Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 8, Issue 2 (2015): 21-28

CHAIN RATIO ESTIMATOR IN TWO STAGE SAMPLING

Sitanshu Shekhar Mishra[1] and B.K. Pradhan[2]
[1]Department of Statistics, F. M. (Auto) College, Baleswar, Odisha, India.
[2]Department of Statistics, Utkal University, Bhubaneswar, Odisha, India.
E Mail[1]: sitanshumishra0@gmail.com

Received May 30, 2014
Modified June 28, 2015
Accepted August 27, 2015

Abstract

In this paper, a chain ratio estimator is suggested in a two stage sampling set up in the presence of two auxiliary variables and its efficiency is compared with the use of single auxiliary variable or without use of any auxiliary information.

Key Words: Ratio Estimator, Chain Ratio Estimator, Auxiliary Information, Two Stage Sampling.

AMS Classification: 62D05.

1. Introduction

In two stage sampling using ratio method of estimation, it is required to have advance information on the population mean $\overline{X}$ of the auxiliary variable x to estimate the population mean $\overline{Y}$ of the study variable (y). It may so happen that the population mean $\overline{X}$ of the auxiliary variable x is not known, but from the official records one can obtain first stage information on x. Besides, there may be availability of first stage information on another auxiliary variable z, which is assumed to be correlated with x.

In the following discussion, we consider a chain type estimator for $\overline{Y}$ to make use of the first stage information x and z with advance knowledge on the population mean $\overline{Z}$ of z. Let there be a finite population subdivided into N clusters $U_1, U_2, \ldots, U_N$. These are termed as first stage units (f.s.u.). The i^{th} cluster U_i contains M_i second stage units (s.s.u.), i = 1, 2, …, N.

We select a simple random sample (without replacement) ‘s’ of n f.s.u.’s and from the i^{th} f.s.u. in the sample containing M_i ssu’s, we select a simple random sample (without replacement) ‘s_i’ of m_i s.s.u’s.

Let y_{ij} and x_{ij} be the values of y and x corresponding to j^{th} s.s.u. of i^{th} f.s.u. in the population.

Define, $\overline{Y}_i = \frac{1}{M_i}\sum_{j=1}^{M_i} y_{ij}$ and $\overline{X}_i = \frac{1}{M_i}\sum_{j=1}^{M_i} x_{ij}$ (i = 1, 2, … , N)

$$\bar{Y} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^{N} M_i} \quad \text{and} \quad \bar{X} = \frac{\sum_{i=1}^{N}\sum_{j=i}^{M_i} x_{ij}}{\sum_{i=1}^{N} M_i}$$

For the sample define,

$$\bar{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij} \quad \text{and} \quad \bar{x}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} x_{ij}$$

Further, define

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} u_i \bar{y}_i \quad \text{and} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} u_i \bar{x}_i$$

Where $u_i = \frac{M_i}{\bar{M}}$ with $\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i$

$\bar{y}$ and $\bar{x}$ are unbiased estimators of $\bar{Y}$ and $\bar{X}$ respectively.

Further define,

$$S'^2_{by} = \frac{1}{N-1}\sum_{i=1}^{N}\left(u_i\bar{Y}_i - \bar{Y}\right)^2$$

$$S'^2_{bx} = \frac{1}{N-1}\sum_{i=1}^{N}\left(u_i\bar{X}_i - \bar{X}\right)^2$$

$$S'_{byx} = \frac{1}{N-1}\sum_{i=1}^{N}\left(u_i\bar{X}_i - \bar{X}\right)\left(u_i\bar{Y} - \bar{Y}\right)$$

$$S^2_{iy} = \frac{1}{M_i - 1}\sum_{j=1}^{M_i}\left(y_{ij} - \bar{Y}_i\right)^2$$

$$S^2_{ix} = \frac{1}{M_i - 1}\sum_{j=1}^{M_i}\left(x_{ij} - \bar{X}_i\right)^2$$

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

$$R_1 = \frac{\bar{Y}}{\bar{X}} \quad \text{and} \quad R_{1i} = \frac{\bar{Y}_i}{\bar{X}_i}$$

and $$\rho_{iyx} = \frac{S_{iyx}}{S_{iy}S_{ix}}, \ (i = 1, 2, ..., N)$$

Let $\bar{Z}_1, \bar{Z}_2, ..., \bar{Z}_n$ be the mean of the first stage information on second auxiliary variable Z.

Then, define

$$S'^2_{bz} = \frac{1}{N-1}\sum_{i=1}^{N}\left(u_i\bar{Z}_i - \bar{Z}\right)^2,$$

$$S'^2_{byz} = \frac{1}{N-1}\sum_{i=1}^{N}\left(u_i\bar{Y}_i - \bar{Y}\right)\left(u_i\bar{Z}_i - \bar{Z}\right),$$

$$\rho_{byz} = \frac{S'_{byz}}{S'_{by} S'_{bz}}, \quad R_2 = \frac{\bar{Y}}{\bar{Z}}, \quad R_{2i} = \frac{\bar{Y}_i}{\bar{Z}_i} \qquad (i = 1, 2, ..., N),$$

where $\bar{Z}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} z_{ij}$ and $\bar{Z} = \frac{1}{N} \sum_{i=1}^{N} u_i \bar{Z}_i$.

2. Chain Ratio type Estimator

When $\bar{X}$ is not known, a ratio estimator of $\bar{Y}$ may be proposed as

$$\hat{\bar{Y}}_1 = \frac{\sum_{i=1}^{n} u_i \bar{y}_i}{\sum_{i=1}^{n} u_i \bar{x}_i} \; \frac{1}{n} \sum_{i=1}^{n} u_i \bar{X}_i \tag{1}$$

Alternatively, if first stage information on z in the form of knowledge about $\bar{Z}_1, \bar{Z}_2, ..., \bar{Z}_n$ where $\bar{Z}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} z_{ij} (i = 1, 2, ..., n)$, z_{ij} being the value corresponding to j^{th} s.s.u. of i^{th} f.s.u, are known from official records or can be obtained cheaply along with knowledge about the population mean $\bar{Z}$ of z, the chain type ratio estimator may be proposed as

$$\hat{\bar{Y}}_C = \frac{\sum_{i=1}^{n} u_i \bar{y}_i}{\sum_{i=1}^{n} u_i \bar{x}_i} . \frac{\sum_{i=1}^{n} u_i \bar{X}_i}{\sum_{i=1}^{n} u_i \bar{Z}_i} \bar{Z} \tag{2}$$

To derive the bias and MSE of $\hat{\bar{Y}}_1$ we proceed as follow:

Let $\frac{1}{n} \sum_{i=1}^{n} u_i \bar{y}_i = \bar{Y} + \epsilon_0$

$$\frac{1}{n} \sum_{i=1}^{n} u_i \bar{x}_i = \bar{X} + \epsilon_1$$

$$\frac{1}{n} \sum_{i=1}^{n} u_i \bar{X}_i = \bar{X} + \epsilon_2$$

Where $E(\epsilon_0) = E(\epsilon_1) = E(\epsilon_2) = 0$

Then $\hat{\bar{Y}}_1 = \left(\frac{\bar{Y} + \epsilon_0}{\bar{X} + \epsilon_1} \right) \left(\bar{X} + \epsilon_2 \right)$

$$= \bar{Y} \left(1 + \frac{\epsilon_0}{\bar{Y}} \right) \left(1 + \frac{\epsilon_1}{\bar{X}} \right)^{-1} \left(1 + \frac{\epsilon_2}{\bar{X}} \right)$$

Assuming $\left| \frac{\epsilon_1}{\bar{X}} \right| < 1$, so that the expansion of $\left(1 + \frac{\epsilon_1}{\bar{X}} \right)^{-1}$ is valid and ignoring terms of order higher than two, we find its bias and MSE up to first order approximation.

$$B\left(\hat{\hat{Y}}_1\right)=\bar{Y}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_{bx}'^2}{\bar{X}^2}+\frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left\{\frac{S_{ix}^2}{\bar{X}^2}-\frac{S_{ixy}}{\bar{Y}\bar{X}}\right\}\right] \quad (3)$$

Assuming bias to be negligible for large n, the mean square error (MSE) of $\hat{\hat{Y}}_1$ to $0\left(\frac{1}{n}\right)$ is given by

$$MSE\left(\hat{\hat{Y}}_1\right)\cong\bar{Y}^2E\left[\frac{\in_0}{\bar{Y}}+\frac{\in_2}{\bar{X}}-\frac{\in_1}{\bar{X}}\right]^2$$

$$=\bar{Y}^2E\left[\frac{\in_0^2}{\bar{Y}^2}+\frac{\in_2^2}{\bar{X}^2}+\frac{\in_1^2}{\bar{X}^2}-\frac{2\in_0\in_1}{\bar{X}\bar{Y}}-\frac{2\in_1\in_2}{\bar{X}^2}+\frac{2\in_0\in_2}{\bar{Y}\bar{X}}\right]$$

$$=\left(\frac{1}{n}-\frac{1}{N}\right)S_{by}'^2+\frac{1}{nN}\sum_{i=1}^{N}u_1^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left(S_{iy}^2+R_1^2S_{ix}^2-2R_1S_{iyx}\right) \quad (4)$$

3. Bias and Mean Square Error of $\hat{\hat{Y}}_C$

$$\hat{\hat{Y}}_C=\frac{\sum_{i=1}^{n}u_i\bar{y}_i}{\sum_{i=1}^{n}u_i\bar{x}_i}.\frac{\sum_{i=1}^{n}u_i\bar{X}_i}{\sum_{i=1}^{n}u_1\bar{Z}_i}\bar{Z}$$

To find the expected value and mean square error of the chain ratio estimator $\hat{\hat{Y}}_C$,

let $\frac{1}{n}\sum_{i=1}^{n}u_i\bar{y}_i=\bar{Y}+\in_0$

$$\frac{1}{n}\sum_{i=1}^{n}u_i\bar{x}_i=\bar{X}+\in_1$$

$$\frac{1}{n}\sum_{i=1}^{n}u_i\bar{X}_i=\bar{X}+\in_2$$

$$\frac{1}{n}\sum_{i=1}^{n}u_i\bar{Z}_i=\bar{Z}+\in_3$$

where, $E(\in_0)=E(\in_1)=E(\in_2)=E(\in_3)=0$

Then $\hat{\hat{Y}}_C=\left(\frac{\bar{Y}+\in_0}{\bar{X}+\in_1}\right),\left(\frac{\bar{X}+\in_2}{\bar{Z}+\in_3}\right)\bar{Z}$

$$=\bar{Y}\left(1+\frac{\in_0}{\bar{Y}}\right)\left(1+\frac{\in_1}{\bar{X}}\right)^{-1}\left(1+\frac{\in_2}{\bar{X}}\right)\left(1+\frac{\in_3}{\bar{Z}}\right)^{-1}$$

Assuming that the expansions of $\left(1+\frac{\in_1}{\bar{X}}\right)^{-1}$ and $\left(1+\frac{\in_3}{\bar{Z}}\right)^{-1}$ are valid and ignoring terms of order higher than two, we have

$$E\left(\in_0^2\right) = V\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{y}_i\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_{by}'^2 + \frac{1}{nN}\sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{iy}^2$$

$$E\left(\in_1^2\right) = V\left[\frac{1}{n}\sum u_i \bar{x}_i\right] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{bx}'^2 + \frac{1}{nN}\sum_{i=1}^{N} u_1^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{ix}^2$$

$$E\left(\in_2^2\right) = V\left[\frac{1}{n}\sum u_i \bar{X}_i\right] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{bx}'^2$$

$$E\left(\in_3^2\right) = V\left[\frac{1}{n}\sum u_i \bar{Z}_i\right] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{bz}'^2$$

$$E\left(\in_0 \in_1\right) = COV\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{y}_i,\ \frac{1}{n}\sum_{i=1}^{n} u_1 \bar{x}_i\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_{byx}' + \frac{1}{nN}\sum_{i=1}^{N} u_1^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{iyx}$$

$$E\left(\in_0 \in_2\right) = COV\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{y}_i,\ \frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i\right] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{byx}'$$

$$E\left(\in_0 \in_3\right) = COV\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{y}_i,\ \frac{1}{n}\sum_{i=1}^{n} u_i \bar{Z}_i\right] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{byz}'$$

$$E\left(\in_1 \in_3\right) = COV\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{x}_i,\ \frac{1}{n}\sum_{i=1}^{n} u_i \bar{Z}_i\right] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{bxz}'$$

$$E\left(\in_1 \in_2\right) = COV\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{x}_i,\ \frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i\right] = V\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i\right] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{bx}'^2$$

$$E\left(\in_2 \in_3\right) = COV\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{Z}_i,\ \frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i\right] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{bxz}'$$

The bias of $\hat{\bar{Y}}_C$ to $0\left(\frac{1}{n}\right)$ is given by

$$B\left(\hat{\bar{Y}}_C\right) = \bar{Y}\left(\frac{1}{n} - \frac{1}{N}\right)\left[\frac{S_{bz}'^2}{\bar{Z}^2} - \frac{S_{byz}'}{\bar{Y}\,\bar{Z}}\right] + \bar{Y}\frac{1}{nN}\sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left[\frac{S_{ix}^2}{\bar{X}^2} - \frac{S_{iyz}}{\bar{Y}\,\bar{X}}\right] \qquad (5)$$

which will be negligible if the sample size n is sufficiently large. Again to find the mean square error to a first approximation to $0\left(\frac{1}{n}\right)$,

we have

MSE $\left(\hat{\bar{Y}}_C\right)$

$$\cong \bar{Y}^2 E\left[\frac{\epsilon_0}{\bar{Y}} - \frac{\epsilon_1}{\bar{X}} + \frac{\epsilon_2}{\bar{X}} - \frac{\epsilon_3}{\bar{Z}}\right]^2$$

$$= \bar{Y}^2 E\left[\frac{\epsilon_0^2}{\bar{Y}^2} + \frac{\epsilon_1^2}{\bar{X}^2} + \frac{\epsilon_2^2}{\bar{X}^2} + \frac{\epsilon_3^2}{\bar{Z}^2} - \frac{2\epsilon_0\epsilon_1}{\bar{Y}\bar{X}} + \frac{2\epsilon_0\epsilon_2}{\bar{Y}\bar{X}} - \frac{2\epsilon_0\epsilon_3}{\bar{Y}\bar{Z}} - \frac{2\epsilon_1\epsilon_2}{\bar{X}\bar{X}} + \frac{2\epsilon_1\epsilon_3}{\bar{X}\bar{Z}} - \frac{2\epsilon_2\epsilon_3}{\bar{Z}\bar{X}}\right]$$

Thus, the MSE of $\hat{\bar{Y}}_C$ to $0\left(\frac{1}{n}\right)$ is given by

$\text{MSE}\left(\hat{\bar{Y}}_C\right)$

$$\cong \left(\frac{1}{n} - \frac{1}{N}\right)\left[S'^2_{by} + S'^2_{bz}\frac{\bar{Y}^2}{\bar{Z}^2} - 2S'_{byz}\frac{\bar{Y}}{\bar{Z}}\right]$$

$$+ \frac{1}{nN}\sum_{i=1}^{N} u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left[S_{iy}^2 + R_1^2 S_{ix}^2 - 2R_1 S_{iyx}\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\left[S'^2_{by} + R_2^2 S'^2_{bz} - 2R_2 S'_{byz}\right]$$

$$+ \frac{1}{nN}\sum_{i=1}^{N} u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)\left[S_{iy}^2 + R_1^2 S_{ix}^2 - 2R_1 S_{iyx}\right] \tag{6}$$

4. Comparison of Efficiencies

The usual unbiased estimator of $\bar{Y}$ without use of auxiliary information is given by

$$\hat{\bar{Y}}_0 = \frac{1}{n}\sum_{i=1}^{n} u_i \bar{y}_i$$

$$\text{MSE}\left(\hat{\bar{Y}}_0\right) = \left(\frac{1}{n} - \frac{1}{N}\right)S'^2_{by} + \frac{1}{nN}\sum_{i=1}^{N} u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{iy}^2 \tag{7}$$

(i) $\hat{\bar{Y}}_1$ will be more efficient than $\hat{\bar{Y}}_0$ if $\rho_{iyx} < \frac{R_1}{2}\frac{S_{ix}}{S_{iy}}$ for all i (i = 1, 2, …, N) or equivalently $\beta_{iyx} < \frac{R_1}{2}$ for all i (i = 1, 2, …, N), where $\beta_{iyx} = \rho_{iyx}\frac{S_{iy}}{S_{ix}}$.

(ii) $\hat{\bar{Y}}_C$ will be more efficient than $\hat{\bar{Y}}_0$ if $\rho_{byz} < \frac{1}{2}\frac{C_{bz}}{C_{by}}$ and $\rho_{iyx} < \frac{R_1}{2}\frac{S_{ix}}{S_{iy}}$ or $\beta_{iyx} \le \frac{R_1}{2}$ for all i (i = 1, 2, …, N), and

(iii) $\hat{\bar{Y}}_C$ will be more efficient than $\hat{\bar{Y}}_1$ if $\rho_{byz} < \frac{1}{2}\frac{C_{bz}}{C_{by}}$.

5. Numerical Illustration

To investigate the precision of chain type estimator compared to estimators under comparison empirically, we consider N = 15 wards (fsu's) of Berhampur city of Orisssa divided into 104 blocks (ssu's). The number of the blocks (M_i) in 15 wards are 6, 6, 12, 5, 6, 6, 10, 5, 6, 6, 6, 6, 6, 12, 6. The three variables i.e. number of educated females, female population and number of households are used as y, x and z variables respectively and data are taken from Census of India (1971). The mean square errors and relative efficiencies are given in Table 1.

Estimator	MSE	Relative Efficiency
$\hat{\hat{Y}}_0$	302.105	100
$\hat{\hat{Y}}_1$	295.693	102
$\hat{\hat{Y}}_C$	279.702	108

Table 1: Comparison of Mean Square Errors (MSE)

Remarks

For the numerical illustration under consideration there has been substantial increase in efficiency in case of $\hat{\hat{Y}}_C$, using primary stage information on the second auxiliary variable z.

Also MSE $\left(\hat{\hat{Y}}_C\right) < \text{MSE}\left(\hat{\hat{Y}}_1\right) < \text{MSE}\left(\hat{\hat{Y}}_0\right)$.

6. Conclusion

The proposed chain ratio type estimator $\hat{\hat{Y}}_C$ in two-stage sampling is more efficient than $\hat{\hat{Y}}_0$ if $\rho_{byz} < \frac{1}{2}\frac{C_{bz}}{C_{by}}$ and $\beta_{iyx} < \frac{R_1}{2}$ for all i (i = 1, 2, …,N) and is more efficient than $\hat{\hat{Y}}_1$ if $\rho_{byz} < \frac{1}{2}\frac{C_{bz}}{C_{by}}$. As per the given numerical illustration, there may arise occasions when the use of information on second auxiliary variable 'z' at the primary stage may provide more efficient estimates.

References

1. Chand, L. (1975). Some ratio type estimators based on two or more auxiliary variables, Ph. D thesis submitted to IOWA State University, Ames, IOWA, USA.
2. Kiregyera, B. (1980). A chain ratio type estimator in finite population: double sampling using two auxiliary variables, Metrika, 27, p. 217-223.
3. Sukhatme, P.V. and Sukhatme, B.V. (1970). Sampling theory of surveys with applications, Second edition, Asia Publishing House, Bombay, India.

4. Swain, A.K.P.C. (1970). A note on the use of multiple auxiliary variables in sample surveys, Trabajos de Estadistica, 21, p. 135-141.
5. Murthy, M.N. (1967). Sampling Theory and Methods, Statistical Publishing Society, Calcutta, India.
6. Sahoo, L.N. and Swain, A.K.P.C. (1983). Chain ratio estimator, Jour. Ind. Soc. Agric. Stat., 35, p. 70-79.
7. Sahoo, L.N. and Swain A.K.P. C. (1986). Chain product estimators, Aligarh Jour. Stat., 6, p. 53-58.
8. Srivastava, S. Rani (1988). Some chain type estimators for population parameters using auxiliary characters in sample survey, Unpublished Ph.D. thesis, Banaras Hindu University, Varanasi, India.