ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 1, January 2013

Efficiency of Ratio Estimators in Stratified Random Sampling Using Information on Auxiliary Attribute

Ran Vijay Kumar Singh, Ahmed Audu

Abstract - In the present paper some ratio-type estimators in stratified random sampling using auxiliary attribute has been proposed. Theoretically, we obtain the biases and mean square errors (MSE) for these estimators to first order approximation and compare them with the MSE of the traditional combined ratio estimate. By this comparison, we demonstrate that the proposed estimators are more efficient and less bias than combined ratio estimate in all conditions. In addition, this theoretical result is supported by a numerical example.

Keywords: **Bias, MSE, Proportion, Stratified random sampling.**

I. INTRODUCTION

The precision of an estimator can be increase by use of auxiliary information when study variable is highly correlated with auxiliary variable. There exist situations when information is available in the form of attribute, which is highly correlated with. For example;

(a) Amount of milk produced and a particular breed of cow,
(b) Amount of yield of wheat crop and a particular variety of wheat,
(c) Sex and height of the persons, etc.(Jhajj et. al.)

Consider a random sample of size $n = n_1 + n_2 + ... + n_k$ to be taken from a population of size $N = N_1 + N_2 + ... + N_k$ stratified into k strata. Let a sample of size n_h, $(h = 1, 2, ..., k)$ be drawn by simple random sampling without replacement from a stratum h of size N_h. Let y_i and ϕ_i denote the observations on a random variable y and ϕ respectively for i^{th} unit $(i = 1, 2, ..., N)$. Suppose there is a complete dichotomy in the population with respect to the presence or absence of an attribute, say ϕ, and it is assumed that attribute ϕ takes only two values 0 and 1 as

$\phi_i = 1$, if ith unit of the population possesses attribute ϕ

$= 0$, otherwise

Then we have the following definition;

$A = \sum_{i=1}^{N} \phi_i$ - denotes the total number of units in the population possessing attribute ϕ.

$A_h = \sum_{i=1}^{N_h} \phi_{hi}$ - denotes the total number of units in the stratum h possessing attribute ϕ.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 1, January 2013

$a_h = \sum_{i=1}^{n_h} \phi_{hi}$ - denotes the total number of units in the sample drawn from stratum h possessing attribute ϕ.

$P = \frac{A}{N}$ - denotes the proportion of units in the population possessing attribute ϕ.

$P_h = \frac{A_h}{N_h}$ - denotes the proportion of units in the stratum h possessing attribute ϕ.

$p_h = \frac{a_h}{n_h}$ - denotes the proportion of units in the sample drawn from stratum h possessing attribute ϕ.

In stratified random sampling, the traditional combined ratio estimators for the population mean $\bar{Y}$ using auxiliary attribute is defined as;

$$T = \frac{\bar{y}_{st}}{p_{st}} P = R_n P \tag{1.1}$$

Where $\bar{y}_{st} = \sum_{h=1}^{k} w_h \bar{y}_h$, $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_i$, $p_{st} = \sum_{h=1}^{k} w_h p_h$, $w_h = \frac{N_h}{N}$,

The Bias and Mean Square Error (MSE) of the traditional estimator are given by equations (1.2) and (1-3) respectively as;

$$Bias(T) \cong \frac{1}{P} \sum_{h=1}^{k} w_h{}^2 \gamma_h \left(R S_{\phi h}{}^2 - S_{y\phi h} \right) \tag{1.2}$$

$$MSE(T) \cong \sum_{h=1}^{k} w_h{}^2 \gamma_h \left(S_{yh}{}^2 + R^2 S_{\phi h}{}^2 - 2 R S_{y\phi h} \right) \tag{1.3}$$

Where $\gamma_h = \left(\frac{1}{n_h} - \frac{1}{N_h} \right)$, $w_h = \frac{N_h}{N}$ is the weight of stratum h, $R = \frac{\bar{Y}}{P}$ is the population ratio, n_h is the number of units in sample from stratum h, N_h is the population size of stratum h, S_{yh}^2 is the population variance in stratum h, $S_{\phi h}^2$ is the population variance of auxiliary attribute in stratum h and $S_{y\phi h}$ is the population covariance between auxiliary attribute and variable of interest in stratum h.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 1, January 2013

II. SUGGESTED ESTIMATORS

We suggest the following estimators;

$$T_1 = \frac{\sum_{h=1}^{k} w_h \left(\bar{y}_h - b_{\phi h} \left(p_h - P_h \right) \right)}{\sum_{h=1}^{k} w_h p_h} P$$

$$T_2 = \frac{\sum_{h=1}^{k} w_h \left(\bar{y}_h - b_{\phi h} \left(p_h - P_h \right) \right)}{\sum_{h=1}^{k} w_h \left(p_h + \beta_2 \left(\phi \right) \right)} \left(P + \beta_2 \left(\phi \right) \right)$$

$$T_3 = \frac{\sum_{h=1}^{k} w_h \left(\bar{y}_h - b_{\phi h} \left(p_h - P_h \right) \right)}{\sum_{h=1}^{k} w_h \left(p_h + C_p \right)} \left(P + C_p \right)$$

$$T_4 = \frac{\sum_{h=1}^{k} w_h \left(\bar{y}_h - b_{\phi h} \left(p_h - P_h \right) \right)}{\sum_{h=1}^{k} w_h \left(p_h + \rho_{pb} \right)} \left(P + \rho_{pb} \right)$$

Where C_p the coefficient of variation is, $B_2(\phi)$ is the coefficient of kurtosis and ρ_{pb} is the point biserial correlation coefficient of the auxiliary attribute.

$$b_{\phi h} = \frac{s_{y\phi h}}{s_{\phi h}^2}, \; s_{\phi h}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (\phi_i - p_h)^2 \text{ and } s_{y\phi h} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (\phi_i - p_h)\left(y_i - \bar{Y}_h \right)$$

Remark 1: when we put $b_{\phi h} = 0$, $\beta_2(\phi) = 0$ in T_2, $C_p = 0$ in $T_3 = 0$ and $\rho_{pb} = 0$ in T_4, the proposed estimators reduce to traditional estimator.

The Bias and Mean Square Error (MSE) to first order approximation of the proposed estimators are given by equations (2.1) and (2.2) respectively as;

$$Bias\left(\hat{\bar{T}}_i^* \right) \cong \tau_i^2 \bar{Y} \sum_{h=1}^{k} w_h^2 \gamma_h S_{\phi h}^2 , \; (i = 1, 2, 3, 4) \tag{2.1}$$

Where $\tau_1 = \frac{1}{P}$, $\tau_2 = \frac{1}{P + \beta_2(\phi)}$, $\tau_3 = \frac{1}{P + C_p}$ $\tau_4 = \frac{1}{P + \rho_{pb}}$,

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 1, January 2013

$$MSE\left(\hat{\bar{T}}_i^*\right)=\sum_{h=1}^{k}w_h^2\gamma_h\left(R_i^2S_{\phi h}^2+S_{yh}^2\left(1-\rho_{pbh}^2\right)\right),\ \left(i=1,2,3,4\right) \qquad (2.2)$$

Where $R_1=\dfrac{\bar{Y}}{P}$, $R_2=\dfrac{\bar{Y}}{P+\beta_2\left(\phi\right)}$, $R_3=\dfrac{\bar{Y}}{P+C_p}$, $R_4=\dfrac{\bar{Y}}{P+\rho_{pb}}$

III. EFFICIENCY COMPARISONS

We compare the traditional estimator T with the proposed estimators T_i $(i=1,2,3,4)$ and the conditions for which the proposed estimators will have the least mean square errors were obtained as follows;

$$MSE\left(T_i\right)<MSE\left(T\right)$$

$$\sum_{h=1}^{k}w_h^2\gamma_h\left(R_i^2S_{\phi h}^2+S_{yh}^2-\rho_{pbh}^2S_{yh}^2\right)<\sum_{h=1}^{k}w_h^2\gamma_h\left(S_{yh}^2+R^2S_{\phi h}^2-2RS_{y\phi h}\right)$$

$$R_i^2\sum_{h=1}^{k}w_h^2\gamma_hS_{\phi h}^2-\sum_{h=1}^{k}w_h^2\gamma_h\rho_{pbh}^2S_{yh}^2<R^2\sum_{h=1}^{k}w_h^2\gamma_hS_{\phi h}^2-2R\sum_{h=1}^{k}w_h^2\gamma_hS_{y\phi h}$$

Let $A=\sum_{h=1}^{k}w_h^2\gamma_hS_{\phi h}^2$, $B=\sum_{h=1}^{k}w_h^2\gamma_h\rho_{pbh}^2S_{yh}^2$, $C=\sum_{h=1}^{k}w_h^2\gamma_hS_{y\phi h}$

Then we have

$$R_i^2A-B<R^2A-2RC$$

$$R_i^2A-R^2A-B+2RC<0$$

$$A\left(R_i^2-R^2\right)-B+2RC<0$$

Where there are two conditions as follows;

(i) When $\left(R_i^2-R^2\right)>0$

$$A-\frac{B-2RC}{R_i^2-R^2}<0$$

$$A<\frac{B-2RC}{R_i^2-R^2}$$

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 1, January 2013

(ii) When $\left(R_i^2 - R^2\right) < 0$

$$A - \frac{B - 2RC}{R_i^2 - R^2} > 0$$

$$A > \frac{B - 2RC}{R_i^2 - R^2}$$

When either of these conditions is satisfied, the proposed estimators T_i $\left(i = 1, 2, 3, 4\right)$ will be more efficient than the traditional estimator T .

IV. EMPIRICAL STUDY

The information on 1500 Students taken from Students Pre-Medical Registration, Usmanu Danfodiyo University, Sokoto (2011/2012 Session) was used as data for empirical study. The height of the students is the variable of interest and their gender was used as auxiliary attribute (Male=1 and Female=0). The stratification is based on the faculties. By using Neyman allocation (Cochran, 1977),

$$n_h = n \frac{N_h S_{yh}}{\sum_{h=1}^{k} N_h S_{yh}} \tag{4.1}$$

We have computed sample size in each stratum. The summary information on the empirical data are given below;

$y =$ Height of the students

$\phi =$ Gender of the students

$N = 1500$, $n = 300$, $\bar{Y} = 158.572$, $P = 0.71$, $\rho_{pb} = 0.541$, $C_p = 1.564$, $\beta_2(\phi) = 1.15$

Table 1: Data Statistics

Stratum No.	Faculty	Stratum Size	Sample Size	Stratum Parameters
1	Agric	96	18	$S_{yh}^2 = 69.813$ $S_{\phi h}^2 = 0.1344$ $S_{y\phi h} = 1.562$ $\rho_{pbh} = .51$
2	Vetnary Medicine	100	18	$S_{yh}^2 = 61.328$ $S_{\phi h}^2 = 0.196$ $S_{y\phi h} = 1.471$ $\rho_{pbh} = .424$
3	Education	288	58	$S_{yh}^2 = 76.159$ $S_{\phi h}^2 = 0.189$ $S_{y\phi h} = 1.964$ $\rho_{pbh} = .518$
4	Art and Islamic Studies	198	43	$S_{yh}^2 = 90.651$ $S_{\phi h}^2 = 0.221$ $S_{y\phi h} = 2.452$ $\rho_{pbh} = .548$
5	Law	136	29	$S_{yh}^2 = 85.334$ $S_{\phi h}^2 = 0.22$ $S_{y\phi h} = 2.538$ $\rho_{pbh} = .586$

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 1, January 2013

6	Col.of Health Sci.	95	19	$S_{yh}^2 = 77.099$ $S_{\phi h}^2 = 0.234$ $S_{y\phi h} = 2.384$ $\rho_{pbh} = .561$
7	Social Sciences	96	20	$S_{yh}^2 = 78.99$ $S_{\phi h}^2 = 0.223$ $S_{y\phi h} = 2.707$ $\rho_{pbh} = 0.645$
8	Sciences	299	61	$S_{yh}^2 = 77.54$ $S_{\phi h}^2 = 0.241$ $S_{y\phi h} = 2.472$ $\rho_{pbh} = 0.572$
9	Management Sci.	192	34	$S_{yh}^2 = 60.248$ $S_{\phi h}^2 = 0.176$ $S_{y\phi h} = 1.393$ $\rho_{pbh} = .428$

Table 2: Relative Biases and Efficiencies

Estimator	**Bias**	**Rel. Bias**	**MSE**	**Relative efficiency**	**Popn. Ratio (R)**	$R_i^2 - R^2$	$X = \frac{B-2RC}{R_i^2 - R^2}$	**Cond. for efficiency** $A > X$ **or** $A < X$
$\hat{\bar{T}}$	0.1403	100.0	25.919	100.0	223.34	–	–	–
$\hat{\bar{T}}_1^*$	0.1730	81.10	27.581	93.97	223.34	0	Undefined	Not satisfied
$\hat{\bar{T}}_2^*$	0.0252	556.8	4.142	625.76	85.25	-42613.19<0	$5.73\text{x}10^{-5}$	satisfied
$\hat{\bar{T}}_3^*$	0.0479	292.9	7.440	348.37	117.55	-36062.75<0	$6.78\text{x}10^{-5}$	satisfied
$\hat{\bar{T}}_4^*$	0.0557	251.9	8.982	288.57	126.76	-33812.66<0	$7.23\text{x}10^{-5}$	satisfied

From Table 2, we observed that the proposed estimators which use some known value of population proportion perform better and less bias than the traditional estimator.

V. CONCLUSION

We have proposed some ratio-type estimators in stratified random sampling for estimating population mean. The estimators use some known values of population proportion. We have obtained their MSE equations and compared with MSE of the traditional estimator in theory. By this comparison, the conditions which the estimators have smaller MSE with respect to each other have been found. These theoretical conditions are also satisfied by the results of an application with original data. For practical purposes the choice of the estimator depends upon the availability of the population parameters.

REFERENCES

[1] Jhajj H. S., Sharma M. K. and Grover L. K. (2006): A family of estimators of population Mean using information on auxiliary attribute. Pak. J. Stat., 22 (1), 43-50.

[2] Searls D. T. (1964): The utilization of known coefficient of variation in the estimation procedure. Journal of American Statistical Association, 59, 1125-1226.

[3] Searls D. T. and Intarapanich P. (1990): A note on a estimator for the variance that utilizes the kurtosis. The American Statistician, 44, 4, 295-296.

[4] Sen A. R. (1978): Estimation of the population mean when the coefficient of variation is known. Commun. Statist. , Theory – Meth. A (7), 657-672.

[5] Singh G. N. (2003): On the improvement of product method of estimation in sample surveys. Jour. Ind. Soc. Agric. Statistics, 56(3), 267-275.

[6] Singh H. P. and Tailor R. (2003): Use of known correlation coefficient in estimating the finite population mean. Statistics in Transition, 64, 555-560.

[7] Singh H. P. and Upadhyaya L. N. (1999): Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal, 41, 5, 627-636.

[8] Singh H. P., Tailor R. and Kakran M. S. (2004): An improved estimator of population means using power transformation. Jour. Ind. Soc. Agr. Stat. Math. , 25, 51-55.

[9] Singh J., Pandey B. N. and Hirano K. (1973): On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. Ann. Inst. Stat. Math., 25, 51-55.

[10] Singh V. K., Singh R. V. K. and Singh G. N.(1991): Efficiency comparison of modified class of ratio type estimators using coefficient of variation of auxiliary variable. Proc. Math. Soc., B. H. U. Vol. 7 , 105-112.

[11] Sisoda B. V. S. and Dwivedi V. K. (1981):A modified ratio estimator using coefficient of variation of auxiliary variable. Jour. Ind. Soc. Agr. Statistics, 33, 2, 13-18.

[12] Upadhayaya L. N. and Singh H. P. (1984): On the estimation of the population mean with known coefficient of variation. Biometrical Journal, 26 (8), 915-922.