ON EFFICIENCY OF SOME RATIO ESTIMATORS IN DOUBLE SAMPLING DESIGN USING SOME EXISTING AGRICULTURAL DATA

ADEWARA, A. A.
Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria
aadewara@yahoo.com

ABSTRACT

In this paper, three sampling ratio estimators in double sampling design were proposed with the intention of finding an alternative double sampling design estimator to the conventional ratio estimator in double sampling design discussed by Cochran (1997), Okafor (2002), Raj (1972) and Raj and Chandhok (1999). Their efficiency was also determined based on the conditions attached to their supremacy in terms of their estimated mean square error (mse) obtained to the first degree of approximation. In order to justify this, three existing agricultural and an employment data sets were used which suggested that one of the proposed estimators is better and preferred to the conventional one.

KEYWORDS: efficiency, estimator, product, ratio

INTRODUCTION

In ratio, regression and difference estimation procedures, auxiliary information is used to improve the precision of an estimate of the population mean or total. Here, the advance knowledge of the population mean, \( \bar{X} \), of the auxiliary variate \( x \) is required.

If the information needed to improve on the estimate of the character under study is lacking, and if it is convenient and cheap to do so, then information on the auxiliary variable is collected from a preliminary large sample (\( n' \)) while information on the variable of interest, \( y \), is collected from a second sample (\( n'' \)) may be a sample of the preliminary sample or may be an independent sample selected from the entire population. When the second sample (\( n'' \)) is independent of the preliminary sample, (\( n' \)), (\( n'' > n \)) information on both the auxiliary and the main character is obtained from the second sample.

The disadvantage of this sampling scheme is that there is a reduction in sample size at the second phase, but this is compensated for by the use of the auxiliary information collected at the first phase. Okafor (2002).

The double sampling ratio estimator of the population mean of \( y \) is given by,
\[
\bar{y}_{dr} = \frac{\bar{y}}{\bar{x}} \bar{x'}, \quad \text{where} \quad \bar{x'} = \frac{1}{n'} \sum_{i=1}^{n'} x_i
\]
is the sample estimate of \( \bar{X} \) obtained from the first phase sample, \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \) and \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \) are the sample estimates of \( \bar{X} \) and \( \bar{Y} \) respectively, obtained from the second phase sample.
Bias and mean square error of the conventional estimator

Conventionally,

1. \[ t_1 = \frac{\bar{y} - \bar{x}^*}{s_{xy}} \]  \hspace{1cm} \ldots 1(a)

discussed by Cochran (1997), Okafor (2002), Raj (1972) and Raj and Chandhok (1999), where

\[ \bar{y} = \bar{Y}(1 + \Lambda_{\bar{y}}), \bar{x} = \bar{X}(1 + \Lambda_{\bar{x}}), \bar{x}^* = \bar{X}(1 + \Lambda_{\bar{x}^*}) \]

\[ \Delta_{\bar{y}} = \frac{|\bar{y} - \bar{x}^*|}{\bar{X}} \quad \text{and} \quad \Delta_{\bar{x}} = \frac{|\bar{y} - \bar{Y}|}{\bar{Y}}. \]

Then,

\[ t_1 = \frac{\bar{Y}(1 + \Lambda_{\bar{y}})\bar{X}(1 + \Lambda_{\bar{x}^*})}{\bar{X}(1 + \Lambda_{\bar{x})}} \]

\[ \bar{Y}(1 + \Lambda_{\bar{y}})(1 + \Lambda_{\bar{x}^*})(1 + \Lambda_{\bar{x}^*}) - 1 \]  \hspace{1cm} \ldots 1(b)

\[ Y(1 + \Lambda_{\bar{y}})(1 + \Lambda_{\bar{x}^*})(1 + \Lambda_{\bar{x}^*}) - \Delta_{\bar{y}}\Delta_{\bar{x}^*} \]

Therefore,

\[ \text{Bias}(t_1) = E(t_1 - \bar{Y}) \]

\[ = E[(\bar{Y}(1 + \Lambda_{\bar{y}})(1 + \Lambda_{\bar{x}^*})(1 + \Lambda_{\bar{x}^*}) - Y)] \]

\[ = \bar{Y}E[(1 + \Lambda_{\bar{y}})(1 + \Lambda_{\bar{x}^*})(1 + \Lambda_{\bar{x}^*}) - 1] \]

\[ = \bar{Y}E[\Lambda_{\bar{y}^2} - \Lambda_{\bar{x}^*}\Lambda_{\bar{y}} + \Lambda_{\bar{x}^*}\Lambda_{\bar{x}^*} - \Delta_{\bar{y}}\Delta_{\bar{x}^*}] \]  \hspace{1cm} \ldots 1(c)

where,

\[ E(\Lambda_{\bar{x}}) = \frac{1}{n}\sum_{i=1}^{n} \left( \frac{1}{N} - \frac{1}{N} \right) S_{x}^2, \quad E(\Lambda_{\bar{x}^*}) = \frac{1}{n}\sum_{i=1}^{n} \left( \frac{1}{N} - \frac{1}{N} \right) S_{x^*}^2, \quad E(\Lambda_{\bar{y}}) = \frac{1}{n}\sum_{i=1}^{n} \left( \frac{1}{N} - \frac{1}{N} \right) S_{y}^2, \]

\[ E(\Lambda_{\bar{x}}\Lambda_{\bar{y}}) = \frac{1}{n}\sum_{i=1}^{n} \left( \frac{1}{N} - \frac{1}{N} \right) S_{xy}, \quad E(\Lambda_{\bar{x}^*}\Lambda_{\bar{y}}) = \frac{1}{n}\sum_{i=1}^{n} \left( \frac{1}{N} - \frac{1}{N} \right) S_{xy^*} \]

\[ E(\Lambda_{\bar{x}^*}\Lambda_{\bar{x}}) = \frac{1}{n}\sum_{i=1}^{n} \left( \frac{1}{N} - \frac{1}{N} \right) S_{x^*}^2, \quad \text{and} \quad E(\Lambda_{\bar{y}}^2) = E(\Lambda_{\bar{x}}^2) = E(\Lambda_{\bar{x}^*}) = 0 \]

Substituting these in 1(c) above, then,

\[ \text{Bias}(t_1) = \frac{1}{n}\sum_{i=1}^{n} \left( \frac{1}{N} - \frac{1}{N} \right) (RS_{x}^2 - S_{xy}) \]  \hspace{1cm} \ldots 1(d)

where \( R = \frac{\bar{y}}{\bar{X}} \).

Similarly,

\[ \text{Mse}(t_1) = E(t_1 - \bar{Y})^2 \]

\[ = E[(\bar{Y}(1 + \Lambda_{\bar{y}})(1 + \Lambda_{\bar{x}})(1 + \Lambda_{\bar{x}^*} + \Delta_{\bar{x}^*}^2) - \bar{Y}] \]

\[ = \bar{Y}^2 E[(1 + \Lambda_{\bar{y}})(1 + \Lambda_{\bar{x}})(1 + \Lambda_{\bar{x}^*} + \Delta_{\bar{x}^*}^2) - 1] \]

\[ = \bar{Y}^2 E[\Lambda_{\bar{x}^2} + \Lambda_{\bar{y}^2} + \Lambda_{\bar{x}^*}^2 - 2\Lambda_{\bar{y}}\Lambda_{\bar{x}} + 2\Lambda_{\bar{y}}\Lambda_{\bar{x}^*} - 2\Lambda_{\bar{x}}\Lambda_{\bar{x}^*}] \]
Bias and mean square error of the proposed estimators

Case (a): When \( X = 0 \hat{x} + (1 - \theta) \hat{x}^*, \ Y = \theta \hat{y} + (1 - \theta) \hat{y}^* \), \( \theta = \frac{N - n'}{Nn'} \)

\[
\hat{x} = \bar{X}(1 + \Delta \hat{x}), \quad \hat{y} = \bar{Y}(1 + \Delta \hat{y}), \quad \hat{x}' = \bar{X}(1 - (\frac{N - n'}{Nn' - N + n'})\Delta \hat{x}) \quad \text{and} \quad \\
\hat{y}' = Y(1 - (\frac{N - n'}{Nn' - N + n'})\Delta \hat{y}) \text{by simple expansion.}
\]

Let,

\[
t_2 = \frac{\hat{y}'}{\hat{x}'} \cdot \bar{Y}(1 - (\frac{N - n'}{Nn' - N + n'})\Delta \hat{y})(1 + \Delta \hat{x})(1 - (\frac{N - n'}{Nn' - N + n'})\Delta \hat{y})^{-1} \quad \ldots (f)
\]

\[
t_3 = \frac{\hat{y}'}{\hat{x}'} \cdot Y(1 - (\frac{N - n'}{Nn' - N + n'})\Delta \hat{y})(1 + \Delta \hat{x})(1 + \Delta \hat{x})^{-1} \quad \ldots (g)
\]

This estimator is similar to Sodipo (2003) estimator but in different dimension.

While, Sodipo (2003) considered this estimator in the presence of non-response, ours differs from that.

\[
t_4 = \frac{\hat{y}'}{\hat{x}'} \cdot Y(1 + \Delta \hat{x})(1 + \Delta \hat{x})(1 - (\frac{N - n'}{Nn' - N + n'})\Delta \hat{y})^{-1} \quad \ldots (h)
\]

Comparing eqs. 1(b), 1(f) – 1(h), one would see that they only differs in terms of their coefficients of \( \Delta \hat{x} \) and \( \Delta \hat{y} \). While that of \( t_4 \) are \(+1\) and \(+1\), that of \( t_3 \) are

\[-(\frac{N - n'}{Nn' - N + n'}) \quad \text{and} \quad -(\frac{N - n'}{Nn' - N + n'}) \]

respectively. Also, while that of \( t_3 \) are \(+1\) and \(+1\)

\[-(\frac{N - n'}{Nn' - N + n'}) \quad \text{and} \quad -(\frac{N - n'}{Nn' - N + n'}) \quad \text{respectively.}
\]

So, the proposed estimators \( t_2, t_3 \) and \( t_4 \) originated from the conventional estimator, \( t_1 \) with the same intention.

Hence,

\[
Bias(t_2) = -\frac{1}{X} \left( \frac{n' - n}{nn'} \right) \left( \frac{N - n'}{Nn' - N + n'} \right)^2 (RS^2 x + S_{xy}) . \quad \ldots (2a)
\]

\[
Bias(t_3) = -\frac{1}{X} \left( \frac{n' - n}{nn'} \right) (RS^2 x + (\frac{N - n'}{Nn' - N + n'})S_{xy}) . \quad \ldots (2b)
\]

\[
Bias(t_4) = -\frac{1}{X} \left( \frac{n' - n}{nn'} \right) \left( \frac{N - n'}{Nn' - N + n'} \right) S_{xy} - R \left( \frac{N - n'}{Nn' - N + n'} \right)^2 S^2 x . \quad \ldots (2c)
\]

\[
Mset(t_2) = \left( \frac{N - n'}{Nn' - N + n'} \right)^2 S^2 y + \frac{N - n'}{Nn' - N + n'} \left( \frac{N - n'}{Nn' - N + n'} \right)^2 (S^2 y - 2RS_{xy} + R^2 S^2 x) . \quad \ldots (2d)
\]

29
\[
M^2\text{se}(t_3) = \left(\frac{N - n'}{nn'}\right)^2 S^2_{xy} + \left(\frac{N - n'}{nn'}\right)^2 S^2_{x^2} + 2\left(\frac{N - n'}{nn'}\right)R S_{xy} + R^2 S^2_{x^2}.
\]

...2(e)

\[
M^2\text{se}(t_4) = \left(\frac{N - n'}{nn'}\right)^2 S^2_{xy} + \left(\frac{N - n'}{nn'}\right)S^2_{y^2} + R^2\left(\frac{n'}{nn'}\right)S^2_{y^2} - 2R\left(\frac{N - n'}{nn'}\right)S_{xy} + 2 \left(\frac{N - n'}{nn'}\right)^2 S^2_{x^2}.
\]

...2(f)

and

\[
M^2\text{se}(y) = \left(\frac{N - n'}{nn'}\right)^2 S^2_{y^2}.
\]

...2(g)

**Case (b):** When \( X = f \bar{x} + (1 - f') \bar{x}' \), \( \bar{Y} = f \bar{y} + (1 - f') \bar{y}' \), \( f' = \frac{n'}{N} \)

\( x = X(1 + \Lambda_x) \), \( \bar{y} = \bar{y}(1 + \Lambda_y) \), \( \bar{x}' = \bar{x}'(1 - \frac{n'}{N - n'} \Lambda_x) \)

By simple expansion. Then, let,

\[
t_3 = \frac{\bar{y}'}{y'} X\left[\frac{1}{n'} X \left(\frac{n'}{n} \Lambda_x \right)(1 + \Lambda_x) (1 - \left(\frac{n'}{N - n'} \Lambda_x\right)^{-1})\right]
\]

...3(a)

\[
t_3 = \frac{y'}{x'} \bar{y}\left(1 + \Lambda_x\right) \left(1 + \Lambda_x\right)^{-1}
\]

...3(b)

\[
t_4 = \frac{y'}{x'} \bar{y}\left(1 + \Lambda_x\right) \left(1 + \Lambda_x\right)^{-1}
\]

...3(c)

Hence,

\[
\text{Bias}(t_3) = -\frac{1}{X} \left(\frac{n'}{n'}\right)^2 \left(\frac{n'}{n} \Lambda_x \right)^2 (RS^2_{x} + S^2_{y}).
\]

...3(d)

\[
\text{Bias}(t_4) = -\frac{1}{X} \left(\frac{n'}{n'}\right)^2 (RS^2_{x} + S^2_{y}).
\]

...3(e)

\[
\text{Bias}(t_4) = -\frac{1}{X} \left(\frac{n'}{n'}\right)^2 (RS^2_{x} + S^2_{y} - R\left(\frac{n'}{n} \Lambda_x\right)^2 S^2_{x}).
\]

...3(f)

\[
M^2\text{se}(t_3) = \left(\frac{N - n'}{nn'}\right)^2 S^2_{xy} + \left(\frac{N - n'}{nn'}\right)^2 S^2_{x^2} + 2\left(\frac{N - n'}{nn'}\right)R S_{xy} + R^2 S^2_{x^2}.
\]

...3(g)
Illustrations

We illustrated the above derivations using three existing agricultural data sets called populations 1, 3 and an employment data set called population 4 already used by Okafor (2002) and Raj (1972) for their own estimators as shown below:-

Population 1:- From a total of 912 villages, a simple random sample of 100 villages are selected. The area in hectares of each of the 100 villages x is then obtained from official record of the land survey office. The sample mean area of these 100 villages is calculated as $x' = 1402.42$ hectares. Another simple random sample of 30 villages is selected from the 100 sample villages, and the farms in each of the 30 villages. The data obtained are presented in Okafor (2002).

Table 8.1. Here, N = 912, n = 100,

$n = 30$, $x' = 1402.42$, $x = 1372.614$, $y = 733.347$, $R = 0.53427$,

$S^2_x = 791376.06$, $S^2_y = 151031.29$, $S_{xy} = 268108.86$

Population 2:- In order to estimate the total area of land (in hectares) planted by rural farmers, a simple random sample of 134 registered cooperative farmers was selected from a total of 481 cooperative farmers in 3 rural villages. The average size of household as estimated was 5 members. A simple random subsample of 28 farmers was selected from the 134 farmers in order to collect information on the land area planted (y). The result obtained are presented in Okafor (2002), Exercise 42(i). Here,

$N = 481$, $n' = 134$, $n = 28$, $x' = 5$, $x = 7.2857$, $y = 6.6946$, $R = 0.9189$,

$S^2_x = 7.6190$, $S^2_y = 18.2101$, $S_{xy} = 4.6261$

Population 3:- In order to estimate the total cultivated area in a commune containing $N = 850$ parcels of land, a random sample of $n' = 100$ parcels is selected at eye estimates x of cultivated area obtained by going from parcel to parcel. A subsample of $n = 30$ parcels is selected and each parcel measured for cultivated area y. The sample mean of eye estimates on the 100 parcels was found to be 4.31 (10stremmas = 1 hectare). The data obtained are presented in Raj (1972), Table 6.1. Here, $N = 850$, $n' = 100$, $n = 30$, $x' = 4.31$, $x = 3.93$, $y = 4.25$, $R = 1.0814$,

$S^2_x = 9.03$, $S^2_y = 9.0274$, $S_{xy} = 8.6353$

Population 4:- From a directory listing 3,500 large manufacturing establishments, a random sample of $n' = 158$ establishments was taken and questionnaire mailed to obtain information on the number of paid employees x. This gives an average of $x = 46.99$ employees per establishment. A random subsample of $n = 30$ establishments was taken and interviewers there set to examine the payroll of establishments and obtain more accurate data on employment y. The data obtained on the reported x and actual y employment are shown in Raj (1972), Table 6.7. Here,

$N = 3500$, $n' = 158$, $n = 30$, $x' = 46.99$, $x = 47.87$, $y = 50.97$, $R = 1.0648$,

$S^2_x = 850.0506$, $S^2_y = 870.6551$, $S_{xy} = 97.846$
RESULT:
The empirical results obtained using the above data sets are presented in the Tables below:

Table 1: Mse of $t_1, t_2, t_3$ and $t_4$ when $X = \theta \bar{x} + (1-\theta)\bar{x}^*$, $Y = \theta \bar{y} + (1-\theta)\bar{y}^*$.

\[
\begin{align*}
\theta &= \frac{N-n'}{Nn'} \\
\bar{x} &= \bar{X}(1+\Delta_x) \\
\bar{y} &= \bar{Y}(1+\Delta_y) \\
\bar{x}^* &= \bar{X}(1-\frac{N-n'}{Nn'-N+n'})
\end{align*}
\]

and

\[
\bar{y}^* = \bar{Y}(1-\frac{N-n'}{Nn'-N+n'})
\]

<table>
<thead>
<tr>
<th></th>
<th>Pop 1</th>
<th>pop 2</th>
<th>pop 3</th>
<th>pop 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>Bias(t_1)</td>
<td>$</td>
<td>4.5576</td>
<td>0.0179</td>
</tr>
<tr>
<td>$</td>
<td>Bias(t_2)</td>
<td>$</td>
<td>0.0004</td>
<td>0</td>
</tr>
<tr>
<td>$</td>
<td>Bias(t_3)</td>
<td>$</td>
<td>0.0409</td>
<td>0.0001</td>
</tr>
<tr>
<td>$</td>
<td>Bias(t_4)</td>
<td>$</td>
<td>0.0406</td>
<td>0.0001</td>
</tr>
<tr>
<td>$Mse(\bar{y})$</td>
<td>4868.7718</td>
<td>6.1250</td>
<td>0.0290</td>
<td>28.7731</td>
</tr>
<tr>
<td>$Mse(t_1)$</td>
<td>3454.978</td>
<td>0.5541</td>
<td>0.10</td>
<td>26.0054</td>
</tr>
<tr>
<td>$Mse(t_2)$</td>
<td>0.2788</td>
<td>0.00002</td>
<td>0.0001</td>
<td>0.9010</td>
</tr>
<tr>
<td>$Mse(t_3)$</td>
<td>5211.199</td>
<td>0.1805</td>
<td>0.2426</td>
<td>2.8261</td>
</tr>
<tr>
<td>$Mse(t_4)$</td>
<td>4809.146</td>
<td>0.6112</td>
<td>0.2864</td>
<td>28.7390</td>
</tr>
<tr>
<td>$\frac{N-n'}{Nn'}$</td>
<td>0.00890</td>
<td>0.00538</td>
<td>0.00882</td>
<td>0.00604</td>
</tr>
<tr>
<td>$\frac{N-n'}{Nn'-N+n'}$</td>
<td>0.00898</td>
<td>0.00541</td>
<td>0.00890</td>
<td>0.00608</td>
</tr>
<tr>
<td>$\frac{n'}{N-n'}$</td>
<td>0.123</td>
<td>0.38617</td>
<td>0.1333</td>
<td>0.04728</td>
</tr>
</tbody>
</table>

32
Table 2:- Mse of $t_1, t_2, t_3$ and $t_4$ when $\bar{X} = f\bar{x} + (1 - f')\bar{x}^*$, $\bar{y} = f\bar{y} + (1 - f')\bar{y}^*$,

$f' = \frac{n'}{N}, \bar{x} = \bar{X}(1 + \Lambda_x), \bar{y} = \bar{Y}(1 + \Lambda_y), \bar{x}^* = \bar{X}(1 - \frac{n'}{N - n'})\Lambda_x$ and

$\bar{y}^* = \bar{Y}(1 - \frac{n'}{N - n'})\Lambda_y$.

| $|Bias(t_1)|$ | $|Bias(t_2)|$ | $|Bias(t_3)|$ | $|Bias(t_4)|$ |
|------------|-------------|-------------|-------------|
| Pop 1      | 4.5576      | 0.0179      | 0.0513      | 0.0552      |
| Pop 2      | 0.0691      | 0.0027      | 0.0009      | 0.0001      |
| Pop 3      | 0.5613      | 0.0069      | 0.0068      | 0.0026      |
| Pop 4      | 0.4922      | 0.0043      | 0.0001      | 0.0025      |

$Mse(\hat{y})$ | 4868.7718 | 6.1250 | 0.0290 | 28.7731 |

$Mse(\bar{t}_1)$ | 3454.978 | 0.5541 | 0.10 | 26.0054 |

$Mse(\bar{t}_2)$ | 52.4002 | 0.0826 | 0.0018 | 0.0581 |

$Mse(\bar{t}_3)$ | 4521.467 | 0.1803 | 0.1935 | 2.6576 |

$Mse(\bar{t}_4)$ | 4125.48 | 0.5469 | 0.2366 | 28.5134 |

$\frac{N - n'}{Nn'}$ | 0.00890 | 0.00538 | 0.00882 | 0.00604 |

$\frac{N - n'}{Nn' - N + n'}$ | 0.00898 | 0.00541 | 0.00890 | 0.00608 |

$\frac{n'}{N - n'}$ | 0.123 | 0.38617 | 0.1333 | 0.04728 |

Discussion

From the Tables 1 and 2, we observed that estimator has the (i) least estimated mean square error, (ii) least estimated absolute error and (iii) least estimated mean square error ratio over mean per unit in the four populations considered in this study. Hence, among the four ratio estimators in double sampling design considered in this study, estimator is better and therefore, it is preferred.

REFERENCES


