CHAIN RATIO ESTIMATOR

BY

L.N. SAHOO* AND A.K.P.C. SWAIN

Utkal University, Bhubaneswar

(Received : February, 1981)

SUMMARY

In this paper chain ratio estimator in two stage sampling is compared with the usual ratio estimator in two stage sampling with respect to bias and mean square error. Also a modified chain ratio estimator is proposed. An example is provided to demonstrate that the gain in efficiency of the modified chain ratio estimator over the usual ratio estimator, may turn out substantial.

1. INTRODUCTION AND THE CHAIN RATIO ESTIMATOR

Let there be a finite population consisting of N first stage units. The ith first stage unit contains M_i second stage units. Let y_{ij} be the value of the variable under investigation for the jth second stage unit of the ith first stage unit ($j=1, 2, ..., M_i$; $i=1, 2, ..., N$)

and
$$\bar{Y}_i=\frac{1}{M_i}\sum_{j=1}^{M_i} y_{ij},$$

$$\bar{Y}=\frac{1}{N\bar{M}}\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij},$$

where
$$\bar{M}=\frac{1}{N}\sum_{i=1}^{N} M_i.$$

To estimate population mean $\bar{Y}$, a simple random sample (WOR) of n first stage units is selected from N first stage units. From the ith selected first stage unit a simple random sample (WOR) of m_i second stage units is again selected.

* Present address :—Department of Statistics, Orissa University of Agriculture & Technology, Bhubaneswar (Orissa)

Let $$\bar{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$$

and $$\bar{y} = \frac{1}{n\bar{m}}\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij},$$

where $$\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i .$$

Let x_{ij} be the value of the auxiliary variable corresponding to y_{ij} and quantities $\bar{X}_i$, $\bar{X}$, $\bar{x}_i$ and $\bar{x}$ are defined similarly.

Let $r_i = \bar{y}_i/\bar{x}_i$ and $u_i = M_i/\bar{M}$. Then, Murthy's [2] chain ratio estimator for the population mean $\bar{Y}$ is

$$\hat{\bar{Y}}_{CR} = \frac{\sum_{i=1}^{n} u_i \bar{X}_i r_i}{\sum_{i=1}^{n} u_i \bar{X}_i}\,\bar{X} .$$

2. The Results

From $$\text{Cov}\left[\frac{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i},\ \frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i\right]$$

$$- E\left[\frac{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i}\right] E\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i\right]$$

it follows that

$$E(\hat{Y}_{CR}) = E\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i\right]$$

$$-\text{Cov}\left[\frac{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i}, \frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i\right] \quad \ldots(1)$$

Now, $E\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i\right] = \bar{Y} - \frac{1}{N}\sum_{i=1}^{N} u_i(\rho_{i_r\bar{x}}\, \sigma_{i_r} \sigma_{i\bar{x}}) \quad \ldots(2)$

where, $\rho_{i_r\bar{x}}$ is the correlation coefficient between r_i and $\bar{x}_i$; σ_{i_r} and $\sigma_{i\bar{x}}$ are standard deviations of r_i and $\bar{x}_i$ respectively.

Further,

$$\text{Cov}\left[\frac{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i}, \frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i\right]$$

$$= \text{Cov}\left[E\left\{\frac{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i} \Big/ i\right\}, E\left\{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i / i\right\}\right]$$

$$+ E\left[\text{Cov}\left\{\frac{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i r_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i}, \frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i / i\right\}\right]$$

$$=\mathrm{Cov}\left[\frac{\frac{1}{n}\sum_{i=1}^{n} u_i\overline{Y}_i}{\frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i},\ \frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i\right]$$

$$-\mathrm{Cov}\left[\frac{\frac{1}{n}\sum_{i=1}^{n} u_i\sigma_{ir\bar{x}}}{\frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i},\ \frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i\right]$$

where $\sigma_{ir\bar{x}}=\mathrm{Cov}(r_i, \bar{x}_i)$

$$=\rho_{br\bar{x}}\,\sigma_{br}\sigma_{b\bar{x}}-\rho_{bt\bar{x}}\,\sigma_{bt}\sigma_{b\bar{x}} \qquad ...(3)$$

where $\rho_{br\bar{x}}$ is the correlation coefficient between

$$\frac{\frac{1}{n}\sum_{i=1}^{n} u_i\overline{Y}_i}{\frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i},\ \frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i$$

and $\rho_{bt\bar{x}}$ is the correlation coefficient between

$$\frac{\frac{1}{n}\sum_{i=1}^{n} u_i\sigma_{ir\bar{x}}}{\frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i},\ \frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i\ .$$

Also σ_{br}, $\sigma_{b\bar{x}}$ and σ_{bt} are standard deviations of the quantities

$$\frac{\frac{1}{n}\sum_{i=1}^{n} u_i\overline{Y}_i}{\frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i},\ \frac{1}{n}\sum_{i=1}^{n} u_i\overline{X}_i$$

and
$$\frac{\frac{1}{n}\sum_{i=1}^{n} u_i \sigma_{i_r \bar{x}}}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i} \text{ respectively.}$$

From (1), (2) and (3) we have the exact expression for the bias of $\hat{Y}_{CR}$ as,

$$|\text{ Bias in } \hat{Y}_{CR}| = |E(\hat{Y}_{CR}) - \bar{Y}|$$

$$= \left| \frac{1}{N}\sum_{i=1}^{N} u_i \rho_{i_r \bar{x}} \sigma_{i_r} \sigma_{i\bar{x}} + (\rho_{br\bar{x}} \sigma_{br} \sigma_{b\bar{x}} - \rho_{bt\bar{x}} \sigma_{bt} \sigma_{b\bar{x}}) \right|$$

Its behaviour *vis-a-vis* sample-size is hard to guess.

An upper bound for the magnitude of the bias is given by,

$$\left| \text{ Bias in } \hat{Y}_{CR} \right| \leqslant \frac{1}{N}\sum_{i=1}^{N} u_i \quad \sigma_{i_r} \sigma_{i\bar{x}} + \sigma_{b\bar{x}} \left| \sigma_{br} - \sigma_{bt} \right|.$$

Assuming n and m_i's are large, to the first order of approximation,

$$E(\hat{Y}_{CR}) = \bar{Y} + \frac{1}{N} \sum_{i=1}^{N} u_i \frac{1-f_i}{m_i} \bar{Y}_i \left(\frac{S_{ix}^2}{\bar{X}_i^2} - \frac{S_{ixy}}{\bar{X}_i \bar{Y}_i} \right)$$

$$+ \frac{1-f}{n} \bar{Y} \left(\frac{S_{bx}^2}{\bar{X}^2} - \frac{S_{bxy}}{\bar{X}\bar{Y}} \right) + \frac{1-f}{nN} \sum_{i=1}^{N} u_i \frac{1-f_i}{m_i} \bar{Y}_i$$

$$\left(\frac{S_{ix}^2}{\bar{X}_i^2} - \frac{S_{ixy}}{\bar{X}_i \bar{Y}_i} \right) \times \left\{ \frac{S_{bx}^2}{\bar{X}^2} - \frac{N}{(N-1)\bar{X}} (u_i \bar{X}_i - \bar{X}) \right\}$$

where $f=n/N$, $f_i=m_i/M_i$,

$$S_{ixy}=\frac{1}{M_i-1}\sum_{j=1}^{M_i}(x_{ij}-\bar{X}_i)(y_{ij}-\bar{Y}_i),$$

$$S_{ix}^2=\frac{1}{M_i-1}\sum_{j=1}^{M_i}(x_{ij}-\bar{X}_i)^2,$$

$$S_{bxy}=\frac{1}{N-1}\sum_{i=1}^{N}(u_i\bar{X}_i-\bar{X})(u_i\bar{Y}_i-\bar{Y})$$

and $$S_{bx}^2=\frac{1}{N-1}\sum_{i=1}^{N}(u_i\bar{X}_i-\bar{X})^2.$$

So, the approximate bias is zero if

$$\beta_i=R_i\ (i=1,2,\ldots,N) \text{ and } \beta_b=R$$

where $R_i=\bar{Y}_i/\bar{X}_i$, $R=\bar{Y}/\bar{X}$, $\beta_i=S_{ixy}/S_{ix}^2$

and $\beta_b=S_{bxy}/S_{bx}^2$.

With increasing sample-size $|E(\hat{\bar{Y}}_{CR})-\bar{Y}|$, of course, diminishes, as expected.

Again, to the first order of approximation, we have, for large n, m_i's,

$$MSE(\hat{\bar{Y}}_{CR})=\frac{1-f}{n}(S_{by}^2-2RS_{bxy}+R^2S_{bx}^2)$$
$$+\frac{1}{nN}\sum_{i=1}^{N}u_i^2\frac{1-f_i}{m_i}(S_{iy}^2-2R_iS_{ixy}+R_i^2S_{ix}^2),$$

where $$S_{iy}^2=\frac{1}{M_i-1}\sum_{j=1}^{M_i}(y_{ij}-\bar{Y}_i)^2$$

and $$S_{by}^2=\frac{1}{N-1}\sum_{i=1}^{N}(u_i\bar{Y}_i-\bar{Y})^2.$$

For the unbiased estimator $\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} u_i\bar{y}_i$, the variance is

$$V(\hat{Y}) = \frac{1-f}{n} S_{by}^2 + \frac{1}{nN}\sum_{i=1}^{N} u_i^2 \frac{1-f_i}{m_1} S_{iy}^2 .$$

Thus, $V(\hat{Y}) - MSE(\hat{Y}_{CR}) = 2\frac{1-f}{n}\bar{Y}^2 C_{bx}^2\left(\rho_b \frac{C_{by}}{C_{bx}} - \frac{1}{2}\right)$

$$+ \frac{2}{nN}\sum_{i=1}^{N} u_i^2 \frac{1-f_i}{m_i} \bar{Y}_i^2 \; C_{ix}^2\left(\rho_i \frac{C_{iy}}{C_{ix}} - \frac{1}{2}\right),$$

where $\rho_b = S_{bxy}/S_{bx} S_{by}$, $\rho_i = S_{ixy}/S_{ix}S_{iy}$,

$C_{bx} = S_{bx}/\bar{X}$, $C_{by} = S_{by}/\bar{Y}$,

$C_{ix} = S_{ix}/\bar{X}_i$ and $C_{iy} = S_{iy}/\bar{Y}_i$.

So, $\rho_b \frac{C_{by}}{C_{bx}} > \frac{1}{2}$ and $\rho_i \frac{C_{iy}}{C_{ix}} > \frac{1}{2}$ are together sufficient to make $\hat{Y}_{CR}$ more efficient than $\hat{Y}$.

The usual ratio estimator

$$\hat{Y}_R = \frac{\sum_{i=1}^{n} u_i \bar{y}_i}{\sum_{i=1}^{n} u_i \bar{x}_i} \bar{X},$$

is approximately unbiased if $\beta_b = \beta_i = R$ $(i=1,2,\ldots,N)$ and to the first order of approximation, for large n, m_i's,

$$MSE(\hat{Y}_R) = \frac{1-f}{n}(S_{by}^2 - 2RS_{bxy} + R^2 S_{bx}^2)$$

$$+ \frac{1}{nN}\sum_{i=1}^{N} u_i^2 \frac{1-f_i}{m_i}(S_{iy}^2 - 2RS_{ixy} + R^2 S_{ix}^2)$$

Then,

$$MSE(\hat{\bar{Y}}_R)-MSE(\hat{\bar{Y}}_{CR})$$

$$=\frac{1}{nN}\sum_{i=1}^{N} u_i^2\frac{1-f_i}{m_i}S_{ix}^2[(R-\beta_i)^2-(R_i-\beta_i)^2].$$

So, $\hat{\bar{Y}}_{CR}$ is more efficient than $\hat{\bar{Y}}_R$ if each β_i is nearer to R_i than to R.

Using usual methods of estimating S_{by}^2, S_{bxy} etc. [c.f Sukhatme and Sukhatme [3]] one may employ the following consistent estimator for $MSE\ (\hat{\bar{Y}}_{CR})$, namely,

$$\text{Est } MSE(\hat{\bar{Y}}_{CR}) = \frac{1-f}{n}(s_{by}^2-2\hat{R}s_{bxy}+\hat{R}^2s_{bx}^2)$$

$$-\frac{1-f}{n^2}\sum_{i=1}^{n}u_i^2\frac{1-fi}{mi}(s_{iy}^2-2\hat{R}s_{ixy}+\hat{R}^2s_{ix}^2)$$

$$+\frac{1}{n^2}\sum_{i=1}^{n}u_i^2\frac{1-f_i}{m_i}(s_{iy}^2-2\hat{R}_is_{ixy}+\hat{R}_i\,s_{ixy})$$

where $\hat{R}=\bar{y}'/\bar{x}',\hat{R}_i=\bar{y}_i/\bar{x}_i$ and $\bar{y}'=\frac{1}{n}\sum_{i=1}^{n}u_i\bar{y}_i,$

$$\bar{x}'=\frac{1}{n}\sum_{i=1}^{n}u_i\bar{x}_i.$$

While studying the precision of chain ratio estimator we assumed m_i's large to neglect bias. Often this may not be feasible. So, we modify $\hat{\bar{Y}}_{CR}$ by replacing r_i's therein by $\bar{r}_i'$'s which are unbiased Hartley–Ross (1954) type estimators of R_i's, given below.

Let $r_{ij}=y_{ij}/x_{ij}$, $(j=1,2,\ldots M_i;\ i=1,2,\ldots,N)$

$$\bar{R}_i=\frac{1}{M_i}\sum_{j=1}^{M_i}r_{ij},\ (i=1,2,\ldots,N)$$

$$\bar{r}_i=\frac{1}{m_i}\sum_{j=1}^{m_i}r_{ij},\ (i=1,2,\ldots,n)$$

Then, from Hartley-Ross [1] it follows that

$$\bar{r}_i' = \bar{r}_i + \frac{(M_i - 1)m_i}{\bar{X}_i M_i (m_i - 1)} (\bar{y}_i - \bar{r}_i \bar{x}_i)$$

is an unbiased estimator of $R_i \; \forall \; i$.

Replacing r_i in $\hat{\bar{Y}}_{CR}$ by $\bar{r}_i'$ our proposed modified chain ratio estimator is

$$\hat{\bar{Y}}'_{CR} = \frac{\sum_{i=1}^{n} u_i \bar{X}_i \bar{r}_i'}{\sum_{i=1}^{n} u_i \bar{X}_i} \bar{X}.$$

It follows that,

$$| \text{Bias in } \hat{\bar{Y}}'_{CR} | \leqslant \sigma_{br} \, \sigma_{b\bar{x}}$$

and to the first order of approximations for large n,

$$E(\hat{\bar{Y}}'_{CR}) \cong \bar{Y} \left[1 + \frac{1-f}{n} \left(\frac{S_{bx}^2}{\bar{X}^2} - \frac{S_{bxy}}{\bar{X}\bar{Y}} \right) \right]$$

and $$MSE(\hat{\bar{Y}}'_{CR}) \cong \frac{1-f}{n} (S_{by}^2 - 2RS_{bxy} + R^2 S_{bx}^2)$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} u_i^2 \frac{1-f_i}{m_i} (S_{iy}^2 - 2\bar{R}_i S_{ixy} + \bar{R}_i^2 S_{ix}^2).$$

So, $\hat{\bar{Y}}'_{CR}$ is approximately unbiased if $R = \beta_b$ and more efficient than usual ratio estimator if in ith first stage unit β_i is nearer to $\bar{R}_i$ than to R.

3. An Example

For the purpose of illustration, we consider data from a survey carried out in 1978 to estimate the total cultivated area in Balikuda Block of Cuttack District where villages are taken as first stage units and households within the villages as second stage units. A simple random sample (WOR) of 20 villages was selected from 273 villages n the Block; x denotes the number of bullocks per household and y

the cultivated area per household. The total number of households in the Block is known to be 22534. The values of m_i's are 13,15,14, 17, 16, 15, 14, 13, 12, 12, 15, 12, 14, 14, 15, 17, 14, 13, 12 and 13.

The estimated efficiencies of various estimators are shown in Table—I.

TABLE 1

Efficiencies of Different Estimators

Estimator	*Percentage Efficiency*
1. Unbiased estimator : $(N\overline{M}\hat{\overline{Y}})$	100.00
2. Ratio estimator : $(N\overline{M}\hat{\overline{Y}}_R)$	930.00
3. Modified chain ratio estimator : $(N\bar{M}\hat{\overline{Y}}'_{CR})$	1001.28

ACKNOWLEDGEMENT

The authors are grateful to the referees for suggesting valuable changes in the original draft.

REFERENCES

[1] Hartley, H.O. and Ross, A. (1954) : Unbiased ratio estimates. *Nature*, **174**, 270-271.

[2] Murthy, M.N. (1967) : *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta.

[3] Sukhatme, P.V. and Sukhatme, B.V. (1970) : *Sampling Theory of Surveys with Applitations*. Asia Publishing House, Bombay.