International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 04 | July-2015 www.irjet.net p-ISSN: 2395-0072

Comparison of Some Almost Unbiased Ratio Estimators

Priyaranjan Dash

Department of Statistics, Tripura University, India.

---***---

Abstract - *A survey sampler is always desperate for designing a best estimator and devotes endless effort to achieve that. As a result a reasonable number of estimators developed especially for the most urged parameter - Average. A notable development occurred in designing such estimators by exploiting the ancillary information at hand. It recognized several non linear estimators, which are analyzed and compared in terms of biases and mean square errors using linearization techniques, in which the higher order terms are sacrificed. So, it is worth investigating their actual performance before using these. The present paper considered some unbiased ratio-type estimators (URE) and their performance were accessed by simulating under bootstrap method. It is recommended that the Beale's (1962) estimator as a reasonable one for estimating population mean under the assumption of bivariate normal law for the universe.*

Key Words: Auxiliary Information, Bootstrap Method, Bias and Mean Square Error.

AMS Subject Classification: 62D05

1. INTRODUCTION

Let the universe under study is composed of N distinct and identifiable units with (y_i, x_i) be the pairs of values associated with the i^{th} unit $(i = 1,2,\cdots,N)$ for the study variable y and the auxiliary variable x respectively. Most frequently We are desperate for estimating the population mean $\bar{Y}$ of y −values on the basis of a finite sample s, of size n, drawn from the universe. This sample s is selected following simple random sampling without replacement scheme with the supposition that the population mean $\bar{X}$ of x −values is known to us. By utilizing this information, the commonly used estimator for estimating $\bar{Y}$ is the usual ratio estimator $\bar{y}_R = \frac{\bar{y}}{\bar{x}}\bar{X}$, where $\bar{y}$ and $\bar{x}$ are the means of y − and x − values respectively for the sample. The major setback of this estimator y is that it fails to satisfy the unbiased property. However, a notable contributions dealing with bias reduction techniques can be seen Hartley and Ross (1954), Quenouille (1956), Goodman and Hartley (1958), Durbin (1959), Beale (1962), Searl (1964), Tin (1965), Rao (1965). Its bias and mean square errors are obtained by the Taylor's linearization technique in which the higher order terms are neglected. So, we do not have any scope to judge about the actual performance of these estimators in practical situations. In different attempts to reduce the bias and mean square error, different estimators came into existence and we have to choose the best among these.

2. CONTROVERSIES WITH SMALL SAMPLES

Cochran (2011) pointed out that in samples of moderate size the distribution of the ratio $\frac{\bar{y}}{\bar{x}}$ shows a tendency to positive skewness in the kinds of populations for which the method is most often used. Again, we have an exact formula for the bias of this estimator but the sampling variance of the estimate only an approximation valid in large samples i.e. when the sample size exceeds 10 and the coefficient of variation of x and y are both less than 10%.

Again, the mean square error of an estimator is a widely used measure of performance in sample surveys, but there exists two alternative formulas for the sample estimate of the variance of the ratio estimator $\bar{y}_R = \bar{X}\hat{R}$, one form for $\hat{R} = \frac{\bar{y}}{\bar{x}}$ is

$$\widehat{V_1(\hat{R})} = \frac{(1-f)}{N\bar{X}^2}\left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{yx}\right) \quad (1)$$

and the other one is

$$\widehat{V_2(\hat{R})} = \frac{(1-f)}{N\bar{x}^2}\left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{yx}\right) \quad (2)$$

We use the second one when the $\bar{X}$ is not known in advance in estimating R and we can have $V_2(\hat{R})$. We are till now confused in preferring one the above two forms in case $\bar{X}$ is known to us (put forward by Cochran (2011). Rao et al. (1971) discussed about this for an infinite population under the linear regression model
$y_i = \alpha + \beta x_i + \epsilon_i$
with $E(\epsilon_i|x_i) = 0$, $V(\epsilon_i|x_i) = \delta x_i^t$, $E(\epsilon_i\epsilon_j|x_i x_j) = 0$, with $0 \le t \le 2$, and $x_i \sim ax^{h-1}e^{-x}$. Huaizhen and Lili (2001) also discussed this problem following the idea of

estimated loss approach and derived certain results for the preferences between these two variances.
Fieller (1932), Paulson (1942), James et al. (1974) analyzed the behavior of ratio estimator for small samples under the assumption of bivariate normal distribution and pointed out some of the practical difficulties. Kovar et al. (1988) emphasized the application of bootstrap method to measure the errors in survey estimates.

Here we analyze the performance of some unbiased ratio estimators for small samples in sampling from a bivariate normal population using bootstrap method.

3. SELECTED ESTIMATORS

In the present study, we consider the following estimators for estimating $\bar{Y}$ for their performance in the small sample case.

Simple mean estimator: In no information case, we usually use the simple mean estimator $\bar{y}$.

Usual ratio estimator: $\bar{y}_R = \frac{\bar{y}}{\bar{x}}\bar{X}$
Beale's estimator: Beale (1962) suggested an approximately unbiased estimator of $o(n^{-2})$ as

$$\bar{y}_b = \bar{y}_R \frac{1 + \theta \frac{S_{yx}}{\bar{y}\,\bar{x}}}{1 + \theta \frac{s_x^2}{\bar{x}^2}}$$

; where $\theta = n^{-1} - N^{-1}$.

Searl's estimator: Searl (1964) proposed another estimator with known value of coefficient of variation of y i.e., C_y using the finite population correction factor as

$$\bar{y}_s = \frac{y}{1 + \theta C_y^2}.$$

Tin's estimator: Tin (1965) proposed an estimator considering the sample coefficient of variations as

$$\bar{y}_t = \bar{y}_R \left[1 + \theta\left(\frac{S_{yx}}{\bar{y}\,\bar{x}} - \frac{s_x^2}{\bar{x}^2}\right)\right]$$

Reddy's estimator: Reddy (1974) proposed an estimator by considering a linear transformation for minimizing the mean square error as

$$\bar{y}_{re} = \bar{y}\frac{\bar{X}}{\bar{x} + a(\bar{x} - \bar{X})},$$

where the optimum value of a is $\rho\, {C_y}/{C_x}$.

Sissodia and Dwivedi estimator: Being motivated by Searl (1964), Sisodia and Dwivedi (1981) proposed an estimator considering the known value of coefficient of variation C_x for the auxiliary variable x as

$$\bar{y}_{sd} = \bar{y}\frac{\bar{X} + C_x}{\bar{x} + C_x}.$$

4. PERFORMANCE ANALYSIS

In order to analyze the performance of these estimators, we have considered four populations with sizes $N = 1500; 2000; 2500; 3000$ and from each population 5 samples of sizes $n = 10; 15; 20; 25$ are selected by using SRSWOR scheme independently. Then bootstrap method is used with 3500 resamples selected by SRSWR scheme from each one of the 20 samples separately. The values of y and x are chosen randomly from a bivariate normal population $N(0;\ 0;\ 1;\ 1;\ 0{:}8)$. We have chosen a moderately high positive correlation $(\rho = 0{:}8)$ each time neutralize the effect of correlation on performance of estimators.

In order to choose a better estimator on the basis of the three criteria: (i) Bias (ii) Standard error and (iii) Approach towards normality.

Performance based on Bias: The Table 1 gives the bias $(\times 10^3)$ of different estimators. Among different estimators the Searl (1964) estimator is almost unbiased and also has a minimum bias in all sample sizes within each population. Again, the estimator due to Tin (1965) shows a very high bias as compared to other estimators.

Performance based on Standard Error: The Table 2 gives the standard error calculated for all the estimators using bootstrap method. The results clearly supports the minimum standard error of Searl (1964) estimator among others following Beale (1962). So, when the population coefficient of variation for study variable C_y is not known in advance, then it is better to use Beale (1962) estimator than other estimators.

Performance based on Normality Behavior: Besides the performance based on biased and standard error, there is another important consideration i.e., following asymptotically Normal distribution by an estimator for a large number of samples. Approach towards normality by different estimators was again considered under bootstrap method and the graph showing the quantiles of the standard normal distribution against that for different estimators are shown in Table 3 and 4. From the graphs, it is clear that the simple mean estimator, Searl (1964) estimator and Sisodia and Dwivedi (1981) estimator follows asymptotically Normal distribution but the others fail.

5. CONCLUSIONS

The bootstrap method applied in this paper clearly distinguishes different estimators for making a choice between them. The performance of the Searl (1964) estimator basing upon all the three qualities viz. bias, standard error and the asymptotically normal behavior should be preferred among other competitive estimators when the value of the population coefficient of variation for the auxiliary variable is known in advance. Otherwise, we could prefer Beale (1962) estimator than other estimators as it has a very low bias as well as standard error in comparison to others. Also, it is asymptotically normally distributed as evidenced from the Q-Q plot for standard normal distribution obtained by using bootstrap technique.

Table 1. Bias $(\times 10^3)$ of Different Estimators

N	n	$\bar{y}$	$\bar{y}_R$	$\bar{y}_b$	$\bar{y}_s$	$\bar{y}_t$	$\bar{y}_{re}$	$\bar{y}_{sd}$
1500	10	−1.7637	2.3171	0.0407	0.0000	#	−4.3694	−11.8083
	15	−0.4772	21.3271	0.0839	0.0000	#	42.5650	3.8251
	20	6.2845	−94.1324	−0.2634	0.0000	#	−125.3357	2.4063
	25	1.3746	19.8516	0.5837	0.0000	#	168.9028	0.3443
	30	1.2931	519.3564	0.1950	0.0000	#	662.616	−5.3052
2000	10	0.1928	24.4339	−0.0099	0.0000	#	−34.7376	0.2223
	15	10.1028	−23.1545	0.0377	0.0000	#	3963.791	4.6403
	20	1.7218	14.3021	−0.3163	0.0000	#	−35.8608	0.5403
	25	−1.6905	82.5623	−0.6099	0.0000	#	8.1854	0.5761
	30	1.6297	−30.4015	−0.3722	0.0000	#	−34.9133	0.6918
2500	10	1.6049	46.8826	−0.4724	0.0000	#	36.0201	8.7719
	15	−4.2763	116.5565	1.4742	0.0000	#	82.0201	0.0871
	20	−5.0251	14.3021	0.9834	0.0000	#	36279.12	3.3634
	25	0.3281	−82.5623	−0.0090	0.0000	#	−127.0847	1.8497
	30	−2.5991	−0.4768	−0.7005	0.0000	#	7751.13	6.6772
3000	10	6.0484	−61.1437	0.7520	0.0000	#	146.0768	7.3740
	15	−8.5963	−2.1653	1.2093	0.0000	#	−14.0511	8.9792
	20	−2.4418	1.7932	1.2874	0.0000	#	−8.1828	6.8241
	25	0.0938	−1.1543	0.4816	0.0000	#	−9.7948	1.6126
	30	−0.5734	6.1577	0.5696	0.0000	#	7.5964	6.3283

: Very high Bias in comparison to others.

Table 2. Standard Error $(\times 10^4)$ of Different Estimators

N	n	$\bar{y}$	$\bar{y}_R$	$\bar{y}_b$	$\bar{y}_s$	$\bar{y}_t$	$\bar{y}_{re}$	$\bar{y}_{sd}$
1500	10	3597	1426	19	0.0026	#	3876	3493
	15	2878	2268	26	0.0032	#	2601	2906
	20	2422	1007	32	0.0035	#	4902	2424
	25	2011	3548	31	0.0038	#	35287	2039
	30	1885	282727	31	0.0043	#	12505	1882
2000	10	3921	1997	27	0.0000	#	5582	3903
	15	2871	1827	24	0.0009	#	22549	2827
	20	2348	5231	17	0.0010	#	17063	2378
	25	1899	1087	19	0.0010	#	5197	1924
	30	1654	9213	22	0.0010	#	11001	1690
2500	10	2949	26505	18	0.0007	#	3683	2926
	15	2460	55559	30	0.0009	#	29276	2412
	20	2364	2046	25	0.0011	#	10877	2354
	25	1891	15866	22	0.0011	#	7692	1934
	30	1707	1477	17	0.0012	#	10505	1749
3000	10	4055	28263	15	0.0034	#	114591	4000
	15	3078	4559	25	0.0039	#	6108	3064
	20	2553	1620	24	0.0043	#	8809	2573
	25	2284	1216	22	0.0047	#	4158	2230
	30	1892	12559	24	0.0048	#	5190	1908

: Very high Standard Error in comparison to others.

Table 3: Histogram and Quantiles of the Standard Normal Distribution of Estimators

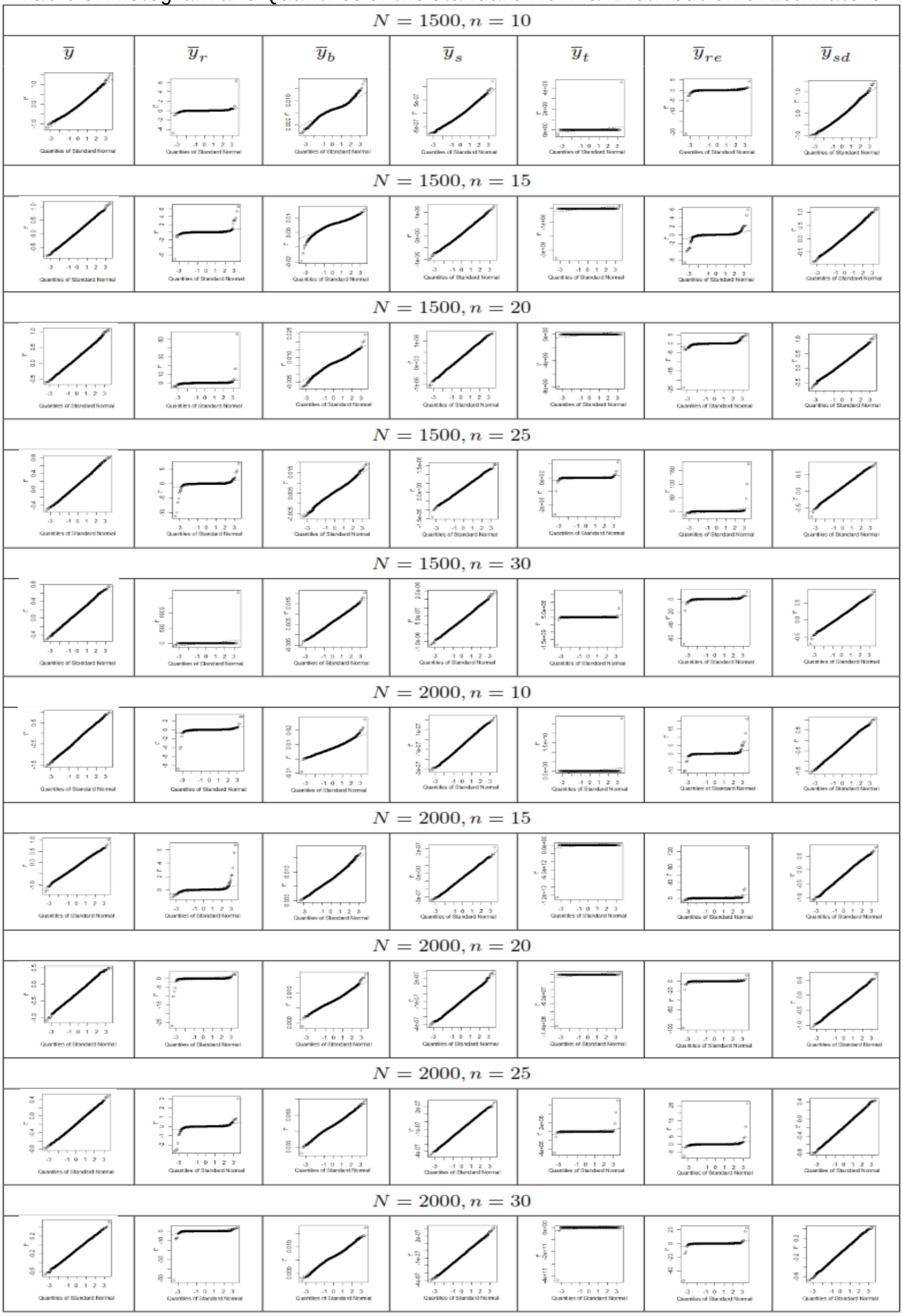

Table 4: Histogram and Quantiles of the Standard Normal Distribution of Estimators

REFERENCES

[1] Beale, E.M.L.(1962). Some uses of computers in operational research. Industrielle organization ,31 ,51-52.
[2] Casella, G. and Berger, R.L. (2007). Statistical Inference. Thomson and Duxbury, Second Edition, 244-245.
[3] Cochran, W.G. (2011). Sampling Techniques. John Wiley & Sons, New York, Third Edition, 153-164.
[4] Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to estimation of ratios. Biometrika, 46, 477-80.
[5] Fieller, E.C. (1932). The distribution of the index in a normal bivariate population. Biometrika, 24, 428-440.
[6] Goodman, L. A. and Hartley, H. O. (1958). The precision of unbiased ratio-type estimators. Jour. Amer. Statist. Assoc., 53,491-508.
[7] Hartley, H. O. and Ross, A. (1954). Unbiased ratio estimators. Nature, 174, 270-271.
[8] Huaizhen, Q., Lili, L. (2001). Comparison of variance estimators for the ratio estimator based on small samples. Acta Mathematicae Applicatae Sinica, 17(4), 449456.
[9] James, A.T., Wilkinson, G.N., and Venables, W.N. (1974). Interval estimates for a ratio of means. Sankhya, A 36, 177-183.
[10] Kovar, J.G., Rao, J.N.K., and Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. The Canadian Journal of Statistics, 16, 2545.
[11] Paulson, E. (1942). A note on the estimation of some mean values for a bivariate distribution. Ann. Math. Stat., 13, 440-444.
[12] Quenouille, M. H. (1956). Notes on bias in estimation. Biometrika, 43, 353-60.
[13] Rao, J. N. K. (1965). A note on estimation of ratios by Quenouille's method. Biometrika, 52, 647-649.
[14] Rao, P.R.P.S., and Rao, J.N.K. (1971). Small sample results for ratio estimators. Biometrika, 58, 625-630.
[15] Rao, J. N. K. and Webster, J. T. (1966). On two methods of bias reduction in the estimation of ratios. Biometrika , 53 (3/4), 571-577.
[16] Reddy, V. N. (1974). On a transformed ratio methods of estimation. Sankhya, 36c, 59-70.
[17] Sahoo, L.N. (1983). On a method of bias reduction in ratio estimation. Jour. Statist. Res., 17, 1-6.
[18] Sahoo, L.N. (1987). On a class of almost unbiased estimators for population ratio. Statistics, 18, 119-121.
[19] Searl D.T. (1964).The utilization of a known coefficient of variation in the estimation procedure. Jour. Amer. Statist. Assoc., 59, 1225-1226.
[20] Sisodia, B.V.S., and Dwivedi, V.K. (1981). A Modified Ratio Estimator using Known Coefficient of Variation of Auxiliary Variable. Journal of the Indian Society of Agricultural Statistics, 33, 13-18.
[21] Srivastava, S.K. (1967). An estimator using auxiliary information in sample surveys. Cal. Stat. Assoc. Bull., 16, 121-132.
[22] Tin, M. (1965). Comparison of some ratio estimators. Jour. Amer. Statist. Assoc., 60, 294-307.

BIOGRAPHY

The author is working in Department of Statistics, Tripura University (A Central University). He has more than 15 years of teaching and research experience. and has several publications.

