Use of Auxiliary Variables and Asymptotically Optimum Estimators in Double Sampling

M. E. Kanwai¹, O. E. Asiribo² & A. Isah³

¹Department of Mathematics, Ahmadu Bello University, Zaria, Kaduna State, Nigeria.
²Department of Statistics, University of Agriculture, Abeokuta, Ogun State, Nigeria.
³Department 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
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 \( U = (U_1, U_2, \ldots, U_N) \) of size \( N \) with the triple characters \((y, x, \text{and } z)\), taking values \( y_i, x_i, \text{and } z_i \) respectively on the unit \( U_i (i = 1, 2, \ldots, N) \). 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:

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

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

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

Also, let \( \rho_{yx} = \frac{s_{yx}}{S_x S_y} \), \( \rho_{yz} = \frac{s_{yz}}{S_y S_z} \), 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, \( b_{yx} = \frac{s_{yx}}{S_x} \) and \( b_{yz} = \frac{s_{yz}}{S_z} \) is the regression coefficient of \( y \) on \( x \) and \( y \) on \( z \) respectively and \( C_y = \frac{S_y}{\bar{Y}}, C_x = \frac{S_x}{\bar{X}}, \text{and } C_z = \frac{S_z}{\bar{Z}} \) 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 “\( \alpha \)” to estimator “\( \beta \)” is expressed as:

\[
PRE_{\alpha\beta} = \frac{\text{Var}(\beta)}{\text{Var}(\alpha)} \times 100\%
\]

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

\[
PRE_{\alpha\beta} = \frac{\text{MSE}_\beta}{\text{MSE}_\alpha} \times 100\%
\]

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 \( \text{PRE} = \frac{\text{MSE}(T_{ij})}{\text{MSE}(T_{ij})} \times 100\% \) 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:

\[
T_{01} = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}
\]  \hspace{1cm} (2.1)

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

\[
MSE(T_{01}) = \theta_2^2 \frac{C_y^2}{\text{var}(y)}
\]  \hspace{1cm} (2.2)

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

\[
T_{02} = k\bar{y}
\]  \hspace{1cm} (2.3)

where \( k \) is a constant which is determined by minimizing mean square error of

\[
MSE(T_{02}) = \frac{\theta_2^2 C_y^2}{1 + \theta_2 C_y^2}
\]  \hspace{1cm} (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:

\[
T_{11} = \bar{y}_2 \frac{\bar{x}_2}{\bar{x}_1}
\]  \hspace{1cm} (2.5)

\[
MSE(T_{11}) = \bar{y}^2 \theta_2 C_y^2 + \left( \theta_2 - \theta_1 \right) \left( C_x^2 + 2C_x C_y \rho_{xy} \right)
\]  \hspace{1cm} (2.6)

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

\[
T_{12} = \bar{y}_2 \frac{\bar{x}_1}{\bar{x}_2}
\]  \hspace{1cm} (2.7)

\[
MSE(T_{12}) = \bar{y}^2 \left( \theta_2 C_y^2 + \left( \theta_2 - \theta_1 \right) \left( C_x^2 - 2C_x C_y \rho_{xy} \right) \right)
\]  \hspace{1cm} (2.8)

Srivastava (1971), developed a general ratio estimator:

\[
T_{13} = \bar{y}_2 \left( \frac{\bar{x}_1}{\bar{x}_2} \right)^\alpha
\]  \hspace{1cm} (2.9)

\[
MSE(T_{13}) = \bar{y}^2 C_y^2 \left[ \theta_2 - \left( \theta_2 - \theta_1 \right) \rho_{xy}^2 \right]
\]  \hspace{1cm} (2.10)

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

\[
T_{14} = \bar{y}_2 \left\{ \frac{k \bar{x}_1}{\bar{x}_2} + \left( 1 - k \right) \bar{x}_2 \right\}
\]  \hspace{1cm} (2.11)

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

\[
MSE(T_{14}) = \bar{y}^2 C_y^2 \left[ \theta_2 \left( 1 - \rho_{xy}^2 \right) + \theta_1 \rho_{xy}^2 \right]
\]  \hspace{1cm} (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:

\[
T_{21} = \left[ \bar{y}_2 + b_{xy}(\bar{x}_1 - \bar{x}_2) \right] \frac{x_{i1}}{z_{i2}}
\]  
(2.13)

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

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

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

\[
MSE(T_{22}) = \bar{V}^2 C_y^2 \left[ \theta_2 - (\theta_2 - \theta_1) \left( \rho_{zy}^2 + \rho_{yz}^2 - 2 \rho_{zy} \rho_{yx} \rho_{zx} \right) \right]
\]  
(2.16)

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

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

where, \( k = \frac{1}{2} + \frac{1}{2} \frac{C_y}{C_z} \rho_{yz} - \frac{1}{2} \frac{C_y}{C_z} \rho_{yx} \rho_{zx} \)

\[
MSE(T_{23}) = \bar{V}^2 C_y^2 \left[ \theta_2 - (\theta_2 - \theta_1) \left( \rho_{xy}^2 + \rho_{yx}^2 - \rho_{xy} \rho_{yx} \rho_{zx} \right) \right]
\]  
(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:

\[
\bar{y}_2 = \bar{V}(1 + e_0), \quad \bar{x}_1 = \bar{X}(1 + e_1), \quad \bar{z}_1 = \bar{Z}(1 + e_2)
\]  
(i)

Where: \( \bar{x}_1 \) 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:

\[
\theta_i = \frac{1}{n_i} - \frac{1}{N}
\]  
(ii)

Where, \( i = 1, 2 \) and \( \theta_2 > \theta_1 \)

\[
E(e_0) = E(e_1) = E(e'_1) = E(e_2) = E(e'_2) = 0
\]

\[
E(e_0^2) = \theta_2 C_y^2, \quad E(e'_2) = \theta_2 C_y^2, \quad E(e_1^2) = \theta_1 C_x^2, \quad E(e'_1^2) = \theta_1 C_x^2, \quad E(e_1 e_2) = \theta_1 \rho_{xy} C_y C_x, \quad E(e_0 e'_2) = \theta_1 \rho_{xy} C_y C_x, \quad E(e_0 e'_1) = \theta_1 \rho_{xy} C_y C_x, \quad E(e_0 e'_2) = \theta_1 \rho_{xy} C_y C_x
\]

\[
E(e'_1 e'_2) = \theta_1 \rho_{xz} C_x C_z, \quad E(e'_1 e'_2) = \theta_1 \rho_{xz} C_x C_z, \quad E(e'_1 e'_2) = \theta_1 \rho_{xz} C_x C_z
\]

Therefore, the proposed estimators are:

(a) \( T_{p1} = \{(\bar{y}_2 + b_{yz}(\bar{z}_1 - \bar{z}_2)\} \left\{ k \frac{x_{i1}}{z_{i2}} + (1 - k) \frac{x_{i2}}{z_{i1}} \right\} \)

(3.1)

(b) \( T_{p2} = \{(\bar{y}_2 + b_{yx}(\bar{x}_1 - \bar{x}_2) + b_{yz}(\bar{z}_1 - \bar{z}_2)\} \left\{ \beta \frac{x_{i1}}{z_{i1}} + (1 - \beta) \frac{x_{i2}}{z_{i2}} \right\} \)

(3.2)

where \( \alpha \) and \( \beta \) are suitable constants, \( 0 \leq \alpha \leq 1 \) and \( 0 \leq \beta \leq 1 \)

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

\[
T_{p1} = \left[ (\bar{y}(1 + e_0) + b_{yz}(\bar{Z}(1 + e_2) - \bar{Z}(1 + e_2)) \left\{ \alpha \frac{(1 + e'_1)}{x(1 + e'_1)} + (1 - \alpha) \frac{\bar{y}(1 + e_2)}{(1 + e'_2)} \right\} \right]
\]

(3.3)

\[
T_{p1} = \left\{ (\bar{y} + Ye_0) + b_{yz}(\bar{Z}e'_2 - Z e_2)\} \left\{ \alpha(1 + e'_1)(1 + e_2)^{-1} + (1 - \alpha)(1 + e_1)(1 + e'_2)^{-1} \right\} \right\}
\]

(3.4)
The negative exponential of (3.4) is expanded using the method of indeterminate coefficients
\[ T_{p1} = \left\{ (\bar{Y} + e_0) + b_{yx}(Ze'_2 - Ze_2)\right\} \left[ \alpha(1 + e'_1)(1 - e_1 + e'_1 \ldots )\right. \\
+ (1 - \alpha)(1 + e_1)(1 - e_1 + e^2 \ldots \ldots )\right\} \]  
(3.5)

Expanding the right hand side of (3.5), substituting (i) and retaining terms in first degree of \(e\)'s, we have:
\[ T_{p1} = \{\bar{Y} - \bar{Y}ae_1 + \bar{Y}ae'_1 + \bar{Y} - \bar{Y}e'_1 + \bar{Y}e_1 - \bar{Y}a + \bar{Y}ae'_1 \\
- \bar{Y}ae_1 + \bar{Y}ae_0 + \bar{Y}e_0 - \bar{Y}ae_0 + Zab_{yx}e'_2 + Zb_{yx}e'_2 \\
- Zab_{yx}e'_1 + \bar{Y}b_{yx}e_2 + Zb_{yx}e_2 - Zab_{yx}e_2\}
\]  
(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
\[ MSE(T_1 - \bar{Y})^2 = MSE(T_{p1}) = E\{\left(\bar{Y}e_0 + \bar{Y}e_1 - \bar{Y}e'_1 - 2\bar{Y}ae_1 + 2\bar{Y}ae'_1 + Zb_{yx}(e'_2 - e_2)\right)^2\} \]  
(3.7)

Expanding the right hand side of (3.7) and applying the notations of (ii) and (iii) we have:
\[ MSE(T_{p1}) = E\{\left(\bar{Y}c_0^2 + (\theta_2 - \theta_1)c_0^2 + (\theta_2 - \theta_1)4\alpha c_0^2 + (\theta_2 - \theta_1)2\rho_{yx}c_0c_x \right) + (\theta_2 - \theta_1)4\alpha c_0^2 + (\theta_2 - \theta_1)\rho_{yx}c_0^2 \}
\]  
(3.8)

The optimum value of “\(\alpha\)” is obtained by differentiating (3.8), which gives it minimum value as:
\[ \alpha = \frac{1}{2} \left( 1 + \frac{c_y}{c_x} \right) \]  
(3.9)

Substituting equation (3.9) in (3.8) and simplifying, the Mean Square Error of (3.1) we have:
\[ MSE(T_{p1}) = \bar{y}^2c_0^2 \left\{ \theta_2 - (\theta_2 - \theta_1) \left[ \rho_{yx}^2 + (\rho_{yx} - \rho_{yx}c_{xx})^2 \right] \right\} \]  
(3.10)

Similarly, to obtain the MSE \(T_{p2}\), from (3.2) above to the first degree of approximation, substituting (i), we have:
\[ T_{p2} = \{\bar{Y}(1 + e_0) + b_{yx}(\bar{X}(1 + e'_1) - \bar{X}(1 + e_1)) + b_{yx}(\bar{Z}(1 + e'_2) - \bar{Z}(1 + e_2))\}
\]  
(3.11)

The negative exponential of (3.11) is expanded using the method of indeterminate coefficients
\[ T_{p2} = \left\{ (\bar{Y} + e_0) + \bar{X}b_{yx}(e'_1 - e_1) + \bar{Z}b_{yx}(e'_2 - e_2)\right\} \]
\[ \left\{ (\beta + \beta e'_2)(1 - e_2 + e'_2 \ldots \ldots ) + (1 + e_2 - \beta - \beta e_2)(1 - e'_2 + e'_2 \ldots \ldots )\right\} \]  
(3.12)

Expanding the right hand side of (3.12) and retaining terms of first degree of \(e\)'s, we have:
\[ T_{p2} = \{\bar{Y}k - \bar{Y}e_0 + \bar{Y}e_1 - \bar{Y}e'_1 + \bar{Y}e_2 - \bar{Y}e'_2 - \bar{Y} + \bar{Y}e_0 - \bar{Y}e'_0 + \bar{Y}e_1 - \bar{Y}e'_1 + \bar{Y}b_{yx}e'_2 + \bar{X}b_{yx}e'_1 + \bar{X}b_{yx}e'_1 - \bar{X}b_{yx}e'_1 \}
\]  
(3.13)

Subtracting \(\bar{Y}\) from both sides of (3.13), squaring both sides and then taking expectations of both sides we:
\[ MSE(T_{p2} - \bar{Y})^2 = \]  
(3.13)
\[
MSE(T_{p2}) = E\{\bar{Y}^2(e_0 + e'_0 - e_2 - 2\beta e'_2 + 2\beta e_2)^2 + \bar{X}^2 b_{Yx}^2(e'_1 - e_1)^2
+ 2\bar{X}\bar{Y} b_{Yx}(e'_1 - e_1)(e_0 + e'_0 - e_2 - 2\beta e'_2 + 2\beta e_2) + \bar{Z}^2 b_{Yz}^2(e'_2 - e_2)^2 + 2\bar{Y}\bar{Z} b_{Yz}(e'_2 - e_2)(e_0 + e'_0 - e_2 - 2\beta e'_2 + 2\beta e_2)
+ 2XZ b_{Yz} b_{Yx} (e'_1 - e_1)(e'_2 - e_2)\} \quad (3.14)
\]

Expanding the right hand side of (3.14) and applying the notations of (ii) we have:
\[
MSE(T_{p2}) = \bar{Y}^2\{(\theta_2 - \theta_1)C_y^2 + \theta_2 C_y (\theta_2 - \theta_1)C_y^2 + (\theta_2 - \theta_1)\rho_{yx}\rho_{xz}C_y C_x - (\theta_2 - \theta_1)\rho_{yx}\rho_{xx}\rho_{xz}C_y C_x
+ \rho_{yx}\rho_{xx}\rho_{xz}\rho_{xz}C_y C_x\} \quad (3.15)
\]

The optimum value of “\(\beta\)” is obtain by differentiating (3.15), which gives it minimum value as:
\[
\beta = \frac{1}{2} \left( 1 + \frac{C_y}{C_x} \rho_{yx} \rho_{xx} \right) \quad (3.16)
\]

Substituting equation (3.16) in (3.15) and simplifying the Mean Square Error of (3.2) is:
\[
MSE(T_{p2}) = \bar{Y}^2 C_y^2 \left\{ (\theta_2 - \theta_1) \rho_{yx}^2 + (\rho_{yx} - \rho_{yx}\rho_{xx})^2 \right\} \quad (3.17)
\]

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]
\[y = \text{Household net disposal income, } x = \text{the household consumption and } z = \text{the number of household income-earners}\]
\[N = 8011, \ n_1 = 700, \ n_2 = 250, \ \bar{Y} = 28229.427, \ \bar{X} = 20418.618, \ \bar{Z} = 1.6897, \ C_y = 0.787, \ C_x = 0.668, \ C_z = 0.4596, \ \rho_{xy} = 0.74, \ \rho_{yx} = 0.458, \ \rho_{xz} = 0.348\]

Data 2. [Source: Perry (2007), page 63]
\[y = \text{The sale area (in square metres), } x = \text{the number of employees and } z = \text{the amount of soft drinks sales (in 1000 euros in a year)}\]
\[N = 2376, \ n_1 = 200, \ n_2 = 70, \ \bar{Y} = 1701.946, \ \bar{X} = 40.617, \ \bar{Z} = 615.637, \ C_y = 1.285, \ C_x = 2.35, \ C_z = 1.651, \ \rho_{xy} = 0.898, \ \rho_{yx} = 0.861, \ \rho_{xz} = 0.773\]

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

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Auxiliary Variable Used</th>
<th>MSE of Population I</th>
<th>MSE of Population II</th>
<th>PRE of Population I</th>
<th>PRE of Population II</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T_{01})</td>
<td>none</td>
<td>1912690</td>
<td>66315.13</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>(T_{02})</td>
<td>none</td>
<td>1908110</td>
<td>64830.9</td>
<td>100.24</td>
<td>102.2894</td>
</tr>
<tr>
<td>(T_{11})</td>
<td>one</td>
<td>4421459</td>
<td>360731</td>
<td>155.1625</td>
<td>155.1625</td>
</tr>
<tr>
<td>(T_{12})</td>
<td>one</td>
<td>1232701</td>
<td>68978.73</td>
<td>155.1625</td>
<td>155.1625</td>
</tr>
<tr>
<td>(T_{13})</td>
<td>one</td>
<td>1217679</td>
<td>30500.06</td>
<td>157.0767</td>
<td>157.0767</td>
</tr>
<tr>
<td>(T_{14})</td>
<td>one</td>
<td>1217679</td>
<td>30500.06</td>
<td>157.0767</td>
<td>157.0767</td>
</tr>
<tr>
<td>(T_{21})</td>
<td>two</td>
<td>1353340</td>
<td>84774.87</td>
<td>141.331</td>
<td>141.331</td>
</tr>
<tr>
<td>(T_{22})</td>
<td>two</td>
<td>1250836</td>
<td>50664.25</td>
<td>152.9129</td>
<td>152.9129</td>
</tr>
<tr>
<td>(T_{23})</td>
<td>two</td>
<td>1166667</td>
<td>29263.7</td>
<td>163.9447</td>
<td>163.9447</td>
</tr>
<tr>
<td>(T_{p1})</td>
<td>two</td>
<td>1218594</td>
<td>30990.89</td>
<td>156.9587</td>
<td>156.9587</td>
</tr>
<tr>
<td>(T_{p2})</td>
<td>two</td>
<td>1166667</td>
<td>29263.7</td>
<td>163.9447</td>
<td>163.9447</td>
</tr>
</tbody>
</table>

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


Copyrights
Copyright for this article is retained by the author(s), with first publication rights granted to the journal.
This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).