Ratio-Type Estimators in Stratified Random Sampling using Auxiliary Attribute

R .V. K .Singh, A .Ahmed, *Member*, *IAENG*

Abstract— Some ratio-type estimators have been proposed in stratified random sampling using auxiliary attribute. The expressions for the bias and mean square errors of the proposed estimators have been derived up to first order of approximation. Comparisons have been made with traditional combined ratio estimator and it is shown that the proposed estimators are more efficient than combined ratio estimator under certain condition. For illustration, an empirical study has been carried out.

Keywords: Proportion, Bias, MSE, Ratio-type estimators, Stratified random sampling.

I. Introduction

The ratio estimator was observed to be more precise than the usual sample mean estimator under different conditions for estimating the population mean of the study character. Several researchers diverted their attention in the direction of using prior value of certain population parameters to find the estimates that are more precise. Searls (1964) used coefficient of variation of study character at estimation stage. In practice, coefficient of variation is seldom known. Motivated by Searls (1964) work, various authors including Sen (1978), Sesodiya and Dewivedi (1981) Singh et al (1991) and Upadhyaya and Singh (1984) used the known coefficient of variation of auxiliary character for estimating population mean of the study character in ratio method of estimation. Singh et al (1973) first made the use of prior value of coefficient of kurtosis in estimating the population variance of study character. Later used by Searls and Interapanich (1990). Recently Singh and Tailor (2003) proposed a modified ratio estimator by using the known value of correlation coefficient. Taking into consideration the point biserial correlation coefficient between auxiliary attribute and study variable, Jhajj et al (2006) and Singh, et al (2008) defined ratio estimators of population mean when the priori information on auxiliary variable possessing some attribute is available.

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 I 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.

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

Manuscript received Aug 22 , 2013; revised Jan15, 2014. .
R.V.K.Singh, Senior lecturer and Head, Department of Mathematics, Kebbi State University of Science and Technology, Aliero,Nnigeria, (e-mail: singhrvk13@gmail.com)
A.Aahmed, Assistant lecturer, Usmanu danfodiyo university, Sokoto, Nigeria (e-mail:ahmedgborom@yahoo.com).

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

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

$p_h=\dfrac{a_h}{n_h}$ - 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 using auxiliary attribute is defined as;

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

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

The Bias and Mean Square Error (MSE) of the traditional estimator are given by equations (1.2) and (1.3) respectively as;

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

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

Where $\gamma_h=\left(\dfrac{1}{n_h}-\dfrac{1}{N_h}\right)$, $w_h=\dfrac{N_h}{N}$ is the weight of stratum h, $R=\dfrac{\bar{Y}}{P}$ 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.

II. PROPOSED ESTIMATORS

The proposed estimators can be written in the form below;

$$\hat{\bar{T}}_i^*=\frac{\sum_{h=1}^{k}w_h(\bar{y}_h-b_{\phi h}(p_h-P_h))}{\sum_{h=1}^{k}w_h(m_1p_h+m_2)}(m_1P+m_2),\ (i=1,2,...,10) \tag{2.1}$$

Where $m_1(\neq 0)$ and m_2 are either real numbers or the functions of the known parameters of the attribute such as coefficient of variation C_p, coefficient of kurtosis $B_2(\phi)$ and point biserial correlation coefficient ρ_{pb}.

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

Remark 1: when we put $b_{\phi h}=0$, $m_1=1$ and $m_2=0$, the proposed estimators reduce to traditional estimator.

The following scheme presents the estimators of the population mean, which can be obtain by suitable choices of constants m_1 and m_2.

ESTIMATOR	VALUES OF	
	m_1	m_2
$\hat{\bar{T}}_1^*=\dfrac{\sum_{h=1}^{k}w_h\left(\bar{y}_h-b_{\phi h}\left(p_h-P_h\right)\right)}{\sum_{h=1}^{k}w_hp_h}P$	1	0
$\hat{\bar{T}}_2^*=\dfrac{\sum_{h=1}^{k}w_h\left(\bar{y}_h-b_{\phi h}\left(p_h-P_h\right)\right)}{\sum_{h=1}^{k}w_h\left(p_h+\beta_2(\phi)\right)}\left(P+\beta_2(\phi)\right)$	1	$\beta_2(\phi)$
$\hat{\bar{T}}_3^*=\dfrac{\sum_{h=1}^{k}w_h\left(\bar{y}_h-b_{\phi h}\left(p_h-P_h\right)\right)}{\sum_{h=1}^{k}w_h\left(p_h+C_p\right)}\left(P+C_p\right)$	1	C_p
$\hat{\bar{T}}_4^*=\dfrac{\sum_{h=1}^{k}w_h\left(\bar{y}_h-b_{\phi h}\left(p_h-P_h\right)\right)}{\sum_{h=1}^{k}w_h\left(p_h+\rho_{pb}\right)}\left(P+\rho_{pb}\right)$	1	ρ_{pb}
$\hat{\bar{T}}_5^*=\dfrac{\sum_{h=1}^{k}w_h\left(\bar{y}_h-b_{\phi h}\left(p_h-P_h\right)\right)}{\sum_{h=1}^{k}w_h\left(\beta(\phi)p_h+C_p\right)}\left(\beta(\phi)P+C_p\right)$	$\beta_2(\phi)$	C_p
$\hat{\bar{T}}_6^*=\dfrac{\sum_{h=1}^{k}w_h\left(\bar{y}_h-b_{\phi h}\left(p_h-P_h\right)\right)}{\sum_{h=1}^{k}w_h\left(C_pp_h+\beta_2(\phi)\right)}\left(C_pP+\beta_2(\phi)\right)$	C_p	$\beta_2(\phi)$

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$\hat{\bar{T}}_7^* = \dfrac{\sum_{h=1}^{k} w_h\left(\bar{y}_h - b_{\phi h}\left(p_h - P_h\right)\right)}{\sum_{h=1}^{k} w_h\left(C_p p_h + \rho_{pb}\right)}\left(C_p P + \rho_{pb}\right)$	C_p	ρ_{pb}
$\hat{\bar{T}}_8^* = \dfrac{\sum_{h=1}^{k} w_h\left(\bar{y}_h - b_{\phi h}\left(p_h - P_h\right)\right)}{\sum_{h=1}^{k} w_h\left(\rho_{pb} p_h + C_p\right)}\left(\rho_{pb} P + C_p\right)$	ρ_{pb}	C_p
$\hat{\bar{T}}_9^* = \dfrac{\sum_{h=1}^{k} w_h\left(\bar{y}_h - b_{\phi h}\left(p_h - P_h\right)\right)}{\sum_{h=1}^{k} w_h\left(\beta_2(\phi) p_h + \rho_{pb}\right)}\left(\beta_2(\phi) P + \rho_{pb}\right)$	$\beta_2(\phi)$	ρ_{pb}
$\hat{\bar{T}}_{10}^* = \dfrac{\sum_{h=1}^{k} w_h\left(\bar{y}_h - b_{\phi h}\left(p_h - P_h\right)\right)}{\sum_{h=1}^{k} w_h\left(\rho_{pb} p_h + \beta_2(\phi)\right)}\left(\rho_{pb} P + \beta_2(\phi)\right)$	ρ_{pb}	$\beta_2(\phi)$

The Bias and Mean Square Error (MSE) up to first order approximation of the proposed estimators are given by equations (2.2) and (2.3) respectively as;

$$Bias\left(\hat{\bar{T}}_i^*\right) \cong \tau_i^2 \bar{Y} \sum_{h=1}^{k} w_h^2 \gamma_h S_{\phi h}^2 \text{ , } (i = 1, 2, ..., 10) \qquad (2.2)$$

Where $\tau_1 = \dfrac{1}{P}$, $\tau_2 = \dfrac{1}{P + \beta_2(\phi)}$, $\tau_3 = \dfrac{1}{P + C_p}$,

$\tau_4 = \dfrac{1}{P + \rho_{pb}}$, $\tau_5 = \dfrac{\beta_2(\phi)}{P\beta_2(\phi) + C_p}$,

$\tau_6 = \dfrac{C_p}{PC_p + \beta_2(\phi)}$ $\tau_7 = \dfrac{C_p}{PC_p + \rho_{pb}}$,

$\tau_8 = \dfrac{\rho_{pb}}{P\rho_{pb} + C_p}$, $\tau_9 = \dfrac{\beta_2(\phi)}{P\beta_2(\phi) + \rho_{pb}}$,.

$$MSE\left(\hat{\bar{T}}_i^*\right) = \sum_{h=1}^{k} w_h^2 \gamma_h \left(R_i^2 S_{\phi h}^2 + S_{yh}^2\left(1 - \rho_{pbh}^2\right)\right),$$
$$(i = 1, 2, ..., 10) \qquad (2.3)$$

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

$R_4 = \dfrac{\bar{Y}}{P + \rho_{pb}}$, $R_5 = \dfrac{\bar{Y}\beta_2(\phi)}{P\beta_2(\phi) + C_p}$ $R_6 = \dfrac{\bar{Y}C_p}{PC_p + \beta_2(\phi)}$,

$R_7 = \dfrac{\bar{Y}C_p}{PC_p + \rho_{pb}}$, $R_8 = \dfrac{\bar{Y}\rho_{pb}}{P\rho_{pb} + C_p}$,

$R_9 = \dfrac{\bar{Y}\beta_2(\phi)}{P\beta_2(\phi) + \rho_{pb}}$, $R_{10} = \dfrac{\bar{Y}\rho_{pb}}{P\rho_{pb} + \beta_2(\phi)}$

III. EFFICIENCY COMPARISONS

We compare the traditional estimator $\hat{\bar{T}}$ with the proposed estimators $\hat{\bar{T}}_i^*$ $(i = 1, 2, ..., 10)$ and the conditions for which the proposed estimators will have the least mean square errors were obtained as follows;

$$MSE\left(\hat{\bar{T}}_i^*\right) < MSE\left(\hat{\bar{T}}\right)$$

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

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

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

$$C = \sum_{h=1}^{k} w_h^2 \gamma_h S_{y\phi h}$$

Then we have

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

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

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

Where there are two conditions as follows;

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

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

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

(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 $\hat{\bar{T}}_i^*$ $(i = 1, 2, ..., 10)$ will be more efficient than the traditional estimator $\hat{\bar{T}}$.

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, we have

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

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

$N = 1500$, $n = 300$, $\bar{Y} = 158.572$, $P = 0.71$, $\rho_{pb} = 0.541$, $C_p = 0.639$, $\beta_2(\phi) = 1.15$

Table II

RELATIVE BIASES AND EFFICIENCIES

Estimator	Bias	Rel. Bias	MSE	Relative efficiency	Popn. Ratio (R)	$X = \frac{B-2RC}{R_i^2 - R^2}$	Cond. for efficiency $A > X$ or $A < X$
$\hat{\bar{T}}$	0.14	100.0	25.9	100.0	223.3	–	–
$\hat{\bar{T}}_1^*$	0.17	81.10	27.6	94.0	223.3	Undefined	Not satisfied
$\hat{\bar{T}}_2^*$	0.03	556.8	4.1	625.8	85.3	5.7x10^{-5}	satisfied
$\hat{\bar{T}}_3^*$	0.05	292.9	7.4	348.3	117.4	6.8x10^{-5}	satisfied
$\hat{\bar{T}}_4^*$	0.06	251.9	8.9	288.6	126.8	7.2x10^{-5}	satisfied
$\hat{\bar{T}}_5^*$	0.05	257.4	8.8	295.2	125.3	7.2x10^{-5}	satisfied
$\hat{\bar{T}}_6^*$	0.01	1016.7	2.3	1107.6	63.2	5.3x10^{-5}	satisfied
$\hat{\bar{T}}_7^*$	0.03	389.7	5.9	442.9	101.9	6.2x10^{-5}	satisfied
$\hat{\bar{T}}_8^*$	0.02	575	4.0	646.2	83.9	5.7x10^{-5}	satisfied
$\hat{\bar{T}}_9^*$	0.06	224.1	10.1	257.4	134.3	7.7x10^{-5}	satisfied
$\hat{\bar{T}}_{10}^*$	0.01	1299.1	1.9	1390.5	55.9	5.2x10^{-5}	satisfied

Table I

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	$S_{yh}^2 = 85.334$ $S_{\phi h}^2 = 0.22$ $S_{y\phi h} = 2.538$ $\rho_{pbh} = .586$
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	$S_{yh}^2 = 60.248$ $S_{\phi h}^2 = 0.176$ $S_{y\phi h} = 1.393$ $\rho_{pbh} = .428$

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

V. CONCLUSION

In the present paper, we have developed some ratio-type estimators in stratified random sampling for estimating population mean by using information on auxiliary attributes. By comparison, it is found that the proposed estimators are more efficient and less bias than the traditional combined ratio estimator is. 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] A. R. Sen,“Estimation of the population mean when the coefficient of variation is known,” *Commun. Statist. ,Theory – Meth*. A (7), 657-672,1978.

[2} B. V. S.Sisodiya and V. K. Dewivedi,” modified ratio estimator using coefficient of variation of auxiliary variable,” *Jour. Ind. Soc. Agr. Statistics*, 33, 2, 13-18,1981.

[3] D. T Searls, “The utilization of known coefficient of variation in the estimation procedure,” *Journal of American Statistical Association*, 59, 1125-1226,1964.

[4] D. T. Searls and P.Intarapanich,“A note on an estimator for the variance that utilizes the kurtosis,” *The American Statistician*, 44, 4, 295-296,1990.

[5] H. P. Singh and R.Tailor,”Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*,6 4, 555-560,2003.

[6] H. S Jhajj, M. K Sharma and L. K., Grover,“A family of estimators of population mean using information on auxiliary attribute,” *Pak. J. Statist*., 22 (1), 43-50,2006.

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

[8] L. N. Upadhayaya and H. P. Singh,“On the estimation of the population mean with known coefficient of variation,” *Biometrical Journal*, 26 (8), 915-922.1984.

[9] R.Singh, P. Chauhan ,N. Sawan and F.Smarandache,“Ratio estimators insimple random sampling using information on auxiliary attribute,” *Pak. J. Stat. Oper.Res*. 4(1), 47-53,2008.

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

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)