Pak. J. Statist.
2006 Vol. 22 (3) pp 209-215

A NOTE ON THE PERFORMANCE OF SEVERAL ALMOST UNBIASED RATIO ESTIMATORS

L.N. Sahoo and **B.C. Das**
Department of Statistics
Utkal University
Bhubaneswar 751004, India.

and

J. Sahoo
Department of Statistics
S.K.C.G. College
Parlakhemundi 761200, India.

ABSTRACT

In this paper, a Monte Carlo study was undertaken to compare the performance of nine almost unbiased ratio estimators. The quality of an estimator is measured in the traditional way; namely with consideration of relative bias, efficiency, achieved coverage rate of the nominal 95% confidence intervals and approach to normality.

KEY WORDS

Almost unbiased ratio estimator, auxiliary variable, confidence interval, efficiency, kurtosis, Monte Carlo study, relative bias, simple random sampling, skewness.

AMS Subject Classification: 62 D05

1. INTRODUCTION

Consider a finite population of N units with observations $\{(y_i, x_i), i = 1, 2, ..., N\}$ on the study variable y and a positively correlated auxiliary variable x. We seek to estimate the population mean $\bar{Y}$ of $y-$values from a sample s, of size n, obtained through simple random sampling without replacement when the population mean $\bar{X}$ of $x-$values is known. A commonly used estimator in this context is the classical ratio estimator $t_R = \bar{y}\bar{X}/\bar{x}$, where $\bar{y}$ and $\bar{x}$ denote the sample means of $y-$ and $x-$values respectively. In general, t_R is biased. Therefore, several ratio-type estimators have been derived to satisfy the criterion of approximate unbiasedness called almost unbiased ratio (AUR) estimators. Many design- and model-motivated theoretical comparisons have been provided by several authors to compare the performances of various AUR estimators. But, the results of some comparisons derived through the Taylor linearization method [*e.g.*, Tin (1965)] are only in asymptotical forms. These results may not be useful for a choice among different estimators when very small samples are taken from populations of small or moderate size. As a counterpart to these theoretical comparisons, extensive empirical investigations of different ratio-type estimators have been undertaken in several papers, for example by Rao and Beegle (1967), Rao, J.N.K. (1969), Hutchinson (1971), Rao and Kuzik (1974), Royall and Cumberland (1981), Wu and Deng (1983). Our present study is empirical in nature and considers the following AUR estimators:

Quenouille's (1956) Estimator

$$t_1 = 2t_R - \tfrac{1}{2}\left(t_R^{(1)} + t_R^{(2)}\right),$$

where $t_R^{(1)}$ and $t_R^{(2)}$ are classical ratio estimators based on first and second halves of the sample.

Murthy-Nanjamma (1959) Estimator

$$t_2 = \frac{n(N-1)t_R - (N-n)\overline{r}\overline{X}}{(n-1)N}, \text{ where } \overline{r} = n^{-1}\sum_{i\in s}\frac{y_i}{x_i}.$$

Pascual's (1961) Estimator

$$t_3 = t_R + \frac{N-1}{N(n-1)}\left(\overline{y} - \overline{r}\overline{x}\right).$$

Beale's (1962) Estimator

$$t_4 = t_R\left(1+\theta\frac{s_{yx}}{\overline{y}\overline{x}}\right)\Bigg/\left(1+\theta\frac{s_x^2}{\overline{x}^2}\right),$$

where $\theta = n^{-1} - N^{-1}$, $s_{yx} = (n-1)^{-1}\sum_{i\in s}(y_i-\overline{y})(x_i-\overline{x})$ and $s_x^2 = (n-1)^{-1}\sum_{i\in s}(x_i-\overline{x})^2$.

Tin's (1965) Estimator

$$t_5 = t_R\left[1+\theta\left(\frac{s_{yx}}{\overline{y}\overline{x}} - \frac{s_x^2}{\overline{x}^2}\right)\right].$$

Sahoo's (1983) Estimator

$$t_6 = t_R\Bigg/\left[1+\theta\left(\frac{s_x^2}{\overline{x}^2} - \frac{s_{yx}}{\overline{y}\overline{x}}\right)\right].$$

Three New AUR Estimators of Sahoo (1987)

$$t_7 = t_R \left(1+\theta \frac{s_{yx}}{\bar{y}\,\bar{x}}\right)\left(1-\theta \frac{s_x^2}{\bar{x}^2}\right),$$

$$t_8 = t_R \left(1-\theta \frac{s_x^2}{\bar{x}^2}\right)\Bigg/\left(1-\theta \frac{s_{yx}}{\bar{y}\,\bar{x}}\right)$$

and $$t_9 = t_R \Bigg/ \left[\left(1-\theta \frac{s_{yx}}{\bar{y}\,\bar{x}}\right)\left(1+\theta \frac{s_x^2}{\bar{x}^2}\right)\right].$$

The six estimators $t_4, t_5, ..., t_9$ are virtually equivalent in the sense that they have the same approximate mean square error to $O(n^{-2})$. Moreover, the same statistics like $\bar{y}, \bar{x}, s_{yx}$ and s_x^2 are also used for their computation. Henceforth, we call these estimators as equivalent almost unbiased ratio (EAUR) estimators. Dalabehera and Sahoo (1995) compared the efficiencies of these EAUR estimators under a proportional regression model considering their approximate mean square errors to order n^{-3}.

In this paper, we carry out a Monte Carlo study to compare the performance of the nine AUR estimators together with t_R under a simulated sampling experiment. The following performance measures were taken into consideration:

i) *Relative Bias (RB)* = $|bias| / \bar{Y}$

ii) *Relative Efficiency (RE)*
We accept mean square error as the measure of the efficiency of an estimator.

iii) *Coverage Rate (CR)* based on $100(1-\alpha)\%$ (95% or 99%) confidence interval

$$t \pm u_{1-\frac{\alpha}{2}} \sqrt{v(t)}\ ,$$

where $u_{1-\frac{\alpha}{2}}$ is exceeded with probability $\alpha/2$ by the unit normal variate under the assumption that the sampling distribution of an estimator t is approximately normal, and $v(t)$ is the common approximate variance estimator of all ratio-type estimators defined by

$$v(t) = \theta \sum_{i \in s} \left(y_i - \frac{\bar{y}}{\bar{x}} x_i\right)^2 \Bigg/ (n-1)\ .$$

This performance measure gives an idea about which percentage of the so constructed confidence intervals covers the true value of $\bar{Y}$ under repeated draws of samples from a population.

iv) *Approach to Normality (Asymmetry)*

The coefficients of skewness and kurtosis, *i.e.*, β_1 and β_2 values of the sampling distribution of the estimators are taken as the indices for the measurements of the symmetry. For a standard symmetrical (*e.g.*, normal) distribution, these coefficients are respectively 0 and 3.

2. DESCRIPTION OF THE MONTE CARLO SIMULATION

Our Monte Carlo study involves repeated draws of simple random without replacement samples from a population by relying on the software Splus. The necessary guidance for this was generously provided to us by Alan H. Drofman, Office of Survey Methods Research, U.S. Department of Labor, to whom we wish to express our sincere thanks.

We created artificial populations of size $N = 500$ in which the auxiliary variable x was generated by a standard *lognormal* distribution (*i.e.*, $x \sim \exp(z)/2$ with z standard normal) to give a right skewed population that is frequently encountered in practice. The $y-$ values were then generated according to the following linear model:

$$y_i = a + bx_i + e_i,\ i = 1, 2, ..., N,$$

with $a = 0$, 0.5 and 1.2, $b = 1$; e_i's are independent drawings from a normal distribution such that $E(e_i/x_i) = 0$ and $V(e_i/x_i) = \sigma^2$ (=0.2).1000 independent samples each of size 10, 20 or 40 were selected from a population. For each sample, several estimators were calculated. The average behavior of the estimators summarized through different performance measures are presented in Tables 1 to 4 and the major findings of the study are discussed in subsections 2.1 to 2.4.

Table 1
Relative Biases ($\times 10^4$) of the Estimators

a	n	t_R	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
0.0	10	0.06	0.07	0.05	0.06	0.04	0.07	3.58	4.01	4.31	8.81
	20	0.03	0.03	0.05	0.05	0.04	0.03	0.85	0.87	0.87	3.34
	40	0.03	0.03	0.03	0.05	0.03	0.02	0.39	0.40	0.41	1.05
0.5	10	63.25	12.01	8.83	14.75	3.88	6.19	5.08	6.25	6.36	9.58
	20	45.64	6.84	6.66	10.35	2.95	4.35	3.49	3.51	4.03	8.65
	40	21.95	4.59	3.01	4.09	2.06	3.31	2.86	2.90	2.98	6.35
1.2	10	84.93	13.00	10.75	18.96	4.65	8.09	6.85	7.01	8.39	10.51
	20	59.05	7.90	7.48	12.36	3.08	6.10	4.72	5.32	5.38	9.49
	40	31.47	5.89	3.98	6.31	2.84	4.61	2.98	2.99	3.03	5.00

2.1 Results Based on the RB

The numerical values on the relative biases shown in Table 1 reveals that the bias of an estimator diminishes gradually with enlargement of sample size and increases with departure of the regression line from the origin. When $a = 0$, RB of t_R, t_1, t_2, t_3, t_4 and t_5 are either negligible or zero thereby showing that these estimators are unbiased when regression line passes through the origin, and on the other hand t_6 is the least biased among the remaining ones. When $a \neq 0$, the estimators in the EAUR-group (except t_9) are usually less biased than others with t_4 being the least biased followed by t_6.

Table 2
Relative Efficiencies of the Estimators

a	n	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
0.0	10	113.8	110.8	113.2	118.5	119.2	118.7	119.8	120.2	115.6
	20	117.7	106.9	112.3	120.4	122.0	121.3	122.2	122.9	116.2
	40	107.2	105.8	109.1	122.6	124.3	123.8	124.5	125.5	116.5
0.5	10	115.6	111.5	113.4	120.1	121.3	120.2	122.6	121.8	119.4
	20	113.6	107.3	112.2	121.4	122.7	122.3	123.9	123.0	120.0
	40	112.6	106.0	110.4	123.4	124.4	123.4	126.8	125.9	120.8
1.2	10	117.4	112.4	113.6	121.2	121.5	121.4	123.0	121.9	120.4
	20	114.5	107.9	113.0	123.8	124.0	123.5	125.5	124.8	121.6
	40	113.2	107.6	111.5	125.6	125.9	124.2	127.4	126.6	121.9

2.2 Results Based on the RE

The percentage relative efficiencies of the nine AUR estimators compared to t_R are presented in Table 2. There is some indication that the RE of t_1, t_2 and t_3 deteriorates as the sample size increases. But, with increase in the value of a, the RE of all AUR estimators improves gradually. In general, the EAUR estimators are consistently more efficient than t_1, t_2 and t_3, even though the performance of t_9 is slightly poor within the EAUR-group. t_7 or t_8 is regarded as the most efficient according as $a = 0$ or $a \neq 0$ and they follow each other giving third place to t_5.

Table 3
Coverage Rates of the Estimators (Nominal 95% Confidence Interval)

a	n	t_R	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
0.0	10	84.6	84.2	79.9	84.4	86.3	88.3	86.5	89.5	90.3	87.8
	20	85.5	84.6	82.1	85.6	86.4	89.3	87.2	90.2	91.3	88.5
	40	85.8	85.3	83.3	86.3	87.8	89.9	88.3	90.6	91.9	89.6
0.5	10	84.5	84.1	79.8	83.5	86.0	85.3	86.0	89.0	88.5	85.0
	20	84.9	84.2	81.3	84.5	86.9	85.8	87.3	90.0	89.2	86.0
	40	85.6	85.0	82.8	84.7	88.3	87.3	88.9	90.2	89.8	87.3
1.2	10	84.0	83.8	79.0	82.4	85.3	84.3	86.0	88.3	87.4	83.0
	20	84.8	83.9	80.8	82.8	85.8	85.0	87.1	88.6	88.3	83.8
	40	85.3	84.2	81.5	80.9	86.8	86.5	87.8	89.7	88.8	87.9

2.3 Results Based on the CR

Using several estimators, the coverage rates of nominal 95% confidence intervals for $\bar{Y}$ are shown in Table 3 (Results for the nominal level 99% are not shown, as they confirm more or less the tendencies found in the case of 95% confidence interval). Table 3 gives some indication of improvement in the performance of an estimator as the sample size increases. But, as the regression line deviates more from the origin, the quality of an estimator deteriorates. The AUR estimators t_1, t_2 and t_3 can not assure good CR compared to t_R and EAUR estimators. The achieved CR for t_7 and t_8 are much closer to the nominal level and follow each other for $a = 0$ and $a \neq 0$ respectively.

2.4 Results Based on the Skewness and Kurtosis

An examination of the numerical results for the coefficient of skewness (Table 4) reveals that the sampling distribution of all estimators tend towards normality when the regression line passes through the origin. Although the EAUR estimators perform about equally well, their distributions are usually less skewed than those of t_1, t_2 and t_3. However, the distribution of the classical ratio estimator is always less skewed than the distributions of the nine AUR estimators.

The coefficients of kurtosis of the estimators have also been compared, but the numerical results are not given here. In respect of this criterion the estimators behave very much erratically and there is no clear indication that one of them would have a decidedly better overall performance compared to the others.

Table 4

Skewness of the Estimators ($\beta_1 \times 10^2$)

a	n	t_R	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
0.0	10	0.03	0.03	0.08	0.07	0.05	0.03	0.07	0.00	0.04	0.02
	20	0.02	0.04	0.03	0.07	0.04	0.04	0.00	0.05	0.00	0.01
	40	0.00	0.01	0.00	0.03	0.03	0.00	0.03	0.00	0.00	0.00
0.5	10	7.23	45.12	56.31	48.23	38.68	38.79	38.70	38.81	38.80	38.73
	20	5.03	43.21	55.12	45.60	31.60	31.59	31.62	31.58	31.62	31.63
	40	2.33	33.89	48.60	40.03	25.51	25.53	25.52	25.50	25.56	25.58
1.2	10	8.56	56.45	66.89	57.00	42.98	42.89	42.96	42.98	42.88	42.90
	20	6.05	43.85	61.25	50.35	39.32	39.36	39.40	39.45	39.28	39.36
	40	3.33	34.65	57.00	48.32	30.16	30.18	30.15	30.12	30.18	30.17

3. CONCLUSIONS

Our Monte Carlo study shows that the overall performance of the six EAUR estimators compared to t_1, t_2 and t_3 on the grounds of RE, achieved CR and skewness are highly satisfactory. When $a \neq 0$, the RB of the EAUR estimators (except t_9) are also less than those of t_R, t_1, t_2 and t_3. The nine AUR estimators are usually inferior to t_R in respect of skewness. Results of the present study lead to the major conclusion that t_7 and

t_8 may be preferable to other estimators when $a = 0$ and $a \neq 0$, respectively, in respect of RE and CR. But, with respect to RB, t_4 may be preferable to others when $a \neq 0$.

Although the conclusions of this Monte Carlo investigation may not be applicable to all situations, they provide some guidelines on the overall performances of the estimators under consideration. Clearly therefore, further work, mathematical as well empirical, using other models and actual data is required to have better understanding of the relative performances of the estimators.

ACKNOWLEDGEMENT

Thanks to the referee for useful comments.

REFERENCES

1. Beale, E.M.L. (1962). Some uses of computers in operational research. *Industrielle Organisation*, 31, 51-52.
2. Dalabehera, M. and Sahoo, L.N. (1995). Efficiencies of six almost unbiased ratio estimators under a particular model. *Statistical Papers*, 36, 61-67.
3. Hutchinson, M.C. (1971). A Monte Carlo comparison of some ratio estimators. *Biometrika*, 58, 313-321.
4. Murthy, M.N. and Nanjamma, N.S. (1959). Almost unbiased ratio estimators based on interpenetrating sub-sample estimators. *Sankhyā*, 21, 381-392.
5. Pascual, J.N. (1961). Unbiased ratio estimators in stratified sampling. *J. Amer. Statist. Assoc.*, 56, 70-87.
6. Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.
7. Rao, J.N.K. and Beegle, L.D. (1967). A Monte Carlo study of some ratio estimators. *Sankhyā*, B29, 47-56.
8. Rao, J.N.K. (1969). Ratio and regression estimators. *New Developments in Survey Sampling*, N.L. Johnson and H. Smith (eds), New York: John Wiley and Sons, 213-234.
9. Rao, J.N.K. and Kuzik, R.A. (1974). Sampling errors in ratio estimation. *Sankhyā*, C36, 43-58.
10. Royall, R.M. and Cumberland, R.A. (1981). An empirical study of the ratio estimator and estimators of its variance. *J. Amer. Statist. Assoc*., 76, 66-88.
11. Sahoo, L.N. (1983). On a method of bias reduction in ratio estimation. *Jour. Statist. Res*., 17, 1-6.
12. Sahoo, L.N. (1987). On a class of almost unbiased estimators for population ratio. *Statistics*, 18, 119-121.
13. Tin, M. (1965). Comparison of some ratio estimators. *J. Amer. Statist. Assoc*., 60, 294-307.
14. Wu, C.F.J. and Deng, L.Y. (1983). Estimation of variance of the ratio estimator: An empirical study. *Scientific Inference, Data Analysis and Robustness*, G.E.P. Box *et al.* (eds), New York: Academic Press, 245-277.