An Efficient Variant of Dual to Ratio and Product Estimator in Sample Surveys

Gajendra K. Vishwakarma
Department of Applied Mathematics
Indian School of Mines

Raj K. Gangele
Department of Mathematics and Statistics
Dr. Harisingh Gour University (Central University)

Ravendra Singh
Department of Mathematics & Statistics
Banasthali University

In this paper, we propose a dual to ratio and product estimator for estimating finite population mean of study variable on applying simple transformation to auxiliary variable by using its average values in the population that are generally available in practice. The mean squared error of the proposed estimator have been obtained to the first degree of approximation. It has also been shown that the proposed estimator has greater applicability and is more efficient than the usual estimator. An empirical study is carried out to demonstrate the performance of proposed estimator.

Keywords: *auxiliary variable, study variable, mean square error, population mean, simple random sampling*

1. Introduction

The ratio method for estimation of population parameters are generally used when auxiliary information about study variable is available. The ratio estimators can be effective when auxiliary variable is positively correlated with study variable. The product estimators behaves similarly as ratio, however, they can be used in case negatively correlated auxiliary variable is available. In order to improve the efficiency of estimation, combination of ratio and product estimators can be used. Some of the important work in this direction is made. Shukla (1966) presented an alternative multivariate ratio estimator, Singh (1967) have defined multivariate product method of estimation, Singh (1969) worked on ratio-cum-product estimator, dual variables for estimation of population parameters were introduced by Bandyopadhyay(1980) and Srivenkataramana (1980), dual to ratio-cum-product estimator is presented in Singh et al(2005),Singh and Vishwakarma (2008) proposed some estimators using transformation on auxiliary variable.

Let $U = \{U_1, U_2,\ldots, U_N\}$ be a finite population of N (given) units. Let y and x denote the study variable and the auxiliary variable taking the values y_i and x_i respectively on the unit U_i (i = 1,2,…, N). Let the variable y and x be highly correlated. Then it is possible to improve upon the conventional unbiased estimator of the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ by the simple ratio method of estimation (or product method of estimation) provided the correlation coefficient ρ, between study variable y and the auxiliary variable x has high positive value (or high negative value). For efficient use of a given auxiliary variable, Murthy (1964) has suggested the use of usual ratio estimator as

$$\bar{y}_R = \bar{y}\frac{\bar{X}}{\bar{x}} \tag{1.1}$$

usual product estimator

$$\bar{y}_P = \bar{y}\frac{\bar{x}}{\bar{X}} \tag{1.2}$$

Here $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ are sample means of y and x respectively based on a simple random sample of size n drawn without replacement (SRSWOR) from the population U, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ is the known population mean of the auxiliary variable x.

In SRSWOR sampling scheme, let N and $n < N$ be the population and the sample sizes respectively. Then clearly, $\bar{x}^* = \left(N\bar{X} - n\bar{x}\right)/\left(N - n\right)$ (is the mean of x-values of (N-n) unobserved units) is also unbiased estimator of the population mean $\bar{X}$. If y and x has positive correlation, then correlation between $\bar{y}$ and $\bar{x}^*$ is negative $\left\{i.e.\ Corr\left(\bar{y},\bar{x}^*\right) = -\rho\right\}$. In such a situation, Srivenkataramana (1980) suggested a dual to ratio estimator for $\bar{Y}$ as

$$\bar{y}_R^* = \bar{y}\frac{\bar{x}^*}{\bar{X}} \tag{1.3}$$

On the other hand, if the correlation between y and x is negative, Bandyopadhyay (1980) suggested a dual to product estimator for $\bar{Y}$ as

$$\bar{y}_P^* = \bar{y}\frac{\bar{X}}{\bar{x}^*} \tag{1.4}$$

In this paper, we have proposed an estimator for estimating population mean $\bar{Y}$ using simple transformations to auxiliary variable when, straight line of y on x does not pass through the neighborhood of the origin. The properties of proposed estimator have been derived. The comparisons of the proposed estimator with the existing estimators have been made with respect to their mean squared errors.

2. Proposed Estimator

We propose the following dual to ratio and product estimator for estimating the population mean $\bar{Y}$ as

$$\hat{\bar{Y}}_{RP}^* = \bar{y}\left\{\frac{\bar{x}^* + \theta\bar{X}}{\bar{X} + \theta\bar{x}^*}\right\} \tag{2.1}$$

where θ is a scalar used as the design parameter. It is worth noting that, for $\theta = 1$, $\hat{\bar{Y}}_{RP}^* = \bar{y}$ and that, for $\theta = 1$, $\hat{\bar{Y}}_{RP}^* = \bar{y}_R^*$. Moreover, if θ is very large, $\hat{\bar{Y}}_{RP}^*$ is almost the same as $\bar{y}_P^*$. Thus $\bar{y}$, $\bar{y}_R^*$ and $\bar{y}_P^*$ are particular cases of proposed estimator $\hat{\bar{Y}}_{RP}^*$.

To evaluate expected value of $\hat{\bar{Y}}_{RP}^*$, let $\bar{y} = \bar{Y}\left(1+e_0\right)$ and $\bar{x} = \bar{X}\left(1+e_1\right)$ such that $E(e_0) = E(e_1) = 0$ and under *SRSWOR* sampling scheme

$$\left.\begin{aligned} E\left(e_0^2\right) &= \frac{\left(1-f\right)}{n}C_y^2 \\ E\left(e_1^2\right) &= \frac{\left(1-f\right)}{n}C_x^2 \\ E\left(e_0 e_1\right) &= \frac{\left(1-f\right)}{n}\rho\, C_y C_x \end{aligned}\right\} \tag{2.2}$$

Now expressing $\hat{\bar{Y}}_{RP}^{*}$ in terms of e's and we can reasonably assume that the sample sizes are large enough to make $|e_0| << 1$, and $|e_1| << 1$. Now we have

$$\hat{\bar{Y}}_{RP}^{*} - \bar{Y} = \bar{Y}\left[\left(1+e_0\right)\left\{1-\frac{g\,e_1}{(1+\theta)}\right\}\left\{1-\frac{\theta\,g\,e_1}{(1+\theta)}\right\}^{-1} - 1\right] \tag{2.3}$$

Expanding the right hand side of (2.3) and neglecting the terms involving e_0 and/or e_1 in a degree greater than two, we have

$$\left(\hat{\bar{Y}}_{RP}^{*} - \bar{Y}\right) = \bar{Y}\left[e_0 - Fge_1 - Fge_0e_1 - F\frac{\theta}{(1+\theta)}g^2e_1^2\right] \tag{2.4}$$

Further squaring both sides of (2.4) and neglecting the terms involving e_0 and/or e_1 in a degree greater than two, we have

$$\left(\hat{\bar{Y}}_{RP}^{*} - \bar{Y}\right)^2 = \bar{Y}^2\left[e_0^2 + F^2g^2e_1^2 - 2\,Fg\,e_0e_1\right] \tag{2.5}$$

Taking expectations both sides of (2.5) and using the results given by (2.2), we get the MSE of $\hat{\bar{Y}}_{RP}^{*}$ to the first degree of approximation

$$MSE\left(\bar{Y}_{RP}^{*}\right) = \lambda\bar{Y}^2\left[C_y^2 + FgC_x^2\left(Fg - 2K\right)\right] \tag{2.6}$$

where, $\lambda = \left(\frac{1}{n} - \frac{1}{N}\right)$, $g = \frac{n}{N-n}$, $K = \rho\frac{C_y}{C_x}$, $F = \left(\frac{1-\theta}{1+\theta}\right)$, $C_y = \frac{S_y}{\bar{Y}}$,

$C_x = \frac{S_x}{\bar{X}}$, $S_y^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$, $S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$.

Remark 2.1. Mean squared errors of $\bar{y}$ and $\bar{y}_R^{*}$ can be obtained by putting $\theta = 1$ and $\theta = 0$ respectively in (2.6). If θ is very large in (2.6), it can be easily obtained the MSE of $\bar{\ }$.

To the first degree of approximation, the *MSEs* of $\bar{y}_R$, $\bar{y}_P$, $\bar{y}_R^{*}$, $\bar{y}_P^{*}$ and $\hat{\bar{Y}}_{RP}$ are respectively given by

$$MSE\left(\bar{y}_R\right)=\lambda\bar{Y}^2\left[C_y^2+C_x^2\left(1-2K\right)\right] \tag{2.7}$$

$$MSE\left(\bar{y}_P\right)=\lambda\bar{Y}^2\left[C_y^2+C_x^2\left(1+2K\right)\right] \tag{2.8}$$

$$MSE\left(\bar{y}_R^*\right)=\lambda\bar{Y}^2\left[C_y^2+C_x^2 g\left(g-2K\right)\right] \tag{2.9}$$

$$MSE\left(\bar{y}_P^*\right)=\lambda\bar{Y}^2\left[C_y^2+C_x^2 g\left(g+2K\right)\right] \tag{2.10}$$

Minimizing MSE of $\hat{\bar{Y}}_{RP}^*$ with respect to F, we get the optimum value of F as

$$F=\frac{\rho C_y}{g C_x}=F_o \tag{2.11}$$

hence 'optimum' value of θ is given by

$$\theta=\left(\frac{1-F_0}{1+F_0}\right)=\theta_0 \text{ (say)}$$

Putting (2.11) in (2.6) we get the minimum *MSE* of $\hat{\bar{Y}}_{RP}^*$ as

$$\min MSE\left(\hat{\bar{Y}}_{RP}^*\right)=\lambda S_y^2\left(1-\rho^2\right) \tag{2.12}$$

which is equal to the approximate variance of the regression estimator

$$\bar{y}_{LR}=\bar{y}+\hat{\beta}\left(\bar{X}-\bar{x}\right), V\left(\bar{y}_{LR}\right)=\lambda S_y^2\left(1-\rho^2\right) \tag{2.13}$$

where $\hat{\beta}=s_{yx}/s_x^2$ is the estimate of the population regression coefficient $\beta=S_{yx}/S_x^2$ of y on x.

Thus we established the following theorem.

Theorem 2.1. To the first degree of approximation, the min. *MSE* of the estimator $\hat{\bar{Y}}_{RP}^*$ "based on estimated optimum value" is same as that of the variance of the estimator $\bar{y}_{LR}$.

Remark 2.2. We have the following rule of conduct. If a good guess of F_0, say F, is available, we use, $\theta = (1 - F)/(1 + F)$ unless $|F_0| \cong 1$, which implies $\theta \cong 0$ or $|\theta|$ very large. In these two cases the modified estimator will be almost the same as the product or ratio estimator respectively and, therefore, then we use the latter methods.

3. Efficiency Comparisons

It is well known under *SRSWOR* sampling scheme, that

$$Var\left(\bar{y}\right) = \lambda \bar{Y}^2 C_y^2 \tag{3.1}$$

Reddy (1978) has shown that the values of parameter $K = \rho\ (C_y / C_x)$ remain stable in any repetitive survey. So we find the conditions on the values of K under which the proposed estimators are superior to the existing ones. For the present situation, we note that $0 < K < (C_y / C_x)$. In the survey sampling situations, usually $\{n / (N - n)\} < 0.1$ but we assume that $\{n / (N - n)\} < 1$ which hold good in almost all the situations of survey sampling.

From (2.6), (2.7), (2.8), (2.9), (2.10) and (3.1) we note that

(i) $MSE\left(\hat{\bar{Y}}_{RP}^{*}\right) < Var\left(\bar{y}\right)$ for $K > \dfrac{F\,g}{2}$

i.e. if $|F - F_0| < |F_0|$ (3.2)

(ii) $MSE\left(\hat{\bar{Y}}_{RP}^{*}\right) < MSE\left(\bar{y}_R\right)$ for $K > \dfrac{F\,g + 1}{2}$

i.e. if $\left|F - F_0\right| < \left|\dfrac{1}{g} + F_0\right|$ (3.3)

(iii) $MSE\left(\hat{\bar{Y}}_{RP}\right) \quad MSE\left(\bar{y}_P\right)$ for $K > \dfrac{F\,g - 1}{2}$

i.e. if $\left|F - F_0\right| < \left|\dfrac{1}{g} - F_0\right|$ (3.4)

(iv) $MSE\left(\hat{\bar{Y}}_{RP}^{*}\right) < MSE\left(\bar{y}_R^{*}\right)$ for $K > \dfrac{g\ (F + 1)}{2}$

i.e. if $|F - F_0| < |1 - F_0|$ (3.5)

(v) $MSE\left(\hat{\bar{Y}}_{RP}^{*}\right) < MSE\left(\bar{y}_P^{*}\right)$ for $K > \dfrac{g\ (F - 1)}{2}$

i.e. if $|F - F_0| < |1 + F_0|$ (3.6)

4. Empirical Study

To illustrate the performance of the constructed estimator $\hat{\bar{Y}}^{*}_{RP}$ over $\bar{y}$, $\bar{y}_R$, $\bar{y}_P$, $\bar{y}^{*}_R$ and $\bar{y}^{*}_P$ numerically, we consider six populations along with different parametric values. The percent relative efficiency of an estimator $t\,;\left(t=\bar{y}_R,\,\bar{y}_P,\,\bar{y}^{*}_R,\,\bar{y}^{*}_P,\,\bar{y}^{*}_{RP}\right)$ with respect to usual unbiased estimator $\bar{y}$ is defined by

$$PRE(t,\,\bar{y})=\frac{V(\bar{y})}{MSE(t)}\times 100 \tag{4.1}$$

We have computed the percent relative efficiencies of $\bar{y}$, $\bar{y}_R$, $\bar{y}_P$, $\bar{y}^{*}_R$, $\bar{y}^{*}_P$ and $\hat{\bar{Y}}^{*}_{RP}$, with respect to $\bar{y}$ which is presented in Table 1. The description of the populations and values of required parameters are given below:

Population I - [Source: Murthy (1967, p. 228)]
x: Fixed capital, : Output
N = 80, n = 20, $\bar{Y}$ = 51.8264, Cy = 0.3542, Cx = 0.7507, ρ = 0.9413.

Population II- [Source: Murthy (1967, p. 228)]
x: Number of workers, y: Output
N = 80, n = 20, $\bar{Y}$ = 51.8264, Cy = 0.3542, Cx = 0.9484, ρ = 0.9150.

Population III - [Source: Das (1988)]
x: Number of agricultural laborers for 1961,
y: Number of agricultural laborers for 1971,
N = 278, n = 30, $\bar{Y}$ = 39.0680, Cy = 1.4451, Cx = 1.6198, ρ = 0.7213.

Population IV - [Source: Johnston (1972, p. 171)]
x: Percentage of HIVs affected by disease,
y: Date of flowering of a particular summer species
(number of days from January 1)
N = 10, n = 4, $\bar{Y}$ = 200, Cy = 0.0458, Cx = 0.1562, ρ = – 0.94.

Population V - [Source: Johnston (1972, p. 171)]
x: Mean January temperature,
y: Date of flowering of a particular summer species
(number of days from January 1)
N = 10, n = 4, $\bar{Y}$ = 200, Cy = 0.0458, Cx = 0.1304, ρ = – 0.73.

Population VI - [Source: Maddala (1977)]
x: Deflated prices of veal, y: Consumption per capita,
N = 16, n = 4, $\overline{Y}$ = 7.6375, Cy = 0.2278, Cx = 0.0986, ρ = – 0.6823.

Table 1.
Percent relative efficiencies of $\overline{y}$, $\overline{y}_R$, $\overline{y}_P$, $\overline{y}_R^*$, $\overline{y}_P^*$ and $\hat{\overline{Y}}_{RP}^*$ with respect to $\overline{y}$

Estimator	Population					
	I	II	III	IV	V	VI
$\overline{y}$	100.00	100.00	100.00	100.00	100.00	100.00
$\overline{y}_R$	66.58	30.59	156.40	5.25	7.54	56.24
$\overline{y}_P$	10.55	7.65	25.82	16.08	20.20	167.59
$\overline{y}_R^*$	191.69	224.06	121.54	9.57	13.56	82.12
$\overline{y}_P^*$	60.18	54.16	82.37	52.77	54.60	121.37
$\hat{\overline{Y}}_{RP}^*$	877.54	614.34	208.45	859.11	214.09	187.10

5. Conclusion

On the basis of the theoretical and empirical results derived in this paper, it may be concluded that for the known value of K in any population, we can choose the most efficient estimator among the estimators namely $\overline{y}$, $\overline{y}_R$, $\overline{y}_P$, $\overline{y}_R^*$, $\overline{y}_P^*$ and $\hat{\overline{Y}}_{RP}^*$. Table 1 clearly indicates that the proposed dual to ratio and product estimator $\hat{\overline{Y}}_{RP}^*$ is more efficient than $\overline{y}$, $\overline{y}_R$, $\overline{y}_P$, $\overline{y}_R^*$ and $\overline{y}_P^*$ with considerable gain in efficiency in each population data sets. Thus the use of suggested estimator $\hat{\overline{Y}}_{RP}^*$ is preferable over other estimators in practice.

ACKNOWLEDGEMENT

The authors are heartily thankful to the editor Professor Erniel B. Barrios and learned referee for their comments which have made substantial improvement to bring the original manuscript to its present form.

REFERENCES

BANDYOPADHYAY, S., 1980, Improved ratio and product estimators. *Sankhya,* 42, C, 45-49.

DAS, A.K., 1988,Contribution to the theory of sampling strategies based on auxiliary information. Ph.D. thesis submitted to BCKV, Mohanpur, Nadia, West-Bengal, India.

JOHNSTON, J., 1972, *Econometric methods, (2nd ed),*Mc-Graw-Hill, Tokyo.

MADDALA, G.S., 1977, *Econometrics*. Mc-Graw Hills Pub. Co. New York.

MURTHY, M.N., 1964, Product method of estimation. *Sankhya,* 26, A, 69-74.

_____________, 1967, *Sampling theory and methods*. Statistical Publishing Society, Calcutta, India.

REDDY, V. N., 1978, A study on the use of prior knowledge on certain population parameters in estimation, *Sankhya C*, 40, 29–37.

SHUKLA, G.K., 1966, An alternative multivariate ratio estimator for finite population. *Calcutta Statistical Association Bulletin,* 15, 127-134.

SINGH, M.P., 1967, Multivariate product method of estimation for finite populations. *Journal of the Indian Society of Agricultural Statistics,* 19, 2, 1-10.

___________, 1969, Comparison of some ratio-cum-product estimators. *Sankhya,* 31,B, 375-382.

SINGH, H. P., SINGH, R., ESPEJO, M. R., PINEDA, M. D., NADARAJAN, S., 2005, On the efficiency of the dual to ratio-cum-product estimator. *Mathematical Proceedings of the RoyalIrishAcademy*,105, A(2), 51-56.

SINGH, H.P. AND VISHWAKARMA, G.K., 2008,Some estimators of finite population mean using auxiliary information in sample surveys. *Journal of Applied Statistical Sciences,* 16, 4, 11-22.

SRIVENKATARAMANA, T., 1980, A dual to ratio estimator in sample surveys. *Biometrika*, 67, 199-204.