ISSN 2250-3153

A Note on Ratio Estimators in two Stage Sampling

Sitanshu Shekhar Mishra

Lecturer in Statistics, Trident Academy of Creative Technology (TACT), Bhubaneswar, Odisha, India.

Abstract- In this paper a number of ratio estimators in two stage sampling are considered and their efficiencies are compared with an estimator without use auxiliary information. Numerical illustration is provided to compare the efficiencies.

I. Introduction

In large scale sample surveys it is usual practice to adopt multi stage sampling to estimate the population mean or total of the study variable y. The main purpose of using multi stage sampling in place of unrestricted unistage sampling is to reduce the cost of survey operations even if estimates derived from multi stage sampling are likely to be less efficient then those of the unrestricted uni-stage sampling. Sometimes, all auxiliary variable x compared with y may be available either at primary stage or of secondary stage or at both the stages to in case the efficiency of estimate of the finite population parametric functions such as population means or total. Thus, in the following we consider different ratio estimators of population mean of the study variable y in two –stage sampling using knowledge on a single auxiliary variable x.

II. Notations

Now, consider a finite population U partitioned into N first stage units (fsu) denoted by $U_1, U_2, \ldots\ldots U_N$. Let M_i be the number of second stage units in $U_i (i = 1, 2, \ldots., N)$. Define $M = \sum_{i=1}^{N} M_i$ and $\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i$. Let y_{ij} and x_{ij} denote values of the study variable y and the auxiliary variable x respectively for the jth ssu of U_i = (j = 1, 2, …, M_i, i= 1, 2, ….,N) .

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

The population mean of y, $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} u_i \bar{Y}_i$ and the population mean of x, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} u_i \bar{X}_i$, where $u_i = \frac{M_i}{\bar{M}}$.

Further, define $R = \frac{\bar{Y}}{\bar{X}}$ and $R_i = \frac{\bar{Y}_i}{\bar{X}_i}$ (i = 1, 2, ….., N)

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

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

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

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

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

$$S_{ixy}=\frac{1}{M_i-1}\sum_{j=1}^{M_i}\left(x_{ij}-\bar{X}_i\right)\left(y_{ij}-\bar{Y}_i\right),$$ i = 1, 2, …, N.

Define,

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

$$\bar{x}_i=\frac{1}{m_i}\sum_{j\in s_i}x_{ij}=\frac{1}{m_i}\sum_{j=1}^{m_i}x_{ij}$$

$$\bar{y}=\frac{1}{n}\sum_{i\in s}u_i\bar{y}_i=\frac{1}{n}\sum_{i=1}^{n}u_i\bar{y}_i$$

$$\bar{x}=\frac{1}{n}\sum_{i\in s}u_i\bar{x}_i=\frac{1}{n}\sum_{i=1}^{n}u_i\bar{x}_i$$

$$s_{by}'^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(u_i\bar{y}_i-\bar{y}\right)^2$$

$$s_{bx}'^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(u_i\bar{x}_i-\bar{x}\right)^2$$

$$s_{bxy}'=\frac{1}{n-1}\sum_{i=1}^{n}\left(u_i\bar{y}_i-\bar{y}\right)\left(u_i\bar{x}_i-\bar{x}\right)$$

$$s_{iy}^2=\frac{1}{m_i-1}\sum_{j=1}^{m_i}\left(y_{ij}-\bar{y}_i\right)^2$$

$$s_{ix}^2=\frac{1}{m_i-1}\sum_{j=1}^{m_i}\left(x_{ij}-\bar{x}_i\right)^2$$

$$s_{ixy}=\frac{1}{m_i-1}\sum_{j=1}^{m_i}\left(y_{ij}-\bar{y}_i\right)\left(x_{ij}-\bar{x}_i\right)$$

$$\rho_b=\frac{S_{bxy}'}{S_{by}'S_{bx}'},\quad \rho_i=\frac{S_{ixy}}{S_{iy}S_{ix}}$$ (i = 1, 2, ….,N)

$$C_{bx}=\frac{S_{bx}'}{\bar{X}},\quad C_{by}=\frac{S_{by}'}{\bar{Y}}$$

$$C_{ix}=\frac{S_{ix}}{\bar{X}_i},\quad C_{iy}=\frac{S_{iy}}{\bar{Y}_i},$$ (i = 1, 2, …., N)

III. Ratio Estimators

Consider the following ratio type estimators under two-stage sampling scheme.

$$T_0=\frac{1}{n}\sum_{i=1}^{n}u_i\bar{y}_i,$$ without using auxiliary information on x.

$$T_1=\frac{\frac{1}{n}\sum_{i=1}^{n}u_i\bar{y}_i}{\frac{1}{n}\sum_{i=1}^{n}u_i\bar{x}_i}\bar{X}.$$

$$T_2 = \frac{1}{n}\sum_{i=1}^{n} u_i \frac{\bar{y}_i}{\bar{x}_i}\bar{X}_i$$

$$T_3 = \frac{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{y}_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i}\bar{X},$$

Smith (1969)

$$T_4 = \frac{\frac{1}{n}\sum_{i=1}^{n} u_i \frac{\bar{y}_i}{\bar{x}_i}\bar{X}_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i}\bar{X}$$

, Murthy (1967)

$$T_5 = \frac{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{y}_i}{\frac{1}{n}\sum_{i=1}^{n} u_i \bar{x}_i}\frac{1}{n}\sum_{i=1}^{n} u_i \bar{X}_i$$

These estimators belong to the class of estimators considered by Panda (1998).

T_1 requires advance knowledge on $\bar{X}_j$ T_2 and T_5 require advance knowledge on $\bar{X}_i (i = 1, 2, ..., n)$; T3 and T_4 require advance knowledge on $\bar{X}$ and $\bar{X}_i \left(i = 1, 2, ..., n\right)$.

IV. Biases and Mean Square Errors of Estimators

As known, T_0 is an unbiased estimator of $\bar{Y}$. T_1, T_2, T_3, T_4, and T_5 are biased estimates of $\bar{Y}$ and upto $0\left(\frac{1}{n}\right)$, the biases are :

$$B(T_1) = \bar{Y}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\left\{\frac{S'^2_{bx}}{\bar{X}^2}-\frac{S'_{bxy}}{\bar{Y}\,\bar{X}}\right\}+\frac{1}{nN}\sum u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left\{\frac{S^2_{ix}}{\bar{X}^2}-\frac{S_{ixy}}{\bar{Y}\,\bar{X}}\right\}\right]$$

$$B(T_2) = \frac{1}{N}\sum_{i=1}^{N} u_i \bar{Y}_i\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left(\frac{S^2_{ix}}{\bar{X}_i^2}-\frac{S_{ixy}}{\bar{X}_i\,\bar{Y}_i}\right)$$

$$B(T_3) = \bar{Y}\left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{S'^2_{bx}}{\bar{X}^2}-\frac{S'_{bxy}}{\bar{Y}.\bar{X}}\right]$$

$$B(T_4) = \left(\frac{1}{n}-\frac{1}{N}\right)\bar{Y}\left\{\frac{S'^2_{bx}}{\bar{X}^2}-\frac{S'_{bxy}}{\bar{X}_i\bar{Y}}\right\}-COV\left[\frac{1}{n}\sum_{i=1}^{n} u_i\bar{Y}_i\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left(\frac{S^2_{ix}}{\bar{X}_i^2}-\frac{S_{ixy}}{\bar{X}_i\,\bar{Y}_i}\right),\frac{1}{n}\sum_{i=1}^{n} u_i\bar{X}_i\right]$$

$$B(T_5) = \bar{Y}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S'^2_{bx}}{\bar{X}^2}+\frac{1}{nN}\sum_{i=1}^{N} u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left\{\frac{S^2_{ix}}{\bar{X}^2}-\frac{S_{ixy}}{\bar{Y}.\bar{X}}\right\}\right]$$

To $0\left(\frac{1}{n}\right)$ the mean square errors (MSE) of T_1, T_2, T_3 , T_4 and T_5 are given by

$$MSE(T_0) = \left(\frac{1}{n}-\frac{1}{N}\right)S'^2_{by}+\frac{1}{nN}\sum_{i=1}^{N} u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S^2_{iy}$$

$$MSE(T_1)=\left(\frac{1}{n}-\frac{1}{N}\right)\left[S_{by}'^2+R^2S_{bx}'^2-2RS_{xy}'\right]+\frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left(S_{iy}^2+R^2S_{ix}^2-2RS_{ixy}\right)$$

$$MSE(T_2)=\left(\frac{1}{n}-\frac{1}{N}\right)S_{by}'^2+\frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left(S_{iy}^2+R_i^2S_{ix}^2-2R_iS_{ixy}\right)$$

$$MSE(T_3)=\left(\frac{1}{n}-\frac{1}{N}\right)\left[S_{by}'^2+R^2S_{bx}'^2-2RS_{xy}'\right]+\frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S_{iy}^2$$

$$MSE(T_4)=\left(\frac{1}{n}-\frac{1}{N}\right)\left[S_{by}'^2+R^2S_{bx}'^2-2RS_{yx}'\right]+\frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)(S_{iy}^2+R_i^2S_{ix}^2-2R_iS_{ixy})$$

$$MSE(T_5)=\left(\frac{1}{n}-\frac{1}{N}\right)S_{by}'^2+\frac{1}{nN}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)(S_{iy}^2+R^2S_{ix}^2-2RS_{ixy})$$

V. Comparison of Efficiencies

The sufficient conditions under which T_1, T_2, T_3, T_4, T_5 would be more efficient than T_0 (without using auxiliary variable) are given in Table 1.

Table 1

$T_1:\rho_b>\frac{1}{2}\frac{C_{bx}}{C_{by}}$	$\rho_i>\frac{1}{2}\left(\frac{R}{R_i}\right)\frac{C_{ix}}{C_{iy}}$ for all i (i= 1, 2, ...N)
$T_2:$ -----	$\rho_i>\frac{1}{2}\frac{C_{ix}}{C_{iy}}$
$T_3:\rho_b>\frac{1}{2}\frac{C_{bx}}{C_{by}}$	-----
$T_4:\rho_b>\frac{1}{2}\frac{C_{bx}}{C_{by}}$	$\rho_i>\frac{1}{2}\frac{C_{ix}}{C_{iy}}$ for all i. (i = 1, 2, ..., N)
$T_5:$ ----	$\rho_i>\frac{1}{2}\left(\frac{R}{R_i}\right)\frac{C_{ix}}{C_{iy}}$ for all i. (i = 1, 2, ...,N)

VI. Estimation of Variances

We have

$$Est(S_{bxy}')=s_{bxy}'-\frac{1}{n}\sum_{i=1}^{n}u_1^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)s_{ixy}$$

$$Est(S_{bx}'^2)=s_{bx}'^2-\frac{1}{n}\sum_{i=1}^{n}u_1^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)s_{ix}^2$$

$$Est(S_{by}'^2)=s_{by}'^2-\frac{1}{n}\sum_{i=1}^{n}u_1^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)s_{iy}^2$$

$$\text{Est}\left[\frac{1}{N}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S_{ixy}\right]=\frac{1}{n}\sum_{i=1}^{n}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)s_{ixy}$$

$$\text{Est}\left[\frac{1}{N}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S_{ix}^2\right]=\frac{1}{n}\sum_{i=1}^{n}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)s_{ix}^2$$

$$\text{Est}\left[\frac{1}{N}\sum_{i=1}^{N}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S_{iy}^2\right]=\frac{1}{n}\sum_{i=1}^{n}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)s_{iy}^2$$

Thus, Est MSE(T_1)

$$=\left(\frac{1}{n}-\frac{1}{N}\right)\left\{s_{by}'^2+\hat{R}^2s_{bx}'^2-2\hat{R}s_{bxy}'\right\}$$
$$+\frac{1}{nN}\sum_{i=1}^{n}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left\{s_{iy}^2+\hat{R}^2s_{ix}^2-2\hat{R}s_{ixy}\right\}$$

Est MSE (T_2)

$$=\left(\frac{1}{n}-\frac{1}{N}\right)s_{by}'^2+\frac{1}{nN}\sum_{i=1}^{n}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)\left\{s_{iy}^2+\hat{R}_i^2s_{ix}^2-2\hat{R}_is_{ixy}\right\}$$

Est MSE(T_3)

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

Est MSE(T_4)

$$=\left(\frac{1}{n}-\frac{1}{N}\right)\left\{s_{by}'^2+\hat{R}^2s_{bx}'^2-2\hat{R}s_{bxy}'\right\}+\frac{1}{nN}\sum_{i=1}^{n}u_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)d_i^2$$

where $d_i^2=s_{iy}^2+\hat{R}_i^2s_{ix}^2-2\hat{R}_is_{ixy}$

Est MSE (T_5)

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

VII. Numerical Illustrations

For numerical illustration of efficiencies of different estimators, we consider data from 1971 census of India, described below. The population consists of 104 blocks (ssu) divided in to N = 15 wards (fsu) of Berhampur City of Odisha. The number of block (M_i) in 15 wards are 6, 6, 12, 5, 6, 6, 10, 5, 6, 6, 6, 6, 6, 12, 6. The two variables i.e. number of educated females, female population are used as y, x, respectively. For comparison of Mean square error (MSE) of T_0, T_1, T_2, T_3, T_4, T_5 , we consider 30% sampling fraction at both the stages. Let the first stage sample size be n= 5 and the sizes of the second stage sample i.e. m_i (i = 1, 2, …, 15) are assumed to be 2, 2, 4, 2, 2, 2, 3, 2, 3, 3, 2,2, 2, 4 and 3 respectively.

Table 2: Statistical Computations

i	M_i	m_i	u_i	R_i	$(u_i\bar{Y}_i-\bar{Y})^2$	$(u_i\bar{X}_i-\bar{X})^2$	$(u_i\bar{X}_i-\bar{X})(u_i\bar{Y}_i-\bar{Y})$	S^2_{iy}	S^2_{ix}	S_{ixy}
1	6	2	0.865	0.124	4400.504	620.313	1652.178	506.4	405.067	-40
2	6	2	0.865	0.355	21.2281	503.926	103.428	1537.066	361.1	138.8
3	12	4	1.730	0.282	2144.551	51236.223	10482.304	1937.174	2095	676.318
4	5	2	0.721	0.309	1618.903	9612.665	3944.867	807.2	212.32	116.3
5	6	2	0.865	0.435	15.246	2690.694	-202.544	1538.167	169.5	-176.9
6	6	2	0.865	0.386	40.172	2330.176	305.956	2183.867	440.267	898.134
7	10	3	1.442	0.198	681.048	6231.397	-2060.068	1114.678	328.844	76.378
8	5	2	0.721	0.242	2390.114	6541.897	3954.223	1934.2	1966.0	1836.5
9	6	3	0.865	0.355	260.599	3014.097	886.269	3097.6	2033.2	-310.2
10	6	3	0.865	0.332	680.959	4746.080	1797.745	1108.3	269.767	257.1
11	6	2	0.865	0.505	1428.824	297.766	-652.269	2505.866	601.9	1006
12	6	2	0.865	0.334	58.273	219.167	113.011	1007.766	488.987	589.866
13	6	2	0.865	0.496	1342.951	219.167	-542.522	3470.4	3470.4	1017
14	12	4	1.730	0.344	5160.784	43098.881	14913.887	1966.697	1079.69	-
15	6	3	0.865	0.531	2104.662	250.088	-725.501	1062.267	869.367	405.424
										475.534

Calculated Results:

N = 15 n = 5

$\bar{X} = 2888.558$ $\bar{Y} = 99.221$ R = 0.343

$S'^2_{bx} = 9400.895$ $S'^2_{by} = 1596.344$ $S'_{bxy} = 2426.497$

Table 3: Comparison of Mean Square Errors (MSES)

Estimator	MSE	Efficiency
T_0	302.105	100
T_1	221.403	136.45
T_2	295.624	102.19
T_3	227.860	132.58
T_4	221.334	136.49
T_5	295.693	102.17

Remarks : For the given illustration, it is observed that

$$MSE(T_4) < MSE(T_1) < MSE(T_3) < MSE(T_2) < MSE(T_5) < MSE(T_0)$$

VIII. Conclusion

The ratio estimators considered in a two stage sampling set up have been compared with the usual unbiased estimator without use of auxiliary information, as regards their efficiencies. The numerical illustration shows that T_1 and T_4 have nearly equal efficiency and are more efficient than other competitive estimators.

References

[1] Murthy, M.N. (1967). Sampling Theory and Methods. Statistical Publishing Society, Calcutta, India.

[2] Panda, P. (1998). Some Strategies in two-stage sampling using auxiliary information. Unpublished Ph.D. dissertation, Utkal University, Odisha, India.

[3] Sukhatme, P.V., and Sukhatme, B. V. (1970) Sampling theory of Surveys with Applications. Ind. Soc. Agri. Stat., New Delhi

Authors

First Author – Sitanshu Shekhar Mishra, MSC, M.phil(Statistics), CIC, LLB, Assistant Professor, Dept. MCA TACT, F-2, Chandaka, Industrial Estate, Bhubaneswar-24, Odisha, India, Mail-ID-sitanshumishra0@gmail.com