

$$E(e^2_{h.T_1}) = \frac{\lambda_h}{\bar{Y}_h^2}\left[S^2_{hy} + \frac{2}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx^2_{ih} + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih}\right)\right]$$

$$E(e^2_{h.T_2}) = \frac{1}{\bar{X}_h^2}\sum_{h=1}^{L} W_h^2 \lambda_h S^2_{hx}$$

$$E(e_{st.T_1}, e_{st.T_2}) = \frac{1}{\bar{Y}\bar{X}}\sum_{h=1}^{L} W_h^2 \lambda_h \left[S_{hyx} + \frac{1}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right)\right]$$

$$E(e_{h.T_1}, e_{h.T_2}) = \frac{\lambda_h}{\bar{Y}_h\bar{X}_h}\left[S_{hyx} + \frac{1}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right)\right]$$

4.5 Proposed Combined Estimators.

We derived the bias and mean square errors of the combined ratio, product and regression estimators.

Proposed combined ratio estimator:

Rewriting the expression $\bar{y}_{CRT}$ in terms of $e_i's$, allows us to get the expressions for bias and MSE

$$\bar{y}_{CRT} = \bar{Y}(1 + e_{st.T_1})(1 + e_{st.T_2})^{-1}$$

When we expand and rearrange the aforementioned statement up to the first degree of approximation, we obtain

$$(\bar{y}_{RCT} - \bar{Y}) \approx \bar{Y}(e_{st.T1} - e_{st.T2} + e_{st.T2}^2 - e_{st.T1}e_{st.T2}) \quad (4.19)$$

By applying the expectation on the aforementioned formula, the bias of $\bar{y}_{RCT}$ is as follows;

$$Bias(\bar{y}_{RCT}) \approx \frac{\bar{Y}}{\bar{X}^2}\sum_{h=1}^{L} W_h^2\lambda_h S_{hx}^2 - \frac{1}{\bar{X}}\sum_{h=1}^{L} W_h^2\lambda_h\left[S_{yx} + \frac{1}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right)\right] \quad (4.20)$$

By taking the square on both sides of (4.19), preserving just the terms up to the first degree of approximation, and computing the MSE, we may obtain

$$(\bar{y}_{RCT} - \bar{Y})^2 \approx \bar{Y}^2[e_{st.T1}^2 + e_{st.T2}^2 - 2e_{st.T1}e_{st.T2}]$$

We obtain MSE by applying the expectations on the two sides of the aforementioned equation.

$$MSE(\bar{y}_{RCT}) \approx \sum_{h=1}^{L} W_h^2\lambda_h\left(S_{hy}^2 + \delta^2 S_{hx}^2 - 2\delta S_{hxy}\right) + \frac{2\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih}fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi}fx_{ih}\right) - \frac{2\delta\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right) \quad (4.21)$$

where $\delta = \bar{Y}/\bar{X}$

Proposed combined product estimator:

Similarly, we can construct the following formulas for the bias and mean square error of the combined product estimator;

$$Bias(\bar{y}_{CPT}) \approx \frac{1}{\bar{X}}\sum_{h=1}^{L} W_h^2\lambda_h\left[S_{hyx} + \frac{1}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right)\right] \quad (4.22)$$

and MSE of the $\bar{y}_{CPT}$ is given by

$$MSE(\bar{y}_{CPT}) \approx \sum_{h=1}^{L} W_h^2 \lambda_h \left(S_{hy}^2 + \delta^2 S_{hx}^2 + 2\delta S_{hxy} \right) + \frac{2\sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih} \right) - \frac{2\delta \sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} fx_{hi} X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi} X_{ih} \right) \tag{4.23}$$

Proposed combined Regression estimator in case of positive correlation:

Similarly, we can construct the following formula for the mean square error of the combined linear regression estimator;

$$MSE(\bar{y}_{ClrT}) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 [1 - \rho_c^2] + \frac{2\sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih} \right) - \frac{2b_c \sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} fx_{hi} X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi} X_{ih} \right)$$

Proposed combined Regression estimator in case of negative correlation:

$$MSE(\bar{y}_{ClrT}) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 [1 - \rho_c^2] + \frac{2\sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih} \right) + \frac{2b_c \sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} fx_{hi} X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi} X_{ih} \right)$$

and the general expression is as follows:

$$MSE(\bar{y}_{ClrT}) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 [1 - \rho_c^2] + \frac{2\sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih} \right) - \frac{2|b_c| \sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} fx_{hi} X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi} X_{ih} \right) \tag{4.24}$$

where $b_c = \sum_{h=1}^{L} W_h^2((N_h - n_h)/N_h n_h) S_{hyx} / \sum_{h=1}^{L} W_h^2((N_h - n_h)/N_h n_h) S_{hx}^2$

Across all strata, b_c represents the population regression coefficient.

where $\rho_c = \sum_{h=1}^{L} W_h^2 \lambda_h S_{hyx} / \sqrt{\sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 \sum_{h=1}^{L} W_h^2 \lambda_h S_{hx}^2}$

p_c represents the population correlation coefficient between study and auxiliary variable.

4.6 Proposed Seperate Estimators.

We derived the bias and mean square errors of the seperate ratio, product and regression estimators.

Proposed seperate ratio estimator:

Rewriting the expression $\bar{y}_{SPT}$ in terms of $e_i's$, allows us to get the expressions for bias and MSE

$$\bar{y}_{RST} = \sum_{h=1}^{L} W_h \bar{Y}_h (1 + e_{hT1})(1 + e_{h.hT2})^{-1}$$

When we expand and rearrange the aforementioned statement up to the first degree of approximation,we obtain

$$(\bar{y}_{RST} - \bar{Y}_h) \approx \sum_{h=1}^{L} W_h \bar{Y}_h (e_{hT1} - e_{hT2} + e_{hT2}^2 - e_{h.T1} e_{h.T2}) \tag{4.25}$$

By applying the expectation on the aforementioned formula, the bias of $\bar{y}_{RCT}$ is as follows;

$$Bias(\bar{y}_{RST}) \approx \sum_{h=1}^{L} W_h \bar{Y}_h \left[\frac{\lambda_h S_{hx}^2}{\bar{X}_h^2} - \frac{\lambda_h}{\bar{Y}_h \bar{X}_h} \left(S_{hyx} + \frac{1}{N_h - 1} \left(\sum_{i=1}^{jh-1} f x_{hi} X_{ih} - \sum_{i=jh+1}^{N_h} f x_{hi} X_{ih} \right) \right) \right] \tag{4.26}$$

By taking the square on both sides of (4.25), preserving just the terms up to the first degree of approximation, and computing the MSE, we may obtain

$$(\bar{y}_{RST} - \bar{Y}_h)^2 \approx \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 [e_{hT1}^2 + e_{hT2}^2 - 2 e_{hT1} e_{hT2}]$$

We obtain MSE by applying the expectations on the two sides of the aforementioned equation.

$$MSE(\bar{y}_{RST}) \approx \sum_{h=1}^{L} W_h^2 \lambda_h \left(S_{hy}^2 + \delta_h^2 S_{hx}^2 - 2\delta_h S_{hxy} \right) + \frac{2 \sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} f x_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} f x_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} f x_{ih} \right) - \frac{2\delta_h \sum_{h=1}^{L} W_h^2 \lambda_h}{N_h - 1} \left(\sum_{i=1}^{jh-1} f x_{hi} X_{ih} - \sum_{i=jh+1}^{N_h} f x_{hi} X_{ih} \right) \tag{4.27}$$

Proposed seperate product estimator:

Similarly, we can construct the following formulas for the bias and mean square error of the seperate product estimator;

$$Bias(\bar{y}_{SPT}) \approx \frac{\bar{Y}_h}{\bar{X}_h \bar{Y}_h} \sum_{h=1}^{L} W_h \lambda_h \left[S_{hyx} - \frac{1}{N_h - 1} \left(\sum_{i=1}^{jh-1} f x_{hi} X_{ih} - \sum_{i=jh+1}^{N_h} f x_{hi} X_{ih} \right) \right] \tag{4.28}$$

and MSE of the $\bar{y}_{CPT}$ is given by

$$MSE(\bar{y}_{SPT}) \approx \sum_{h=1}^{L} W_h^2\lambda_h\left(S_{hy}^2 + \delta_h^2 S_{hx}^2 + 2\delta_h S_{hxy}\right) + \frac{2\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih}\right) - \frac{2\delta_h\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right) \quad (4.29)$$

Proposed seperate linear Regression estimator in case of positive correlation:

Similarly, we can construct the following formula for the variance of the seperate linear regression estimator;

$$MSE(\bar{y}_{SlrT}) = \sum_{h=1}^{L} W_h^2\lambda_h S_{hy}^2[1-\rho_h^2] + \frac{2\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih}\right) - \frac{2b_h\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right)$$

Proposed separate linear Regression estimator in case of negative correlation:

Similarly, we can construct the following formula for the variance of the separate linear regression estimator;

$$MSE(\bar{y}_{SlrT}) = \sum_{h=1}^{L} W_h^2\lambda_h S_{hy}^2[1-\rho_h^2] + \sum_{h=1}^{L}\frac{2W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih}\right) + \frac{2b_h\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right)$$

and the general expression is as follows:

$$MSE(\bar{y}_{SlrT}) = \sum_{h=1}^{L} W_h^2\lambda_h S_{hy}^2[1-\rho_h^2] + \frac{2\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{ih}^2 + \sum_{i=1}^{jh-1} y_{ih} fx_{ih} - \sum_{i=jh+1}^{Nh} y_{hi} fx_{ih}\right) - \frac{2|b_c|\sum_{h=1}^{L} W_h^2\lambda_h}{N_h - 1}\left(\sum_{i=1}^{jh-1} fx_{hi}X_{ih} - \sum_{i=jh+1}^{N_h} fx_{hi}X_{ih}\right) \quad (4.30)$$

where $b_h = S_{hyx}/S_{hx}^2$ represents the population regression coefficient in the h_{th} stratum.

where $\rho_h = \dfrac{S_{hyx}}{S_{hy}S_{hx}}$ represents the correlation coefficient between the study variable and the auxiliary variable in h_{th} stratum.

4.7 Real data sets for empirical comparison:

We investigated three real data sets in this part for the purpose of numerically comparing the current and recommended estimators.while tables 4.1, 4.2 and 4.3 display the summary statistics for Data 1, 2, and 3.

Data 1:(Source : Singh and Mangat (2013), p.219)

y=Milch cows during March 1993

x=Milch cows during March 1990.

Table 4.1: Summary Statistics for Data 1

	N_h	n_h	W_h	$\bar{X}_h$	$\bar{Y}_h$	S^2_{hx}	S^2_{hy}
stratum 1	7	4	0.29	15.29	17.43	20.90	17.62
stratum 2	12	6	0.5	17.25	19.58	30.20	16.63
stratum 3	5	2	0.21	17.8	20.6	10.70	13.30
	ρ_{hyx}	C_{yh}	C_{xh}	λ_h	S_{hyx}	δ_h	b_h
stratum 1	0.77	0.24	0.30	0.11	13	1.14	0.70
stratum 2	0.41	0.21	0.32	0.08	18	1.14	0.30
stratum 3	0.49	0.18	0.18	0.3	8	1.16	0.55

Data 2:(Source : Singh and Mangat (2013), p.180)

y= Orange yields for the current year

x=Orange yields for the previous year

Table 4.2: Summary Statistics for Data 2

	N_h	n_h	W_h	$\bar{X}_h$	$\bar{Y}_h$	S^2_{hx}	S^2_{hy}
stratum 1	12	6	0.5	92.77	96.79	157.16	136.86
stratum 2	12	6	0.5	77.99	86.99	496.93	377.52
	ρ_{hyx}	C_{yh}	C_{xh}	λ_h	S_{hyx}	δ_h	b_h
stratum 1	0.95	0.12	0.14	0.08	138.85	1.04	0.88
stratum 2	0.97	0.22	0.29	0.08	421.58	1.12	0.85

Data 3:(Source : Singh and Mangat (2013), p.208)

y= number of refrigerators sold during last summer (LS)

x=expected sale for the current summer (CS)

Table 4.3: Summary Statistics for Data 3

	N_h	n_h	W_h	$\bar{X}_h$	$\bar{Y}_h$	S^2_{hx}	S^2_{hy}
stratum 1	14	5	0.33	76.21	79.36	211.10	166.71
stratum 2	9	3	0.21	58.11	59.44	197.11	174.28
stratum 3	12	4	0.29	69.08	76.67	192.99	226.61
stratum 4	7	2	0.17	63.71	76.67	408.09	170.62
	ρ_{hyx}	C_{yh}	C_{xh}	λ_h	S_{hyx}	δ_h	b_h
stratum 1	0.79	0.16	0.19	0.13	148.76	1.04	0.70
stratum 2	0.87	0.22	0.24	0.22	161.19	1.02	0.82
stratum 3	0.92	0.20	0.20	0.17	192.21	1.11	0.99
stratum 4	0.91	0.17	0.32	0.36	238.96	1.20	0.59

Based on extreme data, Table 4.4 displays the percentage relative efficiency of the proposed and MEAN PER UNIT estimators. For this objective, three populations have been chosen. The suggested estimator outperforms its competitors,according on the results of these data sets.Because the correlation between the study and the auxiliary variable is positive in the above three data sets, the product estimators do not produce satisfactory results.For data sets 2, the efficiency of the proposed estimators $\bar{y}_{ClrT}$ and $\bar{y}_{SLRT}$ are extremely high.

4.8 Conclusion:

In this chapter, we expanded the previous chapter's approach and suggested a finite population mean estimator based on transformation under stratified random sampling. Bias and MSE of recommended estimators were derived up to the first degree of approximation using stratified random sampling. The proposed estimator works well in theory under particular conditions, and numerical findings for certain populations support the theoretical results.

Table 4.4: PREs of the different Estimators

Estimators	Data 1	Data 2	Data 3
$\bar{y}_{st}$	100	100	100
$\hat{\bar{Y}}_{CR}$	66.9929	782.7733	257.2536
$\hat{\bar{Y}}_{CRT}$	84.7559	1033.3070	266.2523
$\hat{\bar{Y}}_{CP}$	23.9143	20.7542	22.1360
$\hat{\bar{Y}}_{CPT}$	25.8481	20.8885	22.2005
$\hat{\bar{Y}}_{Clr}$	102.9663	1332.9210	322.4822
$\hat{\bar{Y}}_{ClrT}$	132.6210	1503.8850	374.8214
$\hat{\bar{Y}}_{SR}$	67.2622	699.7250	217.6535
$\hat{\bar{Y}}_{SRT}$	85.2699	897.8404	227.5351
$\hat{\bar{Y}}_{SP}$	23.9526	20.2996	20.9235
$\hat{\bar{Y}}_{SPT}$	25.9005	20.4304	21.0113
$\hat{\bar{Y}}_{Slr}$	111.1688	1368.1050	368.6722
$\hat{\bar{Y}}_{SLRT}$	139.8274	1510.3960	425.4589

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Survey sampling is used by survey statisticians to accurately estimate the finite population mean. This aim is achievable with efficient and unbiased estimators. The judicious use of the auxiliary information may increase an estimator's efficiency and accuracy. We focused on using the auxiliary variable and transformation to build novel estimators of the finite population mean under simple and stratified random sampling.

In this research we used the idea of antithetic variable technique in transforming the study variable to reduce its variability, thereby estimating the finite population mean accurately. The basic objective of variance reduction approaches is to reduce the variability of estimates derived from sampling procedures, resulting in more accurate findings with fewer samples. Through the use of correlations between sample pairs, the antithetic variable method achieves this.

To construct efficient estimators of the population mean, previous research relied solely on two extreme values, namely the population minimum and maximum values. However, more extreme observations in a population may affect the estimation of the population mean. Similarly, previous research has expanded the concept to account for a population's two lowest and two maximum observations. Inclusion of more extreme points complicates the estimator and its statistical characteristics derivation. The new study has expanded the concept to account for all of a population observations.

In Chapter 3, we provided new improved mean per unit, ratio, product, and regression type estimators for estimating the finite population mean using simple random sampling. Up to the first level of approximation, expressions for biases and MSEs have been derived. The newly suggested estimators were theoretically compared to other competing estimators that account for transformation. Conditions that prove the superiority of the proposed estimators over the existing estimators have been derived. Various real data sets were utilised to demonstrate the performance of the proposed estimators.

The concept created in stratified random sampling, which is used to reduced population variability for better estimate, was expanded upon in Chapter 4. Biases and MSEs of the suggested estimators were derived up to the first degree of approximation. The suggested estimator outperforms the current estimators, according to theoretical and numerical evidence. A few real data sets have been included as part of the numerical demonstration.

An estimator's effectiveness is increased by using the auxiliary data. Additionally, The use of antithetic variable techniques could further enhance the efficiency of an estimator. It is advised to employ the proposed estimators for population mean estimate.

5.2 Future work

- The current research may be expanded to families of estimators used for estimation of the population parameters.
- This study might be expanded to include different sampling techniques, such as probability proportional to size, etc.

.

References

Ahmad, S. and Shabbir, J. (2018). Use of extreme values to estimate finite population mean under pps sampling scheme. *Journal of Reliability and Statistical Studies*, 11(2):99–112.

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.

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.

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.

Khan, M. and Shabbir, J. (2013). Some improved ratio, product, and regression estimators of finite population mean when using minimum and maximum values. *The Scientific World Journal*, 2013(1):1–7.

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.

Mohanty, S. and Sahoo, J. (1995). A note on improving the ratio method of estimation through linear transformation using certain known population parameters. *Sankhyā: The Indian Journal of Statistics, Series B*, 57(1):93–102.

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

Mursala, K., Ullah, S., Al-Hossain, A. Y., and Bashir, N. (2015). Improved ratio-type estimators using maximum and minimum values under simple random sampling scheme. *Hacettepe Journal of Mathematics and Statistics*, 44(4):923–931.

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

Robson, D. (1957). Applications of multivariate polykays to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*, 52(280):511–522.

Särndal, C.-E. (1972). Sample survey theory vs. general statistical theory: Estimation of the population mean. *International Statistical Review/Revue Internationale De Statistique*, 40(1):1–12.

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

Shoaib, A., Manzoor, K., and Shabbir, J. (2018). Using extreme values and fractional raw moments for mean estimation in stratified random sampling. *Hacettepe Journal of Mathematics and Statistics*, 47(2):383–402.

Singh, R. and Mangat, N. S. (2013). *Elements of survey sampling*, volume 15. Springer Science & Business Media.