Estimation of Population Variance Using the Coefficient of Kurtosis and Median of an Auxiliary Variable Under Simple Random Sampling

Tonui Kiplangat Milton
MS300-0006/2016

A Thesis Submitted to Pan African University, Institute of Basic Sciences, Technology and Innovation in Partial fulfillment of the requirement for the award of Master of Science in Mathematics(Statistics Option) of the Pan African University

2018

DECLARATION

This thesis is my original work and has not been submitted to any other University for any award.

Tonui Kiplangat Milton

MS300-0006/2016

Signature..Date..........................

This thesis has been submitted for examination with our approval as University supervisors

Prof. Romanus Otieno Odhiambo

Department of Statistics and Actuarial Sciences

Jomo Kenyatta University Of Agriculture and Technology

P.O. Box 62000-00200, Nairobi, Kenya.

Signature..Date...................................

Prof. George Otieno Orwa

Department of Statistics and Actuarial Sciences

Jomo Kenyatta University Of Agriculture and Technology

P.O. Box 62000-00200, Nairobi, Kenya.

Signature..Date...................................

ACKNOWLEDGMENTS

First, I would like to thank God, The Almighty, for the gift of life, grace and strength that has enabled me reach this far.

Secondly, I would like to express my sincere gratitude to Prof. Romanus Odhiambo for the continuous unwavering support and encouragement in my MSc. study and related research. His guidance, motivation and immense knowledge helped me in all the time of research and writing of this thesis. I could not have imagined having a better guide and mentor for my MSc. study.

Besides my guide, I would like to thank Prof. George Orwa for his insightful comments and encouragement. My sincere thanks also goes to African Union Commission, who provided me an opportunity to undertake my graduate studies. Without their precious support it would not be possible to conduct this research.

Last but not least, I would like to thank my family: my mother and my siblings for supporting me spiritually throughout writing this proposal and my life in general. Special thanks to my late father who tirelessly supported me in my education.

Contents

DEDICATION

I dedicate this work to my friends and siblings who have been affected in every way possible by this quest. Thank you. My love for you all can never be quantified. GOD BLESS YOU.

List of Tables

List of Appendices

ABSTRACT

In this study we have proposed a modified ratio type estimator for population variance of the study variable y under simple random sampling without replacement making use of coefficient of kurtosis and median of an auxiliary variable x. The estimator's properties have been derived up to first order of Taylor's series expansion. The efficiency conditions are derived theoretically under which the proposed estimator performs better than existing estimators. Empirical studies have been done using real populations to demonstrate the performance of the developed estimator in comparison with the existing estimators. The proposed estimator as illustrated by the empirical studies performs better than the existing estimators under some specified conditions i.e. it has the smallest Mean Squared Error and the highest Percentage Relative Efficiency. The proposed estimator is therefore suitable to be applied to situations in which the variable of interest has a positive correlation with the auxiliary variable.

CHAPTER 0NE

1 INTRODUCTION

1.1 Background of the study

It is notable that the appropriate use of auxiliary information in probability sampling designs yields considerable reduction in the variance of the estimators of population parameters namely, population mean, median,variance,regression coefficient and population correlation coefficient. Cochran (1940) was the first to show the contribution of known auxiliary information in improving the efficiency of the estimator of the population mean $\bar{Y}$ in survey sampling.

In this study we are interested in the estimation of population variance using known auxiliary information under simple random sampling without replacement(SRSWOR) sampling scheme. The precision of estimators under this situation is always increased, the ratio, product and regression estimators gives better outcome than those of simple random sampling.

Variance estimation has become a priority as many surveys require that the quality of the statistics be assessed. Sampling variance which is an estimate of the population variance is a key indicator of quality in sample surveys and estimation. Variance helps the user to draw more accurate conclusions about the statistics produced and it is also important for the design and estimation phases of surveys.

However due to the complexity of the methods used for the design and analysis of the survey like the sampling design, weighting, and the type of estimators involved the calculations are not straightforward.

Variance estimation in sample survey is crucial for future surveys either for determination of sample size or stratification. Usually in the estimation of the finite population mean survey data is used, however in many situations the mean may not be an appropriate average since it fluctuates from large to small observations or outliers in a set of data. Hence the need for the population variance to overcome the difficulty.

On regular instances we encounter surveys in which an auxiliary variable x is relatively cheap (with regard to time and money) to observe than the study variable y. Use of auxiliary information can increase the precision of an estimator when the study variable y is highly correlated with auxiliary variable x. In reality such situations do occur when information is available in the form of auxiliary variable, which is highly correlated with study variable, for example:

(a) Sex and height of the persons,

(b) Amount of milk produced and a particular breed of the cow,

(c) Amount of yield of wheat crop and a particular variety of wheat etc.

(d) Number of trees in an orchard and the yield of fruits.

Many authors have come up with more precise estimators by employing prior knowledge of certain population parameter(s). For instance Searls (1964) used coefficient of variation of study variable at estimation stage. In practice however, this coefficient of variation is seldom known. Motivated by Searls (1964) work, Sen (1978), Sisodia and Dwivedi (1981) and Upadhyaya and Singh (1984) used the known coefficient of variation of the auxiliary variable for estimating population mean of study variable in ratio method of estimation. Reasoning along the same path Hirano et al. (1973) used the prior value of coefficient of kurtosis in estimating the population variance of the study variable y.

Kurtosis in most cases is not reported or used in many research articles, in spite of the fact that virtually speaking every statistical package provides a measure of kurtosis. This maybe attributed to the likelihood that kurtosis is not well understood or its importance in various aspects of statistical analysis has not been explored fully. Kurtosis can simply be expressed as

$$\kappa = \frac{E(x-\mu)^4}{(E(x-\mu)^2)^2} = \frac{\mu^4}{\sigma^4}$$

where E is the expectation operator, μ is the mean, μ^4 is the fourth moment about the mean and σ is the standard deviation.

Median being the middlemost value in a distribution (when the values are arranged in ascending or descending order) has the advantage of being less affected by the outliers and skewed data, thus is preferred to the mean especially when the distribution is not symmetrical. We can therefore utilize the median and the coefficient of kurtosis of the auxiliary variable to derive a more precise ratio type estimator for population variance.

1.2 Problem Statement

The theory and applications of survey sampling have grown tremendously in the last 7 decades. Many authors have considered the estimation of population variance, from the initial works of Evans (1951), Hansen et al. (1953), Isaki (1983), Das and Tripathi (1978), Srivastava and Jhajj (1980), Upadhyaya and Singh (1983), Upadhyaya and Singh (1999), Singh (2001), Singh et al. (2003), Kadilar and Cingi (2006), Gupta and Shabbir (2008), Grover (2010), Singh et al. (2011), Khan and Shabbir (2013a), and recently Yadav et al. (2016). High number of surveys are now carried out every year in the various governmental agencies, the private sector and the academic community, both in Kenya and the entire world at large. For instance the nationwide surveys about health care, economic activity, poverty(people's wellbeing), energy usage and unemployment; market researches and public opinion surveys; and surveys associated with academic research studies.

In the current world, survey sampling touch almost every field of scientific study, including demography, education, energy, transportation, health care, economics, forestry, sociology, politics and so on. In fact it is not an exaggeration to say that much of the data undergoing any form of statistical analysis are collected in surveys. It is imperative to note that as the number and uses of sample surveys increase, so is the need for methods of analyzing and interpreting the resulting data. A central requirement for nearly all forms of analysis and indeed the prime requirement of good survey practice, is that measure of precision be provided for each estimate derived from the survey data.

The most common and widely used measure of precision is the variance of the survey estimator. In reality however, population variances are always not known but must be estimated from the survey data themselves. The problem of constructing such estimate of the population variance which is more efficient using both the coefficient of kurtosis and median has not been explored. As a result of the necessity to offer solutions to fill the gap in methodological problems encountered in the estimation of population variance of the study variable, this study is undertaken utilizing the population coefficient of kurtosis and the median of the auxiliary variable.

1.3 Justification of the study

The approach employed in the development of proposed estimator is numerical studies and existing literature. We not only propose a theoretically more efficient population variance estimator but also test its efficiency using real data from natural population existing in literature; as a consequence of a number of factors that a good estimator for the population variance estimator should possess; numerical studies strengthens, "puts flesh on the bones" of a survey estimator.

1.4 Objectives of the study

1.4.1 General Objective

The main objective of this study is to estimate the population variance using the coefficient of kurtosis and median of an auxiliary Variable under simple random sampling.

1.4.2 Specific Objectives

The above general objective is accomplished by fulfilling the following research objectives:

1. To develop a modified ratio type population variance estimator using the coefficient of kurtosis and median of the auxiliary variable.

2. To evaluate the bias and Mean Squared Error (MSE) of the proposed modified ratio type population variance estimator.

3. To perform empirical study to assess the performance of the proposed estimator vis-a-vis the existing estimators using Percentage Relative Efficiencies (PREs).

1.5 Significance of the study

The mathematical results obtained in this study adds value and knowledge to the field of sample surveys, a new more efficient modified ratio type population variance estimator has been developed making useful use of coefficient of kurtosis and the median of the auxiliary variable. Further to the society considering the fact that mathematics plays an important role in our day to day activities that involve statistical analyses. The greater demand for more precision in the use of survey data justifies the need to develop more efficient estimators with high precision. Thus, using the approach of estimation derived from this study achieves better results than the existing estimators.

1.6 The Scope of the study

This study focused on estimation of population variance under simple random sampling utilizing the knowledge of known coefficient of kurtosis and median of the auxiliary variable. Assuming simple random sampling, Bias and Mean squared error has been obtained up to first order of approximations. Efficiency comparison of existing and proposed modified ratio type population variance estimators using the MSEs has been implemented on the data from the natural populations existing in the literature using percent relative efficiency (PRE).

CHAPTER TWO

2 LITERATURE REVIEW

2.1 Existing Population Variance Estimators

In this section we have reviewed some of the existing estimators available in literature which will help in the construction and development of the proposed estimator. When there is no auxiliary information the usual unbiased estimator to the population variance of the study variable is

$$t_1 = s_y^2 \tag{1}$$

Population variance, S_y^2 estimation using auxiliary information was considered by Isaki (1983), and proposed ratio type population variance estimator, given by

$$t_2 = s_y^2 \frac{S_x^2}{s_x^2} \tag{2}$$

Usage of prior value of coefficient of kurtosis in estimating population variance of study variable y was first done by Hirano et al. (1973). Later, the coefficient of kurtosis was used by Sen (1978), Upadhyaya and Singh (1984), Searls and Intarapanich (1990) in the estimation of population mean of study variable.

Srivastava and Jhajj (1980), proposed a general class of ratio type estimators for estimating the finite population variance S_y^2 as

$$\hat{S}^2{}_{SJ} = s^2 G(u, v) \tag{3}$$

where $u = \frac{\bar{x}}{\bar{X}}$, $v = \frac{s_x^2}{S_x^2}$ and $G(u, v)$ is a function of u and v such that

(i) The point (u, v) assumes a value in a closed convex subset R_2 of two dimensional real space containing the point $(1, 1)$.

(ii) The function $G(u, v)$ is continuous and bounded in R_2

(iii) $G(1,1) = 1$

(iv) The first and second order partial derivatives of $G(u,v)$ exist and are continuous and bounded in R_2.

We note that all ratio or product type estimators of population variance considered by Das and Tripathi (1978) and Kaur and Singh (1982) are special cases of class of estimators of Srivastava and Jhajj (1980). The knowledge of coefficient of kurtosis of a variable under study is seldom available. However, the coefficient of kurtosis of an auxiliary variable can be obtained easily.

In order to have the survey estimate for population mean $\bar{Y}$ of the study variable y for instance assuming the knowledge of population mean $\bar{X}$ of the auxiliary variable x we have the well known ratio estimator.

$$\hat{\bar{Y}}_R = \bar{y}(\frac{\bar{X}}{\bar{x}}) \quad (4)$$

where $\bar{y}$ and $\bar{x}$ are the unweighted sample mean of the variable y and x respectively. The Bias and MSE of $\hat{\bar{Y}}_R$ to first order approximation are given by

$B(\hat{\bar{Y}}_R) = \theta \bar{Y} C_x^2 (1-K)$

$MSE(\hat{\bar{Y}}_R) = \theta \bar{Y}^2 [C_y^2 + C_x^2(1-2K)]$

where $\theta = 1 - \frac{n}{N}$, $K = \rho(\frac{C_y}{C_x})$, C_y and C_x are coefficients of variation of y and x respectively and ρ is the correlation coefficient between y and x.

Prasad and Singh (1990) considered a ratio type for estimating the finite population variance by improving on the Isaki's estimator(1983) in terms of bias and precision.

Singh (1991) considered a general class of estimators for estimating the finite population variance S_y^2 and defined his estimator, $\hat{S^2}_{S91}$ as

$$\hat{S^2}_{S91} = s_y^2 G(u,v) \quad (5)$$

where $u = \frac{\bar{x}}{\bar{x}^*}$, $v = \frac{s_x^2}{s_x^{2*}}$ and $G(u,v)$ is a parametric function satisfying the following regularity conditions:

(i) $G(1,1) = 1$

(ii) The first and second order partial derivatives of G with respect to u and v exist and are continuous and known constants.

Upadhyaya and Singh (1999) using the known information on both S_x^2 and κ_x suggested modified ratio type population variance estimator for S_y^2 as

$$t_3 = s_y^2[\frac{S_x^2 + \kappa_x}{s_x^2 + \kappa_x}] \tag{6}$$

Upadhyaya and Singh (2001) utilized the mean of the auxiliary variable and proposed the following modified ratio estimator of population variance

$$\hat{S^2}_{U01} = s_y^2[\frac{\bar{X}}{\bar{x}}] \tag{7}$$

Singh et al. (2004) assuming known coefficient of kurtosis κ_x and using the transformation $\mu_i = x_i + \kappa_x$,(i=1,2,....,N) suggested the following modified ratio estimator for the population mean $\bar{Y}$ as

$$\hat{\bar{Y}}_M = \bar{y}(\frac{\bar{X} + \kappa_x}{\bar{x} + \kappa_x}) \tag{8}$$

To first order approximation the bias and MSE of $\hat{\bar{Y}}_M$ was obtained by letting $\bar{y} = \bar{Y}(1+\xi_0)$, $\bar{x} = \bar{X}(1+\xi_1)$ so that $E(\xi_0)$=$E(\xi_1)$=0 and $V(\xi_0) = \frac{1-f}{n}C_y^2$, $V(\xi_1) = \frac{1-f}{n}C_x^2$ and $Cov(\xi_0, \xi_1) = \frac{1-f}{n}\rho C_y C_x$. Assumption is made that the sample size n is large enough to make $|\xi_0|$ and $|\xi_1| < 1$ so as to validate the first degree approximation i.e. the terms involving ξ_0 and/or ξ_1 having powers greater than two will be negligible. Then

$$\hat{\bar{Y}}_M = \bar{Y}(1 + \xi_0)(1 + \lambda\xi_1)^{-1} \tag{9}$$

where $\lambda = \frac{\bar{X}}{\bar{x}+\kappa_x}$. Suppose that $|\lambda\xi_1| < 1$ so that $(1 + \lambda\xi_1)^{-1}$ converges. Then the Bias and MSE of $\hat{\bar{Y}}_M$ to first degree of approximation, respectively are given by

$$Bias(\hat{\bar{Y}}_M) = \frac{1-f}{n}\bar{Y}\lambda C_x^2(\lambda - K) \tag{10}$$

$$MSE(\hat{\bar{Y}}_M) = \frac{1-f}{n}\bar{Y}^2[C_y^2 + \lambda C_x^2(\lambda - 2K)] \tag{11}$$

Arcos et al. (2005) also came up with another type of modified ratio estimator that improved on Isaki's estimator (1983) which is less biased and more precise than the previous existing estimators, given by

$$\hat{S}^2{}_{Ar} = s_y^2 + c(S_x^2 - s_x^2) + d(\bar{X} - \bar{x}) \tag{12}$$

Kadilar and Cingi (2006) suggested four modified ratio type variance estimators using known values of coefficient of variation variation C_x and coefficient of kurtosis κ_x of an auxiliary variable X as follows

$$t_4 = s_y^2\{\frac{S_x^2 - C_x}{s_x^2 - C_x}\} \tag{13}$$

$$t_5 = s_y^2\{\frac{S_x^2 - \kappa_x}{s_x^2 - \kappa_x}\} \tag{14}$$

$$t_6 = s_y^2\{\frac{S_x^2\kappa_x - C_x}{s_x^2\kappa_x - C_x}\} \tag{15}$$

$$t_7 = s_y^2\{\frac{S_x^2 C_x - \kappa_x}{s_x^2 C_x - \kappa_x}\} \tag{16}$$

Singh et al. (2011) improved Bahl and Tuteja (1991) exponential ratio type estimator for the population mean defined as, $\bar{Y} = \bar{y}\exp[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}]$ and proposed the following exponential ratio type estimator for the population variance as:

$$\hat{S}^2{}_{S11} = s_y^2 exp\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \tag{17}$$

Using the known value of population median M_x of the auxiliary variable x Subramani and Kumarapandiyan (2012a) have suggested the modified ratio type estimator of the population variance S_y^2 of study variable as

$$t_8 = s_y^2\{\frac{S_x^2 + M_x}{s_x^2 + M_x}\} \tag{18}$$

Subramani and Kumarapandiyan (2012b) have proposed the modified ratio type estimators of population variance S_y^2 using the known quartiles of the auxiliary variable x as

$$t_9 = s_y^2\{\frac{S_x^2 + Q_1}{s_x^2 + Q_1}\} \tag{19}$$

$$t_{10} = s_y^2\{\frac{S_x^2 + Q_3}{s_x^2 + Q_3}\} \tag{20}$$

Motivated by Kadilar and Cingi (2006) and Subramani and Kumarapandiyan (2012a), Subramani and Kumarapandiyan (2013) considered the estimation of finite population variance using known coefficient of variation and median of an auxiliary variable, proposed an estimator, given as:

$$t_{11} = s_y^2[\frac{C_x S_x^2 + M_x}{C_x s_x^2 + M_x}] \tag{21}$$

Khan and Shabbir (2013b) gave a ratio type estimator of population variance using coefficient of correlation and upper quartile of auxiliary variable x. The problem herein was built on Isaki's known parameter variance estimator. The estimator postulated is given as:

$$t_{12} = s_y^2 \left[\frac{S_x^2 \rho_{xy} + Q_3}{s_x^2 \rho_{xy} + Q_3}\right] \tag{22}$$

Khan (2015) proposed an improved modified ratio type estimator for finite population variance using the transformation of variables.

$$\hat{S}^2{}_{K15} = s_y^2[\alpha\{2 - (\frac{S_x^2 + \kappa_x}{S_x^2 + \kappa_x})\} + (1-\alpha)\{2 - (\frac{S_x^2 + \kappa_x}{S_x^2 + \kappa_x})\}] \tag{23}$$

The mean squared error of his proposed estimator is less than the mean squared errors of previously suggested existing estimators meaning that it got some good gain in efficiency. Yadav et al. (2016) considered an efficient dual to ratio and product estimator of the population variance, making use of the coefficient of kurtosis and mean of the auxiliary variable and proposed the following improved ratio type estimator of the population variance

$$\hat{S}^2{}_{Y16} = s_y^2[\frac{\bar{x}^* + \alpha \bar{X}}{\bar{X} + \alpha \bar{x}^*}] \tag{24}$$

where α is a suitably chosen characterizing constant and is obtained by minimizing the MSE of the proposed estimator t_Y and $\bar{x}^* = \frac{N\bar{X} - n\bar{x}}{N-n} = (1+g)\bar{X} - g\bar{x}, g = \frac{n}{N-n}$.

Bhat et al. (2017) estimated variance using Tri-mean(TM) and semi-quartile range of the auxiliary variable x, defined as $TM = \frac{Q_1 + 2M_x + Q_3}{4}$ and $Q_a = \frac{Q_3 + Q_1}{2}$ respectively. The estimator is given by:

$$t_{13} = s_y^2 \left[\frac{S_x^2 + (TM + Q_a)}{s_x^2 + (TM + Q_a)}\right] \tag{25}$$

2.2 Statistical Properties(Bias and MSE)

First, we define the notations we are using in this section:

$\mu_{rs} = \frac{1}{N-1}\Sigma_{i=1}^{n}(y_i - \bar{y})^r(x_i - \bar{x})^s$, $\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{\frac{r}{2}}\mu_{02}^{\frac{s}{2}}}$. Thus we note the following;

$\mu_{20} = S_y^2$, $\mu_{02} = S_x^2$, and $\mu_{11} = S_{xy}$; $\lambda_{22} = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$, $\lambda_{21} = \frac{\mu_{21}}{\mu_{20}\mu_{02}^{\frac{1}{2}}}$ such that;

$C_y = \frac{S_y^2}{Y^2} = \frac{\mu_{20}}{Y^2}$ coefficient of variation for the study variable y, $C_x = \frac{S_x^2}{X^2} = \frac{\mu_{02}}{X^2}$ coefficient of variation for the auxiliary variable x and $\rho_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\mu_{11}}{\sqrt{\mu_{20}}\sqrt{\mu_{02}}}$ coefficient of correlation between x and y, $\kappa_{(y)} = \lambda_{40} = \frac{\mu_{40}}{\mu_{20}^2}$ coeffcient of kurtosis for the study variable, $\kappa_{(x)} = \lambda_{04} = \frac{\mu_{04}}{\mu_{02}^2}$ coefficient of kurtosis for the auxiliary variable and M_x population median of the auxiliary variable.

The bias and variance of t_1 to first order approximation are given by:

$$Bias(t_1) = \frac{1-f}{n}S_y^2\{(\kappa_x - 1)\Psi_1(\Psi_1 - \frac{\lambda_{22}-1}{\kappa_x - 1})\} = 0 \tag{26}$$

$$\begin{aligned} MSE(t_1) = Var(t_1) &= \frac{1-f}{n}S_y^4\{(\kappa_y - 1) + (\kappa_x - 1)\Psi_1(\Psi_1 - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \\ &= \frac{(1-f)}{n}S_y^4(\kappa_y - 1) \end{aligned} \tag{27}$$

where $\Psi_1 = 0$

Prasad and Singh (1990) obtained the bias and Mean Squared Error of Isaki's estimator, to first order of approximation as follows

$$Bias(t_2) = \frac{1-f}{n}S_y^2\{(\kappa_x - 1)\Psi_2(\Psi_2 - \frac{\lambda_{22}-1}{\kappa_x - 1})\} = \frac{(1-f)}{n}S_y^2[(\kappa_x - 1) - (\lambda_{22} - 1)] \tag{28}$$

$$\begin{aligned} MSE(t_2) &= \frac{1-f}{n}S_y^4\{(\kappa_y - 1) + (\kappa_x - 1)\Psi_2(\Psi_2 - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \\ &= \frac{(1-f)}{n}S_y^4[(\kappa_y - 1) + (\kappa_x - 1) - 2(\lambda_{22} - 1)] \end{aligned} \tag{29}$$

where $\Psi_2 = 1$

Upadhyaya and Singh (1999) estimator using the known information on both S_x^2 and κ_x obtained the bias and MSE of their estimator t_3 to first order of approximation

$$Bias(t_3) = \frac{1-f}{n}S_y^2[(\{\kappa_x - 1\})\Psi_3(\Psi_3 - \frac{\lambda_{22}-1}{\kappa_x - 1})] \tag{30}$$

$$MSE(t_3) = \frac{1-f}{n} S_y^4[\{\kappa_y - 1\} + \{\kappa_x - 1\}\Psi_3(\Psi_3 - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))] \quad (31)$$

where $\Psi_3 = \frac{S_x^2}{S_x^2+\kappa_x}$

Upadhyaya and Singh (2001) obtained the bias and MSE of their modified ratio type population variance $\hat{S^2}_{U01}$ estimator up to first order approximations

$$Bias(\hat{S^2}_{U01}) = \frac{1-f}{n} S_y^2[C_x^2 - \lambda_{21}C_x] \quad (32)$$

$$MSE(\hat{S^2}_{U01}) = \frac{1-f}{n} S_y^4[(\lambda_{40} - 1) + C_x^2 - 2\lambda_{21}C_x] \quad (33)$$

Kadilar and Cingi (2006), derived the biases and MSE of their four modified ratio type variance estimators to first order approximations to get;

$$Bias(t_4) = \frac{1-f}{n} S_y^2(\kappa_x - 1)\{\Psi_4(\Psi_4 - \frac{\lambda_{22}-1}{\kappa_x - 1})\} \quad (34)$$

$$MSE(t_4) = \frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \Psi_4(\kappa_x - 1)(\Psi_4 - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \quad (35)$$

$$Bias(t_5) = \frac{1-f}{n} S_y^2(\kappa_x - 1)\{\Psi_5(\Psi_5 - (\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \quad (36)$$

$$MSE(t_5) = \frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \Psi_5(\kappa_x - 1)(\Psi_5 - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \quad (37)$$

$$Bias(t_6) = \frac{1-f}{n} S_y^2(\kappa_x - 1)\{\Psi_6(\Psi_6 - (\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \quad (38)$$

$$MSE(t_6) = \frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \Psi_6(\kappa_x - 1)(\Psi_6 - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \quad (39)$$

$$Bias(t_7) = \frac{1-f}{n} S_y^2(\kappa_x - 1)\{\Psi_7(\Psi_7 - (\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \quad (40)$$

$$MSE(t_7) = \frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \Psi_7(\kappa_x - 1)(\Psi_7 - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))\} \quad (41)$$

where;

$\Psi_4 = \frac{S_x^2}{S_x^2 - C_x}$; $\Psi_5 = \frac{S_x^2}{S_x^2 - \kappa_x}$; $\Psi_6 = \frac{S_x^2\kappa_x}{S_x^2\kappa_x - C_x}$; $\Psi_7 = \frac{S_x^2 C_x}{S_x^2 C_x - \kappa_x}$.

Subramani and Kumarapandiyan (2013) obtained the bias and MSE of their estimator t_8 to first order approximation as:

$Bias(t_8) = \frac{1-f}{n} S_y^2(\kappa_x - 1)\{\Psi_8(\Psi_8 - (\frac{\lambda_{22}-1}{\kappa_x - 1}))\}$

$MSE(t_8) = \frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \Psi_8(\kappa_x - 1)(\Psi_8 - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))\}$

where, $\Psi_8 = \frac{S_x^2}{S_x^2+M_x}$.

This estimator is more efficient in terms of bias and mean squared error than the traditional ratio type and preceding modified ratio type population variance estimators under specified conditions.

Subramani and Kumarapandiyan (2012b) in their proposed modified ratio type population variance estimators using the known quartiles of the auxiliary variable x (upper and lower quartile Q_3 and Q_1 respectively) came up with the bias and MSE of their estimators t_9 and t_{10} as follows

$Bias(t_9) = \frac{1-f}{n} S_y^2(\kappa_x - 1)\{\Psi_9(\Psi_9 - (\frac{\lambda_{22}-1}{\kappa_x-1}))\}$

$MSE(t_9) = \frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \Psi_9(\kappa_x - 1)(\Psi_9 - 2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$

$Bias(t_{10}) = \frac{1-f}{n} S_y^2(\kappa_x - 1)\{\Psi_{10}(\Psi_{10} - (\frac{\lambda_{22}-1}{\kappa_x-1}))\}$

$MSE(t_{10}) = \frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \Psi_{10}(\kappa_x - 1)(\Psi_{10} - 2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$

where $\Psi_9 = \frac{S_x^2}{S_x^2+Q_1}$ and $\Psi_{10} = \frac{S_x^2}{S_x^2+Q_3}$.

The modified ratio type estimator by Subramani and Kumarapandiyan (2013) taking motivation from Kadilar and Cingi (2006) and Subramani and Kumarapandiyan (2012a) obtained bias and MSE of their proposed population variance estimator that utilizes the coefficient of variation and median of auxiliary variable as follows:

$$Bias(t_{11}) = \frac{1-f}{n} S_y^2(\kappa_x - 1)\{\Psi_{11}(\Psi_{11} - (\frac{\lambda_{22}-1}{\kappa_x-1}))\} \tag{42}$$

$$MSE(t_{11}) = \frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \Psi_{11}(\kappa_x - 1)(\Psi_{11} - 2(\frac{\lambda_{22}-1}{\kappa_x-1}))\} \tag{43}$$

where $\Psi_{11} = \frac{C_x S_x^2}{C_x S_x^2+M_x}$.

The bias and MSE of t_{12} to first order of approximations is given by:

$$Bias(t_{12}) = \frac{1-f}{n} S_y^2 \left[(\kappa_x - 1)\Psi_{12}\left(\Psi_{12} - \left(\frac{\lambda_{22}-1}{\kappa_x-1}\right)\right)\right] \tag{44}$$

$$MSE(t_{12}) = \frac{1-f}{n} S_y^4 \left[(\kappa_y - 1) + \Psi_{12}(\kappa_x - 1)\left(\Psi_{12} - 2\left(\frac{\lambda_{22}-1}{\kappa_x-1}\right)\right)\right] \tag{45}$$

where $\Psi_{12} = \frac{S_x^2\rho_{xy}}{S_x^2\rho_{xy}+Q_3}$.

Yadav et al. (2016) derived the bias and MSE of their efficient dual to ratio and product estimator of the population variance, $\hat{S}^2{}_{Y16}$ to first order approximations as

$$Bias(\hat{S}^2{}_{Y16}) = -\frac{1-f}{n}S_y^2[Fg\lambda_{21}C_x + \frac{F\alpha}{1+\alpha}g^2C_x^2] \tag{46}$$

where $F = \frac{1-\alpha}{1+\alpha}$

$$MSE(\hat{S}^2{}_{Y16}) = \frac{1-f}{n}S_y^4[(\lambda_{40}-1) + F^2g^2C_x^2 - 2Fg\lambda_{21}C_x] \tag{47}$$

which is minimum for $F = \frac{\lambda_{21}}{gC_x}$ and the minimum mean squared error of t_Y for this optimum value of F is,

$$MSE_{min}(\hat{S}^2{}_{Y16}) = \frac{1-f}{n}S_y^4[(\lambda_{40}-1) - \lambda_{21}^2] \tag{48}$$

The Bias and MSE of t_{13} to first order of approximations is given by:

$$Bias(t_{13}) = \frac{1-f}{n}S_y^2\left[(\kappa_x - 1)\Psi_{13}\left(\Psi_{13} - \left(\frac{\lambda_{22}-1}{\kappa_x - 1}\right)\right)\right] \tag{49}$$

$$MSE(t_{13}) = \frac{1-f}{n}S_y^4\left[(\kappa_y - 1) + \Psi_{13}(\kappa_x - 1)\left(\Psi_{13} - 2\left(\frac{\lambda_{22}-1}{\kappa_x - 1}\right)\right)\right] \tag{50}$$

where $\Psi_{13} = \frac{S_x^2}{S_x^2 + TM + Q_a}$

2.3 Empirical studies

The performance of proposed modified ratio type variance estimator is always assessed by many authors using empirical studies comparing it with the traditional and existing modified ratio type variance estimators.

Subramani and Kumarapandiyan (2013) used real data from the Italian Bureau for Enviroment Protection (APAT) 2004 Report on Waste 2004 to assess the performance of their estimator. Their results showed that the bias and mean squared error of their proposed estimator is less than the biases and mean squared errors of the traditional and existing estimators.

Khan and Shabbir (2013a) considered two natural populations from the literature of survey to perform efficiency comparison of their proposed estimator with the existing estimators.

Population 1 from Das (1988) and population 2 from (Cochran, 1977, p.325). They conclude out of their empirical studies that their estimator under optimizing conditions was more efficient than the existing estimators.

2.4 Taylor's Linearization Method

Applying the Taylor Linearization method, non-linear statistics are approximated by linear forms of the observations (by taking the first-order terms in an appropriate Taylor-series expansion). Second or even higher-order approximations could be developed by extending the Taylor series expansion. However, in practice, the first-order approximation usually yields satisfactory results, with the exception of highly skewed populations Wolter (2007).

After applying Taylor's approximation, standard variance estimation techniques can then be applied to the linearized statistic. This implies that Taylor Linearization is not 'in itself' method for variance estimation, it simply provides approximate linear forms of the statistics of interest (e.g. a weighted total) and then other methods should be deployed for the estimation of variance itself (either analytic or approximate ones).

Taylor linearization method is a widely applied method because it is quite straightforward for any case where an estimator already exists for totals. However, the Taylor linearization variance estimator is a biased estimator. Its bias stems from its tendency to underestimate the true value and it depends on the size of the sample and the complexity of the estimated parameter. Though, if the statistic is fairly simple, like the weighted sample mean, the bias is negligible even for small samples, while it becomes nil for large samples Sarndal et al. (1992). On the other hand for a complex estimator for a parameter like the variance, large samples are needed for the bias to be small.

It is the most popular method of variance estimation for complex statistics such as ratio and regression estimators and logistic regression coefficient estimators. Generally applicable to any sampling design that permits unbiased variance estimation for linear estimators. Its advantage is that it is computationally simpler and more compatible with many existing programs and softwares than the resampling methods such as the jackknife.

CHAPTER THREE

3 METHODOLOGY

Consider a finite population $V = \{V_1, V_2, V_3, ..., V_N\}$ of N distinct identifiable units. Let Y be our study variable and X be its corresponding auxiliary variable. Suppose we take a random sample of size n from this bivariate population (Y, X) that is (y_i, x_i), for $i = 1, 2, 3, ..., n$ using a Simple Random Sampling Without Replacement ($SRSWOR$) method. Let $\bar{Y}$ and $\bar{X}$ be the population means of the study and auxiliary variable respectively and their corresponding sample means be $\bar{y}$ and $\bar{x}$.

This study considers the problem of estimating the population variance,defined as $S_y^2 = \frac{1}{N-1}\Sigma_{i=1}^{N}(Y_i - \bar{Y})^2$ and uses auxiliary information to improve the efficiency of the population variance estimator.

We define the following notations that we will make use of throughout the thesis. For the population observations we have;

$\bar{Y} = \frac{1}{N}\Sigma_{i=1}^{N}Y_i, \qquad \bar{X} = \frac{1}{N}\Sigma_{i=1}^{N}X_i, \qquad S_y^2 = \frac{1}{N-1}\Sigma_{i=1}^{N}(Y_i - \bar{Y})^2,$

$S_x^2 = \frac{1}{N-1}\Sigma_{i=1}^{N}(X_i - \bar{X})^2, \qquad S_{xy} = \frac{1}{N-1}\Sigma_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X}).$

Also we define the following from the sample observations:

$\bar{y} = \frac{1}{n}\Sigma_{i=1}^{n}y_i, \qquad \bar{x} = \frac{1}{n}\Sigma_{i=1}^{n}x_i, \qquad s_y^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(y_i - \bar{y})^2,$

$s_x^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(x_i - \bar{x})^2, \qquad s_{xy} = \frac{1}{n-1}\Sigma_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}).$

In general, we recall the following parameters we defined in section (2.2):

$\mu_{rs} = \frac{1}{N-1}\Sigma_{i=1}^{n}(y_i - \bar{y})^r(x_i - \bar{x})^s, \qquad \lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{\frac{r}{2}}\mu_{02}^{\frac{s}{2}}}$. Thus we note the following;

$\mu_{20} = S_y^2, \qquad \mu_{02} = S_x^2$, and $\mu_{11} = S_{xy}$; $\lambda_{22} = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$, $\lambda_{21} = \frac{\mu_{21}}{\mu_{20}\mu_{02}^{\frac{1}{2}}}$ such that;

$C_y = \frac{S_y^2}{\bar{Y}^2} = \frac{\mu_{20}}{\bar{Y}^2}$ coefficient of variation for the study variable y, $C_x = \frac{S_x^2}{\bar{X}^2} = \frac{\mu_{02}}{\bar{X}^2}$ coefficient of variation for the auxiliary variable x and $\rho_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\mu_{11}}{\sqrt{\mu_{20}}\sqrt{\mu_{02}}}$ coefficient of correlation between x and y, $\kappa_{(y)} = \lambda_{40} = \frac{\mu_{40}}{\mu_{20}^2}$ coeffcient of kurtosis for the study variable, $\kappa_{(x)} = \lambda_{04} = \frac{\mu_{04}}{\mu_{02}^2}$ coefficient of kurtosis for the auxiliary variable and M_x population median of the auxiliary variable.

3.1 Linearity of Expectation

Following the works of Karr (1993) we have

Theorem 1

Let X and Y be random variables. We have that

$$E(X+Y) = E(X) + E(Y) \tag{51}$$

and note that true for any X and Y even when they are dependent.

Proof

We first show that

$$E(X+Y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} (i+j) P\{X = i, Y = j\} \tag{52}$$

$$E(X+Y) = \sum_{K=-\infty}^{\infty} K.P\{X+Y = K\} \tag{53}$$

$$= \sum_{K=-\infty}^{\infty} K.(\sum_{i=-\infty}^{\infty} P\{X = i, Y = K - i\}) \tag{54}$$

$$= \sum_{i=-\infty}^{\infty} K.(\sum_{k=-\infty}^{\infty} P\{X = i, Y = K - i\}) \tag{55}$$

setting $K - i = j \Rightarrow K = j + i$

$$= \sum_{i=-\infty}^{\infty} (i+j)(\sum_{j=-\infty}^{\infty} P\{X = i, Y = j\}) \tag{56}$$

We now have

$$E(X+Y) = \sum_{i=-\infty}^{\infty} (i+j)(\sum_{j=-\infty}^{\infty} P\{X = i, Y = j\}) \tag{57}$$

$$= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} i.P\{X = i, Y = j\} + \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} j.P\{X = i, Y = j\} \tag{58}$$

Considering the first part of equation (58)

$$\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} i.P\{X=i,Y=j\} = \sum_{i=-\infty}^{\infty} i.\underbrace{\sum_{j=-\infty}^{\infty} P\{X=i,Y=j\}} \tag{59}$$

$$= \sum_{i=-\infty}^{\infty} i.P\{X=i,Y=j\} \tag{60}$$

$$= E(X) \tag{61}$$

and the second part of equation (58)

$$\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} j.P\{X=i,Y=j\} = \sum_{i=-\infty}^{\infty} j.(\underbrace{\sum_{j=-\infty}^{\infty} P\{X=i,Y=j\})} \tag{62}$$

$$= \sum_{i=-\infty}^{\infty} j.P\{X=i,Y=j\} \tag{63}$$

$$= E(Y) \tag{64}$$

Therefore

$$E(X+Y) = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} (i+j)P\{X=i,Y=j\} \tag{65}$$

$$= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} i.P\{X=i,Y=j\} + \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} j.P\{X=i,Y=j\} \tag{66}$$

$$= E(X) + E(Y) \tag{67}$$

3.2 Expected value of the ratio of correlated random variables

Consider random variables m and n which are correlated. Suppose we defined them as

$$\begin{aligned} m &= E(m) + m* \\ n &= E(n) + n* \end{aligned} \tag{68}$$

by simply interchanging variables m and n with new variables $m*$ and $n*$ hence still measuring the same things; we have just shifted the axes so that 0 is the expected value (for instance if the expected number of descendants is 2, then we measure the actual number by how much it differs from 2; if the individual ends up just leaving just 1 descendant, then $m*$=-1).

We note that $E(\frac{m}{n})$ is undefined for any nonzero probability that $n = 0$. Therefore we calculate $E(\frac{m}{n}|n \neq 0)$, the expected value of the ratio conditional on n not equaling zero. This condition makes complete sense in evolutionary theory; since $n = 0$ iff the population goes extinct- hence the case where the result become undefined.

Using the definitions in equation (68) we can write:

$$E\left(\frac{m}{n}|n \neq 0\right) = E\left(\frac{E(m)+m*}{E(n)+n*}\right) = E\left(\frac{E(m)}{E(n)}\frac{[1+\frac{m*}{E(m)}]}{[1+\frac{n*}{E(n)}]}\right) \tag{69}$$

Noting that the expected values $E(m)$ and $E(n)$ are not random variables we can remove outside the expectation on the right hand side of equation (69) yielding:

$$E\left(\frac{m}{n}|n \neq 0\right) = \frac{E(m)}{E(n)}E\left(\frac{1+\frac{m*}{E(m)}}{1+\frac{n*}{E(n)}}\right) = \frac{E(m)}{E(n)}E\left[\left(1+\frac{m*}{E(m)}\right)\left(1+\frac{n*}{E(n)}\right)^{-1}\right] \tag{70}$$

Multiplying out the terms in the square brackets yields:

$$E\left(\frac{m}{n}|n \neq 0\right) = \frac{E(m)}{E(n)}E\left[\left(1+\frac{n*}{E(n)}\right)^{-1}\right] + \frac{1}{E(n)}E\left[m*\left(1+\frac{n*}{E(n)}\right)^{-1}\right] \tag{71}$$

By the definition of harmonic mean $E(\frac{1}{n}) = \frac{1}{H(n)}$, where $H(n)$ is the harmonic mean of n. We can use equation (70) to find $E(\frac{1}{n})$ by setting $m = 1$ (so that $E(m) = 1$ and $m* = 0$). Thus we will obtain:

$$E\left(\frac{m}{n}|n \neq 0\right) \equiv \frac{1}{H(n)} = \frac{1}{E(n)}E\left[\left(1+\frac{n*}{E(n)}\right)^{-1}\right] \tag{72}$$

We can rewrite now the right hand side of equation (71) by using equation (72)

$$E\left(\frac{m}{n}|n \neq 0\right) = \frac{E(m)}{H(n)} + \frac{1}{E(n)}E\left[m*\left(1+\frac{n*}{E(n)}\right)^{-1}\right] \tag{73}$$

Now we have to deal with the term $E\left[m*\left(1+\frac{n*}{E(m)}\right)^{-1}\right]$. Provided that $n* < E(n)$ i.e. $n < 2E(n)$, we can expand $\left(1+\frac{n*}{E(n)}\right)^{-1}$ as a Taylor series in $n*$. If we define:

$$f_{n*} = \left(1+\frac{n*}{E(n)}\right)^{-1} \tag{74}$$

Taylor's theorem will yield:

$$f_{n*} = 1 + \sum_{i=1}^{\infty} (-1)^i \frac{n*^i}{E(n)^i} \tag{75}$$

Importantly we note that the use of Taylor's theorem is not applicable in all cases. Precisely, equation (74) does not converge to equation (75) if $n* \geq E(n)$, hence in such situations we fall back and use the calculus of finite differences.

When we can apply the Taylor expansion in equation (75), we will have:

$$E\left[m * \left(1 + \frac{n*}{E(n)}\right)^{-1}\right] = E\left[m * + \sum_{i=1}^{\infty} (-1)^i \frac{m * b*^i}{E(n)^i}\right] \tag{76}$$

From the definitions of $m*$ and $n*$ in equation (68), we know that $E(m*) = 0$, $E(m * n*) = cov(m, n)$ and in general, $E(m*n*^i)$ is the mixed central moment defined as $E[m - E(m)][n - E(n)]^i$, which can simply use the notation $\langle\langle n,^i m \rangle\rangle$. Hence we can now write equation (76) as

$$E\left[m * \left(1 + \frac{n*}{E(n)}\right)^{-1}\right] = \sum_{i=1}^{\infty} (-1)^i \frac{\langle\langle n,^i m \rangle\rangle}{E(n)^i} \tag{77}$$

Substituting equation (77) into equation (73) gives the equation for the expected value of the ratio:

$$E\left(\frac{m}{n}|n \neq 0\right) = \frac{E(m)}{H(n)} + \sum_{i=1}^{\infty} (-1)^i \frac{\langle\langle n,^i m \rangle\rangle}{E(n)^{i+1}} \tag{78}$$

For other situations, it is useful to have a result in which the first term does not involve the harmonic mean. To do this we simply substitute the series in equation (75) directly into the far right hand part of equation (70). Denoting the i^{th} central moment of n by $\langle\langle^i n \rangle\rangle$, so that $\langle\langle^1 n \rangle\rangle = 0$. Thus

$$E\left(\frac{m}{n}|n \neq 0\right) = \frac{E(m)}{E(n)} + \sum_{i=1}^{\infty} (-1)^i \frac{E(m)\langle\langle^i n \rangle\rangle + \langle\langle n,^i m \rangle\rangle}{E(n)^{i+1}} \tag{79}$$

3.3 Taylor's approximation method

Suppose we have an estimator $Z = g(X, Y)$ a function of two variables. Suppose that we can measure X and determine its population parameters such as mean and variance but really be interested in Y which is related to X in some way. We might be interested to know $Var(Y)$

at least approximately in order to assess the accuracy of indirect measurement process. Since we cannot in general find $E(y) = \mu_y$ and $Var(Y) = \sigma_y$ from $E(X) = \mu_x$ and $Var(X) = \sigma_x$ unless the function g is linear. As in our case of estimating the population variance we involve ratio which is non-linear so let us suppose g is non-linear thus we have to linearize.

Using Taylors' series expansion of g about $\mu = (\mu_x, \mu_y)$ in order to approximate the mean or variance of Z. To first order

$$Z = g(X,Y) \approx g(\mu) + (X - \mu_x)\frac{\partial g(\mu)}{\partial x} + (Y - \mu_y)\frac{\partial g(\mu)}{\partial y} \quad (80)$$

The notation $\frac{\partial g(\mu)}{\partial y}$ and $\frac{\partial g(\mu)}{\partial x}$ means that the partial derivative is evaluated at the point (μ_x, μ_y).

Z having been expressed approximately equal to a linear function of X and Y. The mean and variance of this linear function are easily calculated to be

$E(Z) \approx \mu$ and

$Var(Z) = \sigma_x^2(\frac{\partial g(\mu)}{\partial x})^2 + \sigma_y^2(\frac{\partial g(\mu)}{\partial y}) + 2\sigma_{xy}(\frac{\partial g(\mu)}{\partial x})(\frac{\partial g(\mu)}{\partial y})$

Illustration

Using Isaki (1983) ratio type population variance estimator for our illustration

$$t_R = s_y^2 \frac{S_x^2}{s_x^2}$$

Let us define

$\xi_0 = \frac{s_y^2}{S_y^2} - 1$, $\xi_1 = \frac{s_x^2}{S_x^2} - 1$, $\xi_2 = \frac{\bar{y}}{\bar{Y}} - 1$, $\xi_3 = \frac{\bar{x}}{\bar{X}} - 1$, $\xi_4 = \frac{s_{xy}}{S_{xy}} - 1$ such that

$E(\xi_0) = E(\xi_1) = E(\xi_2) = E(\xi_3) = E(\xi_4) = 0$ and

$E(\xi_2^2) = (\frac{1-f}{n})C_y^2$, $E(\xi_3^2) = (\frac{1-f}{n})C_x^2$, $E(\xi_2\xi_3) = (\frac{1-f}{n})\rho_{xy}C_yC_x$.

Expressing the estimator t_R in terms of ξ_0 and ξ_1 can easily be written as

$$t_R = S_y^2(1+\xi_0)(1+\xi_1)^{-1} = S_y^2(1+\xi_0)(1-\xi_1+\xi_1^2+...) = S_y^2[1+\xi_0-\xi_1+\xi_1^2-\xi_0\xi_1+...] \quad (81)$$

To the first order of Taylor's approximations we have;

$E(\xi_0^2) = (\frac{1-f}{n})(\lambda_{40} - 1)$, $E(\xi_1^2) = (\frac{1-f}{n})(\lambda_{04} - 1)$,

$E(\xi_4^2) = (\frac{1-f}{n})(\frac{\lambda_{22}}{\rho_{xy}^2} - 1)$, $E(\xi_2\xi_0) = (\frac{1-f}{n})C_y\lambda_{30}$, $E(\xi_2\xi_1) = (\frac{1-f}{n})C_y\lambda_{12}$, $E(\xi_2\xi_4) = (\frac{1-f}{n})C_y\frac{\lambda_{21}}{\rho_{xy}}$, $E(\xi_3\xi_0) = (\frac{1-f}{n})C_x\lambda_{21}$, $E(\xi_3\xi_1) = (\frac{1-f}{n})C_x\lambda_{03}$, $E(\xi_3\xi_4) = (\frac{1-f}{n})C_x\frac{\lambda_{12}}{\rho_{xy}}$, $E(\xi_0\xi_1) = (\frac{1-f}{n})(\lambda_{22} - 1)$, $E(\xi_0\xi_4) = (\frac{1-f}{n})(\frac{\lambda_{31}}{\rho_{xy}} - 1)$, and $E(\xi_1\xi_4) = (\frac{1-f}{n})(\frac{\lambda_{13}}{\rho_{xy}} - 1)$, where f is the finite population correction (f.p.c) factor. Thus we have the following theorems as stated by Singh (2003)

Theorem 2

Bias upto order $O(n^{-1})$ in the estimator of t_R is

$$Bias(t_R) = \frac{1-f}{n}S_y^2(\lambda_{04} - \lambda_{22}) \tag{82}$$

Proof

Taking the expectation on both sides of (81) we have

$E(t_R) = S_y^2 E[1 + \xi_0 - \xi_1 + \xi_1^2 - \xi_0\xi_1] = S_y^2[1 + (\frac{1-f}{n})(\lambda_{04} - 1) - (\lambda_{22} - 1)]$

and using the result $B(t_R) = E(t_R) - S_y^2$ we have (82).

Theorem 3

The MSE of the estimator t_R up to first order of approximations is

$$MSE(t_R) = (\frac{1-f}{n})S_y^4[\lambda_{40} + \lambda_{04} - 2\lambda_{22}] \tag{83}$$

Proof

We have

$MSE(t_R) = E(t_R - S_y^2)^2$

$\approx E[S_y^2(1 + \xi_0 - \xi_1 + \xi_1^2 - \xi_0\xi_1 + ...) - S_y^2]^2$

$\approx S_y^4 E(\xi_0 - \xi_1)^2$

$= S_y^4 E[\xi_0^2 + \xi_1^2 - 2\xi_0\xi_1]$

$= \frac{1-f}{n}S_y^4[(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)]$.

3.4 Proposed Estimator

Motivated by the works of Khan and Shabbir (2013a), Upadhyaya and Singh (1999), Singh et al. (2004), Subramani and Kumarapandiyan (2013), Kadilar and Cingi (2006), and Yadav et al. (2016) in the improvement of the performance of the population variance estimator of the study variable using known population parameters of an auxiliary variable. We propose the following modified ratio type estimator for the population variance S_y^2 using known values of population coefficient of kurtosis κ_x and median M_x of an auxiliary variable.

$$\hat{S}_{PM}^2 = s_y^2\{\frac{S_x^2\kappa_x + M_x^2}{s_x^2\kappa_x + M_x^2}\} \tag{84}$$

To obtain the bias and the MSE of our proposed estimator $\hat{S}_{PM}^2$,

We define $s_y^2 = S_y^2(1+\xi_0)$ and $s_x^2 = S_x^2(1+\xi_1)$ or $\xi_0 = \frac{s_y^2}{S_y^2} - 1$ and $\xi_1 = \frac{s_x^2}{S_x^2} - 1$ such that $E(\xi_0) = E(\frac{s_y^2}{S_y^2}) - E(1) = 0$ and $E(\xi_1) = E(\frac{s_x^2}{S_x^2}) - E(1) = 0$ and to the first degreee of approximations we have

$E(\xi_0^2) = \frac{1-f}{n}(\lambda_{40} - 1)$, $E(\xi_1^2) = \frac{1-f}{n}(\lambda_{04} - 1)$, $E(\xi_0\xi_1) = \frac{1-f}{n}(\lambda_{22} - 1)$.

The above expectations are obtained following the works of Sukhatme (1944), Sukhatme and Sukhatme (1970), Srivastava and Jhajj (1981), Tracy (1984) and Withers and Nadarajah (2014).

Now expressing $\hat{S}_{PM}^2$ in terms of $\xi's$ we have

$$\hat{S}_{PM}^2 = S_y^2(1+\xi_0)\{\tfrac{\kappa_x S_x^2 + M_x^2}{\kappa_x S_x^2(1+\xi_1) + M_x^2}\}$$

$$= S_y^2(1+\xi_0)(1+\varrho^*\xi_1)^{-1} \tag{85}$$

where $\varrho^* = \kappa_x S_x^2(\kappa_x S_x^2 + M_x^2)^{-1}$, we assume that $|\varrho^*\xi_1| < 1$ so that $(1+\varrho^*\xi_1)^{-1}$ is expandable. Expanding the right hand side of (85) and multiplying out we have

$$\hat{S}_{PM}^2 = S_y^2(1+\xi_0)(1 - \varrho^*\xi_1 + \varrho^{*2}\xi_1^2 ...)$$
$$= S_y^2(1 + \xi_0 - \varrho^*\xi_1 - \varrho^*\xi_0\xi_1 + \varrho^{*2}\xi_1^2 + \varrho^{*2}\xi_0\xi_1^2 - ...)$$

Neglecting terms of $\xi's$ having power greater than two we have

$$\hat{S}^2_{PM} \cong S^2_y(1 + \xi_0 - \varrho^*\xi_1 - \varrho^*\xi_0\xi_1 + \varrho^{*2}\xi_1^2) \text{ or}$$

$$\hat{S}^2_{PM} - S^2_y \cong S^2_y(\xi_0 - \varrho^*\xi_1 - \varrho^*\xi_0\xi_1 + \varrho^{*2}\xi_1^2) \tag{86}$$

Taking the expectation on both sides of (86)

$E(\hat{S}^2_{PM} - S^2_y) \cong E(S^2_y(\xi_0 - \varrho^*\xi_1 - \varrho^*\xi_0\xi_1 + \varrho^{*2}\xi_1^2))$ We get the bias of the estimator $\hat{S}^2_{PM}$ to the first degree of approximation as

$$Bias(\hat{S}^2_{PM}) = \frac{1-f}{n} S^2_y(\kappa_x - 1)\varrho^*\{\varrho^* - \frac{(\lambda_{22} - 1)}{(\kappa_x - 1)}\} \tag{87}$$

Squaring both sides of (86) and neglecting terms of $\xi's$ having power greater than two we have

$$(\hat{S}^2_{PM} - S^2_y)^2 \cong S^4_y(\xi_0^2 + \varrho^{*2}\xi_1^2 - 2\varrho^*\xi_0\xi_1) \tag{88}$$

Taking the expectation on both sides of (88)

$E((\hat{S}^2_{PM} - S^2_y)^2) \cong E(S^4_y(\xi_0^2 + \varrho^{*2}\xi_1^2 - 2\varrho^*\xi_0\xi_1))$ We get the $(\hat{S^2_{PM}})$ estimator's Mean Squared Error to first degree of approximation as

$$MSE(\hat{S}^2_{PM}) = \frac{1-f}{n} S^4_y\{(\kappa_y - 1) + \varrho^*(\kappa_x - 1)(\varrho^* - 2\frac{(\lambda_{22} - 1)}{(\kappa_x - 1)})\} \tag{89}$$

Theoretical Conditions for our Proposed Estimator

Consider our proposed estimator

$$\hat{S}^2_{PM} = s^2_y\{\frac{S^2_x\kappa_x + M^2_x}{s^2_x\kappa_x + M^2_x}\} \tag{90}$$

Suppose we rewrite it as

$$\hat{S}^2_{PM} = \frac{m}{n} \tag{91}$$

i.e. we let $m = s^2_y(S^2_x\kappa_x + M^2_x)$ and $n = s^2_x\kappa_x + M^2_x$ and invoke the condition in finding the expectation of a ratio of correlated random variables in equation (69).

We note that $E(\hat{S}^2_{PM}) = E(\frac{m}{n})$ is undefined if there is any nonzero probability that $n = 0$.

Thus we will calculate $E(\frac{m}{n}|n \neq 0)$ the expected value of the ratio, conditional n not equaling zero. This means that

$$\kappa_x s_x^2 + M_x^2 \neq 0 \tag{92}$$

in order for our proposed estimator to be applicable.

3.5 Expressions for Bias and Mean Squared Errors of the Proposed and Existing Estimators

Table 1: Summary of Expressions of Statistical Properties (Bias and Mean Squared Errors(MSE))

ESTIMATOR	BIAS(.)	MEAN SQUARED ERROR(MSE)
s_y^2	$\frac{1-f}{n}S_y^2\{(\kappa_x-1)\Psi_1(\Psi_1-\frac{\lambda_{22}-1}{\kappa_x-1})\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+(\kappa_x-1)\Psi_1(\Psi_1-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2}{s_x^2})$	$\frac{1-f}{n}S_y^2\{(\kappa_x-1)\Psi_2(\Psi_2-\frac{\lambda_{22}-1}{\kappa_x-1})\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+(\kappa_x-1)\Psi_2(\Psi_2-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2+\kappa_x}{s_x^2+\kappa_x})$	$\frac{1-f}{n}S_y^2[(\{\kappa_x-1\})\Psi_3(\Psi_3-\frac{\lambda_{22}-1}{\kappa_x-1})]$	$\frac{1-f}{n}S_y^4[\{\kappa_y-1\}+\{\kappa_x-1\}\Psi_3(\Psi_3-2(\frac{\lambda_{22}-1}{\kappa_x-1}))]$
$s_y^2(\frac{S_x^2-C_x}{s_x^2-C_x})$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_4(\Psi_4-\frac{\lambda_{22}-1}{\kappa_x-1})\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\Psi_4(\kappa_x-1)(\Psi_4-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2-\kappa_x}{s_x^2-\kappa_x})$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_5(\Psi_5-(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\Psi_5(\kappa_x-1)(\Psi_5-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2\kappa_x-C_x}{s_x^2\kappa_x-C_x})$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_6(\Psi_6-(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\Psi_6(\kappa_x-1)(\Psi_6-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2C_x-\kappa_x}{s_x^2C_x-\kappa_x})$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_7(\Psi_7-(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\Psi_7(\kappa_x-1)(\Psi_7-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2+M_x}{s_x^2+M_x})$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_8(\Psi_8-(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\Psi_8(\kappa_x-1)(\Psi_8-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2+Q_1}{s_x^2+Q_1})$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_9(\Psi_9-(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\Psi_9(\kappa_x-1)(\Psi_9-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2+Q_3}{s_x^2+Q_3})$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_{10}(\Psi_{10}-(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\Psi_{10}(\kappa_x-1)(\Psi_{10}-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2(\frac{S_x^2C_x+M_x}{s_x^2C_x+M_x})$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_{11}(\Psi_{11}-(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\Psi_{11}(\kappa_x-1)(\Psi_{11}-2(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$
$s_y^2\left[\frac{S_x^2\rho_{xy}+Q_3}{s_x^2\rho_{xy}+Q_3}\right]$	$\frac{1-f}{n}S_y^2\left[(\kappa_x-1)\Psi_{12}\left(\Psi_{12}-\left(\frac{\lambda_{22}-1}{\kappa_x-1}\right)\right)\right]$	$\frac{1-f}{n}S_y^4\left[(\kappa_y-1)+\Psi_{12}(\kappa_x-1)\left(\Psi_{12}-2\left(\frac{\lambda_{22}-1}{\kappa_x-1}\right)\right)\right]$
$s_y^2\left[\frac{S_x^2+(TM+Q_a)}{s_x^2+(TM+Q_a)}\right]$	$\frac{1-f}{n}S_y^2\left[(\kappa_x-1)\Psi_{13}\left(\Psi_{13}-\left(\frac{\lambda_{22}-1}{\kappa_x-1}\right)\right)\right]$	$\frac{1-f}{n}S_y^4\left[(\kappa_y-1)+\Psi_{13}(\kappa_x-1)\left(\Psi_{13}-2\left(\frac{\lambda_{22}-1}{\kappa_x-1}\right)\right)\right]$
$s_y^2\{\frac{S_x^2\kappa_x+M_x^2}{s_x^2\kappa_x+M_x^2}\}$	$\frac{1-f}{n}S_y^2(\kappa_x-1)\varrho^*\{\varrho^*-\frac{(\lambda_{22}-1)}{(\kappa_x-1)}\}$	$\frac{1-f}{n}S_y^4\{(\kappa_y-1)+\varrho^*(\kappa_x-1)(\varrho^*-2\frac{(\lambda_{22}-1)}{(\kappa_x-1)})\}$

In general the Bias and MSE of existing modified ratio estimators $t_j, j=1,2,...,13$ is

$Bias(t_j)=\frac{1-f}{n}S_y^2(\kappa_x-1)\{\Psi_j(\Psi_j-(\frac{\lambda_{22}-1}{\kappa_x-1}))\}$

$MSE(t_j)=\frac{1-f}{n}S_y^4[(\kappa_y-1)+\Psi_j(\kappa_x-1)(\Psi_j-2(\frac{\lambda_{22}-1}{\kappa_x-1}))]$

where

$\Psi_1=0$; $\Psi_2=1$; $\Psi_3=\frac{S_x^2}{S_x^2+\kappa_x}$; $\Psi_4=\frac{S_x^2}{S_x^2-C_x}$; $\Psi_5=\frac{S_x^2}{S_x^2-\kappa_x}$;

$\Psi_6=\frac{S_x^2\kappa_x}{S_x^2\kappa_x-C_x}$; $\Psi_7=\frac{S_x^2C_x}{S_x^2C_x-\kappa_x}$; $\Psi_8=\frac{S_x^2}{S_x^2+M_x}$; $\Psi_9=\frac{S_x^2}{S_x^2+Q_1}$;

$\Psi_{10}=\frac{S_x^2}{S_x^2+Q_3}$; $\Psi_{11}=\frac{C_xS_x^2}{C_xS_x^2+M_x}$; $\Psi_{12}=\frac{S_x^2\rho_{xy}}{S_x^2\rho_{xy}+Q_3}$; $\Psi_{13}=\frac{S_x^2}{S_x^2+TM+Q_a}$.

CHAPTER FOUR

4 RESULTS AND DISCUSSION

4.1 Theoretical Evaluation

The theoretical conditions under which the proposed modified ratio type estimators $\hat{S}^2_{PM}$ is more efficient than the other existing estimators $t_j, j = 1, 2, ..., 13$, from MSE of $t_j, j = 1, 2, ..., 13$ given to first degree of approximation in general as

$$MSE(t_j) = \frac{1-f}{n} S_y^4[(\kappa_y - 1) + \Psi_j(\kappa_x - 1)(\Psi_j - 2(\frac{\lambda_{22} - 1}{\kappa_x - 1}))] \quad (93)$$

Using equation (89) and (93) we have that $MSE(\hat{S}^2_{PM}) < MSE(t_j)$,
if $\varrho^*(\varrho^* - 2(\frac{\lambda_{22}-1}{\kappa_x-1})) < \Psi_j(\Psi_j - 2(\frac{\lambda_{22}-1}{\kappa_x-1}))$

4.2 Numerical Studies

Using the data from Population I (Source:(Murthy, 1967, p.228)), Population II (source:(Daroga and Chaudhary, 1986, p.177)) and Population III (source:(Cochran, 1977, p.152)). We assess the performance of the proposed estimator when simple random sampling without replacement (SRSWOR) scheme is used with that of sample variance and existing estimators.We apply the proposed and existing estimators to this data set and the data summaries are given below:

Population I(Dataset in Appendix I)

X= Fixed capital

Y= output of 80 factories

$N = 80$, $n = 20$ $\bar{X} = 11.265$, $\bar{Y} = 51.826$,

$S_x^2 = 71.504$, $S_y^2 = 336.979$, $S_{xy} = 146.068$,

$\lambda_{04} = \kappa_x = 2.866$, $\lambda_{40} = \kappa_y = 2.267$, $\lambda_{22} = 2.221$,

$\rho_{xy} = 0.941, \qquad C_y = 0.354, \qquad C_x = 0.751$

$M_x = 10.300 \qquad Q_1 = 5.150 \qquad Q_3 = 16.975$

$TM = 10.68125, \qquad Q_a = 11.0625.$

Population II

X = acreage under wheat crop in 1973

Y = acreage under wheat crop in 1974 ,

$N = 70, \qquad n = 25 \qquad \bar{X} = 175.2671, \qquad \bar{Y} = 96.700,$

$S_x^2 = 19840.7508, \qquad S_y^2 = 3686.1898,$

$\lambda_{04} = \kappa_x = 7.0952, \qquad \lambda_{40} = \kappa_y = 4.7596, \qquad \lambda_{22} = 4.6038,$

$\rho_{xy} = 0.7293, \qquad C_y = 0.6254, \qquad C_x = 0.8037$

$M_x = 72.4375 \qquad Q_1 = 80.1500 \qquad Q_3 = 225.0250.$

$TM = 112.5125, \qquad Q_a = 152.5875.$

Population III(Data Set in Appendix II)

X = Total number of inhabitants in the 196 cities in 1920

Y = Total number of inhabitants in the 196 cities in 1930 ,

$N = 49, \qquad n = 20 \qquad \bar{X} = 98.6765, \qquad \bar{Y} = 116.1633,$

$S_x^2 = 10603.0063, \qquad S_y^2 = 9767.0922,$

$\lambda_{04} = \kappa_x = 5.9878, \qquad \lambda_{40} = \kappa_y = 4.9245, \qquad \lambda_{22} = 4.6977,$

$\rho_{xy} = 0.6904, \qquad C_y = 0.8508, \qquad C_x = 1.0435$

$M_x = 64.0000 \qquad Q_1 = 43.0000 \qquad Q_3 = 120.0000$

$TM = 72.75, \qquad Q_a = 81.5.$

Using these summary values to obtain the Bias and MSE of the existing estimators and our proposed estimator we have

4.2.1 Bias and Mean Squared Errors

Table 2: Bias and Mean Squared Errors(MSE) of Existing and Proposed Estimators for Population Variance

	Population I		Population II		Population III	
ESTIMATOR	BIAS(.)	MSE	BIAS(.)	MSE	BIAS(.)	MSE
$t_1 = s_y^2$	0	5395.289	0	1313625.261	0	11078650
$t_2 = s_y^2(\frac{S_x^2}{s_x^2})$	8.151	3276.421	236.154	924946.481	372.873	4282126
$t_3 = s_y^2(\frac{S_x^2+\kappa_x}{s_x^2+\kappa_x})$	6.956	2740.349	235.656	924324.375	371.849	4278020
$t_4 = s_y^2(\frac{S_x^2-C_x}{s_x^2-C_x})$	8.512	3006.373	236.187	925017.011	373.051	4282843
$t_5 = s_y^2(\frac{S_x^2-\kappa_x}{s_x^2-\kappa_x})$	9.518	3186.399	236.445	925569.577	373.898	4286246
$t_6 = s_y^2(\frac{S_x^2\kappa_x-C_x}{s_x^2\kappa_x-C_x})$	8.279	2965.067	236.159	924956.421	372.903	4282246
$t_7 = s_y^2(\frac{S_x^2C_x-\kappa_x}{s_x^2C_x-\kappa_x})$	10.002	3275.722	236.517	925721.916	373.856	4286074
$t_8 = s_y^2(\frac{S_x^2+M_x}{s_x^2+M_x})$	4.530	2377.418	233.201	918641.426	362.038	4238932
$t_9 = s_y^2(\frac{S_x^2+Q_1}{s_x^2+Q_1})$	6.126	2609.91	232.889	917976.121	365.567	4252936
$t_{10} = s_y^2(\frac{S_x^2+Q_3}{s_x^2+Q_3})$	2.934	2181.488	227.099	905689.896	352.748	4202378
$t_{11} = s_y^2(\frac{S_x^2C_x+M_x}{s_x^2C_x+M_x})$	3.656	2314.033	232.485	917116.922	362.485	4240702
$t_{12} = s_y^2\left[\frac{S_x^2\rho_{xy}+Q_3}{s_x^2\rho_{xy}+Q_3}\right]$	2.715	2158.326	223.826	898785.405	343.983	4168313.595
$t_{13} = s_y^2\left[\frac{S_x^2+(TM+Q_a)}{s_x^2+(TM+Q_a)}\right]$	2.034	2093.625	225.523	902361.029	347.151	4180578.085
$\hat{S}_{PM}^2 = s_y^2\{\frac{S_x^2\kappa_x+M_x^2}{s_x^2\kappa_x+M_x^2}\}$	0.708	1993.270	207.653	865134.030	268.201	3892407

From the above table Mean Squared Errors it is clear that our proposed modified ratio type population variance estimator $\hat{S}_{PM}^2$ has the least Mean Squared Error(MSE).

4.2.2 Efficiency Comparison

The performance of the proposed modified ratio type variance estimator evaluated against the usual unbiased estimator s_y^2 and the existing estimators $t_j, j = 1, 2, ..., 13$ using real population from (Murthy, 1967, p.228), (source:Daroga and Chaudhary (1986)) and (source:(Cochran, 1977, p.152)).

We have computed the Percent Relative Efficiencies (PREs) of the estimators $t_j, j = 1, 2, ..., 13$ using the formulae

$$\begin{aligned} PRE(t_j, s_y^2) &= \frac{MSE(s_y^2)}{MSE(t_j)} \times 100 && (94) \\ &= \{\frac{\frac{(1-f)}{n} S_y^4(\kappa_y - 1)}{\frac{1-f}{n} S_y^4[\{\kappa_y - 1\} + \{\kappa_x - 1\}\Psi_j(\Psi_j - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))]}\} \times 100 && (95) \\ &= \frac{(\kappa_y - 1)}{[\{\kappa_y - 1\} + \{\kappa_x - 1\}\Psi_j(\Psi_j - 2(\frac{\lambda_{22}-1}{\kappa_x - 1}))]} \times 100 && (96) \end{aligned}$$

Then PRE for our proposed estimator is subsequently,

$$\begin{aligned} PRE(\hat{S}_{PM}^2, s_y^2) &= \frac{MSE(s_y^2)}{MSE(\hat{S}_{PM}^2)} \times 100 && (97) \\ &= \frac{\frac{(1-f)}{n} S_y^4(\kappa_y - 1)}{\frac{1-f}{n} S_y^4\{(\kappa_y - 1) + \varrho^*(\kappa_x - 1)(\varrho^* - 2\frac{(\lambda_{22}-1)}{(\kappa_x - 1)})\}} \times 100 && (98) \\ &= \frac{(\kappa_y - 1)}{\{(\kappa_y - 1) + \varrho^*(\kappa_x - 1)(\varrho^* - 2\frac{(\lambda_{22}-1)}{(\kappa_x - 1)})\}} \times 100 && (99) \end{aligned}$$

Using formula (96) and (99) we computed the Percent Relative Efficiencies and presented in table 3 below

Percent Relative Efficiencies

Table 3: Percent Relative Efficiencies(PRE) of Existing and Proposed Estimators

ESTIMATOR	POPULATION I	POPULATION II	POPULATION III
t_1	100	100	100
t_2	164.67	142.02	258.72
t_3	196.88	142.12	258.97
t_4	179.46	142.01	258.66
t_5	169.32	141.93	258.47
t_6	181.96	142.02	258.71
t_7	164.71	141.90	258.48
t_8	226.94	143.00	261.35
t_9	206.72	143.10	260.49
t_{10}	247.32	145.04	263.63
t_{11}	233.16	143.23	261.25
t_{12}	249.98	146.16	265.78
t_{13}	257.70	145.58	265.00
$\hat{S}^2_{PM}$	270.68	151.84	284.622

From the findings summarized in the table above it is clear that our proposed estimator $\hat{S}^2_{PM}$ performed best, that is it has the highest PRE among all the other estimators.

CHAPTER FIVE

5 CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

In this study we have suggested a modified ratio type estimator of population variance S_y^2 of the study variable y using known population parameters of the auxiliary variable x, the coefficient of kurtosis and the median. The bias and mean squared error of the proposed estimator has been obtained to first order degree of approximation and consequently compared with that of the usual unbiased estimator and the estimators due to Isaki (1983), Kadilar and Cingi (2006),Subramani and Kumarapandiyan (2013), Subramani and Kumarapandiyan (2012a), Subramani and Kumarapandiyan (2012b), Upadhyaya and Singh (1999) Khan and Shabbir (2013b) and Bhat et al. (2017).

We have also assessed the performance of our proposed estimator using known natural population data sets and found out that the performance of our proposed estimator is better than the other existing estimators for the data sets by comparing their Percent Relative Efficiencies. Based on the results of our studies, it is evident that our proposed estimator has the highest Percent Relative Efficiency.

5.2 Recommendations

We recommend that our proposed estimator can be applied to practical applications, where knowledge of population parameters of auxiliary variable positively correlated with study variable is available. We further recommend that our proposed estimator can be improved by extending the number of Taylor's series terms to be more than order one or be protracted to Stratified Sampling Scheme.

References

Arcos, A., Rueda, M., and Martinez-Miranda, M. (2005). Using multiparametric auxiliary information at the estimation stage. *Statistical Papers*, 46(3):339–358.

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

Bhat, M., Maqbool, S., and Raja, T. (2017). Variance estimation using linear combination of tri-mean and quartiles. *International Research Journal of Agricultural Economics and Statistics*, 8(1):143–145.

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(02):262–275.

Cochran, W. G. (1977). *Sampling Techniques: 3d Ed*. Wiley.

Daroga, S. and Chaudhary, F. (1986). *Theory and Analysis of sample survey designs*. New York, Wiley Eastern Limited.

Das, A. (1988). *Contributions to the theory of sampling strategies based on auxiliary information, unpublished*. PhD thesis, Bidhan Chandra Krishi Vishwavidyalaya, Mohanpur, Nadia, West Bengal, India.

Das, A. K. and Tripathi, T. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya*, 40(C):139–148.

Evans, W. D. (1951). On the variance of estimates of the standard deviation and variance. *Journal of the American Statistical Association*, 46(254):220–224.

Grover, L. K. (2010). A correction note on improvement in variance estimation using auxiliary information. *Communications in Statistics?Theory and Methods*, 39(5):753–764.

Gupta, S. and Shabbir, J. (2008). Variance estimation in simple random sampling using auxiliary information. *Hacettepe Journal of mathematics and Statistics*, 37(1):57–67.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample survey methods and theory*, volume 1. Wiley New York.

Hirano, K., Pandey, B. N., and Singh, J. (1973). On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. *Annals of the Institute of Statistical Mathematics*, 25(1):51–55.

Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, 78(381):117–123.

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

Karr, A. (1993). Probability theory, ser.

Kaur, P. and Singh, G. (1982). A note on estimating variance in a finite population. *J. Stat. Res*, 16.

Khan, M. (2015). An improved exponential-type estimator for finite population variance using transformation of variables. *Journal of Statistics Applications & Probability*, 4(1):31–35.

Khan, M. and Shabbir, J. (2013a). A ratio type estimator for the estimation of population variance using quartiles of an auxiliary variable. *Journal of Statistics Applications and Probability*, 2(3):157–162.

Khan, M. and Shabbir, J. (2013b). A ratio type estimator for the estimation of population variance using quartiles of an auxiliary variable. *Journal of Statistics Applications & Probability*, 2(3):319–325.

Murthy, M. N. (1967). *Sampling theory and methods.* Calcutta-35: Statistical Publishing Society, 204/1, Barrackpore Trunk Road, India.

Prasad, B. and Singh, H. (1990). Some improved ratio-type estimators of finite population variance in sample surveys. *Communications in Statistics-Theory and Methods*, 19(3):1127–1139.

Sarndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. Springer-Verlag.

Searls, D. T. (1964). The utilization of a known coefficient of variation in the estimation procedure. *Journal of the American Statistical Association*, 59(308):1225–1226.

Searls, D. T. and Intarapanich, P. (1990). A note on an estimator for the variance that utilizes the kurtosis. *The American Statistician*, 44(4):295–296.

Sen, A. R. (1978). Estimation of the population mean when the coefficient of variation is known. *Communications in Statistics - Theory and Methods*, 7(7):657–672.

Singh, H. P., Chandra, P., and Singh, S. (2003). Variance estimation using multiauxiliary information for random non-response in survey sampling. *Statistica*, 63(1):23–40.

Singh, H. P., Tailor, R., Singh, S., and Kim, J.-M. (2011). Estimation of population variance in successive sampling. *Quality & Quantity*, 45(3):477–494.

Singh, H. P., Tailor, R., Tailor, R., and Kakran, M. (2004). An improved estimator of population mean using power transformation. *Journal of the Indian Society of Agricultural Statistics*, 58(2):223–230.

Singh, S. (1991). Estimation of finite population variance using double sampling. *Aligarh J. Statist*, 11:53–65.

Singh, S. (2001). Generalized calibration approach for estimating variance in survey sampling. *Annals of the Institute of Statistical Mathematics*, 53(2):404–417.

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

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

Srivastava, S. K. and Jhajj, H. S. (1980). A class of estimators using auxiliary information for estimating finite population variance. *Sankhya, C*, 42(1-2):87–96.

Srivastava, S. K. and Jhajj, H. S. (1981). A class of estimators of the population mean in survey sampling using auxiliary information. *Biometrika*, pages 341–343.

Subramani, J. and Kumarapandiyan, G. (2012a). Variance estimation using median of the auxiliary variable. *International Journal of Probability and Statistics*, 1(3):36–40.

Subramani, J. and Kumarapandiyan, G. (2012b). Variance estimation using quartiles and their functions of an auxiliary variable. *International Journal of Statistics and Applications*, 2(5):67–72.

Subramani, J. and Kumarapandiyan, G. (2013). Estimation of variance using known coefficient of variation and median of an auxiliary variable. *Journal of Modern Applied Statistical Methods*, 12(1):58–64.

Sukhatme, P. (1944). Moments and product moments of moment-statistics for samples of the finite and infinite populations. *Sankhyā: The Indian Journal of Statistics*, 6:363–382.

Sukhatme, P. and Sukhatme, B. (1970). *Sampling theory of surveys with applications:2nd Ed.* Bombay,India,Asia Publishing House.

Tracy, D. S. (1984). Moments of sample moments. *Communications in Statistics-Theory and Methods*, 13(5):553–562.

Upadhyaya, L. and Singh, G. (2001). Chain-type estimators using transformed auxiliary variable in two-phase sampling. *ADVANCES IN MODELLING AND ANALYSIS-A-*, 38(1/2):1a–10a.

Upadhyaya, L. and Singh, H. (1983). Use of auxiliary information in the estimation of population variance. In *Mathematical forum*, volume 4, pages 33–36.

Upadhyaya, L. and Singh, H. (1999). An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample surveys. *Vikram Mathematical Journal*, 19(1):14–17.

Upadhyaya, L. N. and Singh, H. P. (1984). On the estimation of the population mean with known coefficient of variation. *Biometrical Journal*, 26(8):915–922.

Withers, C. S. and Nadarajah, S. (2014). Unbiased estimates for products of moments and cumulants for finite and infinite populations. *arXiv preprint arXiv:1410.7154*.

Wolter, K. (2007). *Introduction to variance estimation.* New York, NY,Springer Science & Business Media.

Yadav, S. K., Misra, S., and Mishra, S. (2016). Efficient estimator for population variance using auxiliary variable. *American Journal of Operational Research*, 6(1):9–15.

APPENDICES

Appendix I, Murthy Population Data Set

x_i	y_i	x_i	y_i	x_i	y_i
2.04	13.5	4.52	41.09	15	65.67
2.04	11.76	4.64	42.16	15.48	67.19
2.08	18.41	4.76	49.5	17	67.52
2.08	26.06	4.84	43.02	17.72	66.6
2.12	26.56	5	43.85	18.08	68.54
2.16	25.46	5.08	44.26	18.64	67.6
2.28	29.11	5.08	45.3	19.24	68.25
2.4	32.8	5.24	46.89	19.8	69.4
2.6	34.25	5.36	53.86	21.12	72.95
2.68	34.16	5.4	49.61	21.76	70.7
2.72	33.9	5.56	48.22	22.52	71.52
2.8	33.95	5.76	50.97	23.4	71.86
2.84	34.17	6.08	51.24	23.92	72.15
2.92	32.9	6.4	52.86	25.76	72.88
2.96	34.81	6.48	51.13	26.68	75.4
3.04	35.2	6.64	52.3	28.2	74.16
3.12	35.7	6.92	53.3	30	76.1
3.2	37.4	7.4	47.62	31	78.94
3.24	35.2	7.68	54.2	32.96	80.63
3.4	36.01	7.92	55.62	34.8	81.8
3.48	37.17	8.44	56.3	36.52	83.15
3.52	37.5	9.68	55.82	38.04	85.76
3.68	37.3	10.12	56.84	39.2	86.75
3.72	37.67	11.4	57.9	43.8	92.5
3.88	38.21	11.64	58.39		
4	38.86	12.56	59.2		
4.28	39.72	13.4	63.15		
4.4	40.65	14.08	65.1		

Appendix II, Daroga Population Data Set

X_i	Y_i	X_i	Y_i
58.73	24.24	37.76	13.09
136.77	72.24	473.24	249.70
268.50	137.70	199.70	120.73
369.19	184.73	77.19	41.21
209.77	134.79	207.25	107.15
104.88	53.82	112.44	64.49
468.20	307.40	109.92	69.82
213.12	134.79	108.24	49.94
84.75	54.30	159.42	84.85
301.23	172.12	304.58	162.43
91.46	48.00	197.18	106.18
403.59	241.46	61.25	30.06
104.88	53.82	52.02	38.30
4.20	2.91	59.57	29.09
358.28	164.37	114.95	48.49
65.45	38.79	164.46	68.36
62.93	50.91	213.96	127.52

Appendix III, Cochran Population Data Set

xi	yi	xi	yi	xi	yi
76	80	2	50	243	291
138	143	507	634	87	105
67	67	179	260	30	111
29	50	121	113	71	79
381	464	50	64	256	288
23	48	44	58	43	61
37	63	77	89	25	57
120	115	64	63	94	85
61	69	64	77	43	50
387	459	56	142	298	317
93	104	40	60	36	46
172	183	40	64	161	232
78	106	38	52	74	93
66	86	136	139	45	53
60	57	116	130	36	54
46	65	46	53	50	58
				48	75