ALTERNATIVE MULTIVARIATE RATIO ESTIMATORS USING GEOMETRIC AND HARMONIC MEANS

By

S. MOHANTY AND L.M. PATTANAIK

O. U. A. T., Bhubaneswar

(Received : May, 1982)

SUMMARY

It has been shown through the present investigation that in case of multi-auxiliary variables, estimates based on geometric mean and harmonic mean are less biased than Olkin's estimate based on arithmetic mean under certain conditions usually satisfied in practice. However, the mean square error of these estimates are the same up to $0(n^{-1})$.

1. INTRODUCTION

Olkin [3] has considered the use of multi-auxiliary variables, positively correlated with the variable under study to build up a multi-variate ratio estimator of the population mean $\bar{Y}$. This estimator is efficient under certain conditions usually found in practice.

Olkin's estimator is based on the weighted arithmetic mean of $r_i\bar{X}_i$'s and is given by

$$\bar{y}_{ap} = \sum_{i=1}^{p} w_i \, r_i \, \bar{X}_i, \qquad \ldots(1.1)$$

where (i) w_i's are weights such that $\sum_{i=1}^{p} w_i = 1$, (ii) $\bar{X}_i$'s are the population means of the auxiliary variables and are assumed to be known and (iii) $r_i = \bar{y}/\bar{x}_i$, $\bar{y}$ and $\bar{x}_i$'s are the sample means of the study variable Y and the auxiliary variables X_i's respectively

* Present Address : Department of Statistics, Utkal University, Bhubaneswar-4.

based on a simple random sample of size n drawn without replacement from a population of size N.

Following Olkin's estimator, several other estimators using multi-auxiliary variables have been proposed in recent years. Singh[7], has extended Olkin's estimator to the case where auxiliary variables are negatively correlated with the variable under study. Srivastava [9], and Rao and Mudholkar [5] have given estimates, where some of the characters are positively and others are negatively correlated with the character under study. Raj [4], Shukla [6], Srivastava [10] and John [1], have considered several Olkin-type eastimators which are linear combinations of the eastimators based on each auxiliary character separately. Singh [8] has given the ratio-cum-product estimator and Mohanty [2] has given regression-cum-ratio eastimator using multi-auxiliary variables. The main objective of presenting these estimators are to reduce the biases and mean square errors.

In this paper we have suggested two alternative eastimators based on geometric mean and harmonic mean, These estimators are

$$\bar{y}_{gp}=\prod_{i=1}^{p}(r_i\,\bar{X}_i)^{w_i} \qquad \ldots(1.2)$$

and

$$\bar{y}_{hp}=\left(\sum_{i=1}^{p}\frac{w_i}{r_i\,\bar{X}_i}\right)^{-1} \qquad \ldots(1.3)$$

such that $\sum_{i=1}^{p} w_i=1.$

These estimators are based on the assumptions that the auxiliary characters are positively correlated with Y.

Let ρ_{ij} be the correlation coefficient between X_i and X_j $(i\neq j)$ and ρ_{0i} be the correlation coefficient between Y and X_i,

$$S_0^2=(N-1)_{-1}\sum_{t=1}^{N}(Y_t-\bar{Y})^2,$$

$$S_i^2=(N-1)^{-1}\sum_{i=1}^{N}(X_{i_t}-\bar{X}_i)^2,$$

$$C_0^2=S_0^2/\bar{Y}^2 \text{ and } C_i^2=S_i^2/\bar{X}_i^2,\ i=1,2,3,\ldots,p.$$

Similarly C_{oi} and C_{ij} are defined.

Further, let $\underset{\sim}{w}'=(w_1, w_2, \ldots, w_p)$, $C=[C_{ij}]_{p\times p}$,

$\underset{\sim 0}{C}'=(C_{01}, C_{02}, \ldots, C_{op})$ and $\underset{\sim}{e}$ be a $(p\times 1)$ vector of elements as one.

2. Bias and Mean Square Error of the Estimates.

Let $e_0=(\bar{y}-\bar{Y})/\bar{Y}$ and $e_i=(\bar{x}_i-\bar{X}_i)/\bar{X}_i$, $i=1, 2, \ldots, p$. Using Taylor's series expansion under the usual assumption we obtain,

$$\bar{y}_{ap}=\bar{Y}\sum_{i=1}^{p} w_i[1+(e_0-e_i)+(e_i^2-e_i\, e_0)+(e_0\, e_i^2-e_i^3)$$
$$+(e_i^4-e_i^3 e_0)+\ldots] \qquad \ldots(2.1)$$

$$\bar{y}_{gp}=\bar{Y}\prod_{i=1}^{p}\Big[1+e_0-(e_i+e_i\, e_0)w_i+\frac{w_i\,(1+w_i)}{2}(e_i^2+e_0\, e_i^2)$$
$$-\frac{w_i(1+w_i)\,(2+w_i)}{6}(e_i^3+e_i^3\, e_0)+\ldots\Big] \qquad \ldots(2.2)$$

and

$$\bar{y}_{hp}=\bar{Y}[1+e_0-\sum_{i=1}^{p} w_i\, e_i-\sum_{i=1}^{p} w_i\, e_i\, e_0+(\sum_{i=1}^{p} w_i\, e_i)^2+(\sum_{i=1}^{p} w_i\, e_i)^2 e_0$$
$$-(\sum_{i=1}^{p} w_i\, e_i)^3-(\sum_{i=1}^{p} w_i\, e_i)^3\, e_0+(\sum_{i=1}^{p} w_i\, e_i)^4+\ldots] \qquad \ldots(2.3)$$

To calculate the bias and mean square error, we have considered only the terms involving powers up to second degree as the calculations become very much complicated when higher order terms are considered.

Thus, from (2.1), (2.2) and (2.3), the biases and mean square errors of the estimates up to $0\,(1/n)$ are obtained as follows :

$$B\,(\bar{y}_{ap})=\theta\bar{Y}[\sum_{i=1}^{p} w_i\, C_i^2-\sum_{i=1}^{p} w_i\, C_{0i}] \qquad \ldots(2.4)$$

$$B(\bar{y}_{gp})=\theta\bar{Y}\Big[\frac{1}{2}\sum_{i=1}^{p} w_i(w_i+1)\ C_i^2+\sum\sum_{i<j} w_i\, w_j\, C_{ij}$$

$$-\sum_{i=1}^{p} w_i\, C_{0i}] \qquad \ldots(2.5)$$

$$B(\bar{y}_{hp})=\theta\bar{Y}(\sum_{i=1}^{p} w_i^2\ C_i^2+2\ \sum\sum_{i>j} w_i\, w_j\, C_{ij}-\sum_{i=1}^{p} w_i\, C_{0i}] \quad \ldots(2.6)$$

and $MSE(\bar{y}_{ap})=MSE(\bar{y}_{gp})=MSE(\bar{y}_{hp})$

$$=\theta\bar{Y}^2[C_0^2+\sum_{i=1}^{p} w_i^2\ C_i^2-2\sum_{i=1}^{p} w_i C_{0i}+2\sum\sum_{i<j} w_i\, w_j\, C_{ij}]$$

$$=\theta\bar{Y}^2[C_0^2+\underset{\sim}{w}'C\underset{\sim}{w}-2\underset{\sim}{w}'\ \underset{\sim}{C_0^*}]_s \qquad \ldots(2.7)$$

where $\theta=\left(\frac{1}{n}-\frac{1}{N}\right)$.

From this we note that mean square errors of these estimates are the same where as biases are different.

It is well known that in univariate case the usual ratio estimator $\bar{y}_R$ for the ith auxiliary variable is superior to the mean per unit estimator $\bar{y}$, when

$$\frac{C_0}{C_1}\rho_{0i}>\tfrac{1}{2}. \qquad \ldots(2.8)$$

Comparing the variance of $\bar{y}(=\theta\bar{Y}^2\ C_0^2)$ with the mean square error given in (2.7), we note that the ratio estimator given in (1.1), (1.2) and (1.3) are more efficient than $\bar{y}$, when

$$\frac{\text{w}'\ \underset{\sim}{C^*}}{\underset{\sim}{w}'C\underset{\sim}{w}}>\tfrac{1}{2}. \qquad \ldots(2.9)$$

3. Comparison of Biases.

Biases may be either positive of negative. So, for the comparison purpose, we have compared the absolute biases of the estimates when these are more efficient than $\bar{y}$. The bias of the estimate $\bar{y}_{gp}$ is smaller than that of $\bar{y}_{ap}$ when

$$|\ B(\bar{y}_{ap})\ |>B(\bar{y}_{gp})\ |\ . \qquad \ldots(3.1)$$

Squaring and simplifying (3.1), we observe that, bias of $\bar{y}_{gp}$ is less than that of $\bar{y}_{ap}$ if

$$\left[\frac{1}{2}\sum_{i=1}^{p} w_i^2 C_i^2+\sum\sum_{i<j} w_i w_j C_{ij}+\frac{3}{2}\sum_{i=1}^{p} w_i C_i^2-2\sum_{i=1}^{p} w_i C_{0i}\right]$$

$$\times\left[\frac{1}{2}\sum_{i=1}^{p} w_i C_i^2-\frac{1}{2}\sum_{i=1}^{p} w_i^2 C_i^2-\sum\sum_{i<j} w_i w_j C_{ij}\right]>0. \quad ...(3.2)$$

Thus above inequality is true when both the factors are either positive or negative. The first factor of (3.2) i.e.

$$\frac{1}{2}\sum_{i=1}^{p} w_i^2 C_i^2+\sum\sum_{i<j} w_i w_j C_{ij}+\frac{3}{2}\sum_{i=1}^{p} w_i C_i^2-2\sum_{i=1}^{p} w_i C_{0i}$$

is positive, when (using 2.9)

$$\frac{\sum_{i=1}^{p} w_i C_i^2}{\underset{\sim}{w}'C\underset{\sim}{w}}>\frac{1}{3}. \quad ...(3.3)$$

Similarly, it can be shown that the second factor of (3.2) is also positive when

$$\frac{\sum_{i=1}^{p} w_i C_i^2}{\underset{\sim}{w}'C\underset{\sim}{w}}>1. \quad ...(3.4)$$

When both the factors of (3.2) are negative, the sign of inequalities of (3.3) and (3.4) are just reversed.

Further, comparing the square of the biases of $\bar{y}_{gp}$ and $\bar{y}_{hp}$, we find that, $\bar{y}_{gp}$ is more biased than $\bar{y}_{hp}$ when the inequalities (2.9) and (3.4) hold good.

Hence, we may conclude that under the situations where $\bar{y}_{ap}$, $\bar{y}_{gp}$ and $\bar{y}_{hp}$ are more efficient than $\bar{y}$ and the relation (3.4) or

$$\frac{\sum_{i=1}^{p} w_i C_i^2}{\underset{\sim}{w}'C\underset{\sim}{w}}<\frac{1}{3}$$

is satisfied, the biases of the estimates satisfy the relation

$$|B(\bar{y}_{hp})<|B(\bar{y}_{gp})|<|B(\bar{y}_{ap})|. \quad ...(3.5)$$

Usually the weights w_i's are so chosen so as to minimise the mean square error of an esitimator subject to the condition $\sum_{i=1}^{p} w_i = 1$.

Following the procedure used by Olkin [3], for determination of optimum weights it can be established that

$$\underset{\sim}{w} = C^{-1} \underset{\sim}{C_0^*} - \left(\frac{\underset{\sim}{e}' C^{-1} \underset{\sim}{C_0^*} - 1}{\underset{\sim}{e}' C^{-1} \underset{\sim}{e}} \right) C^{-1} \underset{\sim}{e}. \quad \ldots(3.6)$$

Comparison of the biases using these optimum weights given by (3.6) is very cumbersome as it involves complicated matrices. For making comparison easier, let us assume $C_i = C$, $\rho_{oi} = \rho_o$ and $\rho_{ij} = \rho$ for $i, j = 1, 2, \ldots\ldots, P$. Under these assumptions the optimum weights are uniform and equal to $1/p$. Substituting these values in (2.9), we note that the estimate based on arithmetic mean, geometric mean and harmonic mean are more efficient than $\bar{y}$ when,

$$\frac{p\,\rho_o}{1+(p-1)\rho} \cdot \frac{C_o}{C} > \tfrac{1}{2}. \quad \ldots(3.7)$$

4. Illustrations

The first part of this section deals with the illustration of the methodology using a natural population. The second part deals with the study of the efficiencies of the estimates based on all possible samples, for which we have considered one natural and one artificial population.

(*a*) The data of the natural population reported by Shukla [6] relate to the 50 sunhemp plants sown at Sunhemp Research Station, Pratapgarh in the year 1965-66. The three characters Y, X_1 and X_2 are yield of fiber per plant, base diameter and plant weight respectively.

The following population values are available :

$$\bar{Y} = 2.584,\ \bar{X}_1 = 0.808,\ \bar{X}_2 = 0.708$$

and

$$C = \begin{pmatrix} 0.08664 & 0.02340 & 0.06952 \\ & 0.01705 & 0.03807 \\ & & 0.11695 \end{pmatrix}.$$

Biases and mean square errors of different estimators under comparison, based on the above data are given in Table 4.1.

TABLE—4.1

Estimator	Auxiliary variate used	$\frac{\text{Bias}}{\theta \bar{Y}}$	$\frac{\text{MSE}}{\theta \bar{Y}^2}$
1. Mean per unit $\bar{y}$	none	0	0.08664
2. Ratio $\bar{y}\frac{\bar{X}_1}{\bar{x}_1}$	X_1	0.00635	0.05689
3. Ratio $\bar{y}\frac{\bar{X}_2}{\bar{x}_2}$	X_2	0.04743	0.06455
4. Olkin $\bar{y}_{ap}$	X_1 & X_2	0.01698	0.04600
5. Suggested $\bar{y}_{gp}$	X_1 & X_2	0.00987	0.04600
6. Suggested $\bar{y}_{hp}$	X_1 & X_2	0.00275	0.04600
7. $\bar{y}_{RS}=\bar{y}\frac{w_2\bar{X}_1+w_2\bar{X}_2}{w_1\bar{x}_1+w_2\bar{x}_1}$ (Shukla, 1966)	X_1 & X_2	0.00276	0.04600
8. $\bar{y}_{RP}=\bar{y}\frac{\bar{X}_1}{\bar{x}_1}\frac{\bar{X}_2}{\bar{x}_2}$ (Singh, 1967 b)	X_1 & X_2	0.07915	0 11094

From the given example it is seen that the multivariate ratio estimate based on harmonic mean is the least biased. However the mean square errors of $\bar{y}_{ap}$, $\bar{y}_{gp}$, $\bar{y}_{hp}$ and $\bar{y}_{RS}$ are the same and also there is considerable reduction in the mean square error of these estimators from the simple ratio estimator (using X_1 or X_2) and mean per unit estimator $\bar{y}$ as the conditions given in (2.9) and (3.4) are satisfied.

(*b*) the data of the first population relate to three characters, area under wheat in 1977 (Y), total cultivated are in 1974 (X_1) and area under wheat in 1976 (X_2) in hectare of six villages. The values of (Y, X_1 & X_2) are (79, 337, 78), (27, 186, 45), (249, 604, 238), (85, 701, 92), (221, 524, 247) and (133, 571, 134).

The second population is an artificial population consists of five units having values of (Y, X_1 & X_2) as (2, 4, 3), (2, 7, 4), (6, 7, 5), (9, 8, 5) and (13, 10, 6).

The exact values of biases and mean square errors of different estimators under, comparison are given in Table 4.2 for sample of size 2.

TABLE 4.2

Eastimator	*Auxiliary Variate Used*	*Population–1*		*Population–2*	
		Bias	*MSE*	*Bias*	*MSE*
1. Mean per unit $\bar{y}$	none	0	2510.222	0	5.610
2. Ratio	X_1	0.321	1709.191	0.189	2.051
3. Ratio	X_2	1.414	82.468	0.184	2.498
4. $\bar{y}_{ap}$	X_1 & X_2	1.423	81.817	0.194	1.875
5. $\bar{y}_{gp}$	X_1 & X_2	1.369	81.393	0.170	1.841
6. $\bar{y}_{hp}$	X_1 & X_2	1.305	80.866	0.143	1.823
7. $\bar{y}_{RS}$	X_1 & X_2	1.672	91.397	0.184	3.243
8. $\bar{y}_{RP}$	X_1 & X_2	3.762	855.868	0.290	1.973

From Table 4.2 it is seen that $\bar{y}_{hp}$ is more efficient than other estimators under comparison for both the populations. In populations 2, $\bar{y}_{hp}$ is the least biased where as in population-1 only ratio estimator using X_1 as an auxilary variable is less biased than $\bar{y}_{hp}$.

From the above illustrations we note that $\bar{y}_{hp}$ is less biased and more efficient than $\bar{y}_{ap}$, $\bar{y}_{gp}$, $\bar{y}_{RS}$ and $\bar{y}_{RP}$. Hence, we may conclude that when more than one auxiliary variables are used for estimating the population parameters, it is better to use $\bar{y}_{hp}$ as an estimate.

REFERENCES

[1] John, S (1969) : On multivariate ratio and product estimators. *Biometrika*, **56**, 533-536.

[2] Mohanty, S. (1967) : Combination of ratio and regression estimates. *Jour. Ind. Stat. Assoc.*, **5**, 1-14.

[3] Olkin, I. (1958) : Multivariate ratio estimation for finite populations. *Biometrika*, **45**, 154-165.

[4] Raj, D. (1965) : On a method of using mutli-auxiliary information in sample surveys. *Jour. Amer. Stat. Assoc.*, 60, 317-319,

[5] Rao, P.S.R.S. and Mudholkar, G.S. (1967) : Generalized multivariate estimator for the mean of a finite population. *Jour. Amer. Stat. Assoc.*, 62, 1009-1012.

[6] Shulka, G.K. (1966) : An alternative multivariate ratio estimate for finite population. *Bull. Cal. Stat. Ass.*, **15**, 127-134.

[7] Singh, M.P. (1967 a) : Multivariate product method of estimation *Jour. Ind. Soc. Agri. Stat.* 9, 1-10.

[8] Singh, M.P. (1967 b) : Ratio-cum-product method of estimation. *Metrika*, **12**, 34-42.

[9] Srivastava, S. K. (1965) : An estimate of the mean of a finite population using several auxiliary variables. *Jour. Ind. Stat. Assoc.*, **3**, 189-194.

[10] Srivastava, S.K. (1966) : On ratio and linear regression method of estimation with several auxiliary variables. *Jour. Ind. Soc. Agrl. Stat.*, **4**, 66-72.