Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 7, Issue 1 (2014): 93-101

RATIO –CUM- PRODUCT ESTIMATOR OF FINITE POPULATION MEAN IN DOUBLE SAMPLING FOR STRATIFICATION

Rajesh Tailor and Hilal A. Lone
School of Studies in Statistics,
Vikram University, Ujjain-456010, M.P. India

Received October 31, 2013
Modified May 30, 2014
Accepted June 05, 2014

Abstract

In this paper a class of ratio-cum-product estimator of population mean using auxiliary information is suggested in double sampling for stratification with their properties. The bias and mean squared error of the suggested estimator is obtained up to the first degree of approximation. The suggested estimator has been compared with ratio and product estimators given by Ige and Tripathi (1987) and usual unbiased estimator of population mean in double sampling for stratification. Asymptotic optimum estimator is identified. Estimator based on estimated optimum value is also obtained. An empirical study has been carried out to assess the performance of the suggested estimator.

Key Words: Finite Population Mean, Double Sampling For Stratification, Bias, Mean Squared Error.

1. Introduction

Stratification not only provides better representation of the population under study but also improves the precision of the estimators. There might be a situation when strata weights are not available or if available, strata weights are outdated and can't be used. This type of situation occurs during the household survey, when investigator does not have information about newly added household in different colonies. This situation leads investigator to use double sampling for stratification. Neyman (1938) developed the theory of double sampling. Singh and Vishwakarma (2007) have studied the properties of Bahl and Tuteja(1991) estimators in case of double sampling. Sharma (2012) has also studied same estimators of population mean in case of double sampling.Chouhan (2012) has discussed ratio-cum-product type exponential estimators of population mean in double sampling for stratification. Ige and Tripathi (1987) ,Tripathi and Bahl (1991) , contributed well in the field double sampling for stratification. Singh and Ruiz Espejo (2003) defined ratio-cum-product estimator of a finite population mean. Singh and Tailor (2005) have developed ratio-cum-product estimator using coefficient of variation. Singh and Vishwakarma (2006) defined combined ratio-product estimator of finite population mean in stratified random sampling. Motivated by Singh and Ruiz Espejo (2003) and Singh and Vishwakarma (2006) , a ratio-cum-product estimator for population mean in double sampling for stratification is suggested in this paper.

Let us consider a finite population $U=\{U_1, U_2, U_3, ... U_N\}$ of size N in which strata weight $\frac{N_h}{N}$, $\{h=1,2,3,...L\}$ are unknown. In double sampling for stratification

(a) a first phase of sample S' of size n' using simple random sampling without replacement is drawn and only auxiliary variate x is observed.

(b) the samples is stratified into L strata on the basis of observed variable x .Let n_h' denotes the number of units in h^{th} stratum $(h=1,2,3,...,L)$ such that $n' = \sum_{h=1}^{L} n_h'$.

(c) from each n_h' unit, a sample of size $n_h = v_h n_h'$ is drawn where $0 < v_h < 1$, $\{h=1,2,3,...,L\}$, is the predetermined probability of selecting a sample of size n_h from each strata of size n_h' and it constitutes a sample S of size $n = \sum_{h=1}^{L} n_h$. In S both study variate y and auxiliary variate x are observed.

Let y and x be the study variate and the auxiliary variate respectively. Let us define

$\bar{X} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} x_{hi}$: Population mean of the auxiliary variate x ,

$\bar{Y} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi}$: Population mean of the study variate y ,

$\bar{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$: h^{th} stratum mean for the auxiliary variate x ,

$\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$: h^{th} stratum mean for the study variate y ,

$S_x^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(x_{hi} - \bar{X}_h\right)^2$: Population mean square of the auxiliary variate x ,

$S_y^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(y_{hi} - \bar{Y}_h\right)^2$: Population mean square of the study variate y ,

$S_{xh}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(x_{hi} - \bar{X}_h\right)^2$: h^{th} stratum population mean square of the auxiliary variate x ,

$S_{yh}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(y_{hi} - \bar{Y}_h\right)^2$: h^{th} stratum population measure of the study variate y ,

$\rho_{yxh} = \dfrac{S_{yxh}}{S_{yh}S_{xh}}$: Correlation coefficient between y and x in the stratum h ,

$\bar{x}'_h = \dfrac{1}{n'_h}\sum_{h=1}^{n_h} \bar{x}_{hi}$:First phase sample mean of the h^{th} stratum for the auxiliary variate x ,

$f = \dfrac{n'}{N}$: First phase sampling fraction.

$n = \sum_{h=1}^{L} n_h$: Size of the sample S

$w_h = \dfrac{n_h}{n'}$: h^{th} stratum weight in the second phase sample ,

$\bar{x}_{ds} = \sum_{h=1}^{L} w_h \bar{x}_h$: unbiased estimator of population mean $\bar{X}$, at second phase or double sampling mean of the auxiliary variate x ,

$\bar{y}_{ds} = \sum_{h=1}^{L} w_h \bar{y}_h$: unbiased estimator of population mean $\bar{Y}$ at second phase or double sampling mean of the study variate x ,

$\bar{x}_h = \dfrac{1}{n_h}\sum_{h=1}^{n_h} \bar{x}_{hi}$:Mean of the second phase sample taken from h^{th} stratum for the auxiliary variate x ,

$\bar{y}_h = \dfrac{1}{n_h}\sum_{h=1}^{n_h} \bar{y}_{hi}$:Mean of the second phase sample taken from h^{th} stratum for the study variate y ,

$\bar{x}' = \dfrac{1}{n'_h}\sum_{h=1}^{n_h} w_h \bar{x}'_h$: is unbiased estimator of population mean $\bar{X}$.

Ige and Tripathi (1987) have defined classical ratio and product estimators in double sampling for stratification as

$$\hat{\bar{Y}}_{Rd} = \bar{y}_{ds}\left(\frac{\bar{x}'}{\bar{x}_{ds}}\right) \tag{1.1}$$

$$\hat{\bar{Y}}_{Pd} = \bar{y}_{ds}\left(\frac{\bar{x}_{ds}}{\bar{x}'}\right). \tag{1.2}$$

The biases and mean squared errors of estimators $\hat{\bar{Y}}_{Rd}$ and $\hat{\bar{Y}}_{Pd}$ up to the first degree of approximation are defined as

$$B\left(\hat{\bar{Y}}_{Rd}\right)=\frac{1}{\bar{X}}\left[\sum_{h=1}^{L}\frac{W_h}{n'}\left(\frac{1}{v_h}-1\right)\left\{RS_{yh}^2-S_{yxh}\right\}\right], \tag{1.3}$$

$$B\left(\hat{\bar{Y}}_{Pd}\right)=\frac{1}{\bar{X}}\left[\sum_{h=1}^{L}\frac{W_h}{n'}\left(\frac{1}{v_h}-1\right)S_{yxh}\right], \tag{1.4}$$

$$MSE\left(\hat{\bar{Y}}_{Rd}\right)=S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+R^2S_{yh}^2-2RS_{yxh}\right], \tag{1.5}$$

$$MSE\left(\hat{\bar{Y}}_{Pd}\right)=S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)\left[S_{yh}^2+R^2S_{xh}^2+2RS_{yxh}\right]. \tag{1.6}$$

Where $R=\frac{\bar{Y}}{\bar{X}}$ and $v_h=\frac{n_h}{n_h'}$.

2. Suggested Estimators

Motivated by Singh and Ruiz Rspejo (2003) and Singh and Vishwakarma (2006), we have suggest ratio-cum-product estimator for population mean $\bar{Y}$ in double sampling for stratification as

$$\hat{\bar{Y}}_{RP}^{(\alpha)}=\bar{y}_{ds}\left(\alpha\frac{\bar{x}'}{\bar{x}_{ds}}+(1-\alpha)\frac{\bar{x}_{ds}}{\bar{x}'}\right). \tag{2.1}$$

where α is suitably chosen real constant can be determined such that mean squared error of $\hat{\bar{Y}}_{RP}^{(\alpha)}$ is minimum. It is to be noted that

(i) for $\alpha=1$ in (2.1), the estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$ reduces to the ratio estimator proposed by Ige and Tripathi (1987) as

$$\hat{\bar{Y}}_{RP}^{(1)}=\bar{y}_{ds}\left(\frac{\bar{x}'}{\bar{x}_{ds}}\right). \tag{2.2}$$

(ii) for $\alpha=0$ in (2.1),the estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$ reduces to the product estimator proposed by Ige and Tripathi (1987) as

$$\hat{\bar{Y}}_{RP}^{(0)}=\bar{y}_{ds}\left(\frac{\bar{x}_{ds}}{\bar{x}'}\right). \tag{2.3}$$

To obtain the bias and mean squared error of the suggested estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$, we write

$\bar{y}_{ds}=\bar{Y}\left(1+e_o\right)$, $\bar{x}_{ds}=\bar{X}\left(1+e_1\right)$ and $\bar{x}'=\bar{X}\left(1+e_1'\right)$

such that $E\left(e_o\right)=E(e_1)=0=E\left(e_1'\right)=0$ and

$$E\left(e_0^2\right) = \frac{1}{\bar{Y}^2}\left[S_y^2\left(\frac{1-f}{n'}\right) + \frac{1}{n'}\sum_{h=1}^{L} W_h\, S_{yh}^2\left(\frac{1}{v_h} - 1\right)\right],$$

$$E\left(e_1^2\right) = \frac{1}{\bar{X}^2}\left[S_x^2\left(\frac{1-f}{n'}\right) + \frac{1}{n'}\sum_{h=1}^{L} W_h\, S_{xh}^2\left(\frac{1}{v_h} - 1\right)\right],$$

$$E\left(e_1'^2\right) = \frac{1}{\bar{X}^2} S_x^2\left(\frac{1-f}{n'}\right),$$

$$E\left(e_0 e_1\right) = \frac{1}{\bar{Y}\,\bar{X}}\left[\left(\frac{1-f}{n'}\right)S_{yx} + \frac{1}{n'}\sum_{h=1}^{L} W_h\, S_{yxh}\left(\frac{1}{v_h} - 1\right)\right],$$

$$E\left(e_0 e_1'\right) = \frac{1}{\bar{Y}\,\bar{X}}\left(\frac{1-f}{n'}\right)S_{yx}, \quad \text{and}$$

$$E\left(e_1 e_1'\right) = \frac{1}{\bar{X}^2}\left(\frac{1-f}{n'}\right)S_x^2.$$

Adopting the usual procedure for finding the bias and mean squared error, the bias and mean squared errors of the proposed estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$ up to the first degree of approximation are obtained as

$$B\left(\hat{\bar{Y}}_{RP}^{(\alpha)}\right) = \frac{1}{\bar{X}}\frac{1}{n'}\sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right)\left[S_{yxh} + \alpha\left(RS_{xh}^2 - 2S_{yxh}\right)\right], \tag{2.4}$$

$$MSE\left(\hat{\bar{Y}}_{RP}^{(\alpha)}\right) = MSE\left(\hat{\bar{Y}}_{Pd}\right) + 4\alpha\frac{1}{n'}\sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right)\left[S_{xh}^2 R^2(\alpha - 1) - RS_{yxh}\right], \tag{2.5}$$

which is minimized for

$$\alpha = \frac{1}{2}\left(\frac{\beta}{R} + 1\right) = \alpha_o \ (say) \tag{2.6}$$

Putting (2.6) in (2.5), we get the minimum mean square error of $\hat{\bar{Y}}_{RP}^{(\alpha)}$ as

$$MSE\left(\hat{\bar{Y}}_{RP}^{(\alpha)}\right) = \left(\frac{1-f}{n'}\right)S_y^2 + \frac{1}{n'}\sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right)S_{yh}^2\left(1 - \rho^2\right), \tag{2.7}$$

Let us define

$$A = \sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right)S_{xh}^2, \quad B = \sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right)S_{yxh}, \quad C = \sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right)S_{yh}^2$$

where $\beta = \frac{B}{A}$, $\rho^2 = \frac{B}{\sqrt{A}\sqrt{C}}$ have their usual meaning.

Substitution of (2.6) in (2.5) yields the asymptotic optimum estimator (AOE) of $\bar{Y}$ as

$$\hat{\bar{Y}}_{RP}^{(\alpha_o)} = \bar{y}_{ds}\left(\left(\frac{1}{2}+\frac{\beta}{2R}\right)\frac{\bar{x}'}{\bar{x}_{ds}}+\left(\frac{1}{2}-\frac{\beta}{2R}\right)\frac{\bar{x}_{ds}}{\bar{x}'}\right). \quad (2.8)$$

With same mean squared error as given in (2.7)
it is obvious that the estimator $\hat{\bar{Y}}_{RP}^{(\alpha_o)}$ in (2.8) requires the prior information of (R, β), which can be obtained easily from the previous surveys.

3. Efficiency comparisons

Variance of usual unbiased estimator $\bar{y}_{ds}$ in double sampling for stratification is given as

$$V(\bar{y}_{ds}) = S_y^2\left(\frac{1-f}{n'}\right)+\frac{1}{n'}\sum_{h=1}^{L}W_h\, S_{yh}^2\left(\frac{1}{v_h}-1\right). \quad (3.1)$$

Efficiency comparisons for the suggested estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$

From (1.5), (1.6), (3.1) and (2.5) it is observed that the suggested estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$ would be more efficient than

(i) usual unbiased estimator $\bar{y}_{ds}$ if

$$\alpha < Max\left\{\frac{1}{2},\left(\frac{1}{2}+\frac{B}{AR}\right)\right\}, \quad (3.2)$$

(ii) Ige and Tripathi (1987) ratio type estimator $\hat{\bar{Y}}_{Rd}$ if

$$\alpha < Max\left\{1,\frac{B}{AR}\right\}, \quad (3.3)$$

(iii) Ige and Tripathi (1987) product type estimator $\hat{\bar{Y}}_{Pd}$ if

$$0 < \alpha < \frac{B}{AR}+1 \cdot \quad (3.4)$$

4. Estimator Based on Estimated optimum

If the investigator failed to get the value of (R, β), the only alternative left for the investigator is to replace (R, β) by its consistent estimate $(\hat{R}, \hat{\beta})$ .hence the estimator based on estimated optimum is

$$\hat{\bar{Y}}_{RP}^{(\hat{\alpha}_o)} = \bar{y}_{ds}\left(\hat{\alpha}\,\frac{\bar{x}'}{\bar{x}_{ds}} + (1-\hat{\alpha})\frac{\bar{x}_{ds}}{\bar{x}'}\right). \quad (4.1)$$

where $\hat{\alpha} = \frac{1}{2}\left(\frac{\hat{\beta}}{\hat{R}} + 1\right)$ has their usual meaning.

Upto the first degree of approximation, mean squared error of the estimator $\hat{\bar{Y}}_{RP}^{(\hat{\alpha}_o)}$ is obtained as

$$MSE\left(\hat{\bar{Y}}_{RP}^{(\hat{\alpha}_o)}\right) = \left(\frac{1-f}{n'}\right)S_y^2 + \frac{1}{n'}\sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right)S_{yh}^2\left(1-\rho^2\right), \quad (4.2)$$

which is same mean squared error as given in (2.7).

5. Empirical study

To exhibit the performance of the suggested estimator in comparison to other estimator, two natural population data sets are being considered. The descriptions of population are given below.

Population I- [Source: National horticulture Board]

y : Productivity (MT/Hectare) and x : Production in '000 Tons

N =20	n =8	n_1 =4	n_2 =4
n_1' =7	n_2' =7	N_1 =10	N_2 =10
$\bar{Y}_1$ =1.70	$\bar{Y}_2$ =3.65	$\bar{X}_1$ =10.41	$\bar{X}_2$ =289.14
S_{x1} =3.53	S_{x2} =111.61	S_{y1} =0.50	S_{y2} =1.41
S_{yx1} =1.60	S_{yx2} =144.87	S_y^2 =2.20	

Population II- [Source: Murthy (1967), p 228]

y : Output and x : Fixed capital

N =10	n =4	n_1 =2	n_2 =2
n_1' =4	n_2' =4	N_1 =5	N_2 =5
$\bar{Y}_1$ =1925.8	$\bar{Y}_2$ =3115.6	$\bar{X}_1$ =214.4	$\bar{X}_2$ =333.8
S_{x1} =74.87	S_{x2} =66.35	S_{y1} =615.92	S_{y2} =340.38
S_{yx1} =39360.68	S_{yx2} =22356.50	S_y^2 =668351.00	

Estimators	**Range of α**	
	Population	
	I	**II**
$\bar{y}_{ds}$	$a \in (0.50,1.155)$	$a \in (0.50,1.170)$
$\hat{\bar{Y}}_{Rd}$	$a \in (0.655,1.0)$	$a \in (0.670,1.0)$
$\hat{\bar{Y}}_{Pd}$	$a \in (0,1.655)$	$a \in (0,1.670)$
Common range of 'α'for $\hat{\bar{Y}}_{RP}^{(\alpha)}$ to be more efficient than $\bar{y}_{ds}$, $\hat{\bar{Y}}_{Rd}$ and $\hat{\bar{Y}}_{Pd}$	$a \in (0.655,1.0)$	$a \in (0.670,1.0)$
Optimum value of 'α'	0.827633	0.835300

Table 1: Range of α in which $\hat{\bar{Y}}_{RP}^{(\alpha)}$ is better than $\bar{y}_{ds}$, $\hat{\bar{Y}}_{Rd}$ and $\hat{\bar{Y}}_{Pd}$

Estimators	$\bar{y}_{ds}$	$\hat{\bar{Y}}_{Rd}$	$\hat{\bar{Y}}_{Pd}$	$\hat{\bar{Y}}_{RP}^{(\hat{\alpha}_o)}$
Population I	100.00	115.46	50.08	122.73
Population II	100.00	138.96	34.20	158.62

Table 2: Percent relative Efficiency of $\bar{y}_{ds}$, $\hat{\bar{Y}}_{Rd}$, $\hat{\bar{Y}}_{Pd}$ and $\hat{\bar{Y}}_{RP}^{(\hat{\alpha}_o)}$ w.r.t. $\bar{y}_{ds}$

6. Conclusion

Table I provides the wide range of α in which suggested estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$ is more efficient than $\bar{y}_{ds}$, $\hat{\bar{Y}}_{Rd}$ and $\hat{\bar{Y}}_{Pd}$. If the scalar α even deviates from its optimum value, the suggested class of estimators $\hat{\bar{Y}}_{RP}^{(\alpha)}$ will yield more efficient estimators. Section 3 deals with the theoretical efficiency comparisons of considered estimators, provided the condition under which suggested estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$ has less mean squared error in comparisons to ratio and product estimators given by Ige and Tripathi (1987) and usual unbiased estimator of population mean in double sampling for stratification.

Table 2 exhibits thus there is a significant gain in efficiency by using optimum estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$ $(or\, \hat{\bar{Y}}_{RP}^{(\hat{\alpha}_o)})$ over $\bar{y}_{ds}$, $\hat{\bar{Y}}_{Rd}$ and $\hat{\bar{Y}}_{Pd}$. Thus the suggested estimator $\hat{\bar{Y}}_{RP}^{(\alpha)}$ (or $\hat{\bar{Y}}_{RP}^{(\hat{\alpha}_o)}$) is recommended for use in practice for estimating the population mean provided conditions given in section 3 are satisfied.

References

1. Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimator, Inform.and optimiz. science, 12(1), p. 159-163.
2. Chouhan, S. (2012). Improved estimation of parameters using auxiliary information in sample surveys, Unpublished Ph.D.Thesis, Vikram University, Ujjain, M.P. India.
3. http://nbd.gov.in/statistics/area-production-statistics.html: [Website of National horticulture Board].
4. Ige, A. F. and Tripathi, T. P. (1987). On doubling for stratification and use of auxiliary information, J. Ind. soc. Agricul. statist., 39, p. 191-201.
5. Murthy, M. N. (1967). Sampling Theory and Methods, Statistical Publishing Society, Calcutta, India, p.228
6. Neyman, J. (1938). Contribution to the theory of sampling human population, Amer. Statist. Assoc., 33, p. 101-116.
7. Sharma, B. K. (2012). Use of auxiliary information for estimation of some population parameters in sample surveys, Unpublished Ph.D.Thesis, Vikram University, Ujjain, M.P. India.
8. Singh, H. P. and Ruiz Espejo, M. (2003). On linear regression and ratio-product estimation of finite population mean, The Statistician, 52(1), p. 59-67.
9. Singh, H. P. and Tailor, R. (2005). Estimation of finite population mean using known correlation coefficient between auxiliary characters, Statist., 65(4), p. 407-418.
10. Singh, H. P. and Vishwakarma, G. K.(2006). Combined ratio-product estimator of finite population mean in stratified sampling, Metodologia de Encuestas, 7(1), p. 32-40.
11. Singh. H. P. and Vishwakarma, G. (2007). Modified exponential ratio and product estimators of finite population mean in double sampling, Aust. J. of statist., 36(3), p. 217-225.
12. Tripathi, T. P. and Bahl, S. (1991). Estimation of mean using double sampling for stratification and multivariate auxiliary information, Commu. Statist., Theo. Meth., 20, p. 2589-2602.