University of the Punjab Lahore

DEVELOPMENT OF MULTIVARIATE ESTIMATORS IN MULTI-PHASE SAMPLING AND THEIR APPLICATIONS

by

NADEEM SHAFIQUE BUTT

DOCTOR OF PHILOSOPHY

IN

STATISTICS

UNIVERSITY OF THE PUNJAB

DEVELOPMENT OF MULTIVARIATE ESTIMATORS IN MULTI-PHASE SAMPLING AND THEIR APPLICATIONS

By

NADEEM SHAFIQUE BUTT

A dissertation submitted to

College of Statistical and Actuarial Sciences

University of the Punjab, Lahore

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

IN

STATISTICS

UNIVERSITY OF THE PUNJAB, LAHORE

DEVELOPMENT OF MULTIVARIATE ESTIMATORS IN MULTI-PHASE SAMPLING AND THEIR APPLICATIONS

BY

Nadeem Shafique Butt

A dissertation submitted to College of Statistical and Actuarial Sciences, University of the Punjab in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN STATISTICS

Dissertation Committee:

____
(Chairman)

____
(Member)

____
(Member)

____
Principal
College of Statistical and Actuarial Sciences, University of the Punjab, Lahore

DECLARATION

This is to certify that this research work has not been submitted and shall not be submitted in future for obtaining similar degree from any other university / college.

NADEEM SHAFIQUE BUTT

DEDICATED

TO

My friend Muhammad Qaiser Shahbaz who is always source of Inspiration and Encouragement for me

ACKNOWLEDGEMENT

All praise for ***Allah Almighty***, the most magnificent, the most beneficent. I pray to Allah for His guidance and Mercy throughout my life. It is with His grace and help that I have been able to complete this work.

First and foremost I offer my sincerest gratitude to my supervisor, Prof. Dr. Shahid Kamal (Principal College of Statistical and Actuarial Sciences, University of the Punjab, Lahore) who has supported me throughout my dissertation with his guidance, knowledge and encouragement. I am also thankful to Dr. Muhammad Qasier Shahbaz (Associate Professor, CIIT, Lahore) as this work would not have been completed without his thought provoking ideas, guidance, valuable comments and critical appreciation.

I am also thankful to all faculty members of College of Statistical and Actuarial Sciences, University of the Punjab for their support. Special thanks to Dr. Ahmed Saeed Akhter, Dr. Ghausia Masood Gillani and Mr. Muhammad Ahmad for their kind guidance and encouragement.

I am highly indebted to the prayers of my mother that have enabled me to complete this work. Thanks to my wife, brother, Bhabi and sisters for their encouragement. I am also thankful to all my teachers for providing me their guidance to accomplish this task.

Finally, I would like to thank my friends Yasir Hassan, Muhammad Farooq, Asif Iftikhar, Javid Akhter, Imtiaz-ul-Din, Munawar Iqbal Rana and Rehan Ahmad Khan for their moral support. They are always source of encouragement and guidance for me.

Nadeem Shafique Butt

RESEARCH COMPLETION CERTIFICATE

Certified that the research work contained in this thesis entitled **"Development of Multivariate Estimators in Multi-Phase Sampling and their Applications"** has been carried out and completed by **"Nadeem Shafique Butt"** under my supervision during his PhD**.** Statistics programme.

(Dr. Shahid Kamal)

Supervisor

SUMMARY

This thesis deals with the development of new univariate and multivariate estimators for single-phase, two-phase and multi-phase sampling based on auxiliary variables and as well as auxiliary attributes. Some available popular estimators have been discussed in chapter 1 and 2 of this thesis. In chapter 3 new univariate estimators for two phase sampling has been proposed. The proposed estimators are extension of the estimator proposed by Roy(2003). The proposed estimator use information on multiple auxiliary variables as well as on multiple auxiliary attributes. Shrinkage versions of proposed estimators have also been given in Chapter 3. The empirical study of proposed estimators has been conducted see its performance as compared with classical regression estimator. It has been observed that the proposed estimators are always more precise as compared with classical regression estimator for both quantitative and qualitative auxiliary variables.

In chapter 4 new multivariate estimators for two phase sampling has been proposed which are the multivariate versions of Roy(2003) estimator. The proposed estimators use information on multiple quantitative variables and as well as multiple qualitative variables. The empirical study based on Eigen values of variance-covariance matrices has also been conducted to see the performance of the proposed estimators over the estimator proposed by Ahmed, Hussin, & Hanif(2010). The results of empirical study shows that proposed estimator perform far better than the multivariate regression estimator proposed by Ahmed, et al.(2010)

Multivariate estimators proposed in chapter 4; as well as proposed by Ahmed, et al.(2010); for simultaneous estimation of several study variables require that all variables depend upon same set of auxiliary variables $\mathbf{X}$. This situation is not always feasible as different response variables may depend on different set of predictors. In this situation different estimation mechanisms are required. The seemingly unrelated regression models of Zellner(1962) have been popular models for simultaneous prediction of multiple response variables which depends on different set of predictors. The concept of seemingly unrelated regression models has been used for simultaneous estimation of multiple response variables which depends on different predictors. Seemingly Unrelated Regression Estimators (SURE's) have been proposed in Chapter 5 of this thesis. SURE has been developed for Sing-Phase, two-Phase and Multiphase sampling. The applicability of SURE's is much wider as compared with multivariate regression estimators available in literature.

Table of Contents

Chapter 1: Introduction

Estimation problem has always been vital in all the domains of life. Effective planning depends upon preciseness of the estimates that's why researchers are always in process of developing methods that can produce more precise estimates. Several methods are available in literature that can be used for efficient estimation of the characteristic under study, these methods are collectively called sampling methods. The scientific development in the field of survey sampling has long history but the groundbreaking work in this field was done by Neyman(1934). The work of Neyman(1934) guided the number of statisticians for significant development in various areas of survey sampling. The historical work done by Hansen & Hurwitz(1943) and by Horvitz & Thompson(1952) in the development of unequal probability sampling is also based upon the ideas given by Neyman(1934).

History of sampling has been discussed by many survey statisticians, some notable references are Chang(1976), Dalenius(1962), Duncan & Shelton(1978), Hansen(1987), Kruskal & Mosteller(1980), Seng(1951) and Stephan(1948). Kiaer(1895) in a meeting of International Statistical Institute (ISI) put forward the idea that a partial investigation could provide useful information. Detailed discussions of Kiaer(1895) work and its impact on sampling methodology may be found in Seng(1951) and Kruskal & Mosteller(1980). The initial reaction to Kiaer(1895) work was negative and generally not receptive; however in 1901 and 1903 Kiaer was supported by C. D. Wright and later by A. L. Bowley. Kiaer(1897) mentions the possibility of randomization, in his words a sample 'selected through the drawing of lots', but does not develop the idea further in his writings.

Like Kiaer, Bowley actively promoted his ideas on sampling and randomization specially. Bowley(1906) paper containing an empirical verification to simple random sampling, at this point Bowley has probably equated random sampling to any sampling scheme in which the inclusion probabilities are the same for every sampling units. Bowley(1913) used a systematic sample of buildings of "Reading" from street listing in the local directory of residential buildings. Bowley(1913) also checked the representativeness of his samples by comparing his sample results to known population counts. For two cases in which Bowley(1913) found a discrepancy between his sample and official statistics, on further checking it was discovered that the official statistics contained error. This work is discussed in details by Seng(1951) and Kruskal & Mosteller(1980). Also Bowley(1926) provided a theoretical monograph summarizing the known results in random and purposive selection.

The work of Neyman(1934) paper has been recognized as an important contribution to the field of survey sampling. Kruskal & Mosteller(1980) have discussed the work as “the Neyman watershed” and Hansen, Dalenius, & Tepping(1985) have commented that the “paper played a paramount role in promoting theoretical research, developments, and application of what is now known as probability sampling”. The work of Neyman(1934) is considered as a classic work on two grounds; Firstly Neyman was able to provide valid reasons, both theoretically and with practical examples that why randomization gave a much more reasonable solution than purposive selection. Secondly, the paper provides a paradigm in the history of sampling is that the theory of point and interval estimation is provided under randomization. Neyman(1938) introduced the use of cost function into survey sampling in connection with two-phase sampling.

In early 1940’s Hansen & Hurwitz(1943) made some fundamental contribution to theory of sampling, they took an important step forward by extending the idea of sampling with unequal inclusion probabilities for units in different strata. This allowed the development of very complex multi-stage designs that are the backbone of large scale social and economic survey research.

1.1 Introduction to Multiphase Sampling

Information related to the variable of interest is termed as auxiliary information, which can be utilized to improve the efficiency of the estimators. In Multiphase sampling certain items of information are drawn from the whole sampling units and certain other items of information are taken from the subsample. Multiphase sampling is used when it is expensive to collect data on the variable of interest but it is relatively inexpensive to collect data on variables that are correlated with the variables of interest. For example, in forest surveys, it is difficult to travel to remote area to make on ground determination. However, aerial photographs of forest are relatively inexpensive, which can be used to decide about forest type; a strongly correlated variable with ground determination.

1.1.1 Notation of Multiphase Sampling

Let a population of N units is designated as $U_1, U_2, \cdots, U_N$, Y_I is the value of variable of interest associated with Y_{ij}, X_I and W_I be information of auxiliary variables associated with U_{ij} $I = 1,2,3,.....,N$. Let $\bar{X}$, $\bar{W}$ and $\bar{Y}$ be population mean of *X*, *W* and *Y* respectively. Further let S_x^2, S_w^2 and S_y^2 are corresponding variances. Also ρ_{xy}, ρ_{xw} and ρ_{yw} are population correlation coefficients between *X* & *Y*, *X* & *W* and *Y* & *W* respectively. Let a first phase sample of n_1 units is drawn from the population and information on auxiliary variables is recoded; further let a sub-sample of $n_2 < n_1$ units is drawn from the first phase sample and information of auxiliary variables alongside variable of interest.

The sample means of auxiliary variables based on n_1 units are denoted by $\bar{x}_1$ and $\bar{w}_1$ etc and sample means of second phase are denoted by $\bar{x}_2$ and $\bar{w}_2$. We will also use $\theta_1 = n_1^{-1} - N^{-1}$ and $\theta_2 = n_2^{-1} - N^{-1}$ such that $\theta_2 > \theta_1$. For notational purpose it will be assumed that the mean of estimand and auxiliary variables can be approximated from their population means so that $\bar{x}_h = \bar{X} + \bar{e}_{x_h}$ where $\bar{x}_h$ is sample mean of auxiliary variable *X* at *h*–th phase; *h* = 1 and 2. Similar notation will be used for other quantitative auxiliary variables. For qualitative auxiliary variables we will use $p_h = P + \bar{e}_{\tau_h}$ will be used and for variable of interest we will use the notation $\bar{y}_h = \bar{Y} + \bar{e}_{y_h}$ with usual assumptions.

Following expectations for deriving the mean square error of estimators which are based upon quantitative auxiliary variables will be used:

$$
\left.\begin{aligned}
&\bar{\mathbf{w}}_1=\bar{\mathbf{w}}+\bar{\mathbf{e}}_{w_1};\ \bar{\mathbf{w}}_2=\bar{\mathbf{w}}+\bar{\mathbf{e}}_{w_2}\\
&\bar{x}_1=\bar{x}+\bar{e}_{\bar{x}_1};\ \bar{x}_2=\bar{x}+\bar{e}_{\bar{x}_2};\ \bar{y}_2=\bar{Y}+\bar{e}_{\bar{y}2}\\
&E\left(\bar{e}_{y_1}^2\right)=\theta_1 S_y^2;\ E\left(\bar{e}_{y_2}^2\right)=\theta_2 S_y^2\\
&E\left(\bar{e}_{x_1}^2\right)=\theta_1 S_x^2;\ E\left(\bar{e}_{x_2}^2\right)=\theta_2 S_x^2\ ;\ E\left(\bar{e}_{x_1}-\bar{e}_{x_2}\right)^2=\left(\theta_2-\theta_1\right)S_x^2\\
&E\left(\bar{e}_{x_1}\bar{e}_{y_1}\right)=\theta_1 S_{xy}\ ;\ E\left(\bar{e}_{x_1}\bar{e}_{y_2}\right)=E\left(\bar{e}_{x_2}\bar{e}_{y_1}\right)=\theta_1 S_{xy};\ E\left(\bar{e}_{x_2}\bar{e}_{y_2}\right)=\theta_2 S_{xy}\\
&E\left(\bar{e}_{\mathbf{w}_1}\bar{e}_{\mathbf{w}_1}^{/}\right)=\theta_1\mathbf{S}_{\mathbf{w}}\ ;\ E\left(\bar{e}_{\mathbf{w}_1}\bar{e}_{\mathbf{w}_2}^{/}\right)=E\left(\bar{e}_{\mathbf{w}_2}\bar{e}_{\mathbf{w}_1}^{/}\right)=\theta_1\mathbf{S}_{\mathbf{w}}\ ;\ E\left(\bar{e}_{\mathbf{w}_2}\bar{e}_{\mathbf{w}_2}^{/}\right)=\theta_2\mathbf{S}_{\mathbf{w}}\\
&E\left(\bar{e}_{y_2}\bar{e}_{\mathbf{w}_1}\right)=\theta_1\mathbf{s}_{y\mathbf{w}}\ ;\ E\left(\bar{e}_{y_2}\bar{e}_{\mathbf{w}_2}\right)=\theta_2\mathbf{s}_{y\mathbf{w}}\\
&E\left(\bar{e}_{x_1}\bar{e}_{\mathbf{w}_1}\right)=\theta_1\mathbf{s}_{x\mathbf{w}}\ ;\ E\left(\bar{e}_{x_1}\bar{e}_{\mathbf{w}_2}\right)=E\left(\bar{e}_{x_2}\bar{e}_{\mathbf{w}_1}\right)=\theta_1\mathbf{s}_{x\mathbf{w}}\ ;\ E\left(\bar{e}_{x_2}\bar{e}_{\mathbf{w}_2}\right)=\theta_2\mathbf{s}_{x\mathbf{w}}
\end{aligned}\right\}\quad (1.1.1.1)
$$

In case of several auxiliary variables; say q; the sample mean of i-th auxiliary at h-th phase will be denoted by $\bar{x}_{(i)h}=\bar{X}_i+\bar{e}_{x_{(i)h}}$. The vector notations in case of multiple auxiliary variables and sample mean vector of auxiliary variables at h-th phase will be denoted by $\bar{\mathbf{x}}_h$ with relation $\bar{\mathbf{x}}_h=\bar{\mathbf{X}}+\bar{\mathbf{e}}_{x_h}$. Following additional expectations are also useful:

$$
E\left(\bar{\mathbf{e}}_{x_h}\bar{\mathbf{e}}_{x_h}^{/}\right)=\theta_h\mathbf{S}_x\ ,\ E\left(\bar{e}_{y_1}\bar{\mathbf{e}}_{x_h}\right)=\theta_1\mathbf{s}_{yx}\ ;h\geq 1
$$

Similar expectations for qualitative auxiliary variables are:

$$
\left.\begin{aligned}
&\mathbf{p}_{\delta_1}=\mathbf{p}_{\delta}+\bar{\mathbf{e}}_{\delta_1};\ \mathbf{p}_{\delta_2}=\mathbf{p}_{\delta}+\bar{\mathbf{e}}_{\delta_2}\\
&p_{\tau_1}=p_{\tau}+\bar{e}_{\tau_1};p_{\tau_2}=p_{\tau}+\bar{e}_{\tau_2};\ \bar{y}_2=\bar{Y}+\bar{e}_{\bar{y}2}\\
&E\left(\bar{e}_{y_1}^2\right)=\theta_1 S_y^2;\ E\left(\bar{e}_{y_2}^2\right)=\theta_2 S_y^2\\
&E\left(\bar{e}_{\tau_1}^2\right)=\theta_1 S_{\tau}^2;\ E\left(\bar{e}_{\tau_2}^2\right)=\theta_2 S_{\tau}^2\ ;\ E\left(\bar{e}_{\tau_1}-\bar{e}_{\tau_2}\right)^2=\left(\theta_2-\theta_1\right)S_{\tau}^2\\
&E\left(\bar{e}_{\tau_1}\bar{e}_{y_1}\right)=\theta_1 S_{\tau y}\ ;\ E\left(\bar{e}_{\tau_1}\bar{e}_{y_2}\right)=E\left(\bar{e}_{\tau_2}\bar{e}_{y_1}\right)=\theta_1 S_{\tau y};\ E\left(\bar{e}_{\tau_2}\bar{e}_{y_2}\right)=\theta_2 S_{\tau y}\\
&E\left(\bar{e}_{\boldsymbol{\delta}_1}\bar{e}_{\boldsymbol{\delta}_1}^{/}\right)=\theta_1\mathbf{S}_{\boldsymbol{\delta}}\ ;\ E\left(\bar{e}_{\boldsymbol{\delta}_1}\bar{e}_{\boldsymbol{\delta}_2}^{/}\right)=E\left(\bar{e}_{\boldsymbol{\delta}_2}\bar{e}_{\boldsymbol{\delta}_1}^{/}\right)=\theta_1\mathbf{S}_{\boldsymbol{\delta}}\ ;\ E\left(\bar{e}_{\boldsymbol{\delta}_2}\bar{e}_{\boldsymbol{\delta}_2}^{/}\right)=\theta_2\mathbf{S}_{\boldsymbol{\delta}}\\
&E\left(\bar{e}_{y_2}\bar{e}_{\boldsymbol{\delta}_1}\right)=\theta_1\mathbf{s}_{y\boldsymbol{\delta}}\ ;\ E\left(\bar{e}_{y_2}\bar{e}_{\boldsymbol{\delta}_2}\right)=\theta_2\mathbf{s}_{y\boldsymbol{\delta}}\\
&E\left(\bar{e}_{\tau_1}\bar{e}_{\boldsymbol{\delta}_1}\right)=\theta_1\mathbf{s}_{\tau\boldsymbol{\delta}}\ ;\ E\left(\bar{e}_{\tau_1}\bar{e}_{\boldsymbol{\delta}_2}\right)=E\left(\bar{e}_{\tau_2}\bar{e}_{\boldsymbol{\delta}_1}\right)=\theta_1\mathbf{s}_{\tau\boldsymbol{\delta}}\ ;\ E\left(\bar{e}_{\tau_2}\bar{e}_{\boldsymbol{\delta}_2}\right)=\theta_2\mathbf{s}_{x\boldsymbol{\delta}}
\end{aligned}\right\}\quad (1.1.1.2)
$$

For multivariate and Zelner estimator we will use following notations:

$$\left.\begin{aligned}
&E\left(\overline{\mathbf{e}}_{\mathbf{y}_1}\overline{\mathbf{e}}_{\mathbf{y}_1}^{/}\right)=\theta_1\mathbf{S}_{\mathbf{y}}\ ;\ E\left(\overline{\mathbf{e}}_{\mathbf{x}_1}\overline{\mathbf{e}}_{\mathbf{x}_1}^{/}\right)=\theta_1\mathbf{S}_{\mathbf{x}}\\
&E\left(\overline{\mathbf{e}}_{\mathbf{y}_2}\overline{\mathbf{e}}_{\mathbf{y}_2}^{/}\right)=\theta_2\mathbf{S}_{\mathbf{y}}\ ;\ E\left(\overline{\mathbf{e}}_{\mathbf{x}_2}\overline{\mathbf{e}}_{\mathbf{x}_2}^{/}\right)=\theta_2\mathbf{S}_{\mathbf{x}}\\
&E\left(\overline{\mathbf{e}}_{\mathbf{y}_1}\overline{\mathbf{e}}_{\mathbf{x}_1}^{/}\right)=\theta_1\mathbf{S}_{\mathbf{yx}}\ ;\ E\left(\overline{\mathbf{e}}_{\mathbf{x}_1}\overline{\mathbf{e}}_{\mathbf{y}_1}^{/}\right)=\theta_1\mathbf{S}_{\mathbf{xy}}\\
&E\left(\overline{\mathbf{e}}_{\mathbf{y}_2}\overline{\mathbf{e}}_{\mathbf{x}_2}^{/}\right)=\theta_2\mathbf{S}_{\mathbf{yx}}\ ;\ E\left(\overline{\mathbf{e}}_{\mathbf{x}_2}\overline{\mathbf{e}}_{\mathbf{y}_2}^{/}\right)=\theta_2\mathbf{S}_{\mathbf{xy}}\\
&E\left(\overline{\mathbf{e}}_{\mathbf{y}_1}\overline{\mathbf{e}}_{\mathbf{x}_2}^{/}\right)=\theta_1\mathbf{S}_{\mathbf{yx}}=E\left(\overline{\mathbf{e}}_{\mathbf{y}_2}\overline{\mathbf{e}}_{\mathbf{x}_1}^{/}\right)\\
&E\left(\overline{\mathbf{e}}_{\mathbf{x}_1}\overline{\mathbf{e}}_{\mathbf{y}_2}^{/}\right)=\theta_1\mathbf{S}_{\mathbf{xy}}=E\left(\overline{\mathbf{e}}_{\mathbf{x}_2}\overline{\mathbf{e}}_{\mathbf{y}_1}^{/}\right)
\end{aligned}\right\}\qquad(1.1.1.3)$$

where $\mathbf{S}_{\mathrm{y}}$ is covariance matrix of variables of interest, the notation of $t_{Nm(2)}$ will be used for new estimators under two phase sampling.

1.2 Some Popular Univarate Estimators in Multiphase Sampling based on Quantitative Predictors

In this section some well-known ratio and regression estimators for estimating population mean along with their mean square errors for two-phase sampling using one and two quantitative auxiliary variables are discussed.

The traditional ratio estimator for unknown population mean $\bar{Y}$ suggested by Cochran(1977) is

$$t_1 = \frac{\bar{y}_2}{\bar{x}_2}\bar{x}_1 \qquad (1.2.1)$$

and its mean square error is

$$MSE(t_1) = \bar{Y}^2\left[\theta_2 C_y^2\left(1-\rho_{yx}^2\right)+(\theta_2-\theta_1)\left(C_x - C_y\,\rho_{xy}\right)^2 + \theta_1 C_y^2\rho_{yx}^2\right], \qquad (1.2.2)$$

where C_x is coefficient of variation of x and ρ_{xy} is correlation coefficient between x and y.

Cochran(1977) suggested following simple regression estimator for unknown $\bar{Y}$ as under:

$$t_2 = \bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right), \qquad (1.2.3)$$

where b_{yx} is based on second phase sample and expression for mean square error is

$$MSE(t_2) = \bar{Y}^2 C_y^2\left[\theta_2\left(1-\rho_{xy}^2\right)+\theta_1\,\rho_{xy}^2\right]. \qquad (1.2.4)$$

Another simple regression estimator of $\bar{Y}$; suggested by Cochran(1977) when $\bar{X}$ is known:

$$t_3 = \bar{y}_2 + b_{yx}\left(\bar{X} - \bar{x}_2\right), \qquad (1.2.5)$$

With mean square error:

$$MSE(t_3) = \theta_2\bar{Y}^2 C_y^2\left(1-\rho_{yx}^2\right). \qquad (1.2.6)$$

It can be immediately seen that $MSE(t_3) < MSE(t_2)$.

Mohanty(1967) suggested the following regression-cum-ratio estimator by combining the regression and ratio method when information on $\bar{X}$ is not available:

$$t_4 = \left[\bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right)\right]\frac{\bar{Z}}{\bar{z}_2}, \qquad (1.2.7)$$

where b_{yx} is calculated from second phase sample and the expression for mean square error of t_4

$$MSE(t_4) = \bar{Y}^2\left[\theta_2\left\{C_y^2\left(1-\rho_{xz}^2\right)+\left(C_z - C_y\rho_{yz}\right)^2\right\}\right.$$
$$\left.+(\theta_2-\theta_1)\left\{C_z^2\rho_{xz}^2 - \left(C_y\rho_{xy} - C_z\rho_{xz}\right)^2\right\}\right]. \quad (1.2.8)$$

Another regression-cum-ratio estimator suggested by Mohanty(1967) when information on both auxiliary variables is unavailable:

$$t_5 = \left[\bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right)\right]\frac{\bar{z}_1}{\bar{z}_2}, \quad (1.2.9)$$

where b_{yx} is calculated from second phase sample and expression for mean square error is

$$MSE(t_5) = \bar{Y}^2\left[\theta_2 C_y^2 + (\theta_2-\theta_1)\left\{\rho_{xz}^2 C_z^2 - \left(\rho_{xy} C_y - \rho_{xz} C_z\right)^2\right.\right.$$
$$\left.\left.+\left(C_z - C_y\rho_{yz}\right)^2 - C_y^2\rho_{yz}^2\right\}\right]. \quad (1.2.10)$$

The chain ratio-type estimator proposed by Chand(1975) for two-phase sampling using two-auxiliary variables; when population mean $\bar{Z}$ is known:

$$t_6 = \bar{y}_2\frac{\bar{x}_1}{\bar{x}_2}\frac{\bar{Z}}{\bar{z}_1} \quad (1.2.11)$$

and its mean square error is:

$$MSE(t_6) = \bar{Y}^2\left[\theta_2 C_y^2 + (\theta_2-\theta_1)\left\{\left(C_x - C_y\rho_{xy}\right)^2 - C_y^2\rho_{xy}^2\right\}\right.$$
$$\left.+\theta_1\left\{\left(C_z - C_y\rho_{yz}\right)^2 - C_y^2\rho_{yz}^2\right\}\right]. \quad (1.2.12)$$

Another chain ratio-type estimator suggested by Sahoo & Sahoo(1992) when information on auxiliary variable X is unavailable for population is

$$t_7 = \bar{y}_2\frac{\bar{x}_2}{\bar{x}_1}\frac{\bar{z}_1}{\bar{Z}} \quad (1.2.13)$$

and its mean square error is

$$\mathrm{MSE}\left(t_7\right)=\bar{Y}^2\left[\theta_2 C_y^2+\left(\theta_2-\theta_1\right)\left\{\left(C_x+\rho_{xy} C_y\right)^2-C_y^2 \rho_{xy}^2\right\}\right.$$

$$\left.+\theta_1\left\{\left(C_z+\rho_{yz} C_z\right)^2-\rho_{yz}^2 C_z^2\right\}\right]. \qquad (1.2.14)$$

The ratio-to-regression estimator suggested by Kiregyera(1980) is:

$$t_8=\frac{\bar{y}_2}{\bar{x}_2}\left[\bar{x}_1+b_{xz}\left(\bar{Z}-\bar{z}_1\right)\right], \qquad (1.2.15)$$

where b_{xz} is calculated from the first phase sample. The mean square error of t_8 can be written as:

$$\mathrm{MSE}\left(t_8\right)=\bar{Y}^2\left[\theta_2 C_y^2+\left(\theta_2-\theta_1\right)\left\{\left(C_x-\rho_{xy} C_y\right)^2-\rho_{xy}^2 C_y^2\right\}\right.$$

$$\left.+\theta_1\left\{\left(C_z\rho_{xz}-C_y\rho_{yz}\right)-C_y^2\rho_{yz}^2\right\}\right]. \qquad (1.2.16)$$

The ratio-in-regression estimator developed by Kiregyera(1984) is

$$t_9=\bar{y}_2+b_{yx}\left(\frac{\bar{x}_1}{\bar{z}_1}\bar{Z}-\bar{x}_2\right), \qquad (1.2.17)$$

where b_{yx} computed from second phase sample. The mean square error of (1.2.17) can be written as:

$$\mathrm{MSE}\left(t_9\right)=\bar{Y}^2 C_y^2\left[\theta_2-\left(\theta_2-\theta_1\right)\rho_{xy}^2+\theta_1\left\{\left(\rho_{xy}\frac{C_z}{C_x}-\rho_{yz}\right)^2-\rho_{yz}^2\right\}\right]. \qquad (1.2.18)$$

Another regression-in-regression estimator suggested by Kiregyera(1984) is:

$$t_{10}=\bar{y}_2+b_{yx}\left\{\left(\bar{x}_1-\bar{x}_2\right)-b_{xz}\left(\bar{z}_1-\bar{Z}\right)\right\}, \qquad (1.2.19)$$

where b_{yx} is based on second phase sample while b_{xz} is based on first phase sample and mean square error of t_{10} is:

$$\mathrm{MSE}\left(t_{10}\right)=\bar{Y}^2 C_y^2\left[\theta_2-\left(\theta_2-\theta_1\right)\rho_{xy}^2-\theta_1\rho_{yz}^2+\theta_1\left(\rho_{yz}-\rho_{xy}\rho_{xz}\right)^2\right]. \qquad (1.2.20)$$

Following the construction of regression-in-regression estimator by Kiregyera(1984), Mukerjee, Rao, & Vijayan(1987) developed the estimator when information on both auxiliary variables is unavailable . The regression estimator using two auxiliary variables is:

$$t_{11} = \bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right) + b_{yz}\left(\bar{z}_1 - \bar{z}_2\right), \tag{1.2.21}$$

where b_{yx} and b_{yz} are based on second phase sample. The mean square error of (1.2.21) can be written as:

$$MSE\left(t_{11}\right) = \bar{Y}^2 C_y^2\left[\theta_2 - \left(\theta_2 - \theta_1\right)\left(\rho_{xy}^2 + \rho_{yz}^2 - 2\rho_{xy}\rho_{yz}\rho_{xz}\right)\right]. \tag{1.2.22}$$

Mukerjee, et al.(1987) also proposed following estimator when population information on auxiliary variable Z is available:

$$t_{12} = \bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right) + b_{yz}\left(\bar{Z} - \bar{z}_2\right), \tag{1.2.23}$$

where b_{yx} and b_{yz} are based on second phase sample. The mean square error t_{12}:

$$MSE\left(t_{12}\right) = \bar{Y}^2 C_y^2\left[\theta_2 - \theta_1\rho_{yz}^2 - \left(\theta_2 - \theta_1\right)\left(\rho_{xy}^2 + \rho_{yz}^2 - 2\rho_{xy}\rho_{yz}\rho_{xz}\right)\right]. \tag{1.2.24}$$

Mukerjee, et al.(1987) developed the third estimator when information on auxiliary variable Z is available for population as:

$$t_{13} = \bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right) + b_{yx}b_{xz}\left(\bar{Z} - \bar{z}_1\right) + b_{yz}\left(\bar{Z} - \bar{z}_2\right), \tag{1.2.25}$$

where b_{yx} and b_{yz} are based on second phase sample while b_{xz} is based on first phase sample. The mean square error for this estimator can be written as:

$$MSE\left(t_{13}\right) = \bar{Y}^2 C_y^2\left[\theta_1\left(\rho_{yz} - \rho_{xy}\rho_{yz}\right)^2 + \theta_2\left(1 - \rho_{yz}^2 - \rho_{xy}^2 + 2\rho_{xy}\rho_{yz}\rho_{xz}\right)\right] \tag{1.2.26}$$

J. Sahoo, Sahoo, & Mohanty(1993) suggested the following estimator

$$t_{14} = \bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right) + b_{yz}\left(\bar{Z} - \bar{z}_1\right), \tag{1.2.27}$$

where b_{yx} and b_{yz} are based on second phase sample. The mean square error for this estimator can be written as:

$$MSE\left(t_{14}\right) = \bar{Y}^2 C_y^2\left[\theta_2\left(1 - \rho_{xy}^2\right) + \theta_1\left(\rho_{xy}^2 - \rho_{yz}^2\right)\right]. \tag{1.2.28}$$

J. Sahoo & Sahoo(1994) proposed three regression type estimators using information of two auxiliary variables. The first estimator proposed by J. Sahoo & Sahoo(1994) when information on auxiliary variable “z” is available for population is

$$t_{15} = \bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right) + b_{yz}\left(\bar{Z} - \bar{z}_2\right), \tag{1.2.29}$$

where b_{yx} and b_{yz} are based on second phase sample. The mean square error for this estimator is:

$$MSE\left(t_{15}\right)=\bar{Y}^2C_y^2\left[\theta_2\left(1-\rho_{yz}^2\right)-\left(\theta_2-\theta_1\right)\left(\rho_{xy}^2-2\rho_{xy}\rho_{yz}\rho_{xz}\right)\right]. \qquad (1.2.30)$$

The second estimator proposed by J. Sahoo & Sahoo(1994) when information on auxiliary variable "z" is available for population is:

$$t_{16}=\bar{y}_2+b_{yx}\left(\bar{x}_1-\bar{x}_2\right)-b_{yx}b_{xz}\left(\bar{z}_1-\bar{z}_2\right)+b_{yx}b_{xz}\left(\bar{Z}-\bar{z}_1\right), \qquad (1.2.31)$$

where b_{yx} and b_{yz} are based on second phase sample while b_{xz} is based on the first phase sample. The mean square error for this estimator is:

$$MSE\left(t_{16}\right)=\bar{Y}^2C_y^2\left[\theta_2+\theta_1\rho_{xy}^2\rho_{xz}^2-\left(\theta_2-\theta_1\right)\left\{\rho_{xy}^2\left(1-\rho_{xz}^2\right)-2\rho_{xy}\rho_{yz}\rho_{xz}\right\}\right]. \qquad (1.2.32)$$

Third estimator proposed by J. Sahoo & Sahoo(1994) when information on auxiliary Z is available for population, is:

$$t_{17}=\bar{y}_2+b_{yx}\left(\bar{x}_1-\bar{x}_2\right)+b_{yx}b_{xz}\left(\bar{Z}-\bar{z}_2\right)+b_{yx}b_{xz}\left(\bar{Z}-\bar{z}_1\right), \qquad (1.2.33)$$

where b_{yx} and b_{yz} are based on the second phase sample while b_{xz} is based on the first phase sample. The mean square error for this estimator is:

$$MSE\left(t_{17}\right)=\bar{Y}^2C_y^2\left[\theta_2-\left(\theta_2-\theta_1\right)\left(\rho_{xy}^2+\rho_{xy}^2\rho_{xz}^2-2\rho_{xy}\rho_{yz}\rho_{xz}\right)\right]. \qquad (1.2.34)$$

S. K. Srivastava(1970) suggested the following general ratio estimator using single auxiliary variable as:

$$t_{18}=\bar{y}_2\left(\frac{\bar{x}_1}{\bar{x}_2}\right)^{\alpha}, \qquad (1.2.35)$$

where α is unknown constant and the value of α for which the mean square error of t_{18} is minimum is $\alpha=\frac{C_y}{C_x}\rho_{xy}$ and the mean square error of t_{18} is

$$MSE\left(t_{18}\right)=\bar{Y}^2C_y^2\left[\theta_2-\left(\theta_2-\theta_1\right)\rho_{yx}^2\right]. \qquad (1.2.36)$$

Another general ratio estimator suggested by S. R. Srivastava, Khare, & Srivastava(1990) using information of two auxiliary variables is:

$$t_{19} = \bar{y}_2 \left(\frac{\bar{x}_1}{\bar{x}_2} \right)^{\alpha_1} \left(\frac{\bar{Z}}{\bar{z}_1} \right)^{\alpha_2}. \tag{1.2.37}$$

The values of α_1 and α_2 for which the mean square error of t_{19} is minimum are $\alpha_1 = \frac{C_y}{C_x}\rho_{xy}$ and $\alpha_2 = \frac{C_y}{C_z}\rho_{yz}$ respectively. The mean square error of t_{19} can be written as:

$$MSE(t_{19}) = \bar{Y}^2 C_y^2 \left[\theta_2 - (\theta_2 - \theta_1)\rho_{xy}^2 - \theta_1 \rho_{yz}^2 \right]. \tag{1.2.38}$$

Roy(2003) suggested the following general regression estimator for two phase sampling when information on Z is available:

$$t_{20} = \bar{y}_2 + k_1 \left[\left\{ \bar{x}_1 + k_2 \left(\bar{Z} - \bar{z}_1 \right) \right\} - \left\{ \bar{x}_2 + k_3 \left(\bar{Z} - \bar{z}_2 \right) \right\} \right]. \tag{1.2.39}$$

The optimum values of unknown constants are

$$k_1 = \frac{\bar{Y}\, C_y}{\bar{X}\, C_x} \frac{\left(\rho_{yx} - \rho_{xz}\rho_{yz} \right)}{\left(1 - \rho_{xz}^2 \right)},\ k_2 = \frac{\bar{X}\, C_x}{\bar{Z} C_z}\rho_{xz} \text{ and } k_3 = -\frac{\bar{X}\, C_x \left(\rho_{yz} - \rho_{xz}\rho_{yz} \right)}{\bar{Z} C_z \left(\rho_{xy} - \rho_{yz}\rho_{xz} \right)},$$

and the expression for mean square error is:

$$MSE(t_{20}) = \bar{Y}^2 C_y^2 \left[\theta_1 \left(1 - \rho_{y.xz}^2 \right) + \theta_2 \left(1 - \rho_{zy}^2 \right) \rho_{yx.z}^2 \right]. \tag{1.2.40}$$

H. P. Singh, Upadhyaya, & Chandra(2004) proposed following generalized estimator when information on auxiliary variable Z is available:

$$t_{21} = \bar{y}_2 \left(\frac{\bar{x}_1}{\bar{x}_2} \right)^{\alpha_1} \left(\frac{a\bar{Z} + b}{a\bar{z}_1 + b} \right)^{\alpha_2} \left(\frac{a\bar{Z} + b}{a\bar{z}_2 + b} \right)^{\alpha_3}. \tag{1.2.41}$$

The optimum values of unknown constants are

$$\alpha_1 = \frac{C_y}{C_x} \left(\frac{\rho_{yx} - \rho_{yz}\rho_{xz}}{1 - \rho^2{}_{xz}} \right),\ \alpha_2 = \frac{1}{\phi} \frac{C_y}{C_z} \left(\frac{\rho_{xz}\left(\rho_{yx} - \rho_{yz}\rho_{xz} \right)}{1 - \rho^2{}_{xz}} \right)$$

$$\text{and } \alpha_3 = \frac{1}{\phi} \frac{C_y}{C_z} \frac{\left(\rho_{yz} - \rho_{yx}\rho_{xz} \right)}{1 - \rho^2{}_{xz}},\ \text{where } \phi = \left(\frac{a\bar{Z}}{a\bar{Z} + b} \right)$$

Mean square error of t_{21} is:

$$MSE(t_{21}) = \bar{Y}^2 C_y^2 \left[\theta_2 - \theta_1 \rho_{yz}^2 - (\theta_2 - \theta_1) \rho_{y.xz}^2 \right]. \quad (1.2.42)$$

Further, H. P. Singh, et al.(2004) investigated that for different values of α_1, α_2, α_3, a and b, the mean per unit estimator, the usual two-phase sampling ratio estimator, the usual two-phase sampling product estimator, S. K. Srivastava(1971) estimator, Chand(1975) ratio-type estimator, S. R. Srivastava, et al.(1990) estimator, G. N. Singh & Upadhyaya(1995) estimator, Upadhyay and Singh (2001) estimators are special cases of their estimator.

Samiuddin & Hanif(2007) has proposed different estimators by considering following situation in two phase sampling:

a) In addition to the sample, the population means of both auxiliary variables are known. They called it the "*Full Information Case*".

b) In addition to the sample, $\bar{X}$ is given only, ($\bar{Z}$ being unknown). They called it the "*Partial Information Case*".

c) When $\bar{X}$ and $\bar{Z}$ are unknown, they called it the "*No Information Case*".

The regression estimator suggested by Samiuddin & Hanif(2007) for *Full information Case* is:

$$t_{22} = \bar{y}_2 + \alpha_1 (\bar{X} - \bar{x}_2) + \alpha_2 (\bar{Z} - \bar{z}_2). \quad (1.2.43)$$

The optimum values of unknown constants are

$$\alpha_1 = \frac{\bar{Y} C_y (\rho_{xy} - \rho_{yz}\rho_{xz})}{\bar{X} C_x (1 - \rho_{xz}^2)} \text{ and } \alpha_2 = \frac{\bar{Y} C_y (\rho_{yz} - \rho_{xy}\rho_{xz})}{\bar{Z} C_z (1 - \rho_{xz}^2)}$$

and mean square error of t_{22} is

$$MSE(t_{22}) = \theta_2 \bar{Y}^2 C_y^2 \left[1 - \rho_{y.xz}^2 \right], \quad (1.2.44)$$

where $\rho_{y.xz}^2$ is the partial correlation coefficient of "y" and combined effects of "x" and "z"

The following regression estimator has been suggested by Samiuddin & Hanif(2007) for *Partial Information Case*.

$$t_{23} = \bar{y}_2 + \alpha_1(\bar{x}_1 - \bar{x}_2) + \alpha_2(\bar{z}_1 - \bar{z}_2) + \alpha_3(\bar{Z} - \bar{z}_2). \quad (1.2.45)$$

The optimum values for unknown constants are

$$\alpha_1 = \frac{\bar{Y}C_y(\rho_{xy} - \rho_{yz}\rho_{xz})}{\bar{X}C_x(1-\rho_{xz}^2)}, \quad \alpha_2 = -\frac{\bar{Y}C_y\rho_{xz}(\rho_{xy} - \rho_{yz}\rho_{xz})}{\bar{Z}C_z(1-\rho_{xz}^2)}$$

and $\alpha_3 = \frac{\bar{Y}C_y}{\bar{Z}C_z}\rho_{yz}$.

The mean square error of (1.2.45) is:

$$MSE(t_{23}) = \bar{Y}^2C_y^2\left[\theta_2(1-\rho_{y.xz}^2) + \theta_1(1-\rho_{yz}^2)\rho_{yx.z}^2\right]. \quad (1.2.46)$$

Samiuddin & Hanif(2007) proposed following regression estimator for *No Information Case*

$$t_{24} = \bar{y}_2 + \alpha_1(\bar{x}_1 - \bar{x}_2) + \alpha_2(\bar{z}_1 - \bar{z}_2) \quad (1.2.47)$$

The optimum values of unknown constants are

$$\alpha_1 = \frac{\bar{Y}C_y(\rho_{xy} - \rho_{xz}\rho_{yz})}{\bar{X}C_x(1-\rho_{xz}^2)} \text{ and } \alpha_2 = \frac{\bar{Y}C_y(\rho_{yz} - \rho_{xz}\rho_{xy})}{\bar{Z}C_z(1-\rho_{xz}^2)}$$

The minimum mean square error is:

$$MSE(t_{24}) = \bar{Y}^2C_y^2\left[\theta_2(1-\rho_{y.xz}^2) + \theta_1\rho_{y.xz}^2\right]. \quad (1.2.48)$$

The ratio estimator suggested by Samiuddin & Hanif(2007) for *Full information Case* is:

$$t_{25} = \bar{y}_2\left(\frac{\bar{X}}{\bar{x}_2}\right)^{\alpha_1}\left(\frac{\bar{Z}}{\bar{z}_2}\right)^{\alpha_2}. \quad (1.2.49)$$

The optimum values of unknown constants are

$$\alpha_1 = \frac{C_y(\rho_{xy} - \rho_{xz}\rho_{yz})}{C_x(1-\rho_{xz}^2)} \text{ and } \alpha_2 = \frac{C_y(\rho_{yz} - \rho_{yx}\rho_{xz})}{C_z(1-\rho_{xz}^2)}$$

The mean square error of t_{25} is:

$$MSE(t_{25}) = \theta_2\bar{Y}^2C_y^2(1-\rho_{y.xz}^2). \quad (1.2.50)$$

The ratio estimator suggested by Samiuddin & Hanif(2007) for *Partial information Case* is:

$$t_{26} = \bar{y}_2 \left(\frac{\bar{x}_1}{\bar{x}_2}\right)^{\alpha_1} \left(\frac{\bar{z}_1}{\bar{z}_2}\right)^{\alpha_2} \left(\frac{\bar{Z}}{\bar{z}_2}\right)^{\alpha_3}, \quad (1.2.51)$$

The optimum values of unknown constants are

$$\alpha_1 = \frac{C_y\left(\rho_{xy} - \rho_{yz}\rho_{xz}\right)}{C_x\left(1-\rho_{xz}^2\right)}, \; \alpha_2 = -\frac{C_y\rho_{xz}\left(\rho_{xy} - \rho_{yz}\rho_{xz}\right)}{C_z\left(1-\rho_{xz}^2\right)}$$

and $\alpha_3 = \frac{C_y}{C_z}\rho_{yz}$

The mean square error of t_{26} is:

$$MSE\left(t_{26}\right) = \bar{Y}^2 C_y^2\left[\theta_2\left(1-\rho_{y.xz}^2\right) + \theta_1\left(1-\rho_{yz}^2\right)\rho_{yx.z}^2\right]. \quad (1.2.52)$$

Ratio estimator proposed by Samiuddin & Hanif(2007) for *No information Case* is:

$$t_{27} = \bar{y}_2 \left(\frac{\bar{x}_1}{\bar{x}_2}\right)^{\alpha_1} \left(\frac{\bar{z}_1}{\bar{z}_2}\right)^{\alpha_2}. \quad (1.2.53)$$

The optimum values of unknown constants are

$$\alpha_1 = \frac{C_y\left(\rho_{xy} - \rho_{xz}\rho_{yz}\right)}{C_x\left(1-\rho_{xz}^2\right)} \text{ and } \alpha_2 = \frac{C_y\left(\rho_{yz} - \rho_{xy}\rho_{xz}\right)}{C_z\left(1-\rho_{xz}^2\right)}$$

mean square error of (1.2.53) is:

$$MSE\left(t_{27}\right) = \bar{Y}^2 C_y^2\left[\theta_2\left(1-\rho_{y.xz}^2\right) + \theta_1\rho_{y.xz}^2\right]. \quad (1.2.54)$$

Z. Ahmed, Hanif, & Ahmad(2009) suggested three classes of regression-cum-ratio estimators for estimating population mean of variable of interest for two-phase sampling using multi-auxiliary variables for full, partial and no information cases.

The proposed estimator by Z. Ahmed, et al.(2009) are:

$$t_{28} = \left[\bar{y}_2 + \sum_{i=1}^{r} \alpha_i \left(\bar{X}_i - \bar{x}_{(2)i} \right) \right] \prod_{i=r+1}^{r+s=q} \left(\frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\gamma_i} \tag{1.2.55}$$

Regressions-Cum-Ratio Estimator for Partial Information Case is:

$$t_{29} = \left[\bar{y}_2 + \sum_{i=1}^{r} \alpha_i'' \left(\bar{x}_{(1)i} - \bar{x}_{(2)i} \right) \right] \prod_{i=r+1}^{r+s=q} \left(\frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i''} \prod_{i=r+1}^{r+s=q} \left(\frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\delta_i''} \tag{1.2.56}$$

Regressions-Cum-Ratio Estimator for No Information Case is:

$$t_{30} = \left[\bar{y}_2 + \sum_{i=1}^{r} \alpha_i' \left(\bar{x}_{(1)i} - \bar{x}_{(2)i} \right) \right] \prod_{i=r+1}^{r+s=q} \left(\frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i'} \tag{1.2.57}$$

The mean square errors of above estimators are:

$$MSE\left(t_{28}\right) = \theta_2 \bar{Y}^2 C_y^2 \left(1 - \rho^2_{y.\underset{\sim}{x}_q}\right) \tag{1.2.58}$$

$$MSE\left(t_{29}\right) = \bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.\underset{\sim}{x}_q}\right) + \theta_1 \left(\rho^2_{y.\underset{\sim}{x}_q} - \rho^2_{y.\underset{\sim}{x}_s} \right) \right] \tag{1.2.59}$$

$$MSE\left(t_{30}\right) = \bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.\underset{\sim}{x}_q}\right) + \theta_1 \rho^2_{y.\underset{\sim}{x}_q} \right] \tag{1.2.60}$$

1.3 Some Popular Univariate Estimators in Multiphase Sampling based on Qualitative Predictors

In this section some estimators in multiphase sampling have been discussed which used information on auxiliary attributes. The pioneering work in multiphase sampling based on auxiliary attributes has been the work of Naik & Gupta(1996).

The family of estimators for two-phase sampling for no information case by Jhajj, Sharma, & Grover(2006) under same regularity conditions is

$$T'_{1(2)} = g_{\omega}\left(\bar{y}_2, v_{1d}\right), \tag{1.3.1}$$

where $v_{1d} = \dfrac{p_{1(2)}}{p_{1(1)}}$, and $g_{\omega}\left(\bar{Y},1\right) = \bar{Y}$.

The followings are some functions (estimators) of (1.3.11).

i) $g_{\omega}\left(\bar{y}_2, v_{1d}\right) = \bar{y}_2\left(v_{1d}\right)^{\alpha}$,

ii) $g_{\omega}\left(\bar{y}_2, v_{1d}\right) = \bar{y}_2 + \alpha\left(v_{1d} - 1\right)$,

iii) $g_{\omega}\left(\bar{y}_2, v_{1d}\right) = \bar{y}_2 + e^{\alpha\left(v_{1d}-1\right)}$,

iv) $g_{\omega}\left(\bar{y}_2, v_{1d}\right) = \bar{y}_2 + e^{\alpha\left(v_{1d}-1\right)} v_{1d}^{\alpha}$,

where α is unknown constant. Many other functions (estimators) may be constructed.
The mean square error of each estimator to the terms of order $1/n$ of this family is,

$$MSE\left(T'_{1(2)}\right) \approx \left(\theta_2 - \theta_3 \rho^2_{Pb_1}\right) S^2_y . \tag{1.3.2}$$

Shabbir and Gupta (2007) proposed an estimator which utilize the attribute auxiliary information:

$$t'_{2(2)} = \left[W_1 \bar{y}_2 + W_2\left(p_{1(1)} - p_{1(2)}\right)\right] \frac{p_{1(1)}}{p_{1(2)}}, \quad \text{for } p_{1(2)} > 0 \tag{1.3.3}$$

where W_1 and W_2 are unknown constants.

The mean square error of (1.3.13) to the terms of order $1/n^2$ is,

$$MSE\left(t'_{2(2)}\right) \approx \frac{\left(\theta_2 - \theta_3\rho^2_{Pb_1}\right)S^2_y}{1+\left(\theta_2 - \theta_3\rho^2_{Pb_1}\right)C^2_y}. \tag{1.3.4}$$

Hanif, Haq, & Shahbaz(2009) proposed a generalized family of estimators based on the information of "k" auxiliary attributes and discussed the estimator for full, partial and no information cases. Hanif, Haq, et al.(2009) showed that the proposed family has smaller mean square error than given by Jhajj, et al.(2006). The proposed estimator for Partial Information Case is:

$$t'_{3(2)} = \bar{y}_2 + \sum_{j=1}^{m}\alpha_j\left(v_j - 1\right) + \sum_{j=m+1}^{k}\alpha_j\left(v_{jd} - 1\right) \tag{1.3.5}$$

The mean square error of (1.3.17) is:

$$MSE\left(t'_{3(2)}\right) = \theta_2 S_y^2 + \theta_1 a_1^{/} S_{\tau_1} a_1 + \left(\theta_2 - \theta_1\right) a_2^{/} S_{\tau_2} a_2 + 2\theta_1 a_1^{/} s_{y\tau_1} + 2\left(\theta_2 - \theta_1\right) a_2^{/} s_{y\tau_2} \tag{1.3.6}$$

The proposed estimator for No Information Case is:

$$t'_{4(2)} = \bar{y}_2 + \sum_{i=1}^{k}\alpha_j\left(v_{jd} - 1\right) \tag{1.3.7}$$

The mean square error of (1.3.19) is:

$$MSE\left(t'_{4(2)}\right) = \theta_2 S_y^2 + \left(\theta_2 - \theta_1\right) a^{/} S_\tau\, a \; - 2\left(\theta_2 - \theta_1\right) a^{/} s_{y\tau} \tag{1.3.8}$$

Hanif, Haq, & Shahbaz(2010) proposed some ratio estimators for single phase and two phase sampling by using information on multiple auxiliary attributes. The proposed estimators are generalization of the estimator proposed by Naik & Gupta(1996). Hanif, et al.(2010) also drive the shrinkage version of the proposed estimators by using the method given Shahbaz & Hanif(2009). The estimator for two phase sampling is:

$$t_{5(2)} = \bar{y}_2 \prod_{j=1}^{m}\left(\frac{P_j}{p_{j(1)}}\right)\prod_{h=m+1}^{k}\left(\frac{p_{h(1)}}{p_{h(2)}}\right) \tag{1.3.9}$$

The mean square error of (1.3.23) up to first order approximation is:

$$MSE\left(t'_{5(2)}\right) \approx \bar{Y}^2 \left[\theta_2 \left\{ C_y^2 + \sum_{j=m+1}^{k} C_{\tau_j}^2 - 2 \sum_{j=m+1}^{k} C_y C_{\tau_j} \rho P b_j + 2 \sum_{j \neq \psi = m+1}^{k} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} \right\} \right.$$

$$\theta_1 \left\{ \left(\sum_{j=1}^{m} C_{\tau_j}^2 - \sum_{j=m+1}^{k} C_{\tau_j}^2 \right) - 2 \left(\sum_{j=1}^{m} C_y C_{\tau_j} \rho P b_j - \sum_{j=m+1}^{k} C_y C_{\tau_j} \rho P b_j \right) \right. \quad (1.3.10)$$

$$\left. \left. +2 \left(\sum_{j \neq \psi = 1}^{m} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} - \sum_{j \neq \psi = m+1}^{k} C_{\tau_j} C_{\tau_\psi} Q_{j\psi} \right) \right\} \right]$$

1.4 Multivariate Estimators

Hanif, Ahmed, & Ahmad(2009) proposed a number of generalized multivariate ratio estimators for two-phase and multi-phase sampling in the presence of multi-auxiliary variables for estimating population mean for a single variable and a vector of variables of interest. Hanif, Ahmed, et al.(2009) proposed more general ratio estimator when information on all auxiliary variables are not available for population (No Information Situation), the estimator is:

$$T_{hk(1\times p)} = \left[\bar{y}_{(k)1} \prod_{i=1}^{q} \left(\frac{\bar{x}_{(h)i}}{\bar{x}_{(k)i}} \right)^{\alpha_{i1}} \bar{y}_{(k)2} \prod_{i=1}^{q} \left(\frac{\bar{x}_{(h)i}}{\bar{x}_{(k)i}} \right)^{\alpha_{i2}} \cdots \bar{y}_{(k)p} \prod_{i=1}^{q} \left(\frac{\bar{x}_{(h)i}}{\bar{x}_{(k)i}} \right)^{\alpha_{ip}} \right] \quad (1.4.1)$$

The variance-covariance matrix of the estimator is of the following form:

$$\Sigma_{T_{hk}(p\times p)} = \theta_k \Sigma_{y(p\times p)} - (\theta_k - \theta_h) \Sigma'_{y(p\times p)} \Sigma^{-1}_{x\ (q\times q)} \Sigma_{yx(q\times p)} \quad (1.4.2)$$

Where Σ_y is covariance matrix of $\mathbf{y}$.

Z. Ahmed, Hussin, & Hanif(2010) also following multivariate regression estimator by using information of multiple auxiliary variables:

$$T_{hk(1\times p)} = \left[\bar{y}_{(k)1}\ \bar{y}_{(k)2}\ \ldots\ \bar{y}_{(k)p} \right] + \left[\sum_{i=1}^{q} \alpha_{i1} \left(\bar{x}_{(h)i} - \bar{x}_{(k)i} \right) \sum_{i=1}^{q} \alpha_{i2} \left(\bar{x}_{(h)i} - \bar{x}_{(k)i} \right) \ldots \sum_{i=1}^{q} \alpha_{ip} \left(\bar{x}_{(h)i} - \bar{x}_{(k)i} \right) \right] \quad (1.4.3)$$

The variance-covariance matrix of the estimator is of the following form:

$$\Sigma_{T_{hk}(p\times p)} = \theta_k \Sigma_{y(p\times p)} - (\theta_k - \theta_h) \Sigma_{yx(p\times q)} \Sigma^{-1}_{x(q\times q)} \Sigma'_{x(q\times p)} \quad (1.4.4)$$

1.5 Introduction to Zellner Models

Seemingly unrelated regression equations (SURE) model, proposed by Zellner(1962), is a generalization of a linear regression model that consists of several regression equations, each having its own dependent variable and potentially different sets of independent variables. Each equation is a linear regression model in its own and can be estimated separately, that's why the system is called seemingly unrelated regression models Greene(2003).

The model can be estimated equation by equation using ordinary least squares (OLS) method. Such estimates are consistent, however generally not as efficient as estimators obtained by SUR method, which amounts to feasible generalized least squares with a specific structure of the variance-covariance matrix. Two situations when SUR is equivalent to OLS, are: either when the error terms are uncorrelated between the equations (truly unrelated), or when each model contains exactly the same set of predictors on the right-hand-side.

1.5.1 The SURE Model

Suppose there are k regression equations

$$\{y_{it} = x_{it}^{/} \beta_i + \varepsilon_{it}, \qquad i = 1, 2, \ldots, k.$$

Where i represents the equation number, and $t = 1, 2 \ldots, T$ is the observations index. The number of observations is assumed to be large enough, such that in the analysis we take $T \to \infty$, whereas the number of models k remains same.

Each i^{th} equation has a single dependent variable y_{it}, and a k_i-dimensional vector of predictors x_{it}. If we stack observations corresponding to the i^{th} equation into T-dimensional vectors and matrices, then the regression model can be written in vector form as:

$$\{y_i = x_i \ \beta_i + \varepsilon_i, \qquad i = 1, 2, \ldots, k,$$

where y_i and ε_i are $T\times1$ vectors, X_i is a $T\times k_i$ matrix, and β_i is a $k_i\times1$ vector.

Finally, if we stack these k vector equations on top of each other, the system will take form Zellner(1962)

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_k \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \qquad (1.5.1.1)$$

The model (1.5.1.1) can be collectively estimated by using Feasible Generalized Least Square(FGLS)

1.6 The Shrinkage Estimator

Shrinkage estimator is an estimator that, either implicitly or explicitly, incorporates the effects of shrinkage. In simple words this means that a raw estimate is improved by combining it with other information. One general result is that many standard estimators can be improved, in terms of mean squared error (MSE), by shrinking them towards zero. It is assumed that the expected value of raw estimate is not zero, and consider other estimators obtained by multiplying the raw estimate by some certain parameter. Value for this parameter may be specified by minimizing the Mean Square Error of the new estimate. For this value of the parameter, the new estimate will have a smaller Mean Square Error as compared to raw estimate. Thus it has been improved. The effect here may be to convert an unbiased raw estimate to an improved biased one. A good example can be considered in the case of estimation of the population variance based on a sample; for a sample size of n, the use of a divisor $\mathrm{n}-1$ in the usual formula gives an unbiased estimator while a divisor of $n+1$ gives one which has the minimum mean square error.

1.6.2 General Shrinkage Estimator Shahbaz & Hanif(2009)

Let a population parameter Θ can be estimated by using an estimator $\hat{\eta}^*$ whit mean square error $MSE\left(\hat{\eta}^*\right)$. Shahbaz & Hanif(2009) has defined a general shrinkage estimator as $\hat{\eta}_s^* = d\hat{\eta}^*$ where d is a constant to be determined such that mean square error of $\hat{\eta}_s^*$ is minimized.

$$MSE\left(\hat{\eta}_s^*\right) = \frac{MSE\left(\hat{\eta}^*\right)}{1+\Theta^{-2}MSE\left(\hat{\eta}^*\right)} \qquad (1.6.2.1)$$

The expression for mean square error given in (1.6.2.1) can be used to obtain the mean square error of shrinkage version of any estimator. The estimator proposed by Searl (1964) turned out to be special case of shrinkage estimator proposed by Shahbaz and Hanif (2009) by using $\hat{\eta}^* = \bar{y}$.

1.7 Study Objectives

The core objectives of this study are to:

- generalize the Roy (2003) estimator to multiple auxiliary variables
- generalize the Roy (2003) estimator to multiple auxiliary attributes
- propose multivariate version of Roy (2003) estimator for multiple auxiliary variables
- propose multivariate version of Roy (2003) estimator for multiple auxiliary attributes
- propose the Seemingly Unrelated Regression Estimators (SURE) in single phase, two phase and multiphase sampling
- conduct theoretical and empirical comparison of the proposed estimators with the existing estimators.

Chapter 2: Literature Review

Neyman(1938) was the first one who gave the concept of two-phase sampling as:

"A more accurate estimate of the original character may be obtained for the same total expenditure by arranging the sampling of population in two steps. The first step is to secure data, for the second character only, from a relatively large random sample of the population in order to obtain an accurate estimate of the distribution of this character.

The second step is to divide this sample, as in stratified sampling into classes or strata according to the value of the second character and to draw at random from each of the strata, a small sample for the costly intensive interviewing necessary to secure data regarding the first character.

An estimate of the first character based on these samples may be more accurate than based on an equally expensive sample drawn at random without stratification. The question is to determine for a given expenditure, the sizes of the initial sample and the subsequent samples which yield the most accurate estimate of the first character".

Cochran(1940) developed Ratio estimator for estimating population total by utilizing the auxiliary information and discussed the relative efficiency of the estimator. The ratio estimator is an efficient estimator of population total if there exist strong linear relationship between variable of interest and auxiliary variable. The regression estimator is always more efficient than the ratio estimator if population regression coefficient is used as a building block of the estimator. Both estimators are equally precise if the regression line passes through origin. Use of auxiliary variable are well studied in literature of survey sampling as discussed in the standard books on survey sampling by various authors including Hartley & Ross(1954) , Yates(1960), Kish(1965), Murthy(1967), Raj(1968), Cochran(1977) and P. V. Sukhatme, Sukhatme, Sukhatme, & Ashok(1984).

Hartley & Ross(1954) developed exact ratio estimator. Rao & Rao(1971) studied performance of the ratio estimator based on small samples. B. V. Sukhatme(1962) developed a general ratio-type estimator in two-phase sampling. Mohanty(1967) discussed that the precision in estimating the population mean may be increased by using another auxiliary variable which was correlated with variable of interest. Swain(2000) constructed chain regression estimator in which the auxiliary variable with known population mean was used to estimate the unknown population mean of another auxiliary variable say "x" then this estimated mean of "x" was used to estimate the population mean of study variable "y". Chand(1975) developed two chain ratio-type estimators by using the information of two auxiliary variables for estimating finite population mean. Kiregyera(1980) constructed a chain ratio-to-regression type estimator by using two auxiliary variables and discussed the relative efficiency with Chand(1975) chain ratio-type estimator.

S. K. Srivastava(1970) suggested a general family of ratio-type estimators for estimating mean of a finite population by using single auxiliary variable. Kiregyera(1984) developed two estimators, one is ratio-in-regression and other is regression-in-regression estimator; both use two auxiliary variables. The efficiency of estimators was investigated empirically as well as under super-population model, both constructed estimators performed better than regression estimator using one auxiliary variable for two-phase sampling. The regression-in-regression estimator performed better than ratio-in-regression estimator and their performance was better than Kiregyera(1980) estimator. Mukerjee, et al.(1987) developed three estimators following the method of Kiregyera(1984). Mukerjee, et al.(1987) also extended their results to the case when multi-auxiliary information was utilized.

H. P. Singh(1987) proposed a regression estimator for estimating population mean in two-phase sampling by using prior knowledge of correlation coefficient between variable of interest and auxiliary variable. H. P. Singh(1987) proposed his estimator and demonstrated that the proposed estimator is more efficient than usual regression estimator in two-phase sampling. Tripathi, Singh, & Upadhyaya(1988) provided a general framework for estimating a general function of parameters with the help of a general function of supplementary parameter, for bivariate population, variance of study variable was estimated through general results derived from estimating general function of parameters. An asymptotically optimum subclass of the wider class was also identified in it. H. P. Singh & Namjoshi(1988) suggested a class of multivariate regression estimators of population mean of study variable in two-phase sampling. H. P. Singh & Namjoshi(1988) provided exact expression of mean square error and optimum estimator of the proposed class. H. P. Singh, Tripathi, &

Upadhyaya(1989) proposed a general class of estimators for population mean and discussed that usual ratio, regression and product estimators in two-phase sampling may always be improved under moderate conditions. H. P. Singh, et al.(1989) also provided a general condition under which two-phase sampling estimators were preferable over usual unbiased estimator for single sample for a linear cost structure.

Tripathi & Khattree(1989) discussed the estimation of means of several variables of interest, using multi-auxiliary variables, under simple random sampling. Further Tripathi(1989) extends the results to the case of two occasions. Tripathi & Chaubey(1993) have considered the problem of obtaining optimum probabilities of selection, based on multi-auxiliary variables, in unequal probability sampling for estimating the finite population mean.

S. R. Srivastava, et al.(1990) developed a general family of chain ratio-type estimators for estimating population mean by using two auxiliary variables.

H. P. Singh, Upadhyaya, & Iachan(1990) proposed a class of estimators based on general sampling designs for population parameter utilizing auxiliary information of some other parameters. They also discussed the properties of the suggested class and find the asymptotic lower bound to the mean square error of the estimators belonging to the class. H. P. Singh, et al.(1990) also proposed several unbiased ratio and product estimators with their expressions of asymptotic variances using Jackknife technique in two-phase sampling.

H. P. Singh, Singh, & Kushwaha(1992) suggested a class of chain ratio-to-regression estimators in two-phase sampling for finite population mean of variable of interest. Optimum estimator was identified from this class. The performance of optimum estimator is investigated theoretically as well as empirically.

L. N. Upadhyaya, Dubey, & Singh(1992) suggested a class of ratio-in-regression estimators for population mean of the study variable using two auxiliary variables in two-phase sampling and investigated its asymptotic properties.

J. Sahoo & Sahoo(1993) gave a general frame work of estimation of population mean of variable of interest by using an additional auxiliary variable for two-phase sampling when the population mean of the main auxiliary variable was unknown. Chand(1975) and Kiregyera(1980, 1984) estimators can be seen as the special cases of J. Sahoo & Sahoo(1993) class of estimators.

H. P. Singh(1993) developed a class of chain ratio-cum-difference estimator for mean of a finite population using two auxiliary variables with asymptotic expressions for its bias and mean square error in two-phase sampling. H. P. Singh(1993) also theoretically and empirically proved that the constructed class of estimator was more efficient than Chand(1975) and S. R. Srivastava, et al.(1990) estimators.

J. Sahoo, et al.(1993) suggested a regression-type estimator based upon the information on second auxiliary variable when population mean of the main auxiliary variable was unknown. H. P. Singh & Biradar(1994) developed general class of unbiased ratio-type estimators in two phase sampling and derived expression of its asymptotic variance.

J. Sahoo & Sahoo(1994) discussed relative efficiency of four chain-type estimators in two-phase sampling under super-population model. J. Sahoo, Sahoo, & Mohanty(1994a) provided a regression approach for estimation using two auxiliary variables for two-phase sampling. J. Sahoo, Sahoo, & Mohanty(1994b) considered an alternative approach for estimating mean in two-phase sampling using two auxiliary variables. V. K. Singh & Singh(1994) proposed a class of estimators for estimating ratio and product of means of two finite populations in two-phase sampling. V. K. Singh & Singh(1994) obtained the asymptotic expression for bias and mean square error.

G. N. Singh & Upadhyaya(1995) developed a generalized estimator for estimating the population mean in two-phase sampling using two-auxiliary variables. H. P. Singh & Gangele(1995) suggested an estimator using information of coefficient of variation and information on two-auxiliary variables for population mean in two phase sampling. Their proposed estimator was efficient than Chand(1975), Chand(1975; Kiregyera(1980, 1984) and J. Sahoo & Sahoo(1993) estimators.

H. P. Singh, Katyar, & Gangwar(1996) discussed a class of almost unbiased regression type estimators in two-phase sampling by using Quenouille(1956) and Jack-Knife technique. Naik & Gupta(1996) proposed ratio, product and regression estimators for the population mean when auxiliary attribute information is available.

Hidiroglou & Särndal(1998) discussed that two-phase sampling is cost effective and precision of ratio and regression estimates under two-phase sampling increases if there exist high correlation between the auxiliary variable and variable under study.

M. S. Ahmed(1998) interpreted the regression coefficients correctly for the estimators suggested by Mukerjee, et al.(1987). M. S. Ahmed(1998) mentioned that the corrected mean square errors of Kiregyera(1984) estimators are computed assuming that the regression coefficient b_{yx} and b_{yz} are ordinary not partial regression coefficient. Furthermore he proved that Kiregyera(1984) estimators were better than Mukerjee, et al.(1987) estimators and also showed that estimator suggested by Tripathi & Ahmed(1995) was more efficient than Kiregyera(1984) estimators. H. P. Singh & Gangele(1999) suggested almost unbiased ratio-type and product-type estimators for population mean in two-phase sampling. The performance of suggested estimators was empirically evaluated. Tracy & Singh(1999) proposed a class of chain regression estimators with asymptotic expression of bias and mean squared error for estimating the population mean of variable of interest in two-phase sampling by using two-auxiliary variables.
Tracy & Singh(1999) also derived asymptotic optimum unbiased ratio-type estimator with its variance in two-phase sampling and also in successive sampling with the use of two auxiliary variables. Tracy & Singh(1999) proved that proposed estimator is better than Olkin(1958) and Sen(1971) estimators.

J. Sahoo & Sahoo(1999a) developed a class of estimators by using two phase sampling and J. Sahoo & Sahoo(1999b) conducted a comparative study of the estimators considered by Chand(1975), Kiregyera(1980, 1984), Mukerjee, et al.(1987), J. Sahoo, et al.(1993) and J. Sahoo, et al.(1994a) under the super population model using two auxiliary variables.

H. P. Singh & Tailor(2000) suggested some ratio-type estimators of population mean of study variable using two auxiliary variables in two-phase sampling with coefficient of variation of the second auxiliary variable was known. H. P. Singh & Tailor(2000) obtained the conditions in which proposed estimators were more efficient than usual two-phase sampling ratio-estimator, Chand(1975) estimator, and G. N. Singh & Upadhyaya(1995).

A. K. Singh, Singh, & Upadhyaya(2001) proposed two classes of chain ratio-type estimators and also derived expressions of bias and mean square errors in two-phase sampling by using two-auxiliary variables. L. N. Sahoo & Sahoo(2001) proposed estimators of finite population mean by using predictive approach in two-phase sampling using two auxiliary variables. A. K. Singh, et al.(2001) considered a generalized chain estimator for finite population mean using two auxiliary variables in two phase sampling.

Radhey, Singh, & Singh(2002) provided a modified ratio estimator with approximate expressions for its bias and mean square error in two-phase sampling for population mean of variable of interest by using two-auxiliary variables. Radhey, et al.(2002) investigated empirically that asymptotic optimum estimators performed better than conventional unbiased ratio, traditional ratio, Chand(1975), Kiregyera(1980) and L. Upadhyaya, Kushwaha, & Singh(1990) estimators. H. P. Singh & Singh(2002) estimated the population coefficient of variation of study variable with chain ratio-type estimator using two auxiliary variables in two-phase sampling and also derived expressions for the bias and mean squared error.

Chandra & Singh(2003) discussed a class of unbiased estimators with its properties for the population mean of study variable in two-phase sampling using two-auxiliary variables when information for the mean of main auxiliary variable was not available. The unbiased estimators suggested by Chand(1975) and Dalbehera & Sahoo(2000) found to be the special cases of proposed class. Diana & Tommasi(2003) proposed a general class of estimators for finite population mean in two-phase sampling. Diana & Tommasi(2003) class of estimators was based on the sample means and variances of two auxiliary variables. Diana & Tommasi(2003) also provide the minimum variance bound for any member of the class.

H. P. Singh & Espejo(2003) proposed a class of ratio–product estimators for estimating a finite population mean in two-phase sampling and identified an asymptotically optimum estimator in their class along with its approximate mean-square error by using the prior knowledge of the parameter $C = \rho C_y / C_x$. H. P. Singh & Espejo(2003) also found that the estimators are equally efficient for known value of C as well as for consistent estimator of C.

R. Singh & Singh(2003) proposed a regression-type estimator in two-phase sampling for population mean when information on second variable was known and variance of main auxiliary variable was not known. The proposed estimator was more efficient than Chand(1975), Kiregyera(1980, 1984) and usual ratio, regression estimators.

Roy(2003) constructed a regression-type estimator of population mean of the main variable in the presence of available information on second auxiliary variable, when the population mean of the first auxiliary variable was not known. Roy(2003) estimator was more efficient than Mohanty(1967), Chand(1975), Kiregyera(1980, 1984) and J. Sahoo, et al.(1993)

.

M. S. Ahmed(2003) proposed chain based general estimators for finite population mean using multivariate auxiliary information under multiphase sampling. M. S. Ahmed(2003) considered a number of auxiliary variables in each phase under a general sampling design and studied the properties of these estimators and presented the results for simple random sampling without replacement schemes. M. S. Ahmed(2003) also derived the optimum sample sizes using a modified cost.

H. P. Singh, et al.(2004) proposed a family of estimators, which is more efficient than those considered by S. K. Srivastava(1970), Chand(1975), S. R. Srivastava, et al.(1990), H. P. Singh & Biradar(1994), H. P. Singh & Gangele(1995) and A. K. Singh, et al.(2001). H. P.

Singh & Vishwakarma(2005-2006) suggested a modified version of Sahai(1979) estimator in two-phase sampling and discussed its properties. L. N. Upadhyaya, Singh, & Tailor(2006) proposed a family of chain ratio-type estimators for population mean by utilizing information of mean for first auxiliary variable and coefficient of variation for second auxiliary variable.

H. P. Singh, Singh, & Kim(2006) considered chain ratio and regression type estimators of median and provided expressions for its variance. The optimum sample sizes were also obtained for first phase and second phase using fixed cost of the survey. Comparison was made with estimators suggested by A. K. Singh & Singh(2001) . Jhajj, et al.(2006) has proposed a family of estimators in single and two phase sampling using information on a single auxiliary attributes, the proposed family is based upon a general function. Shabbir & Gupta(2007) have also proposed an estimator for population mean in single phase sampling using information of single auxiliary attribute.
H. P. Singh & Espejo(2007) suggested a class of ratio-product estimators in two-phase sampling for population mean in the presence of two-auxiliary variables and also discussed their properties. H. P. Singh & Espejo(2007) also identified asymptotically optimum estimators with their variances and compared their efficiency with two-phase ratio, product and mean per unit estimator under some conditions. Shabbir & Gupta(2007) have also proposed an estimator for population mean in single phase sampling using information of single auxiliary attribute.

Samiuddin & Hanif(2007) introduced ratio and regression estimation procedures for estimating population mean in two-phase sampling for different three situations depending upon the availability of information on two auxiliary variables for population. Samiuddin & Hanif(2007) considered three situations, first when information on both auxiliary variables was not available, second when information on one auxiliary variable was available and third, when information was available on both auxiliary variables. Samiuddin & Hanif(2007) estimators developed in second situation were found to be as efficient as H. P. Singh, et al.(2004) and Roy(2003). But the estimators developed in third situation were more efficient then H. P. Singh, et al.(2004) and Roy(2003) as well as their own estimators developed in first two situations.

Z. Ahmed, et al.(2009) proposed generalized regression-cum-ratio estimators for two-phase sampling using multi-auxiliary variables. Z. Ahmed, et al.(2009)suggested three classes of regression-cum-ratio estimators for estimating population mean of variable of interest for two-phase sampling based on multi-auxiliary variables for full information, partial information and no information cases. Hanif, Ahmed, et al.(2009) proposed a number of generalized multivariate ratio estimators for two-phase and multi-phase sampling in the utilizing multi-auxiliary variables for estimating population mean for a single variable and a vector of variables of interest(s). Hanif, Ahmed, et al.(2009) also made theoretical and empirical to check the efficiencies of the estimators.

Hanif, Haq, et al.(2009) proposed general family of estimators and derived general expression of mean square error of estimators proposed by Jhajj, et al.(2006). The family has been proposed for single-phase sampling in case of full information and for two-phase sampling in case of partial and no information cases. Hanif, Haq, et al.(2009) discussed that the proposed family has smaller mean square error than given by Jhajj, et al.(2006).

Z. Ahmed, et al.(2010) suggested a number of generalized multivariate regression estimators for two-phase and multi-phase sampling in the presence of multi-auxiliary variables for estimating population mean for a single variable and a vector of variables. Hanif, et al.(2010) proposed some ratio estimators for single phase and two phase sampling using information on multiple auxiliary attributes. The proposed estimators are generalization of the estimator proposed by Naik & Gupta(1996). Hanif, et al.(2010) proposed some ratio estimators for single phase and two phase sampling by using information on multiple auxiliary attributes. The proposed estimators are generalization of the estimator proposed by Naik & Gupta(1996). Hanif, et al.(2010) also drive the shrinkage version of the proposed estimators by using the method given Shahbaz & Hanif(2009).

Chapter 3: Univariate Estimators

3.1 Introduction

In this chapter some new univariate estimators for two phase sampling have been proposed. The proposed estimators are extension of the estimator proposed by Roy(2003). The proposed estimator use information on multiple variables as well as on multiple attributes. The empirical study has also been conducted to see the performance of proposed estimators.

3.2 New Estimator with Quantitative Predictors

Suppose we have a random sample of n_1 observations and information for a set of m+1 auxiliary variable(s) is recorded from that sample. Suppose further that a subsample of n_2 observation is drawn and information for auxiliary variables as well as variable of interest y is recorded from that subsample. Based upon the available information following estimator for population mean as under:

$$t_{N_1(2)} = \bar{y}_2 + k\left[\bar{x}_1 + \boldsymbol{\alpha}^{/}\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}_1\right) - \left\{\bar{x}_2 + \boldsymbol{\beta}^{/}\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}_2\right)\right\}\right] \qquad (3.2.1)$$

where $\bar{x}_1$ is mean of auxiliary variable at first phase and $\bar{x}_2$ is mean of same variable at second phase. The vector $\bar{\mathbf{w}}$ is an (m x 1) vector which contains means of m auxiliary variables for the population, $\bar{\mathbf{w}}_1$ is vector of means for same variables at first phase and $\bar{\mathbf{w}}_2$ is vector of means at the second phase. Let us write these quantities as:

$$\bar{\mathbf{w}}_1 = \bar{\mathbf{w}} + \bar{\mathbf{e}}_{w_1}; \quad \bar{\mathbf{w}}_2 = \bar{\mathbf{w}} + \bar{\mathbf{e}}_{w_2}$$

$$\bar{x}_1 = \bar{x} + \bar{e}_{\bar{x}_1};\ \bar{x}_2 = \bar{x} + \bar{e}_{\bar{x}_2}$$

$$\bar{y}_2 = \bar{Y} + \bar{e}_{\bar{y}2}$$

Using above representations, the estimator (3.2.1) can be put in the following form

$$t_{N_1(2)} - \bar{Y} = \bar{e}_{y_2} + k\left[\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - \boldsymbol{\alpha}^{/}\bar{\mathbf{e}}_{w_1} + \boldsymbol{\beta}^{/}\bar{\mathbf{e}}_{w_2}^{/}\right]$$

Squaring above equation:

$$\left(t_{N_1(2)}-\bar{Y}\right)^2=\bar{e}_{y_2}^2+k^2\left[\bar{e}_{x_1}-\bar{e}_{x_2}-\boldsymbol{\alpha}^/\bar{\mathbf{e}}_{w_1}+\boldsymbol{\beta}\bar{\mathbf{e}}_{w_2}\right]^2+2k\,\bar{e}_{y_2}\left[\bar{e}_{x_1}-\bar{e}_{x_2}-\boldsymbol{\alpha}^/\bar{\mathbf{e}}_{w_1}+\boldsymbol{\beta}\bar{\mathbf{e}}_{w_2}\right]$$

$$=\bar{e}_{y_2}^2+k^2\left[\bar{e}_{x_1}^2+\bar{e}_{x_2}^2+\boldsymbol{\alpha}^/\bar{\mathbf{e}}_{w_1}\bar{\mathbf{e}}_{w_2}\boldsymbol{\alpha}+\boldsymbol{\beta}^/\bar{\mathbf{e}}_{w_1}\bar{\mathbf{e}}_{w_2}\boldsymbol{\beta}-2\bar{e}_{x_1}\bar{e}_{x_2}-2\boldsymbol{\alpha}^/\bar{\mathbf{e}}_{w_1}\bar{e}_{x_1}\right.$$
$$\left.+2\boldsymbol{\beta}^/\bar{\mathbf{e}}_{w_2}\bar{e}_{x_1}+2\boldsymbol{\alpha}^/\bar{\mathbf{e}}_{w_1}\bar{e}_{x_2}-2\boldsymbol{\beta}^/\bar{\mathbf{e}}_{w_2}\bar{e}_{x_2}-2\boldsymbol{\alpha}^/\bar{\mathbf{e}}_{w_1}\bar{\mathbf{e}}_{w_2}^/\right]+2k\left[\bar{e}_{x_1}\bar{e}_{y_2}-\bar{e}_{y_2}\bar{e}_{x_2}-\boldsymbol{\alpha}^/\bar{\mathbf{e}}_{w_1}\bar{e}_{y_2}+\boldsymbol{\beta}^/\bar{\mathbf{e}}_{w_2}\bar{e}_{y_2}\right]$$

Applying the expectation, the mean square error of $t_{N_1(2)}$ is:

$$S=\theta_2S_y^2+k^2\left[\theta_1S_x^2+\theta_2S_x^2+\theta_1\boldsymbol{\alpha}^/\mathbf{S}_w\boldsymbol{\alpha}+\theta_2\boldsymbol{\beta}^/\mathbf{S}_w\boldsymbol{\beta}-2\theta_1S_x^2-2\theta_1\boldsymbol{\alpha}^/\mathbf{s}_{wx}+2\theta_1\boldsymbol{\beta}^/\mathbf{s}_{wx}\right.$$
$$\left.+2\theta_1\boldsymbol{\alpha}^/\mathbf{s}_{wx}-2\theta_2\boldsymbol{\beta}^/\mathbf{s}_{wx}-2\theta_1\boldsymbol{\alpha}^/\mathbf{S}_w\boldsymbol{\beta}\right]+2k\left[\theta_1S_{xy}-\theta_2S_{xy}-\theta_1\boldsymbol{\alpha}^/\mathbf{s}_{wy}+\theta_2\boldsymbol{\beta}^/\mathbf{s}_{wy}\right] \quad (3.2.2)$$

The optimum values of unknown quantities which minimize the S are obtained by differentiating (3.2.2) with respect to unknown quantities. The partial derivatives are:

$$\frac{\partial S}{\partial k}=\theta_2S_y^2+k^2\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\boldsymbol{\alpha}^/\mathbf{S}_w\boldsymbol{\alpha}^/+\theta_2\boldsymbol{\beta}^/\mathbf{S}_w\boldsymbol{\beta}+2\theta_1\boldsymbol{\beta}^/\mathbf{s}_{wx}-2\theta_2\boldsymbol{\beta}^/\mathbf{s}_{wx}-2\theta_1\boldsymbol{\alpha}^/\mathbf{S}_w\boldsymbol{\beta}\right]$$
$$+2k\left[\left(\theta_2-\theta_1\right)S_{xy}-\theta_1\boldsymbol{\alpha}^/\mathbf{s}_{wy}+\theta_2\boldsymbol{\beta}^/\mathbf{s}_{wy}\right]$$

Setting $\frac{\partial S}{\partial k}=0$:

$$=2k\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\boldsymbol{\alpha}^/\mathbf{S}_w\boldsymbol{\alpha}+\theta_2\boldsymbol{\beta}^/\mathbf{S}_w\boldsymbol{\beta}+2\theta_1\boldsymbol{\beta}^/\mathbf{s}_{wx}-2\theta_2\boldsymbol{\beta}^/\mathbf{s}_{wx}-2\theta_1\boldsymbol{\alpha}^/\mathbf{S}_w\boldsymbol{\beta}\right]$$
$$+2\left[\left(\theta_1-\theta_2\right)S_{xy}-\theta_1\boldsymbol{\alpha}^/\mathbf{s}_{wy}+\theta_2\boldsymbol{\beta}^/\mathbf{s}_{wy}\right]=0$$

(3.2.3)

Again

$$\frac{\partial S}{\partial\boldsymbol{\alpha}}=k^2\left[\theta_1^2\mathbf{S}_w\boldsymbol{\alpha}-2\theta_1\mathbf{S}_w\boldsymbol{\beta}\right]-2k\theta_1\mathbf{s}_{wy}$$

$$\frac{\partial S}{\partial\boldsymbol{\alpha}}=\mathbf{0}:$$

$$2\theta_1k^2\mathbf{S}_w\boldsymbol{\alpha}-2\theta_1k^2\mathbf{S}_w\boldsymbol{\beta}-2\theta_1k\mathbf{s}_{wy}=\mathbf{0}$$

or

$$k\mathbf{S}_w\boldsymbol{\alpha}-k\mathbf{S}_w\boldsymbol{\beta}-\mathbf{s}_{wy}=\mathbf{0} \quad (3.2.4)$$

Finally,

$$\frac{\partial S}{\partial\boldsymbol{\beta}}=k^2\left[2\theta_2\mathbf{S}_w\boldsymbol{\beta}+2\theta_1\mathbf{s}_{wx}-2\theta_2\mathbf{s}_{wx}-2\theta_1\mathbf{S}_w\boldsymbol{\alpha}\right]+2k\theta_2\mathbf{s}_{wy}$$

Setting $\frac{\partial S}{\partial \boldsymbol{\beta}} = \mathbf{0}$

$$2\theta_2 k^2 \mathbf{S}_w \boldsymbol{\beta} + 2\theta_1 k^2 \mathbf{s}_{wx} - 2\theta_2 k^2 \mathbf{s}_{wx} - 2\theta_1 k^2 \mathbf{S}_w \boldsymbol{\alpha} + 2k\ \theta_2 \mathbf{s}_{wy} = \mathbf{0}$$

$$2k\left[\theta_2 k^2 \mathbf{S}_w \boldsymbol{\beta} + \theta_1 k \mathbf{s}_{wx} - \theta_2 k \mathbf{s}_{wx} - \theta_1 k \mathbf{S}_w \boldsymbol{\alpha} + \theta_2 \mathbf{s}_{wy}\right] = \mathbf{0}$$

$$\theta_2 k \mathbf{S}_w \boldsymbol{\beta} - \theta_1 k \mathbf{S}_w \boldsymbol{\alpha} + k\left(\theta_1 - \theta_2\right)\mathbf{s}_{wx} + \theta_2 \mathbf{s}_{wy} = \mathbf{0}$$

$$k\left(\theta_2 \mathbf{S}_w \boldsymbol{\beta} - \theta_1 \mathbf{S}_w \boldsymbol{\alpha}\right) - k\left(\theta_2 - \theta_1\right)\mathbf{s}_{wx} + \theta_2 \mathbf{s}_{wy} = \mathbf{0} \tag{3.2.5}$$

Solving (3.2.4) and (3.2.5)

$$\boldsymbol{\alpha} = \mathbf{S}_w^{-1}\mathbf{s}_{wx} \tag{3.2.6}$$

Putting (3.2.6) in (3.2.5)

$$k\left(\theta_2 \mathbf{S}_w \boldsymbol{\beta} - \theta_1 \mathbf{S}_w \mathbf{S}_w^{-1}\mathbf{s}_{wx}\right) - k\left(\theta_2 - \theta_1\right)\mathbf{s}_{wx} + \theta_2 \mathbf{s}_{wy} = \mathbf{0}$$

$$\theta_2 k \mathbf{S}_w \boldsymbol{\beta} - \theta_1 k \mathbf{s}_{wx} - \theta_2 k \mathbf{s}_{wx} + \theta_1 k \mathbf{s}_{wx} + \theta_2 \mathbf{s}_{wy} = \mathbf{0}$$

$$k \mathbf{S}_w \boldsymbol{\beta} - k \mathbf{s}_{wx} = -\mathbf{s}_{wy}$$

$$k \mathbf{S}_w \boldsymbol{\beta} = k \mathbf{s}_{wx} - \mathbf{s}_{wy}$$

$$\mathbf{S}_w \boldsymbol{\beta} = \mathbf{s}_{wx} - \frac{1}{k}\mathbf{s}_{wy}$$

$$\boldsymbol{\beta} = \mathbf{S}_w^{-1}\mathbf{s}_{wx} - \frac{1}{k}\mathbf{S}_w^{-1}\mathbf{s}_{wy}; \boldsymbol{\beta}' = \mathbf{s}'_{wx}\mathbf{S}_w^{-1} - \frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1} \tag{3.2.7}$$

Using (3.2.6) and (3.2.7) in (3.2.3):

$$k\Bigg[\left(\theta_2 - \theta_1\right)S_x^2 + \theta_1 \mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{S}_w\mathbf{S}_w^{-1}\mathbf{s}_{wx} + \theta_2\left(\mathbf{s}_{wx}\mathbf{S}_w^{-1} - \frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\right)\mathbf{S}_w\left(\mathbf{S}_w^{-1}\mathbf{s}_{wx} - \frac{1}{k}\mathbf{S}_w^{-1} - \mathbf{s}_{wx}\right)$$
$$+2\theta_1\left(\mathbf{s}'_{wx}\mathbf{S}_w^{-1} - \frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\right)\mathbf{s}_{wx} - 2\theta_2\left(\mathbf{s}'_{wx}\mathbf{S}_w^{-1} - \frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\right)\mathbf{s}_{wx} - 2\theta_1\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{S}_w\left(\mathbf{S}_w^{-1}\mathbf{s}_{wx} - \frac{1}{k}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right)\Bigg]$$
$$+\left[\left(\theta_1 - \theta_2\right)S_{xy} - \theta_1\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy} + \theta_2\left(\mathbf{s}'_{wx}\mathbf{S}_w^{-1} - \frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\right)\mathbf{s}_{wy}\right] = 0$$

$$k\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}+\theta_2\left(\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}-\frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wx}-\frac{1}{k}\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}+\frac{1}{k^2}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right)\right.$$
$$\left.+2\theta_1\left(\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}-\frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wx}\right)-2\theta_2\left(\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}-\frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wx}\right)-2\theta_1\left(\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}-\frac{1}{k}\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right)\right]$$
$$+\left[\left(\theta_2-\theta_1\right)S_{xy}-\theta_1\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}+\theta_2\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}-\theta_2\frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right]=0$$

$$k\left[\left(\theta_1-\theta_2\right)S_x^2+\theta_1\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}+\theta_2\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}-\theta_2\frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wx}-\theta_2\frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wx}\right.$$
$$\left.+\theta_2\frac{1}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}-2\theta_2\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}+\frac{2\theta_2}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wx}\right]+\left[\left(\theta_1-\theta_2\right)S_{xy}+\left(\theta_2-\theta_1\right)\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}-\frac{\theta_2}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right]=0$$

$$k\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}-\theta_2\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}+\frac{\theta_2}{k^2}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right]$$
$$+\left[\left(\theta_1-\theta_2\right)S_{xy}+\left(\theta_2-\theta_1\right)\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}-\frac{\theta_2}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right]=0$$

$$k\left[\left(\theta_2-\theta_1\right)S_x^2-\left(\theta_2-\theta_1\right)\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}\right]-\left(\theta_2-\theta_1\right)S_{xy}+\left(\theta_2-\theta_1\right)\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}=0$$

$$k=\frac{S_{xy}-\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}}{S_x^2-\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}}=\beta_{yx.\mathbf{w}}=\frac{S_{xy.\mathbf{w}}}{S_{xx.\mathbf{w}}} \tag{3.2.8}$$

Using (3.2.6), (3.2.7) and (3.2.8) in (3.2.2), the mean square error of $t_{N_1(2)}$ is:

$$S=\theta_2S_y^2+k^2\left[\left(\theta_2-\theta_1\right)S_x^2-\left(\theta_2-\theta_1\right)\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}+\frac{\theta_2}{k^2}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right]$$
$$+2k\left[\left(\theta_1-\theta_2\right)S_{xy}+\left(\theta_2-\theta_1\right)\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}-\frac{\theta_2}{k}\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right]$$
$$=\theta_2S_y^2+\left(\theta_2-\theta_1\right)k^2S_x^2-\left(\theta_2-\theta_1\right)k^2\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}+\theta_2\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}-2k\ \left(\theta_2-\theta_1\right)S_{xy}$$
$$+2k\ \left(\theta_2-\theta_1\right)\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}-2\theta_2\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}$$
$$=\theta_2S_y^2-\theta_2\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}+\left(\theta_2-\theta_1\right)k^2\left(S_x^2-\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}\right)-2k\ \left(\theta_2-\theta_1\right)\left(S_{xy}-\mathbf{s}_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right)$$
$$=\theta_2S_y^2-\theta_2\mathbf{s}'_{wy}\mathbf{S}_w^{-1}\mathbf{s}_{wy}+\left(\theta_2-\theta_1\right)\frac{\left(S_{xy}-\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right)^2}{S_x^2-\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}}-2\left(\theta_2-\theta_1\right)\frac{\left(S_{xy}-\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wy}\right)^2}{S_x^2-\mathbf{s}'_{wx}\mathbf{S}_w^{-1}\mathbf{s}_{wx}}$$

$$= \theta_2 S_y^2 - \theta_2 \mathbf{s}_{wy}^{/} \mathbf{S}_w^{-1} \mathbf{s}_{wy} - (\theta_2 - \theta_1) \frac{\left(S_{xy} - \mathbf{s}_{wx}^{/} \mathbf{S}_w^{-1} \mathbf{s}_{wy}\right)^2}{S_x^2 - \mathbf{s}_{wx} \mathbf{S}_w^{-1} \mathbf{s}_{wx}}$$

$$= \theta_2 S_y^2 - \theta_2 \mathbf{s}_{wy}^{/} \mathbf{S}_w^{-1} \mathbf{s}_{wy} - \theta_2 \frac{\left(S_{xy} - \mathbf{s}_{wx}^{/} \mathbf{S}_w^{-1} \mathbf{s}_{wy}\right)^2}{S_x^2 - \mathbf{s}_{wx}^{/} \mathbf{S}_w^{-1} \mathbf{s}_{wx}} + \theta_1 \frac{\left(S_{xy} - \mathbf{s}_{wx}^{/} \mathbf{S}_w^{-1} \mathbf{s}_{wy}\right)^2}{S_x^2 \left(1 - \frac{\mathbf{s}_{wx}^{/} \mathbf{S}_w^{-1} \mathbf{s}_{wx}}{S_x^2}\right)}$$

$$= \theta_2 S_y^2 - \theta_2 \mathbf{s}_{wy}^{/} \mathbf{S}_w^{-1} \mathbf{s}_{wy} - \theta_2 \frac{S_{xy.\mathbf{w}}^2}{S_{x.\mathbf{w}}^2} + \theta_1 \frac{S_{xy.\mathbf{w}}^2}{S_{x.\mathbf{w}}^2}$$

$$= \theta_2 S_{y.\mathbf{w}}^2 - \theta_2 \frac{S_{xy.\mathbf{w}}^2}{S_{x.\mathbf{w}}^2} + \theta_1 \frac{S_{xy.\mathbf{w}}^2}{S_{x.\mathbf{w}}^2}$$

$$= S_{y.\mathbf{w}}^2 \left[\theta_2 - \theta_2 \rho_{xy.\mathbf{w}}^2 + \theta_1 \rho_{xy.\mathbf{w}}^2\right]$$

$$= S_{y.\mathbf{w}}^2 \left[\theta_2 \left(1 - \rho_{xy.\mathbf{w}}^2\right) + \theta_1 \rho_{xy.\mathbf{w}}^2\right] \tag{3.2.9}$$

Using $S_{y.\mathbf{w}}^2 = S_y^2 \left(1 - \rho_{y.\mathbf{w}}^2\right)$ in (3.2.9) and simplifying, the mean square error of t_1 may be written as

$$S = S_y^2 \left[\theta_2 \left(1 - \rho_{y.x\mathbf{w}}^2\right) + \theta_1 \rho_{xy.\mathbf{w}}^2 \left(1 - \rho_{y.\mathbf{w}}^2\right)\right] \tag{3.2.10}$$

The mean square error given in (3.2.10) is natural extension of mean square error of estimator proposed by Roy(2003).

3.2.1 Comparison of New Estimator with Classical Regression Estimator

In this section the proposed estimator has been compared with classical regression estimator based on several auxiliary variables.

Let t_0 be Classical Regression Estimator with mean square error given as:

$$MSE(t_0) = S_y^2 \left[\theta_2 \left(1 - \rho_{y.x\mathbf{w}}^2\right) + \theta_1 \rho_{y.x\mathbf{w}}^2\right] \tag{3.2.1.1}$$

The mean square error of proposed estimator is given in (3.2.10). Using $\left(1 - \rho_{y.x\mathbf{w}}^2\right) = \left(1 - \rho_{y.\mathbf{w}}^2\right)\left(1 - \rho_{yx.\mathbf{w}}^2\right)$ in (3.2.10) the mean square error of $t_{N_1(2)}$ can be written as:

$$MSE\left(t_{N_1(2)}\right) = S_y^2 \left[\theta_2 \left(1 - \rho_{y.x\mathbf{w}}^2\right) + \theta_1 \left(\rho_{y.x\mathbf{w}}^2 - \rho_{y.\mathbf{w}}^2\right)\right] \tag{3.2.1.2}$$

Comparing (3.2.1.1) with (3.2.1.2)

$$MSE(t_0) - MSE\left(t_{N_1(2)}\right) = S_y^2\left[\theta_2\left(1-\rho_{y.x\mathbf{w}}^2\right) + \theta_1\rho_{y.x\mathbf{w}}^2\right] - S_y^2\left[\theta_2\left(1-\rho_{y.x\mathbf{w}}^2\right) + \theta_1\left(\rho_{y.x\mathbf{w}}^2 - \rho_{y.\mathbf{w}}^2\right)\right]$$

$$= S_y^2\left[\theta_2\left(1-\rho_{y.x\mathbf{w}}^2\right) - \theta_2\left(1-\rho_{y.x\mathbf{w}}^2\right) + \theta_1\rho_{y.x\mathbf{w}}^2 - \theta_1\left(\rho_{y.x\mathbf{w}}^2 - \rho_{y.\mathbf{w}}^2\right)\right]$$

$$= \theta_1 S_y^2 \rho_{y.\mathbf{w}}^2 > 0 \qquad (3.2.1.3)$$

The equation (3.2.1.3) shows that the proposed estimator is always more precise as compared with the classical regression estimator.

3.2.2 Shrinkage version of the proposed Estimator

In this section the shrinkage version of the proposed estimator has been given following the method of Shahbaz & Hanif(2009). Using (3.2.10) in (1.6.2.5) the MSE of the proposed estimator can be given as:

$$MSE\left(t_{N_1(2)}\right) = \frac{S_y^2\left[\theta_2\left(1-\rho_{y.x\mathbf{w}}^2\right) + \theta_1\rho_{xy.\mathbf{w}}^2\left(1-\rho_{y.\mathbf{w}}^2\right)\right]}{1+\Theta^{-2}S_y^2\left[\theta_2\left(1-\rho_{y.x\mathbf{w}}^2\right) + \theta_1\rho_{xy.\mathbf{w}}^2\left(1-\rho_{y.\mathbf{w}}^2\right)\right]}$$

3.3 New Estimator with Qualitative Predictors

In this section a new estimator has been proposed using information of the multiple auxiliary attributes. Suppose a random sample of n_1 observations is drawn from a population of "N" units and information for a set of m+1 binary auxiliary variable(s) is recorded from that sample. Suppose further that a subsample of n_2 observation is drawn and information for binary auxiliary variables as well as variable of interest y is recorded from that subsample. Based upon the available information following estimator of population mean has been proposed:

$$t_{N_2(2)} = \overline{y}_2 + k\left[p_{\tau_1} + \boldsymbol{\gamma}'\left(\mathbf{p}_\delta - \mathbf{p}_{\delta_1}\right) - \left\{p_{\tau_2} + \boldsymbol{\eta}'\left(\mathbf{p}_\delta - \mathbf{p}_{\delta_2}\right)\right\}\right] \tag{3.3.1}$$

where p_{τ_1} is proportion of auxiliary attribute τ at first phase and p_{τ_2} is proportion of same variable at second phase. The vector $\mathbf{p}_\delta$ is an (m x 1) vector which contains proportions of m auxiliary attributes for the population, $\mathbf{p}_{\delta_1}$ is vector of proportion for same attributes at first phase and $\mathbf{p}_{\delta_2}$ is vector of proportions at the second phase. Proceeding as in previous section, let us write these quantities as:

$$\mathbf{p}_{\delta_1} = \mathbf{p}_\delta + \overline{\mathbf{e}}_{\delta_1}$$

$$\mathbf{p}_{\delta_2} = \mathbf{p}_\delta + \overline{\mathbf{e}}_{\delta_2}$$

$$p_{\tau_1} = p_\tau + \overline{e}_{\tau_1}; p_{\tau_2} = p_\tau + \overline{e}_{\tau_2}$$

Using the above representations, the estimator (3.3.1) can be put in the following form

$$t_{N_2(2)} - \overline{Y} = \overline{e}_{y_2} + k\left[\left(\overline{e}_{\tau_1} - \overline{e}_{\tau_2}\right) - \boldsymbol{\gamma}'\overline{\mathbf{e}}_{\delta_1} + \boldsymbol{\eta}'\overline{\mathbf{e}}'_{\delta_2}\right]$$

Squaring above equation:

$$\left(t_{N_2(2)} - \overline{Y}\right)^2 = \overline{e}_{y_2}^2 + k^2\left[\overline{e}_{\tau_1} - \overline{e}_{\tau_2} - \boldsymbol{\gamma}'\overline{\mathbf{e}}_{\delta_1} + \boldsymbol{\eta}\overline{\mathbf{e}}_{\delta_2}\right]^2 + 2k\overline{e}_{y_2}\left[\overline{e}_{\tau_1} - \overline{e}_{\tau_2} - \boldsymbol{\gamma}'\overline{\mathbf{e}}_{\delta_1} + \boldsymbol{\eta}\overline{\mathbf{e}}_{\delta_2}\right]$$

$$= \overline{e}_{y_2}^2 + k^2\left[\overline{e}_{\tau_1}^2 + \overline{e}_{\tau_2}^2 + \boldsymbol{\gamma}'\overline{\mathbf{e}}_{\delta_1}\overline{\mathbf{e}}_{\delta_2}\boldsymbol{\gamma} + \boldsymbol{\eta}'\overline{\mathbf{e}}_{\delta_1}\overline{\mathbf{e}}_{\delta_2}\boldsymbol{\eta} - 2\overline{e}_{\tau_1}\overline{e}_{\tau_2} - 2\boldsymbol{\gamma}'\overline{\mathbf{e}}_{\delta_1}\overline{e}_{\tau_1}\right.$$

$$\left.+2\boldsymbol{\eta}'\overline{\mathbf{e}}_{\delta_2}\overline{e}_{\tau_1} + 2\boldsymbol{\gamma}'\overline{\mathbf{e}}_{\delta_1}\overline{e}_{\tau_2} - 2\boldsymbol{\eta}'\overline{\mathbf{e}}_{\delta_2}\overline{e}_{\tau_2} - 2\boldsymbol{\gamma}'\overline{\mathbf{e}}_{\delta_1}\overline{\mathbf{e}}'_{\delta_2}\right] + 2k\left[\overline{e}_{\tau_1}\overline{e}_{y_2} - \overline{e}_{y_2}\overline{e}_{\tau_2} - \boldsymbol{\gamma}'\overline{\mathbf{e}}_{\delta_1}\overline{e}_{y_2} + \boldsymbol{\eta}'\overline{\mathbf{e}}_{\delta_2}\overline{e}_{y_2}\right]$$

Applying the expectation, the mean square error of $t_{N_2(2)}$ is:

$$S=\theta_2 S_y^2+k_1^2\left[\theta_1 S_\tau^2+\theta_2 S_\tau^2+\theta_1\boldsymbol{\gamma}'\mathbf{S}_\delta\boldsymbol{\gamma}+\theta_2\boldsymbol{\eta}'\mathbf{S}_\delta\boldsymbol{\eta}-2\theta_1 S_\tau^2-2\theta_1\boldsymbol{\gamma}'\mathbf{s}_{\delta\tau}+2\theta_1\boldsymbol{\eta}'\mathbf{s}_{\delta\tau}\right.$$
$$\left.+2\theta_1\boldsymbol{\gamma}'\mathbf{s}_{\delta\tau}-2\theta_2\boldsymbol{\eta}'\mathbf{s}_{\delta\tau}-2\theta_1\boldsymbol{\gamma}'\mathbf{S}_\delta\boldsymbol{\eta}\right]+2k\left[\theta_1 S_{\tau y}-\theta_2 S_{\tau y}-\theta_1\boldsymbol{\gamma}'\mathbf{s}_{\delta y}+\theta_2\boldsymbol{\eta}'\mathbf{s}_{\delta y}\right] \quad (3.3.2)$$

The optimum values of unknown quantities which minimizes the S are obtained by differentiating (3.3.2). The partial derivatives of (3.3.2) with respect to unknown quantities are:

$$\frac{\partial S}{\partial k}=\theta_2 S_y^2+k^2\left[\left(\theta_2-\theta_1\right)S_\tau^2+\theta_1\boldsymbol{\gamma}'\mathbf{S}_\delta\boldsymbol{\gamma}+\theta_2\boldsymbol{\eta}'\mathbf{S}_\delta\boldsymbol{\eta}+2\theta_1\boldsymbol{\eta}'\mathbf{s}_{\delta\tau}-2\theta_2\boldsymbol{\eta}'\mathbf{s}_{\delta\tau}-2\theta_1\boldsymbol{\gamma}'\mathbf{S}_\delta\boldsymbol{\eta}\right]$$
$$+2k\left[\left(\theta_2-\theta_1\right)S_{\tau y}-\theta_1\boldsymbol{\gamma}'\mathbf{s}_{\delta y}+\theta_2\boldsymbol{\eta}'\mathbf{s}_{\delta y}\right]$$

$$\frac{\partial S}{\partial k}=0:$$

$$=2k\left[\left(\theta_2-\theta_1\right)S_\tau^2+\theta_1\boldsymbol{\gamma}'\mathbf{S}_\delta\boldsymbol{\gamma}+\theta_2\boldsymbol{\eta}'\mathbf{S}_\delta\boldsymbol{\eta}+2\theta_1\boldsymbol{\eta}'\mathbf{s}_{\delta\tau}-2\theta_2\boldsymbol{\eta}'\mathbf{s}_{\delta\tau}-2\theta_1\boldsymbol{\gamma}'\mathbf{S}_\delta\boldsymbol{\eta}\right]$$
$$+2\left[\left(\theta_1-\theta_2\right)S_{\tau y}-\theta_1\boldsymbol{\gamma}'\mathbf{s}_{\delta y}+\theta_2\boldsymbol{\eta}'\mathbf{s}_{\delta y}\right]=0 \quad (3.3.3)$$

Again

$$\frac{\partial S}{\partial \boldsymbol{\gamma}}=k^2\left[\theta_1^2\mathbf{S}_\delta\boldsymbol{\gamma}-2\theta_1\mathbf{S}_\delta\boldsymbol{\eta}\right]-2k\theta_1\mathbf{s}_{\delta y}$$

$$\frac{\partial S}{\partial \boldsymbol{\gamma}}=\mathbf{0}:$$

$$2\theta_1 k^2\mathbf{S}_\delta\boldsymbol{\gamma}-2\theta_1 k^2\mathbf{S}_\delta\boldsymbol{\eta}-2\theta_1 k\mathbf{s}_{\delta y}=\mathbf{0}$$

or

$$k\mathbf{S}_\delta\boldsymbol{\gamma}-k\mathbf{S}_\delta\boldsymbol{\eta}-\mathbf{s}_{\delta y}=\mathbf{0} \quad (3.3.4)$$

Finally,

$$\frac{\partial S}{\partial \boldsymbol{\eta}}=k^2\left[2\theta_2\mathbf{S}_\delta\boldsymbol{\eta}+2\theta_1\mathbf{s}_{\delta\tau}-2\theta_2\mathbf{s}_{\delta\tau}-2\theta_1\mathbf{S}_\delta\boldsymbol{\gamma}\right]+2k\theta_2\mathbf{s}_{\delta y}$$

$$\frac{\partial S}{\partial \boldsymbol{\eta}}=\mathbf{0}:$$

$$2\theta_2 k^2\mathbf{S}_\delta\boldsymbol{\eta}+2\theta_1 k^2\mathbf{s}_{\delta\tau}-2\theta_2 k^2\mathbf{s}_{\delta\tau}-2\theta_1 k^2\mathbf{S}_\delta\boldsymbol{\gamma}+2k\ \theta_2\mathbf{s}_{\delta y}=\mathbf{0}$$

$$2k\left[\theta_2 k^2\mathbf{S}_\delta\boldsymbol{\eta}+\theta_1 k\mathbf{s}_{\delta\tau}-\theta_2 k\mathbf{s}_{\delta\tau}-\theta_1 k\mathbf{S}_\delta\boldsymbol{\gamma}+\theta_2\mathbf{s}_{\delta y}\right]=\mathbf{0}$$

$$\theta_2 k\mathbf{S}_\delta\boldsymbol{\eta}-\theta_1 k\mathbf{S}_\delta\boldsymbol{\gamma}+k\left(\theta_1-\theta_2\right)\mathbf{s}_{\delta\tau}+\theta_2\mathbf{s}_{\delta y}=\mathbf{0}$$

$$k\left(\theta_2\mathbf{S}_\delta\boldsymbol{\eta}-\theta_1\mathbf{S}_\delta\boldsymbol{\gamma}\right)-k\left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau}+\theta_2\mathbf{s}_{\delta y}=\mathbf{0} \quad (3.3.5)$$

Solving (3.3.4) and (3.3.5)

$$\boldsymbol{\gamma} = \mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} . \tag{3.3.6}$$

Put (3.3.6) in (3.3.5)

$$k\left(\theta_2\mathbf{S}_{\delta}\boldsymbol{\eta} - \theta_1\mathbf{S}_{\delta}\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau}\right) - k\left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau} + \theta_2\mathbf{s}_{\delta y} = \mathbf{0}$$

$$\theta_2 k\mathbf{S}_{\delta}\boldsymbol{\eta} - \theta_1 k\mathbf{s}_{\delta\tau} - \theta_2 k\mathbf{s}_{\delta\tau} + \theta_1 k\mathbf{s}_{\delta\tau} + \theta_2\mathbf{s}_{\delta y} = \mathbf{0}$$

$$k\mathbf{S}_{\delta}\boldsymbol{\eta} - k\mathbf{s}_{\delta\tau} = -\mathbf{s}_{\delta y}$$

$$k\mathbf{S}_{\delta}\boldsymbol{\eta} = k\mathbf{s}_{\delta\tau} - \mathbf{s}_{\delta y}$$

$$\mathbf{S}_{\delta}\boldsymbol{\eta} = \mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{s}_{\delta y}$$

$$\boldsymbol{\eta} = \mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y} ; \boldsymbol{\eta}' = \mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1} - \frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1} \tag{3.3.7}$$

Using (3.2.6) and (3.3.7) in (3.3.3):

$$k\Bigg[\left(\theta_2-\theta_1\right)S_{\tau}^2 + \theta_1\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{S}_{\delta}\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} + \theta_2\left(\mathbf{s}_{\delta x}\mathbf{S}_{\delta}^{-1} - \frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\right)\mathbf{S}_{\delta}\left(\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{S}_{\delta}^{-1} - \mathbf{s}_{\delta\tau}\right)$$
$$+2\theta_1\left(\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1} - \frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\right)\mathbf{s}_{\delta\tau} - 2\theta_2\left(\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1} - \frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\right)\mathbf{s}_{\delta\tau} - 2\theta_1\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{S}_{\delta}\left(\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y}\right)\Bigg]$$
$$+\left[\left(\theta_1-\theta_2\right)s_{\tau y} - \theta_1\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y} + \theta_2\left(\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1} - \frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\right)\mathbf{s}_{\delta y}\right] = 0$$

$$k\Bigg[\left(\theta_2-\theta_1\right)S_{\tau}^2 + \theta_1\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} + \theta_2\left(\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y} + \frac{1}{k^2}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y}\right)$$
$$+2\theta_1\left(\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau}\right) - 2\theta_2\left(\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau}\right) - 2\theta_1\left(\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \frac{1}{k}\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y}\right)\Bigg]$$
$$+\left[\left(\theta_2-\theta_1\right)S_{\tau y} - \theta_1\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y} + \theta_2\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y} - \theta_2\frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y}\right] = 0$$

$$k\Bigg[\left(\theta_1-\theta_2\right)S_{\tau}^2 + \theta_1\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} + \theta_2\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \theta_2\frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} - \theta_2\frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau}$$
$$+\theta_2\frac{1}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y} - 2\theta_2\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau} + \frac{2\theta_2}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta\tau}\Bigg] + \left[\left(\theta_1-\theta_2\right)S_{\tau y} + \left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y} - \frac{\theta_2}{k}\mathbf{s}_{\delta y}'\mathbf{S}_{\delta}^{-1}\mathbf{s}_{\delta y}\right] = 0$$

$$k\left[\left(\theta_2-\theta_1\right)S_\tau^2+\theta_1\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}-\theta_2\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}+\frac{\theta_2}{k^2}\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right]$$

$$+\left[\left(\theta_1-\theta_2\right)S_{\tau y}+\left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}-\frac{\theta_2}{k}\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right]=0$$

$$k\left[\left(\theta_2-\theta_1\right)S_\tau^2-\left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}\right]-\left(\theta_2-\theta_1\right)S_{\tau y}+\left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}=0$$

$$k=\frac{S_{\tau y}-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}}{S_\tau^2-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}}=\beta_{y\tau.\mathbf{w}}=\frac{S_{\tau y.\boldsymbol{\delta}}}{S_{\tau\tau.\boldsymbol{\delta}}} \tag{3.3.8}$$

Using the values of (3.3.6), (3.3.7) and (3.3.8) in (3.3.2), the mean square error of $t_{N_2(2)}$ is:

$$S=\theta_2S_y^2+k^2\left[\left(\theta_2-\theta_1\right)S_\tau^2-\left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}+\frac{\theta_2}{k^2}\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right]$$

$$+2k\left[\left(\theta_1-\theta_2\right)S_{\tau y}+\left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}-\frac{\theta_2}{k}\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right]$$

$$=\theta_2S_y^2+\left(\theta_2-\theta_1\right)k^2S_\tau^2-\left(\theta_2-\theta_1\right)k_\tau^2\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}+\theta_2\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}-2k\left(\theta_2-\theta_1\right)S_{\tau y}$$

$$+2k\left(\theta_2-\theta_1\right)\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}-2\theta_2\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}$$

$$=\theta_2S_y^2-\theta_2\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}+\left(\theta_2-\theta_1\right)k^2\left(S_\tau^2-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}\right)-2k\left(\theta_2-\theta_1\right)\left(S_{\tau y}-\mathbf{s}_{\delta\tau}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right)$$

$$=\theta_2S_y^2-\theta_2\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}+\left(\theta_2-\theta_1\right)\frac{\left(S_{\tau y}-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right)^2}{S_\tau^2-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}}-2\left(\theta_2-\theta_1\right)\frac{\left(S_{\tau y}-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right)^2}{S_\tau^2-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}}$$

$$=\theta_2S_y^2-\theta_2\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}-\left(\theta_2-\theta_1\right)\frac{\left(S_{\tau y}-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right)^2}{S_\tau^2-\mathbf{s}_{\delta\tau}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}}$$

$$=\theta_2S_y^2-\theta_2\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}-\theta_2\frac{\left(S_{\tau y}-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right)^2}{S_\tau^2-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}}+\theta_1\frac{\left(S_{\tau y}-\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}\right)^2}{S_\tau^2\left(1-\frac{\mathbf{s}_{\delta\tau}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta\tau}}{s_\tau^2}\right)}$$

$$=\theta_2S_y^2-\theta_2\mathbf{s}_{\delta y}^{/}\mathbf{S}_\delta^{-1}\mathbf{s}_{\delta y}-\theta_2\frac{S_{\tau y.\boldsymbol{\delta}}^2}{S_{\tau.\boldsymbol{\delta}}^2}+\theta_1\frac{S_{\tau y.\boldsymbol{\delta}}^2}{S_{\tau.\boldsymbol{\delta}}^2}$$

$$=\theta_2S_{y.\boldsymbol{\delta}}^2-\theta_2\frac{S_{\tau y.\boldsymbol{\delta}}^2}{S_{\tau.\boldsymbol{\delta}}^2}+\theta_1\frac{S_{\tau y.\boldsymbol{\delta}}^2}{S_{\tau.\boldsymbol{\delta}}^2}$$

$$=S_{y.\boldsymbol{\delta}}^2\left[\theta_2-\theta_2\rho_{\tau y.\boldsymbol{\delta}}^2+\theta_1\rho_{\tau y.\boldsymbol{\delta}}^2\right]$$

$$=S_{y.\boldsymbol{\delta}}^2\left[\theta_2\left(1-\rho_{\tau y.\boldsymbol{\delta}}^2\right)+\theta_1\rho_{xy.\boldsymbol{\delta}}^2\right] \tag{3.3.9}$$

Using $S_{y.\boldsymbol{\delta}}^2 = S_y^2\left(1-\rho_{y.\boldsymbol{\delta}}^2\right)$ in (3.3.9)

$$S = S_y^2\left[\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\rho_{\tau y.\boldsymbol{\delta}}^2\left(1-\rho_{y.\boldsymbol{\delta}}^2\right)\right] \qquad (3.3.10)$$

3.3.1 Comparison of New Estimator with Classical Regression Estimator

In this section the proposed estimator has been compared with classical regression estimator based on several auxiliary variables.

Let t_0 be Classical Regression Estimator with mean square error given as:

$$MSE\left(t_0\right) = S_y^2\left[\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\rho_{y.\tau\boldsymbol{\delta}}^2\right] \qquad (3.3.11)$$

The mean square error of proposed estimator given in (3.3.10). Using $\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)=\left(1-\rho_{y.\boldsymbol{\delta}}^2\right)\left(1-\rho_{y\tau.\boldsymbol{\delta}}^2\right)$ in (3.3.10) the mean square error of $t_{N_2(2)}$ can be written as:

$$MSE\left(t_{N_2(2)}\right) = S_y^2\left[\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\left(\rho_{y.\tau\boldsymbol{\delta}}^2-\rho_{y.\boldsymbol{\delta}}^2\right)\right] \qquad (3.3.12)$$

$$\begin{aligned} MSE\left(t_0\right)-MSE\left(t_{N_2(2)}\right) &= S_y^2\left[\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\rho_{y.\tau\boldsymbol{\delta}}^2\right]-S_y^2\left[\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\left(\rho_{y.\tau\boldsymbol{\delta}}^2-\rho_{y.\boldsymbol{\delta}}^2\right)\right] \\ &= S_y^2\left[\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)-\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\rho_{y.\tau\boldsymbol{\delta}}^2-\theta_1\left(\rho_{y.\tau\boldsymbol{\delta}}^2-\rho_{y.\boldsymbol{\delta}}^2\right)\right] \\ &= \theta_1 S_y^2\rho_{y.\boldsymbol{\delta}}^2 > 0 \qquad (3.3.12) \end{aligned}$$

The equation (3.3.12) shows that the proposed estimator is always more precise as compared with the classical regression estimator.

3.3.2 Shrinkage version of the proposed estimator

In this section the shrinkage version of the proposed estimator has been given following the method of Shahbaz & Hanif(2009). Using (3.3.10) in (1.6.2.5) the MSE of the proposed estimator can be given as:

$$MSE\left(t_{N_2(2)}\right) = \frac{S_y^2\left[\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\rho_{\tau y.\boldsymbol{\delta}}^2\left(1-\rho_{y.\boldsymbol{\delta}}^2\right)\right]}{1+\Theta^{-2}S_y^2\left[\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\rho_{\tau y.\boldsymbol{\delta}}^2\left(1-\rho_{y.\boldsymbol{\delta}}^2\right)\right]}$$

3.4 Numerical Study Quantitative Predictors

In this section empirical study has been carried out by using natural population available in literature (Annex-A). R code (Annex-B) is used to compute the mean Square error of proposed and classical regression estimators alongside and relative efficiencies are compared for various values of θ_1 and θ_2. The results of empirical study clearly shows that for all values of θ_1 and θ_2, the proposed estimator is more precise as compared with the regression estimator. It can also be seen that for fixed θ_1 the efficiency decreases with the increase in θ_2. Also for fixed θ_2 the efficiency increases with increase in θ_1.

Table 1: Mean Square Errors and relative efficiency of proposed estimator for Population -1

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
	0.2	0.45							
	0.3	0.67	0.68						
	0.4	0.89	0.90	0.92					
θ_2	**0.5**	1.11	1.12	1.13	1.15				
	0.6	1.33	1.34	1.35	1.37	1.38			
	0.7	1.55	1.56	1.57	1.59	1.60	1.61		
	0.8	1.76	1.78	1.79	1.81	1.82	1.83	1.85	
	0.9	1.98	2.00	2.01	2.02	2.04	2.05	2.06	2.08
		MSE of Classical Regression Estimator							
	0.2	0.58							
	0.3	0.80	0.95						
	0.4	1.02	1.17	1.32					
θ_2	**0.5**	1.24	1.39	1.53	1.68				
	0.6	1.46	1.61	1.75	1.90	2.05			
	0.7	1.68	1.83	1.97	2.12	2.26	2.41		
	0.8	1.90	2.04	2.19	2.34	2.48	2.63	2.78	
	0.9	2.12	2.26	2.41	2.56	2.70	2.85	3.00	3.14
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
	0.2	0.77							
	0.3	0.83	0.72						
	0.4	0.87	0.77	0.70					
θ_2	**0.5**	0.89	0.81	0.74	0.68				
	0.6	0.91	0.83	0.77	0.72	0.67			
	0.7	0.92	0.85	0.80	0.75	0.71	0.67		
	0.8	0.93	0.87	0.82	0.77	0.73	0.70	0.66	
	0.9	0.94	0.88	0.83	0.79	0.75	0.72	0.69	0.66

Table 2: Mean Square Errors and relative efficiency of proposed estimator for Population-2

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	266.6							
	0.3	367.2	432.6						
	0.4	467.7	533.2	598.6					
	0.5	568.3	633.7	699.2	764.6				
	0.6	668.9	734.3	799.7	865.2	930.6			
	0.7	769.4	834.9	900.3	965.7	1031.2	1096.6		
	0.8	870.0	935.4	1000.9	1066.3	1131.7	1197.2	1262.6	
	0.9	970.6	1036.0	1101.5	1166.9	1232.3	1297.8	1363.2	1428.6
		MSE of Classical Regression Estimator							
θ_2	**0.2**	266.9							
	0.3	367.5	433.2						
	0.4	468.0	533.8	599.5					
	0.5	568.6	634.4	700.1	765.9				
	0.6	669.2	734.9	800.7	866.4	932.2			
	0.7	769.8	835.5	901.3	967.0	1032.8	1098.5		
	0.8	870.3	936.1	1001.8	1067.6	1133.3	1199.1	1264.8	
	0.9	970.9	1036.7	1102.4	1168.2	1233.9	1299.7	1365.4	1431.2
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.9988							
	0.3	0.9991	0.9985						
	0.4	0.9993	0.9988	0.9984					
	0.5	0.9994	0.9990	0.9986	0.9983				
	0.6	0.9995	0.9991	0.9988	0.9985	0.9983			
	0.7	0.9996	0.9992	0.9989	0.9987	0.9985	0.9983		
	0.8	0.9996	0.9993	0.9990	0.9988	0.9986	0.9984	0.9982	
	0.9	0.9997	0.9994	0.9991	0.9989	0.9987	0.9985	0.9984	0.9982

Table 3: Mean Square Errors and relative efficiency of proposed estimator for Population-3

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	396.2							
	0.3	465.7	722.8						
	0.4	535.3	792.3	1049.3					
	0.5	604.9	861.9	1118.9	1375.9				
	0.6	674.5	931.5	1188.5	1445.5	1702.5			
	0.7	744.1	1001.1	1258.1	1515.1	1772.1	2029.1		
	0.8	813.7	1070.7	1327.7	1584.7	1841.7	2098.7	2355.7	
	0.9	883.2	1140.2	1397.2	1654.3	1911.3	2168.3	2425.3	2682.3
		MSE of Classical Regression Estimator							
θ_2	**0.2**	408.7							
	0.3	478.3	747.9						
	0.4	547.9	817.5	1087.1					
	0.5	617.5	887.1	1156.6	1426.2				
	0.6	687.1	956.6	1226.2	1495.8	1765.4			
	0.7	756.7	1026.2	1295.8	1565.4	1835.0	2104.5		
	0.8	826.2	1095.8	1365.4	1635.0	1904.5	2174.1	2443.7	
	0.9	895.8	1165.4	1435.0	1704.6	1974.1	2243.7	2513.3	2782.9
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.969							
	0.3	0.974	0.966						
	0.4	0.977	0.969	0.965					
	0.5	0.980	0.972	0.967	0.965				
	0.6	0.982	0.974	0.969	0.966	0.964			
	0.7	0.983	0.975	0.971	0.968	0.966	0.964		
	0.8	0.985	0.977	0.972	0.969	0.967	0.965	0.964	
	0.9	0.986	0.978	0.974	0.970	0.968	0.966	0.965	0.964

Table 4: Mean Square Errors and relative efficiency of proposed estimator for Population-4

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	0.0017							
	0.3	0.0022	0.0028						
	0.4	0.0027	0.0033	0.0039					
	0.5	0.0033	0.0038	0.0044	0.0050				
	0.6	0.0038	0.0044	0.0050	0.0055	0.0061			
	0.7	0.0043	0.0049	0.0055	0.0061	0.0066	0.0072		
	0.8	0.0049	0.0054	0.0060	0.0066	0.0072	0.0078	0.0083	
	0.9	0.0054	0.0060	0.0066	0.0071	0.0077	0.0083	0.0089	0.0095
		MSE of Classical Regression Estimator							
θ_2	**0.2**	0.0024							
	0.3	0.0030	0.0043						
	0.4	0.0035	0.0049	0.0062					
	0.5	0.0040	0.0054	0.0068	0.0081				
	0.6	0.0046	0.0059	0.0073	0.0087	0.0100			
	0.7	0.0051	0.0065	0.0078	0.0092	0.0106	0.0119		
	0.8	0.0057	0.0070	0.0084	0.0098	0.0111	0.0125	0.0139	
	0.9	0.0062	0.0076	0.0089	0.0103	0.0117	0.0130	0.0144	0.0158
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.677							
	0.3	0.735	0.637						
	0.4	0.776	0.677	0.622					
	0.5	0.805	0.709	0.652	0.613				
	0.6	0.828	0.735	0.677	0.637	0.608			
	0.7	0.846	0.757	0.699	0.658	0.628	0.605		
	0.8	0.861	0.776	0.718	0.677	0.646	0.622	0.602	
	0.9	0.873	0.792	0.735	0.694	0.662	0.637	0.617	0.600

Table 5: Mean Square Errors and relative efficiency of proposed estimator for Population-5

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	0.0670							
	0.3	0.0990	0.1018						
	0.4	0.1311	0.1339	0.1367					
	0.5	0.1632	0.1660	0.1688	0.1716				
	0.6	0.1953	0.1981	0.2009	0.2037	0.2065			
	0.7	0.2273	0.2301	0.2329	0.2357	0.2386	0.2414		
	0.8	0.2594	0.2622	0.2650	0.2678	0.2706	0.2734	0.2762	
	0.9	0.2915	0.2943	0.2971	0.2999	0.3027	0.3055	0.3083	0.3111
		MSE of Classical Regression Estimator							
θ_2	**0.2**	0.0723							
	0.3	0.1044	0.1126						
	0.4	0.1365	0.1447	0.1529					
	0.5	0.1686	0.1768	0.1849	0.1931				
	0.6	0.2006	0.2088	0.2170	0.2252	0.2334			
	0.7	0.2327	0.2409	0.2491	0.2573	0.2655	0.2737		
	0.8	0.2648	0.2730	0.2812	0.2894	0.2975	0.3057	0.3139	
	0.9	0.2969	0.3051	0.3132	0.3214	0.3296	0.3378	0.3460	0.3542
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.93							
	0.3	0.95	0.90						
	0.4	0.96	0.93	0.89					
	0.5	0.97	0.94	0.91	0.89				
	0.6	0.97	0.95	0.93	0.90	0.88			
	0.7	0.98	0.96	0.94	0.92	0.90	0.88		
	0.8	0.98	0.96	0.94	0.93	0.91	0.89	0.88	
	0.9	0.98	0.96	0.95	0.93	0.92	0.90	0.89	0.88

3.5 Numerical Study Qualitative Predictors

In this section empirical study has been carried out by using natural population available in literature (Annex-A). R code (Annex-B) is used to compute the mean Square error of proposed and classical regression estimators has been computed and relative efficiencies are compared for various values of θ_1 and θ_2. The results of empirical study clearly shows that for all values of θ_1 and θ_2, the proposed estimator is more precise as compared with the regression estimator. It can also be seen that for fixed θ_1 the efficiency decreases with the increase in θ_2. Also for fixed θ_2 the efficiency increases with increase in θ_1.

Table 6: Mean Square Errors and relative efficiency of proposed estimator for Population -6

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	0.0013							
	0.3	0.0021	0.0019						
	0.4	0.0029	0.0027	0.0024					
	0.5	0.0037	0.0035	0.0032	0.0030				
	0.6	0.0045	0.0042	0.0040	0.0038	0.0035			
	0.7	0.0053	0.0050	0.0048	0.0045	0.0043	0.0041		
	0.8	0.0061	0.0058	0.0056	0.0053	0.0051	0.0049	0.0046	
	0.9	0.0068	0.0066	0.0064	0.0061	0.0059	0.0056	0.0054	0.0052
		MSE of Classical Regression Estimator							
θ_2	**0.2**	0.0027							
	0.3	0.0035	0.0046						
	0.4	0.0043	0.0054	0.0065					
	0.5	0.0050	0.0062	0.0073	0.0084				
	0.6	0.0058	0.0070	0.0081	0.0092	0.0103			
	0.7	0.0066	0.0077	0.0089	0.0100	0.0111	0.0122		
	0.8	0.0074	0.0085	0.0096	0.0108	0.0119	0.0130	0.0141	
	0.9	0.0082	0.0093	0.0104	0.0115	0.0127	0.0138	0.0149	0.0160
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.4956							
	0.3	0.6098	0.4092						
	0.4	0.6818	0.4956	0.3735					
	0.5	0.7314	0.5600	0.4412	0.3539				
	0.6	0.7676	0.6098	0.4956	0.4092	0.3416			
	0.7	0.7952	0.6494	0.5404	0.4558	0.3883	0.3331		
	0.8	0.8169	0.6818	0.5779	0.4956	0.4288	0.3735	0.3269	
	0.9	0.8345	0.7087	0.6098	0.5300	0.4643	0.4092	0.3624	0.3222

Table 7: Mean Square Errors and relative efficiency of proposed estimator for Population-7

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	0.0037							
	0.3	0.0055	0.0055						
	0.4	0.0073	0.0073	0.0073					
	0.5	0.0091	0.0091	0.0092	0.0092				
	0.6	0.0110	0.0110	0.0110	0.0110	0.0110			
	0.7	0.0128	0.0128	0.0128	0.0128	0.0128	0.0128		
	0.8	0.0146	0.0146	0.0146	0.0146	0.0146	0.0146	0.0147	
	0.9	0.0165	0.0165	0.0165	0.0165	0.0165	0.0165	0.0165	0.0165
		MSE of Classical Regression Estimator							
θ_2	**0.2**	0.0037							
	0.3	0.0056	0.0056						
	0.4	0.0074	0.0075	0.0075					
	0.5	0.0092	0.0093	0.0094	0.0094				
	0.6	0.0110	0.0111	0.0112	0.0113	0.0113			
	0.7	0.0129	0.0129	0.0130	0.0131	0.0132	0.0132		
	0.8	0.0147	0.0148	0.0148	0.0149	0.0150	0.0151	0.0151	
	0.9	0.0165	0.0166	0.0167	0.0167	0.0168	0.0169	0.0170	0.0170
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.9813							
	0.3	0.9875	0.9752						
	0.4	0.9906	0.9813	0.9722					
	0.5	0.9924	0.9850	0.9777	0.9704				
	0.6	0.9937	0.9875	0.9813	0.9752	0.9693			
	0.7	0.9946	0.9892	0.9839	0.9787	0.9735	0.9684		
	0.8	0.9953	0.9906	0.9859	0.9813	0.9767	0.9722	0.9678	
	0.9	0.9958	0.9916	0.9875	0.9833	0.9793	0.9752	0.9712	0.9673

Table 8: Mean Square Errors and relative efficiency of proposed estimator for Population-8

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	8.55							
	0.3	12.75	12.90						
	0.4	16.96	17.10	17.25					
	0.5	21.16	21.31	21.45	21.60				
	0.6	25.36	25.51	25.65	25.80	25.95			
	0.7	29.56	29.71	29.86	30.00	30.15	30.30		
	0.8	33.76	33.91	34.06	34.21	34.35	34.50	34.65	
	0.9	37.97	38.11	38.26	38.41	38.56	38.70	38.85	39.00
		MSE of Classical Regression Estimator							
θ_2	**0.2**	8.68							
	0.3	12.88	13.15						
	0.4	17.08	17.36	17.63					
	0.5	21.28	21.56	21.83	22.11				
	0.6	25.49	25.76	26.03	26.31	26.58			
	0.7	29.69	29.96	30.24	30.51	30.78	31.06		
	0.8	33.89	34.16	34.44	34.71	34.99	35.26	35.53	
	0.9	38.09	38.37	38.64	38.91	39.19	39.46	39.74	40.01
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.9854							
	0.3	0.9902	0.9808						
	0.4	0.9926	0.9854	0.9785					
	0.5	0.9941	0.9883	0.9826	0.9771				
	0.6	0.9950	0.9902	0.9854	0.9808	0.9762			
	0.7	0.9957	0.9916	0.9875	0.9834	0.9795	0.9756		
	0.8	0.9963	0.9926	0.9890	0.9854	0.9819	0.9785	0.9751	
	0.9	0.9967	0.9934	0.9902	0.9870	0.9839	0.9808	0.9777	0.9747

Table 9: Mean Square Errors and relative efficiency of proposed estimator for Population-9

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	0.0451							
	0.3	0.0676	0.0678						
	0.4	0.0900	0.0902	0.0904					
	0.5	0.1125	0.1127	0.1129	0.1131				
	0.6	0.1349	0.1352	0.1354	0.1356	0.1358			
	0.7	0.1574	0.1576	0.1578	0.1580	0.1582	0.1584		
	0.8	0.1799	0.1801	0.1803	0.1805	0.1807	0.1809	0.1811	
	0.9	0.2023	0.2025	0.2027	0.2029	0.2031	0.2033	0.2035	0.2037
		MSE of Classical Regression Estimator							
θ_2	**0.2**	0.0794							
	0.3	0.1019	0.1364						
	0.4	0.1244	0.1589	0.1934					
	0.5	0.1468	0.1813	0.2158	0.2504				
	0.6	0.1693	0.2038	0.2383	0.2728	0.3073			
	0.7	0.1917	0.2262	0.2608	0.2953	0.3298	0.3643		
	0.8	0.2142	0.2487	0.2832	0.3177	0.3523	0.3868	0.4213	
	0.9	0.2366	0.2712	0.3057	0.3402	0.3747	0.4092	0.4438	0.4783
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.5680							
	0.3	0.6632	0.4969						
	0.4	0.7240	0.5680	0.4676					
	0.5	0.7662	0.6215	0.5230	0.4517				
	0.6	0.7973	0.6632	0.5680	0.4969	0.4417			
	0.7	0.8210	0.6966	0.6052	0.5351	0.4797	0.4348		
	0.8	0.8398	0.7240	0.6365	0.5680	0.5129	0.4676	0.4298	
	0.9	0.8550	0.7469	0.6632	0.5965	0.5421	0.4969	0.4587	0.4260

Table 10: Mean Square Errors and relative efficiency of proposed estimator for Population-10

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		MSE of Proposed Estimator							
θ_2	**0.2**	42.40							
	0.3	65.30	61.89						
	0.4	88.21	84.80	81.38					
	0.5	111.11	107.70	104.29	100.88				
	0.6	134.02	130.61	127.19	123.78	120.37			
	0.7	156.92	153.51	150.10	146.69	143.28	139.86		
	0.8	179.83	176.42	173.01	169.59	166.18	162.77	159.36	
	0.9	202.73	199.32	195.91	192.50	189.09	185.67	182.26	178.85
		MSE of Classical Regression Estimator							
θ_2	**0.2**	49.71							
	0.3	72.61	76.51						
	0.4	95.52	99.42	103.32					
	0.5	118.42	122.32	126.22	130.12				
	0.6	141.33	145.23	149.13	153.03	156.92			
	0.7	164.23	168.13	172.03	175.93	179.83	183.73		
	0.8	187.14	191.04	194.94	198.84	202.73	206.63	210.53	
	0.9	210.04	213.94	217.84	221.74	225.64	229.54	233.44	237.34
		Relative Efficiency of Proposed Estimator over Classical Regression Estimator							
θ_2	**0.2**	0.8529							
	0.3	0.8993	0.8089						
	0.4	0.9235	0.8529	0.7877					
	0.5	0.9383	0.8805	0.8262	0.7753				
	0.6	0.9483	0.8993	0.8529	0.8089	0.7671			
	0.7	0.9555	0.9130	0.8725	0.8338	0.7967	0.7613		
	0.8	0.9609	0.9235	0.8875	0.8529	0.8197	0.7877	0.7569	
	0.9	0.9652	0.9317	0.8993	0.8681	0.8380	0.8089	0.7808	0.7536

Chapter 4: New Multivariate Estimators

4.1 Introduction

In this chapter some new multivariate estimators for two phase sampling has been proposed. The proposed estimators use information on multiple quantitative variables and as well as multiple qualitative variables. Empirical study has also been conducted to see the performance of the proposed estimators.

4.2 New Multivariate Estimator with Quantitative Predictors

In this section the multivariate extension of Roy(2003) estimator has been proposed. The multivariate extension has been proposed by using information on two auxiliary variables and can be used for simultaneous estimation of several variables. The multivariate extension is proposed below:

Suppose a first phase random sample of size n_1 is available and information on auxiliary variables X and W is recorded. Suppose further that a second phase random sample of size n_2 is available and information on auxiliary variables X and W has been collected alongside information of multiple response variables $Y_1, Y_2..., Y_p$. Suppose further that $\bar{\mathbf{y}}_2$ is mean vector of estimates based upon second phase sample, $\mathbf{k}$ is a vector of constants and $\mathbf{A}$ & $\mathbf{B}$ are diagonal matrices with diagonal entries α_i & β_i respectively. Based upon these information, the multivariate estimator is defined below:

$$\underline{\mathbf{t}}_{N(2)} = \bar{\mathbf{y}}_2 + \left(\bar{x}_1 - \bar{x}_2\right)\mathbf{k} + \left(\bar{W} - \bar{w}_1\right)\mathbf{Ak} + \left(\bar{W} - \bar{w}_2\right)\mathbf{Bk} \tag{4.2.1}$$

The ith component of (4.2.1) is given as:

$$t_{N_i(2)} = \bar{y}_{i2} + k_i\left[\left\{\bar{x}_1 + \alpha_i\left(\bar{W} - \bar{w}_1\right)\right\} - \left\{\bar{x}_2 + \beta_i\left(\bar{W} - \bar{w}_2\right)\right\}\right] \tag{4.2.2}$$

Using conventional transformation

$$\bar{w}_1 = \bar{W} - \bar{e}_{w_1}; \bar{w}_2 = \bar{W} - \bar{e}_{w_2}; \bar{y}_{i2} = \bar{Y}_i + \bar{e}_{y_{i2}}; \bar{x}_1 = \bar{X} - \bar{e}_{x_1}; \bar{x}_2 = \bar{X} - \bar{e}_{x_2}$$

the estimator (4.2.2) can be written in the following form:

$$t_{N_i(2)} = \left(\bar{y}_i + \bar{e}_{y_{i2}}\right) + k_i\left[\left(\bar{x} + \bar{e}_{x_1}\right) + \alpha_i\left(\bar{W} - \bar{W} - \bar{e}_{w_1}\right) - \left\{\left(\bar{x} + \bar{e}_{x_2}\right) + \beta_i\left(\bar{W} - \bar{W} - \bar{e}_{w_2}\right)\right\}\right]$$

or $t_{N_i(2)} = \left(\bar{y}_i + \bar{e}_{y_{i2}}\right) + k_i\left[\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - \alpha_i\bar{e}_{w_1} + \beta_i\bar{e}_{w_2}\right]$

or $t_{N_i(2)} - y_i = \bar{e}_{y_{i2}} + k_i\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - k_i\alpha_i\bar{e}_{w_1} + k_i\beta_i\bar{e}_{w_2}$

Squaring above equation:

$$\begin{aligned}\left(t_i - y_i\right)^2 = {} & \bar{e}_{y_{i2}}^2 + k_i^2\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right)^2 + k_i^2\alpha_i^2\bar{e}_{w_1}^2 + k_i^2\beta_i^2\bar{e}_{w_2}^2 + 2k_i\bar{e}_{y_{i2}}^2\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - 2k_i\alpha_i\bar{e}_{y_{i2}}\bar{e}_{w_1} \\ & + 2k_i\beta_i\bar{e}_{y_{i2}}\bar{e}_{w_2} - 2k_i^2\alpha_i\bar{e}_{w_1}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) + 2k_i^2\beta_i\bar{e}_{w_2}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - 2k_i^2\alpha_i\beta_i\bar{e}_{w_1}\bar{e}_{w_2}\end{aligned}$$

By applying expectation, the mean square error of $t_{N_i(2)}$ is:

$$S_i = MSE\left(t_i\right) = E\left(t_i - \bar{y}_i\right)^2$$

$$\begin{aligned}S_i = {} & \theta_2 S_{y_i}^2 + \left(\theta_2 - \theta_1\right)k_i^2 S_x^2 + \theta_1 k_i^2\alpha_i^2 S_w^2 + \theta_2 k_i^2\beta_i^2 S_w^2 + 2\left(\theta_1 - \theta_2\right)k_i S_{xy_i} - 2\theta_1 k_i\alpha_i S_{wy_i} \\ & + 2\theta_2 k_i\beta_i S_{wy_i} + 2\left(\theta_1 - \theta_2\right)k_i^2\beta_i^2 S_{wx} - 2\theta_1 k_i^2\alpha_i\beta_i S_w^2\end{aligned} \qquad (4.2.3)$$

Optimum values of unknown quantities which minimize S_i can be obtain by differentiating (4.2.3) with respect to unknown quantities. The partial derivative of (4.2.3) with respect to α_i, β_i and k_i are:

$$\begin{aligned}\frac{\partial S_i}{\partial k_i} = {} & 2k_i\left(\theta_2 - \theta_1\right)s_x^2 + 2\theta_1 k_i\alpha_i^2 S_w^2 + 2\theta_2 k_i\beta_i^2 S_w^2 + 2\left(\theta_1 - \theta_2\right)S_{xy_i} - 2\theta_1\alpha_i S_{wy_i} \\ & + 2\theta_2\beta_i S_{wy_i} + 4\left(\theta_1 - \theta_2\right)k_i\beta_i S_{wx} - 4\theta_1 k_i\alpha_i\beta_i S_w^2\end{aligned}$$

$\frac{\partial S_i}{\partial k_i} = 0$ gives

$$\begin{aligned}& 2k_i\left(\theta_2 - \theta_1\right)s_x^2 + 2\theta_1 k_i\alpha_i^2 S_w^2 + 2\theta_2 k_i\beta_i^2 S_w^2 + 2\left(\theta_1 - \theta_2\right)S_{xy_i} - 2\theta_1\alpha_i S_{wy_i} \\ & + 2\theta_2\beta_i S_{wy_i} + 4\left(\theta_1 - \theta_2\right)k_i\beta_i S_{wx} - 4\theta_1 k_i\alpha_i\beta_i S_w^2 = 0\end{aligned}$$

or

$$\begin{aligned}& 2k_i\left(\theta_2 - \theta_1\right)S_x^2 + 2\theta_1 k_i\alpha_i^2 S_w^2 + 2\theta_2 k_i\beta_i^2 S_w^2 + 2\left(\theta_1 - \theta_2\right)S_{xy_i} - 2\theta_1\alpha_i S_{wy_i} + 2\theta_2\beta_i S_{wy_i} \\ & + 4\left(\theta_1 - \theta_2\right)k_i\beta_i S_{wx} - 4\theta_1 k_i\alpha_i\beta_i S_w^2 = 0\end{aligned} \qquad (4.2.4)$$

Again

$$\frac{\partial S_i}{\partial \alpha_i} = 2\theta_1 k_i^2 \alpha_i S_w^2 - 2\theta_1 k_i S_{wy_i} - 2\theta_1 k_i^2 \beta_i S_w^2$$

$\frac{\partial S_i}{\partial \alpha_i} = 0$ gives

$$2\theta_1 k_i^2 \alpha_i S_w^2 - 2\theta_1 k_i^2 \beta_i S_w^2 - 2\theta_1 k_i S_{wy_i} = 0$$

or $k_i(\alpha_i - \beta_i) S_w^2 - S_{wy_i} = 0$ (4.2.5)

Finally

$$\frac{\partial S_i}{\partial \beta_i} = 2\theta_2 k_i^2 \beta_i S_w^2 + 2\theta_2 k_i S_{wy_i} + 2(\theta_1 - \theta_2) k_i^2 S_{wx} - 2\theta_1 k_i^2 \alpha_i S_w^2$$

$\frac{\partial S_i}{\partial \beta_i} = 0$ gives:

$$2\theta_2 k_i^2 \beta_i S_w^2 - 2\theta_1 k_i^2 \alpha_i S_w^2 - 2(\theta_2 - \theta_1) k_i^2 S_{wx} + 2\theta_2 k_i S_{wy_i} = 0$$

$$k_i(\theta_2 \beta_i - \theta_1 \alpha_i) S_w^2 - k_i(\theta_2 - \theta_1) S_{wx} + \theta_2 S_{wy_i} = 0 \quad (4.2.6)$$

By solving (4.2.5) and (4.2.6) we get

$$\alpha_i = \frac{S_x}{S_w^2} = \beta_{xw} \quad (4.2.7)$$

Put (4.2.7) in (4.2.6)

$$k_i\left(\theta_2 \beta_i - \theta_1 \frac{S_{wx}}{S_w^2}\right) S_w^2 - k_i(\theta_2 - \theta_1) S_{wx} + \theta_2 S_{wy_i} = 0$$

$$k_i(\theta_2 \beta_i S_w^2 - \theta_1 S_{wx}) - k_i(\theta_2 - \theta_1) S_{wx} + \theta_2 S_{wy_i} = 0$$

$$k_i \beta_i \theta_2 S_w^2 - k_i \theta_2 S_{wx} + \theta_2 S_{wy_i} = 0$$

$$k_i \beta_i S_w^2 = k_i S_{wx} - S_{wy_i}$$

$$\beta_i = \frac{S_{wx}}{S_w^2} - \frac{1}{k_i}\frac{S_{wy_i}}{S_w^2} = \beta_{xw} - \frac{1}{k_i}\beta_{y_i w} \quad (4.2.8)$$

Using (4.2.7) and (4.2.8) in (4.2.4):

$$k_i\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\frac{S_{wx}^2}{S_w^4}S_w^2+\theta_2\left(\frac{S_{wx}}{S_w^2}-\frac{1}{k_i}\frac{S_{wy_i}}{S_w^2}\right)^2S_w^2+2\left(\theta_1-\theta_2\right)\left(\frac{S_{wx}}{S_w^2}-\frac{1}{k_i}\frac{S_{wy_i}}{S_w^2}\right)S_{wx}\right.$$

$$\left.-2\theta_1\frac{S_{wx}}{S_w^2}\left(\frac{S_{wx}}{S_w^2}-\frac{1}{k_i}\frac{S_{wy_i}}{S_w^2}\right)S_w^2\right]+2\left[\left(\theta_1-\theta_2\right)S_{xy_i}-\theta_1\frac{S_{wx}}{S_w^2}S_{wy_i}+\theta_2\left(\frac{S_{wx}}{S_w^2}-\frac{1}{k_i}\frac{S_{wy_i}}{S_w^2}\right)S_{wy_i}\right]=0$$

$$2k_i\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\frac{S_{wx}^2}{S_w^2}+\theta_2\left(\frac{S_{wx}^2}{S_w^2}+\frac{1}{k_i}\frac{S_{wy_i}^2}{S_w^2}-\frac{2}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}\right)+2\left(\theta_1-\theta_2\right)\left(\frac{S_{wx}^2}{S_w^2}-\frac{1}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}\right)\right.$$

$$\left.-2\theta_1\left(\frac{S_{wx}^2}{S_w^2}-\frac{1}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}\right)\right]+2\left[\left(\theta_1-\theta_2\right)S_{xy_i}-\theta_1\frac{S_{wx}S_{wy_i}}{S_w^2}+\theta_2\frac{S_{wx}S_{wy_i}}{S_w^2}-\frac{\theta_2}{k_i}\frac{S_{wy_i}^2}{S_w^2}\right]=0$$

$$k_i\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\frac{S_{wx}^2}{S_w^2}+\theta_2\frac{S_{wx}^2}{S_w^2}+\theta_2\frac{S_{wy_i}^2}{S_w^2}-\frac{2\theta_2}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}+2\theta_1\frac{S_{wx}^2}{S_w^2}-\frac{2\theta_1}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}-2\theta_2\frac{S_{wx}^2}{S_w^2}\right.$$

$$\left.+\frac{2\theta_2}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}-2\theta_1\frac{S_{wx}^2}{S_w^2}+\frac{2\theta_1}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}\right]+\left[\left(\theta_1-\theta_2\right)S_{xy_i}+\left(\theta_2-\theta_1\right)\frac{S_{wx}S_{wy_i}}{S_w^2}-\frac{\theta_2}{k_i}\frac{S_{wy_i}^2}{S_w^2}\right]=0$$

$$k_i\left[\left(\theta_2-\theta_1\right)S_x^2-\left(\theta_2-\theta_1\right)\frac{S_{wx}^2}{S_w^2}\right]-\left(\theta_2-\theta_1\right)S_{xy_i}+\left(\theta_2-\theta_1\right)\frac{S_{wx}S_{wy_i}}{S_w^2}=0$$

$$k_i=\frac{S_{xy_i}-\dfrac{S_{wx}S_{wy_i}}{S_w^2}}{S_x^2-\dfrac{S_{wx}^2}{S_w^2}}=\frac{\rho_{xy_i}S_xS_{y_i}-\rho_{wx}\rho_{wy_i}S_xS_{y_i}}{S_x^2-\rho_{wx}^2S_x^2}=\frac{S_{y_ix.w}}{S_{x.w}^2}$$

$$k_i=\left(\frac{\rho_{xy_i}-\rho_{wx}\rho_{wy_i}}{1-S_{wx}^2}\right)\frac{S_y}{S_x}=\beta_{y_ix.w} \qquad (4.2.9)$$

Using the values of (4.2.7), (4.2.8) and (4.2.9) in (4.2.3); the MSE becomes

$$S_i=\theta_2S_{y_i}^2+k_i^2\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\alpha_i^2S_w^2+\theta_2\beta_i^2S_w^2-2\theta_1\alpha_i\beta_iS_w^2+2\left(\theta_1-\theta_2\right)\beta_iS_{wx}\right]$$

$$+2k_i\left[\left(\theta_1-\theta_2\right)S_{xy_i}-\theta_1\alpha_iS_{wy_i}+\theta_2\beta_iS_{wy_i}\right]$$

$$=\theta_2S_{y_i}^2+k_i^2\left[\left(\theta_2-\theta_1\right)S_x^2+\theta_1\frac{S_{wx}^2}{S_w^2}+\theta_2\frac{S_{wx}^2}{S_w^2}+\frac{\theta_2}{k_i^2}\frac{S_{wy_i}^2}{S_w^2}-\frac{2\theta_2}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}-2\theta_1\frac{S_{wx}^2}{S_w^2}+\frac{2\theta_1}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}+2\theta_1\frac{S_{wx}^2}{S_w^2}\right.$$

$$\left.-\frac{2\theta_1}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}-2\theta_2\frac{S_{wx}^2}{S_w^2}+\frac{2\theta_2}{k_i}\frac{S_{wx}S_{wy_i}}{S_w^2}\right]+2k_i\left[\left(\theta_1-\theta_2\right)S_{xy_i}-\theta_1\frac{S_{wx}S_{wy_i}}{S_w^2}+\theta_2\frac{S_{wx}S_{wy_i}}{S_w^2}-\frac{\theta_2}{k_i}\frac{S_{wy_i}^2}{S_w^2}\right]$$

$$=\theta_2 S_{y_i}^2+k_i^2\left[(\theta_2-\theta_1)S_x^2-(\theta_2-\theta_1)\frac{S_{wx}^2}{S_w^2}\right]-2k_i(\theta_2-\theta_1)S_{xy_i}+2k_i(\theta_2-\theta_1)\frac{S_{wx}S_{wy_i}}{S_w^2}-\theta_2\frac{S_{wy_i}^2}{S_w^2}$$

$$=\theta_2 S_{y_i}^2+(\theta_2-\theta_1)k_i^2\left(S_x^2-\frac{S_{wx}^2}{S_w^2}\right)-2(\theta_2-\theta_1)k_i\left(S_{xy_i}-\frac{S_{wx}S_{wy_i}}{S_w^2}\right)-\theta_2\frac{S_{wy_i}^2}{S_w^2}$$

$$=\theta_2 S_{y_i}^2-\theta_2\frac{S_{wy_i}^2}{S_w^2}+(\theta_2-\theta_1)k_i^2\left(S_x^2-\frac{S_{wx}^2}{S_w^2}\right)-2(\theta_2-\theta_1)k_i^2\left(S_x^2-\frac{S_{wx}^2}{S_w^2}\right)$$

$$=\theta_2\left(S_{y_i}^2-\frac{S_{wy_i}^2}{S_w^2}\right)-(\theta_2-\theta_1)k_i^2 S_{wx}^2$$

$$=\theta_2 S_{y_i.w}^2-(\theta_2-\theta_1)\frac{S_{xy_i.w}}{S_{x.w}^2}$$

$$=S_{y_i.w}^2\left[\theta_2-(\theta_2-\theta_1)\rho_{xy_i.w}^2\right]$$

$$=S_{y_i.w}^2\left[\theta_2\left(1-\rho_{xy_i.w}^2\right)+\theta_1\rho_{xy_i.w}^2\right]$$

$$=S_{y_i}^2\left(1-\rho_{wy_i}^2\right)\left[\theta_2\left(1-\rho_{xy_i.w}^2\right)+\theta_1\rho_{xy_i.w}^2\right]$$

$$=S_{y_i}^2\left[\theta_2\left(1-\rho_{wy_i}^2\right)\left(1-\rho_{xy_i.w}^2\right)+\theta_1\rho_{xy_i.w}^2\left(1-\rho_{wy_i}^2\right)\right]$$

$$S_i=S_{y_i}^2\left[\theta_2\left(1-\rho_{y_i.wx}^2\right)+\theta_1\rho_{xy_i.w}^2\left(1-\rho_{wy_i}^2\right)\right] \qquad (4.2.10)$$

The covariance between any two components of (4.2.1) is derived as under:

$$t_{N_i(2)}=\bar{y}_{i2}+k_i\left[\left\{\bar{x}_1+\alpha_i\left(\bar{w}-\bar{w}_1\right)\right\}-\left\{\bar{x}_2+\beta_i\left(\bar{w}-\bar{w}_2\right)\right\}\right]$$

$$t_{N_j(2)}=\bar{y}_{j2}+k_j\left[\left\{\bar{x}_1+\alpha_j\left(\bar{w}-\bar{w}_1\right)\right\}-\left\{\bar{x}_2+\beta_j\left(\bar{w}-\bar{w}_2\right)\right\}\right]$$

Using conventional transformations; above estimators can be written as:

$$t_{N_i(2)}=\left(\bar{y}_i+\bar{e}_{y_{i2}}\right)+k_i\left[\left(\bar{x}+\bar{e}_{x_1}\right)+\alpha_i\left(\bar{w}-\bar{w}-\bar{e}_{w_1}\right)-\left\{\left(\bar{x}+\bar{e}_{x_2}\right)+\beta_i\left(\bar{w}-\bar{w}-\bar{e}_{w_2}\right)\right\}\right]$$

or $t_{N_i(2)}-y_i=\bar{e}_{y_{i2}}+k_i\left(\bar{e}_{x_1}+\bar{e}_{x_2}\right)-k_i\alpha_i\bar{e}_{w_1}+k_i\beta_i\bar{e}_{w_2}$

Similarly:

$$t_{N_j(2)}-y_j=\bar{e}_{y_{j2}}+k_j\left(\bar{e}_{x_1}+\bar{e}_{x_2}\right)-k_j\alpha_j\bar{e}_{w_1}+k_j\beta_j\bar{e}_{w_2}$$

Now

$$\left(t_{N_i(2)} - y_i\right)\left(t_{N_j(2)} - y_j\right) = \bar{e}_{y_{i2}}\bar{e}_{y_{j2}} + k_i\bar{e}_{y_{j2}}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - \alpha_i k_i \bar{e}_{y_{j2}}\bar{e}_{w_1} + \beta_i k_i \bar{e}_{y_{j2}}\bar{e}_{w_2} + k_j\bar{e}_{y_{i2}}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right)$$
$$+ k_i k_j\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right)^2 - \alpha_i k_i k_j \bar{e}_{w_1}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) + \beta_i k_i k_j \bar{e}_{w_2}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - \alpha_j k_j \bar{e}_{w_1}\bar{e}_{y_{i2}} - \alpha_j k_i k_j \bar{e}_{w_1}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right)$$
$$+ \alpha_i\alpha_j k_i k_j \bar{e}_{w_1}^2 + \beta_i k_i \alpha_j k_j \bar{e}_{w_1}\bar{e}_{w2} + \beta_j k_j \bar{e}_{w2}\bar{e}_{y_{i2}} + \beta_j k_i k_j \bar{e}_{w2}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - \alpha_i k_i \beta_j k_j \bar{e}_{w_1}\bar{e}_{w2} + \beta_i\beta_j k_i k_j \bar{e}_{w2}^2$$

By applying expectation to above equation, the covariance is:

$$S_{ij} = Cov\left(t_{N_i(2)}, t_{N_j(2)}\right) = E\left(t_{N_i(2)} - \bar{y}_i\right)\left(t_{N_j(2)} - \bar{y}_j\right)$$

$$S_{ij} = \theta_2 S_{y_i y_j} + k_i\left(\theta_1 - \theta_2\right)S_{xy_j} - \theta_1\alpha_i k_i S_{wy_j} + \theta_2\beta_i k_i S_{wy_j} + k_j\left(\theta_1 - \theta_2\right)S_{xy_i} + k_i k_j\left(\theta_2 - \theta_1\right)S_x^2$$
$$+\left(\theta_1 - \theta_2\right)\beta_i k_i k_j S_{wx} - \theta_1\alpha_j k_j S_{wy_i} + \theta_1\alpha_i\alpha_j k_i k_j S_w^2 + \theta_1\alpha_j\beta_i k_i k_j S_w^2 + \theta_2\beta_j k_j S_{wy_i}$$
$$+\left(\theta_1 - \theta_2\right)\beta_j k_i k_j S_{wx} - \theta_1\alpha_i\beta_j k_i k_j S_w^2 + \theta_2\beta_i\beta_j k_i k_j S_w^2$$

(4.2.11) Using (4.2.7), (4.2.8) and (4.2.9) in (4.2.11):

$$S_{ij} = \theta_2 S_{y_i y_j} + \beta_{xy_i w}\left(\theta_1 - \theta_2\right)S_{xy_j} + \theta_1\beta_{xw}\beta_{xy_i w}S_{wy_j} + \theta_2\left(\beta_{xw} - \frac{\beta_{wy_i}}{\beta_{xy_i w}}\right)\beta_{xy_i w}S_{wy_j} + \left(\theta_1 - \theta_2\right)\beta_{xy_j w}S_{xy_i}$$
$$+\left(\theta_2 - \theta_1\right)\beta_{xy_i w}\beta_{xy_j w}S_x^2 + \left(\theta_1 - \theta_2\right)\left(\beta_{xw} - \frac{\beta_{wy_i}}{\beta_{xy_i w}}\right)\beta_{xy_i w}\beta_{xy_j w}S_{wx} - \theta_1\beta_{xw}\beta_{xy_j w}S_{wy_i} + \theta_1\beta_{xw}^2\beta_{xy_i w}S_w^2$$
$$+\theta_1\beta_{xw}\left(\beta_{xw} - \frac{\beta_{wy_i}}{\beta_{xy_i w}}\right)\beta_{xy_i w}\beta_{xy_j w}S_w^2 + \theta_2\left(\beta_{xw} - \frac{\beta_{xy_j}}{\beta_{xy_j w}}\right)\beta_{xy_j w}S_{wy_i} + \left(\theta_1 - \theta_2\right)\left(\beta_{xw} - \frac{\beta_{wy_i}}{\beta_{xy_i w}}\right)\beta_{xy_i w}\beta_{xy_j w}S_{wx}$$
$$-\theta_1\beta_{xy}\left(\beta_{xw} - \frac{\beta_{wy_i}}{\beta_{xy_i w}}\right)\beta_{xy_i w}\beta_{xy_j w}S_w^2 + \theta_2\left(\beta_{xw} - \frac{\beta_{wy_i}}{\beta_{xy_i w}}\right)\left(\beta_{xw} - \frac{\beta_{wy_j}}{\beta_{xy_j w}}\right)\beta_{xy_i w}\beta_{xy_j w}S_w^2$$

$$S_{ij} = S_{y_i}S_{y_j}\left[\left\{\theta_2\left(\rho_{y_i y_j}\left(1 - \rho_{wx}^2\right) - \rho_{xy_i}\rho_{xy_j} - \rho_{wy_i}\rho_{wy_j} + \rho_{xy_i}\rho_{xy_j}\rho_{wx} + \rho_{xy_j}\rho_{wy_i}\rho_{wx}\right)\right.\right.$$
$$\left.\left. + \theta_1\left(\rho_{xy_i} - \rho_{wy_i}\right)\left(\rho_{xy_j} - \rho_{wy_j}\rho_{wx}\right)\right\} / \left(1 - \rho_{wx}^2\right)\right]$$

$$S_{ij} = S_{y_i}S_{y_j}\left[\theta_2\left\{\rho_{y_i y_j} - \frac{\rho_{xy_i}\rho_{xy_j} + \rho_{wy_i}\rho_{wy_j} - \rho_{xy_i}\rho_{wy_j}\rho_{wx} - \rho_{xy_j}\rho_{wy_i}\rho_{wx}}{1 - \rho_{wx}^2}\right\}\right.$$
$$\left. + \theta_1\rho_{xy_i . w}\rho_{xy_j . w}\sqrt{1 - \rho_{wy_i}^2}\sqrt{1 - \rho_{wy_j}^2}\right] \qquad (4.2.12)$$

The covariance matrix of (4.2.1) can be written by using (4.2.10) and (4.2.12)

4.3 New Multivariate Estimator with Qualitative Predictors

In this section the multivariate extension of Roy(2003) estimator has been proposed. The multivariate extension has been proposed by using information on two auxiliary attributes and can be used for simultaneous estimation of several variables. The multivariate extension is proposed as:

Suppose a first phase random sample of size n_1 is available and information on auxiliary attributes τ and W is recorded. Further a second phase random sample of size n_2 is available and information on auxiliary attributes τ and W has been collected alongside information of multiple response variables $Y_{1,} Y_2 ..., Y_p$. Suppose that $\bar{\mathbf{y}}_2$ is the mean vector of estimates based upon second phase, $\mathbf{k}$ is a vector of constants and $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices with diagonal entries γ_i and η_i respectively. Based upon these information, the multivariate estimator is defined below:

$$\underline{\mathbf{t}}_{c(2)} = \bar{\mathbf{y}}_2 + \left(\tau_1 - \tau_2\right)\mathbf{k} + \left(p_\delta - p_{\delta_1}\right)\mathbf{Ak} + \left(p_\delta - p_{\delta_2}\right)\mathbf{Bk} \quad (4.3.1)$$

The i th component of (4.3.1) is given as

$$t_{C_i(2)} = \bar{y}_{i2} + k_i\left[\left\{\tau_1 + \gamma_i\left(p_\delta - p_{\delta_1}\right)\right\} - \left\{\tau_2 + \eta_i\left(p_\delta - p_{\delta_2}\right)\right\}\right] \quad (4.3.2)$$

using conventional transformation

$\bar{y}_{i2} = \bar{y}_i + \bar{e}_{y_{i2}}$; $\tau_1 = \tau + \bar{e}_{\tau_1}$; $\tau_2 = \tau + \bar{e}_{\tau_2}$; $p_{\delta_1} = p_\delta - \bar{e}_{\delta_1}$ and $p_{\delta_2} = p_\delta - \bar{e}_{\delta_2}$

Using above representations, the estimator (4.3.2) can be put in the following form

$$t_{C_i(2)} = \left(\bar{y}_i + \bar{e}_{y_{i2}}\right) + k_i\left[\left(\tau + \bar{e}_{\tau_1}\right) + \gamma_i\left(p_\delta - p_\delta - \bar{e}_{\delta_1}\right) - \left\{\left(\tau + \bar{e}_{\tau_2}\right) + \eta_i\left(p_\delta - p_\delta - \bar{e}_{\delta_2}\right)\right\}\right]$$

or $t_{C_i(2)} = \left(\bar{y}_i + \bar{e}_{y_{i2}}\right) + k_i\left[\left(\bar{e}_{\tau_1} - \bar{e}_{\tau_2}\right) - \gamma_i\bar{e}_{\delta_1} + \eta_i\bar{e}_{\delta_2}\right]$

or $t_{C_i(2)} - y_i = \bar{e}_{y_{i2}} + k_i\left(\bar{e}_{\tau_1} - \bar{e}_{\tau_2}\right) - k_i\gamma_i\bar{e}_{\delta_1} + k_i\eta_i\bar{e}_{\delta_2}$

Squaring above equation:

$$\left(t_{C_i(2)} - y_i\right)^2 = \bar{e}_{y_{i2}}^2 + k_i^2\left(\bar{e}_{\tau_1} - \bar{e}_{\tau_2}\right)^2 + k_i^2\gamma_i^2\bar{e}_{\delta_1}^2 + k_i^2\eta_i^2\bar{e}_{\delta_2}^2 + 2k_i\bar{e}_{y_{i2}}^2\left(\bar{e}_{\tau_1} - \bar{e}_{\tau_2}\right) - 2k_i\gamma_i\bar{e}_{y_{i2}}\bar{e}_{\delta_1}$$
$$+ 2k_i\eta_i\bar{e}_{y_{i2}}\bar{e}_{\delta_2} - 2k_i^2\gamma_i\bar{e}_{\delta_1}\left(\bar{e}_{\tau_1} - \bar{e}_{\tau_2}\right) + 2k_i^2\eta_i\bar{e}_{\delta_2}\left(\bar{e}_{\tau_1} - \bar{e}_{\tau_2}\right) - 2k_i^2\gamma_i\eta_i\bar{e}_{\delta_1}\bar{e}_{\delta_2}$$

By applying the expectation, the mean square error of $t_{C_i(2)}$ is:

$S_i = MSE(t_i) = E(t_i - \bar{y}_i)^2$

or

$$S_i = \theta_2 S_{y_i}^2 + (\theta_2 - \theta_1) k_i^2 S_\tau^2 + \theta_1 k_i^2 \gamma_i^2 S_\delta^2 + \theta_2 k_i^2 \eta_i^2 S_\delta^2 + 2(\theta_1 - \theta_2) k_i S_{\tau y_i} - 2\theta_1 k_i \gamma_i S_{\delta y_i} + 2\theta_2 k_i \eta_i S_{\delta y_i} + 2(\theta_1 - \theta_2) k_i^2 \eta_i^2 S_{\delta\tau} - 2\theta_1 k_i^2 \gamma_i \eta_i S_\delta^2 \quad (4.3.3)$$

Optimum values of unknown quantities which minimize S_i can be obtained by differentiating (4.3.3) with respect to unknown quantities. The partial derivatives of (4.3.3) with respect of γ_i, η_i and k_i are:

$$\frac{\partial S_i}{\partial k_i} = 2k_i(\theta_2 - \theta_1) S_\tau^2 + 2\theta_1 k_i \gamma_i^2 S_\delta^2 + 2\theta_2 k_i \eta_i^2 S_\delta^2 + 2(\theta_1 - \theta_2) S_{\tau y_i} - 2\theta_1 \gamma_i S_{\delta y_i} + 2\theta_2 \eta_i S_{\delta y_i} + 4(\theta_1 - \theta_2) k_i \eta_i S_{\delta\tau} - 4\theta_1 k_i \gamma_i \eta_i S_\delta^2$$

$\frac{\partial S_i}{\partial k_i} = 0$ gives

$$2k_i(\theta_2 - \theta_1) S_\tau^2 + 2\theta_1 k_i \gamma_i^2 S_\delta^2 + 2\theta_2 k_i \eta_i^2 S_\delta^2 + 2(\theta_1 - \theta_2) S_{\tau y_i} - 2\theta_1 \gamma_i S_{\delta y_i} + 2\theta_2 \eta_i S_{\delta y_i} + 4(\theta_1 - \theta_2) k_i \beta_i S_{\delta\tau} - 4\theta_1 k_i \gamma_i \eta_i S_\delta^2 = 0$$

$$2k_i\left[(\theta_2 - \theta_1) S_\tau^2 + \theta_1 \gamma_i^2 S_\delta^2 + \theta_2 \eta_i^2 S_\delta^2 + 2(\theta_1 - \theta_2) \eta_i S_{\delta\tau} - 2\theta_1 \gamma_i \eta_i S_\delta^2\right] + 2\left[(\theta_1 - \theta_2) S_{\tau y_i} - \theta_1 \gamma_i S_{\delta y_i} + \theta_2 \eta_i S_{\delta y_i}\right] = 0 \quad (4.3.4)$$

Again

$$\frac{\partial S_i}{\partial \gamma_i} = 2\theta_1 k_i^2 \gamma_i S_\delta^2 - 2\theta_1 k_i S_{\delta y_i} - 2\theta_1 k_i^2 \eta_i S_\delta^2$$

$\frac{\partial S_i}{\partial \gamma_i} = 0$ gives

$$2\theta_1 k_i^2 \gamma_i S_\delta^2 - 2\theta_1 k_i^2 \eta_i S_\delta^2 - 2\theta_1 k_i S_{\delta y_i} = 0$$

$$k_i(\gamma_i - \eta_i) S_\delta^2 - S_{\delta y_i} = 0$$

(4.3.5)

$$\frac{\partial S_i}{\partial \eta_i} = 2\theta_2 k_i^2 \eta_i S_\delta^2 + 2\theta_2 k_i S_{\delta y_i} + 2(\theta_1 - \theta_2) k_i^2 S_{\delta\tau} - 2\theta_1 k_i^2 \gamma_i S_\delta^2$$

$\frac{\partial S_i}{\partial \eta_i} = 0$ gives

$$2\theta_2 k_i^2 \eta_i S_\delta^2 - 2\theta_1 k_i^2 \gamma_i S_\delta^2 - 2(\theta_2 - \theta_1) k_i^2 S_{\delta\tau} + 2\theta_2 k_i S_{\delta y_i} = 0$$

$$k_i(\theta_2 \eta_i - \theta_1 \gamma_i) S_\delta^2 - k_i(\theta_2 - \theta_1) S_{\delta\tau} + \theta_2 S_{\delta y_i} = 0 \tag{4.3.6}$$

By solving (4.3.5) and (4.3.6) we get

$$\gamma_i = \frac{S_{\delta\tau}}{S_\delta^2} = \beta_{\tau\delta} \tag{4.3.7}$$

Put (4.3.7) in (4.3.6)

$$k_i\left(\theta_2 \eta_i - \theta_1 \frac{S_{\delta\tau}}{S_\delta^2}\right) S_\delta^2 - k_i(\theta_2 - \theta_1) S_{\delta\tau} + \theta_2 S_{\delta y_i} = 0$$

$$k_i\left(\theta_2 \eta_i S_\delta^2 - \theta_1 S_{\delta\tau}\right) - k_i(\theta_2 - \theta_1) S_{\delta\tau} + \theta_2 S_{\delta y_i} = 0$$

$$k_i \eta_i \theta_2 S_\delta^2 - k_i \theta_2 S_{\delta\tau} + \theta_2 S_{\delta y_i} = 0$$

$$k_i \eta_i S_\delta^2 = k_i S_{\delta\tau} - S_{\delta y_i}$$

$$\eta_i = \frac{S_{\delta\tau}}{S_\delta^2} - \frac{1}{k_i}\frac{S_{\delta y_i}}{S_\delta^2} = \beta_{\tau\delta} - \frac{1}{k_i}\beta_{y_i\delta} \tag{4.3.8}$$

Put (4.3.7) and (4.2.8) in (4.3.4):

$$k_i\left[(\theta_2 - \theta_1) S_\tau^2 + \theta_1 \frac{S_{\delta\tau}^2}{S_\delta^4} s_\delta^2 + \theta_2\left(\frac{S_{\delta\tau}}{S_\delta^2} - \frac{1}{k_i}\frac{S_{\delta y_i}}{S_\delta^2}\right)^2 S_\delta^2 + 2(\theta_1 - \theta_2)\left(\frac{S_{\delta\tau}}{S_\delta^2} - \frac{1}{k_i}\frac{S_{\delta y_i}}{S_\delta^2}\right) S_{\delta\tau} \right.$$
$$\left. -2\theta_1 \frac{S_{\delta\tau}}{S_\delta^2}\left(\frac{S_{\delta\tau}}{S_\delta^2} - \frac{1}{k_i}\frac{S_{\delta y_i}}{S_\delta^2}\right) S_\delta^2\right] + 2\left[(\theta_1 - \theta_2) S_{\tau y_i} - \theta_1 \frac{S_{\delta\tau}}{S_\delta^2} S_{\delta y_i} + \theta_2\left(\frac{S_{\delta\tau}}{S_\delta^2} - \frac{1}{k_i}\frac{S_{\delta y_i}}{S_\delta^2}\right) S_{\delta y_i}\right] = 0$$

$$2k_i\left[(\theta_2 - \theta_1) S_\tau^2 + \theta_1 \frac{S_{\delta\tau}^2}{S_\delta^2} + \theta_2\left(\frac{S_{\delta\tau}^2}{S_\delta^2} + \frac{1}{k_i}\frac{S_{\delta y_i}^2}{S_\delta^2} - \frac{2}{k_i}\frac{S_{\delta\tau} S_{\delta y_i}}{S_\delta^2}\right) + 2(\theta_1 - \theta_2)\left(\frac{S_{\delta\tau}^2}{S_\delta^2} - \frac{1}{k_i}\frac{S_{\delta\tau} S_{\delta y_i}}{S_\delta^2}\right) \right.$$
$$\left. -2\theta_1\left(\frac{S_{\delta\tau}^2}{S_\delta^2} - \frac{1}{k_i}\frac{S_{\delta\tau} S_{\delta y_i}}{S_\delta^2}\right)\right] + 2\left[(\theta_1 - \theta_2) S_{\tau y_i} - \theta_1 \frac{S_{\delta\tau} S_{\delta y_i}}{S_\delta^2} + \theta_2 \frac{S_{\delta\tau} S_{\delta y_i}}{S_\delta^2} - \frac{\theta_2}{k_i}\frac{S_{\delta y_i}^2}{S_\delta^2}\right] = 0$$

$$k_i\left[\left(\theta_2-\theta_1\right)S_\tau^2+\theta_1\frac{S_{\delta\tau}^2}{S_\delta^2}+\theta_2\frac{S_{\delta\tau}^2}{S_\delta^2}+\theta_2\frac{S_{\delta y_i}^2}{S_\delta^2}-\frac{2\theta_2}{k_i}\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}+2\theta_1\frac{S_{\delta\tau}^2}{S_\delta^2}-\frac{2\theta_1}{k_i}\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}-2\theta_2\frac{S_{\delta\tau}^2}{S_\delta^2}\right.$$

$$\left.+\frac{2\theta_2}{k_i}\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}-2\theta_1\frac{S_{\delta\tau}^2}{S_\delta^2}+\frac{2\theta_1}{k_i}\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}\right]+\left[\left(\theta_1-\theta_2\right)S_{\tau y_i}+\left(\theta_2-\theta_1\right)\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}-\frac{\theta_2}{k_i}\frac{S_{\delta y_i}^2}{S_\delta^2}\right]=0$$

$$k_i\left[\left(\theta_2-\theta_1\right)S_\tau^2-\left(\theta_2-\theta_1\right)\frac{S_{\delta\tau}^2}{S_\delta^2}\right]-\left(\theta_2-\theta_1\right)S_{\tau y_i}+\left(\theta_2-\theta_1\right)\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}=0$$

$$k_i=\frac{S_{\tau y_i}-\dfrac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}}{S_\tau^2-\dfrac{S_{\delta\tau}^2}{S_\delta^2}}=\frac{\rho_{\tau y_i}S_\tau S_{y_i}-\rho_{\delta\tau}\rho_{\delta y_i}S_\tau S_{y_i}}{S_\tau^2-\rho_{\delta\tau}^2S_\tau^2}=\frac{S_{y_i\tau.\delta}}{S_{\tau.\delta}^2}$$

$$k_i=\left(\frac{\rho_{\tau y_i}-\rho_{\delta\tau}\rho_{\delta y_i}}{1-S_{\delta\tau}^2}\right)\frac{S_y}{S_\tau}=\beta_{y_i\tau.\delta} \tag{4.3.9}$$

Using the values of (4.3.7), (4.3.8) and (4.3.9) in (4.3.3); the MSE becomes

$$S_i=\theta_2S_{y_i}^2+k_i^2\left[\left(\theta_2-\theta_1\right)S_\tau^2+\theta_1\gamma_i^2S_\delta^2+\theta_2\eta_i^2S_\delta^2-2\theta_1\gamma_i\eta_iS_\delta^2+2\left(\theta_1-\theta_2\right)\eta_iS_{\delta\tau}\right]$$

$$+2k_i\left[\left(\theta_1-\theta_2\right)S_{\tau y_i}-\theta_1\gamma_iS_{\delta y_i}+\theta_2\eta_iS_{\delta y_i}\right]$$

$$=\theta_2S_{y_i}^2+k_i^2\left[\left(\theta_2-\theta_1\right)S_\tau^2+\theta_1\frac{S_{\delta\tau}^2}{S_\delta^2}+\theta_2\frac{S_{\delta\tau}^2}{S_\delta^2}+\frac{\theta_2}{k_i^2}\frac{S_{\delta y_i}^2}{S_\delta^2}-\frac{2\theta_2}{k_i}\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}-2\theta_1\frac{S_{\delta\tau}^2}{S_\delta^2}+\frac{2\theta_1}{k_i}\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}+2\theta_1\frac{S_{\delta\tau}^2}{S_\delta^2}\right.$$

$$\left.-\frac{2\theta_1}{k_i}\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}-2\theta_2\frac{S_{\delta\tau}^2}{S_\delta^2}+\frac{2\theta_2}{k_i}\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}\right]+2k_i\left[\left(\theta_1-\theta_2\right)S_{\tau y_i}-\theta_1\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}+\theta_2\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}-\frac{\theta_2}{k_i}\frac{S_{\delta y_i}^2}{S_\delta^2}\right]$$

$$S_i=\theta_2S_{y_i}^2+k_i^2\left[\left(\theta_2-\theta_1\right)S_\tau^2-\left(\theta_2-\theta_1\right)\frac{S_{\delta\tau}^2}{S_\delta^2}\right]-2k_i\left(\theta_2-\theta_1\right)S_{\tau y_i}+2k_i\left(\theta_2-\theta_1\right)\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}-\theta_2\frac{S_{\delta y_i}^2}{S_\delta^2}$$

$$=\theta_2S_{y_i}^2+\left(\theta_2-\theta_1\right)k_i^2\left(S_\tau^2-\frac{S_{\delta\tau}^2}{S_\delta^2}\right)-2\left(\theta_2-\theta_1\right)k_i\left(S_{\tau y_i}-\frac{S_{\delta\tau}S_{\delta y_i}}{S_\delta^2}\right)-\theta_2\frac{S_{\delta y_i}^2}{S_\delta^2}$$

$$=\theta_2S_{y_i}^2-\theta_2\frac{S_{\delta y_i}^2}{S_\delta^2}+\left(\theta_2-\theta_1\right)k_i^2\left(S_\tau^2-\frac{S_{\delta\tau}^2}{S_\delta^2}\right)-2\left(\theta_2-\theta_1\right)k_i^2\left(S_\tau^2-\frac{S_{\delta\tau}^2}{S_\delta^2}\right)$$

$$=\theta_2\left(S_{y_i}^2-\frac{S_{\delta y_i}^2}{S_\delta^2}\right)-\left(\theta_2-\theta_1\right)k_i^2S_{\delta\tau}^2$$

$$=\theta_2 S_{y_i.\delta}^2-\left(\theta_2-\theta_1\right)\frac{S_{\tau y_i.\delta}}{S_{\tau.\delta}^2}$$

$$=S_{y_i.\delta}^2\left[\theta_2-\left(\theta_2-\theta_1\right)\rho_{\tau y_i.\delta}^2\right]$$

$$=S_{y_i.\delta}^2\left[\theta_2\left(1-\rho_{\tau y_i.\delta}^2\right)+\theta_1\rho_{\tau y_i.\delta}^2\right]$$

$$=S_{y_i}^2\left(1-\rho_{\delta y_i}^2\right)\left[\theta_2\left(1-\rho_{\tau y_i.\delta}^2\right)+\theta_1\rho_{\tau y_i.\delta}^2\right]$$

$$=S_{y_i}^2\left[\theta_2\left(1-\rho_{\delta y_i}^2\right)\left(1-\rho_{\tau y_i.\delta}^2\right)+\theta_1\rho_{\tau y_i.\delta}^2\left(1-\rho_{\delta y_i}^2\right)\right]$$

$$S_i=S_{y_i}^2\left[\theta_2\left(1-\rho_{y_i.\delta\tau}^2\right)+\theta_1\rho_{\tau y_i.\delta}^2\left(1-\rho_{\delta y_i}^2\right)\right] \tag{4.3.10}$$

The covariance between any two components of (4.3.1) is derived as under:

$$t_{C_i(2)}=\overline{y}_{i2}+k_i\left[\left\{\tau_1+\gamma_i\left(p_\delta-p_{\delta_1}\right)\right\}-\left\{\tau_2+\eta_i\left(p_\delta-p_{\delta_2}\right)\right\}\right]$$

$$t_{C_j(2)}=\overline{y}_{j2}+k_j\left[\left\{\tau_1+\gamma_j\left(p_\delta-p_{\delta_1}\right)\right\}-\left\{\tau_2+\eta_j\left(p_\delta-p_{\delta_2}\right)\right\}\right]$$

Using conventional transformations:

$$t_{C_i(2)}=\left(\overline{y}_i+\overline{e}_{y_{i2}}\right)+k_i\left[\left(\tau+\overline{e}_{\tau_1}\right)+\gamma_i\left(p_\delta-p_\delta-\overline{e}_{\delta_1}\right)-\left\{\left(\tau+\overline{e}_{\tau_2}\right)+\eta_i\left(p_\delta-p_\delta-\overline{e}_{\delta_2}\right)\right\}\right]$$

$$t_{C_i(2)}-y_i=\overline{e}_{y_{i2}}+k_i\left(\overline{e}_{\tau_1}+\overline{e}_{\tau_2}\right)-k_i\gamma_i\overline{e}_{\delta_1}+k_i\eta_i\overline{e}_{\delta_2}$$

Similarly:

$$t_{C_j(2)}-y_j=\overline{e}_{y_{j2}}+k_j\left(\overline{e}_{\tau_1}+\overline{e}_{\tau_2}\right)-k_j\gamma_j\overline{e}_{\delta_1}+k_j\eta_j\overline{e}_{\delta_2}$$

Now

$$\begin{aligned}\left(t_{C_i(2)}-y_i\right)\left(t_{C_j(2)}-y_j\right)&=\overline{e}_{y_{i2}}\overline{e}_{y_{j2}}+k_i\overline{e}_{y_{j2}}\left(\overline{e}_{\tau_1}-\overline{e}_{\tau_2}\right)-\gamma_i k_i\overline{e}_{y_{j2}}\overline{e}_{\delta_1}+\eta_i k_i\overline{e}_{y_{j2}}\overline{e}_{\delta_2}+k_j\overline{e}_{y_{i2}}\left(\overline{e}_{\tau_1}-\overline{e}_{\tau_2}\right)\\&+k_ik_j\left(\overline{e}_{\tau_1}-\overline{e}_{\tau_2}\right)^2-\gamma_ik_ik_j\overline{e}_{\delta_1}\left(\overline{e}_{\tau_1}-\overline{e}_{\tau_2}\right)+\eta_ik_ik_j\overline{e}_{\delta_2}\left(\overline{e}_{\tau_1}-\overline{e}_{\tau_2}\right)-\gamma_jk_j\overline{e}_{\delta_1}\overline{e}_{y_{i2}}-\alpha_jk_ik_j\overline{e}_{\delta_1}\left(\overline{e}_{\tau_1}-\overline{e}_{\tau_2}\right)\\&+\gamma_i\gamma_jk_ik_j\overline{e}_{\delta_1}^2+\eta_ik_i\gamma_jk_j\overline{e}_{\delta_1}\overline{e}_{\delta2}+\eta_jk_j\overline{e}_{\delta2}\overline{e}_{y_{i2}}+\eta_jk_ik_j\overline{e}_{\delta_2}\left(\overline{e}_{\tau_1}-\overline{e}_{\tau_2}\right)-\gamma_ik_i\eta_jk_j\overline{e}_{\delta_1}\overline{e}_{\delta2}+\eta_i\eta_jk_ik_j\overline{e}_{\delta2}^2\end{aligned}$$

By applying expectation to above equation we get:

$$S_{ij}=Cov\left(t_{C_i(2)},t_{C_j(2)}\right)=E\left(t_{C_i(2)}-\overline{y}_i\right)\left(t_{C_j(2)}-\overline{y}_j\right)$$

$$S_{ij}=\theta_2 S_{y_iy_j}+k_i\left(\theta_1-\theta_2\right)S_{\tau y_j}-\theta_1\gamma_i k_i S_{\delta y_j}+\theta_2\eta_i k_i S_{\delta y_j}+k_j\left(\theta_1-\theta_2\right)S_{\tau y_i}+k_ik_j\left(\theta_2-\theta_1\right)S_\tau^2$$
$$+\left(\theta_1-\theta_2\right)\eta_i k_i k_j S_{\delta\tau}-\theta_1\gamma_j k_j S_{\delta y_i}+\theta_1\gamma_i\gamma_j k_i k_j S_\delta^2+\theta_1\gamma_j\eta_i k_i k_j S_\delta^2+\theta_2\eta_j k_j S_{\delta y_i} \qquad (4.3.11)$$
$$+\left(\theta_1-\theta_2\right)\eta_j k_i k_j S_{\delta\tau}-\theta_1\gamma_i\eta_j k_i k_j S_\delta^2+\theta_2\eta_i\eta_j k_i k_j S_\delta^2$$

Using (4.3.7), (4.3.8) and (4.3.9) in (4.3.11):

$$S_{ij}=\theta_2 S_{y_iy_j}+\beta_{\tau y_i\delta}\left(\theta_1-\theta_2\right)S_{\tau y_j}+\theta_1\beta_{\tau\delta}\beta_{\tau y_i.\delta}S_{wy_j}+\theta_2\left(\beta_{\tau w}-\frac{\beta_{\delta y_i}}{\beta_{\tau y_i.\delta}}\right)\beta_{\tau y_i.\delta}S_{\delta y_j}+\left(\theta_1-\theta_2\right)\beta_{\tau y_j.\delta}S_{\tau y_i}$$

$$+\left(\theta_2-\theta_1\right)\beta_{\tau y_i.\delta}\beta_{\tau y_j.\delta}S_\tau^2+\left(\theta_1-\theta_2\right)\left(\beta_{\tau\delta}-\frac{\beta_{\delta y_i}}{\beta_{\tau y_i.\delta}}\right)\beta_{\tau y_i.\delta}\beta_{\tau y_j.\delta}S_{\delta\tau}-\theta_1\beta_{\tau\delta}\beta_{\tau y_j.\delta}S_{\delta y_i}+\theta_1\beta_{\tau\delta}^2\beta_{\tau y_i.\delta}S_\delta^2$$

$$+\theta_1\beta_{\tau\delta}\left(\beta_{\tau\delta}-\frac{\beta_{\delta y_i}}{\beta_{\tau y_i.\delta}}\right)\beta_{\tau y_i.\delta}\beta_{\tau y_j.\delta}S_\delta^2+\theta_2\left(\beta_{\tau\delta}-\frac{\beta_{\tau y_j}}{\beta_{\tau y_j.\delta}}\right)\beta_{\tau y_j.\delta}S_{\delta y_i}+\left(\theta_1-\theta_2\right)\left(\beta_{\tau\delta}-\frac{\beta_{\delta y_i}}{\beta_{\tau y_i.\delta}}\right)\beta_{\tau y_i.\delta}\beta_{\tau y_j.\delta}S_{\delta\tau}$$

$$-\theta_1\beta_{\tau y}\left(\beta_{\tau\delta}-\frac{\beta_{\delta y_i}}{\beta_{\tau y_i.\delta}}\right)\beta_{\tau y_i.\delta}\beta_{\tau y_j.\delta}S_\delta^2+\theta_2\left(\beta_{\tau\delta}-\frac{\beta_{\delta y_i}}{\beta_{\tau y_i.\delta}}\right)\left(\beta_{\tau\delta}-\frac{\beta_{\delta y_j}}{\beta_{\tau y_j.\delta}}\right)\beta_{\tau y_i.\delta}\beta_{\tau y_j.\delta}S_\delta^2$$

$$S_{ij}=S_{y_i}S_{y_j}\Big[\Big\{\theta_2\left(\rho_{y_iy_j}\left(1-\rho_{\delta\tau}^2\right)-\rho_{\tau y_i}\rho_{\tau y_j}-\rho_{\delta y_i}\rho_{\delta y_j}+\rho_{\tau y_i}\rho_{\tau y_j}\rho_{\delta\tau}+\rho_{\tau y_j}\rho_{\delta y_i}\rho_{\delta\tau}\right)$$
$$+\theta_1\left(\rho_{\tau y_i}-\rho_{\delta y_i}\right)\left(\rho_{\tau y_j}-\rho_{\delta y_j}\rho_{\delta\tau}\right)\Big\}/\left(1-\rho_{\delta\tau}^2\right)\Big]$$

$$S_{ij}=S_{y_i}S_{y_j}\left[\theta_2\left\{\rho_{y_iy_j}-\frac{\rho_{\tau y_i}\rho_{\tau y_j}+\rho_{\delta y_i}\rho_{\delta y_j}-\rho_{\tau y_i}\rho_{\delta y_j}\rho_{\delta\tau}-\rho_{\tau y_j}\rho_{\delta y_i}\rho_{\delta\tau}}{1-\rho_{\delta\tau}^2}\right\}\right.$$
$$\left.+\theta_1\rho_{\tau y_i.\delta}\rho_{\tau y_j.\delta}\sqrt{1-\rho_{\delta y_i}^2}\sqrt{1-\rho_{\delta y_j}^2}\right] \qquad (4.3.12)$$

The covariance matrix of (4.3.1) can be written by using (4.3.10) and (4.3.12)

4.4 Numerical Study

In this section empirical study has been carried out by using natural population available in literature (Annex-A). R code (Annex-B) is used to compute the mean Square error of proposed and estimator and estimator proposed by Z. Ahmed, et al.(2010). Relative efficiencies are compared for various values of θ_1 and θ_2. The results of empirical study clearly shows that for all values of θ_1 and θ_2, the proposed multivariate estimator is more precise as compared with the estimator proposed by Z. Ahmed, et al.(2010). It can also be seen that for fixed θ_1 the efficiency decreases with the increase in θ_2. Also for fixed θ_2 the efficiency increases with increase in θ_1.

Table 4.4.1: Variance-Covariance matrices of proposed estimator for various values of θ_1 and θ_2 for Population 11

θ_1	θ_2	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
0.1	0.2	9.7	3.9	2.2	3.9	11.9	1.3	2.2	1.3	13.4
0.1	0.3	13.9	5.3	2.8	5.3	17.3	1.3	2.8	1.3	19.5
0.1	0.4	18.2	6.7	3.3	6.7	22.6	1.4	3.3	1.4	25.6
0.1	0.5	22.4	8.0	3.8	8.0	28.0	1.4	3.8	1.4	31.7
0.1	0.6	26.7	9.4	4.3	9.4	33.3	1.4	4.3	1.4	37.8
0.1	0.7	31.0	10.7	4.9	10.7	38.6	1.5	4.9	1.5	43.9
0.1	0.8	35.2	12.1	5.4	12.1	44.0	1.5	5.4	1.5	50.0
0.1	0.9	39.5	13.4	5.9	13.4	49.3	1.5	5.9	1.5	56.1
0.2	0.3	15.1	6.5	4.0	6.5	18.6	2.6	4.0	2.6	20.7
0.2	0.4	19.4	7.9	4.5	7.9	23.9	2.6	4.5	2.6	26.8
0.2	0.5	23.6	9.2	5.0	9.2	29.2	2.6	5.0	2.6	32.9
0.2	0.6	27.9	10.6	5.5	10.6	34.6	2.7	5.5	2.7	39.0
0.2	0.7	32.1	12.0	6.1	12.0	39.9	2.7	6.1	2.7	45.1
0.2	0.8	36.4	13.3	6.6	13.3	45.2	2.7	6.6	2.7	51.2
0.2	0.9	40.6	14.7	7.1	14.7	50.6	2.8	7.1	2.8	57.3
0.3	0.4	20.5	9.1	5.7	9.1	25.2	3.9	5.7	3.9	28.0
0.3	0.5	24.8	10.5	6.2	10.5	30.5	3.9	6.2	3.9	34.1
0.3	0.6	29.1	11.8	6.7	11.8	35.8	3.9	6.7	3.9	40.2
0.3	0.7	33.3	13.2	7.3	13.2	41.2	4.0	7.3	4.0	46.3
0.3	0.8	37.6	14.5	7.8	14.5	46.5	4.0	7.8	4.0	52.4
0.3	0.9	41.8	15.9	8.3	15.9	51.8	4.0	8.3	4.0	58.5
0.4	0.5	26.0	11.7	7.4	11.7	31.8	5.1	7.4	5.1	35.3
0.4	0.6	30.2	13.0	7.9	13.0	37.1	5.2	7.9	5.2	41.4
0.4	0.7	34.5	14.4	8.5	14.4	42.5	5.2	8.5	5.2	47.5
0.4	0.8	38.7	15.8	9.0	15.8	47.8	5.2	9.0	5.2	53.6
0.4	0.9	43.0	17.1	9.5	17.1	53.1	5.3	9.5	5.3	59.7
0.5	0.6	31.4	14.3	9.1	14.3	38.4	6.4	9.1	6.4	42.6
0.5	0.7	35.7	15.6	9.7	15.6	43.7	6.4	9.7	6.4	48.7
0.5	0.8	39.9	17.0	10.2	17.0	49.1	6.5	10.2	6.5	54.8
0.5	0.9	44.2	18.3	10.7	18.3	54.4	6.5	10.7	6.5	60.9
0.6	0.7	36.8	16.9	10.8	16.9	45.0	7.7	10.8	7.7	50.0
0.6	0.8	41.1	18.2	11.4	18.2	50.3	7.7	11.4	7.7	56.0
0.6	0.9	45.3	19.6	11.9	19.6	55.7	7.8	11.9	7.8	62.1
0.7	0.8	42.3	19.4	12.6	19.4	51.6	9.0	12.6	9.0	57.3
0.7	0.9	46.5	20.8	13.1	20.8	57.0	9.0	13.1	9.0	63.4
0.8	0.9	47.7	22.0	14.3	22.0	58.2	10.2	14.3	10.2	64.6

Table 4.4.2: Variance-Covariance matrices of Z. Ahmed, et al.(2010) estimator for various values of θ_1 and θ_2 for Population 11

θ_1	θ_2	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
0.1	0.2	13.0	2.9	1.0	2.9	15.1	-0.3	1.0	-0.3	17.6
0.1	0.3	17.3	2.1	-1.0	2.1	20.5	-2.9	-1.0	-2.9	23.7
0.1	0.4	21.5	1.3	-2.9	1.3	25.8	-5.5	-2.9	-5.5	29.8
0.1	0.5	25.8	0.5	-4.9	0.5	31.1	-8.2	-4.9	-8.2	35.9
0.1	0.6	30.0	-0.3	-6.9	-0.3	36.5	-10.8	-6.9	-10.8	42.0
0.1	0.7	34.3	-1.1	-8.8	-1.1	41.8	-13.4	-8.8	-13.4	48.1
0.1	0.8	38.6	-1.9	-10.8	-1.9	47.1	-16.0	-10.8	-16.0	54.2
0.1	0.9	42.8	-2.7	-12.8	-2.7	52.5	-18.6	-12.8	-18.6	60.3
0.2	0.3	21.8	6.6	4.0	6.6	24.9	2.0	4.0	2.0	29.1
0.2	0.4	26.1	5.8	2.0	5.8	30.3	-0.6	2.0	-0.6	35.2
0.2	0.5	30.3	5.0	0.1	5.0	35.6	-3.2	0.1	-3.2	41.3
0.2	0.6	34.6	4.2	-1.9	4.2	40.9	-5.9	-1.9	-5.9	47.4
0.2	0.7	38.8	3.4	-3.9	3.4	46.3	-8.5	-3.9	-8.5	53.5
0.2	0.8	43.1	2.6	-5.9	2.6	51.6	-11.1	-5.9	-11.1	59.6
0.2	0.9	47.3	1.8	-7.8	1.8	56.9	-13.7	-7.8	-13.7	65.7
0.3	0.4	30.6	10.3	7.0	10.3	34.7	4.3	7.0	4.3	40.7
0.3	0.5	34.8	9.5	5.0	9.5	40.1	1.7	5.0	1.7	46.8
0.3	0.6	39.1	8.7	3.0	8.7	45.4	-0.9	3.0	-0.9	52.9
0.3	0.7	43.3	7.9	1.1	7.9	50.7	-3.6	1.1	-3.6	59.0
0.3	0.8	47.6	7.1	-0.9	7.1	56.1	-6.2	-0.9	-6.2	65.1
0.3	0.9	51.9	6.3	-2.9	6.3	61.4	-8.8	-2.9	-8.8	71.1
0.4	0.5	39.4	14.0	10.0	14.0	44.5	6.6	10.0	6.6	52.2
0.4	0.6	43.6	13.2	8.0	13.2	49.9	4.0	8.0	4.0	58.3
0.4	0.7	47.9	12.4	6.0	12.4	55.2	1.4	6.0	1.4	64.4
0.4	0.8	52.1	11.5	4.0	11.5	60.6	-1.3	4.0	-1.3	70.5
0.4	0.9	56.4	10.7	2.1	10.7	65.9	-3.9	2.1	-3.9	76.6
0.5	0.6	48.1	17.6	12.9	17.6	54.3	8.9	12.9	8.9	63.7
0.5	0.7	52.4	16.8	11.0	16.8	59.7	6.3	11.0	6.3	69.8
0.5	0.8	56.6	16.0	9.0	16.0	65.0	3.7	9.0	3.7	75.9
0.5	0.9	60.9	15.2	7.0	15.2	70.4	1.0	7.0	1.0	82.0
0.6	0.7	56.9	21.3	15.9	21.3	64.2	11.2	15.9	11.2	75.2
0.6	0.8	61.2	20.5	14.0	20.5	69.5	8.6	14.0	8.6	81.3
0.6	0.9	65.4	19.7	12.0	19.7	74.8	6.0	12.0	6.0	87.4
0.7	0.8	65.7	25.0	18.9	25.0	74.0	13.5	18.9	13.5	86.8
0.7	0.9	69.9	24.2	16.9	24.2	79.3	10.9	16.9	10.9	92.9
0.8	0.9	74.5	28.7	21.9	28.7	83.8	15.8	21.9	15.8	98.3

Table 4.4.3: Eigen values for matrices given in table 4.4.1

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		λ_1							
θ_2	**0.2**	16.71							
	0.3	23.23	26.89						
	0.4	29.75	33.41	37.08					
	0.5	36.28	39.94	43.60	47.26				
	0.6	42.80	46.46	50.12	53.79	57.45			
	0.7	49.33	52.98	56.64	60.31	63.97	67.64		
	0.8	55.85	59.51	63.17	66.83	70.49	74.16	77.82	
	0.9	62.38	66.03	69.69	73.35	77.01	80.68	84.34	88.01
		λ_2							
θ_2	**0.2**	11.78							
	0.3	17.67	17.67						
	0.4	23.55	23.56	23.56					
	0.5	29.44	29.44	29.44	29.45				
	0.6	35.33	35.33	35.33	35.33	35.33			
	0.7	41.21	41.22	41.22	41.22	41.22	41.22		
	0.8	47.10	47.11	47.11	47.11	47.11	47.11	47.11	
	0.9	52.99	52.99	53.00	53.00	53.00	53.00	53.00	53.00
		λ_3							
θ_2	**0.2**	6.55							
	0.3	9.82	9.83						
	0.4	13.10	13.10	13.11					
	0.5	16.37	16.38	16.38	16.39				
	0.6	19.64	19.65	19.66	19.66	19.67			
	0.7	22.91	22.92	22.93	22.94	22.94	22.94		
	0.8	26.18	26.19	26.20	26.21	26.21	26.22	26.22	
	0.9	29.46	29.47	29.47	29.48	29.49	29.49	29.50	29.50

Table 4.4.4: Eigen values for matrices given in table 4.4.2

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		μ_1							
	0.2	17.91							
	0.3	25.87	33.80						
	0.4	34.46	35.81	49.85					
θ_2	**0.5**	43.07	43.18	51.57	65.90				
	0.6	51.67	51.73	53.72	67.59	81.96			
	0.7	60.28	60.32	60.54	69.36	83.64	98.01		
	0.8	68.88	68.92	69.03	71.63	85.36	99.69	114.07	
	0.9	77.49	77.53	77.60	77.97	87.17	101.39	115.75	130.13
		μ_2							
	0.2	17.02							
	0.3	19.38	25.79						
	0.4	22.18	34.04	34.41					
θ_2	**0.5**	27.37	36.97	42.95	43.03				
	0.6	32.78	38.77	51.07	51.59	51.64			
	0.7	38.21	40.82	54.53	60.09	60.21	60.25		
	0.8	43.65	44.37	56.35	68.09	68.75	68.82	68.86	
	0.9	49.09	49.41	58.15	72.02	77.20	77.38	77.44	77.47
		μ_3							
	0.2	10.86							
	0.3	16.22	16.30						
	0.4	20.51	21.72	21.74					
	0.5	22.40	27.11	27.17	27.18				
θ_2	**0.6**	24.07	32.44	32.58	32.61	32.62			
	0.7	25.72	37.48	37.98	38.03	38.04	38.05		
	0.8	27.36	41.02	43.35	43.44	43.47	43.48	43.49	
	0.9	28.99	43.05	48.67	48.84	48.89	48.91	48.92	48.93

Table 4.4.5: Relative Efficiencies of the proposed estimator over Z. Ahmed, et al.(2010) estimator

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		$\frac{\sum \lambda_i}{\sum \mu_i}$							
	0.2	0.77							
	0.3	0.83	0.72						
	0.4	0.86	0.77	0.70					
θ_2	**0.5**	0.88	0.80	0.73	0.68				
	0.6	0.90	0.83	0.77	0.72	0.68			
	0.7	0.91	0.84	0.79	0.74	0.70	0.67		
	0.8	0.92	0.86	0.81	0.77	0.73	0.70	0.67	
	0.9	0.93	0.87	0.83	0.78	0.75	0.72	0.69	0.66

Table 4.4.5 shows efficiency comparison of proposed multivariate estimator with estimator proposed by Z. Ahmed, et al.(2010). Sum of Eigen values of covariance matrices is used to calculate relative efficiencies. The above entries clearly indicate that the proposed estimator is more efficient as compared with the estimator proposed by Z. Ahmed, et al.(2010) for all combinations of θ_1 and θ_2.

Table 4.4.6: Variance-Covariance matrices of proposed estimator for various values of θ_1 and θ_2 for Population 12

θ_1	θ_2	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
0.1	0.2	0.105	0.207	-0.015	0.207	1.723	-0.090	-0.015	-0.090	0.950
0.1	0.3	0.157	0.311	-0.023	0.311	2.583	-0.130	-0.023	-0.130	1.411
0.1	0.4	0.210	0.414	-0.030	0.414	3.443	-0.170	-0.030	-0.170	1.871
0.1	0.5	0.262	0.518	-0.038	0.518	4.302	-0.210	-0.038	-0.210	2.332
0.1	0.6	0.315	0.622	-0.045	0.622	5.162	-0.250	-0.045	-0.250	2.792
0.1	0.7	0.367	0.725	-0.052	0.725	6.022	-0.290	-0.052	-0.290	3.253
0.1	0.8	0.420	0.829	-0.060	0.829	6.881	-0.329	-0.060	-0.329	3.713
0.1	0.9	0.472	0.932	-0.067	0.932	7.741	-0.369	-0.067	-0.369	4.174
0.2	0.3	0.157	0.311	-0.024	0.311	2.587	-0.141	-0.024	-0.141	1.440
0.2	0.4	0.210	0.415	-0.031	0.415	3.447	-0.181	-0.031	-0.181	1.901
0.2	0.5	0.262	0.518	-0.038	0.518	4.306	-0.221	-0.038	-0.221	2.361
0.2	0.6	0.315	0.622	-0.046	0.622	5.166	-0.261	-0.046	-0.261	2.822
0.2	0.7	0.367	0.725	-0.053	0.725	6.026	-0.300	-0.053	-0.300	3.282
0.2	0.8	0.420	0.829	-0.061	0.829	6.885	-0.340	-0.061	-0.340	3.743
0.2	0.9	0.472	0.932	-0.068	0.932	7.745	-0.380	-0.068	-0.380	4.203
0.3	0.4	0.210	0.415	-0.032	0.415	3.451	-0.192	-0.032	-0.192	1.930
0.3	0.5	0.262	0.519	-0.039	0.519	4.310	-0.232	-0.039	-0.232	2.390
0.3	0.6	0.315	0.622	-0.046	0.622	5.170	-0.271	-0.046	-0.271	2.851
0.3	0.7	0.367	0.726	-0.054	0.726	6.030	-0.311	-0.054	-0.311	3.311
0.3	0.8	0.420	0.829	-0.061	0.829	6.889	-0.351	-0.061	-0.351	3.772
0.3	0.9	0.472	0.933	-0.069	0.933	7.749	-0.391	-0.069	-0.391	4.232
0.4	0.5	0.262	0.519	-0.040	0.519	4.314	-0.242	-0.040	-0.242	2.420
0.4	0.6	0.315	0.622	-0.047	0.622	5.174	-0.282	-0.047	-0.282	2.880
0.4	0.7	0.367	0.726	-0.055	0.726	6.034	-0.322	-0.055	-0.322	3.341
0.4	0.8	0.420	0.829	-0.062	0.829	6.893	-0.362	-0.062	-0.362	3.801
0.4	0.9	0.472	0.933	-0.069	0.933	7.753	-0.402	-0.069	-0.402	4.262
0.5	0.6	0.315	0.623	-0.048	0.623	5.178	-0.293	-0.048	-0.293	2.909
0.5	0.7	0.367	0.726	-0.055	0.726	6.038	-0.333	-0.055	-0.333	3.370
0.5	0.8	0.420	0.830	-0.063	0.830	6.897	-0.373	-0.063	-0.373	3.830
0.5	0.9	0.472	0.933	-0.070	0.933	7.757	-0.412	-0.070	-0.412	4.291
0.6	0.7	0.367	0.726	-0.056	0.726	6.042	-0.344	-0.056	-0.344	3.399
0.6	0.8	0.420	0.830	-0.063	0.830	6.901	-0.383	-0.063	-0.383	3.860
0.6	0.9	0.472	0.934	-0.071	0.934	7.761	-0.423	-0.071	-0.423	4.320
0.7	0.8	0.420	0.830	-0.064	0.830	6.905	-0.394	-0.064	-0.394	3.889
0.7	0.9	0.472	0.934	-0.071	0.934	7.765	-0.434	-0.071	-0.434	4.349
0.8	0.9	0.472	0.934	-0.072	0.934	7.769	-0.445	-0.072	-0.445	4.379

Table 4.4.7: Variance-Covariance matrices of Z. Ahmed, et al.(2010) estimator for various values of θ_1 and θ_2 for Population 12

θ_1	θ_2	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
0.1	0.2	0.124	0.181	-0.031	0.181	2.689	0.189	-0.031	0.189	2.396
0.1	0.3	0.177	0.204	-0.130	0.204	3.549	-0.302	-0.130	-0.302	2.856
0.1	0.4	0.229	0.226	-0.228	0.226	4.409	-0.793	-0.228	-0.793	3.317
0.1	0.5	0.281	0.249	-0.326	0.249	5.268	-1.284	-0.326	-1.284	3.778
0.1	0.6	0.334	0.272	-0.425	0.272	6.128	-1.775	-0.425	-1.775	4.238
0.1	0.7	0.386	0.294	-0.523	0.294	6.988	-2.266	-0.523	-2.266	4.699
0.1	0.8	0.439	0.317	-0.621	0.317	7.847	-2.757	-0.621	-2.757	5.159
0.1	0.9	0.491	0.339	-0.719	0.339	8.707	-3.249	-0.719	-3.249	5.620
0.2	0.3	0.196	0.340	0.036	0.340	4.519	0.869	0.036	0.869	4.331
0.2	0.4	0.248	0.362	-0.063	0.362	5.379	0.378	-0.063	0.378	4.792
0.2	0.5	0.301	0.385	-0.161	0.385	6.238	-0.113	-0.161	-0.113	5.252
0.2	0.6	0.353	0.408	-0.259	0.408	7.098	-0.605	-0.259	-0.605	5.713
0.2	0.7	0.406	0.430	-0.357	0.430	7.958	-1.096	-0.357	-1.096	6.173
0.2	0.8	0.458	0.453	-0.456	0.453	8.817	-1.587	-0.456	-1.587	6.634
0.2	0.9	0.510	0.475	-0.554	0.475	9.677	-2.078	-0.554	-2.078	7.095
0.3	0.4	0.267	0.498	0.103	0.498	6.349	1.548	0.103	1.548	6.267
0.3	0.5	0.320	0.521	0.005	0.521	7.208	1.057	0.005	1.057	6.727
0.3	0.6	0.372	0.544	-0.094	0.544	8.068	0.566	-0.094	0.566	7.188
0.3	0.7	0.425	0.566	-0.192	0.566	8.928	0.075	-0.192	0.075	7.648
0.3	0.8	0.477	0.589	-0.290	0.589	9.787	-0.416	-0.290	-0.416	8.109
0.3	0.9	0.530	0.611	-0.389	0.611	10.647	-0.907	-0.389	-0.907	8.569
0.4	0.5	0.339	0.657	0.170	0.657	8.178	2.228	0.170	2.228	8.202
0.4	0.6	0.391	0.680	0.072	0.680	9.038	1.737	0.072	1.737	8.663
0.4	0.7	0.444	0.702	-0.027	0.702	9.898	1.246	-0.027	1.246	9.123
0.4	0.8	0.496	0.725	-0.125	0.725	10.757	0.755	-0.125	0.755	9.584
0.4	0.9	0.549	0.747	-0.223	0.747	11.617	0.264	-0.223	0.264	10.044
0.5	0.6	0.410	0.816	0.237	0.816	10.008	2.908	0.237	2.908	10.138
0.5	0.7	0.463	0.838	0.139	0.838	10.867	2.417	0.139	2.417	10.598
0.5	0.8	0.515	0.861	0.040	0.861	11.727	1.926	0.040	1.926	11.059
0.5	0.9	0.568	0.883	-0.058	0.883	12.587	1.435	-0.058	1.435	11.519
0.6	0.7	0.482	0.974	0.304	0.974	11.837	3.588	0.304	3.588	12.073
0.6	0.8	0.534	0.997	0.206	0.997	12.697	3.097	0.206	3.097	12.534
0.6	0.9	0.587	1.019	0.107	1.019	13.557	2.606	0.107	2.606	12.994
0.7	0.8	0.553	1.133	0.371	1.133	13.667	4.268	0.371	4.268	14.009
0.7	0.9	0.606	1.155	0.273	1.155	14.527	3.777	0.273	3.777	14.469
0.8	0.9	0.625	1.291	0.438	1.291	15.497	4.948	0.438	4.948	15.944

Table 4.4.8: Eigen values for matrices given in table 4.4.6

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		λ_1							
θ_2	**0.2**	1.76							
	0.3	2.64	2.64						
	0.4	3.51	3.52	3.53					
	0.5	4.39	4.40	4.40	4.41				
	0.6	5.27	5.27	5.28	5.29	5.29			
	0.7	6.14	6.15	6.16	6.16	6.17	6.18		
	0.8	7.02	7.03	7.03	7.04	7.05	7.05	7.06	
	0.9	7.90	7.90	7.91	7.92	7.92	7.93	7.94	7.94
		λ_2							
θ_2	**0.2**	0.94							
	0.3	1.40	1.42						
	0.4	1.85	1.88	1.91					
	0.5	2.31	2.34	2.36	2.39				
	0.6	2.77	2.79	2.82	2.85	2.87			
	0.7	3.22	3.25	3.28	3.30	3.33	3.36		
	0.8	3.68	3.71	3.73	3.76	3.79	3.81	3.84	
	0.9	4.14	4.16	4.19	4.22	4.24	4.27	4.30	4.32
		λ_3							
θ_2	**0.2**	0.08							
	0.3	0.12	0.12						
	0.4	0.16	0.16	0.16					
	0.5	0.20	0.20	0.20	0.20				
	0.6	0.24	0.24	0.24	0.24	0.24			
	0.7	0.28	0.28	0.28	0.28	0.28	0.28		
	0.8	0.32	0.32	0.32	0.32	0.32	0.32	0.32	
	0.9	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35

Table 4.4.9: Eigen values for matrices given in table 4.4.7

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		μ_1							
	0.2	2.79							
	0.3	3.68	5.31						
	0.4	4.85	5.58	7.88					
θ_2	**0.5**	6.03	6.28	8.07	10.45				
	0.6	7.22	7.36	8.37	10.63	13.03			
	0.7	8.42	8.51	8.97	10.84	13.19	15.60		
	0.8	9.61	9.69	9.93	11.16	13.38	15.76	18.17	
	0.9	10.80	10.88	11.03	11.70	13.62	15.94	18.33	20.75
		μ_2							
	0.2	2.31							
	0.3	2.74	3.56						
	0.4	2.90	4.62	4.78					
θ_2	**0.5**	3.05	5.24	5.90	5.98				
	0.6	3.19	5.49	6.93	7.13	7.19			
	0.7	3.33	5.66	7.65	8.23	8.34	8.40		
	0.8	3.48	5.81	8.01	9.24	9.47	9.55	9.60	
	0.9	3.62	5.95	8.23	10.01	10.55	10.69	10.76	10.81
		μ_3							
	0.2	0.11							
	0.3	0.16	0.17						
	0.4	0.21	0.22	0.23					
	0.5	0.25	0.27	0.28	0.28				
θ_2	**0.6**	0.29	0.32	0.33	0.34	0.34			
	0.7	0.32	0.37	0.38	0.39	0.40	0.40		
	0.8	0.36	0.41	0.43	0.44	0.45	0.45	0.46	
	0.9	0.39	0.45	0.48	0.49	0.50	0.51	0.51	0.51

Table 4.4.10: Relative Efficiencies of the proposed estimator over Z. Ahmed, et al.(2010) estimator

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		$\frac{\sum \lambda_i}{\sum \mu_i}$							
	0.2	0.53							
	0.3	0.63	0.46						
	0.4	0.69	0.53	0.43					
θ_2	**0.5**	0.74	0.59	0.49	0.42				
	0.6	0.77	0.63	0.53	0.46	0.41			
	0.7	0.80	0.67	0.57	0.50	0.45	0.40		
	0.8	0.82	0.69	0.60	0.53	0.48	0.43	0.40	
	0.9	0.84	0.72	0.63	0.56	0.51	0.46	0.43	0.39

Table 4.4.10 shows efficiency comparison of proposed multivariate estimator with estimator proposed by Z. Ahmed, et al.(2010). Sum of Eigen values of covariance matrices is used to calculate relative efficiencies. The above entries clearly indicate that the proposed estimator is more efficient as compared with the estimator proposed by Z. Ahmed, et al.(2010) for all combinations of θ_1 and θ_2.

Table 4.4.11: Variance-Covariance matrices of proposed estimator for various values of θ_1 and θ_2 for Population 13

θ_1	θ_2	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
0.1	0.2	0.076	0.014	0.017	0.014	0.099	0.032	0.017	0.032	0.128
0.1	0.3	0.114	0.021	0.025	0.021	0.149	0.048	0.025	0.048	0.191
0.1	0.4	0.151	0.029	0.033	0.029	0.198	0.065	0.033	0.065	0.254
0.1	0.5	0.189	0.036	0.040	0.036	0.248	0.081	0.040	0.081	0.317
0.1	0.6	0.226	0.043	0.048	0.043	0.297	0.097	0.048	0.097	0.380
0.1	0.7	0.264	0.050	0.056	0.050	0.347	0.113	0.056	0.113	0.442
0.1	0.8	0.301	0.057	0.064	0.057	0.397	0.130	0.064	0.130	0.505
0.1	0.9	0.339	0.065	0.072	0.065	0.446	0.146	0.072	0.146	0.568
0.2	0.3	0.115	0.021	0.026	0.021	0.149	0.048	0.026	0.048	0.192
0.2	0.4	0.152	0.028	0.034	0.028	0.198	0.064	0.034	0.064	0.255
0.2	0.5	0.190	0.036	0.042	0.036	0.248	0.081	0.042	0.081	0.318
0.2	0.6	0.227	0.043	0.050	0.043	0.297	0.097	0.050	0.097	0.381
0.2	0.7	0.265	0.050	0.057	0.050	0.347	0.113	0.057	0.113	0.444
0.2	0.8	0.302	0.057	0.065	0.057	0.397	0.129	0.065	0.129	0.507
0.2	0.9	0.340	0.064	0.073	0.064	0.446	0.146	0.073	0.146	0.570
0.3	0.4	0.153	0.028	0.035	0.028	0.198	0.064	0.035	0.064	0.257
0.3	0.5	0.191	0.035	0.043	0.035	0.248	0.080	0.043	0.080	0.320
0.3	0.6	0.229	0.043	0.051	0.043	0.297	0.097	0.051	0.097	0.383
0.3	0.7	0.266	0.050	0.059	0.050	0.347	0.113	0.059	0.113	0.446
0.3	0.8	0.304	0.057	0.067	0.057	0.397	0.129	0.067	0.129	0.509
0.3	0.9	0.341	0.064	0.074	0.064	0.446	0.145	0.074	0.145	0.572
0.4	0.5	0.192	0.035	0.045	0.035	0.248	0.080	0.045	0.080	0.322
0.4	0.6	0.230	0.043	0.052	0.043	0.297	0.096	0.052	0.096	0.384
0.4	0.7	0.267	0.050	0.060	0.050	0.347	0.113	0.060	0.113	0.447
0.4	0.8	0.305	0.057	0.068	0.057	0.397	0.129	0.068	0.129	0.510
0.4	0.9	0.342	0.064	0.076	0.064	0.446	0.145	0.076	0.145	0.573
0.5	0.6	0.231	0.042	0.054	0.042	0.298	0.096	0.054	0.096	0.386
0.5	0.7	0.268	0.050	0.061	0.050	0.347	0.112	0.061	0.112	0.449
0.5	0.8	0.306	0.057	0.069	0.057	0.397	0.129	0.069	0.129	0.512
0.5	0.9	0.343	0.064	0.077	0.064	0.446	0.145	0.077	0.145	0.575
0.6	0.7	0.269	0.049	0.063	0.049	0.347	0.112	0.063	0.112	0.451
0.6	0.8	0.307	0.057	0.071	0.057	0.397	0.128	0.071	0.128	0.514
0.6	0.9	0.344	0.064	0.078	0.064	0.446	0.145	0.078	0.145	0.577
0.7	0.8	0.308	0.056	0.072	0.056	0.397	0.128	0.072	0.128	0.515
0.7	0.9	0.346	0.064	0.080	0.064	0.446	0.144	0.080	0.144	0.578
0.8	0.9	0.347	0.063	0.081	0.063	0.446	0.144	0.081	0.144	0.580

Table 4.4.12: Variance-Covariance matrices of Z. Ahmed, et al.(2010) estimator for various values of θ_1 and θ_2 for Population 13

θ_1	θ_2	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
0.1	0.2	0.082	0.002	0.001	0.002	0.099	0.009	0.001	0.009	0.131
0.1	0.3	0.120	0.002	-0.001	0.002	0.149	0.013	-0.001	0.013	0.194
0.1	0.4	0.157	0.003	-0.004	0.003	0.198	0.018	-0.004	0.018	0.257
0.1	0.5	0.195	0.003	-0.006	0.003	0.248	0.022	-0.006	0.022	0.320
0.1	0.6	0.233	0.004	-0.009	0.004	0.298	0.026	-0.009	0.026	0.383
0.1	0.7	0.270	0.004	-0.011	0.004	0.347	0.031	-0.011	0.031	0.445
0.1	0.8	0.308	0.005	-0.014	0.005	0.397	0.035	-0.014	0.035	0.508
0.1	0.9	0.345	0.005	-0.016	0.005	0.446	0.039	-0.016	0.039	0.571
0.2	0.3	0.127	0.003	0.004	0.003	0.149	0.014	0.004	0.014	0.198
0.2	0.4	0.165	0.004	0.002	0.004	0.199	0.018	0.002	0.018	0.261
0.2	0.5	0.202	0.004	-0.001	0.004	0.248	0.023	-0.001	0.023	0.324
0.2	0.6	0.240	0.005	-0.003	0.005	0.298	0.027	-0.003	0.027	0.387
0.2	0.7	0.277	0.005	-0.005	0.005	0.347	0.031	-0.005	0.031	0.450
0.2	0.8	0.315	0.006	-0.008	0.006	0.397	0.036	-0.008	0.036	0.513
0.2	0.9	0.352	0.006	-0.010	0.006	0.446	0.040	-0.010	0.040	0.576
0.3	0.4	0.172	0.005	0.008	0.005	0.199	0.019	0.008	0.019	0.266
0.3	0.5	0.210	0.005	0.005	0.005	0.248	0.023	0.005	0.023	0.329
0.3	0.6	0.247	0.006	0.003	0.006	0.298	0.027	0.003	0.027	0.392
0.3	0.7	0.285	0.006	0.000	0.006	0.348	0.032	0.000	0.032	0.455
0.3	0.8	0.322	0.007	-0.002	0.007	0.397	0.036	-0.002	0.036	0.518
0.3	0.9	0.360	0.007	-0.004	0.007	0.447	0.040	-0.004	0.040	0.581
0.4	0.5	0.217	0.006	0.011	0.006	0.249	0.024	0.011	0.024	0.334
0.4	0.6	0.255	0.007	0.008	0.007	0.298	0.028	0.008	0.028	0.396
0.4	0.7	0.292	0.007	0.006	0.007	0.348	0.032	0.006	0.032	0.459
0.4	0.8	0.330	0.008	0.004	0.008	0.397	0.037	0.004	0.037	0.522
0.4	0.9	0.367	0.008	0.001	0.008	0.447	0.041	0.001	0.041	0.585
0.5	0.6	0.262	0.007	0.014	0.007	0.298	0.028	0.014	0.028	0.401
0.5	0.7	0.300	0.008	0.012	0.008	0.348	0.033	0.012	0.033	0.464
0.5	0.8	0.337	0.008	0.009	0.008	0.398	0.037	0.009	0.037	0.527
0.5	0.9	0.375	0.009	0.007	0.009	0.447	0.041	0.007	0.041	0.590
0.6	0.7	0.307	0.009	0.017	0.009	0.348	0.033	0.017	0.033	0.469
0.6	0.8	0.345	0.009	0.015	0.009	0.398	0.038	0.015	0.038	0.532
0.6	0.9	0.382	0.010	0.013	0.010	0.447	0.042	0.013	0.042	0.595
0.7	0.8	0.352	0.010	0.021	0.010	0.398	0.038	0.021	0.038	0.536
0.7	0.9	0.390	0.011	0.018	0.011	0.448	0.042	0.018	0.042	0.599
0.8	0.9	0.397	0.012	0.024	0.012	0.448	0.043	0.024	0.043	0.604

Table 4.4.13: Eigen values for matrices given in table 4.4.11

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		λ_1							
	0.2	0.080							
	0.3	0.081	0.158						
	0.4	0.083	0.160	0.236					
θ_2	**0.5**	0.085	0.161	0.238	0.314				
	0.6	0.087	0.163	0.239	0.316	0.392			
	0.7	0.089	0.165	0.241	0.317	0.394	0.470		
	0.8	0.091	0.167	0.243	0.319	0.396	0.472	0.549	
	0.9	0.093	0.168	0.244	0.321	0.397	0.474	0.550	0.627
		λ_2							
	0.2	0.040							
	0.3	0.041	0.080						
	0.4	0.041	0.080	0.119					
θ_2	**0.5**	0.042	0.081	0.120	0.159				
	0.6	0.043	0.081	0.120	0.160	0.199			
	0.7	0.044	0.082	0.121	0.160	0.199	0.238		
	0.8	0.044	0.083	0.122	0.161	0.200	0.239	0.278	
	0.9	0.045	0.084	0.122	0.161	0.200	0.239	0.278	0.318
		λ_3							
	0.2	0.036							
	0.3	0.036	0.071						
	0.4	0.037	0.072	0.106					
	0.5	0.037	0.072	0.107	0.141				
θ_2	**0.6**	0.037	0.073	0.107	0.142	0.176			
	0.7	0.037	0.073	0.108	0.143	0.177	0.211		
	0.8	0.037	0.073	0.109	0.143	0.178	0.212	0.246	
	0.9	0.037	0.074	0.109	0.144	0.178	0.213	0.247	0.282

Table 4.4.14: Eigen values for matrices given in table 4.4.12

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		μ_1							
θ_2	**0.2**	0.077							
	0.3	0.087	0.146						
	0.4	0.097	0.154	0.215					
	0.5	0.108	0.163	0.223	0.284				
	0.6	0.120	0.173	0.231	0.292	0.353			
	0.7	0.132	0.184	0.240	0.300	0.361	0.423		
	0.8	0.143	0.195	0.250	0.308	0.368	0.430	0.492	
	0.9	0.155	0.206	0.260	0.317	0.377	0.437	0.499	0.561
		μ_2							
θ_2	**0.2**	0.049							
	0.3	0.051	0.097						
	0.4	0.053	0.098	0.146					
	0.5	0.054	0.101	0.146	0.195				
	0.6	0.054	0.103	0.147	0.195	0.243			
	0.7	0.055	0.104	0.150	0.195	0.244	0.292		
	0.8	0.055	0.105	0.152	0.196	0.244	0.292	0.341	
	0.9	0.056	0.106	0.154	0.198	0.244	0.292	0.341	0.389
		μ_3							
θ_2	**0.2**	0.049							
	0.3	0.049	0.094						
	0.4	0.049	0.097	0.138					
	0.5	0.049	0.098	0.143	0.183				
	0.6	0.049	0.098	0.146	0.187	0.227			
	0.7	0.050	0.098	0.146	0.191	0.232	0.272		
	0.8	0.050	0.098	0.147	0.194	0.236	0.277	0.316	
	0.9	0.050	0.098	0.147	0.195	0.240	0.281	0.321	0.360

Table 4.4.15: Relative Efficiencies of the proposed estimator over Z. Ahmed, et al.(2010) estimator

		θ_1							
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
		$\frac{\sum \lambda_i}{\sum \mu_i}$							
θ_2	**0.2**	0.892							
	0.3	0.848	0.916						
	0.4	0.810	0.892	0.924					
	0.5	0.777	0.869	0.908	0.929				
	0.6	0.747	0.848	0.892	0.916	0.931			
	0.7	0.720	0.829	0.877	0.904	0.921	0.933		
	0.8	0.695	0.810	0.862	0.892	0.911	0.924	0.934	
	0.9	0.674	0.793	0.848	0.880	0.901	0.916	0.927	0.935

Table 4.4.15 shows efficiency comparison of proposed multivariate estimator with estimator proposed by Z. Ahmed, et al.(2010). Sum of Eigen values of covariance matrices is used to calculate relative efficiencies. The above entries clearly indicate that the proposed estimator is more efficient as compared with the estimator proposed by Z. Ahmed, et al.(2010) for all combinations of θ_1 and θ_2.

Chapter 5: SURE Estimators in Survey Sampling

5.1 Introduction

Let $Y_1, Y_2, ..., Y_p$ be set of variables to be estimated. Multivariate estimators have been proposed by Ahmad et al. (2010) for simultaneous estimation of several study variables with the exception that all variables depend upon same auxiliary variable *X*. This situation is not always feasible as different response variables may depend on different predictors. In this situation different estimation mechanisms are required. The seemingly unrelated regression models of Zelner (1962) have been popular models for simultaneous prediction of multiple response variables which depends on different set of predictors. The models can be used in survey sampling for simultaneous estimation of multiple response variables which depends on different predictors. These estimators will be called Seemingly Unrelated Regression Estimators (SURE) and are proposed in the following.

5.2 SURE Estimator for Single Phase Sampling using Single Auxiliary Variable

As before, suppose a random sample of *n* observations is drawn from *j*th population and information on study variable Y_j and auxiliary variable X_j is recorded. The regression type estimator for estimation of mean of *j*th population $\bar{Y}_j$ is:

$$t_{SURE-1_j(1)} = \bar{y}_{(j)} + \beta_j \left(\bar{X}_j - \bar{x}_{(j)} \right) \tag{5.2.1}$$

It can be readily shown that the estimator $t_{j(1)}$ is unbiased. The expression for mean square error of $t_{SURE-1_j(1)}$ is derived below.

Writing $\bar{y}_{(j)} = \bar{Y}_j + \bar{e}_{y_{(j)}}$ and $\bar{x}_{(j)} = \bar{X}_j + \bar{e}_{x_{(j)}}$, equation (5.2.1) can be written as:

$$t_{SURE-1_j(1)} - \bar{Y}_j = \bar{e}_{y_{(j)}} - \beta_j \bar{e}_{x_{(j)}}$$

The mean square error is:

$$MSE\left(t_{SURE-1_j}(1)\right)=E\left(t_{j_{(1)}}-\overline{Y}_j\right)^2$$

$$=E\left(\overline{e}_{y_{(j)}}-\beta_j\overline{e}_{x_{(j)}}\right)^2$$

$$=E\left(\overline{e}^2_{y_{(j)}}+\beta_j^2\overline{e}^2_{x_{(j)}}-2\beta_j\overline{e}_{y_{(j)}}\overline{e}_{x_{(j)}}\right)$$

$$=\theta_1 S^2_{y_{(j)}}+\theta_1\beta_j^2 S^2_{x_{(j)}}-2\theta_1\beta_j S_{y_{(j)}x_{(j)}}$$

Using optimum value of $\beta_j=S_{y_{(j)}x_{(j)}}\Big/S^2_{x_{(j)}}$, the mean square error is:

$$MSE\left(t_{SURE-1_j}(1)\right)=\theta_1 S^2_{y_{(j)}}+\theta_1\left(S^2_{y_{(j)}x_{(j)}}\Big/S^2_{x_{(j)}}\right)-2\theta_1\left(S^2_{y_{(j)}x_{(j)}}\Big/S^2_{x_{(j)}}\right)$$

$$=\theta_1 S^2_{y_{(j)}}-\theta_1\left(S^2_{y_{(j)}x_{(j)}}\Big/S^2_{x_{(j)}}\right)$$

$$MSE\left(t_{SURE-1_j}(1)\right)=\theta_1 S^2_{y_{(j)}}\left(1-\rho^2_{y_{(j)}x_{(j)}}\right) \tag{5.2.2}$$

The covariance between two SUE's is:

$$Cov\left(t_{SURE-1_j}(1),t_{SURE-1_k}(1)\right)=E\left(t_{j(1)}-\overline{Y}_j\right)\left(t_{k(1)}-\overline{Y}_k\right)$$

$$=E\left[\left(\overline{e}_{y_{(j)}}-\beta_j\overline{e}_{x_{(j)}}\right)\left(\overline{e}_{y_{(k)}}-\beta_k\overline{e}_{x_{(k)}}\right)\right]$$

$$=E\left[\overline{e}_{y_{(j)}}\overline{e}_{y_{(k)}}-\beta_j\overline{e}_{y_{(k)}}\overline{e}_{x_{(j)}}-\beta_k\overline{e}_{y_{(j)}}\overline{e}_{x_{(k)}}+\beta_j\overline{e}_{x_{(j)}}\overline{e}_{x_{(k)}}\beta_k\right]$$

$$=\theta_1 S_{y_{(j)}y_{(k)}}-\theta_1\beta_j S_{y_{(k)}x_{(j)}}-\theta_1\beta_k S_{y_{(j)}x_{(k)}}+\theta_1\beta_j S_{x_{(j)}x_{(k)}}\beta_k$$

Using $\beta_j=\rho_{y_{(j)}x_{(j)}}\left(S_{y_{(j)}}\Big/S_{x_{(j)}}\right)$, the covariance is:

$$Cov\left(t_{SURE-1_j}(1),t_{SURE-1_k}(1)\right)=\theta_1\rho_{y_{(j)}y_{(k)}}S_{y_{(j)}}S_{y_{(k)}}$$

$$-\theta_1\rho_{y_{(j)}x_{(j)}}\left(S_{y_{(j)}}\Big/S_{x_{(j)}}\right)S_{y_{(k)}x_{(j)}}$$

$$-\theta_1\rho_{y_{(k)}x_{(k)}}\left(S_{y_{(k)}}\Big/S_{x_{(k)}}\right)S_{y_{(j)}x_{(k)}}$$

$$+\theta_1\rho_{y_{(j)}x_{(j)}}\left(S_{y_{(j)}}\Big/S_{x_{(j)}}\right)S_{x_{(j)}x_{(k)}}\rho_{y_{(k)}x_{(k)}}\left(S_{y_{(k)}}\Big/S_{x_{(k)}}\right)$$

$$Cov\left(t_{SURE-1_j}(1), t_{SURE-1_k}(1)\right) = \theta_1 \rho_{y_{(j)}y_{(k)}} S_{y_{(j)}} S_{y_{(k)}} - \theta_1 \rho_{y_{(j)}x_{(j)}} \rho_{y_{(k)}x_{(j)}} S_{y_{(j)}} S_{y_{(k)}}$$
$$-\theta_1 \rho_{y_{(k)}x_{(k)}} \rho_{y_{(j)}x_{(k)}} S_{y_{(j)}} S_{y_{(k)}}$$
$$+\theta_1 \rho_{y_{(j)}x_{(j)}} \rho_{y_{(k)}x_{(k)}} \rho_{x_{(j)}x_{(k)}} S_{y_{(j)}} S_{y_{(k)}}$$

$$Cov\left(t_{SURE-1_j}(1), t_{SURE-1_k}(1)\right) = \theta_1 S_{y_{(j)}} S_{y_{(k)}} \Big[\rho_{y_{(j)}y_{(k)}} - \rho_{y_{(j)}x_{(j)}} \rho_{y_{(k)}x_{(j)}} - \rho_{y_{(k)}x_{(k)}} \rho_{y_{(j)}x_{(k)}} + \rho_{y_{(j)}x_{(j)}} \rho_{y_{(k)}x_{(k)}} \rho_{x_{(j)}x_{(k)}} \Big] \quad (5.2.3)$$

Equations (5.2) and (5.3) can be used to compute mean square error for any estimator and covariance between any two estimators.

5.3 SURE Estimator for Single Phase Sampling using Multiple Auxiliary Variables

The SURE's in single phase sampling using multiple predictors can be given on the above lines. Specifically the SURE in single phase sampling with multiple predictors is:

$$t_{SURE-2_j}(1) = \bar{y}_{(j)} + \boldsymbol{\beta}_j^{/}\left(\bar{\mathbf{X}}_j - \bar{\mathbf{x}}_{(j)}\right) \quad (5.3.1)$$

The mean square error for *i*th estimator can be readily written as:

$$MSE\left(t_{SURE-2_j}(1)\right) = \theta S_j^2 \left(1 - \rho^2_{y.\mathbf{x}_j}\right) \quad (5.3.2)$$

The covariance between two estimators based upon different set of predictors is derived below:

$$Cov\left(t_{SURE-2_j}(1), t_{SURE-2_k}(1)\right) = E\left[\left(t_{SURE-2_j}(1) - \bar{Y}_j\right)\left(t_{SURE-2_k}(1) - \bar{Y}_k\right)\right] \quad (5.3.3)$$

Consider $t_j = \bar{y}_j + \boldsymbol{\beta}_j^{/}\left(\bar{\mathbf{X}}_j - \bar{\mathbf{x}}_j\right)$

Using $\bar{y}_j = \bar{Y}_j + \bar{\mathbf{e}}_{y_j}$; $\bar{x}_j = \bar{X}_j + \bar{\mathbf{e}}_{x_j}$ We have

$$t_j = \bar{Y}_j + \bar{e}_{y_j} + \boldsymbol{\beta}_j^{/}\left(\bar{\mathbf{X}}_j - \bar{\mathbf{X}}_j - \bar{\mathbf{e}}_{x_j}\right)$$

So $t_{SURE-2_j(1)} - \bar{Y}_j = \bar{e}_{y_j} - \boldsymbol{\beta}_j^{/}\bar{\mathbf{e}}_{x_j}$ (5.3.4)

And $t_{SURE-2_j(1)} - \bar{Y}_k = \bar{e}_{y_k} - \boldsymbol{\beta}_k^{/}\bar{\mathbf{e}}_{x_k}$ (5.3.5)

By using (5.3.4) and (5.3.5) in (5.3.3)

$$Cov\left(t_{SURE-2_j(1)}, t_{SURE-2_k(1)}\right) = E\left[\left(\bar{e}_{y_j} - \boldsymbol{\beta}_j^{/}\bar{\mathbf{e}}_{x_j}\right)\left(\bar{e}_{y_k} - \boldsymbol{\beta}_k^{/}\bar{\mathbf{e}}_{x_k}\right)\right]$$
$$= E\left[\bar{e}_{y_j}\bar{e}_{y_k} - \boldsymbol{\beta}_j^{/}\bar{e}_{y_k}\bar{\mathbf{e}}_{x_j} - \boldsymbol{\beta}_k^{/}\bar{e}_{y_j}\bar{\mathbf{e}}_{x_k} + \boldsymbol{\beta}_j^{/}\bar{\mathbf{e}}_{x_j}\bar{\mathbf{e}}_{x_k}^{/}\boldsymbol{\beta}_k\right]$$

Appling expectation we have:

$$Cov\left(t_{SURE-2_j(1)}, t_{SURE-2_k(1)}\right) = \theta\mathbf{S}_{y_j y_k} - \theta\boldsymbol{\beta}_j^{/}\mathbf{S}_{y_k\mathbf{x}_j} - \theta\boldsymbol{\beta}_k^{/}\mathbf{S}_{y_j\mathbf{x}_k} + \theta\boldsymbol{\beta}_j^{/}\mathbf{S}_{jk}\boldsymbol{\beta}_k$$

Using $\boldsymbol{\beta}_j = \mathbf{S}_j^{-1}\mathbf{S}y_j\mathbf{x}_j$ *and* $\boldsymbol{\beta}_k = \mathbf{S}_k^{-1}\mathbf{S}y_k\mathbf{x}_k$

$$Cov\left(t_{SURE-2_j(1)}, t_{SURE-2_k(1)}\right) = \theta\mathbf{S}_{y_j y_k} - \theta\mathbf{S}_{y_j\mathbf{x}_j}^{/}\mathbf{S}_j^{-1}\mathbf{S}_{y_k\mathbf{x}_j} - \theta\mathbf{S}_{y_k\mathbf{x}_k}^{/}\mathbf{S}_k^{-1}\mathbf{S}_{y_j\mathbf{x}_k}$$
$$+\theta\mathbf{S}_{y_j\mathbf{x}_j}^{/}\mathbf{S}_j^{-1}\mathbf{S}_{jk}\mathbf{S}_k^{-1}\mathbf{S}y_k\mathbf{x}_k$$
$$= \theta\left[\mathbf{S}_{y_j y_k} - \mathbf{S}_{y_j\mathbf{x}_j}^{/}\mathbf{S}_j^{-1}\mathbf{S}_{y_k\mathbf{x}_j} - \mathbf{S}_{y_k\mathbf{x}_k}^{/}\mathbf{S}_k^{-1}\mathbf{S}_{y_j\mathbf{x}_k}\right.$$
$$\left.+\mathbf{S}_{y_j\mathbf{x}_j}^{/}\mathbf{S}_j^{-1}\mathbf{S}_{jk}\mathbf{S}_k^{-1}\mathbf{S}y_k\mathbf{x}_k\right]$$

5.4 SURE Estimator for Two-phase Sampling using Single Auxiliary Variable

The seemingly unrelated estimators when a two phase sample is available can be constructed following the method of Zellner (1962). The proposed set of estimators has wider applicability as compared with estimators proposed by Ahmad et al. (2010). The SUE in two phase sampling using single predictor is:

$$t_{SURE-1_j(2)} = \bar{y}_{2(j)} + \beta_j\left(\bar{x}_{1(j)} - \bar{x}_{2(j)}\right) \quad (5.4.1)$$

It can be readily shown that the estimator $t_{j(1)}$ is unbiased. The expression for mean square error of $t_{SURE-1_j(2)}$ is derived below.

Writing $\bar{y}_{(j)} = \bar{Y}_j + \bar{e}_{y_{(j)}}$, $\bar{x}_{1(j)} = \bar{X}_j + \bar{e}_{x_{1(j)}}$ and $\bar{x}_{2(j)} = \bar{X}_j + \bar{e}_{x_{2(j)}}$, equation (5.4.1) can be written as:

$$t_{SURE-1_j(2)} - \bar{Y}_j = \bar{e}_{y_{(j)}} - \beta_j \left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}} \right)$$

The mean square error is:

$$\begin{aligned}
MSE\left(t_{SURE-1_j(2)}\right) &= E\left(t_{j(2)} - \bar{Y}_j\right)^2 \\
&= E\left\{\bar{e}_{y_{2(j)}} + \beta_j \left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right)\right\}^2 \\
&= E\left\{\bar{e}^2_{y_{2(j)}} + \beta_j^2 \left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right)^2 + 2\beta_j \bar{e}_{y_{2(j)}} \left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right)\right\} \\
&= \theta_2 S^2_{y_{(j)}} + (\theta_2 - \theta_1)\beta_j^2 S^2_{x_{(j)}} - 2(\theta_2 - \theta_1)\beta_j S_{y_{(j)}x_{(j)}} \\
&= \theta_2 S^2_{y_{(j)}} - (\theta_2 - \theta_1)\beta_j^2 S^2_{x_{(j)}}
\end{aligned}$$

Using optimum value of $\beta_j = S_{y_{(j)}x_{(j)}} \big/ S^2_{x_{(j)}}$, the mean square error is:

$$\begin{aligned}
MSE\left(t_{SURE-1_j(2)}\right) &= \theta_2 S^2_{y_{(j)}} - (\theta_2 - \theta_1)\left(S^2_{y_{(j)}x_{(j)}} \big/ S^2_{x_{(j)}} \right) \\
&= \theta_2 S^2_{y_{(j)}} - \theta_2 \left(S^2_{y_{(j)}x_{(j)}} \big/ S^2_{x_{(j)}} \right) + \theta_1 \left(S^2_{y_{(j)}x_{(j)}} \big/ S^2_{x_{(j)}} \right) \qquad (5.4.2) \\
&= S^2_{y_{(j)}} \left\{ \theta_2 \left(1 - \rho^2_{y_{(j)}x_{(j)}} \right) + \theta_1 \rho^2_{y_{(j)}x_{(j)}} \right\}
\end{aligned}$$

The covariance between two SURE's is:

$$\begin{aligned}
Cov\left(t_{SURE-1_j(2)}, t_{SURE-1_k(2)}\right) &= E\left(t_{SURE-1_j(2)} - \bar{Y}_j\right)\left(t_{SURE-1_k(2)} - \bar{Y}_k\right) \\
&= E\left[\left\{\bar{e}_{y_{2(j)}} + \beta_j \left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right)\right\}\left\{\bar{e}_{y_{2(k)}} + \beta_k \left(\bar{e}_{x_{1(k)}} - \bar{e}_{x_{2(k)}}\right)\right\}\right]
\end{aligned}$$

$$Cov\left(t_{SURE-1_j}(2), t_{SURE-1_k}(2)\right) = E\left[\bar{e}_{y_{2(j)}}\bar{e}_{y_{2(k)}} + \beta_k \bar{e}_{y_{2(j)}}\left(\bar{e}_{x_{1(k)}} - \bar{e}_{x_{2(k)}}\right)\right.$$
$$\left. + \beta_j \bar{e}_{y_{2(k)}}\left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right) + \beta_j \beta_k \left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right)\left(\bar{e}_{x_{1(k)}} - \bar{e}_{x_{2(k)}}\right)\right]$$
$$= \theta_2 S_{y_{(j)}y_{(k)}} - (\theta_2 - \theta_1)\beta_k S_{y_{(j)}x_{(k)}} - (\theta_2 - \theta_1)\beta_j S_{y_{(k)}x_{(j)}}$$
$$+ (\theta_2 - \theta_1)\beta_j S_{x_{(j)}x_{(k)}}\beta_k$$

Using $\beta_j = \rho_{y_{(j)}x_{(j)}}\left(S_{y_{(j)}} / S_{x_{(j)}}\right)$, the covariance is:

$$Cov\left(t_{SURE-1_j}(2), t_{SURE-1_j}(2)\right) = \theta_2 \rho_{y_{(j)}y_{(k)}} S_{y_{(j)}} S_{y_{(k)}} - (\theta_2 - \theta_1)\rho_{y_{(j)}x_{(j)}}\left(S_{y_{(j)}} / S_{x_{(j)}}\right) S_{y_{(k)}x_{(j)}}$$
$$- (\theta_2 - \theta_1)\rho_{y_{(k)}x_{(k)}}\left(S_{y_{(k)}} / S_{x_{(k)}}\right) S_{y_{(j)}x_{(k)}}$$
$$+ (\theta_2 - \theta_1)\rho_{y_{(j)}x_{(j)}}\left(S_{y_{(j)}} / S_{x_{(j)}}\right) S_{x_{(j)}x_{(k)}} \rho_{y_{(k)}x_{(k)}}\left(S_{y_{(k)}} / S_{x_{(k)}}\right)$$

$$= \theta_2 \rho_{y_{(j)}y_{(k)}} S_{y_{(j)}} S_{y_{(k)}} - (\theta_2 - \theta_1)\rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(j)}} S_{y_{(j)}} S_{y_{(k)}}$$
$$- (\theta_2 - \theta_1)\rho_{y_{(k)}x_{(k)}}\rho_{y_{(j)}x_{(k)}} S_{y_{(j)}} S_{y_{(k)}}$$
$$+ (\theta_2 - \theta_1)\rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(k)}}\rho_{x_{(j)}x_{(k)}} S_{y_{(j)}} S_{y_{(k)}}$$

$$Cov\left(t_{SURE-1_j}(2), t_{SURE-1_j}(2)\right) = S_{y_{(j)}} S_{y_{(k)}}\left[\theta_2\left\{\rho_{y_{(j)}y_{(k)}} - \rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(j)}} - \rho_{y_{(k)}x_{(k)}}\rho_{y_{(j)}x_{(k)}}\right.\right.$$
$$\left. + \rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(k)}}\rho_{x_{(j)}x_{(k)}}\right\} + \theta_1\left\{\rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(j)}}\right.$$
$$\left.\left. - \rho_{y_{(k)}x_{(k)}}\rho_{y_{(j)}x_{(k)}}\rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(k)}}\rho_{x_{(j)}x_{(k)}}\right\}\right] \qquad (5.4.3)$$

Equations (5.4.2) and (5.4.3) can be used to compute mean square error for any estimator and covariance between any two estimators.

5.5 SURE Estimator for Two-phase Sampling using Multiple Auxiliary Variables

The SUE's in two-phase sampling using multiple predictors can be given on the above lines as under:

$$t_{SURE-2_j(2)} = \bar{y}_{2(j)} + \boldsymbol{\beta}_j^{/}\left(\bar{\mathbf{x}}_{1(j)} - \bar{\mathbf{x}}_{2(j)}\right) \quad (5.5.1)$$

With $MSE\left(t_{SURE-2_j(2)}\right) = S_y^2\left[\theta_2\left(1-\rho_{y_j.\mathbf{x}_j}^2\right) + \theta_1\rho_{y_j.x_j}^2\right]$ (5.5.2)

$$Cov\left(t_{SURE-2_j(2)}, t_{SURE-2_k(2)}\right) = E\left[\left(t_{SURE-2_j(2)} - \bar{Y}_j\right)\left(t_{SURE-2_k(2)} - \bar{Y}_k\right)\right] \quad (5.5.3)$$

Consider $t_{SURE-2_j(2)} = \bar{y}_{2(j)} + \boldsymbol{\beta}_j^{/}\left(\bar{x}_{1(j)} - \bar{\mathbf{x}}_{2(j)}\right)$

Using $\bar{y}_{2(j)} = \bar{Y}_j + \bar{e}_{y_{2(j)}}$; $\bar{x}_{1(j)} = \bar{X}_j + \bar{\mathbf{e}}_{x_{1(j)}}$; $\bar{x}_{2(j)} = \bar{X}_j + \bar{\mathbf{e}}_{x_{2(j)}}$ We have

$$t_{SURE-2_j(2)} = \bar{Y}_j + \bar{e}_{y_{2(j)}} + \boldsymbol{\beta}_j^{/}\left(\bar{\mathbf{X}} + \bar{\mathbf{e}}_{x_{1(j)}} - \bar{\mathbf{X}} - \bar{\mathbf{e}}_{x_{2(j)}}\right)$$

So $t_{SURE-2_j(2)} - \bar{Y}_j = \bar{e}_{y_{2(j)}} + \boldsymbol{\beta}_j^{/}\left(\bar{\mathbf{X}} + \bar{e}_{x_{1(j)}} - \bar{\mathbf{X}} - \bar{e}_{x_{2(j)}}\right)$ (5.5.4)

$$t_{SURE-2_k(2)} - \bar{Y}_k = \bar{e}_{y_{2(k)}} + \boldsymbol{\beta}_k^{/}\left(\bar{\mathbf{X}} + \bar{e}_{x_{1(k)}} - \bar{\mathbf{X}} - \bar{e}_{x_{2(k)}}\right) \quad (5.5.5)$$

By using (5.5.4) and (5.5.5) in (5.5.3)

$$Cov\left(t_{SURE-2_j(2)}, t_{SURE-2_k(2)}\right) = E\left[\left\{\bar{e}_{y_{2(j)}} - \boldsymbol{\beta}_j^{/}\left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right)\right\}\left\{\bar{e}_{y_{2(k)}} - \boldsymbol{\beta}_k^{/}\left(\bar{e}_{x_{1(k)}} - \bar{e}_{x_{2(k)}}\right)\right\}\right]$$

$$Cov\left(t_{SURE-2_j(2)}, t_{SURE-2_k(2)}\right) = E\left[\bar{e}_{y_{2(j)}}\bar{e}_{y_{2(k)}} - \boldsymbol{\beta}_j^{/}\left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right)\bar{e}_{y_{2(k)}}\right.$$
$$\left. + \boldsymbol{\beta}_k^{/}\left(\bar{e}_{x_{1(k)}} - \bar{e}_{x_{2(k)}}\right)\bar{e}_{y_{2(j)}} + \boldsymbol{\beta}_j^{/}\left(\bar{e}_{x_{1(j)}} - \bar{e}_{x_{2(j)}}\right)\left(\bar{e}_{x_{1(k)}} - \bar{e}_{x_{2(k)}}\right)\boldsymbol{\beta}_k\right]$$

Appling expectation we have:

$$Cov\left(t_{SURE-2_j}(2), t_{SURE-2_k}(2)\right) = \theta_2 \mathbf{S}_{y_j y_k} - (\theta_2 - \theta_1)\boldsymbol{\beta}_j^{/}\mathbf{S}_{y_k \mathbf{x}_j} - (\theta_2 - \theta_1)\boldsymbol{\beta}_k^{/}\mathbf{S}_{y_j \mathbf{x}_k} + (\theta_2 - \theta_1)\boldsymbol{\beta}_j^{/}\mathbf{S}_{jk}\boldsymbol{\beta}_k$$

$$Cov\left(t_{SURE-2_j}(2), t_{SURE-2_k}(2)\right) = \Big[\theta_2 \left\{S_{y_j y_k} - \boldsymbol{\beta}_j^{/}\mathbf{S}_{y_k \mathbf{x}_j} - \boldsymbol{\beta}_k^{/}\mathbf{S}_{y_j \mathbf{x}_k} + \boldsymbol{\beta}_j^{/}\mathbf{S}_{jk}\boldsymbol{\beta}_k\right\} + \theta_1 \left\{\boldsymbol{\beta}_j^{/}\mathbf{S}_{y_k \mathbf{x}_j} + \boldsymbol{\beta}_k^{/}\mathbf{S}_{y_j \mathbf{x}_k} - \boldsymbol{\beta}_j^{/}\mathbf{S}_{jk}\boldsymbol{\beta}_k\right\}\Big] \tag{5.5.6}$$

Using $\boldsymbol{\beta}_j = \mathbf{S}_j^{-1}\mathbf{S}y_j\mathbf{x}_j$ *and* $\boldsymbol{\beta}_k = \mathbf{S}_k^{-1}\mathbf{S}y_k\mathbf{x}_k$ in (5.5.6):

$$Cov\left(t_{SURE-2_j}(2), t_{SURE-2_j}(2)\right) = \Big[\theta_2 \Big\{\mathbf{S}_{y_j y_k} - \mathbf{S}_{y_j \mathbf{x}_j}^{/}\mathbf{S}_j^{-1}\mathbf{S}_{y_k \mathbf{x}_j} - \mathbf{S}_{y_k \mathbf{x}_k}^{/}\mathbf{S}_k^{-1}\mathbf{S}_{y_j \mathbf{x}_k} + \mathbf{S}_{y_j \mathbf{x}_j}^{/}\mathbf{S}_j^{-1}\mathbf{S}_{jk}\mathbf{S}_k^{-1}\mathbf{S}y_k\mathbf{x}_k\Big\} + \theta_1 \Big\{\mathbf{S}_{y_j \mathbf{x}_j}^{/}\mathbf{S}_j^{-1}\mathbf{S}_{y_k \mathbf{x}_j} - \mathbf{S}_{y_k \mathbf{x}_k}^{/}\mathbf{S}_k^{-1}\mathbf{S}_{y_j \mathbf{x}_k} + \mathbf{S}_{y_j \mathbf{x}_j}^{/}\mathbf{S}_j^{-1}\mathbf{S}_{jk}\mathbf{S}_k^{-1}\mathbf{S}y_k\mathbf{x}_k\Big\}\Big]$$

5.6 SURE Estimator for Multiphase Sampling using Single Auxiliary Variable

The SURE estimator and its mean square error for multiphase sampling can be analogously written from (5.4.1), (5.4.2) and (5.4.3). Specifically if a sample of size n_h is taken at h^{th} phase and a sample of n_q is taken at q^{th} phase with $n_q < n_h$, the estimator of the population mean is:

$$t_{SURE-3_j(2)} = \bar{y}_{2(j)} + \beta_j\left(\bar{x}_{1(j)} - \bar{x}_{2(j)}\right) \tag{5.6.1}$$

The expression of the Multiphase sampling can analogously be written from (5.4.2)

$$MSE\left(t_{SURE-3_j(2)}\right) = S^2_{y_{(j)}}\left\{\theta_q\left(1-\rho^2_{y_{(j)}x_{(j)}}\right) + \theta_h\rho^2_{y_{(j)}x_{(j)}}\right\} \tag{5.6.2}$$

The covariance between two SUE's is:

$$\begin{aligned}Cov\left(t_{SURE-3_j(2)}, t_{SURE-3_k(2)}\right) = S_{y_{(j)}}S_{y_{(k)}}\Big[\theta_q\Big\{&\rho_{y_{(j)}y_{(k)}} - \rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(j)}} - \rho_{y_{(k)}x_{(k)}}\rho_{y_{(j)}x_{(k)}} \\ &+\rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(k)}}\rho_{x_{(j)}x_{(k)}}\Big\} + \theta_h\Big\{\rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(j)}} \\ &-\rho_{y_{(k)}x_{(k)}}\rho_{y_{(j)}x_{(k)}}\rho_{y_{(j)}x_{(j)}}\rho_{y_{(k)}x_{(k)}}\rho_{x_{(j)}x_{(k)}}\Big\}\Big]\end{aligned} \tag{5.6.3}$$

5.7 SURE Estimator for Multiphase Sampling using Multiple Auxiliary Variable

Similarly the SURE in multiphase sampling using multiple predictors can be given on the above lines as under:

$$t_{SURE-4_j(2)} = \bar{y}_{2(j)} + \boldsymbol{\beta}'_j\left(\bar{\mathbf{x}}_{1(j)} - \bar{\mathbf{x}}_{2(j)}\right) \tag{5.7.1}$$

With $MSE\left(t_{SURE-4_j(2)}\right) = S^2_y\left[\theta_q\left(1-\rho^2_{y_j.\mathbf{x}_j}\right) + \theta_h\rho^2_{y_j.x_j}\right]$ (5.7.2)

and

$$Cov\left(t_{SURE-4_j(2)}, t_{SURE-4_k(2)}\right) = \Big[\theta_q \Big\{\mathbf{S}_{y_j y_k} - \mathbf{S}'_{y_j \mathbf{x}_j}\mathbf{S}_j^{-1}\mathbf{S}_{y_k \mathbf{x}_j} - \mathbf{S}'_{y_k \mathbf{x}_k}\mathbf{S}_k^{-1}\mathbf{S}_{y_j \mathbf{x}_k}$$

$$+\mathbf{S}'_{y_j \mathbf{x}_j}\mathbf{S}_j^{-1}\mathbf{S}_{jk}\mathbf{S}_k^{-1}\mathbf{S}y_k\mathbf{x}_k\Big\} + \theta_h \Big\{\mathbf{S}'_{y_j \mathbf{x}_j}\mathbf{S}_j^{-1}\mathbf{S}_{y_k \mathbf{x}_j}$$

$$-\mathbf{S}'_{y_k \mathbf{x}_k}\mathbf{S}_k^{-1}\mathbf{S}_{y_j \mathbf{x}_k} + \mathbf{S}'_{y_j \mathbf{x}_j}\mathbf{S}_j^{-1}\mathbf{S}_{jk}\mathbf{S}_k^{-1}\mathbf{S}y_k\mathbf{x}_k\Big\}\Big] \qquad (5.7.3)$$

Conclusions and Recommendations

Survey statisticians have always been in search of efficient estimations methodologies. The aim in developing these methodologies is to obtain accurate estimates with least information. Non-availability of sampling frame has played key role in development of two phase sampling. Large numbers of estimators have been developed in literature for use under two-phase and multi-phase sampling designs. Some of these popular estimators have been discussed in chapter 1 and 2 of this thesis.

The motivation of this study has been the work of Roy(2003), Z. Ahmed, et al.(2009) and Z. Ahmed, et al.(2010), where these authors have proposed univariate and multivariate estimators. The extension of Roy(2003) estimator has been proposed in chapter 3 of the thesis. The extension has been proposed by using several auxiliary variables as well as attributes. It has been found that Roy(2003) estimator turned out to be special case of $t_{N_1(2)}$ for m = 1. The proposed estimator $t_{N_1(2)}$ has wider applicability when information on several auxiliary variables is available. The extension of Roy(2003) estimator for auxiliary attributes has also been proposed in chapter 3. This extension has been found more efficient as compared with estimators proposed by Shabbir & Gupta(2007). The empirical study of proposed estimators has been conducted for different values of θ_1 and θ_2 to see its performance as compared with classical regression estimator. It has been concluded that the proposed estimators are always more precise as compared with classical regression estimator for both quantitative and qualitative auxiliary variables. It has also been concluded that the efficiency increases for large values of θ_1 and θ_2. It has also been concluded that efficiency is not much higher when difference between θ_1 and θ_2 is large. The shrinkage versions of proposed estimators have also been proposed in chapter 3 following the procedure of Shahbaz & Hanif(2009) and it is concluded that the shrinkage estimators are more efficient as compared with conventional estimators. It is therefore recommended that the proposed estimator should be used when information on multiple auxiliary variables or attributes is available. It is further recommended that the proposed estimator should be used with not much gap between θ_1 and θ_2.

The multivariate version of Roy(2003) estimator has been proposed in chapter 4. The multivariate version has been proposed by using information of several auxiliary variables as well as attributes. It has been concluded that the proposed estimators in chapter 3 turned out to be special case of $\underline{\mathbf{t}}_{N(2)}$ for p=1. The empirical study based upon eigen values of covariance matrix of proposed estimator and multivariate regression estimator of Z. Ahmed, et al.(2010) has also been conducted in chapter 4. From the results of empirical study, it is concluded that the proposed multivariate estimators are more efficient as compared with the estimator proposed by Z. Ahmed, et al.(2010). It is also concluded that efficiency of proposed multivariate estimators increases with increase in the value of θ_1. It is, therefore recommended to use the proposed estimators by using values of θ_1 and θ_2 close to each other.

The multivariate estimators proposed in chapter 4 are limited to the fact that set of same auxiliary variables is to be used for estimation of all variables of interest. This condition has been relaxed in chapter 5 of the thesis where Seemingly Unrelated Regression Estimators (SURE's) have been proposed. It is concluded that applicability of SURE's is much wider as compared with multivariate regression estimators available in literature. It is therefore recommended that the SURE's are to be used when different sets of auxiliary variables are available for estimation of different response variables.

References:

Ahmed, M. S. (1998). A Note on Regression Type Estimators Using Multiple Auxiliary Information. *Australian & New Zealand Journal of Statistics, 40*(3), 373-376.

Ahmed, M. S. (2003). General chain estimators under multi phase sampling. *J. Applied Statist. Science, 12*(4), 243-250.

Ahmed, Z., Hanif, M., & Ahmad, M. (2009). Generalized Regression Cum-Ratio Estimators for Two-Phase Sampling Using Muliti-Auxiliary Variables *Pak. J. Statist, 25*(2), 93-106.

Ahmed, Z., Hussin, A. G., & Hanif, M. (2010). Generalized multivariate regression estimators for multi-phase sampling using multi-auxiliary variables. *Pak. J. Statist, 26*(4), 569-583.

Bowley, A. L. (1906). Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science, York, 1906. *Journal of the Royal Statistical Society, 69*(3), 540-558.

Bowley, A. L. (1913). Working-class households in reading. *Journal of the Royal Statistical Society, 76*(7), 672-701.

Bowley, A. L. (1926). Measurement of the precision attained in sampling. *Bull. Int. Statist. Inst., 22*(1), 6-62.

Chand, L. (1975). *Some ratio-type estimators based on two or more auxiliary variables*: Unpublished Dissertation, Iowa State University.

Chandra, P., & Singh, H. P. (2003). A family of unbiased estimators in two-phase sampling using two auxiliary variables. *Statistics in Transition, 6*(1), 131-141.

Chang, W. C. (1976). *Statistical theories and sampling practice*: Dekker. New York.

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

Cochran, W. G. (1977). *Sampling techniques*. New York: John Wiley & Sons.

Dalbehera, M., & Sahoo, L. N. (2000). An unbiased estimator in two-phase sampling using two-auxiliary variables. *Jour. Ind. Soc. Agri. Statist., 53*(2), 134-140.

Dalenius, T. (1962). Recent advances in sample survey theory and methods. *The Annals of Mathematical Statistics, 33*(2), 325-349.

Diana, G., & Tommasi, C. (2003). On optimal estimation for finite population mean in two-phase sampling. *Statistical Methods and Applications, 12*, 41-48.

Duncan, J. W., & Shelton, W. C. (1978). *Revolution in United States government statistics, 1926-1976*: U.S. Dept. of Commerce, Washington.

Greene, W. H. (2003). *Econometric analysis*: Prentice Hall Upper Saddle River, NJ.

Hanif, M., Ahmed, Z., & Ahmad, M. (2009). Generalized Multivariate Ratio Estimator using Multi-Auxiliary Variables for Multi-Phase Sampling. *Pak. J. Statist, 25*(4), 615-629.

Hanif, M., Haq, I., & Shahbaz, M. Q. (2009). On a new family of estimators using multiple auxiliary attributes. *World Applied Sciences Journal, 7*(11), 1419-1422.

Hanif, M., Haq, I., & Shahbaz, M. Q. (2010). Ratio Estimators using Multiple Auxiliary Attributes. *World Applied Sciences Journal, 8*(1), 133-136.

Hansen, M. H. (1987). Some history and reminiscences on survey sampling. *Statistical Science, 2*(2), 180-190.

Hansen, M. H., Dalenius, T., & Tepping, B. J. (1985). *The development of sample surveys of finite populations*: Springer, New York.

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

Hartley, H. O., & Ross, A. (1954). Unbiased ratio estimators. *Nature, 174*, 207-271.

Hidiroglou, M. A., & Särndal, C. E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology, 24*, 11-20.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association, 47*(260), 663-685.

Jhajj, H. S., Sharma, M. K., & Grover, L. K. (2006). A family of estimators of population mean using information on auxiliary attribute. *Pak J. Statist, 22*(1), 43-50.

Kiaer, A. N. (1895). Observations et experiences concentrant des denombrements resresentatifs. *Amer. Statist. Assoc., 25*(284-294).

Kiaer, A. N. (1897). The Representative Method of Satistical Survery. *Central Bureau of Statistics of norway, Oslo*.

Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika, 27*(1), 217-223.

Kiregyera, B. (1984). Regression-type estimators using two auxiliary variables and the model of double sampling from finite populations. *Metrika, 31*(1), 215-226.

Kish, L. (1965). Survey Sampling. New York: J. *Wiley & Sons*.

Kruskal, W., & Mosteller, F. (1980). Representative sampling, IV: The history of the concept in statistics, 1895-1939. *International Statistical Review, 48*(2), 169-195.

Mohanty, S. (1967). Combination of regression and ratio estimate. *Jour. Ind. Statist. Asso, 5*, 16-19.

Mukerjee, R., Rao, T. J., & Vijayan, K. (1987). Regression type estimators using multiple auxiliary information. *Australian & New Zealand Journal of Statistics, 29*(3), 244-254.

Murthy, M. N. (1967). *Sampling theory and methods*. Calcutta, India: Statistical Publishing Society.

Naik, V. D., & Gupta, P. C. (1996). A note on estimation of mean with known population proportion of an auxiliary character. *Jour. Ind. Soc. Agr. Stat, 48*(2), 151-158.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society, 97*(4), 558-625.

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

Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika, 45*(1/2), 154-165.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika, 43*(3/4), 353-360.

Radhey, B. P., Singh, S., & Singh, H. P. (2002). Modified chain ratio estimators for finite population mean using two auxiliary variables in double sampling. *Statistics in Transition, 5*(6), 1051-1066.

Raj, D. (1968). *Sampling theory*: McGraw-Hill, New York.

Rao, P. S. R. S., & Rao, J. N. K. (1971). Small sample results for ratio estimators. *Biometrika, 58*(3), 625-630.

Roy, D. C. (2003). A regression type estimator in two phase sampling using two auxiliary variables. *Pak J. Statist, 19*(3), 281-290.

Sahai, A. (1979). An efficient variant of the product and ratio estimators. *Statistica Neerlandica, 33*(1), 27-35.

Sahoo, J., & Sahoo, L. N. (1993). A class of estimators in two-phase sampling using two auxiliary variables. *Jour. Ind. Stat. Assoc, 31*, 107-114.

Sahoo, J., & Sahoo, L. N. (1994). On the Efficiency of Four Chain-Type Estimators in Two-Phase Sampling Under a Model. *Statistics, 25*(4), 361-366.

Sahoo, J., & Sahoo, L. N. (1999a). An alternative class of estimators in double sampling procedures. *Bulletin of the Calcutta Statistical Association, 49*(193), 79-84.

Sahoo, J., & Sahoo, L. N. (1999b). A comparative study of some regression -type estimators in double sampling procedures. *Aligarh J. Statist., 19*, 67-76.

Sahoo, J., Sahoo, L. N., & Mohanty, S. (1993). A regression approach to estimation in two-phase sampling using two auxiliary variables. *Current Science, 65*(1), 73-75.

Sahoo, J., Sahoo, L. N., & Mohanty, S. (1994a). A regression approach to estimation in two-phase sampling using two auxiliary variables. *Current Science, 65*(1), 73-75.

Sahoo, J., Sahoo, L. N., & Mohanty, S. (1994b). An alternative approach to estimation in two-phase sampling using two auxiliary variables. *Biometrical J., 3*, 293-298.

Sahoo, L. N., & Sahoo, R. K. (2001). Predictive estimation of finite population mean in two-phase sampling using two auxiliary variables. *J. Ind. Soc. Agri. Statist., 54*(2), 258-264.

Samiuddin, M., & Hanif, M. (2007). Estimation of Population Mean In Single and Two Phase Sampling with or without Additional Information. *Pak. J. Statist, 23*(2), 99.

Sen, A. R. (1971). Successive sampling with two-auxiliary variables. *Sankhya B, 33*, 371-378.

Seng, Y. P. (1951). Historical survey of the development of sampling theories and practice. *Journal of the Royal Statistical Society. Series A (General), 114*(2), 214-231.

Shabbir, J., & Gupta, S. (2007). On estimating the finite population mean with known population proportion of an auxiliary variable. *Pak J. Statist, 23*(1), 1-9.

Shahbaz, M. Q., & Hanif, M. (2009). A General Shrinkage Estimator in Survey Sampling. *World Applied Sciences Journal, 7*(5), 593-596.

Singh, A. K., & Singh, H. P. (2001). Dual to chain ratio type estimator in double sampling using two auxiliary variables. *J. Ravishankar University, 14*, 99-106.

Singh, A. K., Singh, H. P., & Upadhyaya, L. N. (2001). A generalized chain estimator for finite population mean in two phase sampling. *J. Ind. Soc. Agri. Statist., 34*(3), 370-375.

Singh, G. N., & Upadhyaya, L. N. (1995). A class of modified chain type estimators using two auxiliary variables in two-phase sampling. *Metron, 53*, 117-125.

Singh, H. P. (1987). On the estimation of population mean when the correlation coefficients is known in two phase sampling. *Assam Statist. Rev, 1*(1), 17-21.

Singh, H. P. (1993). A chain ratio-cum-difference estimator using two auxiliary variates in double sampling. *Journal of Raishankar University, 6*, 79-83.

Singh, H. P., & Biradar, R. S. (1994). A class of unbiased ratio and product estimators in two-phase sampling. *Statistica, LIV*(3), 349-359.

Singh, H. P., & Espejo, M. R. (2003). On linear regression and ratio-product estimation of a finite population mean. *Journal of the Royal Statistical Society. Series D (The Statistician), 52*(1), 59-67.

Singh, H. P., & Espejo, M. R. (2007). Double sampling ratio-product estimator of a finite population mean in sample surveys. *Journal of Applied Statistics, 34*(1), 71-85.

Singh, H. P., & Gangele, R. K. (1995). An improved regression estimator with known coefficient of variation in two-phase sampling with two auxiliary variables. *J. Vikram University Ujjain, XXV*, 27-40.

Singh, H. P., & Gangele, R. K. (1999). Classes of almost unbiased ration and product-type estimators in two phase sampling. *Statistica, LIX*(1), 109-124.

Singh, H. P., Katyar, N. P., & Gangwar, D. K. (1996). A class of almost unbiased regression-type estimators in two phase sampling applying Quenouille's mehod. *J. Ind. Soc. Agri. Statist., 48*(1), 98-104.

Singh, H. P., & Namjoshi, U. D. (1988). A class of multivariate regression estimators in two-phase sampling. *Assam Statist. Rev, 2*(2), 1-7.

Singh, H. P., Singh, H. P., & Kushwaha, K. S. (1992). On chain ratio-to-regression-type estimator in double sampling. *Assam Statist. Rev., 6*(2), 91-105.

Singh, H. P., & Singh, R. (2002). A class of chain ratio-type estimators for the coefficient of variation of finite population in two phase sampling. *Aligarh J. Statist., 22*, 1-9.

Singh, H. P., Singh, S., & Kim, J. M. (2006). General families of chain ratio type estimators of the population mean with known coefficient of variation of the second auxiliary variable in two phase sampling. *Journal of the Korean Statistical Society, 35*(4), 377-395.

Singh, H. P., & Tailor, R. (2000). Predictive estimation of finite population mean in two phase sampling with known coefficient of variation of second auxiliary variable. *Vikram Math. J., 20*, 5-13.

Singh, H. P., Tripathi, T. P., & Upadhyaya, L. N. (1989). Improved estimators for population mean based on double sampling. *J. Ind. Statist. Assoc., 27* 89-99.

Singh, H. P., Upadhyaya, L. N., & Chandra, P. (2004). A general family of estimators for estimating population mean using two auxiliary variables in two-phase sampling. *Statistics in transition, 6*(7), 1055-1077.

Singh, H. P., Upadhyaya, L. N., & Iachan, R. (1990). An efficient class of estimators using supplementary information in sample survey. *Aligarh J. Statist., 10*, 37-50.

Singh, H. P., & Vishwakarma, G. K. (2005-2006). An Efficient Variant of the Product and ratio estimators in double sampling. *Model Assisted Statistics and Applications, 1*(3), 155-165.

Singh, R., & Singh, H. P. (2003). Estimation of variance through regression approach in two phase sampling. *Aligarh J. Statist. , 23*(13-30).

Singh, V. K., & Singh, H. P. (1994). Estimation of ratio and product of two finite population means in two-phase sampling. *Journal of statistical planning and inference, 41*(2), 163-171.

Srivastava, S. K. (1970). A two-phase sampling estimator in sample surveys. *Australian & New Zealand Journal of Statistics, 12*(1), 23-27.

Srivastava, S. K. (1971). A generalized estimator for the mean of a fininte population using multi-auxiliary information. *J. Amer. Statist. Assoc., 66*, 404-407.

Srivastava, S. R., Khare, B. B., & Srivastava, S. R. (1990). A generalized chain ratio estimator for mean of finite population. *Journal of the Indian Society of Agricultural Statistics, 42*, 108–117.

Stephan, F. F. (1948). History of the uses of modern sampling procedures. *Journal of the American Statistical Association, 43*(241), 12-39.

Sukhatme, B. V. (1962). Some ratio-type estimators in two-phase sampling. *Journal of the American Statistical Association, 57*(299), 628-632.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., & Ashok, C. (1984). *Sampling theory of surveys with applications*: Iowa State University Press.

Swain, A. (2000). *Current Developments in Survey Sampling*. Bhubaneswar, India: The Modern Book Depot.

Tracy, D. S., & Singh, H. P. (1999). Efficient Use of Two Auxiliary Variables in Two Phase Sampling as Well as in Successive Sampling. *Pak J. Statist, 15*(1), 27-40.

Tripathi, T. P. (1989). *Optimum estimation of mean vector for dynamic population.* Paper presented at the International Symposium on Optimization and Statistics, AMU. Aligarh.

Tripathi, T. P., & Ahmed, M. S. (1995). A class of estimators for a finite population mean based on multivariate information and general two-phase sampling. *Calcutta Statistical Association Bulletin, 45*, 179-180.

Tripathi, T. P., & Chaubey, Y. P. (1993). *Optimum probabilities of selection in pps sampling based on super population model and multivariate information,* . ISI, Culcatta.

Tripathi, T. P., & Khattree, R. (1989). *Simultaneous estimation of several means using multivariate auxiliary information*. ISI, Culcatta: Technical Report.

Tripathi, T. P., Singh, H. P., & Upadhyaya, L. N. (1988). A generalized method of estimation in double sampling. *J. Ind. Statist. Assoc., 26*, 91-101.

Upadhyaya, L., Kushwaha, K. S., & Singh, H. P. (1990). A modified chain ratio-type estimator in two phase sampling using multi auxiliary information. *Metron, 48*, 381-393.

Upadhyaya, L. N., Dubey, S. P., & Singh, H. P. (1992). A class of ratio-in-regression estimators using two auxiliary variables in double sampling. *J. Scient. Res, 42*, 127-134.

Upadhyaya, L. N., Singh, H. P., & Tailor, R. (2006). Estimation of mean with known coefficient of variation of an auxiliary variable in two phase sampling. *Statistics in Transition, 7*(6), 1327-1344.

Yates, F. (1960). Sampling Methods for Censuses and Surveys. Charles Griffin and Company. *Limited, London*, 356-364.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association, 57*(298), 348-368.

Appendix-A: Populations Description

Population	Variable	Description
Population-1	**Y= Length of Stay**	Average length of stay of all patients in hospital (in days)
	X= Age	Average age of patients (in years)
	W_1= Infection Risk	Average estimated probability of acquiring infection in hospital (in percent)
	W_2= No. of Beds	Average No. of beds in hospital during study period
	W_3= No. of Nurses	Average No. of full-time registered nurses during study period
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	
Population-2	**Y= PSA level**	Serum prostate-specific antigen level (mg/ml)
	X= Cancer volume	Estimate of prostate cancer volume (cc)
	W_1= Weight	Prostate weight (gm)
	W_2= Age	Age of the patient (years)
	W_3= BPH	Amount of Benign prostatic hyperplasia(cm^2)
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	
Population-3	**Y= Total Serious Crime**	Total no. of serious crime in thousands
	X= Total Population	Estimated population
	W_1= Percent Unemployment	Percent of CDI labor force that is unemployed
	W_2= Income	Per capita income of CDI
	W_3= Population 18-34	Percent of population aged 18-34 years
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	
Population-4	**Y= Price**	Sales price in millions
	X= Finished Square Feet	Finished are of residence (Square Feet)
	W_1= No. of bedrooms	Total no. of bedrooms in residence
	W_2= No. of bathrooms	Total no. of bathrooms in residence
	W_3= Garage size	No. of cars that garage will hold
	W4= Lot Size	Lot size (square feet)
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	

Population-5	**Y= GPA**	Grade-point average
	X= High School Rank	High School class rank as percentile
	W_1= ACT Score	ACT entrance examination score
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	
Population-6	**Y= Market Share**	Average monthly market share
	τ = Nielsen Rating	Gross Nielson rating (1 if High: 0 otherwise)
	δ_1 = Discount price	Presence or absence of discount price (1 if yes: 0 otherwise)
	δ_2 = Package Promotion	Presence or absence of package promotion (1 if yes: 0 otherwise)
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	
Population-7	**Y= Sales Price**	Sales price in millions
	τ = Air Conditioning	Presence or absence of air conditioning (1 if yes: 0 otherwise)
	δ_1 = Pool	Presence or absence of swimming pool (1 if yes: 0 otherwise)
	δ_2 = Quality	Quality of construction (1 if Good: 0 if Not Good)
	δ_3 = Highway	Presence or absence of adjacency of highway (1 if yes: 0 otherwise)
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	
Population-8	**Y= Cost**	Total cost of claims
	τ = Gender	Gender of Subscriber (1 if male: 0 otherwise)
	δ_1 = Age	Age of the subscriber (1 if >= 50 years: 0 otherwise)
	δ_2 = Complications	Complication (1 if yes: 0 No)
	δ_3 = Co morbidities	Co morbidities (1 if yes: 0 No)
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	

Population-9	**Y= Liver press rate**	Total no. of lever presses dived by elapsed time in seconds
	τ = **Unit**	Observation Unit
	δ_1 = **Initial rate**	Initial liver press rate (1 slow : 0 otherwise)
	δ_2 = **Part**	Part of Study (1 if FR-2 : 0 if FR-5)
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	
Population-10	**Y= Weight**	Weight Gain of rats
	τ = **Protein**	The protein source: 1= Beef ; 0=Cereal
	δ_1 = **Amount**	The amount of protein: 1= High; 0=Low
Source	Neter, J., Wasserman, W., Kutner, M. H., & Li, W. (1996). *Applied linear statistical models*: Irwin.	
Population-11	**Y1= math**	Mathematics Score
	Y2= science	Science Score
	Y3= social science	Social Science Score
	x_1 = **Read**	Reading Score
	x_2 = **Write**	Writing Score
Source	UCLA: Academic Technology Services, Statistical Consulting Group. from http://www.ats.ucla.edu .	
Population-12	**Y1= head**	Headroom (in.)
	Y2= trunk	Trunk space (cu. ft.)
	Y3= turn	Turn Circle (ft.)
	x_1 = **weight**	Weight (lbs.)
	x_2 = **length**	Length (in.)
Source	STATA: STATA Corporation LP, from http://www.stata.com	

Population-13	**Y1= locus_of_control**	Locus of Control Score
	Y2= self_concept	Self-Concept Score
	Y3=motivation	Motivation Score
	x_1 = **Read**	Reading Score
	x_2 = **Write**	Writing Score
Source	UCLA: Academic Technology Services, Statistical Consulting Group. from http://www.ats.ucla.edu ..	

Appendix-B: R code

Code to Calcuate Partial Correlation matrix

```
pcor.mat <- function(x,y,z,method="p",na.rm=T){

       x <- c(x)
       y <- c(y)
       z <- as.data.frame(z)

       if(dim(z)[2] == 0){
              stop("There should be given data\n")
       }

       data <- data.frame(x,y,z)

       if(na.rm == T){
              data = na.omit(data)
       }

       xdata <- na.omit(data.frame(data[,c(1,2)]))
       Sxx <- cov(xdata,xdata,m=method)

       xzdata <- na.omit(data)
       xdata <- data.frame(xzdata[,c(1,2)])
       zdata <- data.frame(xzdata[,-c(1,2)])
       Sxz <- cov(xdata,zdata,m=method)

       zdata <- na.omit(data.frame(data[,-c(1,2)]))
       Szz <- cov(zdata,zdata,m=method)

       # is Szz positive definite?
       zz.ev <- eigen(Szz)$values
       if(min(zz.ev)[1]<0){
              stop("\'Szz\' is not positive definite!\n")
       }

       # partial correlation
       Sxx.z <- Sxx - Sxz %*% solve(Szz) %*% t(Sxz)

       rxx.z <- cov2cor(Sxx.z)[1,2]

       rxx.z
}
```

Code to Calcuate Mean Square of the estimator given in (Quantitative)

```
(pcor.mat(x,y,new[,c("w1", "w2", "w3", "w4")]))^2-> r2xy.w
lm(y~x+w1+w2+w3+w4)-> lm.yxw
summary(lm.yxw)$r.squared->r2y.xw
lm(y~w1+w2+w3+w4)->lm.yw
summary(lm.yw)$r.squared->r2y.w
c(r2xy.w=r2xy.w, r2y.xw=r2y.xw, r2y.w=r2y.w)
var(y)->s2y

mse<-s2y*(th2*(1-r2y.xw)+th1*r2xy.w*(1-r2y.w))
mse1<- s2y*(th2*(1-r2y.xw)+th1*r2y.xw)
mse
mse1

 for (th1 in seq(0.1,0.9,0.1)){
  for(th2 in seq(0.1,0.9,0.1)){
  mse<-s2y*(th2*(1-r2y.xw)+th1*r2xy.w*(1-r2y.w))
  mse1<- s2y*(th2*(1-r2y.xw)+th1*r2y.xw)
  c(mse,mse1)-> u
write(u,file="c:/Real.txt",append=T,sep="\t")
  print(u)
   }
    }
```

Code to Calcuate Mean Square of the estimator given in (Qualitative)

```
library(polycor)
hetcor(y,x,w1,w2)$cor->cat
-cat[1,2]/sqrt(cat[1,1]*cat[2,2])->r2xy.w

lm(y~x+w1+w2)-> lm.yxw
summary(lm.yxw)$r.squared->r2y.xw
lm(y~w1+w2)->lm.yw
summary(lm.yw)$r.squared->r2y.w
c(r2xy.w=r2xy.w, r2y.xw=r2y.xw, r2y.w=r2y.w)
var(y)->s2y

 for (th1 in seq(0.1,0.9,0.1)){
  for(th2 in seq(0.1,0.9,0.1)){
  mse<-s2y*(th2*(1-r2y.xw)+th1*r2xy.w*(1-r2y.w))
  mse1<- s2y*(th2*(1-r2y.xw)+th1*r2y.xw)
  c(mse,mse1)-> x
write(x,file="c:/Cdrug.txt",append=T,sep="\t")
  print(x)
   }
    }
```

Code to Calcuate Mean Square of the estimator given in (Multivariate)

```
#for v11
(pcor.mat(x,y1,new[,c("w1")]))^2-> r2xy.w
lm(y1~x+w1)-> lm.yxw
summary(lm.yxw)$r.squared->r2y.xw
lm(y1~w1)->lm.yw
summary(lm.yw)$r.squared->r2y.w
var(y1)->s2y

 for (th1 in seq(0.1,0.9,0.1)){
  for(th2 in seq(0.1,0.9,0.1)){
  mse<-s2y*(th2*(1-r2y.xw)+th1*r2xy.w*(1-r2y.w))
  mse1<- s2y*(th2*(1-r2y.xw)+th1*r2y.xw)
  c(mse,mse1,th1,th2,11)-> u1
write(u1,file="c:/v.txt",append=T,sep="\t")
  print(u1)
   }
    }

#for v22

(pcor.mat(x,y2,new[,c("w1")]))^2-> r2xy.w
lm(y2~x+w1)-> lm.yxw
summary(lm.yxw)$r.squared->r2y.xw
lm(y2~w1)->lm.yw
summary(lm.yw)$r.squared->r2y.w
var(y2)->s2y

 for (th1 in seq(0.1,0.9,0.1)){
  for(th2 in seq(0.1,0.9,0.1)){
  mse<-s2y*(th2*(1-r2y.xw)+th1*r2xy.w*(1-r2y.w))
  mse1<- s2y*(th2*(1-r2y.xw)+th1*r2y.xw)
  c(mse,mse1,th1,th2,22)-> u2
write(u2,file="c:/v.txt",append=T,sep="\t")
  print(u2)
   }
    }

#for v33
 (pcor.mat(x,y3,new[,c("w1")]))^2-> r2xy.w
lm(y3~x+w1)-> lm.yxw
summary(lm.yxw)$r.squared->r2y.xw
lm(y3~w1)->lm.yw
summary(lm.yw)$r.squared->r2y.w
var(y3)->s2y

 for (th1 in seq(0.1,0.9,0.1)){
  for(th2 in seq(0.1,0.9,0.1)){
  mse<-s2y*(th2*(1-r2y.xw)+th1*r2xy.w*(1-r2y.w))
  mse1<- s2y*(th2*(1-r2y.xw)+th1*r2y.xw)
  c(mse,mse1,th1,th2,33)-> u3
write(u3,file="c:/v33.txt",append=T,sep="\t")
  print(u3)
   }
    }
```

```
#for v12

(var(y1))^0.5 -> sy1
(var(y2))^0.5 -> sy2

cor(y1,y2)->ry1y2
cor(x,y1)->rxy1
cor(x,y2)->rxy2
cor(w1,y1)->rwy1
cor(w1,y2)->rwy2
cor(w1,x) -> rwx

(pcor.mat(x,y1,new[,c("w1")]))-> rxy1.w
(pcor.mat(x,y2,new[,c("w1")]))-> rxy2.w

(rxy1*rxy2+rwy1*rwy2-rxy1*rwy2*rwx-rxy2*rwy1*rwx)/(1-rwx^2)-> A

 for (th1 in seq(0.1,0.9,0.1)){
  for(th2 in seq(0.1,0.9,0.1)){
  s12<- sy1*sy2*(th2*(ry1y2-A)+th1*rxy1.w*rxy2.w*sqrt(1-rwy1^2)*sqrt(1-
rwy2^2))
  s12_z<- sy1*sy2*(th2*(ry1y2^2-A)+th1*A)
  c(s12,s12_z,th1,th2,12)-> u12
write(u12,file="c:/v.txt",append=T,sep="\t")
  print(u12)
   }
    }

#for v13

(var(y1))^0.5 -> sy1
(var(y3))^0.5 -> sy2

cor(y1,y3)->ry1y2
cor(x,y1)->rxy1
cor(x,y3)->rxy2
cor(w1,y1)->rwy1
cor(w1,y3)->rwy2
cor(w1,x) -> rwx

(pcor.mat(x,y1,new[,c("w1")]))-> rxy1.w
(pcor.mat(x,y3,new[,c("w1")]))-> rxy2.w

(rxy1*rxy2+rwy1*rwy2-rxy1*rwy2*rwx-rxy2*rwy1*rwx)/(1-rwx^2)-> A

 for (th1 in seq(0.1,0.9,0.1)){
  for(th2 in seq(0.1,0.9,0.1)){
  s12<- sy1*sy2*(th2*(ry1y2-A)+th1*rxy1.w*rxy2.w*sqrt(1-rwy1^2)*sqrt(1-
rwy2^2))
  s12_z<- sy1*sy2*(th2*(ry1y2^2-A)+th1*A)
  c(s12,s12_z,th1,th2,13)-> u13
write(u13,file="c:/v.txt",append=T,sep="\t")
  print(u13)
   }
    }
```

```
#for v23


(var(y2))^0.5 -> sy1
(var(y3))^0.5 -> sy2

cor(y2,y3)->ry1y2
cor(x,y2)->rxy1
cor(x,y3)->rxy2
cor(w1,y2)->rwy1
cor(w1,y3)->rwy2
cor(w1,x) -> rwx

(pcor.mat(x,y2,new[,c("w1")]))-> rxy1.w
(pcor.mat(x,y3,new[,c("w1")]))-> rxy2.w

(rxy1*rxy2+rwy1*rwy2-rxy1*rwy2*rwx-rxy2*rwy1*rwx)/(1-rwx^2)-> A

 for (th1 in seq(0.1,0.9,0.1)){
  for(th2 in seq(0.1,0.9,0.1)){
  s12<- sy1*sy2*(th2*(ry1y2-A)+th1*rxy1.w*rxy2.w*sqrt(1-rwy1^2)*sqrt(1-
rwy2^2))
  s12_z<- sy1*sy2*(th2*(ry1y2^2-A)+th1*A)
  c(s12,s12_z,th1,th2,23)-> u23
write(u23,file="c:/v.txt",append=T,sep="\t")
  print(u23)
   }
    }
```

Appendix-C: Published Work

World Applied Sciences Journal 13 (10): 2216-2223, 2011
ISSN 1818-4952

Multivariate Estimators for Two Phase Sampling

[1] Nadeem Shafique Butt, [1] Shahid Kamal and [2] Muhammad Qaiser Shahbaz

[1]College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan
[2]Department of Mathematics, COMSATS Institute of IT, Lahore, Pakistan

Abstract: In this paper some new multivariate estimators for two phase sampling has been proposed. The proposed estimators use information on multiple quantitative variables and as well as multiple qualitative variables. Empirical study has been carried out to see the performance of proposed estimator over estimator proposed by Ahmed, Hussin [1].

Key words: Multivariate estimator . two phase sampling . multiple auxiliary variables . minimum variance

INTRODUCTION

The auxiliary information has always been a source of improvement in estimation of certain population characteristics. Several estimators have been developed in single and two phase sampling which utilizes information on auxiliary variables as well as auxiliary attributes. The classical estimators which use information on auxiliary variables are the ratio and regression estimators as given in Hansen, Hurwitz [2]. The classical regression estimator of population mean is given as:

$$\bar{y}_{lr} = \bar{y} + \beta\left(\bar{X} - \bar{x}\right) \qquad (1.1)$$

The value of β for which the variance of (1.1) is minimum is $\beta = S_{xy}/S_x^2$. The minimum variance of (1.1) is given as:

$$\text{Var}(\bar{y}_{lr}) = \theta S_y^2\left(1 - \rho_{yx}^2\right) \qquad (1.2)$$

where $\theta = n^{-1}-N^{-1}$ and ρ_{yx} is the correlation coefficient between X and Y. The estimator (1.1) in case of several auxiliary variables has been discussed by number of statisticians and the estimator in this case is given as:

$$\bar{y}_{mlr} = \bar{y} + \beta^{/}\left(\bar{\mathbf{X}} - \bar{\mathbf{x}}\right) \qquad (1.3)$$

where $\bar{\mathbf{x}}$ is vector of sample means for auxiliary variables. The variance of (1.3); reported by Ahmed, Hanif [3] among others; is given as:

$$\text{Var}\left(\bar{y}_{mlr}\right) = \theta S_y^2\left(1 - \rho_{y.\mathbf{x}}^2\right) \qquad (1.4)$$

where $\rho_{y.\mathbf{x}}^2$ is the squared multiple correlation coefficient between Y and x. The classical regression estimator for two phase sampling is given by Hansen, Hurwitz [2] as:

$$\bar{y}_{1(2)} = \bar{y}_2 + \beta\left(\bar{x}_1 - \bar{x}_2\right) \qquad (1.5)$$

where $\bar{x}_1$ and $\bar{x}_2$ are first phase and second phase means of auxiliary variable X and $\bar{y}_2$ is second phase mean of Y. The variance of (1.5) is given as:

$$\text{Var}\left(\bar{y}_{1(2)}\right) = S_y^2\left\{\theta_2\left(1 - \rho_{yx}^2\right) + \theta_1\rho_{yx}^2\right\} \qquad (1.6)$$

where

$$\theta_h = n_h^{-1} - N_h^{-1}$$

and n_h is sample size at h^{th} phase. Ahmed, Hanif [3] has extended the (1.6) the case of several variables. Sahoo, Sahoo [4] has proposed the regression type estimator using information of two auxiliary variables. The estimator proposed by Sahoo, Sahoo [4] is given as:

$$\bar{y}_{ssm} = \bar{y}_2 + \beta_1\left(\bar{x}_1 - \bar{x}_2\right) + \beta_2\left(\bar{Z} - \bar{z}\right) \qquad (1.7)$$

The variance of (1.7) is:

$$\text{Var}\left(\bar{y}_{ssm}\right) = S_y^2\left\{\theta_2\left(1 - \rho_{yx}^2\right) + \theta_1\left(\rho_{yx}^2 - \rho_{yz}^2\right)\right\} \qquad (1.8)$$

where ρ_{yz}^2 is squared correlation coefficient between Y and Z.

Corresponding Author: Dr. Nadeem Shafique Butt, College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan

Jhajj, Sharma [5] have proposed a family of estimators in single and two phase sampling using information on auxiliary attributes. The variance of the proposed family depends upon the point bi-serial correlation coefficient. Samiuddin and Hanif [6] have also proposed several estimators in single and two phase sampling. A regression-in-ratio estimator proposed by Samiuddin and Hanif [6] is:

$$\bar{y}_{s\,I(2)} = \left[\bar{y}_2 + \beta_{yz}\left(\bar{z}_1 - \bar{z}_2\right)\right]\frac{\bar{X}}{\bar{x}_2} \tag{1.9}$$

The variance of (1.9) is:

$$\begin{aligned}MSE\left(\bar{y}_{s\,I(2)}\right) \approx \bar{Y}^2\Big[\theta_2\left\{C_y^2\left(1-\rho_{xy}^2\right) + \left(C_x - C_y\rho_{xy}\right)^2\right\} \\ +\left(\theta_2 - \theta_1\right)\left\{C_x^2\rho_{xz}^2 - \left(C_y\rho_{yz} - C_x\rho_{xz}\right)^2\right\}\Big]\end{aligned} \tag{1.10}$$

Hanif, Haq [7] proposed a generalized family of estimators based on the information of "k" auxiliary attributes and discussed the estimator for full, partial and no information cases. Hanif, Haq [7] showed that the proposed family has smaller mean square error than given by Jhajj, Sharma [5]. Hanif, Haq [8] proposed some ratio estimators for single phase and two phase sampling by using information on multiple auxiliary attributes. The proposed estimators are generalization of the estimator proposed by Naik and Gupta [9]. Hanif, Haq [8] also drive the shrinkage version of the proposed estimators by using the method given Shahbaz and Hanif [10]

Hanif, Ahmed [11] proposed a number of generalized multivariate ratio estimators for two-phase and multi-phase sampling in the presence of multi-auxiliary variables for estimating population mean for a single variable and a vector of variables of interest. Hanif, Ahmed [11] proposed more general ratio estimator when information on all auxiliary variables are not available for population (No Information Situation), the estimator is:

$$T_{hk(1\times p)} = \left[\bar{y}_{(k)1}\prod_{i=1}^{q}\left(\frac{\bar{x}_{(h)i}}{\bar{x}_{(k)i}}\right)^{a_{i1}} \bar{y}_{(k)2}\prod_{i=1}^{q}\left(\frac{\bar{x}_{(h)i}}{\bar{x}_{(k)i}}\right)^{a_{i2}} \cdots \bar{y}_{(k)p}\prod_{i=1}^{q}\left(\frac{\bar{x}_{(h)i}}{\bar{x}_{(k)i}}\right)^{a_{ip}}\right] \tag{1.11}$$

The variance-covariance matrix of the estimator is of the following form:

$$\Sigma_{T_{hk}(p\times p)} = q_k\,\Sigma_{y(p\times p)} - \left(q_k - q_h\right)\Sigma'_{y(p\times p)}\,\Sigma^{-1}_{x(q\times q)}\,\Sigma_{yx(q\times p)} \tag{1.12}$$

Ahmed, Hussin [1] proposed a number of generalized multivariate regression estimators for two-phase and multi-phase sampling in the presence of multi-auxiliary variables for estimating population mean for a single variable and a vector of variables of interest. The proposed estimator is of the following form:

$$T_{hk(1\times p)} = \left[\bar{y}_{(k)1}\,\bar{y}_{(k)2}\,\ldots\,\bar{y}_{(k)p}\right] + \left[\sum_{i=1}^{q}a_{i1}\left(\bar{x}_{(h)i} - \bar{x}_{(k)i}\right)\sum_{i=1}^{q}a_{i2}\left(\bar{x}_{(h)i} - \bar{x}_{(k)i}\right)\ldots\sum_{i=1}^{q}a_{ip}\left(\bar{x}_{(h)i} - \bar{x}_{(k)i}\right)\right] \tag{1.13}$$

The variance-covariance matrix of the estimator is of the following form:

$$\Sigma_{T_{hk}(p\times p)} = q_k\Sigma_{y(p\times p)} - \left(q_k - q_h\right)\Sigma_{yx(p\times q)}\Sigma^{-1}_{x(q\times q)}\Sigma'_{x(q\times p)} \tag{1.14}$$

In this paper we have proposed some multivariate regression estimators using information on several auxiliary variables and as well as auxiliary attributes.

NOTATIONS

In this section we define the notations used for the development of the multivariate estimators and variance covariance matrices. Let "w" and "x" be auxiliary variables and Y be the variable of interest. Let S_{xw} be the covariance between x and w, s_{yw} be the covariance between Y and w. Using these notations we define β_{xw} as regression coefficient between x and w for the i-th response variable and

$$\beta_{y_i x.w} = S_{xy.w} / S^2_{x.w}$$

as partial regression coefficient between Y_i and x keeping the w at constant level. Also $S_{y_i x.w}$ is partial covariance between Y_i and x after removing the effect of e, $S^2_{y_i.w}$ is the partial variance of T and $S^2_{x.w}$ is the partial variance of x. We also define

$$\rho^2_{y_i x.w} = S^2_{y_i x.w} / \left(S^2_{x.w} S^2_{y_i.w}\right)$$

as partial correlation coefficient between Y_i and x after removing the effect of w, $\rho^2_{y_i x.w}$ as squared multiple correlation coefficient between Y_i and combined effects of x and w, $\rho^2_{y_i w}$ as squared multiple correlation coefficient between Y_i and w.

MULTIVARIATE ESTIMATOR WITH QUANTITATIVE PREDICTORS

In this section the multivariate extension of Roy [12] estimator has been proposed. The multivariate extension has been proposed by using information on two auxiliary variables and can be used for simultaneous estimation of several variables. The multivariate extension is proposed below:

Suppose a first phase random sample of size n_1 is available and information on auxiliary variables X and W is recorded. Suppose further that a second phase random sample of size n_2 is available and information on auxiliary variables X and W has been collected alongside information of multiple response variables $Y_1, Y_2, \ldots, Y_p$. Suppose further that $\bar{y}_2$ is mean vector of estimates based upon second phase sample, k is a vector of constants and A & B are diagonal matrices with diagonal entries α_i & β_i respectively. Based upon these information, the multivariate estimator is defined below:

$$\mathbf{t} = \bar{\mathbf{y}}_2 + \left(\bar{x}_1 - \bar{x}_2\right)\mathbf{k} + \left(\bar{W} - \bar{w}_1\right)\mathbf{Ak} + \left(\bar{W} - \bar{w}_2\right)\mathbf{Bk} \quad (3.1)$$

The *i*th component of (3.1) is given as:

$$t_i = \bar{y}_{i2} + k_i\left[\left\{\bar{x}_1 + a_i\left(\bar{W} - \bar{w}_1\right)\right\} - \left\{\bar{x}_2 + b_i\left(\bar{W} - \bar{w}_2\right)\right\}\right] \quad (3.2)$$

Using conventional transformation

$$\bar{w}_1 = \bar{W} - \bar{e}_{w_1};\ \bar{w}_2 = \bar{W} - \bar{e}_{w_2};\ \bar{y}_{i2} = \bar{Y}_i + \bar{e}_{y_{i2}}$$

$$\bar{x}_1 = \bar{X} - \bar{e}_{x_1};\ \bar{x}_2 = \bar{X} - \bar{e}_{x_2}$$

the estimator (3.2) can be written in the following form:

$$t_i - y_i = \bar{e}_{y_2} + k_i\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - k_i a_i \bar{e}_{w_1} + k_i b_i \bar{e}_{w_2}$$

Squaring above equation:

$$\left(t_i - y_i\right)^2 = \bar{e}^2_{y_{i2}} + k_i^2\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right)^2 + k_i^2 a_i^2 \bar{e}^2_{w_1} + k_i^2 b_i^2 \bar{e}^2_{w_2} + 2k_i \bar{e}^2_{y_{i2}}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - 2k_i a_i \bar{e}_{y_{i2}} \bar{e}_{w_1}$$
$$+ 2k_i b_i \bar{e}_{y_{i2}} \bar{e}_{w_2} - 2k_i^2 a_i \bar{e}_{w_1}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) + 2k_i^2 b_i \bar{e}_{w_2}\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - 2k_i^2 a_i b_i \bar{e}_{w_1} \bar{e}_{w_2}$$

By applying expectation, the mean square error of t_i is:

$$S_i = MSE\left(t_i\right) = E\left(t_i - \bar{y}_i\right)^2$$

or

$$S_i = q_2 S^2_{y_i} + \left(q_2 - q_1\right)k_i^2 S_x^2 + q_1 k_i^2 a_i^2 S_w^2 + q_2 k_i^2 b_i^2 S_w^2 + 2\left(q_1 - q_2\right)k_i S_{xy_i} - 2q_1 k_i a_i S_{wy_i}$$
$$+ 2q_2 k_i b_i S_{wy_i} + 2\left(q_1 - q_2\right)k_i^2 b_i^2 S_{wx} - 2q_1 k_i^2 a_i b_i S_w^2 \quad (3.3)$$

Optimum values of α_i, β_i and k_i which minimize S_i can be obtain by differentiating (3.3) with respect to unknown quantities.

$$a_i = \frac{S_x}{S_w^2} = b_{xw} \quad (3.4)$$

$$b_i = \frac{S_{wx}}{S_w^2} - \frac{1}{k_i}\frac{S_{wy_i}}{S_w^2} = b_{xw} - \frac{1}{k_i} b_{y_i w} \quad (3.5)$$

$$k_i = \left(\frac{r_{xy_i} - r_{wx} r_{wy_i}}{1 - S_{wx}^2} \right) \frac{S_y}{S_x} = b_{y_i x.w} \tag{3.6}$$

Using the values of (3.4), (3.5) and (3.6) in (3.3); the MSE becomes

$$S_i = S_{y_i}^2 \left[q_2 \left(1 - r_{y_i.wx}^2 \right) + q_1 r_{xy_i.w}^2 \left(1 - r_{wy_i}^2 \right) \right] \tag{3.7}$$

The covariance between any two components of (3.1) is derived as under:

$$t_i = \overline{y}_{i2} + k_i \left[\left\{ \overline{x}_1 + a_i \left(\overline{w} - \overline{w}_1 \right) \right\} - \left\{ \overline{x}_2 + b_i \left(\overline{w} - \overline{w}_2 \right) \right\} \right]$$

$$t_j = \overline{y}_{j2} + k_j \left[\left\{ \overline{x}_1 + a_j \left(\overline{w} - \overline{w}_1 \right) \right\} - \left\{ \overline{x}_2 + b_j \left(\overline{w} - \overline{w}_2 \right) \right\} \right]$$

Using conventional transformations:

$$t_i - y_i = \overline{e}_{y_2} + k_i \left(\overline{e}_{x_1} + \overline{e}_{x_2} \right) - k_i a_i \overline{e}_{w_1} + k_i b_i \overline{e}_{w_2}$$

Similarly

$$t_j - y_j = \overline{e}_{y_{j2}} + k_j \left(\overline{e}_{x_1} + \overline{e}_{x_2} \right) - k_j a_j \overline{e}_{w_1} + k_j b_j \overline{e}_{w_2}$$

Now

$$\begin{aligned} \left(t_i - y_i \right)\left(t_j - y_j \right) &= \overline{e}_{y_{i2}} \overline{e}_{y_{j2}} + k_i \overline{e}_{y_{j2}} \left(\overline{e}_{x_1} - \overline{e}_{x_2} \right) - a_i k_i \overline{e}_{y_{j2}} \overline{e}_{w_1} + b_i k_i \overline{e}_{y_{j2}} \overline{e}_{w_2} + k_j \overline{e}_{y_{i2}} \left(\overline{e}_{x_1} - \overline{e}_{x_2} \right) \\ &+ k_i k_j \left(\overline{e}_{x_1} - \overline{e}_{x_2} \right)^2 - a_i k_i k_j \overline{e}_{w_1} \left(\overline{e}_{x_1} - \overline{e}_{x_2} \right) + b_i k_i k_j \overline{e}_{w_2} \left(\overline{e}_{x_1} - \overline{e}_{x_2} \right) - a_j k_j \overline{e}_{w_1} \overline{e}_{y_{i2}} - a_j k_i k_j \overline{e}_{w_1} \left(\overline{e}_{x_1} - \overline{e}_{x_2} \right) \\ &+ a_i a_j k_i k_j \overline{e}_{w_1}^2 + b_i k_i a_j k_j \overline{e}_{w_1} \overline{e}_{w2} + b_j k_j \overline{e}_{w2} \overline{e}_{y_{i2}} + b_j k_i k_j \overline{e}_{w2} \left(\overline{e}_{x_1} - \overline{e}_{x_2} \right) - a_i k_i b_j k_j \overline{e}_{w_1} \overline{e}_{w2} + b_i b_j k_i k_j \overline{e}_{w2}^2 \end{aligned}$$

By applying expectation to above equation, the covariance is:

$$\begin{aligned} S_{ij} &= Cov\left(t_i, t_j \right) = E\left(t_i - \overline{y}_i \right)\left(t_j - \overline{y}_j \right) \\ S_{ij} &= q_2 S_{y_i y_j} + k_i \left(q_1 - q_2 \right) S_{xy_j} - q_1 a_i k_i S_{wy_j} + q_2 b_i k_i S_{wy_j} + k_j \left(q_1 - q_2 \right) S_{xy_i} + k_i k_j \left(q_2 - q_1 \right) S_x^2 \\ &+ \left(q_1 - q_2 \right) b_i k_i k_j S_{wx} - q_1 a_j k_j S_{wy_i} + q_1 a_i a_j k_i k_j S_w^2 + q_1 a_j b_i k_i k_j S_w^2 + q_2 b_j k_j S_{wy_i} \\ &+ \left(q_1 - q_2 \right) b_j k_i k_j S_{wx} - q_1 a_i b_j k_i k_j S_w^2 + q_2 b_i b_j k_i k_j S_w^2 \end{aligned} \tag{3.8}$$

Using (3.4), (3.5) and (3.6) in (3.8) we have:

$$\begin{aligned} S_{ij} = S_{y_i} S_{y_j} \Bigg[q_2 \left\{ r_{y_i y_j} - \frac{r_{xy_i} r_{xy_j} + r_{wy_i} r_{wy_j} - r_{xy_i} r_{wy_j} r_{wx} - r_{xy_j} r_{wy_i} r_{wx}}{1 - r_{wx}^2} \right\} \\ + q_1 r_{xy_i.w} r_{xy_j.w} \sqrt{1 - r_{wy_i}^2} \sqrt{1 - r_{wy_j}^2} \Bigg] \end{aligned} \tag{3.9}$$

The covariance matrix can be written by using (3.7) and (3.9)

MULTIVARIATE ESTIMATOR WITH QUALITATIVE PREDICTORS

In this section the multivariate extension of Roy [12] estimator has been proposed. The multivariate extension has been proposed by using information on two auxiliary a ttributes and can be used for simultaneous estimation of several variables. The multivariate extension is proposed as:

Suppose a first phase random sample of size n_1 is available and information on auxiliary attributes τ and W is recorded. Further a second phase random sample of size n_2 is available and information on auxiliary attributes τ and W has been collected alongside information of multiple response variables Y_1, Y_2,…, Y_p. Suppose that $\bar{y}_2$ is the mean vector of estimates based upon second phase, k is a vector of constants and A and B are diagonal matrices with diagonal entries γ_i and η_i respectively. Based upon these information, the multivariate estimator is defined below:

$$\mathbf{t} = \bar{\mathbf{y}}_2 + (t_1 - t_2)\mathbf{k} + \left(p_d - p_{d_1}\right)\mathbf{Ak} + \left(p_d - p_{d_2}\right)\mathbf{Bk} \quad (4.1)$$

The ith component of (4.3.1) is given as

$$t_i = \bar{y}_{i2} + k_i\left[\left\{t_1 + g_i\left(p_d - p_{d_1}\right)\right\} - \left\{t_2 + h_i\left(p_d - p_{d_2}\right)\right\}\right] \quad (4.2)$$

using conventional transformation

$$\bar{y}_{i2} = \bar{y}_i + \bar{e}_{y_{i2}}\ ;\ t_1 = t + \bar{e}_{t_1}\ ;\ t_2 = t + \bar{e}_{t_2}\ ;\ p_{d_1} = p_d - \bar{e}_{d_1}$$

and

$$p_{d_2} = p_d - \bar{e}_{d_2}$$

Using above representations, the estimator (4.3.2) can be put in the following form

$$t_i = \left(\bar{y}_i + \bar{e}_{y_{i2}}\right) + k_i\left[\left(t + \bar{e}_{t_1}\right) + g_i\left(p_d - p_d - \bar{e}_{d_1}\right) - \left\{\left(t + \bar{e}_{t_2}\right) + h_i\left(p_d - p_d - \bar{e}_{d_2}\right)\right\}\right]$$

or

$$t_i - y_i = \bar{e}_{y_{i2}} + k_i\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right) - k_i g_i \bar{e}_{d_1} + k_i h_i \bar{e}_{d_2}$$

Squaring above equation:

$$\left(t_i - y_i\right)^2 = \bar{e}^2_{y_{i2}} + k_i^2\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right)^2 + k_i^2 g_i^2 \bar{e}^2_{d_1} + k_i^2 h_i^2 \bar{e}^2_{d_2} + 2k_i \bar{e}^2_{y_{i2}}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right) - 2k_i g_i \bar{e}_{y_{i2}} \bar{e}_{d_1}$$
$$+ 2k_i h_i \bar{e}_{y_{i2}} \bar{e}_{d_2} - 2k_i^2 g_i \bar{e}_{d_1}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right) + 2k_i^2 h_i \bar{e}_{d_2}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right) - 2k_i^2 g_i h_i \bar{e}_{d_1} \bar{e}_{d_2}$$

By applying the expectation, the mean square error of t_i is:

$$S_i = MSE\left(t_i\right) = E\left(t_i - \bar{y}_i\right)^2$$

$$S_i = q_2 S^2_{y_i} + \left(q_2 - q_1\right) k_i^2 S_t^2 + q_1 k_i^2 g_i^2 S_d^2 + q_2 k_i^2 h_i^2 S_d^2 + 2\left(q_1 - q_2\right) k_i S_{t\,y_i} - 2q_1 k_i g_i S_{d\,y_i}$$
$$+ 2q_2 k_i h_i S_{d\,y_i} + 2\left(q_1 - q_2\right) k_i^2 h_i^2 S_{dt} - 2q_1 k_i^2 g_i h_i S_d^2 \quad (4.3)$$

Optimum values of γ_i, η_i and k_i which minimize S_i can be obtained by differentiating (4.3) with respect to unknown quantities.

$$g_i = \frac{S_{dt}}{S_d^2} = b_{td} \quad (4.4)$$

$$h_i = \frac{S_{dt}}{S_d^2} - \frac{1}{k_i}\frac{S_{d\,y_i}}{S_d^2} = b_{td} - \frac{1}{k_i} b_{y_i d} \quad (4.5)$$

$$k_i = \left(\frac{r_{t\,y_i} - r_{dt}\, r_{d\,y_i}}{1 - S^2_{dt}}\right)\frac{S_y}{S_t} = b_{y_i t.d} \quad (4.6)$$

Using the values of (4.4), (4.5) and (4.6) in (4.3); the MSE becomes

$$S_i = S^2_{y_i}\left[q_2\left(1 - r^2_{y_i.dt}\right) + q_1 r^2_{t\,y_i.d}\left(1 - r^2_{d\,y_i}\right)\right] \quad (4.7)$$

The covariance between any two components of (4.3.1) is derived as under:

Table 1: Eigen values of the variance-covariance matrices of proposed estimator

		θ_1							
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
θ_2		λ_1							
	0.2	1.76							
	0.3	2.64	2.64						
	0.4	3.51	3.52	3.53					
	0.5	4.39	4.40	4.40	4.41				
	0.6	5.27	5.27	5.28	5.29	5.29			
	0.7	6.14	6.15	6.16	6.16	6.17	6.18		
	0.8	7.02	7.03	7.03	7.04	7.05	7.05	7.06	
	0.9	7.90	7.90	7.91	7.92	7.92	7.93	7.94	7.94
θ_2		λ_2							
	0.2	0.94							
	0.3	1.40	1.42						
	0.4	1.85	1.88	1.91					
	0.5	2.31	2.34	2.36	2.39				
	0.6	2.77	2.79	2.82	2.85	2.87			
	0.7	3.22	3.25	3.28	3.30	3.33	3.36		
	0.8	3.68	3.71	3.73	3.76	3.79	3.81	3.84	
	0.9	4.14	4.16	4.19	4.22	4.24	4.27	4.30	4.32
θ_2		λ_3							
	0.2	0.08							
	0.3	0.12	0.12						
	0.4	0.16	0.16	0.16					
	0.5	0.20	0.20	0.20	0.20				
	0.6	0.24	0.24	0.24	0.24	0.24			
	0.7	0.28	0.28	0.28	0.28	0.28	0.28		
	0.8	0.32	0.32	0.32	0.32	0.32	0.32	0.32	
	0.9	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35

$$t_i = \bar{y}_{i2} + k_i\left[\left\{t_1 + g_i\left(p_d - p_{d_1}\right)\right\} - \left\{t_2 + h_i\left(p_d - p_{d_2}\right)\right\}\right]$$

Using conventional transformations:

$$t_i - y_i = \bar{e}_{y_{i2}} + k_i\left(\bar{e}_{t_1} + \bar{e}_{t_2}\right) - k_i g_i \bar{e}_{d_1} + k_i h_i \bar{e}_{d_2}$$

Similarly:

$$t_j - y_j = \bar{e}_{y_{j2}} + k_j\left(\bar{e}_{t_1} + \bar{e}_{t_2}\right) - k_j g_j \bar{e}_{d_1} + k_j h_j \bar{e}_{d_2}$$

Now

$$\begin{aligned}\left(t_i - y_i\right)\left(t_j - y_j\right) &= \bar{e}_{y_{i2}}\bar{e}_{y_{j2}} + k_i\bar{e}_{y_{j2}}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right) - g_i k_i \bar{e}_{y_{j2}}\bar{e}_{d_1} + h_i k_i \bar{e}_{y_{j2}}\bar{e}_{d_2} + k_j\bar{e}_{y_{i2}}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right)\\ &+ k_i k_j\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right)^2 - g_i k_i k_j \bar{e}_{d_1}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right) + h_i k_i k_j \bar{e}_{d_2}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right) - g_j k_j \bar{e}_{d_1}\bar{e}_{y_{i2}} - a_j k_i k_j \bar{e}_{d_1}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right)\\ &+ g_i g_j k_i k_j \bar{e}_{d_1}^2 + h_i k g_j k_j \bar{e}_{d_1}\bar{e}_{d2} + h_j k_j \bar{e}_{d2}\bar{e}_{y_{i2}} + h_j k_i k_j \bar{e}_{d_2}\left(\bar{e}_{t_1} - \bar{e}_{t_2}\right) - g_i k_i h_j k_j \bar{e}_{d_1}\bar{e}_{d2} + h_i h_j k_i k_j \bar{e}_{d2}^2\end{aligned}$$

By applying expectation to above equation we get:

Table 2: Eigen values of the variance-covariance matrices of estimator proposed by Ahmed, Hussin [1]

		θ_1							
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
θ_2		μ_1							
	0.2	2.79							
	0.3	3.68	5.31						
	0.4	4.85	5.58	7.88					
	0.5	6.03	6.28	8.07	10.45				
	0.6	7.22	7.36	8.37	10.63	13.03			
	0.7	8.42	8.51	8.97	10.84	13.19	15.60		
	0.8	9.61	9.69	9.93	11.16	13.38	15.76	18.17	
	0.9	10.80	10.88	11.03	11.70	13.62	15.94	18.33	20.75
θ_2		μ_2							
	0.2	2.31							
	0.3	2.74	3.56						
	0.4	2.90	4.62	4.78					
	0.5	3.05	5.24	5.90	5.98				
	0.6	3.19	5.49	6.93	7.13	7.19			
	0.7	3.33	5.66	7.65	8.23	8.34	8.40		
	0.8	3.48	5.81	8.01	9.24	9.47	9.55	9.60	
	0.9	3.62	5.95	8.23	10.01	10.55	10.69	10.76	10.81
θ_2	μ_3								
	0.2	0.11							
	0.3	0.16	0.17						
	0.4	0.21	0.22	0.23					
	0.5	0.25	0.27	0.28	0.28				
	0.6	0.29	0.32	0.33	0.34	0.34			
	0.7	0.32	0.37	0.38	0.39	0.40	0.40		
	0.8	0.36	0.41	0.43	0.44	0.45	0.45	0.46	
	0.9	0.39	0.45	0.48	0.49	0.50	0.51	0.51	0.51

Table 3: Relative efficiency of proposed estimator over estimator proposed by Ahmed, Hussin [1]

		θ_1							
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
θ_2		$\frac{\sum l_i}{\sum m_i}$							
	0.2	0.53							
	0.3	0.63	0.46						
	0.4	0.69	0.53	0.43					
	0.5	0.74	0.59	0.49	0.42				
	0.6	0.77	0.63	0.53	0.46	0.41			
	0.7	0.80	0.67	0.57	0.50	0.45	0.40		
	0.8	0.82	0.69	0.60	0.53	0.48	0.43	0.40	
	0.9	0.84	0.72	0.63	0.56	0.51	0.46	0.43	0.39

$$S_{ij} = Cov\left(t_i, t_j\right) = E\left(t_i - \overline{y}_i\right)\left(t_j - \overline{y}_j\right)$$

$$\begin{aligned} S_{ij} = {} & q_2 S_{y_i y_j} + k_i\left(q_1 - q_2\right) S_{t y_j} - q_1 g_i k_i S_{d y_j} + q_2 h_i k_i S_{d y_j} + k_j\left(q_1 - q_2\right) S_{t y_i} + k_i k_j\left(q_2 - q_1\right) S_t^2 \\ & + \left(q_1 - q_2\right) h_i k_i k_j S_{dt} - q_1 g_j k_j S_{d y_i} + q_1 g_i g_j k_i k_j S_d^2 + q_1 g_j h_i k_i k_j S_d^2 + q_2 h_j k_j S_{d y_i} \\ & + \left(q_1 - q_2\right) h_j k_i k_j S_{dt} - q_1 g_i h_j k_i k_j S_d^2 + q_2 h_i h_j k_i k_j S_d^2 \end{aligned} \quad (4.8)$$

Using (4.4), (4.5) and (4.6) in (4.8) we have

$$S_{ij} = S_{y_i} S_{y_j} \left[q_2 \left\{ r_{y_i y_j} - \frac{r_{ty_i} r_{ty_j} + r_{dy_i} r_{dy_j} - r_{ty_i} r_{dy_j} s_{dt} - r_{ty_j} r_{dy_i} r_{dt}}{1 - r_{dt}^2} \right\} + q_1 r_{ty_i.d} r_{ty_j.d} \sqrt{1 - r_{dy_i}^2} \sqrt{1 - r_{dy_j}^2} \right] \quad (4.9)$$

The covariance matrix can be written by using (4.7) and (4.9)

NUMERICAL STUDY

In this section empirical study is conducted to see the performance of the proposed estimator over the estimator proposed by Ahmed, Hussin [1]. Ratio of the Sum of Eigen values of variance-covariance matrices is used to calculate relative efficiencies of the proposed estimator for various values of θ_1 and θ_2.

Table 1 contains the Eigen values computed from the variance-covariance matrix of proposed estimator and Table 2 contain the Eigen values computed from the variance-covariance matrix of estimator proposed by Ahmed, Hussin [1]. Table 3 shows efficiency comparison of proposed multivariate estimator with estimator proposed by Ahmed, Hussin [1]. The entries of Table 3 clearly indicate that the proposed estimator is more efficient as compared with the estimator proposed by Ahmed, Hussin [1] for all combinations of θ_1 and θ_2.

REFERENCES

1. Ahmed, Z., A.G. Hussin and M. Hanif, 2010. Generalized multivariate regression estimators for multi-phase sampling using multi-auxiliary variables. Pak. J. Statist., 26 (4): 569-583.
2. Hansen, M.H., W.N. Hurwitz and W.G. Madow, 1953. Sample survey methods and theory. New York: Wiley.
3. Ahmed, Z., M. Hanif and M. Ahmad, 2009. Generalized Regression Cum-Ratio Estimators for Two-Phase Sampling Using Muliti-Auxiliary Variables. Pak. J. Statist., 25 (2): 93-106.
4. Sahoo, J., L.N. Sahoo and S. Mohanty, 1993. A regression approach to estimation in two-phase sampling using two auxiliary variables. Current Science, 65 (1): 73-75.
5. Jhajj, H.S., M.K. Sharma and L.K. Grover, 2006. A family of estimators of population mean using information on auxiliary attribute. Pak. J. Statist., 22 (1): 43-50.
6. Samiuddin, M. and M. Hanif, 2007. Estimation of Population Mean In Single and Two Phase Sampling with or without Additional Information. Pak. J. Statist., 23 (2): 99.
7. Hanif, M., I. Haq and M.Q. Shahbaz, 2009. On a new family of estimators using multiple auxiliary attributes. World Applied Sciences Journal, 7 (11): 1419-1422.
8. Hanif, M., I. Haq and M.Q. Shahbaz, 2010. Ratio Estimators using Multiple Auxiliary Attributes. World Applied Sciences Journal, 8 (1): 133-136.
9. Naik, V.D. and P.C. Gupta, 1996. A note on estimation of mean with known population proportion of an auxiliary character. Jour. Ind. Soc. Agr. Stat., 48 (2): 151-158.
10. Shahbaz, M.Q. and M. Hanif, 2009. A General Shrinkage Estimator in Survey Sampling. World Applied Sciences Journal, 7 (5): 593-596.
11. Hanif, M., Z. Ahmed and M. Ahmad, 2009. Generalized Multivariate Ratio Estimator using Multi-Auxiliary Variables for Multi-Phase Sampling. Pak. J. Statist., 25 (4): 615-629.
12. Roy, D.C., 2003. A regression type estimator in two phase sampling using two auxiliary variables. Pak J. Statist., 19 (3): 281-290.

Middle-East Journal of Scientific Research 6 (2): 198-200, 2010
ISSN 1990-9233

Estimation of Population Mean in Two Phase Sampling

[1]Nadeem Shafique Butt, [1]Shahid Kamal and [2]Muhammad Qaiser Shahbaz

[1]College of Statistical and Actuarial Sciences University of the Punjab, Lahore, Pakistan
[2]Department of Mathematics, COMSATS Institute of IT, Lahore, Pakistan

Abstract: A new estimator for population mean has been proposed in two phase sampling by using information of multiple auxiliary variables. The minimum variance of the proposed estimator has been obtained. Comparison has also been made with some available estimators of two phase sampling.

Key words: Two phase sampling · Multiple auxiliary variables · Minimum variance

INTRODUCTION

The auxiliary information has always been a source of improvement in estimation of certain population characteristics. Several estimators have been developed in single and two phase sampling which utilizes information on auxiliary variables as well as auxiliary attributes. The classical estimators which use information on auxiliary variables are the ratio and regression estimators as given in Hansen, *et al.* [1]. The classical regression estimator of population mean is given as:

$$\bar{y}_{lr} = \bar{y} + \beta(\bar{X} - \bar{x}) \quad (1.1)$$

The value of β for which the variance of (1.1) is minimum is $\beta = S_{xy}/S^2_x$. The minimum variance of (1.1) is given as:

$$Var(\bar{y}_{lr}) = \theta S_y^2\left(1-\rho_{yx}^2\right) \quad (1.2)$$

Where
$\theta = n^{-1} - N^{-1}$ and ρ_{yx} is the correlation coefficient between X and Y. The estimator (1.1) in case of several auxiliary variables has been discussed by number of statisticians and the estimator in this case is given as:

$$\bar{y}_{mlr} = \bar{y} + \beta^{/}\left(\bar{\mathbf{X}} - \bar{\mathbf{x}}\right); \quad (1.3)$$

Where:
$\bar{\mathbf{x}}$ is vector of sample means for auxiliary variables. The variance of (1.3); reported by Ahmad [2] among others; is given as:

$$Var(\bar{y}_{mlr}) = \theta S_y^2\left(1-\rho_{y.\mathbf{x}}^2\right); \quad (1.4)$$

Where:
$\rho_{y.\mathbf{x}}^2$ is the squared multiple correlation coefficient between Y and x. The classical regression estimator for two phase sampling is given by Hansen, *et al.* [1] as:

$$\bar{y}_{lr(2)} = \bar{y}_2 + \beta(\bar{x}_1 - \bar{x}_2); \quad (1.5)$$

Where:
$\bar{x}_1$ and $\bar{x}_2$ are first phase and second phase means of auxiliary variable X and $\bar{y}_2$ is second phase mean of Y. The variance of (1.5) is given as:

$$Var\left(\bar{y}_{lr(2)}\right) = S_y^2\left\{\theta_2\left(1-\rho_{yx}^2\right)+\theta_1\rho_{yx}^2\right\}; \quad (1.6)$$

Where:
$\theta_h = n^{-1}_h - N^{-1}_h$ and n_h is sample size at h^{th} phase. Ahmed [2] has extended the (1.6) the case of several variables. Sahoo, *et al.* [3] has proposed the regression type estimator using information of two auxiliary variables. The estimator proposed by Sahoo, *et al.* [3] is given as:

$$\bar{y}_{ssm} = \bar{y}_2 + \beta_1(\bar{x}_1 - \bar{x}_2) + \beta_2(\bar{Z} - \bar{z}) \quad (1.7)$$

The variance of (1.7) is:

$$Var(\bar{y}_{ssm}) = S_y^2\left\{\theta_2\left(1-\rho_{yx}^2\right)+\theta_1\left(\rho_{yx}^2-\rho_{yz}^2\right)\right\} \quad (1.8)$$

Corresponding Author: Nadeem Shafique Butt, College of Statistical and Actuarial Sciences University of the Punjab, Lahore, Pakistan. E-mail: nadeemshafique@hotmail.com & kamal_shahid@hotmail.com.

Where:

ρ_{yz}^2 is squared correlation coefficient between Y and Z.

Jhajj, *et al.* [4] have proposed a family of estimators in single and two phase sampling using information on auxiliary attributes. The variance of the proposed family depends upon the point bi-serial correlation coefficient. Samiuddin and Hanif [5] have also proposed several estimators in single and two phase sampling. A regression-in-ratio estimator proposed by Samiuddin and Hanif [5] is:

$$\bar{y}_{sh(2)} = \left[\bar{y}_2 + \beta_{yz}\left(\bar{z}_1 - \bar{z}_2\right)\right]\frac{\bar{X}}{\bar{x}_2}. \qquad (1.9)$$

The variance of (1.9) is:

$$MSE\left(\bar{y}_{sh(2)}\right) \approx \bar{Y}^2\left[\theta_2\left\{C_y^2\left(1-\rho_{xy}^2\right)+\left(C_x - C_y\,\rho_{xy}\right)^2\right\}\right.$$
$$\left.+\left(\theta_2-\theta_1\right)\left\{C_x^2\,\rho_{xz}^2-\left(C_y\rho_{yz}-C_x\rho_{xz}\right)^2\right\}\right]. \qquad (1.10)$$

In this paper we have proposed a modified regression type estimator using information on several auxiliary variables.

Notations: In this section we define the notations used for the development of the estimator and its variance. Let **w** be a vector of auxiliary variables with covariance matrix Sw, X be another auxiliary variable and Y be the variable of interest.

Let s_{xw} be the vector of covariances between X and w, s_{yw} be the vector of covariances between Y and w. Using these notations we define $\alpha = S_w^{-1}s_{xw}$ as vector of regression coefficients between X and w, and $\gamma = S_w^{-1}s_{yw}$ as vector of regression coefficients between Y and w. We also define $\beta_{yx.w} = S_{xy.w}/S_{2.w}$ as partial regression coefficient between Y and X keeping the w at constant level. Also $S_{yx.\mathbf{w}} = S_{yx} - \mathbf{s}_{x\mathbf{w}}^{/}\mathbf{S}_{\mathbf{w}}^{-1}\mathbf{s}_{x\mathbf{w}}$ is partial covariance between Y and X after removing the effect of w, $s_{yx.\mathbf{w}} = s_{yx} - \mathbf{s}_{x\mathbf{w}}^{/}\mathbf{S}_{\mathbf{w}}^{-1}\mathbf{s}_{x\mathbf{w}}$ is the partial variance of Y and $s_{x.\mathbf{w}}^2 = s_x^2 - \mathbf{s}_{x\mathbf{w}}^{/}\mathbf{S}_{\mathbf{w}}^{-1}\mathbf{s}_{x\mathbf{w}}$ is the partial variance of X. We also define $\rho_{yx.\mathbf{w}}^2 = s_{yx.\mathbf{w}}^2\Big/\left(s_{x.\mathbf{w}}^2 s_{y.\mathbf{w}}^2\right)$ as partial correlation coefficient between Y and X after removing the effect of w, $\rho^2_{x.w}$ as squared multiple correlation coefficient between Y and combined effects of X and w, $\rho^2_{x.w}$ as squared multiple correlation coefficient between Y and combined effects of w.

Using the above notations we proposed the new estimators in the section 3.

The Proposed Estimator: We propose following unbiased estimator of population mean in two phase sampling using information of several auxiliary variables:

$$t_{nss} = \bar{y}_2 + k\left[\bar{x}_1 + \mathbf{a}^{/}\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}_1\right) - \left\{\bar{x}_2 + \mathbf{ß}^{/}\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}_2\right)\right\}\right] \qquad (3.1)$$

Using $\bar{y}_2 = \bar{Y} + \bar{e}_{y_2}, \bar{x}_1 = \bar{X} + \bar{e}_{x_1}, \bar{x}_2 = \bar{X} + \bar{e}_{x_2}, \bar{\mathbf{w}}_1 = \bar{\mathbf{w}} + \bar{\mathbf{e}}_{w_1}$ and $\bar{\mathbf{w}}_2 = \bar{\mathbf{w}} + \bar{\mathbf{e}}_{w_2}$

in (2.1) we have:

$$t_{nss} - \bar{Y} = \bar{e}_{y_2} + k\left[\left(\bar{e}_{x_1} - \bar{e}_{x_2}\right) - \mathbf{a}^{/}\bar{\mathbf{e}}_{w_1} + \mathbf{ß}^{/}\bar{\mathbf{e}}_{w_2}^{/}\right]$$

Squaring and applying expectation, the variance of (3.1) is given as:

$$S = Var(t_{nss}) = \theta_2 s_y^2 + k^2\left[(\theta_2-\theta_1)s_x^2 + \theta_1\mathbf{a}^{/}\mathbf{S}_{\mathbf{w}}\mathbf{a} + \theta_2\mathbf{ß}^{/}\right.$$
$$\left.\mathbf{S}_{\mathbf{w}}\mathbf{ß} + 2(\theta_1-\theta_2)\mathbf{ß}^{/}\mathbf{s}_{x\mathbf{w}} - 2\theta_1\mathbf{a}^{/}\mathbf{S}_{\mathbf{w}}\mathbf{ß}\right] +$$
$$2k\left[(\theta_1-\theta_2)s_{yx} - \theta_1\mathbf{a}^{/}\mathbf{s}_{yw} + \theta_2\mathbf{ß}^{/}\mathbf{s}_{y\mathbf{w}}\right] \qquad (3.2)$$

The optimum values of α, β and k are obtained by minimizing (3.2). These values are obtained by solving following three equations, obtained by partially differentiating (3.2) and setting the derivative to zero.

$$2k\left[(\theta_2-\theta_1)s_x^2 + \theta_1\mathbf{a}^{/}\mathbf{S}_{\mathbf{w}}\mathbf{a} + \theta_2\mathbf{ß}^{/}\mathbf{S}_{\mathbf{w}}\mathbf{ß} + 2(\theta_1-\theta_2)\mathbf{ß}^{/}\mathbf{s}_{x\mathbf{w}} - 2\theta_1\mathbf{a}^{/}\mathbf{S}_{\mathbf{w}}\mathbf{ß}\right]$$
$$+2\left[(\theta_1-\theta_2)s_{yx} - \theta_1\mathbf{a}^{/}\mathbf{s}_{y\mathbf{w}} + \theta_2\mathbf{ß}^{/}\mathbf{s}_{y\mathbf{w}}\right] = 0 \qquad \text{(I)}$$

$$k\mathbf{S}_{\mathbf{w}}(\mathbf{a} - \mathbf{ß}) - \mathbf{s}_{y\mathbf{w}} = 0 \qquad \text{(ii)}$$

$$k\mathbf{S}_w(\theta_2\mathbf{ß} - \theta_1\mathbf{a}) - k(\theta_2-\theta_1)\mathbf{s}_{x\mathbf{w}} + \theta_2\mathbf{s}_{y\mathbf{w}} = \mathbf{0} \qquad \text{(iii)}$$

Solving the above equations simultaneously, the optimum values of α,β and k are:

$$\mathbf{a} = S_{\mathbf{w}}^{-1}s_{x\mathbf{w}},\ \mathbf{ß} = \mathbf{a} - k^{-1}\gamma \text{ and } k = \beta_{yx.\mathbf{w}} = S_{yx.\mathbf{w}}\Big/S_{x.\mathbf{w}}^2$$

Using the optimum values in (3.2) and simplifying, the variance of proposed estimator is:

$$Var(t_{nss}) = s_{y.\mathbf{w}}^2\left[\theta_2\left(1-\rho_{xy.\mathbf{w}}^2\right) + \theta_1\rho_{xy.\mathbf{w}}^2\right] \qquad (3.3)$$

Further, by using the fact that $s^2_{y.w} = S^2_y(1 - \rho^2_{xy.w})$ and utilizing the relationship that $1 - \rho^2_{y.xw} = (1 - \rho^2_{y.w})(1 - \rho^2_{yx.w})$ the variance of proposed estimator can be written as:

$$Var(t_{nss}) = s_y^2\left\{\theta_2\left(1-\rho^2_{y.x\mathbf{w}}\right)+\theta_1\rho^2_{xy.\mathbf{w}}\left(1-\rho^2_{y.\mathbf{w}}\right)\right\} \quad (3.4)$$

From (3.4) we can see that the variance of (3.1) depends upon the squared multiple and partial correlation coefficients. The estimator and its variance for multiphase sampling can be analogously written from (3.1) and (3.4). Specifically if a sample of size n_h is taken at h^{th} phase and a sample of n_q is taken at q_{th} phase with $n_q<n_h$, the estimator of the population mean is:

$$t_{nss} = \bar{y}_2 + k\left[\bar{x}_h + \mathbf{a}'\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}_h\right) - \left\{\bar{x}_q + \mathbf{ß}'\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}_q\right)\right\}\right] \quad (3.5)$$

The variance of (3.5) can be written from (3.4) as:

$$Var(t_{nss}) = s_y^2\left\{\theta_q\left(1-\rho^2_{y.x\mathbf{w}}\right)+\theta_h\rho^2_{xy.\mathbf{w}}\left(1-\rho^2_{y.\mathbf{w}}\right)\right\} \quad (3.6)$$

For practical applicability, the proposed estimator can be easily modified by using the sample estimates in place of population parameters. The consistent estimate of population mean can be straight-away written as:

$$t_{nss} = \bar{y}_2 + b_{yx.\mathbf{w}}(\bar{x}_1 - \bar{x}_2) + b_{yx.\mathbf{w}}\mathbf{b}_{x\mathbf{w}}(\bar{\mathbf{w}}_2 - \bar{\mathbf{w}}_1) + \mathbf{b}'_{y\mathbf{w}}(\bar{\mathbf{w}} - \bar{\mathbf{w}}_2) \quad (3.7)$$

The estimated standard error of (3.1) is given as:

$$S.E(t_{nss}) = s_y\sqrt{\theta_2\left(1-r^2_{y.x\mathbf{w}}\right)+\theta_1 r^2_{yx.\mathbf{w}}\left(1-r^2_{y.\mathbf{w}}\right)} \quad (3.8)$$

Using (3.7) and (3.8), the confidence interval for true population mean can be constructed.

Comparison with Available Estimators: Ahmed [2] has proposed various estimators for two phase and multiphase sampling using information son several auxiliary variables. We have compared the estimator (3.1) with following estimator given in Ahmed [2]:

$$t_1 = y_2 + \sum_{i=1}^{r}\alpha_i\left(\bar{W}_i - \bar{w}_{i1}\right) + \sum_{i=1}^{r}\beta_i\left(\bar{W}_i - \bar{w}_{i2}\right) + \sum_{i=r+1}^{p}\beta_i\left(\bar{w}_{i1} - \bar{w}_{i2}\right)$$

The variance of above estimator is:

$$Var(t_1) = S_y^2\left[\theta_2\left(1-\rho^2_{y.\mathbf{w}}\right)+\theta_1\left(\rho^2_{y.\mathbf{w}}-\rho^2_{y.\mathbf{w}_1}\right)\right] \quad (4.1)$$

Where:

$\rho^2_{y.w}$ is squared multiple correlation between Y and combined effect of all auxiliary variables and $\rho^2_{y.w1}$ is the squared multiple correlation between Y and first r auxiliary variables. Now comparing (3.4) with (4.1) gives:

$$Var(t_1) - Var(t_{nss}) = \theta_2\left(\rho^2_{y.x\mathbf{w}} - \rho^2_{y.\mathbf{w}}\right) + \theta_1\left[\rho^2_{y.\mathbf{w}}\left(1-\rho^2_{yx.\mathbf{w}}\right) - \rho^2_{yx.\mathbf{w}} - \rho^2_{y.\mathbf{w}_1}\right] > 0 \quad (4.2)$$

From (4.2) we can readily see that the proposed estimator performs well as compared with the estimator proposed by Ahmed [2].

REFERENCES

1. Hansen, M.H., Hurwitz, W.N., and W.G. Madow, 1953. Sample Survey Methods and Theory (Vol. II): John Wiley.
2. Ahmed, Z., 2008. Generalized Ratio and Regression Estimators in Muliphase Sampling. Unpublished PhD thesis.
3. Sahoo, J., L.N. Sahoo and S. Mohanty, 1993. A regression approach to estimation in two phase sampling using two auxiliary variables. Current Sciences, 65(1): 73-75.
4. Jhajj, H.S., M.K. Sharma and L.K. Grover, 2006. A familty of estimators of population mean using information of auxiliary attribute. Pak. J. Stat., 22(1): 43-50.
5. Samiuddin, M. and M. Hanif, 2007. Estimation of population mean in single and two phase sampling with or without additional information. Pak. J. Stat., 23(2): 99-118.

Estimation of Population Mean in Two Phase Sampling using Attribute Auxiliary Information

Nadeem Shafique Butt[1], Muhammad Qaiser Shahbaz[2]

[1] College of Statistical and Actuarial Sciences, University of the Punjab, Pakistan
nadeemshafique@hotmail.com

[2] Department of Mathematics, COMSATS Institute of Information Technology, Pakistan
qshahbaz@gmail.com

Abstract. A new estimator for population mean has been proposed in two phase sampling by using information of multiple auxiliary attributes. The minimum variance of the proposed estimator has been obtained.

1. Introduction

The auxiliary information has always been a source of improvement in estimation of certain population characteristics. Several estimators have been developed in single and two phase sampling which utilizes information on auxiliary variables as well as auxiliary attributes. The classical estimators which use information on auxiliary variables are the ratio and regression estimators as given in Hansen, Hurwitz, & Madow (1953). The classical regression estimator of population mean is given as:

$$\bar{y}_{lr} = \bar{y} + \beta\left(\bar{X} - \bar{x}\right) \tag{1.1}$$

The value of β for which the mean square error of (1.1) is minimum is $\beta = \dfrac{S_{xy}}{S_x^2}$. The minimum mean square error of (1.1) is given as

$$MSE(\bar{y}_{lr}) = \theta S_y^2\left(1 - \rho_{yx}^2\right) \tag{1.2}$$

Where $\theta = n^{-1} - N^{-1}$ and ρ_{yx} is the correlation coefficient between X and Y. The estimator (1.1) in case of auxiliary attribute is discussed by Naik & Gupta (1996), and the estimator in this case is given as:

$$t_{1(1)} = \bar{y} + b\left(p_1 - P_1\right) \tag{1.3}$$

where P_1 is sample proportion for auxiliary variables. The mean square error of (1.3) is:

$$MSE\left(t_{1(1)}\right) = \theta\left(1 - \rho_{Pb_1}^2\right)S_y^2 \tag{1.4}$$

where $\rho^2_{Pb_1}$ is the squared point bi-serial correlation coefficient. Jhajj, Sharma, & Grover (2006) has proposed a family of estimators in single and two phase sampling using information on a single auxiliary attributes. The proposed family is based upon a general function and is given as:

$$t_{2(1)} = g_\omega\left(\bar{y}, v_1\right) \tag{1.5}$$

where $v_1 = \frac{p_1}{P_1}$ and $g_\omega\left(\bar{y}, v_1\right)$ is a parametric function of $\bar{y}$ and v_1 such that $g_\omega\left(\bar{Y},1\right) = \bar{Y}$, for all $\bar{Y}$.

The mean square error of each estimator; to the terms of order $1/n$; of this family is,

$$MSE\left(T_{2(1)}\right) \approx \theta\left(1-\rho^2_{Pb_1}\right)S^2_y . \tag{1.6}$$

The mean square error of the proposed family depends upon the point bi-serial correlation coefficient.

Shabbir & Gupta (2007) have also proposed an estimator for population mean in single phase sampling using information of single auxiliary attribute. The estimator is given as:

$$t_{3(1)} = \left[d_1\bar{y} + d_2\left(P_1 - p_1\right)\right]\frac{P_1}{p_1}, \text{ for } p_1 > 0 \tag{1.7}$$

where d_1 and d_2 are unknown constants. The mean square error of (1.7) is:

$$MSE\left(t_{3(1)}\right) \approx \frac{\theta\left(1-\rho^2_{Pb_1}\right)S^2_y}{1+\theta\left(1-\rho^2_{Pb_1}\right)C^2_y} \tag{1.8}$$

In this paper we have proposed a modified regression type estimator using information on several auxiliary attributes.

2. Notations

In this section we define the notations used for the development of the estimator and its variance. Let $\boldsymbol{\delta}$ be a vector of auxiliary attributes with covariance matrix $\mathbf{S}_{\boldsymbol{\delta}}$, τ be another auxiliary attribute and Y be the variable of interest.

Let $\mathbf{s}_{\tau\boldsymbol{\delta}}$ be the vector of covariances between τ and $\boldsymbol{\delta}$, $\mathbf{s}_{y\boldsymbol{\delta}}$ be the vector of covariances between Y and $\boldsymbol{\delta}$. Using these notations we define $\boldsymbol{\gamma} = \mathbf{S}^{-1}_{\boldsymbol{\delta}}\mathbf{s}_{\tau\boldsymbol{\delta}}$ as

vector of regression coefficients between τ and $\boldsymbol{\delta}$, and $\boldsymbol{\gamma} = \mathbf{S}_{\boldsymbol{\delta}}^{-1}\mathbf{s}_{y\boldsymbol{\delta}}$ as vector of regression coefficients between Y and $\boldsymbol{\delta}$. We also define $\beta_{y\tau.\boldsymbol{\delta}} = S_{\tau y.\boldsymbol{\delta}}/S_{\tau.\boldsymbol{\delta}}^2$ as partial regression coefficient between Y and τ keeping the $\boldsymbol{\delta}$ at constant level. Also $S_{y\tau.\boldsymbol{\delta}} = S_{y\tau} - \mathbf{s}_{\tau\boldsymbol{\delta}}^{/}\mathbf{S}_{\boldsymbol{\delta}}^{-1}\mathbf{s}_{\tau\boldsymbol{\delta}}$ is partial covariance between Y and τ after removing the effect of $\boldsymbol{\delta}$, $\mathbf{S}_{y.\boldsymbol{\delta}}^2 = S_y^2 - \mathbf{s}_{\tau\boldsymbol{\delta}}^{/}\mathbf{S}_{\boldsymbol{\delta}}^{-1}\mathbf{s}_{\tau\boldsymbol{\delta}}$ is the partial variance of Y, and $S_{\tau.\boldsymbol{\delta}}^2 = S_\tau^2 - \mathbf{s}_{\tau\boldsymbol{\delta}}^{/}\mathbf{S}_{\boldsymbol{\delta}}^{-1}\mathbf{s}_{\tau\boldsymbol{\delta}}$ is the partial variance of τ. We also define $\rho_{y\tau.\boldsymbol{\delta}}^2 = S_{y\tau}^2 / \left(S_{\tau.\boldsymbol{\delta}}^2 S_{y.\boldsymbol{\delta}}^2\right)$ as partial correlation coefficient between Y and τ after removing the effect of $\boldsymbol{\delta}$, $\rho_{y\tau.\boldsymbol{\delta}}^2$ as squared multiple bi-serial correlation coefficient between Y and combined effects of τ and $\boldsymbol{\delta}$, $\rho_{y.\boldsymbol{\delta}}^2$ as squared multiple correlation coefficient between Y and combined effects of $\boldsymbol{\delta}$.
Using the above notations we proposed the new estimators in the section 3.

3. The Proposed Estimator

We propose following unbiased estimator of population mean in two phase sampling using information of several auxiliary attributes:

$$t_{nss(A)} = \bar{y}_2 + k\left[p_{\tau_1} + \boldsymbol{\gamma}^{/}\left(\mathbf{p}_\delta - \mathbf{p}_{\delta_1}\right) - \left\{p_{\tau_2} + \boldsymbol{\eta}^{/}\left(\mathbf{p}_\delta - \mathbf{p}_{\delta_2}\right)\right\}\right] \qquad (3.1)$$

Using
$\bar{y}_2 = \bar{Y} + \bar{e}_{y_2}$, $p_{\tau_1} = p_\tau + \bar{e}_{\tau_1}$, $p_{\tau_2} = p_\tau + \bar{e}_{\tau_2}$, $\mathbf{p}_{\delta_1} = \mathbf{p}_\delta + \bar{\mathbf{e}}_{\delta_1}$ and $\mathbf{p}_{\delta_2} = \mathbf{p}_\delta + \bar{\mathbf{e}}_{\delta_2}$ in (3.1) we have:

$$t_{nss(A)} - \bar{Y} = \bar{e}_{y_2} + k\left[\left(\bar{e}_{\tau_1} - \bar{e}_{\tau_2}\right) - \boldsymbol{\gamma}^{/}\bar{\mathbf{e}}_{\delta_1} + \boldsymbol{\eta}^{/}\bar{\mathbf{e}}_{\delta_2}^{/}\right]$$

Squaring and applying expectation, the mean square error of (3.1) is given as:

$$\begin{aligned} S = MSE\left(t_{nss}\right) = \theta_2 s_y^2 + k^2\Big[\left(\theta_2 - \theta_1\right)s_\tau^2 + \theta_1\boldsymbol{\gamma}^{/}\mathbf{S}_{\boldsymbol{\delta}}\boldsymbol{\gamma} + \theta_2\boldsymbol{\eta}^{/}\mathbf{S}_{\boldsymbol{\delta}}\boldsymbol{\eta} + 2\left(\theta_1 - \theta_2\right)\boldsymbol{\eta}^{/}\mathbf{s}_{\tau\boldsymbol{\delta}} \\ -2\theta_1\boldsymbol{\gamma}^{/}\mathbf{S}_{\boldsymbol{\delta}}\boldsymbol{\eta}\Big] + 2k\left[\left(\theta_1 - \theta_2\right)s_{y\tau} - \theta_1\boldsymbol{\gamma}^{/}\mathbf{s}_{y\boldsymbol{\delta}} + \theta_2\boldsymbol{\eta}^{/}\mathbf{s}_{y\boldsymbol{\delta}}\right] \end{aligned} \qquad (3.2)$$

The optimum values of $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ and k are obtained by minimizing (3.2). These values are obtained by solving following three equations, obtained by partially differentiating (3.2) and setting the derivative to zero

$$2k\left[\left(\theta_2-\theta_1\right)s_\tau^2+\theta_1\boldsymbol{\gamma}'\mathbf{S}_{\boldsymbol{\delta}}\boldsymbol{\gamma}+\theta_2\boldsymbol{\eta}'\mathbf{S}_{\boldsymbol{\delta}}\boldsymbol{\eta}+2\left(\theta_1-\theta_2\right)\boldsymbol{\eta}'\mathbf{s}_{\tau\boldsymbol{\delta}}-2\theta_1\boldsymbol{\gamma}'\mathbf{S}_{\boldsymbol{\delta}}\boldsymbol{\eta}\right] \quad (1)$$
$$+2\left[\left(\theta_1-\theta_2\right)s_{y\tau}-\theta_1\boldsymbol{\gamma}'\mathbf{s}_{y\boldsymbol{\delta}}+\theta_2\boldsymbol{\eta}'\mathbf{s}_{y\boldsymbol{\delta}}\right]=0$$

$$k\mathbf{S}_{\boldsymbol{\delta}}\left(\boldsymbol{\gamma}-\boldsymbol{\eta}\right)-\mathbf{s}_{y\boldsymbol{\delta}}=0$$
(2)

$$k\mathbf{S}_{\boldsymbol{\delta}}\left(\theta_2\boldsymbol{\eta}-\theta_1\boldsymbol{\gamma}\right)-k\left(\theta_2-\theta_1\right)\mathbf{s}_{\tau\boldsymbol{\delta}}+\theta_2\mathbf{s}_{y\boldsymbol{\delta}}=\mathbf{0}$$
(3)

Solving the above equations simultaneously, the optimum values of $\boldsymbol{\alpha},\boldsymbol{\beta}$ and k are:

$$\boldsymbol{\gamma}=S_{\boldsymbol{\delta}}^{-1}s_{x\boldsymbol{\delta}},\ \boldsymbol{\eta}=\boldsymbol{\gamma}-k^{-1}\boldsymbol{\gamma} \text{ and } k=\boldsymbol{\eta}_{y\tau.\boldsymbol{\delta}}=\frac{s_{y\tau.\boldsymbol{\delta}}}{s_{\tau\tau.\boldsymbol{\delta}}}$$

Using the optimum values in (3.2) and simplifying, the mean square error of proposed estimator is:

$$MSE\left(t_{nss}\right)=s_{y.\boldsymbol{\delta}}^2\left[\theta_2\left(1-\rho_{\tau y.\boldsymbol{\delta}}^2\right)+\theta_1\rho_{\tau y.\boldsymbol{\delta}}^2\right] \quad (3.3)$$

Further, by using the fact that $s_{y.\boldsymbol{\delta}}^2=S_y^2\left(1-\rho_{\tau y.\boldsymbol{\delta}}^2\right)$ and utilizing the relationship that $1-\rho_{y.\tau\boldsymbol{\delta}}^2=\left(1-\rho_{y.\boldsymbol{\delta}}^2\right)\left(1-\rho_{y\tau.\boldsymbol{\delta}}^2\right)$, the mean square error of proposed estimator can be written as:

$$MSE\left(t_{nss}\right)=s_y^2\left\{\theta_2\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1\rho_{\tau y.\boldsymbol{\delta}}^2\left(1-\rho_{y.\boldsymbol{\delta}}^2\right)\right\} \quad (3.4)$$

From (3.4) we can see that the mean square error of (3.1) depends upon the squared multiple and partial correlation coefficients. The estimator and its mean square error for multiphase sampling can be analogously written from (3.1) and (3.4). Specifically if a sample of size n_h is taken at h^{th} phase and a sample of n_q is taken at q^{th} phase with $n_q < n_h$, the estimator of the population mean is:

$$t_{nss(A)}=\bar{y}_2+k\left[p_{\tau_h}+\boldsymbol{\gamma}'\left(\mathbf{p}_\delta-\mathbf{p}_{\delta_h}\right)-\left\{p_{\tau q}+\boldsymbol{\eta}'\left(\mathbf{p}_\delta-\mathbf{p}_{\delta_q}\right)\right\}\right] \quad (3.5)$$

The mean square error of (3.5) can be written from (3.4) as:

$$MSE\left(t_{nss}\right)=s_y^2\left\{\theta_q\left(1-\rho_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_h\rho_{\tau y.\boldsymbol{\delta}}^2\left(1-\rho_{y.\boldsymbol{\delta}}^2\right)\right\} \quad (3.6)$$

For practical applicability, the proposed estimator can be easily modified by using the sample estimates in place of population parameters. The consistent estimate of population mean can be straight-away written as:

$$t_{nss}=\bar{y}_2+b_{y\tau.\boldsymbol{\delta}}\left(p_{\tau_1}-p_{\tau_2}\right)+b_{y\tau.\boldsymbol{\delta}}\mathbf{b}_{\tau\boldsymbol{\delta}}\left(\mathbf{p}_{\delta_2}-\mathbf{p}_{\delta_1}\right)+\mathbf{b}'_{y\boldsymbol{\delta}}\left(\mathbf{p}_\delta-\mathbf{p}_{\delta_2}\right) \quad (3.7)$$

The estimated standard error of (3.1) is given as:

$$S.E\left(t_{nss}\right)=s_y\sqrt{\theta_2\left(1-r_{y.\tau\boldsymbol{\delta}}^2\right)+\theta_1 r_{y\tau.\boldsymbol{\delta}}^2\left(1-r_{y.\boldsymbol{\delta}}^2\right)} \quad (3.8)$$

Using (3.7) and (3.8), the confidence interval for true population mean can be constructed.

References:

1. Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample Survey Methods and Theory* (Vol. II): John Wiley.
2. Jhajj, H. S., Sharma, M. K., & Grover, L. K. (2006). A familty of estimators of population mean using information of auxiliary attribute. *Pak. J. Stat., 22*(1), 43-50.
3. Naik, V., & Gupta, P. (1996). A note on estimation of mean with known population proportion of an auxiliary character. *Jour. Ind. Soc. Agr. Stat, 48*(2), 151-158.
4. Shabbir, J., & Gupta, S. (2007). On estimating the finite population mean with known population proportion of an auxiliary variable. *Pak. J. Statist., 23*(1), 1-9.