International Journal of Statistics and Probability; Vol. 5, No. 3; May 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Use of Auxiliary Variables and Asymptotically Optimum Estimators in Double Sampling

M. E. Kanwai[1], O. E. Asiribo[2] & A. Isah[3]

[1]Department of Mathematics, Ahmadu Bello University, Zaria, Kaduna State, Nigeria.

[2]Department of Statistics, University of Agriculture, Abeokuta, Ogun State, Nigeria.

[3]Department of Mathematics and Statistics, Federal University of Technology, Minna, Niger State, Nigeria.

Correspondence: M. E. Kanwai, E-mail: mekanwai3@gmail.com

Received: January 17, 2016 Accepted: February 16, 2016 Online Published: April 14, 2016

doi:10.5539/ijsp.v5n3p55 URL: http://dx.doi.org/10.5539/ijsp.v5n3p55

Abstract

This paper explores the need for exploiting auxiliary variables in sample survey and utilizing asymptotically optimum estimator in double sampling to increase the efficiency of estimators. The study proposed two types of estimators with two auxiliary variables for two phase sampling when there is no information about auxiliary variables at population level. The expressions for the Mean Squared Error (MSE) of the proposed estimators were derived to the first order of approximation. An empirical comparative approach of the minimum variances and percent relative efficiency were adopted to study the efficiency of the proposed and existing estimators. It was established that, the proposed estimators performed more efficiently than the mean per unit estimator and other previous estimators that don't use auxiliary variable and that are not asymptotically optimum. Also, it was established that estimators that are asymptotically optimum that utilized single auxiliary variable are more efficient than those that are not asymptotically optimum with two auxiliary variables.

Keywords: auxiliary variable, asymptotically optimum estimator, efficiency, double sampling.

1. Introduction

In survey research, there are times when information is available on every unit in the population. If a variable that is known for every unit of the population is not a variable of interest but is instead employed to improve the sampling plan or to enhance estimation of the variables of interest, it is called an auxiliary variable. The auxiliary variable about any study population may include a known variable to which the variable of interest called the study variable is approximately related. This information may be used at the planning stage of the survey, in the estimation procedure, or at both phases.

The estimation of population parameters with greater precision is an unrelenting issue in sampling theory and the precision of estimates can be improved by increasing the sampling size, but doing so tend to sabotage the benefits of sampling. Therefore, the precision may be increased by using an appropriate estimation procedure that utilizes some auxiliary information which is closely related to the study variable and employing estimators that are asymptotically optimum.

Laplace (1820) was the first to use auxiliary information in ratio type estimator. Watson (1937) used regression method of estimation to estimate the average area of the leaves on a plant. Cochran (1940) used auxiliary information in single-phase sampling to develop the ratio estimator for estimation of population mean. In the ratio estimator, the study variable and the auxiliary variable have high positive correlation and the regression line passes through the origin. Robson (1957) and Murthy (1964) worked independently on usual product estimator of population mean. General intuitive variable of interest, can be improved if the information supplied by a related variable (auxiliary variable, supplementary variable, or concomitant variable).When two or more auxiliary variables are available; many estimators may be defined by linking together different estimators such as ratio, product or regression, each one of them exploiting a single variable. These mixed estimators have been seen performing better as compared with individual estimators. Mohanty (1967), used this methodology for the first time to propose mixed estimator using two auxiliary variables.

Many other contributions are present in sampling literature and, recently, some new estimators appeared that found the asymptotical expression for the mean square error. Here we mention, among others, Upadhyaya *et al* (1992), Tracy and Singh (1999), Radhey *et al* (2002), Singh and Espejo (2007), Samiuddin and Hanif (2007) and Singh *et al* (2010). Also, estimators, with no information case and that utilize two auxiliary variables includes: Samiuddin and Hanif (2007) and

Swain (2012). Motivated by these recent proposals, in this paper we propose, when two auxiliary variables are available, some new estimators obtained from the Mohanty (1967), Mukerjee *et al* (1987), and Singh and Espejo (2007).

This paper explores the need for exploiting auxiliary variables and asymptotically optimum estimator to increase efficiency of estimators in double sampling**.** The paper is organized as follows: Section 2 introduces methods and estimators considered in the study. In Section 3, we present the notations and two proposed estimators and obtained, up to the first degree of approximation, the approximate expressions for mean square errors. Section 4 is devoted to the empirical study of the efficiency of the proposed estimators. Section 5 is on discussion of the results from the empirical analysis. Section 6 is on conclusion and recommendations.

2. Method and Estimators

2.1 Research Design

Consider a finite population $U = (U_1, U_2, ---, U_N)$ of size N with the triple characters $(y, x, and\ z,)$, taking values $y_i,\ x_i,\ and\ z_i$ respectively on the unit $U_i (i = 1,2, ---, N)$. The purpose is to estimate the population mean of a study variable 'y' in the presence of two auxiliary variables $'x'$ and $'z'$. The population means $\bar{X}$ and $\bar{Z}$ of x and z respectively are not known, therefore, there is the need to adopt a double sampling technique. Assuming simple random sampling without replacement (SRSWOR) at each phase, the two phase sampling scheme runs as follows: A first phase sample $s'(s' \subset U)$ of fixed size n_1 is drawn from U to observe both x and z in order to find estimates of $\bar{X}$ and $\bar{Z}$. Given s', a second phase sample $s(s \subset s')$ of fixed size n_2 is drawn from s' to observe y in order to estimate the population mean of $\bar{Y}$.

Now, define the population means of y, x and z respectively as: $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $\bar{Z} = \frac{1}{N}\sum_{i=1}^{N} z_i$

The finite population variances of $x, y\ and\ z$ respectively are:

$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$ and $S_z^2 = \frac{1}{N-1}\sum_{i=1}^{N}(z_i - \bar{Z})^2$

More so, the covariance between y and x, y and z, and x and z are given by:

$S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})$, $S_{yz} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})(z_i - \bar{Z})$ and $S_{xz} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})(z_i - \bar{Z})$

Also, let $\rho_{yx} = \frac{s_{yx}}{s_y s_x}$, $\rho_{yz} = \frac{s_{yz}}{s_y s_z}$ and $\rho_{xz} = \frac{s_{xz}}{s_x s_z}$ denote the sample correlation between y and x, y and z, and x and z respectively, $b_{yx} = \frac{s_{yx}}{s_x^2}$ and $b_{yz} = \frac{s_{yz}}{s_z^2}$ is the regression coefficient of $y$on x and $y$on z respectively and $C_y = \frac{s_y}{\bar{y}}$, $C_x = \frac{s_x}{\bar{x}}$,and $C_z = \frac{s_z}{\bar{z}}$ denote the coefficient of variation of y, x and z respectively.

2.2 Analytical Techniques

The analytical technique adopted in this study is the relative efficiency. It is used where the comparison is made between a given procedure and a notional "best possible" procedure.

Gupta (2011), defined Relative Efficiency as a statistical tool that is used to measure the efficiency of one estimator over another estimator. The percent relative efficiency of estimator "α" to estimator "β" is expressed as:

$$PRE_{\alpha\beta} = \frac{Var(\beta)}{Var(\alpha)} * 100\%$$

According to Singh *et al* (2010), the percent relative efficiency can also be calculated using,

$$PRE_{\alpha\beta} = \frac{MSE_\beta}{MSE_\alpha} * 100\%$$

Therefore, in this research the Percent Relative Efficiency (PRE) is a statistical tool that will be used to measure the efficiency of the proposed and previous estimators with respect to mean per unit estimator.

Thus, the PRE $= \frac{MSE(T_{01})}{MSE(T_{ij})} * 100$, for the none use of auxiliary variable $(i = 0,\ j = 1,2)$; for the use of one auxiliary variable $(i = 1,\ j = 1,2,3,4)$ and for the use of two auxiliary variables $(i = 2,\ j = 1,2,3)$ and for the proposed estimators $(i = p, j = 1,2)$

2.3 *Estimators Used in Sampling Survey*

In this section we analyzed the performance of the proposed estimators and other existing estimators considered popular by means of a numerical evaluation of the first order mean square error (MSE) to the first order of approximation. For a fixed sample size, we considered the efficiency of the estimators with respect to: (i) without the use of any auxiliary variable; (ii) exploiting a single auxiliary variable; (iii) utilizing double auxiliary variables.

2.3.1 Sampling without Auxiliary Variable

The mean per unit estimator is perhaps the oldest estimator in the history of sample survey .The estimator for a sample of size *n* drawn from a population of size *N* is defined as:

$$T_{01} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \tag{2.1}$$

The mean square error (variance; as estimator is unbiased) can be immediately written as:

$$MSE(T_{01}) = \theta_2 \bar{Y}^2 C_y^2 \tag{2.2}$$

Searle (1964) presented a modified version of mean per unit estimator as given below

$$T_{02} = k\bar{y} \tag{2.3}$$

where *k* is a constant which is determined by minimizing mean square error of

$$MSE(T_{02}) = \frac{\bar{Y}^2\theta_2 C_y^2}{1+\theta_2 C_y^2} = \frac{MSE(T_{01})}{1+\bar{Y}^{-2}MSE(T_{01})} \tag{2.4}$$

2.3. 2 Sampling with one Auxiliary Variable

Auxiliary information is often used to improve the efficiency of estimators while using product, regression and ratio methods of estimation in survey sampling.

Robson (1957), Introduced the idea of product estimator when there is highly negative correlation, the estimator is given as:

$$T_{11} = \bar{y}_2 \frac{\bar{x}_2}{\bar{x}_1} \tag{2.5}$$

$$MSE(T_{11}) = \bar{Y}^2\left[\theta_2 C_y^2 + (\theta_2 - \theta_1)\left(C_x^2 + 2C_x C_y \rho_{xy}\right)\right] \tag{2.6}$$

Sukhatme(1962), used auxiliary variable in his ratio type estimator for two-phase sampling as:

$$T_{12} = \bar{y}_2 \frac{\bar{x}_1}{\bar{x}_2} \tag{2.7}$$

$$MSE(T_{12}) = \bar{Y}^2\left[\theta_2 C_y^2 + (\theta_2 - \theta_1)\left(C_x^2 - 2C_x C_y \rho_{xy}\right)\right] \tag{2.8}$$

Srivastava (1971), developed a general ratio estimator:

$$T_{13} = \bar{y}_2 \left(\frac{\bar{x}_1}{\bar{x}_2}\right)^{\alpha} \tag{2.9}$$

$$MSE(T_{13}) = \bar{Y}^2 C_y^2\left[\theta_2 - (\theta_2 - \theta_1)\rho_{xy}^2\right] \tag{2.10}$$

Singh and Espejo (2007), developed a ratio-product estimator:

$$T_{14} = \bar{y}_2 \left\{k\frac{\bar{x}_1}{\bar{x}_2} + (1-k)\frac{\bar{x}_2}{\bar{x}_1}\right\} \tag{2.11}$$

where $k = \frac{1}{2}\left(1 + \frac{C_y}{C_x}\rho_{xy}\right)$ and $0 \le k \le 1$

$$MSE(T_{14)} = \bar{Y}^2 C_y^2\left[\theta_2\left(1-\rho_{xy}^2\right) + \theta_1 \rho_{xy}^2\right] \tag{2.12}$$

2.3.4 Sampling with Two Auxiliary Variables

Various authors have proposed mixed type estimators, (that is, use of both ratio and regression estimators in some fashion). These mixed estimators perform better as compared with individual estimators.

Mohanty (1967) proposed a Regression Ratio estimator:

$$T_{21} = \left[\bar{y}_2 + b_{xy}(\bar{x}_1 - \bar{x}_2)\right]\frac{\bar{z}_1}{\bar{z}_2} \tag{2.13}$$

$$MSE(T_{21}) = \bar{Y}^2\left[\theta_2 C_y^2 + (\theta_2 - \theta_1)\left\{\rho_{xz}^2 C_z^2 - \left(\rho_{xy}C_y - \rho_{xz}C_z\right)^2 + \left(C_z - C_y\rho_{yz}\right)^2 - C_y^2\rho_{yz}^2\right\}\right] \tag{2.14}$$

Mukerjee *et al* (1987), developed three regression type estimators. One was for the situation when no auxiliary information was available:

$$T_{22} = \bar{y}_2 + b_{xy}(\bar{x}_1 - \bar{x}_2) + b_{yz}(\bar{z}_1 - \bar{z}_2) \tag{2.15}$$

$$MSE(T_{22}) = \bar{Y}^2 C_y^2\left[\theta_2 - (\theta_2 - \theta_1)\left(\rho_{xy}^2 + \rho_{yz}^2 - 2\rho_{xy}\rho_{yz}\rho_{xz}\right)\right] \tag{2.16}$$

Hanif *et al* (2010), proposed an estimator in two phase sampling given by:

$$T_{23} = \left(\bar{y}_2 + b_{yx}(\bar{x}_1 - \bar{x}_2)\right)\left\{k\frac{\bar{z}_1}{\bar{z}_2} + (1-k)\frac{\bar{z}_2}{\bar{z}_1}\right\} \tag{2.17}$$

$$where\ , k = \frac{1}{2} + \frac{1}{2}\frac{C_y}{C_z}\rho_{yz} - \frac{1}{2}\frac{C_y}{C_z}\rho_{yx}\rho_{xz}$$

$$MSE(T_{23}) = \bar{Y}^2 C_y^2\left\{\theta_2 - (\theta_2 - \theta_1)\left[\rho_{yx}^2 + \left(\rho_{yz} - \rho_{yx}\rho_{xz}\right)^2\right]\right\} \tag{2.18}$$

3. Notations and the Proposed Estimators

The study is motivated by Mohanty (1967), Mukerjee *et al* (1987), and Singh and Espejo (2007), their estimators and Mean Square error are given in (2.12) and (2.13), (2.14) and (2.15) and (2.10) and (2.11) respectively.

For notational purpose it is assumed that the mean of the estimated variable and auxiliary variables can be approximated from their population mean so that:

$$\bar{y}_2 = \bar{Y}(1+e_o),\ \bar{x}_1 = \bar{X}(1+e_1'),\ \bar{x}_2 = \bar{X}(1+e_1),\ \bar{z}_1 = \bar{Z}(1+e_2'),\quad \bar{z}_2 = \bar{Z}(1+e_2). \tag{i}$$

Where: $\bar{x}_h$ and $\bar{z}_h$ are the sample mean of the auxiliary variables X and Z at h-th phase for $h = 1$ and 2, for the variable of interest $\bar{y}_2$ is the sample mean of the study variable Y for the second phase.

Also:

$$\theta_i = \frac{1}{n_i} - \frac{1}{N} \tag{ii}$$

Where, $i = 1,2$ and $\theta_2 > \theta_1$

$E(e_0) = E(e_1) = E(e_1') = E(e_2) = E(e_2') = 0$

$E(e_0^2) = \theta_2 C^2{}_y$, $E(e_1'^2) = \theta_1 C^2{}_x$, $E(e_1^2) = \theta_2 C^2{}_x$, $E(e_2'^2) = \theta_1 C^2{}_z$, $E(e_2^2) = \theta_2 C^2{}_z$

$E(e_0 e_1') = \theta_1\rho_{yx}C_y C_x$, $E(e_0 e_1) = \theta_2\rho_{yx}C_y C_x$, $E(e_0 e_2') = \theta_1\rho_{yz}C_y C_z$, ,

$E(e_0 e_2) = \theta_2\rho_{yz}C_y C_z$, $E(e_1' e_1) = \theta_1 C^2{}_x$, $E(e_1' e_2') = \theta_1\rho_{xz}C_x C_z$, $E(e_1' e_2) = \theta_1\rho_{xz}C_x C_z$

$E(e_1 e_2') = \theta_1\rho_{xz}C_x C_z$, $E(e_1 e_2) = \theta_2\rho_{xz}C_x C_z$, $E(e_2' e_2) = \theta_1 C^2{}_z$ (iii)

Therefore, the proposed estimators are:

(a) $T_{p1} = \left\{\bar{y}_2 + b_{yz}(\bar{z}_1 - \bar{z}_2)\right\}\left\{\alpha\frac{\bar{x}_1}{\bar{x}_2} + (1-\alpha)\frac{\bar{x}_2}{\bar{x}_1}\right\}$ (3.1)

(b) $T_{p2} = \left\{\bar{y}_2 + b_{yx}(\bar{x}_1 - \bar{x}_2) + b_{yz}(\bar{z}_1 - \bar{z}_2)\right\}\left\{\beta\frac{\bar{z}_2}{\bar{z}_1} + (1-\beta)\frac{\bar{z}_1}{\bar{z}_2}\right\}$ (3.2)

where α and β are suitable constants, $0 \le \alpha \le 1$ and $0 \le \beta \le 1$

To obtain the MSE (T_{p1}) to the first degree of approximation, express equation(3.1), in terms of $e's$, we have:

$$T_{p1} = \left\{\bar{Y}(1+e_0) + b_{yz}[\bar{Z}(1+e_2') - \bar{Z}(1+e_2)]\right\}\left\{\alpha\frac{\bar{X}(1+e_1')}{\bar{X}(1+e_1)} + (1-\alpha)\frac{\bar{X}(1+e_1)}{\bar{X}(1+e_1')}\right\} \tag{3.3}$$

$$T_{p1} = \left\{(\bar{Y} + \bar{Y}e_0) + b_{yz}(\bar{Z}e_2' - \bar{Z}e_2)\right\}\{\alpha(1+e_1')(1+e_1)^{-1} + (1-\alpha)(1+e_1)(1+e_1')^{-1}\} \tag{3.4}$$

The negative exponential of (3.4) is expanded using the method of indeterminate coefficients

$$T_{p1} = \{(\bar{Y} + \bar{Y}e_0) + b_{yz}(\bar{Z}e_2' - \bar{Z}e_2)\}\{\alpha(1 + e_1')(1 - e_1 + e_1^2 \ldots\ldots) \\ +(1-\alpha)(1 + e_1)(1 - e_1' + e_1'^2 \ldots\ldots.)\} \tag{3.5}$$

Expanding the right hand side of (3.5), substituting (i) and retaining terms in first degree of $e's$, we have:

$$T_{p1} = \{\bar{Y}\alpha - \bar{Y}\alpha e_1 + \bar{Y}\alpha e_1' + \bar{Y} - \bar{Y}e_1' + \bar{Y}e_1 - \bar{Y}\alpha + \bar{Y}\alpha e_1' \\ -\bar{Y}\alpha e_1 + \bar{Y}\alpha e_0 + \bar{Y}e_0 - \bar{Y}\alpha e_0 + \bar{Z}\alpha b_{yz}e_2' + \bar{Z}b_{yz}e_2' \\ -\bar{Z}\alpha b_{yz}e_2' + \bar{Z}\alpha b_{yz}e_2 + \bar{Z}b_{yz}e_2 - \bar{Z}\alpha b_{yz}e_2\}$$

$$T_{p1} = \{(\bar{Y} + \bar{Y}e_0 + \bar{Y}e_1 - \bar{Y}e_1' - 2\alpha\bar{Y}e_1 + 2\alpha\bar{Y}e_1') + \bar{Z}b_{yz}(e_2' - e_2)\} \tag{3.6}$$

Subtracting $\bar{Y}$ from both sides of (3.6), squaring both sides and then taking expectations of both sides we get MSE of the estimator T_{p1}, up to the first order of approximation as

$$MSE(T_1 - \bar{Y})^2 = MSE(T_{p1}) = E\{(\bar{Y}e_0 + \bar{Y}e_1 - \bar{Y}e_1' - 2\bar{Y}\alpha e_1 + 2\bar{Y}\alpha e_1') + \bar{Z}b_{yz}(e_2' - e_2)\}^2$$

$$MSE(T_{p1}) = E\{(\bar{Y}^2(e_0 + e_1 - e_1' - 2\alpha e_1 + 2\alpha e_1')^2 + \bar{Z}^2 b_{yz}^2(e_2' - e_2)^2 \\ +2\bar{Y}\bar{Z}b_{yz}(e_2' - e_2)(e_0 + e_1 - e_1' - 2\alpha e_1 + 2\alpha e_1')\} \tag{3.7}$$

Expanding the right hand side of (3.7) and applying the notations of (ii) and (iii) we have:

$$MSE(T_{p1}) = \bar{Y}^2\{\theta_2 C_y^2 + (\theta_2 - \theta_1)C_x^2 + (\theta_2 - \theta_1)4\alpha^2 C_x^2 + (\theta_2 - \theta_1)2\rho_{yx}C_yC_x \\ -(\theta_2 - \theta_1)4\alpha\rho_{yx}C_yC_x - (\theta_2 - \theta_1)4\alpha C_x^2 - (\theta_2 - \theta_1)\rho_{yz}^2 C_y^2 \\ -(\theta_2 - \theta_1)2\rho_{yz}\rho_{xz}C_yC_x + (\theta_2 - \theta_1)4\alpha\rho_{yz}\rho_{xz}C_yC_z\} \tag{3.8}$$

The optimum value of "α" is obtained by differentiating (3.8), which gives it minimum value as:

$$\alpha = \frac{1}{2}\left(1 + \frac{C_y}{C_x}\rho_{yx} - \frac{C_y}{C_x}\rho_{yz}\rho_{xz}\right) \tag{3.9}$$

Substituting equation (3.9) in (3.8) and simplifying, the Mean Square Error of (3.1) we have:

$$MSE(T_{p1}) = \bar{Y}^2 C_y^2\left\{\theta_2 - (\theta_2 - \theta_1)\left[\rho_{yz}^2 + \left(\rho_{yx} - \rho_{yz}\rho_{xz}\right)^2\right]\right\} \tag{3.10}$$

Similarly, to obtain the MSE (T_{p2}), from (3.2) above to the first degree of approximation, substituting (i), we have:

$$T_{p2} = \{\bar{Y}(1 + e_0) + b_{yx}(\bar{X}(1 + e_1') - \bar{X}(1 + e_1)) + b_{yz}(\bar{Z}(1 + e_2') - \bar{Z}(1 + e_2))\} \\ \left\{\beta\frac{\bar{Z}(1 + e_2')}{\bar{Z}(1 + e_2)} + (1 - \beta)\frac{\bar{Z}(1 + e_2)}{\bar{Z}(1 + e_2')}\right\}$$

$$T_{p2} = \{\bar{Y}(1 + e_0) + b_{yx}(\bar{X}(1 + e_1') - \bar{X}(1 + e_1)) + b_{yz}(\bar{Z}(1 + e_2') - \bar{Z}(1 + e_2))\} \\ \{\beta(1 + e_2')(1 + e_2)^{-1} + (1 - \beta)(1 + e_2)(1 + e_2')^{-1}\} \tag{3.11}$$

The negative exponential of (3.11) is expanded using the method of indeterminate coefficients

$$T_{p2} = \{(\bar{Y} + \bar{Y}e_0) + \bar{X}b_{yx}(e_1' - e_1) + \bar{Z}b_{yz}(e_2' - e_2)\} \\ \{(\beta + \beta e_2')(1 - e_2 + e_2^2 \ldots\ldots) + (1 + e_2 - \beta - \beta e_2)(1 - e_2' + e_2'^2 \ldots\ldots.)\} \tag{3.12}$$

Expanding the right hand side of (3.12) and retaining terms of first degree of $e's$, we have:

$$T_{p2} = \{\bar{Y}k - \bar{Y}\beta e_2' + \bar{Y}\beta e_2 + \bar{Y} - \bar{Y}e_2 + \bar{Y}e_2' - \bar{Y}\beta + \bar{Y}\beta e_2 - \bar{Y}\beta e_2' \\ +\bar{Y}\beta e_0 + \bar{Y}e_0 - \bar{Y}\beta e_0 + \bar{X}\beta b_{yx}e_1' + \bar{X}b_{yx}e_1' - \bar{X}\beta b_{yx}e_1' \\ +\bar{X}\beta b_{yx}e_1 - \bar{X}b_{yx}e_1 + \bar{X}\beta b_{yx}e_1 + \bar{Z}\beta b_{yz}e_2' + \bar{Z}b_{yz}e_2' \\ -\bar{Z}\beta b_{yz}e_2' - \bar{Z}\beta b_{yz}e_1 - \bar{Z}b_{yz}e_1 + \bar{Z}\beta b_{yz}e_1\}$$

$$T_{p2} = \{(\bar{Y} + \bar{Y}e_0 + \bar{Y}e_2' - \bar{Y}e_2 - 2\beta\bar{Y}e_2' + 2\beta\bar{Y}e_2) + \bar{X}b_{yx}(e_1' - e_1) + \bar{Z}b_{yz}(e_2' - e_2)\} \tag{3.13}$$

Subtracting $\bar{Y}$ from both sides of (3.13), squaring both sides and then taking expectations of both sides we:

$$MSE(T_{p2} - \bar{Y})^2 =$$

$$MSE(T_{p2}) = E\{(\bar{Y}e_0 + \bar{Y}e_2' - \bar{Y}e_2 - 2\beta\bar{Y}e_2' + 2\beta\bar{Y}e_2) + \bar{X}b_{yx}(e_1' - e_1) + \bar{Z}b_{yz}(e_2' - e_2)\}^2$$

$$MSE(\ T_{p2}) = E\{\bar{Y}^2(e_0 + e_2' - e_2 - 2\beta e_2' + 2\beta e_2)^2 + \bar{X}^2 b_{yx}^2 (e_1' - e_1)^2$$
$$+2\bar{X}\bar{Y} b_{yx}(e_1' - e_1)(e_0 + e_2' - e_2 - 2\beta e_2' + 2\beta e_2)$$
$$+\bar{Z}^2 b_{yz}^2 (e_2' - e_2)^2 + 2\bar{Y}\bar{Z} b_{yz}(e_2' - e_2)(e_0 + e_2' - e_2 - 2\beta e_2' + 2\beta e_2)$$
$$+2\bar{X}\bar{Z} b_{yx} b_{yz}(e_1' - e_1)(e_2' - e_2)\} \tag{3.14}$$

Expanding the right hand side of(3.14) and applying the notations of (ii) we have:

$$MSE(T_{p2}) = \bar{Y}^2\{\theta_2 C_y^2 + (\theta_2 - \theta_1)C_z^2 + (\theta_2 - \theta_1)4\beta^2 C_z^2$$
$$-(\theta_2 - \theta_1)4\beta C_z^2 - (\theta_2 - \theta_1)\rho_{yx}^2 C_y^2$$
$$+(\theta_2 - \theta_1)2\rho_{yx}\rho_{xz}C_y C_z - (\theta_2 - \theta_1)4\beta\rho_{yx}\rho_{xz}C_y C_z \tag{3.15}$$
$$-(\theta_2 - \theta_1)\rho_{yz}^2 C_y^2 + (\theta_2 - \theta_1)2\rho_{yx}\rho_{yz}\rho_{xz}C_y^2$$

The optimum value of "β" is obtain by differentiating (3.15), which gives it minimum value as:

$$\beta = \frac{1}{2}\left(1 + \frac{C_y}{C_z}\rho_{yx}\rho_{xz}\right) \tag{3.16}$$

Substituting equation (3.16) in (3.15) and simplifying the Mean Square Error of (3.2) is:

$$MSE(\ T_{p2}) = \bar{Y}^2 C_y^2 \left\{\theta_2 - (\theta_2 - \theta_1)\left[\rho_{yx}^2 + \left(\rho_{yz} - \rho_{yx}\rho_{xz}\right)^2\right]\right\} \tag{3.17}$$

4. Empirical Study

To analyze the performance of various estimators of population mean $\bar{Y}$of study variable y, we considered the following two data sets:

Data 1. [Source: Perry (2007), page 63]

y= Household net disposal income, x= the household consumption and z =the number of household income-earners

$$N = 8011,\ n_1 = 700,\ n_2 = 250,\ \bar{Y} = 28229.427, \bar{X} = 20418.618,\ \bar{Z} = 1.6897,\ C_y = 0.787,$$
$$C_x = 0.668,\ C_z = 0.4596,\ \rho_{xy} = 0.74,\ \rho_{yz} = 0.458,\ \rho_{xz} = 0.348$$

Data 2. [Source: Perry (2007), page 63]

y= The sale area (in square metres), x= the number of employees and $z =$ the amount of soft drinks sales (in 1000 euros in a year)

$$N = 2376,\ n_1 = 200,\ n_2 = 70,\ \bar{Y} = 1701.946, \bar{X} = 40.617,\ \bar{Z} = 615.637,\ C_y = 1.285,$$
$$C_x = 2.35,\ C_z = 1.651,\ \rho_{xy} = 0.898,\ \rho_{yz} = 0.861,\ \rho_{xz} = 0.773$$

Therefore, in this study the Percent Relative Efficiency (PRE) is used to measure the performance of the proposed and previous estimators with respect to mean per unit estimator.

Table 4.1. Percent relative efficiency of different estimators compared to mean per unit estimator

Estimators	**Auxiliary Variable Used**	**MSE of Population I**	**MSE of Population II**	**PRE of Population I**	**PRE of Population II**
T_{01}	*none*	1912690	66315.13	100	100
T_{02}	*none*	1908110	64830.9	100.24	102.2894
T_{11}	*one*	4421459	360731	43.25925	18.38354
T_{12}	*one*	1232701	68978.73	155.1625	96.13852
T_{13}	*one*	1217679	30500.06	157.0767	217.4262
T_{14}	*one*	1217679	30500.06	157.0767	217.4262
T_{21}	*two*	1353340	84774.87	141.331	78.22499
T_{22}	*two*	1250836	50664.25	152.9129	130.8914
T_{23}	*two*	1166667	29263.7	163.9447	226.6122
T_{p1}	*two*	1218594	30990.89	156.9587	213.9827
T_{p2}	*two*	1166667	29263.7	163.9447	226.6122

***Note*: MSE=Mean Square Error; PRE=Percent Relative Efficiency**

Table 4.2. Percent relative efficiency of non-asymptotical optimum and asymptotical optimum estimators compared to mean per unit Estimator

Non-Asymptotical Optimum Estimators	MSE of Population I	MSE of Population II	PRE of Population I	PRE of Population II
T_{01}	1912690	66315.13	100	100
T_{11}	4421459	360731	43.25925	18.38354
T_{12}	1232701	68978.73	155.1625	96.13852
T_{21}	1353340	84774.87	141.331	78.22499
T_{22}	1250836	50664.25	152.9129	130.8914
Asymptotical Optimum Estimators				
T_{02}	1908110	64830.9	100.24	102.2894
T_{13}	1217679	30500.06	157.0767	217.4262
$T_{14} = T_{13}$	1217679	30500.06	157.0767	217.4262
T_{23}	1166667	29263.7	163.9447	226.6122
T_{p1}	1218594	30990.89	156.9587	213.9827
$T_{p2} = T_{23}$	1166667	29263.7	163.9447	226.6122

5. Discussions of Results

From table 4.1, in the first population, the estimators T_{11} $T_{12}, T_{13}, T_{14}, T_{21}, T_{22,} T_{23}, T_{p1}$, and T_{p2} that utilizes supplementary (auxiliary variable) information has established superiority over the two estimators (T_{01} and T_{02}) that do not use such information. Also, in the second population all the estimators with the exception of T_{11} and T_{12} shown advantage over T_{01} and T_{02} that do not use the auxiliary variables. Probably, the discrepancy in the outcome of T_{11} and T_{12} is caused by the different types of populations considered. The population described in Data Set 2 shows, as compared to that in Data Set 1, a higher variability and higher correlation between the variables. Particularly, the high variability in the auxiliary variables may affect the first order mean square error making it inaccurate. Therefore, there is always the need to ensure that the auxiliary variables is highly correlated with the study variable and the population under consideration is homogeneously distributed and where there is no correlation between the auxiliary and study variables the application of double sampling may be futile. Also, where there is correlation between the study and auxiliary variables and such population is not homogeneously distributed, stratified double sampling will be more appropriate.

Furthermore, utilizing supplementary information to improve the performance of an estimator cannot be overemphasized, but it is worth to note that asymptotical optimum estimators performed better than non- asymptotical optimum estimators. In table 4.2, estimators T_{13} and T_{14} utilized only one auxiliary variable, but performed better than estimators T_{21} and T_{22} that used two auxiliary variables. The two mean per unit estimator considered in this study also shows that, the asymptotical optimum estimators T_{02} have advantage over the non-asymptotical optimum estimators T_{01}. Therefore, the performance of the asymptotical optimum estimators and non- asymptotical optimum estimators is shown in table 4.3. It reveals that the asymptotical optimum estimators ($T_{13}, T_{14}, T_{23}, T_{p1}$ and T_{p2}) perform better than the non-asymptotical optimum estimators ($T_{01}, T_{11}, T_{12}, T_{21}$ and T_{22}), except for the mean per unit estimator T_{02}.

The estimator T_{13} performed equally well as T_{14} and the first proposed estimator T_{p1} used the second auxiliary variable for the regression and the first auxiliary variable for the ratio-product estimator and it performed better than the following estimators $T_{01}, T_{02}, T_{11}, T_{12}, T_{21}$, and T_{22} .The second estimator T_{p2} is regression-cum-regression and product-cum-ratio estimator and it gave a higher precision over all the estimators considered in this study, but gave an equal precision as T_{23}. Though T_{23} is a regression and ratio-cum- product estimator and it uses the first auxiliary variable for the regression and the second auxiliary variable for the ratio-product estimator. Perry (2007), asserted that, when two or more auxiliary variables are available, many estimators may be defined by linking together different estimators such as ratio, product or regression, each one exploiting a single variable.

6. Conclusions

In the course of the research two asymptotical optimum estimators that utilize two auxiliary variables were proposed for increasing the efficiency of estimators in double sampling.

The study reveals that, where there is correlation between the study and auxiliary variables and such population is not homogeneously distributed, stratified double sampling will be more appropriate.

The study also showed that, estimators that are asymptotically optimum and utilized single auxiliary variable are more efficient than those that non-asymptotically optimum with two auxiliary variables. Therefore, estimators that do not use auxiliary variable have high precision; however, estimators that use one or two auxiliary variables have higher precision and estimators that use one or two auxiliary variables and that are asymptotically optimum have highest precision.

The study considered estimators with the no information case for the ratio and regression estimators. It is recommended that a study of estimators with the partial and complete information case and the stratified double sampling can be carried out.

References

Cochran, W. G. (1940). The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *Journal of the Agricultural Science, 30*, 262-275. http://dx.doi.org/10.1017/S0021859600048012

Gupta, S. C. (2011). Fundamentals of Statistics. Mumbai: Himalaya Publishing House PVT.LTD.

Hanif, M., Hamad, N., & Shahbaz, M. Q. (2010). Some New Regression Type estimators in two Phase Sampling. *World Applied Sciences Journal, 8*(7), 799-803.

Laplace, P. S. (1820). A philosophical essay on probabilities. English translation, Dover, 1951.

Mohanty, S. (1967). Combination of Regression and Ratio Estimate. *Journal of Indian Statistical Association, 5*, 16-19.

Mukerjee, R., Rao, T. J., & Vijayan, K. (1987). Regression type estimators using multiple auxiliary information. *Austral. J. Statist., 29*(3), 244-254. http://dx.doi.org/10.1111/j.1467-842X.1987.tb00742.x

Murthy, M. N. (1964). Product method of estimation. *Sankhya A, 26*, 69–74.

Perry, P. F. (2007). Improved ratio-cum-product type estimators. *Statistics in Transition, 8*(1), 51-69.

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.

Robson, D. S. (1957). Application of Multivariate Polykays to the theory of Unbiased Ratio Type Estimators. *Journal of the America Statistical Association, 52*(280), 511-522. http://dx.doi.org/10.1080/01621459.1957.10501407

Samiuddin, M., & Hanif, M. (2007). Estimation of population mean in single phase and two phase sampling with or without additional information. *Pak. J. Statist., 23*(2), 99-118.

Searls, D. T. (1964). The Utilization of a known Coefficient of Variation in the Estimation Procedure. *J. Amer.* http://dx.doi.org/10.1080/01621459.1964.10480765

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. http://dx.doi.org/10.1080/02664760600994562

Singh, H. P., Singh, S., & Kim, J. M. (2010). Efficient Use of Auxiliary Variables in Estimating Finite Population Variance in Two-Phase Sampling. *Communications of the Korean Statistical Society, 17*(2), 165-181. http://dx.doi.org/10.5351/CKSS.2010.17.2.165

Srivastava, S. K. (1971). A Generalized Estimator for the Mean of a Finite Population Using Multi Auxiliary Information. *Journal of the American Statistical Association, 66*(334), 404-407. http://dx.doi.org/10.1080/01621459.1971.10482277

Sukhatme, B. (1962). Some ratio type estimators in two-phase sampling. *J. Amer. Statist. Assoc., 57*, 628-632. http://dx.doi.org/10.1080/01621459.1962.10500551

Swain, A K P C. (2012). On Some Classes of Modified Ratio Type and Regression Type Estimators in Two Phase Sampling. *Pak. J. Statist., 28*(4), 445-466.

Tracy, D., & Singh, H. (1999). A general class of chain regression estimators in two-phase sampling. *Journal of Applied Statistical Science, 8*(4), 205-216.

Upadhyaya, L. N., Dubey, S. P., & Singh, H. P. (1992). A class ratio-in regression estimator using two auxiliary variables in double sampling. *J. Scientific Research, 42*, 128-134.

Watson, D. J. (1937). The Estimation of Leaf Areas. *Journal of Agricultural Science, 27*, 475-504. http://dx.doi.org/10.1017/S002185960005173X

Copyrights