

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
aaadewara@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 were 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 other 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,

$$\bar{y}_{dr} = \frac{\bar{y}}{\bar{x}} \bar{x}'$$
, where $\bar{x}' = \frac{1}{n'} \sum_{i=1}^{n'} x_i$ 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,

$$1. \quad t_1 = \frac{y}{x} x' \quad \dots 1(a)$$

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

$$y = Y(1 + \Lambda_y), \bar{x} = \bar{X}(1 + \Lambda_{\bar{x}}), x' = \bar{X}(1 + \Lambda_{x'}), \Lambda_{\bar{x}} = \left| \frac{\bar{x} - \bar{X}}{\bar{X}} \right| \text{ and } \Lambda_y = \left| \frac{y - \bar{Y}}{\bar{Y}} \right|.$$

Then,

$$t_1 = \frac{Y(1 + \Lambda_y)\bar{X}(1 + \Lambda_{x'})}{\bar{X}(1 + \Lambda_{\bar{x}})} \cdot \frac{\bar{Y}(1 + \Lambda_y)(1 + \Lambda_{x'})(1 + \Lambda_{\bar{x}})^{-1}}{Y(1 + \Lambda_y)(1 + \Lambda_{x'})(1 + \Lambda_{\bar{x}} + \Delta_{\bar{x}}^2)} \quad \dots 1(b)$$

Therefore,

$$\begin{aligned} Bias(t_1) &= E(t_1 - \bar{Y}) \\ &= E[(Y(1 + \Lambda_y)(1 + \Lambda_{x'})(1 + \Lambda_{\bar{x}} + \Delta_{\bar{x}}^2)) - Y] \\ &= \bar{Y}E[((1 + \Lambda_y)(1 + \Lambda_{x'})(1 + \Lambda_{\bar{x}} + \Delta_{\bar{x}}^2)) - 1] \\ &= \bar{Y}E[\Lambda_{\bar{x}}^2 - \Lambda_{\bar{x}}\Lambda_y + \Lambda_y\Lambda_{x'} - \Lambda_{\bar{x}}\Lambda_{x'}] \quad \dots 1(c) \end{aligned}$$

where,

$$E(\Lambda_{\bar{x}}^2) = \frac{1}{\bar{X}^2} \left(\frac{1}{n} - \frac{1}{N} \right) S_x^2, \quad E(\Lambda_{x'}^2) = \frac{1}{\bar{X}'^2} \left(\frac{1}{n'} - \frac{1}{N} \right) S_{x'}^2, \quad E(\Lambda_y^2) = \frac{1}{\bar{Y}^2} \left(\frac{1}{n} - \frac{1}{N} \right) S_y^2,$$

$$E(\Lambda_{\bar{x}}\Lambda_y) = \frac{1}{\bar{X}\bar{Y}} \left(\frac{1}{n} - \frac{1}{N} \right) S_{xy}, \quad E(\Lambda_{x'}\Lambda_y) = \frac{1}{\bar{X}'\bar{Y}} \left(\frac{1}{n'} - \frac{1}{N} \right) S_{xy}$$

$$E(\Lambda_{x'}\Lambda_{\bar{x}}) = \frac{1}{\bar{X}'\bar{X}} \left(\frac{1}{n'} - \frac{1}{N} \right) S_x^2 \text{ and } E(\Lambda_y) = E(\Lambda_{x'}) = E(\Lambda_{\bar{x}}) = 0$$

Substituting for these in 1(c) above, then,

$$Bias(t_1) = \frac{1}{\bar{X}} \left(\frac{1}{n} - \frac{1}{n'} \right) (RS_x^2 - S_{xy}) \quad \dots 1(d)$$

where $R = \frac{\bar{Y}}{\bar{X}}$.

Similarly,

$$\begin{aligned} Mse(t_1) &= E(t_1 - \bar{Y})^2 \\ &= E[(Y(1 + \Lambda_y)(1 + \Lambda_{x'})(1 + \Lambda_{\bar{x}} + \Delta_{\bar{x}}^2)) - \bar{Y}]^2 \\ &= \bar{Y}^2 E[((1 + \Lambda_y)(1 + \Lambda_{x'})(1 + \Lambda_{\bar{x}} + \Delta_{\bar{x}}^2)) - 1]^2 \\ &= \bar{Y}^2 E[\Lambda_{\bar{x}}^2 + \Lambda_y^2 + \Lambda_{x'}^2 - 2\Lambda_{\bar{x}}\Lambda_y + 2\Lambda_y\Lambda_{x'} - 2\Lambda_{\bar{x}}\Lambda_{x'}] \end{aligned}$$

Bias and mean square error of the proposed estimators

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

$$\bar{x} = \bar{X}(1 + \Delta_{\bar{x}}), \quad \bar{y} = \bar{Y}(1 + \Delta_{\bar{y}}), \quad \bar{x}^* = \bar{X}\left(1 - \left(\frac{N-n'}{Nn' - N + n'}\right)\Delta_{\bar{x}}\right) \quad \text{and}$$

$$\bar{y}^* = \bar{Y}\left(1 - \left(\frac{N-n'}{Nn' - N + n'}\right)\Delta_{\bar{y}}\right) \text{ by simple expansion.}$$

Let,

$$t_2 = \frac{\bar{y}^*}{\bar{x}^*} \bar{x}' \cdot \bar{Y}\left[\left(1 - \left(\frac{N-n'}{Nn' - N + n'}\right)\Delta_{\bar{y}}\right)(1 + \Delta_{\bar{x}})\left(1 - \left(\frac{N-n'}{Nn' - N + n'}\right)\Delta_{\bar{x}}\right)^{-1}\right] \dots 1(f)$$

$$t_3 = \frac{\bar{y}^*}{\bar{x}} \bar{x}' \cdot \bar{Y}\left[\left(1 - \left(\frac{N-n'}{Nn' - N + n'}\right)\Delta_{\bar{y}}\right)(1 + \Delta_{\bar{x}})(1 + \Delta_{\bar{x}})^{-1}\right] \dots 1(g)$$

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.

$$t_4 = \frac{\bar{y}}{\bar{x}^*} \bar{x}' = \bar{Y}\left[(1 + \Delta_{\bar{y}})(1 + \Delta_{\bar{x}})\left(1 - \left(\frac{N-n'}{Nn' - N + n'}\right)\Delta_{\bar{x}}\right)^{-1}\right] \dots 1(h)$$

Comparing eqs. 1(b), 1(f)–1(h), one would see that they only differs interms of their coefficients of $\Delta_{\bar{x}}$ and $\Delta_{\bar{y}}$. While that of t_1 are +1 and +1, that of t_2 are

$$-\left(\frac{N-n'}{Nn' - N + n'}\right) \text{ and } -\left(\frac{N-n'}{Nn' - N + n'}\right) \text{ respectively. Also, while that of } t_3 \text{ are +1 and } -\left(\frac{N-n'}{Nn' - N + n'}\right), \text{ that of } t_4 \text{ are } -\left(\frac{N-n'}{Nn' - N + n'}\right) \text{ and +1 respectively.}$$

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

Hence,

$$\text{Bias}(t_2) = -\frac{1}{X} \left(\frac{n'-n}{nn'}\right) \left(\frac{N-n'}{Nn' - N + n'}\right)^2 (RS^2_x + S_{x_y}). \dots 2(a)$$

$$\text{Bias}(t_3) = \frac{1}{\bar{X}} \left(\frac{n'-n}{nn'}\right) \left(RS^2_x + \left(\frac{N-n'}{Nn' - N + n'}\right)S_{x_y}\right). \dots 2(b)$$

$$\text{Bias}(t_4) = \frac{1}{\bar{X}} \left(\frac{n'-n}{nn'}\right) \left(\left(\frac{N-n'}{Nn' - N + n'}\right)S_{x_y} - R\left(\frac{N-n'}{Nn' - N + n'}\right)^2 S^2_x\right). \dots 2(c)$$

$$\begin{aligned} \text{Mse}(t_2) &= \left(\frac{N-n'}{Nn'}\right) \left(\frac{N-n'}{Nn' - N + n'}\right)^2 S^2_y \\ &+ \left(\frac{n'-n}{nn'}\right) \left(\frac{N-n'}{Nn' - N + n'}\right)^2 (S^2_y - 2RS_{x_y} + R^2 S^2_x). \dots 2(d) \end{aligned}$$

$$Mse(t_3) = \left(\frac{N-n'}{Nn'}\right)\left(\frac{N-n'}{Nn'-N+n'}\right)^2 S^2_y + \left(\frac{n'-n}{nn'}\right)\left(\frac{N-n'}{Nn'-N+n'}\right)^2 S^2_y - 2\left(\frac{N-n'}{Nn'-N+n'}\right)RS_{xy} + R^2 S^2_x. \quad \dots 2(c)$$

$$Mse(t_4) = \left(\frac{N-n'}{Nn'}\right)S^2_y + \left(\frac{n'-n}{nn'}\right)(S^2_y - 2R\left(\frac{N-n'}{Nn'-N+n'}\right)S_{xy} + \left(\frac{N-n'}{Nn'-N+n'}\right)^2 S^2_x), \quad \dots 2(f)$$

and

$$Mse(\bar{y}) = \left(\frac{N-n}{Nn}\right)S^2_y. \quad \dots 2(g)$$

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

$$x = X(1 + \Delta_{\bar{x}}), \bar{y} = \bar{Y}(1 + \Delta_{\bar{y}}), \bar{x}^* = \bar{X}\left(1 - \left(\frac{n'}{N-n'}\right)\Delta_{\bar{x}}\right) \quad \text{and}$$

$$y^* = Y\left(1 - \left(\frac{n'}{N-n'}\right)\Delta_{\bar{y}}\right) \text{ by simple expansion.}$$

Then, let,

$$t_2 = \frac{\bar{y}^*}{x^*} x' = Y\left[\left(1 - \left(\frac{n'}{N-n'}\right)\Delta_{\bar{y}}\right)(1 + \Delta_{\bar{x}})\left(1 - \left(\frac{n'}{N-n'}\right)\Delta_{\bar{x}}\right)^{-1}\right] \quad \dots 3(a)$$

$$t_3 = \frac{\bar{y}}{x} x' = Y\left[\left(1 - \left(\frac{n'}{N-n'}\right)\Delta_{\bar{y}}\right)(1 + \Delta_{\bar{x}})(1 + \Delta_{\bar{x}})^{-1}\right] \quad \dots 3(b)$$

$$t_4 = \frac{\bar{y}}{x^*} x' = \bar{Y}\left[\left(1 + \Delta_{\bar{y}}\right)(1 + \Delta_{\bar{x}})\left(1 - \left(\frac{n'}{N-n'}\right)\Delta_{\bar{x}}\right)^{-1}\right] \quad \dots 3(c)$$

Hence,

$$Bias(t_2) = \frac{1}{X} \left(\frac{n'-n}{nn'}\right)\left(\frac{n'}{N-n'}\right)^2 (RS^2_x + S_{xy}). \quad \dots 3(d)$$

$$Bias(t_3) = \frac{1}{X} \left(\frac{n'-n}{nn'}\right)(RS^2_x + \left(\frac{n'}{N-n'}\right)S_{xy}). \quad \dots 3(e)$$

$$Bias(t_4) = \frac{1}{X} \left(\frac{n'-n}{nn'}\right)\left(\left(\frac{n'}{N-n'}\right)S_{xy} - R\left(\frac{n'}{N-n'}\right)^2 S^2_x\right). \quad \dots 3(f)$$

$$Mse(t_2) = \left(\frac{N-n'}{Nn'}\right)\left(\frac{n'}{N-n'}\right)^2 S^2_y + \left(\frac{n'-n}{nn'}\right)\left(\frac{n'}{N-n'}\right)^2 (S^2_y - 2RS_{xy} + R^2 S^2_x). \quad \dots 3(g)$$

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 $\bar{x}' = 1402.42$ 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, $N = 912$, $n' = 100$,

$$n = 30, \quad \bar{x}' = 1402.42, \quad \bar{x} = 1372.614, \quad \bar{y} = 733.347, \quad R = 0.53427, \\ S^2_x = 791376.06, \quad S^2_y = 151031.29, \quad S_{xy} = 268108.86$$

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,

$$N = 481, \quad n' = 134, \quad n = 28, \quad \bar{x}' = 5, \quad \bar{x} = 7.2857, \quad \bar{y} = 6.6946, \quad R = 0.9189, \\ S^2_x = 7.6190, \quad S^2_y = 18.2101, \quad S_{xy} = 4.6261$$

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 stremmas. (10 stremmas = 1 hectare). The data obtained are presented in Raj (1972), Table 6.1. Here, $N = 850$, $n' = 100$, $n = 30$, $\bar{x}' = 4.31$, $\bar{x} = 3.93$, $\bar{y} = 4.25$, $R = 1.0814$,

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

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 $\bar{x}' = 46.99$ 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,

$$N = 3500, \quad n' = 158, \quad n = 30, \quad \bar{x}' = 46.99, \quad \bar{x} = 47.87, \quad \bar{y} = 50.97, \quad R = 1.0648, \\ S^2_x = 850.0506, \quad S^2_y = 870.6551, \quad S_{xy} = 97.846$$

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 $\bar{X} = \theta\bar{x} + (1-\theta)\bar{x}^*$, $\bar{Y} = \theta\bar{y} + (1-\theta)\bar{y}^*$, $\theta = \frac{N-n'}{Nn'}$, $\bar{x} = \bar{X}(1 + \Delta_{\bar{x}})$, $\bar{y} = \bar{Y}(1 + \Delta_{\bar{y}})$, $\bar{x}^* = \bar{X}(1 - (\frac{N-n'}{Nn' - N + n'})\Delta_{\bar{x}})$ and $\bar{y}^* = \bar{Y}(1 - (\frac{N-n'}{Nn' - N + n'})\Delta_{\bar{y}})$.

	Pop 1	pop 2	pop 3	pop 4
$ Bias(t_1) $	4.5576	0.0179	0.0513	0.0552
$ Bias(t_2) $	0.0004	0	0	0
$ Bias(t_3) $	0.0409	0.0001	0.0002	0.0003
$ Bias(t_4) $	0.0406	0.0001	0.0005	0.0003
$Mse(\bar{y})$	4868.7718	6.1250	0.0290	28.7731
$Mse(t_1)$	3454.978	0.5541	0.10	26.0054
$Mse(t_2)$	0.2788	0.00002	0.00001	0.0010
$Mse(t_3)$	5211.199	0.1805	0.2426	2.8261
$Mse(t_4)$	4809.146	0.6112	0.2864	28.7390
$\frac{N-n'}{Nn'}$	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

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

	Pop 1	pop 2	pop 3	pop 4
$ Bias(t_1) $	4.5576	0.0179	0.0513	0.0552
$ Bias(t_2) $	0.0691	0.0027	0.0009	0.0001
$ Bias(t_3) $	0.5613	0.0069	0.0068	0.0026
$ Bias(t_4) $	0.4922	0.0043	0.0001	0.0025
$Mse(\bar{y})$	4868.7718	6.1250	0.0290	28.7731
$Mse(t_1)$	3454.978	0.5541	0.10	26.0054
$Mse(t_2)$	52.4002	0.0826	0.0018	0.0581
$Mse(t_3)$	4521.467	0.1803	0.1935	2.6576
$Mse(t_4)$	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 t_2 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 t_2 is better and therefore, it is preferred.

REFERENCES

- Cochran, W.G(1977): **Sampling Techniques**, 3rd Edition. John Wiley, New York.
 Okafor, F.C.(2002):- **Sample Survey Theory With Application**. AFRO-ORBIS PUBLICATIONS LTD. University of Nigeria. Nsukka.
 Raj, D. (1972):- **The Design of Sample Surveys**. Mc Graw-Hill Inc. USA
 Raj, D. and Chandhok, P. (1999):- **Sample Survey Theory**. Narosa Publishing House, London.
 Sodipo, A.A. (2003):- Estimation of the population mean using difference and regression estimators in a double sampling design with non-response. ABACUS. The Journal of the Mathematical Association of Nigeria. 30, (2B), **Mathematical Series**: 14 - 220.