The International Journal Of Engineering And Science (IJES)
||Volume|| 2 ||Issue|| 11 ||Pages|| 118-123 || 2013 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

On The Use of an Almost Unbiased Ratio Estimator in the Two-Phase Sampling Scheme

Olusegun Alfred O.

ABSTRACT

There has been an extensive use of auxiliary information in ratio estimation for estimating population parameters in forestry. Tree heights, measured during forest inventory programs is an important component required for determining a forest site's productivity. We consider the problem of obtaining precise estimates of the mean height of trees in a finite population. Using simple random sampling without replacement, data on the total height of Tectona grandis (THT), the study variate and its diameter-at-breast height (DBH), which is the auxiliary information were obtained via the two-phase sampling approach at the Oluwa Forest Reserve, Ondo State, Nigeria. The estimators considered in this study are the unbiased estimator, the usual ratio estimator, the Bahl and Tuteja (1991) estimator and the almost unbiased (AU) estimator. The Bias and the mean square error (MSE) of these estimators were then obtained to the first degree of approximation, using the Taylor's linearization method as described by Wolter (2007). The inequality which expresses the relationship obtained from a direct comparison of the MSE of estimators, an observation of findings,indicate that AU ratio estimator has the highest percent relative gain in efficiency.

KEYWORDS: *Auxiliary Information,Mean Square Error, Ratio Estimator, Tectona grandis, Two-Phase Sampling*

Date of Submission**:** 11, September - 2013 Date of Acceptance: 10, November - 2013

I. INTRODUCTION

Demand of statistics on various facets of economy in all the countries is increasing at a fast rate. In developing countries, basic data are needed for formulating development plans and for subsequent assessment of their progress. For this purpose, it becomes necessary to collect relevant information regarding the different sectors of economy for each of the administrative divisions. With respect to wood and timber in the forestry and paper industry, data on various aspects of its cultivation are needed for planning, fixing the target of production and for assessing the progress of various development programmes being conducted in the sector. Census, which is a complicated and costly operation, cannot be operated on a very frequent interval in order to obtain the needed information. Hence, this necessity, which has led in part, to this research work, the considerable development in the applications of efficient and modern sampling techniques based on probability theory.

Teak trees with botanical name *Tectona grandis* is a genus of tropical hardwood trees in the family *verbenaceae* native to south and south-east of Asia, the timber from teak is used in shipbuilding, boat decks and in the manufacture of outdoor furniture and other articles where weather resistance is desired. In south-western Nigeria, *Tectona grandis* is one of the most prominent species in the man-made forests. It is regarded as a very suitable specie for the rapid production of large volumes of timber, fuel-wood and poles of uniform and desirable quality (Akindele 1991), e.g. PHCN poles. Teak is also used extensively to make doors and window frames as it is resistant to termite attack Oyamakin et al (2010).

The concept of double sampling was first introduced by Neyman (1938). He felt the need of double sampling technique while examining the problem of stratification. The technique of stratification improves the precision of estimate of character under study provided it is possible to choose an appropriate character highly correlated with the character under study for the purpose of stratification.

Hidiroglou and Sarndal (1998) said that double sampling is a cost effective sampling design, and precision of ratio and regression estimates of study variable under two-phase sampling increases if there is a high degree of correlation between the auxiliary variable and study variable.

Double sampling is usually presented under the assumption that one of the samples is nested within the other; this method in particular is known as two-phase sampling, which is precisely a sampling scheme involving two distinct phases, in the first of which information about (a) particular variable(s) of interest (the auxiliary information) is collected on all members of the sample, and in the second, information about other variables is collected on a sub-sample of the individuals in the original sample.

This technique consists of taking a large sample of size n' by simple random sampling without replacement (SRSWOR) to estimate population mean of auxiliary variable while a sub-sample of n out of units is drawn by SRSWOR to observe the characteristic under study.

The use of the auxiliary information in ratio estimation can be used to achieve greater precision in estimation as it takes advantage of the correlation between the auxiliary variate $(say\ x)$ and the variate of interest$(say\ y)$. When information about the auxiliary variable is available, it increases the precision of the population variance.

Snedecor and King (1942) mentioned the application of a two-phase sampling procedure for determination of corn yield. They found out that it was easier and cheaper to count the number of ears of corn in a given unit area than to harvest the yield and obtain the dry weight of kernels. The high cost of making dry weight determination led to the use of two-phase sampling in which ears were not counted and measured in many fields but harvested in only a portion of fields, thus, taking advantage of the correlation between study variable (the dry weight of kernels) and the auxiliary variable(length-diameter of the ear).

Hartley and Ross (1954) developed exact ratio estimator. Freese (1962) presented detailed description of its application in forestry. He selected an auxiliary variable to achieve primary objective of two-phase sampling i.e. to reduce total inventory time without sacrificing the precision about the point estimate. Basal area is commonly utilized as auxiliary variable with two-phase sampling for volume estimates. This is because of the high correlation between basal area and volume and also the fact that basal area can be determined very quickly. Singh et al. (2007) have suggested modified ratio estimators by using different pairs of known value of population parameter(s).Singh et al. (2008) then proposed the Almost Unbiased (AU) estimator, a modification which utilises a linear combination of the usual ratio estimator and the exponential estimator due to Bahl and Tuteja (1991).

II. MATERIALS AND METHODS

The data utilised in this study is primary in nature. It was obtained at the Oluwa Forest Reserve (OFR). The study sites are located within Oluwa Forest Reserve (OFR). OFR is located in Odigbo Local government area of Ondo state, Nigeria within latitudes 06^0 52' and 7^0 20' N;and longitude 3^0 45' and 4^0 32' E. Oluwa Forest Reserve is approximately 828 km^2 with much of it lying approximately between 300 and 600m above sea level (Iloeje 1981).

The sampling frame consists of a population of 481 trees, in which 204 initial sample trees of Teak were measured for diameter-at-breast height (DBH) alone and subsequently, 60 second stage samples were then measured for both the Total Height of Tree (THT) and DBH. Both samples were collected using SRSWOR and their sizes determined from a pilot survey of 30 trees using an optimum allocation in two-phase sampling method.

The correlation value ρ which is used in analysing the improvement in efficiency due to ratio estimation and also determining the optimal samples sizes was obtained from a pilot survey of 30 trees in which information was collected on both Diameter-at-Breast Height (DBH) and the Total Height of Tree (THT).

2.1 The Almost Unbiased (AU) Ratio Estimator

Considering a finite population; $U = U_1, U_2, \ldots, U_N$ of N units. Let $(y_i, x_i), i = 1,2, \ldots, n$ denote the values of the units included in a sample of size n, drawn by SRSWOR.
In order to have a survey estimate of the population mean $\bar{Y}$ of the study character y, assuming the knowledge of the population mean $\bar{X}$ of the auxiliary character x, we have;

$$t_o = \bar{y} \quad (1)$$

$$t_{rs} = \bar{y}\left(\frac{a\bar{X}+b}{a\bar{x}+b}\right) \quad (2)$$

And $t_{rse} = \bar{y}exp\left[\frac{(a\bar{X}+b)-(a\bar{x}+b)}{(a\bar{X}+b)+(a\bar{x}+b)}\right]$ (3)

Such that $t_o, t_{rs}, t_{rse} \epsilon W_r$, where W_r denotes the set of all possible ratio type estimators for estimating the population mean $\bar{Y}$.

The Almost Unbiased (AU) estimator is given as;

$$t_w = \omega_0\bar{y} + \omega_1 t_{rs} + \omega_2 t_{rse} \quad (4)$$

And this is a linear variety if; $t_w = \omega_0\bar{y} + \omega_1 t_{rs} + \omega_2 t_{rse} \epsilon W,$ (5)

$$\text{for}\sum_{i=0}^{2}\omega_i = 1, \quad \omega_i \epsilon R \quad (6)$$

Where $\omega_i, i = 0,1,2$ denotes statistical constants and R denotes the set of real numbers.

Noting that setting the constant a and b in $(2)\&(3)$ to 1 and 0 respectively, transforms t_{rs} to the usual ratio estimator, $\bar{y}\frac{\bar{X}}{\bar{x}}$ and t_{rse} correspondingly to the Bahl and Tuteja (1991) estimator, $\bar{y}\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)$.

2.2 The AU ratio estimator in two-phase sampling

When $\bar{X}$ is unknown, it is sometimes estimated from a preliminary large sample of size n' on which only the characteristic x

Let $\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i$ denote the sample mean of x based on the first sample of size n', $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, be the sample means of y and x respectively, based on a second phase sample of size $n(n < n')$. In two-phase sampling, the estimator t_w will take the following form;

$$t_{wd} = \omega_{0d}\,\bar{y} + \omega_{1d}\,t_{rsd} + \omega_{2d}\,t_{rsed} \;\epsilon\; W_d \quad (7)$$

$$\text{for}\sum_{i=0}^{2}\omega_{id} = 1, \quad \omega_{id} \epsilon R \quad (8)$$

Where $t_{rsd} = \bar{y}\left(\frac{a\bar{x}'+b}{a\bar{x}+b}\right)$ and $t_{rsed} = \bar{y}exp\left[\frac{(a\bar{x}'+b)-(a\bar{x}+b)}{(a\bar{x}'+b)+(a\bar{x}+b)}\right]$.

Noting that setting the constants a and b in the above equations to 1 and 0, transforms t_{rsd} to the usual ratio estimator, $\bar{y}\frac{\bar{x}'}{\bar{x}}$ and t_{rsed} correspondingly to the Bahl and Tuteja (1991) estimator, $\bar{y}\left(\frac{\bar{x}'-\bar{x}}{\bar{x}'+\bar{x}}\right)$, both for use in two-phase sampling.

The estimators with their biases and MSE's (under two-phase sampling) are given below:

- **The Usual Ratio Estimator t_{rsd}**

$$t_{rsd} = \bar{y}\frac{\bar{x}'}{\bar{x}} \quad (9)$$

$$Bias(t_{rsd}) = \bar{Y}f_3(\theta^2 C_x^2 - \theta\rho\, C_x C_y) \quad (10)$$

$$MSE(t_{rsd}) = \bar{Y}^2\left[f_1 C_y^2 + f_3\left(\theta^2 C_x^2 - 2\theta\rho C_x C_y\right)\right] \quad (11)$$

Where; $f_1 = \left(\frac{1}{n} - \frac{1}{N}\right), f_3 = \left(\frac{1}{n} - \frac{1}{n'}\right), C_y = \frac{S_y}{\bar{Y}},\ C_x = \frac{S_x}{\bar{X}},\ \theta = \frac{a\bar{X}}{a\bar{X}+b},\ \rho = \frac{S_{xy}}{S_x S_y}$

- **The Bahl and Tuteja (1991) Estimator t_{rsed}**

$$t_{rsed} = \bar{y}\left(\frac{\bar{x}'-\bar{x}}{\bar{x}'+\bar{x}}\right) \quad (12)$$

$$B(t_{rsed}) = \bar{Y}f_3\left(\frac{3\theta^2 C_x^2}{8} - \frac{\theta\rho C_x C_y}{2}\right) \quad (13)$$

$$MSE(t_{rsed}) = \bar{Y}^2\left[f_1 C_y^2 + f_3\left(\frac{\theta^2 C_x^2}{4} - \theta\rho C_x C_y\right)\right] \quad (14)$$

- **The Almost Unbiased Estimator t_{wd}**

$$t_{wd} = \omega_{0d}\,\bar{y} + \omega_{1d}\,t_{rsd} + \omega_{2d}\,t_{rsed}$$

$$B(t_{wd}) = \bar{Y}f_3\left[\theta^2\left(\omega_{1d} + \frac{3\omega_{2d}}{8}\right)C_x^2 - \omega_d\theta\rho\, C_x C_y\right] \quad (15)$$

$$MSE(t_{wd}) = \bar{Y}^2\left[f_1 C_y^2 + f_3\left(\theta^2\omega_d^2 C_x^2 - 2\theta\,\omega_d\rho C_x C_y\right)\right] \quad (16)$$

Given that: $\omega_{0d} = \frac{\Delta_{0d}}{\Delta_{rd}}$, $\omega_{1d} = \frac{\Delta_{1d}}{\Delta_{rd}}$ and $\omega_{2d} = \frac{\Delta_{2d}}{\Delta_{rd}}$

Where: $\Delta_{rd} = B(t_{rsed}) - \frac{1}{2} B(t_{rsd})$ (17)

$$\Delta_{0d} = B(t_{rsed})(1-k) + B(t_{rsd})\left(k - \frac{1}{2}\right) \quad (18)$$

$$\Delta_{1d} = k.B(t_{rsed}) \quad (19)$$

$$\Delta_{2d} = -k.B(t_{rsd}) \quad (20)$$

2.3 Analysis of Improvement In Efficiency Due To Ratio Estimation:

From Cochran (1977), it is shown that the ratio estimator improves efficiency if;

$$\rho_{xy} > \left(\frac{1}{2}\right)\frac{c.v(x)}{c.v(y)} \quad (21)$$

$$\therefore \quad 0.837385 > \left(\frac{1}{2}\right)\frac{0.45887}{0.34714}$$

$$0.837385 > \left(\frac{1}{2}\right)1.32186$$

$$0.837385 > 0.66093$$

Hence, data obtained from the study site pilot survey, shows that the ratio estimator which is being considered for use in estimating the population total in this research will improve efficiency as compared to the mean per unit estimator $\bar{y}$.

2.4 Critical range of assumed Costs:

Using an assumed cost of $c' = N0.20$, being the cost for obtaining information on DBH and a cost $c = N1.00$, for the THT of each tree during the pilot survey.

Equation (22) from Cochran (1977)gives the critical ranges of c/c' for a given ρ that makes two-phase sampling more profitable.

$$\frac{c}{c'} > \frac{\left(1+\sqrt{1-\rho^2}\right)^2}{\rho^2} \quad (22)$$

$$=> \frac{1}{0.2} > \frac{\left(1+\sqrt{1-0.837385^2}\right)^2}{0.837385^2}$$

$$5 > \frac{1.54661354}{0.701214}$$

$$5 > 2.205623$$

It is shown that the assumed costs satisfies the critical ranges c'/c that make double sampling more profitable for a given $\rho = 0.837385$

To compute for the optimum sample sizes, we have;

$$n' = n\sqrt{\left(c/c'\right)\left[\frac{\rho^2}{(1-\rho^2)}\right]} \quad (23)$$

$$n = \frac{c}{c+\sqrt{cc'\rho^2/(1-\rho^2)}} \quad (24)$$

(23) and (24) due to Raj and Chandhok (1998), are the optimal values of n' and n given a correlation value ρ between the auxiliary variable and the study variable. Hence, n' and n were conveniently chosen as 204 and 60 respectively.

Based on the first phase sample, the following parameters were estimated as shown in Table 1.

Table 1: Descriptive Statistics of The First Phase Samples

Name	Variable	n'	Mean$(\bar{x}')$	Variance	Standard Deviation
Diam.-at-Breast Height (DBH)	x'	204	0.28810294	0.0089842244	0.0947851485

A second phase sample n of size 60 was taken from the first phase samplen', which records only the measure of height (in metres) of the study variable y, the following parameters were measured and are shown in Table 2.

Table 2:Descriptive Statistics of The Second Phase Samples

Name	Variable	n	Mean (m)	Variance	Stand. Deviation
DBH	x	60	$\bar{x} = 0.22479833$	$s_x^2 = 0.0066576459$	$s_x = 0.0815943989$
THT	y	60	$\bar{y} = 18.039$	$s_y^2 = 22.761951525$	$s_y = 4.7709487029$

III. RESULTS AND DISCUSSION

The values of the Average Height of Teak trees in Oluwa Forest Reserve, obtained using the various estimators considered in this study is summarized in the table below.

Table 3: Values of The Population Mean Obtained Using Different Estimators

Estimator	Population Mean Estimates
t_0	18.039
t_{rsd}	20.957
t_{rsed}	20.40869
t_{wd}	21.6010

The Bias, a measure of deviation of the estimates from the true value of the population parameter under study in this research work are presented in Table 5.1.2 below;

Table 4: Bias of Estimators

Estimator	Bias
t_0	-
t_{rsd}	0.009612
t_{rsed}	0.001161
t_{wd}	0.00000179

Table 5 shows the Variance of the unbiased mean per unit estimator$t_0 = \bar{y}$ and Mean Square Errors (MSEs) of the usual ratio estimatort_{rsd}, the Bahl and Tuteja (1991) estimatort_{rsed} and the AU ratio estimator.

Table 5: Variance/MSEs of Estimators

Estimators	Var*/MSE
t_0^*	0.372194
t_{rsd}	0.1997189
t_{rsed}	0.1586222
t_{wd}	0.14407924

The inequality below shows the relationship among various estimators being considered;
$Var(t_0) \geq MSE(t_{rsd}) \geq MSE(t_{rsed}) \geq MSE(t_{wd}) \geq MSE\left[(t_{wd})_{opt}\right]$

This indicates an agreement with previous literatures which holds that the Bahl & Tuteja (1991) estimatort_{rsed} is more precise than the usual ratio estimatort_{rsd}.

The Percent Relative Gain in Efficiencies (PRGEs) of the estimators$t_{rsd}, t_{rsed}, t_{wd}$ and $(t_{wd})_{opt}$all computed with respect to the unbiased estimatort_0 are listed in Table 6.

Table 6: Percent Relative Gain in Efficiencies (PRGE) of Estimators

Estimators	$PRGE(.,t_0)$
t_{rsd}	46.34

t_{rsed}	57.38
t_{wd}	61.29

- The correlation between the variables THT And DBH is strong and positive with a value of $\rho = 0.837$
- The estimate of the population mean height of trees in the Oluwa forest reserve, Ondo State, Nigeria, using the almost unbiased ratio estimator is 21.601m
- The Mean Squared Error (MSE) of the estimates based on the AU estimator is relatively lowest with respect to the Bahl and Tuteja (1991) estimator and the usual ratio estimator, all under the two-phase sampling scheme and given the same assumed costs.

IV. CONCLUSION

This shows clearly that the Almost Unbiased Ratio Estimatort_{wd} is more efficient than other estimators considered with considerable gain in efficiency.

Hence we can conclude that the estimatort_{wd} is to be preferred in practice over the unbiased estimator$\bar{y}$, the usual ratio estimatort_{rsd} and the Bahl and Tuteja (1991) estimatort_{rsed}, when estimating population means, not only in agricultural studies but also in food processing, education and other sectors where a decent level of precision is needed for obtaining estimates required for planning purposes and decision making.

For further study, one would suggest the development of a modification of the AU estimator for use in stratification, such that the enhancement of precision which come about as result of classifying the population based on their shared attributes could be exploited.

REFERENCES

[1]. Bahl, S. and Tuteja, R.K. 1991. Ratio and product type exponential estimator. *Journal Information and Optimization Sciences* 12.1: 159-163.

[2]. Wolter, K.M. 2007. *Introduction to Variance Estimation*. 2nd Ed. Springer, New York.

[3]. Akindele, S.O. 1991. Development of a site index equation for teak plantations in south-western Nigeria.*Journal of Tropical Forest Science* 4.2: 162-169.

[4]. Oyamakin, O.S., Julius, A.J., Asaolu, J.O. and Oyeleye, B. 2010. Sampling techniques procedures in sustainable forest management; a two-phase sampling approach.*American Journal of Scientific and Industrial Research*.667-670.

[5]. Neyman, J. 1938. Contribution to the theory of sampling human populations.*Journal of the American Statistical Association*, *33*.201: 101-116.

[6]. Hidiroglou, M.A and Sarndal, C.E. 1998. Use of auxiliary information for two-phase sampling.*Survey Methodology* 24: 873-878.

[7]. Snedecor, G. W. and King, A. J. 1942. Recent developments in sampling for agricultural statistics. *Journal of the American Statistical Association*, *37*.217: 95-102.

[8]. Hartley, H.O. and Ross, A. 1954. Unbiased ratio estimators.*Nature* 174: 270-271.

[9]. Freese, F. (1962). Elementary forest sampling. *Agriculture handbook/Forest Service.USDA* 232.

[10]. Singh, R., Chauhan, P., Smarandache, F., &Sawan, N. (2007). *Auxiliary information and a priori values in construction of improved estimators*.Infinite Study.

[11]. Singh, R, Kumar, M. andSmarandache, F. 2008. Almost unbiased estimator for estimating population mean using known value of some population parameter(s).*Pakistani Journal of Statistics and Operations Research* 4.2: 63-76.

[12]. Iloeje, N.P.: *A New Geography of Nigeria*. New Revised Edition, Longman, Great Britain (1981).

[13]. Cochran, W.G. 1977.*Sampling Techniques* 3rd Ed. John Wiley & Sons, Inc. London.

[14]. Raj, D. and Chandok, P. 1998. *Sample Survey Theory*. Narosa Publishing, New Delhi.