

Ratio and Product Type Exponential Estimators of Population Mean in Double Sampling for Stratification

Rajesh Tailor^a, Sunil Chouhan^b, Jong-Min Kim^{1,c}

^aSchool of Studies in Statistics, Vikram University, India

^bShri Vaishnav Institute of Management, India

^cDivision of Science and Mathematics, University of Minnesota at Morris, USA

Abstract

This paper discusses the problem of estimation of finite population mean in double sampling for stratification. In fact, ratio and product type exponential estimators of population mean are proposed in double sampling for stratification. The biases and mean squared errors of proposed estimators are obtained upto the first degree of approximation. The proposed estimators have been compared with usual unbiased estimator, ratio and product estimators in double sampling for stratification. To judge the performance of the proposed estimators an empirical study has been carried out.

Keywords: Finite population mean, double sampling for stratification, bias, mean squared error.

1. Introduction

Use of auxiliary information in the estimation of population parameters such as population mean, ratio of two population means, product of two population means, coefficient of variation etc. has been in practice. Ratio, product and regression type estimators are good examples in this context. Cochran (1940) initiated the use of auxiliary information at estimation stage and proposed ratio estimator for population mean. It is well established fact that ratio type estimators provide better efficiency in comparison to simple mean estimator if the study variate and auxiliary variate are positively correlated. If the correlation between the study variate and auxiliary variate is negative, product estimator given by Robson (1957) is more efficient than simple mean estimator.

Hansen *et al.* (1946) proposed a combined ratio estimator for population mean in stratified random sampling. Later Kadilar and Cingi (2003) and Singh and Vishwakarma (2006) discussed some ratio and product type estimators using known parameters of auxiliary variate for estimation of population mean in stratified random sampling. Bahl and Tuteja (1991) pioneered ratio and product type exponential estimators using an exponential function in simple random sampling. Later on these estimators were defined in stratified random sampling by Singh *et al.* (2008).

In stratified random sampling, it is assumed that strata weights as well as sampling frame are available in advance. But in many practical situations, strata weights may be available but sampling frame within strata may not be available. For example in household survey in a city, number of households in different colonies may be available but list of households may not be available. In this type of situation, post stratification technique is applied. There might be a situation when strata weights are not available or if available, strata weights are outdated and cannot be used. This type of situation occurs during the household surveys, when investigator does not have information about newly added

¹ Corresponding author: Statistics Discipline, Division of Science and Mathematics, University of Minnesota at Morris, Morris, MN 56267, USA. E-mail: jongmink@morris.umn.edu

households in different colonies. This situation leads investigator to use double sampling for stratification which was developed by Neyman (1938). Ige and Tripathi (1987) defined usual ratio and product estimators in double sampling for stratification. Singh and Vishwakarma (2007) developed a general procedure for estimating the population mean using double sampling for stratification. Recently, Saini and Bahl (2012) worked for the estimation of finite population mean.

Ige and Tripathi (1987) and Bahl and Tuteja (1991) motivated authors to study the ratio and product type exponential estimators in case of double sampling for stratification.

2. Procedure, Notations and Definitions

Let us consider a finite population $U = (U_1, U_2, \dots, U_N)$ of size N which is divided into L strata of size N_h ($h = 1, 2, \dots, L$) with strata weights N_h/N . When strata weights N_h/N are unknown, double sampling for stratification is used. The procedure of double sampling for stratification is described below:

- A first phase sample S of size n' , is drawn using simple random sampling without replacement and only auxiliary variate x is observed.
- The sample S is stratified into L strata on the basis of auxiliary variable x . Let n'_h be the number of units in h^{th} stratum ($h = 1, 2, \dots, L$) such that $\sum_{h=1}^L n'_h = n'$.
- From each n'_h units, a sample of size $n_h = v_h n'_h$ is drawn, where $0 < v_h < 1$ is the predetermined probability of selecting a sample of size n_h from strata of size n'_h and it constitutes a sample S' of size $n = \sum_{h=1}^L n_h$. In sample S' both study variate y and auxiliary variate x are observed.

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

Notation	Description
$n = \sum_{h=1}^L n_h$	Size of the sample S'
$w_h = \frac{n_h}{n'}$	h^{th} stratum weight in the second phase sample
$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$	h^{th} stratum mean for the study variate y
$\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hi}$	h^{th} stratum mean for the study variate x
$S_y^2 = \frac{1}{N-1} \sum_{h=1}^L \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$	Population mean square of the auxiliary variate y
$S_x^2 = \frac{1}{N-1} \sum_{h=1}^L \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$	Population mean square of the auxiliary variate x
$S_{yh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$	h^{th} stratum mean square of the study variate y
$S_{xh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$	h^{th} stratum mean square of the study variate x

$S_{yxh} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(x_{hi} - \bar{X}_h)$	Covariance between y and x in stratum
$\rho_{yxh} = \frac{S_{yxh}}{S_{yh}S_{xh}}$	Correlation coefficient between y and x in the stratum h
$\bar{y}_{ds} = \sum_{h=1}^L w_h \bar{y}_h$	Unbiased estimator of population mean \bar{Y} in second phase or double sampling mean of the study variate y
$\bar{x}_{ds} = \frac{1}{n} \sum_{h=1}^L w_h \bar{x}_h$	Unbiased estimator of population mean \bar{X} in second phase or double sampling mean of the study variate x
$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \bar{y}_{hi}$	Mean of the second phase sample taken from h^{th} stratum for the study variate y
$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \bar{x}_{hi}$	Mean of the second phase sample taken from h^{th} stratum for the study variate x
$\bar{x}'_h = \frac{1}{n_h} \sum_{i=1}^{n'_h} \bar{x}_{hi}$	First phase sample mean of the of h^{th} stratum for the auxiliary variate x
$f = \frac{n'}{N}$	First phase sampling fraction

Ige and Tripathi (1987) defined classical ratio estimator in double sampling for stratification as

$$d_R = \bar{y}_{ds} \frac{\bar{x}'}{\bar{x}_{ds}}. \quad (2.1)$$

Similarly product type estimator in double sampling for stratification can be defined as

$$d_P = \bar{y}_{ds} \frac{\bar{z}_{ds}}{\bar{z}'}, \quad (2.2)$$

where z is an auxiliary variate which is negatively correlated with the study variate y and notations \bar{z}_{ds} and \bar{z}' have their usual meanings.

The biases and mean squared errors of estimators d_R and d_P , upto the first degree of approximation, are

$$B(d_R) = \frac{1}{\bar{X}} \left[\sum_{h=1}^L \frac{W_h}{n'} \left(\frac{1}{v_h} - 1 \right) \{ R_1 S_{yh}^2 - S_{yxh} \} \right], \quad (2.3)$$

$$B(d_P) = \frac{1}{\bar{Z}} \left[\sum_{h=1}^L \frac{W_h}{n'} \left(\frac{1}{v_h} - 1 \right) S_{yxh} \right], \quad (2.4)$$

$$\text{MSE}(d_R) = S_y^2 \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) [S_{yh}^2 + R_1^2 S_{yh}^2 - 2R_1 S_{yxh}], \quad (2.5)$$

$$\text{MSE}(d_P) = S_y^2 \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) [S_{yh}^2 + R_2^2 S_{zh}^2 + 2R_2 S_{yzh}], \quad (2.6)$$

where $W_h = N_h/N$, $R_1 = \bar{Y}/\bar{X}$ and $R_2 = \bar{Y}/\bar{Z}$.

3. Proposed Ratio and Product Type Exponential Estimators

Bahl and Tuteja (1991) proposed ratio and product type exponential estimators of population mean \bar{Y} , in simple random sampling, respectively as

$$t_{Re} = \bar{y} \exp \left[\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right], \quad (3.1)$$

$$t_{Pe} = \bar{y} \exp \left[\frac{\bar{z} - \bar{Z}}{\bar{z} + \bar{Z}} \right], \quad (3.2)$$

where \bar{X} and \bar{Z} are known population means of auxiliary variates x and z respectively.

Motivated by Bahl and Tuteja (1991), the proposed ratio and product type estimators in double sampling for stratification are

$$d_{Re} = \bar{y}_{ds} \exp \left[\frac{\bar{x}' - \bar{x}_{ds}}{\bar{x}' + \bar{x}_{ds}} \right], \quad (3.3)$$

$$d_{Pe} = \bar{y}_{ds} \exp \left[\frac{\bar{z}_{ds} - \bar{z}'}{\bar{z}_{ds} + \bar{z}'} \right], \quad (3.4)$$

where \bar{y}_{ds} , \bar{x}_{ds} and \bar{z}_{ds} are unbiased estimators of population means \bar{Y} , \bar{X} and \bar{Z} respectively. To obtain the bias and mean squared error of the proposed ratio type estimator d_{Re} , we write $\bar{y}_{ds} = \bar{Y}(1 + e_0)$, $\bar{x}_{ds} = \bar{X}(1 + e_1)$ and $\bar{x}' = \bar{X}(1 + e'_1)$ such that

$$E(e_0) = E(e_1) = E(e'_1) = 0,$$

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

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

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

$$E(e_0 e_1) = \frac{1}{\bar{Y} \bar{X}} \text{cov}(\bar{y}_{ds}, \bar{x}_{ds}) = \frac{1}{\bar{Y} \bar{X}} \left[S_y^2 \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) S_{yxh} \right],$$

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

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

Now expressing the proposed estimator d_{Re} in terms of e_i 's, we have

$$d_{Re} = \bar{Y}(1 + e_0) \exp \left[\frac{\bar{X}(1 + e'_1) - \bar{X}(1 + e_1)}{\bar{X}(1 + e'_1) + \bar{X}(1 + e_1)} \right].$$

Therefore upto the terms of order n^{-1}

$$d_{Re} - \bar{Y} = \bar{Y} \left[e_0 + \frac{e'_1 - e_1}{2} + \frac{3e_1^2 - e_1'^2 - 2e'_1 e_1 + 4e'_1 e_0 - 4e_1 e_0}{8} \right]. \quad (3.5)$$

Taking expectations of both the sides of (3.5), we get

$$B(d_{Re}) = \frac{1}{8\bar{X}} \left[3S_{yx} \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) (3S_{xh}^2 - S_{yxh}) \right] \quad (3.6)$$

and finally upto the first degree of approximation, the mean squared error of the proposed estimator d_{Re} is obtained as

$$MSE(d_{Re}) = \left[S_y^2 \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) \left\{ S_{yh}^2 + \frac{R_1^2}{4} S_{xh}^2 \left(1 - \frac{\beta_{yxh}}{R_1} \right) \right\} \right], \quad (3.7)$$

where $\beta_{yxh} = \rho_{yxh}(S_{yh}/S_{xh})$.

Theorem 1. *Upto first degree of approximation*

$$B(d_{Re}) = \frac{1}{8\bar{X}} \left[3S_{yx} \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) (3S_{xh}^2 - S_{yxh}) \right].$$

Theorem 2. *Upto first degree of approximation*

$$MSE(d_{Re}) = S_y^2 \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) \left\{ S_{yh}^2 + \frac{R_1^2}{4} S_{xh}^2 \left(1 - \frac{\beta_{yxh}}{R_1} \right) \right\}.$$

In similar way we can prove the following theorems regarding the bias and mean squared error of the proposed product type estimator d_{Pe} :

Theorem 3. *Upto first degree of approximation*

$$B(d_{Pe}) = \frac{1}{8\bar{Z}} \left[\frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) (4S_{yzh} - R_2 S_{zh}^2) \right], \quad (3.8)$$

and

Theorem 4. *Upto the first degree of approximation*

$$MSE(d_{Pe}) = S_y^2 \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) \left\{ S_{yh}^2 + \frac{R_2^2}{4} S_{zh}^2 \left(1 + \frac{\beta_{yzh}}{R_2} \right) \right\}. \quad (3.9)$$

where $S_z^2 = 1/(N-1) \sum_{h=1}^L \sum_{i=1}^{N_h} (z_{hi} - \bar{Z}_h)^2$, $S_{yzh} = 1/(N_h-1) \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(z_{hi} - \bar{Z}_h)$ and $\beta_{yzh} = \rho_{yzh}(S_{yh}/S_{zh})$.

4. Efficiency Comparisons

The variance of usual unbiased estimator \bar{y}_{ds} in double sampling for stratification is given by

$$V(\bar{y}_{ds}) = S_y^2 \left(\frac{1-f}{n'} \right) + \frac{1}{n'} \sum_{h=1}^L W_h \left(\frac{1}{v_h} - 1 \right) S_{yh}^2. \quad (4.1)$$

Efficiency comparisons for Ratio-Type Estimator d_{Re}

From (3.7) and (4.1), it is observed that the proposed estimator d_{Re} would be more efficient than usual unbiased estimator \bar{y}_{ds} if

$$0 < R_1 < \frac{4A'}{B'}, \quad (4.2)$$

where $A' = \sum_{h=1}^L W_h(1/v_h - 1)S_{yxh}$ and $B' = \sum_{h=1}^L W_h(1/v_h - 1)S_{xh}^2$.

Comparison of (2.5) and (3.7) shows that the proposed estimator d_{Re} would be more efficient than ratio-type estimator d_R given by Ige and Tripathi (1987) if

$$R_1 > \max \left\{ 0, \frac{4A'}{3B'} \right\}. \quad (4.3)$$

Efficiency comparisons for Product-Type Exponential Estimator d_{Pe}

From (2.6), (3.9) and (4.1), it is observed that the proposed product type exponential estimator d_{Pe} would be more efficient than

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

$$0 < R_2 < -\frac{4A'}{B'}. \quad (4.4)$$

(ii) Ige and Tripathi (1987) product type estimator d_P if

$$R_2 > \max \left\{ 0, -\frac{4A'}{3B'} \right\}. \quad (4.5)$$

Expressions (4.2) to (4.5) provide the conditions under which the proposed estimators d_{Re} and d_{Pe} are more efficient than usual unbiased estimators, ratio and product estimators in double sampling for stratification.

5. Empirical Study

To exhibit the performance of the proposed estimators in comparison to other estimators, three population data sets are being considered. The description of populations is given below:

Population I [Source: Murthy (1967)]

y : Output, x : Fixed capital and z : Number of workers

Estimators	Stratum I	Stratum II
n_h	2	2
n'_h	4	4
N_h	5	5
\bar{Y}_h	1925.80	315.60
\bar{X}_h	214.40	333.80
\bar{Z}_h	51.80	60.60
S_{yh}	615.92	340.38
S_{xh}	74.87	66.35
S_{zh}	0.75	4.84
S_{yxh}	39360.68	22356.50
S_{yzh}	411.16	1536.24
S_{xzh}	30.08	287.92
S_y^2	668351.00	

Population II [Source: Official website of National Horticulture Board]

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

Estimators	Stratum I	Stratum II
n_h	2	2
n'_h	4	4
N_h	10	10
\bar{Y}_h	1.70	3.67
\bar{X}_h	10.41	289.14
\bar{Z}_h	6.32	80.67
S_{yh}	0.50	1.41
S_{xh}	3.53	111.61
S_{zh}	1.19	10.82
S_{yxh}	1.61	144.88
S_{yzh}	-0.06	-7.05
S_{xzh}	1.38	-92.02
S_y^2	2.21	

Population III [Source: Website of Japan Meteorological Society]

y : Snowy days, x : Rainy days and z : Total annual sunshine hours

Estimators	Stratum I	Stratum II
n_h	4	4
n'_h	14	14
N_h	10	10
\bar{Y}_h	142.80	102.60
\bar{X}_h	149.70	91.00
\bar{Z}_h	1630	2036
S_{yh}	6.09	12.60
S_{xh}	13.46	6.57
S_{zh}	102.17	103.26
S_{yxh}	18.44	23.30
S_{yzh}	-239.30	-655.30
S_{xzh}	-1073.00	-240.30
S_y^2	528.43	

Table 1: Percent Relative Efficiencies of \bar{y}_{ds} , d_R , d_P , d_{Re} and d_{Pe} with respect to \bar{y}_{ds}

Estimators	\bar{y}_{ds}	d_R	d_P	d_{Re}	d_{Pe}
Population I	100.00	128.63	85.33	137.82	*
Population II	100.00	144.92	111.85	167.84	*
Population III	100.00	80.66	104.24	*	106.22

Notes: * not applicable.

For comparison of different estimators we calculate percent relative efficiencies (PRE) of \bar{y}_{ds} , d_R , d_P , d_{Re} and d_{Pe} with respect to \bar{y}_{ds} as

$$\text{PRE}(d_R, \bar{y}_{ds}) = \frac{V(\bar{y}_{ds})}{\text{MSE}(d_R)} \times 100,$$

$$\text{PRE}(d_P, \bar{y}_{ds}) = \frac{V(\bar{y}_{ds})}{\text{MSE}(d_P)} \times 100,$$

$$\text{PRE}(d_{Re}, \bar{y}_{ds}) = \frac{V(\bar{y}_{ds})}{\text{MSE}(d_{Re})} \times 100,$$

$$\text{PRE}(d_{Pe}, \bar{y}_{ds}) = \frac{V(\bar{y}_{ds})}{\text{MSE}(d_{Pe})} \times 100.$$

Table 1 gives percent relative efficiencies of estimators \bar{y}_{ds} , d_R , d_P , d_{Re} and d_{Pe} with respect to usual unbiased estimator \bar{y}_{ds} . It is observed that for population I, PRE of usual unbiased estimator \bar{y}_{ds} in double sampling for stratification is 100, ratio estimator d_R proposed by Ige and Tripathi (1987) is 128.63 and the proposed ratio type estimator d_P is 137.82, which is highest. For population I the PRE of the proposed product type estimator is not calculated as the correlation between the study variate y and the auxiliary variate z is negative. Also in case of population II, the PRE of the proposed estimator d_{Re} is highest in comparison to other estimators. For population III, the PRE of proposed product type estimator d_{Pe} is highest. In this case the PRE of ratio estimator is not calculated because the correlation between study variate y and auxiliary variate x is not positive. Thus if the correlation between study variate and auxiliary variate is positive the proposed ratio type estimator d_{Re} performs well and if the correlation is negative proposed product type estimator d_{Pe} provides higher efficiency.

6. Conclusion

Section 4 provides the conditions under which the proposed estimators are more efficient than usual unbiased estimator, ratio estimator and product estimator in case of double sampling for stratification. Empirical study reveals that the proposed ratio type estimator d_{Re} has maximum percent relative efficiency in comparison to other considered estimators for populations I and II. The proposed product type estimator d_{Pe} also has highest percent relative efficiency in comparison to usual unbiased estimator \bar{y}_{ds} and product type estimator d_P in population III where the study variate y and the auxiliary variate z are negatively correlated. Therefore, the proposed estimators are recommended to use in practice for estimating the population mean provided conditions given in Section 4 are satisfied.

Acknowledgement

The authors are thankful to the editor in chief and two learned reviewers for their valuable suggestions that helped a lot to improve this paper.

References

- Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimator, *Journal of Information and Optimization Sciences*, **12**, 159–163.
- Cochran, W. G. (1940). The estimation of yield of cereal experiments by sampling for the ratio of gain to total produce, *Journal of Agricultural Science*, **30**, 262–275.
- Hansen, M. H., Hurwitz, W. N. and Gurney, M. (1946). Problems and methods of the sample survey of business, *Journal of the American Statistical Association*, **41**, 173–189.
- Ige, A. F. and Tripathi, T. P. (1987). On double sampling for stratification and use of auxiliary information, *Journal of the Indian Society of Agricultural Statistics*, **39**, 191–201.
- Kadilar, C. and Cingi, H. (2003). Ratio estimators in stratified random sampling, *Biometrical Journal*, **45**, 218–225.
- Murthy, M. N. (1967). *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, India, 228.
- Neyman, J. (1938). Contribution to the theory of sampling human population, *Journal of the American Statistical Association*, **33**, 101–116.
- Robson, D. S. (1957). Applications of multivariate polykays to the theory of unbiased ratio type estimation, *Journal of the American Statistical Association*, **52**, 511–522.
- Saini, M. and Bahl, S. (2012). Estimation of population mean in two stage design using double sampling for stratification and multi-auxiliary information, *International Journal of Computer Applications*, **47**, 17–21.
- Singh, H. P. and Vishwakrama, G. K. (2006). An efficient variant of the product and ratio estimators in stratified random sampling, *Statistics in Transition*, **7**, 1311–1325.
- Singh, H. P. and Vishwakarma, G. K. (2007). A general procedure for estimating the mean using double sampling for stratification, *Model Assisted Statistics and Applications*, **2**, 225–237.
- Singh, H. P., Tailor, R., Singh, S. and Kim, J. M. (2008). A modified estimator of population mean using power transformation, *Statistical Papers*, **49**, 37–58.
- (Official website of National Horticulture Board, India) <http://nhb.gov.in/statistics/area-production-statistics.html>
- (Website of Japan Meteorological Society.) <http://www.data.jma.go.jp/obd/stats/data/en/index.html>

Received January 9, 2013; Revised July 16, 2013; Accepted September 30, 2013