Global Journal of Science Frontier Research
Mathematics and Decision Sciences
Volume 12 Issue 14 Version 1.0 Year 2012
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

A Class of Improved Estimators for Estimating Population Mean Regarding Partial Information in Double Sampling

By Hina Khan, Saleha Shouket & Aamir Sanaullah
GC University, Lahore, Pakistan

Abstract - In this paper a class of improved estimators has been proposed for estimating population mean in two phase (double) sampling when only partial information is available on either of two auxiliary variables. Under simple random sampling (SRWOR), expressions of mean square error and bias have been derived to make comparison of suggested class with wide range of other estimators. Empirical study has also been given using five different natural populations. Empirical study confirmed that the suggested class of improved estimators is more efficient under percent relative efficiency (PRE) criterion.

Keywords : *double sampling, auxiliary variable, partial information, bias, mean square error.*

GJSFR-F Classification : *MSC 2010 : 62F12 , 06A06*

A CLASS OF IMPROVED ESTIMATORS FOR ESTIMATING POPULATION MEAN REGARDING PARTIAL INFORMATION IN DOUBLE SAMPLING

Strictly as per the compliance and regulations of :

Notes

A Class of Improved Estimators for Estimating Population Mean Regarding Partial Information in Double Sampling

Hina Khan [α], Saleha Shouket [σ] & Aamir Sanaullah [ρ]

Abstract - In this paper a class of improved estimators has been proposed for estimating population mean in two phase (double) sampling when only partial information is available on either of two auxiliary variables. Under simple random sampling *(SRWOR)*, expressions of mean square error and bias have been derived to make comparison of suggested class with wide range of other estimators. Empirical study has also been given using five different natural populations. Empirical study confirmed that the suggested class of improved estimators is more efficient under percent relative efficiency (PRE) criterion.

Keywords : *double sampling, auxiliary variable, partial information, bias, mean square error.*

I. Introduction

The history of use of auxiliary information in survey sampling is as old as the history of survey sampling. Bowley (1926) and Neyman (1934, 38) provide foundation stones of modern sampling theory, dealing with stratified random sampling. Hansen and Hurwitz (1943) firstly use auxiliary information in selecting sample with varying probabilities. Snedecor and King (1942), Spurr (1952), Freese (1962), Unnikrishan and Kunte (1995), Armstrong and St-Jean (1994) provide applications of two phase (or double) sampling procedure.

The estimation of the population mean is an unrelenting issue in sampling theory and several efforts have been made to improve the precision of the estimators in the presence of multi-auxiliary variables. Mohanty (1967) suggested regression cum ratio estimator in double sampling using two auxiliary variables. Tripathi (1970) and Das (1988) describe the auxiliary information in four ways. Das and Tripathi (1978) initiate to use population variance of auxiliary variable for estimating the population variance. Srivastava and Jhajj (1980) also consider the use of population mean and variance of auxiliary variable for estimating population variance of the study variable. Several other authors have also used information on the parameters of auxiliary variable to find more precise estimates. Regarding the use of information on C_x, $\bar{Z}$, σ_z, $\beta_1(z)$, and $\beta_2(z)$ the researcher may be referred to Sear (1964), Singh et al. (1973), Sen(1978), Singh (2001), Uphadhyaya and Singh (2001), Singh et al. (2006), Singh et al (2007), and Singh et al. (2011).

Following Chand (1975) and Kiregyera (1980, 1984), Sahoo and Sahoo (1993) and Sahoo et al. (1994) have discussed a general frame work of estimation by using an

Author α σ : GC University, Lahore, Pakistan.
Author ρ : National College of Business Administration and Economics, Lahore, Pakistan.

additional auxiliary variable for double sampling when the population mean of the main auxiliary variable is unknown. Kiregyera (1984) developed two estimators, one is ratio-in-regression and other is regression-in-regression estimator. Mukerjee et al. (1987) developed three estimators following Kiregyera's (1984) technique. Sahoo's (1993) class of estimators covered a large number of estimators. Roy (2003) constructed a regression-type estimator of population mean of the main variable in the presence of available knowledge on second auxiliary variable, when the population mean of the first auxiliary variable was not known. Samiuddin and Hanif (2007) have reported three different methods of estimation in double sampling. These methods are proposed depending whether information of auxiliary variables is available or not at first phase of sampling. Singh et al. (2011) proposed chain ratio type estimator for population mean using some known values of population parameters of secondary auxiliary variable.

Let $U = (U_1, U_2, \ldots\ldots, U_n)$ be a finite population consisting of N units. Let y and (x, z) be the variate of interest and auxiliary characteristics respectively related to y assume real non-negative i^{th} value (y_i, x_i, z_i) $i = 1, 2, \ldots\ldots, N$ with population means $\bar{Y}$, $\bar{X}$, and $\bar{Z}$ respectively. Let a simple random sample without replacement *(SRSWOR)* is drawn in each phase, the two phase (or double) sampling scheme is as follows:

i. The first phase sample S_1 $(S_1 \subset U)$ of size n_1 is drawn to measure x and z say (x_1, z_1).
ii. The second phase sample S_2 $(S_2 \subset S_1)$ of size n_2 $(n_2$ ¡ $n_1)$ is drawn from the first phase sample S to measure y say y_2.

Let $$\bar{x}_1 = \frac{1}{n_1}\sum_{i\in s_1} x_i, \quad \bar{x}_2 = \frac{1}{n_2}\sum_{i\in s_2} x_i, \quad \bar{z}_1 = \frac{1}{n_1}\sum_{i\in s_1} z_i, \quad \bar{z}_2 = \frac{1}{n_2}\sum_{i\in s_2} z_i, \text{ and } \bar{y}_2 = \frac{1}{n_2}\sum_{i\in s_2} y_i$$

For a *SRSWOR*, we have some assumptions as following,

$$\left.\begin{array}{lll}
\bar{y}_1 = \bar{Y}(1+e_{\bar{y}_1}), & \bar{x}_1 = \bar{X}(1+e_{\bar{x}_1}), & \bar{z}_1 = \bar{Z}(1+e_{\bar{z}_1}) \\
\bar{y}_2 = \bar{Y}(1+e_{\bar{y}_2}), & \bar{x}_2 = \bar{X}(1+e_{\bar{x}_2}), \quad \text{and} & \bar{z}_2 = \bar{Z}(1+e_{\bar{z}_2}) \\
E(e_{\bar{y}_2}) = E(e_{\bar{z}_2}) = E(e_{\bar{z}_1}) = E(e_{\bar{x}_1}) = 0 & & \\
E(e_{\bar{y}_2})^2 = \theta_2 C^2{}_y & E(e_{\bar{z}_2})^2 = \theta_2 C_z^2 & \\
E(e_{\bar{z}_1})^2 = \theta_1 C_z^2, & E(e_{\bar{x}_1})^2 = \theta_1 C_x^2 & \\
E(e_{\bar{y}_2} e_{\bar{z}_2}) = \theta_2 C_y C_z \rho_{yz}, & E(e_{\bar{z}_2} e_{\bar{z}_1}) = \theta_1 C_z^2 & \\
E(e_{\bar{z}_1} e_{\bar{x}_1}) = \theta_1 C_z C_x \rho_{zx} & E(e_{\bar{y}_2} e_{\bar{x}_1}) = \theta_2 C_y C_x \rho_{yx} & \\
\theta_1 = \dfrac{1}{n_1} - \dfrac{1}{N} & \theta_2 = \dfrac{1}{n_2} - \dfrac{1}{N} &
\end{array}\right\} \quad (1.1)$$

In many practical situations even if $\bar{X}$ is unknown, information on a secondary auxiliary variable z, closely related to x but compared to x remotely related to y, is readily available on all units of population such that z_i denotes its value on i^{th} unit and $\bar{Z}$

as its known mean see Singh et al. (2004) and Singh et al.(2006). For instance, if the elements of population are hospitals, and y_i, x_i and z_i are respectively the number of deaths, number of patients admitted and number of available beds relating to the i^{th} hospital, then information on z_i's can be collected easily from the official records of the Health Department. This situation has also been discussed by chand (1975), Mukhergee et al. (1987), Sahoo and Sahoo (1993), Roy (2003) and among many others.

Notes

II. Some Available Estimators

In this section we reproduce some well known ratio type estimators for the population mean available for double sampling under *SRWOR* regarding only partial information are available.

1 The variance of the usual unbiased estimator $\bar{y}$ under SRSWOR scheme is as;

$$Var(T_1) = \theta \bar{Y}^2 C^2{}_y \tag{2.1}$$

2 Mohanty (1967) regression to ratio estimator

$$T_2 = [\bar{y}_2 + b_{yx}(\bar{x}_1 - \bar{x}_2)]\frac{\bar{Z}}{\bar{z}_2} \tag{2.2}$$

$$MSE(T_2) = \bar{Y}^2\left[\theta_2\left(C_y^2 + C_z^2 - C_y^2\rho_{xy}^2 - 2C_yC_z\rho_{yz} + 2C_yC_z\rho_{xy}\rho_{xz}\right) + \theta_1\left(C_y^2\rho_{xy}^2 - 2C_yC_z\rho_{xy}\rho_{xz}\right)\right] \tag{2.3}$$

3 Chand (1975) chain ratio estimator

$$T_3 = \bar{y}_2\frac{\bar{x}_2}{\bar{x}_1}\frac{\bar{z}_1}{\bar{Z}} \tag{2.4}$$

$$MSE(T_3) = \bar{Y}^2\left[\theta_2 C_y^2 + (\theta_2 - \theta_1)\left((C_x + \rho_{xy}C_y)^2 - C_y^2\rho_{xy}^2\right) + \theta_1\left((C_z + \rho_{yz}C_z)^2 - C_z^2\rho_{yz}^2\right)\right] \tag{2.5}$$

4 Kiregyera (1984) regression in regression estimator

$$T_4 = \bar{y}_2 + b_{yx}\left[(\bar{x}_1 - \bar{x}_2) - b_{xz}(\bar{z}_1 - \bar{Z})\right] \tag{2.6}$$

$$MSE(T_4) = \bar{Y}^2 C_y^2\left[\theta_2 - (\theta_2 - \theta_1)\rho_{xy}^2 - \theta_1\rho_{yz}^2 + \theta_1(\rho_{yz} - \rho_{xy}\rho_{xz})^2\right] \tag{2.7}$$

5 Bedi (1985) ratio estimator

$$T_5 = \bar{y}_2\left[\frac{\bar{z}_1}{\bar{z}_2}\right]^{\alpha} \tag{2.8}$$

$$MSE(T_5) = \bar{Y}^2 C_y^2\left[\theta_2 - (\theta_2 - \theta_1)\rho_{yz}^2\right] \tag{2.9}$$

6 Mukhergee et at (1987) regression estimator

$$T_6 = \bar{y}_2 + b_{yx}(\bar{x}_1 - \bar{x}_2) + b_{yx}b_{xz}(\bar{Z} - \bar{z}_1) + b_{yz}(\bar{Z} - \bar{z}_2) \tag{2.10}$$

$$MSE(T_6) = \bar{Y}^2 C_y^2 \left[\theta_1(\rho_{yz} - \rho_{xy}\rho_{yz})^2 + \theta_2(1 - \rho_{yz}^2 - \rho_{xy}^2 + 2\rho_{xy}\rho_{yz}\rho_{zx})\right] \tag{2.11}$$

7 Srivastava et al (1990) ratio estimator

$$T_7 = \bar{y}_2 \left[\frac{\bar{x}_1}{\bar{x}_2}\right]^{\alpha_1} \left[\frac{\bar{Z}}{\bar{z}_1}\right]^{\alpha_2} \tag{2.12}$$

$$MSE(T_7) = \bar{Y}^2 C_y^2 \left[\theta_2 - (\theta_2 - \theta_1)\rho_{xy}^2 - \theta_1 \rho_{yz}^2\right] \tag{2.13}$$

8 Sahoo et al (1994a) regression in regression estimator

$$T_8 = \bar{y}_2 + b_{yx}(\bar{x}_1 - \bar{x}_2) + b_{yx}b_{xz}(\bar{z}_1 - \bar{z}_2) + b_{yx}b_{xz}(\bar{Z} - \bar{z}_1) \tag{2.14}$$

$$MSE(T_8) = \bar{Y}^2 C_y^2 \left[\theta_2 + \theta_1 \rho_{xy}^2 \rho_{xz}^2 - (\theta_2 - \theta_1)\left(\rho_{xy}^2(1 - \rho_{xz}^2) - 2\rho_{xy}\rho_{xz}\rho_{yz}\right)\right] \tag{2.15}$$

9 Singh (2001) chain ratio type estimator:

$$T_9 = \bar{y}_2 \left[\frac{\bar{x}_1}{\bar{x}_2}\right] \left[\frac{\alpha\bar{Z} + \sigma_z}{\alpha\bar{z}_1 + \sigma_z}\right]^g \tag{2.16}$$

$$MSE(T_9) = \bar{Y}^2 \left[\theta_2 C_y^2 + (\theta_2 - \theta_1)\left(C_x^2 - 2C_y C_x \rho_{xy}\right) - \theta_1 C_y^2 \rho_{yz}^2\right] \tag{2.17}$$

10 Singh et al. (2004) generalized estimator:

$$T_{10} = \bar{y}_2 \left[\frac{\bar{x}_1}{\bar{x}_2}\right]^{\alpha_1} \left[\frac{a\bar{Z} + b}{a\bar{z}_1 + b}\right]^{\alpha_2} \left[\frac{a\bar{Z} + b}{a\bar{z}_2 + b}\right]^{\alpha_3} \tag{2.18}$$

$$MSE(T_{10}) = \bar{Y}^2 C_y^2 \left[\theta_2 - \theta_1 \rho_{yz}^2 - (\theta_2 - \theta_1)\rho_{y.xz}^2\right] \tag{2.19}$$

11 Samiuddin and Hanif (2006) ratio cum regression estimator:

$$T_{11} = \left[\bar{y}_2 + b_{yz}(\bar{z}_1 - \bar{z}_2)\right] \left[\frac{\bar{X}}{\bar{x}_2}\right] \tag{2.20}$$

$$MSE(T_{11}) = \bar{Y}^2 \left[\theta_2\left(C_y^2(1 - \rho_{yx}^2) + (C_x - C_y\rho_{xy})^2\right) + \theta_3\left(C_x^2\rho_{xz}^2 - (C_y\rho_{yz} - C_x\rho_{xz})^2\right)\right] \tag{2.21}$$

12 Samiuddin and Hanif (2007) chain ratio estimator

$$T_{12} = \bar{y}_2 \left[\frac{\bar{x}_1}{\bar{x}_2}\right]^{\alpha_1} \left[\frac{\bar{z}_1}{\bar{z}_2}\right]^{\alpha_2} \left[\frac{\bar{Z}}{\bar{z}_2}\right]^{\alpha_3} \tag{2.22}$$

Notes

$$MSE(T_{12}) = \bar{Y}^2 C_y^2 \left[\theta_2 (1-\rho_{y.xz}^2) + \theta_1 (1-\rho_{yz}^2)\rho_{yx.z}^2\right] \tag{2.23}$$

13 Singh et al. (2007) general family of ratio estimators

Notes

$$T_{13} = \bar{y}_2 \left[\frac{\bar{x}_1}{\bar{x}_2}\right]^{\alpha_1} \left[\frac{\bar{Z} + \rho_{xz}}{\bar{z}_1 + \rho_{xz}}\right]^{\alpha_2} \tag{2.24}$$

$$MSE(T_{13}) = \bar{Y}^2 C_y^2 \left[\theta_2 - \theta_1 \rho_{yz}^2 - \theta_3 \rho_{yx}^2\right] \tag{2.25}$$

14 H.P.Singh and N.Agnihortie (2008) ratio product estimator

$$T_{13} = \bar{y}_1 \left[\delta\left(\frac{a\bar{X}+b}{a\bar{x}_1+b}\right) + (1-\delta)\left(\frac{a\bar{x}_1+b}{a\bar{X}+b}\right)\right] \tag{2.26}$$

$$\min . MSE(T_{13}) = \theta_1 \bar{Y}^2 C_y^2 \left(1-\rho_{yx}^2\right) \tag{2.27}$$

III. Proposed Estimators

In section-2 mentioned estimators have been widely used in the estimation of population mean in diverse situations regarding partial information. Now following estimators stated above, we have proposed two estimators in this section regarding the availability of partial information. One estimator (section-3.1) has been proposed regarding partial information on main auxiliary variable x and other estimator (section-3.2) have been proposed regarding partial information on secondary variable z for double sampling under *SRWOR*.

a) Proposed estimator in two phase (double) sampling

The new estimator $\hat{\bar{Y}}_1$ has been proposed for two phase sampling using two auxiliary variables regarding partial information on main auxiliary variable x. The estimator has been convinced by Srivastava (1971) and Singh (2001) ratio estimators.

$$\hat{\bar{Y}}_1 = \bar{y}_2 \left(\frac{\bar{x}_1}{\bar{x}_2}\right)^{\alpha} \left(\frac{a\bar{Z}+b}{a\bar{z}_1+b}\right) \tag{3.1}$$

Where α, is an unknown constant whose values is to estimate. $a(\neq 0)$, and b are assumed to be known as either real numbers or (Linear or Non-linear) functions of some known parameters of auxiliary variable z such as standard deviation σ_z, coefficient of variation C_z, skewness $\beta_1(z)$, kurtosis $\beta_2(z)$.

Using the notations given in (1.1), $\hat{\bar{Y}}_1$ is expressed in the form of e's and up to the first degree of approximation as:

$$\hat{\bar{Y}}_1 \approx \bar{Y} + \bar{Y}\left[\alpha\left(e_{\bar{x}_1} - e_{\bar{x}_2}\right) - \omega e_{\bar{z}_1} + e_{\bar{y}_2}\right] \qquad \text{where} \quad \omega = \frac{a\bar{Z}}{a\bar{Z}+b},$$

$$\hat{\bar{Y}}_1 - \bar{Y} \approx \bar{Y}\left[\alpha\left(e_{\bar{x}_1} - e_{\bar{x}_2}\right) - \omega e_{\bar{z}_1} + e_{\bar{y}_2}\right] \tag{3.2}$$

Notes

Taking square and applying expectation, [given in (1.1)], the mean square error of (3.1) is obtained as:

$$MSE(\hat{\bar{Y}}_1) \approx \bar{Y}^2\begin{bmatrix}\theta_2 C_y^2 + \alpha^2(\theta_2-\theta_1)^2 C_x^2 + \theta_1\omega^2 C_z^2 - \\ 2\alpha(\theta_2-\theta_1)C_yC_x\rho_{xy} - 2\omega\theta_1 C_yC_z\rho_{yz}\end{bmatrix} \tag{3.3}$$

Differentiating (3.3) with respect to α and ω and setting equal zero. We have:

$$\alpha = \frac{C_y}{C_x}\rho_{yx} \qquad \textit{and} \qquad \omega = \frac{C_y}{C_z}\rho_{yz}$$

Taking the values of α and ω in equation (3.3), and simplifying. We get *min.MSE*($\hat{\bar{Y}}_1$):

$$\min.MSE(\hat{\bar{Y}}_1) = \bar{Y}^2 C_y^2\left[\theta_2 - \theta_3\rho_{xy}^2 - \theta_1\rho_{xy}^2\right] \qquad \text{where } \theta_3 = \theta_2 - \theta_1 \tag{3.4}$$

In order to derive bias of (3.1), we again use (3.2) upto the 2[nd] order of approximation as:

$$\hat{\bar{Y}}_1 - \bar{Y} \approx \begin{bmatrix}\alpha\left(e_{\bar{x}_1} - e_{\bar{x}_2}\right) - \alpha^2 e_{\bar{x}_1}e_{\bar{x}_2} + \dfrac{\alpha(\alpha-1)}{2!}e_{\bar{x}_1}^2 + \dfrac{\alpha(\alpha+1)}{2!}e_{\bar{x}_2}^2 - \omega e_{\bar{z}_1} - \omega\alpha\left(e_{\bar{x}_1} - e_{\bar{x}_2}\right)e_{\bar{z}_1} \\ + \omega^2 e_{\bar{z}_1}^2 + e_{\bar{y}_2} + \alpha e_{\bar{y}_2}\left(e_{\bar{x}_1} - e_{\bar{x}_2}\right) - \omega e_{\bar{y}_2}e_{\bar{z}_1}\end{bmatrix}$$

After applying expectation and simplifying, the optimum bias of (3.1) is:

$$Bias(\hat{\bar{Y}}_1) = \bar{Y}\left[-\frac{\theta_3}{2}C_y^2\rho_{yx}^2 + \frac{\theta_3}{2}C_x\left(C_y\rho_{xy} - \frac{C_x}{4}\right) - \frac{\theta_1}{4}C_y^2\rho_{yz}^2\right] \; \textit{where}\; \theta_3 = \theta_2 - \theta_1 \tag{3.5}$$

i. *Deduced Family of* $\hat{\bar{Y}}_1$

A large number of estimators have been deduced as a family of proposed estimator $\hat{\bar{Y}}_2$ under certain choices of the constants α, a, and b. These deduced estimators have been presented in the following table.

Notes

Deduced Estimator	α	a	b
$t_0 = \bar{y}_2\left(\frac{\bar{x}_1}{\bar{x}_2}\right)$ Shkhatme's (1962) ratio estimator	**1**	**0**	$b_0 \neq 0$
$t_1 = \bar{y}_2\left(\frac{\bar{x}_2}{\bar{x}_1}\right)$ Two phase product estimator	**-1**	**0**	$b_0 \neq 0$
$t_2 = \bar{y}_2\left(\frac{\bar{x}_1}{\bar{x}_2}\right)\left(\frac{\bar{Z}}{\bar{z}_1}\right)$ Chand (1975) estimator	**1**	$a_0 \neq 0$	**0**
$t_3 = \bar{y}_2\left(\frac{\bar{x}_1}{\bar{x}_2}\right)\left(\frac{\bar{Z}+C_z}{\bar{z}_1+C_z}\right)$ Sing and Upadhayaya's estimator (2001)	**1**	**1**	$\boldsymbol{C_z}$
$t_4 = \bar{y}_2\left(\frac{\bar{x}_1}{\bar{x}_2}\right)\left(\frac{\bar{Z}C_z+\beta_2(z)}{\bar{z}_1C_z+\beta_2(z)}\right)$ Upadhyaya and Singh (2001)	**1**	$\boldsymbol{C_z}$	$\beta_2(z)$
$t_5 = \bar{y}_2\left(\frac{\bar{x}_1}{\bar{x}_2}\right)\left(\frac{\bar{Z}+\sigma_z}{\bar{z}_1+\sigma_z}\right)$ Singh (2001) estimator	**1**	**1**	σ_z
$t_6 = \bar{y}_2\left(\frac{\bar{x}_1}{\bar{x}_2}\right)\left(\frac{\beta_1(z)\bar{Z}+\sigma_z}{\beta_1(z)\bar{z}_1+\sigma_z}\right)$ Singh (2001) estimator	$\beta_1(z)$	**1**	σ_z
$t_7 = \bar{y}_2\left(\frac{\bar{x}_1}{\bar{x}_2}\right)\left(\frac{\bar{Z}+\rho_{xz}}{\bar{z}_1+\rho_{xz}}\right)$ Singh et al.(2007) estimator	**1**	**1**	ρ_{xz}

b) Another proposed estimator for two phase (double) sampling

In this section another estimator denoted by $\hat{\bar{Y}}_2$ has been suggested for double sampling using two auxiliary variables. The suggested estimator has been convinced by H.P. Singh and N. Agnihortie (2008) and Bedi (1985) ratio estimators for two phase sampling regarding partial information on secondary auxiliary variable.

$$\hat{\bar{Y}}_2 = \bar{y}_2\left[\frac{\bar{z}_1}{\bar{z}_2}\right]^{\alpha}\left[\delta\left(\frac{a\bar{X}+b}{a\bar{x}_1+b}\right)+(1-\delta)\left(\frac{a\bar{x}_1+b}{a\bar{X}+b}\right)\right] \tag{3.6}$$

Where α, and δ are the unknown constants whose values are to be estimated. $a(\neq 0)$, and b are assumed to be known as either real numbers or (Linear or Non-linear) functions of some known parameters of auxiliary variable x as (Section-3.1).

$$\hat{\bar{Y}}_2 = \bar{Y}+\bar{Y}\left[e_{\bar{y}_2}+\alpha\left(e_{\bar{z}_1}-e_{\bar{z}_2}\right)+\gamma e_{\bar{x}_1}(1-2\delta)\right] \quad \text{where} \quad \gamma = \frac{a\bar{X}}{a\bar{X}+b}$$

$$\hat{\hat{Y}}_2 - \bar{Y} \approx \bar{Y}\left[e_{\bar{y}_2} + \alpha\left(e_{\bar{z}_1} - e_{\bar{z}_2}\right) + \gamma e_{\bar{x}_1}(1-2\delta)\right] \tag{3.7}$$

Taking square and applying expectation, [given in (1.1)], the mean square error of (3.6) is obtained as:

$$MSE(\hat{\hat{Y}}_2) = \bar{Y}^2\begin{bmatrix}\theta_2 C_y^2 + \alpha^2(\theta_2-\theta_1)C_z^2 + \theta_1\gamma^2(1-2\delta)^2 C_x^2 + \\ 2\alpha(\theta_1-\theta_2)C_y C_z\rho_{yz} + 2\theta_1(1-2\delta)\gamma C_y C_x \rho_{xy}\end{bmatrix} \tag{3.8}$$

Notes

Differentiating (3.8) with respect to α and δ and setting equal zero. We have:

$$\alpha = \frac{C_y\rho_{yz}}{C_z} \qquad \text{and} \qquad \delta = \frac{1}{2}\left(C_x + \frac{C_y\rho_{yx}}{\gamma}\right)$$

Taking the values of α and δ in equation (3.8), and simplifying. We get *min.MSE*($\hat{\hat{Y}}_2$):

$$\min.MSE(\hat{\hat{Y}}_2) = \bar{Y}^2 C_y^2\left[\theta_2 - \theta_1\rho_{xy}^2 - \theta_3\rho_{yz}^2\right] \tag{3.9}$$

In order to derive bias of (3.6), we again use (3.7) upto the 2nd order of approximation as:

$$\hat{\hat{Y}}_2 - \bar{Y} \approx \bar{Y}\begin{bmatrix} e_{\bar{y}_2} + \alpha\left(e_{\bar{z}_1} - e_{\bar{z}_2}\right) + \frac{\alpha(\alpha-1)}{2}e_{\bar{z}_1}^2 + \frac{\alpha(\alpha+1)}{2}e_{\bar{z}_2}^2 - \alpha^2 e_{\bar{z}_1}e_{\bar{z}_2} + \gamma e_1 + \alpha\gamma e_1\left(e_{\bar{z}_1} - e_{\bar{z}_2}\right) - 2\delta\gamma e_{\bar{x}_1} \\ -2\delta\gamma\alpha e_{\bar{x}_2}\left(e_{\bar{z}_1} - e_{\bar{z}_2}\right) + \delta\gamma^2 e_{\bar{x}_1}^2 + \alpha e_{\bar{y}_2}\left(e_{\bar{z}_1} - e_{\bar{z}_2}\right) + \gamma e_{\bar{x}_1}e_{\bar{y}_2}(1-2\delta) \end{bmatrix}$$

After applying expectation and simplifying, the optimum bias of (3.6) is:

$$Bias(\hat{\hat{Y}}_2) = \bar{Y}\left[\frac{\theta_3}{2}\left(C_y\rho_{yz} - \frac{C_z}{2}\right)^2\right] \tag{3.10}$$

i. *Deduced Family of* $\hat{\hat{Y}}_2$

A large number of estimators have been deduced as a family of proposed estimator $\hat{\hat{Y}}_2$ under certain choices of the constants α, a, b and δ. These deduced estimators have been presented in the following table.

Deduced Estimator	α	a	b	δ
$t_0 = \bar{y}_2$ Usual mean per unit	**0**	**0**	**1**	δ_0
$t_1 = \bar{y}_2\frac{\bar{X}}{\bar{x}_1}$ Usual ratio type	**0**	**1**	**0**	**1**
$t_2 = \bar{y}_2\frac{\bar{x}_1}{\bar{X}}$ Usual product type	**0**	**1**	**0**	**0**

Notes

$t_3 = \bar{y}_2 \dfrac{\bar{X} + C_x}{\bar{x}_1 + C_x}$ Sisodia and Diwivedi (1981) type estimator	**0**	**1**	$\boldsymbol{C_x}$	**1**
$t_4 = \bar{y}_2 \dfrac{\bar{x}_1 + C_x}{\bar{X} + C_x}$ Pandey and Dubey (1988) type estimator	**0**	**1**	$\boldsymbol{C_x}$	**0**
$t_5 = \bar{y}_2 \dfrac{\beta_2(x)\bar{x}_1 + C_x}{\beta_2(x)\bar{X} + C_x}$ Upadhyaya and Singh (1999) type estimator	0	$\beta_2(x)$	$\boldsymbol{C_x}$	0
$t_6 = \bar{y}_2 \dfrac{C_x\bar{x}_1 + \beta_2(x)}{C_x\bar{X} + \beta_2(x)}$ Upadhyaya and Singh (1999) type estimator	0	$\boldsymbol{C_x}$	$\beta_2(x)$	0
$t_7 = \bar{y}_2 \dfrac{\bar{x}_1 + \sigma_x}{\bar{X} + \sigma_x}$ G.N. Singh (2003) type estimator	0	1	σ_x	0
$t_8 = \bar{y}_2 \dfrac{\beta_1(x)\bar{x}_1 + \sigma_x}{\beta_1(x)\bar{X} + \sigma_x}$ G.N. Singh (2003) type estimator	0	$\beta_1(x)$	σ_x	0
$t_9 = \bar{y}_2 \dfrac{\beta_2(x)\bar{x}_1 + \sigma_x}{\beta_2(x)\bar{X} + \sigma_x}$ G.N. Singh (2003) type estimator	0	$\beta_2(x)$	σ_x	0
$t_{10} = \bar{y}_2 \dfrac{\bar{X} + \rho}{\bar{x}_1 + \rho}$ Singh and Tailor (2003) type estimator	0	1	ρ	1
$t_{11} = \bar{y}_2 \dfrac{\bar{x}_1 + \rho}{\bar{X} + \rho}$ Singh and Tailor (2003) type estimator	0	1	ρ	0
$t_{12} = \bar{y}_2 \dfrac{\bar{X} + \beta_2(x)}{\bar{x}_1 + \beta_2(x)}$ Singh et al. (2004) type estimator	0	1	$\beta_2(x)$	1
$t_{13} = \bar{y}_2 \dfrac{\bar{x}_1 + \beta_2(x)}{\bar{X} + \beta_2(x)}$ Singh et al. (2004) type estimator	0	1	$\beta_2(x)$	0
$t_{14} = \bar{y}_2 \dfrac{\bar{X}}{\bar{x}_1} \dfrac{\bar{z}_1}{\bar{z}_2}$ Chain ratio type estimator	1	1	0	1
$t_{15} = \bar{y}_2 \dfrac{\bar{x}_1}{\bar{X}} \dfrac{\bar{z}_1}{\bar{z}_2}$ Product to ratiotype estimator	1	1	0	0

In addition to these estimators a large number of estimators can also be deduced from the proposed family of estimators by putting values of α, a, b and δ. It is observed that the expression of the first order approximation of MSE of the given number of the family can be obtained by mere substituting the values of α, a, b and δ in (3.8).

IV. Empirical Illustration

To analyze the performance of proposed estimators in comparison to other estimators, five population data sets are being considered. In two phase sampling under *SRSWOR*, the comparison of proposed estimators $\hat{\hat{\bar{Y}}}_1$ and $\hat{\hat{\bar{Y}}}_2$ with respect to usual unbiased estimator, Mohanty (1967), Chand (1975), Mukhergee et at (1987), Srivastava et al (1990), Sahoo et al (1994a), Singh (2001), Singh et al. (2004), Samiuddin and Hanif (2006), Samiuddin and Hanif (2007), and Singh et al. (2007) have been made regarding the availability of partial information only. The descriptions of populations are given below.

Population-I.
Data used by Anderson (1958)
Y: Head length of second son **X:** Head length of first son **Z:** Head breadth of first son

$N = 25$, $\bar{Y} = 183.84$, $\bar{X} = 185.72$, $\bar{Z} = 151.12$, $C_y = 0.0546$, $C_x = 0.0526$,
$C_z = 0.0488$, $\rho_{xy} = 0.7108$, $\rho_{yz} = 0.6932$, $\rho_{zx} = 0.7346$, $n_1 = 10$, $n_2 = 7$

Population-II.
(Source: Nachtshemim, Neter and Kutner. Advanced applied linear models, 2004)
Y: No of persons below poverty level **X:** No of unemployed persons
Z: Total population

$N = 440$, $\bar{Y} = 119.50$, $\bar{X} = 906.79$, $\bar{Z} = 159.17$, $C_y = 1.9955$, $C_x = 1.7501$,
$C_z = 1.5317$, $\rho_{xy} = 0.956$, $\rho_{yz} = 0.932$, $\rho_{zx} = 0.969$, $n_1 = 88$, $n_2 = 18$

Population-III.
(Source: Population census report of Okara district (1998), Pakistan)
Y: Population Matric and above **X:** Primary but below Matric **Z:** Population both sexes

$N = 300$, $\bar{Y} = 41.5233$, $\bar{X} = 141.58$, $\bar{Z} = 1518.767$, $C_y = 1.2185$, $C_x = 1.088$,
$C_z = 0.9757$, $\rho_{xy} = 0.894$, $\rho_{yz} = 0.84$, $\rho_{zx} = 0.94$, $n_1 = 60$, $n_2 = 12$

Population-IV.
(Source: Population census report of Gujrat district (1998), Pakistan)
Y: Population Matric and above **X:** Primary but below Matric **Z:** Population both sexes

$N = 300$, $\bar{Y} = 131.5133$, $\bar{X} = 356.8433$, $\bar{Z} = 1407.407$, $C_y = 1.2532$, $C_x = 0.991$,
$C_z = 0.9545$, $\rho_{xy} = 0.927$, $\rho_{yz} = 0.893$, $\rho_{zx} = 0.972$, $n_1 = 60$, $n_2 = 12$

Population-V.
(Source: Nachtshemim, Neter and Kutner. Advanced applied linear models, 2004)
Y: Grade-point average following freshman year **Z:** ACT entrance examination score
X: High school class rank as percentile: lower percentile imply higher class rank

$N = 705$, $\bar{Y} = 2.9773$, $\bar{X} = 76.95$, $\bar{Z} = 24.54$, $C_y = 0.213123$, $C_x = 0.242157$,
$C_z = 0.16357$, $\rho_{xy} = 0.398$, $\rho_{yz} = 0.366$, $\rho_{zx} = 0.443$, $n_1 = 141$, $n_2 = 28$

Table 4.1 : PRE's of different proposed estimators of $\bar{Y}$ in double sampling w.r.t $\bar{y}$

Estimator	Population #:				
	1	2	3	4	5
Usual unbiased estimator $t_1 = \bar{y}$	100	100	100	100	100
Mohanty (1967) t_2	135.48	172.44	133.06	154.44	89.21

Notes

Notes

Chand (1975) t_3	32.58	30.15	30.47	33.19	33.32
Mukhergee et al. (1987) t_6	131.30	105.87	110.09	104.91	*118.38*
Srivastava et al. (1990) t_7	196.39	*1066.15*	*462.17*	*662.32*	*118.25*
Sahoo et al. (1994a) t_8	73.32	39.49	45.11	41.03	99.34
Singh (2001) t_9	124.37	415.59	306.62	344.23	75.64
Singh et al. (2004) t_{10}	170.97	631.94	285.64	399.93	106.0
Sammiudin & Hanif(2006) t_{11}	130.35	145.83	126.9	156.58	63.20
Sammiudin & Hanif (2007) t_{12}	137.36	582.78	265.37	370.30	103.07
Singh et al. (2007) t_{13}	196.39	*1066.15*	*462.17*	*662.32*	*118.25*
Proposed Estimator $\hat{\hat{\bar{Y}}}_1$	196.39	1066.15	462.17	662.32	118.25
Proposed Estimator $\hat{\hat{\bar{Y}}}_2$	197.99	808.77	358.69	520.19	116.01

V. Conclusion

We have suggested two improved estimators $\hat{\hat{\bar{Y}}}_1$ and $\hat{\hat{\bar{Y}}}_2$. From table 4.1, we conclude that the proposed estimators are better than usual unbiased estimator $\bar{y}$, Mohanty (1967), Chand (1975), Mukhergee et at (1987), Srivastava et al (1990), Sahoo et al (1994a), Singh (2001), Singh et al. (2004), Samiuddin and Hanif (2006), Samiuddin and Hanif (2007), and Singh et al. (2007). It is also observed that among the class of suggested estimators, $\hat{\hat{\bar{Y}}}_1$ performs more efficiently except in population-I and population-V, in comparison with proposed estimator $\hat{\hat{\bar{Y}}}_2$ and Mukhergee et al. (1987) respectively. It is further observed that $\hat{\hat{\bar{Y}}}_1$, Srivastava et al. (1990) and Singh et al. (2007) are performed equally. It is also observed that the performance of $\hat{\hat{\bar{Y}}}_2$ is also fairly good though it seems slightly less efficient in comparison with $\hat{\hat{\bar{Y}}}_1$. Hence proposed estimators are recommended for their practical use if only partial information are available.

References Références Referencias

1. Armstrong, J. and St-Jean, H. (1993). Generalized regression estimator for a two-phase sample of tax records. *Survey Methodology*, 20, 91-105.
2. Bedi, P.K. (1985). On two-phase multivariate sampling estimator. *Journal of the Indian Society of Agricultural Statistics*, 37(2), 158-162.
3. Bowley, A.L. (1926). Measurements of precision attained in sampling. *Bull. Inst. Inte. Statist.,* 22, 1-62.
4. Chand, L. (1975). *Some Ratio-type Estimators based on two or more Auxiliary Variables.* Unpublished Ph.D. Dissertation. Iowa State University, Iowa.
5. Das, A.K. (1988). *Contribution to the theory of sampling strategies based on auxiliary information.* Ph. D. Thesis Submitted to B.C. K.V., Mohanpur, Nadia, West Bengal, India.
6. Das, A.K. and Tripathi, T.P. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya,* Ser. C, 40, 139-148.

7. Freese, F. (1962). *Elementary forest sampling*. USDA For. Ser. Agric. Handb. 232.
8. Hansen, M.H. and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Ann. Math. Statist.* 14, 333-362.
9. Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika.* 27, 217-223.
10. Kiregyera, B. (1984). Regression-type estimator using two auxiliary variables and model of double sampling from finite populations. *Metrika*. 31, 215-226.
11. Mohanty, S. (1967). Combination of Regression and Ratio Estimate. *J. Ind. Statist. Assoc*. 5, 16-19.
12. Mukerjee, R., Rao, T.J. and Vijayan, K. (1987). Regression type estimators using multiple auxiliary information. *Austral. J. Statist.* 29(3), 244-254.
13. Neyman, J. (1934). On the two different aspects of representative method: The method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.*, 97, 558-606.
14. Neyman, J. (1938). Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.*, 33, 101-116.
15. Pandey, B.N. and Dubey, Vyas (1988): Modified product estimator using coefficient of variation of auxiliary variate, *Assam Statistical Rev.*, 2(2), 64-66.
16. Roy, D.C. (2003). A regression type estimates in two-phase sampling using two auxiliary variables. *Pak. J. Statist.* 19(3), 281-290.
17. Sahoo, J. and Sahoo, L.N. (1993). A class of estimators in two-phase sampling using two auxiliary variables. *J. Ind. Statist. Assoc.* 31, 107-114.
18. Sahoo, J., Sahoo, L. N. and Mohanty, S. (1994a). A regression approach to estimation in two-phase sampling using two auxiliary variables. *Current Science*, 65(1), 73-75.
19. Sahoo. J., Sahoo, L.N. and Mohanty, S. (1994b). An alternative approach to estimation in two-phase sampling using two auxiliary variables. *Biometrical J.,* 3, 293-298.
20. Samiuddin, M. and Hanif, M. (2007). Estimation of population mean in single and two phase sampling with or without additional information. *Pak. J. Statist.,* 23(2), 99-118.
21. Searls, D.T. (1964). The utilization of a known coefficient of variation in the estimation procedure. *J. Amer. Statist. Assoc.*, 59, 1225-1226.
22. Sen, A.R. (1978). Estimation of the population mean when the coefficient of variation is known. *Commun. Statist. Theor. Method*, A 7(7), 657-672.
23. Singh H.P. and Tailor, R. (2003): Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, 6(4), 555-560.
24. Singh H.P. Singh S. and Kim Jong-Min (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.
25. Singh, H.P., Tailor, R. and Kakaran, M.S. (2004): An estimator of Population mean using power transformation. *Jour. Ind. Soc. Agri. Statistics,* 58(2), 223-230.
26. Singh, R. Chauhan P. Swan N. and Smarandache F. (2011). *Italian journal of pure and applied mathematics-N*, 28 135-142.
27. Singh, G.N. (2001). On the use of transformed auxiliary variable in the estimation of population mean in two phase sampling. *Statistics in Transition*. 5(3), 405-416.
28. Singh, G.N. (2003): On the improvement of product method of estimation in sample surveys. *Jour. Ind. Soc. Agri. Statistics,* 56(3), 267-275.

Notes

Notes

29. Singh, H.P. and Agnihotri.N. (2008). A general procedure of estimating population mean using auxiliary information in sample survey. *Statistics in Transition,* 9(1), 71-87.
30. Singh, H.P., Upadhyaya, L.N. and 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.
31. Singh, J. Pandey, B.N. and Hirano, K. (1973). On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. *Ann. Inst. Stat. Math.*, 25, 51-55.
32. Singh, R., Chuhan, P. and Sawan, N. (2007). A Family of Estimators for Estimating Population Mean Using Known Correlation Coefficient in Two Phase Sampling. *Statistics in Transition*, 8(1), 89-96.
33. Sisodia, B.V.S. And Dwivedi, V.K. (1981): A modified ratio estimator using coefficient of variation of auxiliary variable. *Journ. Ind. Soc. Agril. Statist.*, 33(2), 13-18.
34. Snedecor, G.W. and King, A.J. (1942). Recent developments in sampling for agricultural statistics. *J. Amer. Statist. Assoc.,* 37, 95-102.
35. Spurr, S.H. (1952). *Forest Inventory*. The Ronald Press Company.
36. Srivastava, S. Rani, Khare, B.B. and Srivastava, S.R. (1990). A generalized chain ratio estimator for mean of finite population. *J. Ind. Soc. Agri. Statist,* 42(1), 108-117.
37. Srivastava, S.K. and Jhajj, H.S. (1980). A class of estimators using auxiliary information for estimating finite population variance. *Sankhya,* Ser. C. 42, 87-96.
38. Sukhatme, B.V. (1962). Some ratio-type estimators in two-phase sampling, *J. Amer. Statist. Assoc.*, 57, 628-632.
39. Tripathi, T.P. (1970). *Contributions to the sampling theory in multivariate information*. Ph.D. Thesis submitted to Punjabi University, Patiala, India.
40. Unnikrishan, N.K. and Kunte, S. (1995) Optimality of an analogue of Basu's estimator under a double sampling design. *Sankhya*, 57, 103-111.
41. Upadhyaya, L.N. and Singh, G.N. (2001), Chain type estimators using transformed auxiliary variable in two-phase sampling*. Advances in Modeling and Analysis*, 38(1-2), 1-10.
42. Upadhyaya, L.N. and Singh, H.P. (1999): Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, 41(5), 627-636.