J. Stat. Appl. Pro. **3**, No. 3, 363-368 (2014)

Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/030307

Improved Product Cum Dual to Product Estimator of Population Mean in Stratified Random Sampling

Subhash Kumar Yadav[1,*] *and Alok Kumar Shukla*[2]

[1] Department of Mathematics and Statistics (A Centre of Excellence), Dr. RML Avadh University, Faizabad 224001, U.P., India
[2] Department of Statistics, D.A.V College, Kanpur, U.P., India

Received: 5 Jun. 2014, Revised: 12 Oct. 2014, Accepted: 16 Oct. 2014
Published online: 1 Nov. 2014

Abstract: This manuscript deals with the estimation of population mean in stratified random sampling where the units are non-homogeneous for the characteristics under study using product cum dual to product estimator utilizing the negatively correlated auxiliary variable. The expressions for the bias and mean square error (MSE) of proposed estimator have been obtained up to the first degree of approximation. The optimum value of the constant α involved in the estimator that minimizes the mean square error of the proposed estimator has been obtained. The minimum value of the mean square error of proposed estimator for this optimum value of α has also been obtained. An efficiency comparison has also been made of the proposed estimator with other existing mentioned estimators of population mean under stratified random sampling. Finally an empirical study is also carried out to judge the performance of the proposed estimator along with the existing estimators under stratified random sampling.

Keywords: Ratio and Product estimators, Stratified Random Sampling, Bias, MSE, Efficiency

1 Introduction

Stratification is a technique of dividing the whole population into different relatively homogeneous groups known as strata. The technique of stratification is applied when the population units for the characteristics under study are non-homogeneous. It is well established that the use of the population information of auxiliary variable (z) improves the precision of the estimate(s) of the parameter(s) under study for the main characteristic (y) and it is in use since the use of the sampling itself. The auxiliary information is used at both the stages of designing and estimation. In the present investigation, it has been used at estimation stage for the estimation of population mean in stratified random sampling. The auxiliary variable is highly correlated (positively or negatively) with the main variable under study and the product method of estimation giving product and dual to type estimators is used when y and z are highly negatively correlated to each other. While on the other hand, ratio and dual to ratio type estimators are used when main variable and auxiliary variables are highly positively correlated to each other. In the present study we have considered the negatively correlated case only.

Laplace (1820) was the first to use the auxiliary information for the estimation of the total number of inhabitants of the study variable in France. Robson (1956) proposed the traditional product estimator for the estimation of population mean of the study variable utilizing highly negatively correlated auxiliary variable. Simple random sampling technique is the most appropriate technique of sampling when the population units are homogeneous for the characteristic under study. But when the units of the population are heterogeneous for the characteristic under study, then the simple random sampling scheme is not appropriate. In this situation the appropriate sampling technique is stratified random sampling where the whole population is divided into different groups known as strata or stratum in which the units are relatively homogeneous to each other. From these stratums the required sample is drawn either by proportional or optimal allocation method.

Let the finite population $u = (u_1, u_2, ..., u_N)$ consists of N distinct and identifiable units which are heterogeneous for the characteristic under study. Let the whole population is divided into L stratums of sizes N_h $(h = 1, 2, 3...., L)$, in which units

* Corresponding author e-mail: drskystats@gmail.com

are relatively homogeneous to each other. Let the study variable y and the auxiliary variable z under consideration takes the values y_{hi} and $z_{hi}\,(h=1,2,3....,L, i=1,2,3...,N_h)$ respectively for the i^{th} unit of the h^{th} strata. Let the sub-samples of sizes $n_h\,(h=1,2,3....,L)$ are drawn from each stratum using proportional allocation method, constituting the required sample of size $n=\sum\limits_{h=1}^{L} n_h$.

Following notations have been used throughout the manuscript for stratified random sampling which have been used by many authors in the literature,

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

$\bar{Z}_h=\frac{1}{N_h}\sum\limits_{i=1}^{N_h} z_{hi}$: The h^{th} stratum mean for the study variable z,

$\bar{Y}=\frac{1}{N}\sum\limits_{h=1}^{L}\sum\limits_{i=1}^{N_h} y_{hi}=\frac{1}{N}\sum\limits_{h=1}^{L} N_h\bar{Y}_h=\sum\limits_{h=1}^{L} W_h\bar{Y}_h$: The population mean of study variable y,

$\bar{Z}=\frac{1}{N}\sum\limits_{h=1}^{L}\sum\limits_{i=1}^{N_h} z_{hi}=\frac{1}{N}\sum\limits_{h=1}^{L} N_h\bar{Z}_h=\sum\limits_{h=1}^{L} W_h\bar{Z}_h$: The population mean of study variable z,

$\bar{y}_h=\frac{1}{n_h}\sum\limits_{i=1}^{n_h} y_{hi}$: The sample mean of the auxiliary variate y for the h^{th} stratum,

$\bar{z}_h=\frac{1}{n_h}\sum\limits_{i=1}^{n_h} z_{hi}$: The sample mean of the auxiliary variate z for the h^{th} stratum,

$W=\frac{N_h}{N}$: Weight for the h^{th} stratum.

The combined product estimator of population mean in stratified random sampling is defined as,

$$\hat{\bar{Y}}_{PC}=\bar{y}_{st}\left(\frac{\bar{z}_{st}}{\bar{Z}}\right) \tag{1}$$

where $\bar{y}_{st}=\sum\limits_{h=1}^{L} W_h\bar{y}_h$ and $\bar{z}_{st}=\sum\limits_{h=1}^{L} W_h\bar{z}_h$.

The Mean Square Error of the estimator $\hat{\bar{Y}}_{PC}$ up to the first degree of approximation is,

$$MSE(\hat{\bar{Y}}_{PC})=\sum_{h=1}^{L} W_h^2\lambda_h\left[S_{yh}^2+R^2S_{zh}^2+2RS_{yzh}\right], \tag{2}$$

where $\lambda_h=\left(\frac{1}{n_h}-\frac{1}{N_h}\right)$, $R=\frac{\bar{Y}}{\bar{Z}}=\frac{\bar{Y}_{st}}{\bar{Z}_{st}}$ is the population ratio, S_{yh}^2 is the population variance of the main variable y under study, S_{zh}^2 is the population variance of the auxiliary variable and S_{yzh} is the population covariance between study and auxiliary variables in stratum.

Kushwaha et.al (1990), using the estimator $\hat{\bar{Y}}_{PC}$ proposed the following dual to product estimator using Srivenkataramana (1980) transformation as,

$$\hat{\bar{Y}}_{PC}^{*}=\bar{y}_{st}\left(\frac{\bar{Z}}{\bar{z}_{st}^{*}}\right) \tag{3}$$

where $\bar{z}_{st}^{*}=\sum\limits_{h=1}^{L} W_h\bar{z}_h^{*}$ and $\bar{z}_h^{*}=\frac{\bar{Z}_hN_h-\bar{z}_hn_h}{N_h-n_h}$.

The Mean Square Error of $\hat{\bar{Y}}_{PC}^{*}$, up to the first order of approximation is,

$$MSE(\hat{\bar{Y}}_{PC})=\sum_{h=1}^{L} W_h^2\lambda_h\left[S_{yh}^2+R^2g_h^2S_{zh}^2+2Rg_hS_{yzh}\right], \tag{4}$$

where $g_h=\frac{n_h}{N_h-n_h}$.

Singh et.al (2008) suggested the following exponential product type estimators in stratified random sampling by adapting the Bahl and Tuteja (1991) estimator of population mean under simple random sampling as,

$$\hat{\bar{Y}}_{Pe}^{st}=\bar{y}_{st}exp\left(\frac{\bar{z}_{st}-\bar{Z}}{\bar{z}_{st}+\bar{Z}}\right)=\bar{y}_{st}exp\left(\frac{\sum\limits_{h=1}^{L} W_h(\bar{z}_h-\bar{Z}_h)}{\sum\limits_{h=1}^{L} W_h(\bar{z}_h+\bar{Z}_h)}\right), \tag{5}$$

The Mean Square Error of $\widehat{Y}_{Pe}^{st}$, up to the first order of approximation is,

$$MSE\left(\widehat{Y}_{Pe}^{st}\right)=\sum_{h=1}^{L}W_h^2\lambda_h\left[S_{yh}^2+\frac{R^2}{4}S_{zh}^2+RS_{yzh}\right]. \tag{6}$$

Tailor et.al (2013) proposed the following dual to product type estimator using Singh et.al (2008) estimators as,

$$\widehat{Y}_{Pe}^{*st}=\bar{y}_{st}exp\left(\frac{\bar{Z}-\bar{z}_{st}^{*}}{\bar{Z}+\bar{z}_{st}^{*}}\right)=\bar{y}_{st}exp\left(\frac{\sum_{h=1}^{L}W_h(\bar{Z}_h-\bar{z}_h^{*})}{\sum_{h=1}^{L}W_h(\bar{Z}_h+\bar{z}_h^{*})}\right). \tag{7}$$

The Mean Square Error $\widehat{Y}_{Pe}^{*st}$, up to the first order of approximation is,

$$MSE\left(\widehat{Y}_{Pe}^{*st}\right)=\sum_{h=1}^{L}W_h^2\lambda_h\left[S_{yh}^2+\frac{R^2}{4}g_h^2S_{zh}^2+Rg_hS_{yzh}\right]. \tag{8}$$

2 Proposed Estimator

We propose the following product cum dual to product estimator of population mean in stratified random sampling by combining the product and dual to product estimators as,

$$\begin{aligned}t&=\alpha\bar{y}_{st}exp\left(\frac{\bar{z}_{st}-\bar{Z}}{\bar{z}_{st}+\bar{Z}}\right)+(1-\alpha)\bar{y}_{st}exp\left(\frac{\bar{Z}-\bar{z}_{st}^{*}}{\bar{Z}+\bar{z}_{st}^{*}}\right)\\&=\alpha\bar{y}_{st}exp\left[\frac{\sum_{h=1}^{L}W_h(\bar{z}_h-\bar{Z}_h)}{\sum_{h=1}^{L}W_h(\bar{z}_h+\bar{Z}_h)}\right]+(1-\alpha)\bar{y}_{st}exp\left[\frac{\sum_{h=1}^{L}W_h(\bar{Z}_h-\bar{z}_h^{*})}{\sum_{h=1}^{L}W_h(\bar{Z}_h+\bar{z}_h^{*})}\right]\end{aligned} \tag{9}$$

where α is some suitable constant to be determined such that the mean square error of the proposed estimator t is minimum.
It is worth notable that for $\alpha=0$, the proposed estimator becomes the exponential dual to product estimator in (7) and for $\alpha=1$, it reduces to exponential product type estimator in (5).

To study the large sample properties of the proposed estimator t , let us define,
$\bar{y}_h=\bar{Y}_h(1+e_{0h})$ and $\bar{z}_h=\bar{Z}_h(1+e_{1h})$ such that
$E(e_{0h})=E(e_{1h})$, $E(e_{0h}^2)=\lambda_hC_{yh}^2$,$E(e_{1h}^2)=\lambda_hC_{zh}^2$ and $E(e_{0h}e_{1h})=\lambda_h\rho_{yzh}C_{yh}C_{zh}$.
Expressing the proposed estimator t in (9) in terms of $e_{ih}(i=0,1,2)$, we have

$$t=\alpha\sum_{h=1}^{L}W_h\bar{y}_hexp\left[\frac{\sum_{h=1}^{L}W_h(\bar{z}_h-\bar{Z}_h)}{\sum_{h=1}^{L}W_h(\bar{z}_h+\bar{Z}_h)}\right]+(1-\alpha)\sum_{h=1}^{L}W_h\bar{y}_hexp\left[\frac{\sum_{h=1}^{L}W_h(\bar{Z}_h-\bar{z}_h^{*})}{\sum_{h=1}^{L}W_h(\bar{Z}_h+\bar{z}_h^{*})}\right]$$

$$\begin{aligned}&=\alpha\sum_{h=1}^{L}W_h\bar{Y}_h(1+e_{0h})exp\left[\frac{\sum_{h=1}^{L}W_h\bar{Z}_he_{1h}}{2\sum_{h=1}^{L}W_h\bar{Z}_h+\sum_{h=1}^{L}W_h\bar{Z}_he_{1h}}\right]+(1-\alpha)\sum_{h=1}^{L}W_h\bar{Y}_h(1+e_{0h})exp\left[\frac{\sum_{h=1}^{L}W_hg_h\bar{Z}_he_{1h}}{2\sum_{h=1}^{L}W_h\bar{Z}_h-\sum_{h=1}^{L}W_hg_h\bar{Z}_he_{1h}}\right]\\&=\alpha\bar{Y}(1+e_0)exp\left[\left(\frac{e_1}{2}\right)\left(1+\frac{e_1}{2}\right)^{-1}\right]+(1-\alpha)exp\left[\left(\frac{e_{1g}}{2}\right)\left(1-\frac{e_{1g}}{2}\right)^{-1}\right]\end{aligned} \tag{10}$$

Where, $e_0=\frac{\sum_{h=1}^{L}W_h\bar{Y}_he_{0h}}{\bar{Y}}$, $e_1=\frac{\sum_{h=1}^{L}W_h\bar{Z}_he_{1h}}{\bar{Z}}$ and $e_{1g}=\frac{\sum_{h=1}^{L}W_h\bar{Z}_hg_he_{1h}}{\bar{Z}}$ such that
$E(e_0)=E(e_1)=E(e_{1g})=0$ and $E(e_0^2)=\frac{1}{\bar{Y}^2}\sum_{h=1}^{L}W_h^2\lambda_hS_{yh}^2$, $E(e_1^2)=\frac{1}{\bar{Z}^2}\sum_{h=1}^{L}W_h^2\lambda_hS_{zh}^2$, $E(e_{1g}^2)=\frac{1}{\bar{Z}^2}\sum_{h=1}^{L}W_h^2\lambda_hg_h^2S_{zh}^2$,

$E(e_0e_1) = \frac{1}{Y}\frac{1}{Z}\sum_{h=1}^{L} W_h^2\lambda_h S_{yzh}$, $E(e_0e_{1g}) = \frac{1}{Y}\frac{1}{Z}\sum_{h=1}^{L} W_h^2\lambda_h g_h S_{yzh}$, $E(e_1e_1) = \frac{1}{Z^2}\sum_{h=1}^{L} W_h^2\lambda_h g_h S_{zh}^2$.

On simplifying the expressions after expansion on right hand side of (10), up to the first order of approximation, we have

$$t = \bar{Y}\left[1 + e_0 + \alpha\left(\frac{e_1}{2} + \frac{e_0e_1}{2} - \frac{e_1^2}{8}\right) + (1-\alpha)\left(\frac{e_{1g}}{2} + \frac{e_0e_{1g}}{2} + \frac{3e_{1g}^2}{8}\right)\right]$$

$$t - \bar{Y} = \bar{Y}\left[e_0 + \alpha\left(\frac{e_1}{2} + \frac{e_0e_1}{2} - \frac{e_1^2}{8}\right) + \alpha_1\left(\frac{e_{1g}}{2} + \frac{e_0e_{1g}}{2} + \frac{3e_{1g}^2}{8}\right)\right] \tag{11}$$

where $\alpha_1 = 1-\alpha$.

Taking expectation on both sides of (11) and putting the values of different expectations, we have the bias of t , up to the first order of approximation as,

$$B(t) = \bar{Y}\left[\alpha\left\{\frac{1}{Y}\frac{1}{Z}\sum_{h=1}^{L} W_h^2\lambda_h S_{yzh} - \frac{1}{8Z^2}\sum_{h=1}^{L} W_h^2\lambda_h S_{zh}^2\right\} + \alpha_1\left\{\frac{1}{2Y}\frac{1}{Z}\sum_{h=1}^{L} W_h^2\lambda_h g_h S_{yzh} + \frac{3}{8Z^2}\sum_{h=1}^{L} W_h^2\lambda_h g_h^2 S_{zh}^2\right\}\right] \tag{12}$$

From (11), up to the first order of approximation, the MSE of t is,

$$MSE(t) \approx E\left[\bar{Y}\left\{e_0 + \alpha\frac{e_1}{2} + \alpha_1\frac{e_{1g}}{2}\right\}\right]^2$$

$$= \bar{Y}^2 E\left[\left\{e_0^2 + \frac{\alpha^2}{4}e_1^2 + \frac{\alpha_1^2}{4}e_{1g}^2 + \alpha e_0e_1 + \alpha_1 e_0e_{1g} + \frac{1}{2}\alpha\alpha_1 e_1e_{1g}\right\}\right]$$

$$= \bar{Y}^2\left[E(e_0^2) + \frac{\alpha^2}{4}E(e_1^2) + \frac{\alpha_1^2}{4}E(e_{1g}^2) + \alpha E(e_0)E(e_1) + \alpha_1 E(e_0)E(e_{1g}) + \frac{1}{2}\alpha\alpha_1 E(e_1)E(e_{1g})\right] \tag{13}$$

Which is minimum for,

$$\alpha = \frac{E(e_{1g}^2) - 2E(e_0e_1) + 2E(e_0e_{1g}) - E(e_1e_{1g})}{E(e_1^2) + E(e_{1g}^2) - 2E(e_1e_{1g})} = \frac{A}{B} \tag{14}$$

Where,

$A = E(e_{1g}^2 - 2E(e_0e_1) + 2E(e_0e_{1g}) - E(e_1e_{1g})$

$= \frac{1}{Z^2}\sum_{h=1}^{L} W_h^2\lambda_h g_h^2 S_{zh}^2 - \frac{2}{Y}\frac{1}{Z}\sum_{h=1}^{L} W_h^2\lambda_h S_{yzh} + \frac{2}{Y}\frac{1}{Z}\sum_{h=1}^{L} W_h^2\lambda_h g_h S_{yzh} - \frac{1}{Z^2}\sum_{h=1}^{L} W_h^2\lambda_h g_h S_{zh}^2,$

$B = E(e_1^2) + E(e_{1g}^2) - 2E(e_1e_{1g}) = \frac{1}{Z^2}\sum_{h=1}^{L} W_h^2\lambda_h S_{zh}^2 + \frac{1}{Z^2}\sum_{h=1}^{L} W_h^2\lambda_h g_h^2 S_{zh}^2 - \frac{2}{Z^2}\sum_{h=1}^{L} W_h^2\lambda_h g_h S_{zh}^2.$

The minimum MSE of t is,

$$MSE_{min}(t) = \bar{Y}^2\left[\left\{E(e_0^2) + \frac{1}{4}E(e_{1g}^2) + E(e_0e_{1g}) - \frac{A^2}{4B}\right\}\right]$$

$$MSE_{min}(t) = \sum_{h=1}^{L} W_h^2\lambda_h\left[S_{yh}^2 + \frac{1}{4}R^2 g_h^2 S_{zh}^2 + Rg_h S_{yzh}\right] - \bar{Y}^2\frac{A^2}{4B} \tag{15}$$

3 Efficiency Comparison

The proposed estimator in (9) is better than the traditional combined product estimator in (1) if,

$$MSE(\hat{\bar{Y}}_{PC}) - MSE_{min}(t) > 0$$

$$or \sum_{h=1}^{L} W_h^2\lambda_h\left[\left(1 - \frac{g_h^2}{4}\right)R^2 S_{zh}^2 + (2-g_h)RS_{yzh}\right] + \bar{Y}^2\frac{A^2}{4B} > 0 \tag{16}$$

From (15) and (4), we have
$MSE(\hat{\bar{Y}}^{*}_{PC}) - MSE_{min}(t) > 0$, if

$$\sum_{h=1}^{L} W_h^2 \lambda_h \left[\frac{3g_h^2}{4} R^2 S_{zh}^2 + g_h R S_{yzh} \right] + \bar{Y}^2 \frac{A^2}{4B} > 0 \tag{17}$$

From (15) and (6), we have
$MSE(\hat{\bar{Y}}^{st}_{Pe}) - MSE_{min}(t) > 0$, if

$$\sum_{h=1}^{L} W_h^2 \lambda_h \left[(1-g_h^2) \frac{R^2}{4} S_{zh}^2 + (1-g_h) R S_{yzh} \right] + \bar{Y}^2 \frac{A^2}{4B} > 0 \tag{18}$$

From (15) and (8), we have
$MSE(\hat{\bar{Y}}^{*st}_{Pe}) - MSE_{min}(t) > 0$, if

$$\bar{Y}^2 \frac{A^2}{4B} > 0 \tag{19}$$

4 Empirical Study

To examine the efficiency of the proposed estimator, we have considered the data set given in Table 1. The MSE and percent relative efficiency (PRE) values of different estimators are given in Table 2. It can be seen that the proposed estimator is quite efficient than the other estimators for the given population.

Population (Source: Japan Meteorological Society)
Y: Rainy days, Z: Total annual sunshine hours

Table 1: Statistics of Population

	n_h	N_h	$\bar{Y}_h$	$\bar{Z}_h$	S_{yh}	S_{zh}	S_{yzh}
Stratum I	4	10	142.80	1630	6.09	102.17	-239.30
Stratum II	4	10	102.60	2036	12.61	103.26	-655.30

Table 2: MSE and PRE Values of Product Estimators for Population

	$\hat{\bar{Y}}_{PC}$	$\hat{\bar{Y}}^{*}_{PC}$	$\hat{\bar{Y}}^{st}_{Pe}$	$\hat{\bar{Y}}^{*st}_{Pe}$	t
MSE	6.4081	5.9354	5.9945	6.5206	5.7923
PRE	100	107.9648	106.9003	102.5206	110.6325

5 Conclusion

In this study, we have proposed the product cum dual to product estimator of population mean under stratified random sampling scheme utilizing negatively correlated auxiliary variable. The expressions for the bias and mean square errors have been obtained for the proposed estimator as well as different product and dual to product estimators of population mean up to the first order of approximation. By MSE equations, the MSE values and the percentage relative efficiency (PRE) are obtained for different estimators. From the theoretical findings given in Section-3 and from the numerical results given in Table 2, we conclude that the proposed estimator is more efficient than the mentioned estimators in Section 1 under the stratified random sampling scheme as the proposed estimator have lesser MSE value. Thus, the proposed estimator should be preferred for the estimation of the population mean under stratified random sampling.

Acknowledgements

The authors are very much thankful to the editor in chief and the unknown learned referee for critically examining the manuscript and giving valuable suggestions to improve it.

References

[1] S. Bahl and R.K Tuteja, Ratio and product type exponential estimator, Information and optimization Sciences, **XII (I)**, 159-163 (1991).
[2] K.S.Kushwaha, L.N. Upadhyaya, S.P. Dubey, A dual to ratio estimator in stratified random sampling, Proc. Math. Soc., **6**, 11-15 (1990).
[3] P.S. Laplace, A Philosophical Essay on Probabilities. English translation, Dover, (1951).
[4] D.S. Robson, Application of multivariate polykays to the theory of unbiased ratio type estimation, JASA,**52**, 411-422 (1957).
[5] T. Srivenkataramana, A dual to ratio estimator in sample surveys, Biometrika, **67**, 199-204 (1980).
[6] R. Jatwa Tailor, R. N.K. Tailor, N. Garg, Dual to ratio and product type exponential estimators in stratified random sampling using two auxiliary variates, Journal of Reliability and Statistical Studies, **6**, 2, 115-126 (2013).
[7] R. Singh, P. Chauhan, N. Sawan, On linear combination of Ratio-product type exponential estimator for estimating finite population mean, Statistics in Transition,**9**(1),105-115 (2008).