QÜESTIIÓ, Vol 18, 3 pp. 369-375, 1994

COMPARISON OF SIX ALMOST UNBIASED RATIO ESTIMATORS

M. DALABEHERA* AND L.N. SAHOO†

In this paper, we compare six almost unbiased ratio estimators with respect to bias and efficiency for (i) finite populations, and (ii) infinite populations in which the joint distribution of the characters under study is bivariate normal.

Key words: Almost unbiased estimator, bias, efficiency, mean square error, ratio estimator.

AMS classification: 62 DO5

1. INTRODUCTION

Let y and x be real variates taking values y_i and x_i $(1 \leq i \leq N)$ for the ith unit of a population of size N with means μ_y and μ_x respectively. Suppose that a sample of n units is selected from the population by simple random sampling without replacement (SRSWOR) for the purpose of estimating the ratio $R = \mu_y/\mu_x \quad (\mu_x \neq 0)$. Under the assumption that the two variates are positively correlated the usual ratio estimator $r = \overline{y}/\overline{x} \quad (\overline{y}, \overline{x}$ denoting the sample means of y and x) for R is well known in the literature. In general, the estimator is

* M. Dalabehera. Department of Statistics. Orissa University of Agriculture and Technology, Bhubaneswar. India.

† L.N. Sahoo. Department of Statistics. Utkal University, Bhubaneswar. India.

–Article rebut l'octubre de 1993.

–Acceptat l'abril de 1994.

biased and though for large n the bias is negligible as compared to the standard deviation, several "ratio-type" estimators have been put forward that are unbiased or almost unbiased (unbiased upto terms of order n^{-1}).

Beale (1962) and Tin (1965) proposed equivalent almost unbiased ratio (AUR) estimators

$$\lambda_1 = r\left(1+\theta_1 c_{xy}\right) / \left(1+\theta_1 c_x^2\right)$$

and

$$\lambda_2 = r\left[1+\theta_1\left(c_{xy}-c_x^2\right)\right]$$

respectively, where

$$\begin{aligned}
\theta_1 &= n^{-1}-N^{-1}, \\
c_{xy} &= s_{xy}/\overline{x}\overline{y}, \\
c_x^2 &= s_x^2/\overline{x}^2, \\
s_{xy} &= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right),
\end{aligned}$$

and

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2 .$$

Another AUR estimator, similar to those of Beale and Tin suggested by Sahoo (1983) is

$$\lambda_3 = r/\left[1+\theta_1\left(c_x^2-c_{xy}\right)\right].$$

Sahoo (1987) generated a class of AUr estimators considering a function of c_x^2 and c_{xy} satisfying some regularity conditions, which includes λ_1, λ_2 and λ_3 as particular cases. Three new AUR estimators as members of the class were also proposed by Sahoo (1987). These estimators are defined by

$$\begin{aligned}
\lambda_4 &= r\left(1+\theta_1 c_{xy}\right)\left(1-\theta_1 c_x^2\right) \\
\lambda_5 &= r\left(1-\theta_1 c_x^2\right)/\left(1-\theta_1 c_{xy}\right)
\end{aligned}$$

and

$$\lambda_6 = r/\left[\left(1-\theta_1 c_{xy}\right)\left(1+\theta_1 c_x^2\right)\right].$$

But, no attempts have been made to study the design based properties of these estimators.

In this paper, a comparison of these six AUR estimators is perfomed on the basis of biasedness and efficiency for

(i) finite populations considering bias and mean square error to terms of $O(n^{-2})$, and

(ii) infinite populations assuming that the distribution of the two-dimensional variable (x, y) is bivariate normal and considering mean square error upto $O(n^{-3})$.

The comparative study made by Tin (1965) leads to the conclusion that for (i) λ_1 has a smaller bias than λ_2, whereas λ_2 is more efficient than λ_1 for (ii). Sahoo (1983) also judged the efficiency of λ_3 for (ii) and shown that it is more efficient than λ_1 but more efficient than λ_2 when $C_{11} > C_{20}$, where we define $C_{pq} = K_{pq}/\mu_x^p \mu_y^q$, K_{pq} being the (p,q)th cumulant of x and y.

In the present study, the biases and mean square errors of the estimators are obtained by using the same notations and approximations used by Tin (1965) [also see Kendall, Stuart and Ord, vol. III, p. 242].

2. COMPARISON OF BIASES

To terms of order n^{-2}, the expected values of the AUR estimators are given as follows:

$$
\begin{aligned}
E(\lambda_1) &= R\left[1 - H - 2\theta_1^2 C_{20}\left(C_{20} - C_{11}\right)\right] \\
E(\lambda_2) &= R\left[1 - H - 3\theta_1^2 C_{20}\left(C_{20} - C_{11}\right)\right] \\
E(\lambda_3) &= R\left[1 - H - \theta_1^2\left(2C_{20} + C_{11}\right)\left(C_{20} - C_{11}\right)\right] \\
E(\lambda_4) &= R\left[1 - H - \theta_1^2 C_{20}\left(3C_{20} - 2C_{11}\right)\right] \\
E(\lambda_5) &= R\left[1 - H - \theta_1^2\left(3C_{20} + C_{11}\right)\left(C_{20} - C_{11}\right)\right] \\
E(\lambda_6) &= R\left[1 - H - \theta_1^2\left(2C_{20} - C_{11}\right)^2\right]
\end{aligned}
$$

where $H = (2\theta_2 - 3\theta_1/N)(C_{21} - C_{30})$ and $\theta_2 = n^{-2} - N^{-2}$.

From the above results no general conclusions can be expected. However, taking into account the signs of the biases the following tentative conclusions may be drawn:

(a) When $C_{21} > C_{30}$,

$$|B(\lambda_1)| < |B(\lambda_3)| < |B(\lambda_2)| < |B(\lambda_5)| < |B(\lambda_4)| < |B(\lambda_6)|$$

if $C_{11} < C_{20}$, showing that Beale's estimator is the least biased.

(b) When $C_{21} < C_{30}$,

$$|B(\lambda_1)| < |B(\lambda_2)| < |B(\lambda_3)| < |B(\lambda_5)| \quad \text{if } C_{11} > C_{20}$$

and

$$|B(\lambda_4)| < |B(\lambda_1)| < |B(\lambda_2)| < |B(\lambda_3)| < |B(\lambda_5)| \quad \text{if } C_{11} > 1.5C_{20}$$

(c) If the distribution of (x, y) is bivariate normal, $C_{21} = C_{30} = 0$ and we have

$$\begin{aligned} &|B(\lambda_1)| < |B(\lambda_2)| < |B(\lambda_5)|, \\ &|B(\lambda_1)| < |B(\lambda_3)| < |B(\lambda_5)|, \\ &|B(\lambda_2)| \gtrless |B(\lambda_3)| \quad \text{according as } C_{11} \gtrless C_{20}, \\ &|B(\lambda_5)| < |B(\lambda_4)| < |B(\lambda_6)| \quad \text{for } C_{11} < C_{20} \end{aligned}$$

and

$$|B(\lambda_4)| < |B(\lambda_5)| \quad \text{for } C_{11} > 1.5C_{20}.$$

(d) If the regression line of y on x is linear passing through the origin, $C_{11} = C_{20}, C_{21} = C_{30}$ and we have

$$B(\lambda_i) = 0, \quad i = 1,\ 2,\ 3 \text{ and } 5$$

and

$$B(\lambda_4) = B(\lambda_6) \neq 0.$$

3. COMPARISON OF MEAN SQUARE ERRORS

To terms of order n^{-2}, the mean square errors of the six AUR estimators are identical and given by the following expressions:

$$\begin{aligned} M(\lambda_i) &= R^2 \left[\theta_1 (C_{20} - 2C_{11} + C_{02}) + \right. \\ &+ \theta_1^2 \left(2C_{20}^2 - 4C_{20}C_{11} + C_{11}^2 + C_{20}C_{02}\right) + \\ &+ \left. (2\theta_1/N)(C_{30} - 2C_{21} + C_{12})\right], \quad i = 1, 2, \ldots, 6. \end{aligned} \tag{3.1}$$

Using the characteristics C_{pq}, Tin (1965) also expressed the mean square error of r to $O(N^{-2})$ as

$$\begin{aligned} M(r) &= R^2 \left[\theta_1 \left(C_{20} - 2C_{11} + C_{02}\right) + \right. \\ &+ \theta_1^2 \left(9C_{20}^2 - 18C_{20}C_{11} + 6C_{11}^2 + 3C_{20}C_{02}\right) - \\ &- \left. 2\left(\theta_2 - 3\theta_1/N\right)\left(C_{30} - 2C_{21} + C_{12}\right)\right] \end{aligned} \tag{3.2}$$

Thus, from (3.1) and (3.2) we see that the inequality $M(r) \gtreqless M(\lambda_i)$ holds whenever we have

$$7C_{20}^2 - 14C_{20}C_{11} + 5C_{11}^2 + 2C_{20}C_{02} \gtreqless 2\left(C_{30} - 2C_{21} + C_{12}\right)$$

or after adjusting

$$3.5\left(C_{20} - C_{11}\right)^2 + \left(1 - \lambda^2\right) C_{20}C_{02} \gtreqless \delta, \tag{3.3}$$

where λ is correlation coefficient between x and y; $\delta = C_{30} - 2C_{21} + C_{12}$.

In practice, one can not check any of the conditions given in (3.3) to be able to make an appropriate decision, since the value of δ is not known or rarely known. Tin (1965) while comparing the precisions of λ_1 and λ_2 with r, did not point out anything on the contribution of δ to the mean square errors. However, when the distribution of (x, y) is bivariate normal, $\delta = 0$, and six AUR estimators have smaller mean squares than the usual ratio estimator r.

A choice among the six AUR estimators naturally depends on the comparison of mean square errors to $O(n^{-3})$. Unfortunately, the results are too complicated to lead to a useful guide for practical purposes. But, assuming the distribution of (x, y) to be bivariate normal, the results simplify considerably and we get,

$$M(\lambda_1) = \frac{R^2}{n}\left[V_1 + \frac{C_{20}}{n}\left(1 + \frac{1}{n}\right)\left(V_1 + A^2C_{20}\right) - \frac{2C_{20}^2}{n^2}\left(V_1 + 6A^2C_{20}\right)\right]$$

$$M(\lambda_2) = \frac{R^2}{n}\left[V_1 + \frac{C_{20}}{n}\left(1 + \frac{1}{n}\right)\left(V_1 + A^2C_{20}\right) - \frac{4C_{20}^2}{n^2}\left(V_1 + 6A^2C_{20}\right)\right]$$

$$\begin{aligned} M(\lambda_3) = \frac{R^2}{n}&\left[V_1 + \frac{C_{20}}{n}\left(1 + \frac{1}{n}\right)\left(V_1 + A^2C_{20}\right) - \frac{2C_{20}^2}{n^2}\left(V_1 + 6A^2C_{20}\right) - \right. \\ &- \left. \frac{2C_{11}}{n^2}\left(V_1C_{11} + 4A^2C_{20}^2\right)\right] \end{aligned}$$

$$M(\lambda_4) = \frac{R^2}{n}\left[V_1 + \frac{C_{20}}{n}\left(1+\frac{1}{n}\right)(V_1 + A^2C_{20}) - \frac{4C_{20}^2}{n^2}(V_1 + 5A^2C_{20}+ \right.$$
$$\left. + \; 2AC_{11})\right]$$

$$M(\lambda_5) = \frac{R^2}{n}\left[V_1 + \frac{C_{20}}{n}\left(1+\frac{1}{n}\right)(V_1 + A^2C_{20}) - \frac{2C_{20}^2}{n^2}(3V_1 - 2AV_1+ \right.$$
$$\left. + \; A^2V_1 + 14A^2C_{20} - 4A^3C_{20})\right]$$

$$M(\lambda_6) = \frac{R^2}{n}\left[V_1 + \frac{C_{20}}{n}\left(1+\frac{1}{n}\right)(V_1 + A^2C_{20}) - \frac{2C_{20}^2}{n^2}(2V_1 - 2AV_1+ \right.$$
$$\left. + \; A^2V_1 - 4AC_{20} + 14A^2C_{20} - 4A^3C_{20})\right]$$

where $A = (1 - C_{11}/C_{20})$, $V_1 = A^2C_{20} + V_2$, $V_2 = C_{02}(1 - \lambda^2)$.

On comparison of the mean squares to $O(n^{-3})$ the following conclusions may be drawn:

(a) If $C_{11} > C_{20}$,

$$M(\lambda_5) < M(\lambda_2) < M(\lambda_1)$$
$$M(\lambda_6) < M(\lambda_3) < M(\lambda_2)$$

and

$$M(\lambda_6) < M(\lambda_5) < M(\lambda_4).$$

(b) If $C_{11} < C_{20}$,

$$M(\lambda_4) < M(\lambda_2) < M(\lambda_1)$$
$$M(\lambda_4) < M(\lambda_3) < M(\lambda_1)$$

and

$$M(\lambda_5) < M(\lambda_3).$$

(c) If $C_{11} < 2C_{20}$,

$$M(\lambda_6) < M(\lambda_1).$$

(d) If $C_{20} < C_{11} < 2C_{20}$, then λ_6 is uniformly better than other AUR estimators under comparison.

(e) If there is a linear regression of y on x passing through the origin,

$$M(\lambda_5) < M(\lambda_2) = M(\lambda_3) = M(\lambda_4) = M(\lambda_6) < M(\lambda_1).$$

4. CONCLUDING REMARKS

We have seen that, when the regression line of y on x is linear passing through the origin ($C_{11} = C_{20}$), which is of course an optimality condition for ratio estimate to be fruitfully employed in practice, the performance of λ_5 seems to be better than its competitors. But for other situations ($C_{11} \neq C_{20}$) no general conclusions can be obtained and performance of one estimator over other depends mainly on the population parameters. Because, the most part of the resulting expressions is only in an asymptotical form. The asymptotic theory, however, can not be expected to help when very small samples are taken without replacement especially from populations of small and moderate size. Thus, the above comparison clearly indicates the need for an extensive study mathematical as well as empirical, using various models and wide varieties of actual data.

REFERENCES

[1] **Belae, E.M.L.** (1962). "Some uses of computers in operational research". *Industrielle Organisation,* **31**, 51–52.

[2] **Kendall, M., Stuart, A.** and **Ord, J.K.** (1983). *The Adavanced Theory of Statistics.* (vol. III). Charles Griffin and Company Limited. London and High Wycombe.

[3] **Sahoo, L.N.** (1983). "On a method of bias reduction in ratio estimation". *Jour. Stat. Res.,* **17**, 1–6.

[4] **Sahoo, L.N.** (1987). "On a class of almost unbiased estimators for population ratio". *Statistics,* **18**, 119–121.

[5] **Tin, M.** (1965). "Comparison of some ratio estimators". *Jour. Amer. Stat. Assoc.,* **60**, 294–307.