Alternative Ratio Estimator of Population Mean in Simple Random Sampling

Ekaette Inyang Enang¹, Victoria Matthew Akpan¹ & Emmanuel John Ekpenyong²

¹ Department of Mathematics/Statistics & Computer Science, University of Calabar, Calabar, Nigeria
² Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria

Abstract

An alternative ratio estimator is proposed for a finite population mean of a study variable \( Y \) in simple random sampling using information on the mean of an auxiliary variable \( X \), which is highly correlated with \( Y \). Expressions for the bias and the mean square error of the proposed estimator are obtained. Both analytical and numerical comparisons have shown the proposed alternative estimator to be more efficient than some existing ones. The bias of the proposed estimator is also found to be negligible for all populations considered, indicating that the estimator is as good as the regression estimator and better than the other estimators under consideration.

Keywords: ratio estimator, efficiency, bias, simple random sampling, auxiliary variables, precision

1. Introduction

1.1 Background of Study

Estimation theory is an important part of statistical studies, whereby, population parameters are obtained using sample statistics. In any survey work, the experimenters interest is to make use of methods that will improve precisions of estimates of the population parameters both at the design stage and estimation stage. These parameters can be totals, means or proportions of some desired characters.

In sample surveys, auxiliary information is used at selection as well as estimation stages to improve the design as well as obtaining more efficient estimators. Increased precision can be obtained when study variable \( Y \) is highly correlated with auxiliary variable \( X \).

Usually, in a class of efficient estimators, the estimator with minimum variance or mean square error is regarded as the most efficient estimator. A good estimator can also be described by the value of its bias. An estimator with minimum absolute bias is regarded as a better estimator among others in the class (Rajesh et al., 2011).

When the populations mean of an auxiliary variable is known, so many estimators for population parameters of study variable have been discussed in literature. The literature on survey sampling describes a great variety of techniques for using auxiliary information by means of ratio, product and regression methods.

If the regression line of the character of interest \( Y \) on the auxiliary variable, \( X \) is through the origin and when correlation between study and auxiliary variables is positive (high), then the ratio estimate of mean or total may be used (Cochran 1940).

On the other hand, if the regression line used for the estimate does not pass through the origin but makes an intercept along the y-axis, the regression estimation is used (Okafor, 2002). Furthermore, when correlation between study variable and auxiliary variable is negative, the product method of estimation is preferred. Robson (1957), Murthy (1967), Perri (2005), Muhammad et al. (2009) and Solanki et al. (2012) had established that the regression estimator is generally more efficient than the ratio and product estimators except when the regression line of the study variable on the auxiliary variable passes through a suitable neighbourhood of the origin, in which case, the efficiencies of these estimators are almost equal. When the population parameters of the auxiliary variable \( X \) such as population mean, coefficient of variation, coefficient of kurtosis, coefficient of skewness, median are known, a
number of modified estimators such as modified ratio estimators, modified product estimators and modified linear regression estimators have been proposed and is widely acceptable in the literature (Subramani & Kumarpandiyan, 2012). In sampling literature, many estimators have been proposed when a single auxiliary variable is involved. Under some realistic conditions, they are found to be more efficient than the sample mean, the ratio and product estimators and are as efficient as the regression estimator in the optimum case but the problem of the best estimator in terms of both efficiency and biasness has not been fully addressed. This paper is another attempt in solving this problem. An alternative ratio estimators for population mean of the study variable, $Y$, which is more efficient than some of the existing estimators is proposed using information on one auxiliary variable, $X$, that is highly correlated with study variable.

1.2 Summary of Some Existing Estimators

To consider effective comparison, we summarize below some existing estimators, their biases and mean square errors.

Consider a finite population of $N$ distinct and identifiable units $\Pi = \{U_1, U_2, \ldots, U_N\}$. Let a sample of size $n$ be drawn from the population by simple random sampling without replacement. Suppose that interest is to obtain a ratio estimate of the mean of a random variable $Y$ from the sample using a related variable $X$ as supplementary information and assuming that the total of $X$ is known from sources outside the survey.

Table 1. Some existing estimators, their biases and mean square errors

<table>
<thead>
<tr>
<th>S/N</th>
<th>Estimator</th>
<th>Bias</th>
<th>Mean Square Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>$\bar{y}$</td>
<td>0</td>
<td>$\frac{1-f}{n} \hat{\theta}^2 C^2$</td>
</tr>
<tr>
<td>2.</td>
<td>$\bar{y}_{st} = \frac{\bar{y}}{\bar{x} + \rho}$</td>
<td>$\frac{1-f}{n} f[C^2_2 - \rho C, C_1]$</td>
<td>$\frac{1-f}{n} f^2[C^2_2 + C^2_1 - 2\rho C, C_1]$</td>
</tr>
<tr>
<td>3.</td>
<td>$\bar{y}_{kc} = \frac{\bar{y} + b(X - \bar{x})}{\bar{x}}$</td>
<td>$\frac{1-f}{n} f[wC_2 - \rho wC, C_1]$</td>
<td>$\frac{1-f}{n} f^2[C^2_2 + C^2_1 w(w - 2\theta)]$</td>
</tr>
<tr>
<td>4.</td>
<td>$\bar{y}_{reg} = \hat{y} + b(X - X)$</td>
<td>$\frac{1-f}{n} fC^2_1$</td>
<td>$\frac{1-f}{n} f^2[C^2_2 + C^2_1 (1 - \rho^2)]$</td>
</tr>
<tr>
<td>5.</td>
<td>Regression estimator</td>
<td>0</td>
<td>$\frac{1-f}{n} f^2 C^2 (1 - \rho^2)$</td>
</tr>
</tbody>
</table>

Where

- $C_2 = \frac{S_2}{S_1}$; the coefficient of variation of the auxiliary variable, $X$ and the response variable, $Y$;
- $\rho = \frac{S_2}{S_1 S_2}$; the correlation coefficient between $X$ and $Y$;
- $w = \frac{X + \rho}{Y}$; $B = \frac{S_2}{S_1}$; the regression coefficient;
- $\theta = \frac{\rho C_2}{C_1}$ and $f = \frac{n}{N}$, where $S_2^2 = (N - 1)^{-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$, $S_1^2 = (N - 1)^{-1} \sum_{i=1}^{N} (y_i - \bar{y})^2$; the population variances of the auxiliary and study variables respectively;
- $S_{xy} = (N - 1)^{-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$; the population covariance between $X$ and $Y$;
- $\bar{X} = N^{-1} \sum_{i=1}^{N} x_i$, $\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i$; population means of the auxiliary and study variables;
- $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$, $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$; sample means of the auxiliary and study variables are respectively defined wherever they appear.

2. Proposed Estimator

The proposed ratio estimator is obtained by forming linear combination of Singh and Tailor (2003) and Kadilar...
and Cingi (2004) estimators as shown below:

\[
\hat{y}_{pr} = \frac{a\tilde{y}(\bar{X} + \rho)}{\bar{X} + \rho} + \frac{\beta(\tilde{y} + b(\bar{X} - \tilde{x}))}{\bar{X} - \tilde{x}} \bar{X} 
\]

(1)

such that \( \alpha + \beta = 1 \)

2.1 Bias and Mean Square Error of the Proposed Estimator

To obtain the approximate expression for the bias and the mean squared error for the proposed ratio estimator, let

\[
\bar{x} = \bar{X}(1 + e_x); \quad \bar{y} = \bar{Y}(1 + e_y)
\]

(2)

where

\[
e_x = \frac{\bar{x} - \tilde{x}}{\tilde{x}}, \quad e_y = \frac{\bar{y} - \tilde{y}}{\tilde{y}}
\]

So that,

\[
E(e_x) = E(e_y) = 0, \quad E(e_x^2) = \frac{(1-f)}{n} C_x^2, \quad E(e_y^2) = \frac{(1-f)}{n} C_y^2
\]

(3)

Therefore, expressing (1) in terms of (2), we obtain

\[
\hat{y}_{pr} = \frac{a\tilde{y}(1 + e_y)(\bar{X} + \rho)}{\bar{X} + \rho} + (1 - a) \left[ \frac{\tilde{y}(1 + e_y) + b[\bar{X} - \tilde{x}(1 + e_x)]}{\bar{X}(1 + e_x)} \right] \bar{X}
\]

\[
= a\tilde{y}(1 + e_y)(\bar{X} + \rho) + (1 - a) \left[ \tilde{y}(1 + e_y)(1 + e_x)^{-1} - b\bar{X}e_x(1 + e_x)^{-1} \right]
\]

\[
= a\tilde{y}(1 + e_y)(1 + we_x)^{-1} + \tilde{y}(1 - e_x + e^2_x + e_y - e_xe_y) - \tilde{y}(\alpha - ae_x + ae_x^2 + ae_y - ae_ye_x) - b\bar{X}e_x + b\bar{X}e_x^2 + ab\bar{X}e_x - ab\bar{X}e_x^2
\]

By Taylor Series approximation up to order 2, the expression becomes

\[
\hat{y}_{pr} = a\tilde{y}(1 - we_x + w^2e_x^2 + e_y - we_ye_x) + \tilde{y}(1 - e_x + e^2_x + e_y - e_xe_y)
\]

\[
+ \tilde{y}(\alpha - a - e_x^2 - ae_x + ae_xe_x) - B\tilde{X}e_x + B\tilde{X}e_x^2 + a\tilde{X}e_x - a\tilde{X}e_x^2
\]

\[
= \tilde{y}[\alpha - a - awe_x + aw^2e_x^2 + ae_x - awe_xe_x + 1 - e_x + e^2_x + e_y - e_xe_y - \alpha + ae_x]
\]

\[
- ae_x^2 - ae_x + ae_xe_x - B\tilde{X}e_x + B\tilde{X}e_x^2 + a\tilde{X}e_x - a\tilde{X}e_x^2]
\]

The expression for the Bias of this estimator to first order approximation is obtained as follows:

\[
B(\hat{y}_{pr}) = E(\hat{y}_{pr} - \bar{Y})
\]

\[
= \tilde{y}[\alpha - a - aw - 1 + (aw + 1 - a + BK - aBK)e_x^2 - \bar{Y}]
\]

(4)

\[
= \frac{1 - f}{n} \tilde{y}[(a - aw - 1) \rho C_x C_y + (aw^2 + 1 - a + BK - aBK) C_y^2]
\]

The MSE of this estimator is given by:

\[
MSE(\hat{y}_{pr}) = E(\hat{y}_{pr} - \bar{Y})^2
\]

\[
= \tilde{y}^2[C_x^2 + 2(aBK - BK - a - aw) \rho C_x C_y + (aBK - BK - a - aw)^2 C_y^2]
\]

(5)
2.2 Optimal Conditions for the Proposed Estimator

To obtain the value of $\alpha$ that minimizes the MSE, we take partial derivative of Equation (5) with respect to $\alpha$ and equate it to zero as follows:

$$\frac{\partial \text{MSE}(\bar{y}_{\text{pr}})}{\partial \alpha} = \frac{1 - f}{n} \bar{y}^2 \left[ 2(BK + 1 - w) \rho C_y C_x + 2(BK + 1 - w) (\alpha BK - BK - \alpha - aw) C_x^2 \right] = 0$$

$$\Rightarrow \rho C_x C_y + \alpha (BK + 1 - w) C_x^2 - (BK + 1) C_x^2 = 0$$

$$\Rightarrow \alpha = \frac{(BK + 1) C_x^2 - \rho C_x C_y}{(BK + 1 - w) C_x^2}$$

(6)

Substituting for (6) in (5) gives the optimal MSE for $\bar{y}_{\text{pr}}$ as:

$$\text{MSE}(\bar{y}_{\text{pr}}) = \frac{1 - f}{n} \bar{y}^2 \left[ C_y^2 \left(1 - \rho^2\right)\right]$$

(7)

3. Efficiency Comparison

In order to compare the efficiency of the various existing estimators with that of proposed estimator, we require the expressions of mean square error of these estimators, up to first order of approximation. An analytical comparison of the proposed estimator with three of the existing estimators namely: the classical, Singh and Tailor (2003) and Kadilar and Cingi (2004) estimators are carried out.

3.1 Efficiency Comparison of $\bar{y}_{\text{pr}}$ and $\bar{y}_{\text{cl}}$

In this section, the analytical condition under which the proposed estimator will be more efficient than the classical ratio estimator is established.

$$\text{MSE}(\bar{y}_{\text{pr}}) - \text{MSE}(\bar{y}_{\text{cl}}) = \frac{1 - f}{n} \bar{y}^2 \left[ C_y^2 \left(1 - \rho^2\right)\right] - \frac{1 - f}{n} \bar{y}^2 \left[ C_y^2 + C_x^2 - 2\rho C_x C_y \right]$$

$$= \frac{1 - f}{n} \bar{y}^2 \left[ C_y^2 - C_y^2 \rho^2 - C_y^2 + 2\rho C_x C_y \right]$$

$$= \frac{1 - f}{n} \bar{y}^2 \left[ 2\rho C_x C_y - C_y^2 - C_x^2 \rho^2 \right]$$

$$= - \left[ \frac{1 - f}{n} \bar{y}^2 \left(C_x \rho - C_x\right)^2 \right]$$

(8)

Since the expression in the square bracket is always positive, we conclude that the proposed estimator will always be more efficient than the classical ratio estimator.

3.2 Efficiency Comparison of $\bar{y}_{\text{pr}}$ and $\bar{y}_{\text{ST}}$

$$\text{MSE}(\bar{y}_{\text{pr}}) - \text{MSE}(\bar{y}_{\text{ST}}) = \frac{1 - f}{n} \bar{y}^2 \left[ C_y^2 \left(1 - \rho^2\right)\right] - \frac{1 - f}{n} \bar{y}^2 \left[ C_y^2 + C_x^2 w (w - 2\theta) \right]$$

$$= \frac{1 - f}{n} \bar{y}^2 \left[C_y^2 - C_x^2 w \left(w - \frac{2\rho C_y}{C_x}\right) \right]$$

$$= \frac{1 - f}{n} \bar{y}^2 \left[-C_y^2 \rho^2 - C_x^2 w \left(w - \frac{2\rho C_y}{C_x}\right) \right]$$

$$= - \left[ \frac{1 - f}{n} \bar{y}^2 \left[C_x^2 \rho^2 + C_x^2 w (w - 2\theta) \right] \right]$$

(9)

Therefore, for the proposed estimator to be more efficient than Singh and Tailor (2003), the terms in the second bracket must be positive. This implies that:

$$C_x^2 \rho^2 + C_x^2 w (w - 2\theta) > 0$$

(10)
3.3 Efficiency Comparison of \( \bar{y}_{pr} \) and \( \bar{y}_{KC} \)

Since the expression in the square bracket of Equation (11) is always positive, it therefore means that the proposed estimator will always be more efficient than Kadilar and Cingi (2004) estimator of population mean.

\[
MS E \left( \bar{y}_{pr} \right) - MS E \left( \bar{y}_{KC} \right) = \frac{1 - f}{n} Y^2 \left[ C_\gamma^2 \left( 1 - \rho^2 \right) \right] - \frac{1 - f}{n} Y^2 \left[ C_x^2 + C_\gamma^2 \left( 1 - \rho^2 \right) \right] = - \left[ \frac{1 - f}{n} Y^2 C_\gamma^2 \right]
\]

(11)

Since the expression in the square bracket of Equation (11) is always positive, it therefore means that the proposed estimator will always be more efficient than Kadilar and Cingi (2004) estimator of population mean.

4. Numerical Comparison

In this section, to study the performance of the estimator presented in this work, we consider four empirical populations used by others. The source of populations and the values of requisite population parameters are given. We compare the efficiency of the proposed estimator with the existing estimators using the four known population data.

4.1 Data Statistics for Population 1

\[
\begin{align*}
N &= 200 \quad \bar{Y} = 500 \\
n &= 50 \quad \bar{X} = 25 \\
\rho &= 0.90 \quad \theta = \rho \frac{C_y}{C_x} = 6.75 = BK \\
C_y &= 15 \quad w = \frac{\bar{X}}{\bar{X} + \rho} = 0.97 \\
C_x &= 2
\end{align*}
\]

From the above, other statistics derived are

\[
\begin{align*}
S_y &= 7500 \\
S_x &= 25(2) = 50 \\
K &= \frac{\bar{X}}{\bar{Y}} = 0.05 \\
B &= \frac{\rho S_y}{S_x} = 135 \\
R &= \frac{\bar{Y}}{\bar{X}} = 20
\end{align*}
\]


Table 2. Estimators, biases, MSE and relative bias using one auxiliary variable in Population 1

<table>
<thead>
<tr>
<th>Estimator</th>
<th>MSE</th>
<th>Bias of Estimator</th>
<th>%Relative Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{y}_{cl} )</td>
<td>656,250</td>
<td>-172.5</td>
<td>29.29</td>
</tr>
<tr>
<td>( \bar{y}_{ST} )</td>
<td>662,262.3</td>
<td>-167.512</td>
<td>20.58</td>
</tr>
<tr>
<td>( \bar{y}_{KC} )</td>
<td>175,312.50</td>
<td>30</td>
<td>7.16</td>
</tr>
<tr>
<td>( \bar{y}_{pr} )</td>
<td>160,312.5</td>
<td>0.888822</td>
<td>0.22</td>
</tr>
</tbody>
</table>

4.2 Data Statistics for Population 2

\[
\begin{align*}
N &= 106 \quad \bar{Y} = 1536.77 \\
n &= 20 \quad \bar{X} = 24375.59 \\
\rho &= 0.82 \quad \theta = \rho \frac{C_y}{C_x} = 1.696832 = BK \\
C_y &= 4.18 \quad w = \frac{\bar{X}}{\bar{X} + \rho} = 0.99997 \\
C_x &= 2.02
\end{align*}
\]
From the above, other statistics derived are

\[
S_y = 6425.09 \quad S_x = 49189.08 \\
K = \frac{\bar{X}}{\bar{Y}} = 15.86157 \quad B = \frac{\rho S_y}{S_x} = 0.107109 \\
R = \frac{\bar{Y}}{\bar{X}} = 0.063045
\]


Table 3. Estimators, biases, MSE and relative bias using one auxiliary variable in Population 2

<table>
<thead>
<tr>
<th>Estimator</th>
<th>MSE</th>
<th>Bias of Estimator</th>
<th>%Relative Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\bar{y}_{cl})</td>
<td>738,192.6</td>
<td>-177.2565</td>
<td>20.63</td>
</tr>
<tr>
<td>(\bar{y}_{ST})</td>
<td>738,210.9</td>
<td>-177.2591</td>
<td>20.63</td>
</tr>
<tr>
<td>(\bar{y}_{KC})</td>
<td>939,289.6</td>
<td>254.3749</td>
<td>26.24</td>
</tr>
<tr>
<td>(\bar{y}_{pr})</td>
<td>548,373.92</td>
<td>-0.213436</td>
<td>0.02</td>
</tr>
</tbody>
</table>

4.3 Data Statistics for Population 3

\[
N = 80 \quad \bar{Y} = 51.8264 \\
n = 20 \quad \bar{X} = 11.2624 \\
\rho = 0.9413 \quad \theta = \rho \frac{C_y}{C_x} = 0.44413 = BK \\
C_y = 0.3542 \quad w = \frac{\bar{X}}{\bar{X} + \rho} = 0.9228 \\
C_x = 0.7507
\]

From the above, other statistics derived are

\[
S_y = 18.3569 \quad S_x = 25(2) = 8.4563 \\
K = \frac{\bar{X}}{\bar{Y}} = 0.217353 \quad B = \frac{\rho S_y}{S_x} = 2.04337 \\
R = \frac{\bar{Y}}{\bar{X}} = 4.60082
\]

Source: Murthy (1967).

Table 4. Estimators, biases, MSE and relative bias using one auxiliary variable in Population 3

<table>
<thead>
<tr>
<th>Estimator</th>
<th>MSE</th>
<th>Bias of Estimator</th>
<th>%Relative Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\bar{y}_{cl})</td>
<td>18.9731</td>
<td>0.629852</td>
<td>29.29</td>
</tr>
<tr>
<td>(\bar{y}_{ST})</td>
<td>14.45027</td>
<td>0.483917</td>
<td>12.73</td>
</tr>
<tr>
<td>(\bar{y}_{KC})</td>
<td>58.20311</td>
<td>1.095255</td>
<td>14.36</td>
</tr>
<tr>
<td>(\bar{y}_{pr})</td>
<td>1.4399958</td>
<td>-0.077576</td>
<td>6.46</td>
</tr>
</tbody>
</table>

4.4 Data Statistics for Population 4

\[
N = 278 \quad \bar{Y} = 39.068 \\
n = 48 \quad \bar{X} = 25.111 \\
\rho = 0.7213 \quad \theta = \rho \frac{C_y}{C_x} = 0.6435 = BK \\
C_y = 1.4451 \quad w = \frac{\bar{X}}{\bar{X} + \rho} = 0.972078 \\
C_x = 1.6198
\]
From the above, other statistics derived are

\[ S_y = 56.4572 \quad S_x = 40.6748 \]

\[ K = \frac{\bar{X}}{\bar{Y}} = 0.642751 \quad B = \frac{\mu S_y}{S_x} = 1.001175 \]

\[ R = \frac{\bar{Y}}{\bar{X}} = 1.555812 \]

Source: Das (1988).

Table 5. Estimators, biases, MSE and relative bias using one auxiliary variable in Population 4

<table>
<thead>
<tr>
<th>Estimator</th>
<th>MSE</th>
<th>Bias of Estimator</th>
<th>%Relative Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{y}_{cl} )</td>
<td>35.12791</td>
<td>0.629852</td>
<td>29.29</td>
</tr>
<tr>
<td>( \hat{y}_{ST} )</td>
<td>33.80755</td>
<td>0.564309</td>
<td>12.73</td>
</tr>
<tr>
<td>( \hat{y}_{KC} )</td>
<td>95.38072</td>
<td>1.766793</td>
<td>14.36</td>
</tr>
<tr>
<td>( \hat{y}_{pr} )</td>
<td>26.355636</td>
<td>-0.024141</td>
<td>0.47</td>
</tr>
</tbody>
</table>

5. Discussion

The optimal Mean square error (MSE) of the proposed estimator given in Equation (7) has the same expression as the MSE of the regression estimator which is known to be more efficient than the ratio and product estimators. The analytical comparison of the proposed estimator with the three existing ones in Equations (8), (9) and (11) show that the proposed estimator will always be more efficient than the classical and Kadilar and Cingi (2004) estimators and be preferred to Singh and Tailor (2003) estimator when the condition stated in Equation (10) is satisfied.

The empirical results presented in Tables 2, 3, 4 and 5 show that the MSE of the proposed estimator is consistently less than those of classical ratio estimator, Singh and Tailor (2003) and Kadilar and Cingi (2004) estimators in population for all four populations under consideration, showing that the estimator, \( \hat{y}_{pr} \) is more efficient than all the other estimators under consideration. This is due to the fact that the proposed estimator is equally as efficient as the regression estimator and confirms Cochran (1942), Robson (1957), Murthy (1967) and Perri (2005) assertion that the regression estimator is generally more efficient than ratio and product estimators.

Analyses of biases have also shown the proposed estimator to have the smallest bias compared to all the existing estimators for all populations considered. The relative bias of the proposed estimator shown in the four tables is 10% for all the populations under consideration showing that the proposed estimator is almost unbiased. This agrees with Okafor (2002) assertion that any estimator with a relative bias of less than 10% is considered to have a negligible bias.

6. Conclusion

In conclusion, since the proposed estimator gives the same precision as the regression estimator and is consistently better in terms of bias and efficiency than the three estimators under consideration, \( \hat{y}_{pr} \) can always be used as an alternative to the regression estimator and gives a better replacement to some existing ratio estimators.

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/).