International Journal of Scientific Research in Mathematical and Statistical Sciences
Volume-5, Issue-2, pp.65-69, April (2018)
Research Paper
E-ISSN: 2348-4519

Efficient Hierarchic Multivariate Product-Based Estimator

K.B.Panda[1], P.Das[2*]

[1]Department of Statistics, Utkal University, Bhubaneswar, India
[2]Department of Statistics, Utkal University, Bhubaneswar, India

**Corresponding Author: prabhatadas91@gmail.com*

Available online at: www.isroset.org

Received 19/Mar/2018, Revised 29/Mar/2018, Accepted 17/Apr/2018, Online 30/Apr/2018

Abstract-- We, in this paper, have proposed a multivariate product estimator using multi-auxiliary information. The performance of the proposed multivariate product-based estimator of order k, both under one-phase and two-phase sampling, is compared against the customary multivariate product estimator and the simple mean under conditions which hold good in practice very often. Moreover, the estimator is shown to be more efficient than the competing estimators invariably when k is determined optimally. The superiority of the estimator has been numerically illustrated by considering data from two real populations.

Keywords-- *Hierarchic multivariate product-based estimator, auxiliary information, predictive estimation*

I. INTRODUCTION

Auxiliary variables are extensively used in survey sampling to improve the precision of estimates. Agrawal and Panda(1993) and Panda(1994) have, along the lines of Olkin(1958) utilized multi-auxiliary variables negatively correlated with the study variable to propose the customary multivariate product estimator. Employing the estimator due to Agrawal and Panda(1993) in hierarchic estimation introduced by Agrawal and Sthapit(1997), we have arrived at a new multivariate product estimator of order k.

Let $U = (U_1, U_2, \dots\dots\dots, U_N)$ be the finite population of size N out of which a sample of size n is drawn using simple random sampling without replacement scheme. Let y and x_i $(i = 1,2, \dots\dots, p)$ be, respectively, the study and i-th auxiliary variables having population means $\bar{Y}$ and $\bar{X}_i$ (known), and sample means $\bar{y}$ and $\bar{x}_i$, respectively. The auxiliary variable x_i $(i = 1,2, \dots\dots, p)$ is assumed to be negatively correlated with the study variable y. Let ρ_{oi} and ρ_{ij}, respectively, denote the correlation coefficients between y and x_i and x_i and x_j $(i \neq j = 1, \dots\dots, p)$ and C_0 and $C_i (i = 1, \dots\dots, p)$ be, respectively, the coefficients of variation of y and x_i. Let's further suppose that $C_{0i} = \rho_{0i} C_0 C_i$ and $C_{ij} = \rho_{ij} C_i C_j$.

The conventional multivariate product estimator due to Agrawal and Panda(1993) is given by

$$\bar{y}_{MP} = \bar{y} \sum_{i=1}^{p} w_i \bar{x}_{i,} / \bar{X}_i \qquad (1.1)$$

where $w_i's$ are weights such that $\sum_{i=1}^{p} w_i = 1$, its bias and mean square error , to the first degree of approximation, i.e., to $o(n^{-1})$ have been expressed, respectively, as

$$B(\bar{y}_{MP}) = \theta \bar{Y} \left[\sum_{i=1}^{p} w_i C_{0i}\right] \qquad (1.2)$$

and $M(\bar{y}_{MP}) = \theta \bar{Y}^2 \left(C_0^2 + \sum_{i=1}^{P} w_i{}^2 C_i^2 + \sum \sum_{i \neq j}^{P} w_i w_j C_{ij} + 2 \sum_{i=1}^{P} w_i C_{0i}\right)$ (1.3)

$$= \boldsymbol{w B w}^T,$$

where $\boldsymbol{w} = (w_1, w_2, \dots\dots\dots, w_p)$ is a p-vector, $\boldsymbol{B} = (b_{ij})$, $b_{ij} = \theta \bar{Y}^2 \left[C_0^2 + C_{0i} + C_{0j} + C_{ij}\right]$ (i$\neq j = 1, \dots.., p$) and $\theta = \frac{1}{n} - \frac{1}{N}$. The superscript T refers to transpose. Minimization of the mean square error of $\bar{y}_{MP}$ yields the following optimal weight vector :

$$\boldsymbol{W} = \frac{\boldsymbol{eB^{-1}}}{\boldsymbol{eB^{-1}e^T}}, \qquad (1.4)$$

where $\boldsymbol{e} = (1, 1, \dots\dots, 1)$ and $\boldsymbol{W} = (W_1, W_2, \dots\dots\dots, W_p)$ are p-vectors. In what follows, we shall consider multivariate product estimator $\bar{y}_{MP}$ using optimum weights.

While section 2 deals with the newly proposed multivariate product-based estimator, its bias and mean square error and their comparision with that of the competing estimators have been dealt with in section 3. Performance of the estimator in

two-phase sampling has been discussed in section 4. In section 5, empirical study showing the supremacy of the proposed estimator over its competing estimators has been carried out. Finally, brief conclusion of the present work has been presented in section 6.

II. THE NEWLY PROPOSED MULTIVARIATE PRODUCT-BASED ESTIMATOR

Following the predictive approach of Basu(1971) and Smith(1976), we write the population total as

$$Y = \sum_{l \epsilon s} y_l + \sum_{l \epsilon \bar{s}} y_l \,, \qquad (2.1)$$

where s is the sample of selected units and $\bar{s}$ is its complement. Thus, the first part on the right-hand side of equation (2.1) is known and to estimate Y, we need to predict the second part on the right- hand side of the equation . As a matter of fact, the predictive format for estimation of Y becomes

$$\hat{Y} = \sum_{l \epsilon s} y_l + \sum_{l \epsilon \bar{s}} \hat{y}_l \,, \qquad (2.2)$$

where $\hat{y}_l$ is the implied predictor of $y_l (l \in \bar{s})$. If we use the multivariate product estimator due to Agrawal and Panda(1963) given in (1.1) as an intuitive predictor of y_l $(l \in \bar{s})$, then we arrive at

$$\hat{Y} = \sum_{l \epsilon s} y_l + (N - n)\, \bar{y}_{MP}$$

$$\text{or, } \hat{\bar{Y}} = \bar{y}_{MP}^{(1)}, \qquad (2.3)$$

where $\bar{y}_{MP}^{(1)} = \phi_1 \bar{z}_{MP} + \bar{y}_{MP}$,

with $\phi_1 = 1 + \lambda\, \phi_0$, $\phi_0 = 0$, $\lambda = 1 - \frac{n}{N}$

and $\bar{z}_{MP} = \frac{n}{N}\, \bar{y} \left(1 - \sum_{i=1}^{P} w_i \frac{\bar{x}_i}{\bar{X}_i}\right)$.

Now, making use of $\bar{y}_{MP}^{(1)}$ as an intuitive predictor of y_l $(l \in \bar{s})$ in (2.2), we obtain

$$\hat{\bar{Y}} = \bar{y}_{MP}^{(2)},$$

where $\bar{y}_{MP}^{(2)} = \phi_2 \bar{z}_{MP} + \bar{y}_{MP}$ and $\phi_2 = 1 + \lambda \phi_1$.

Proceeding in this manner, we would, at the kth iteration, reach

$$\bar{y}_{MP}^{(k)} = \phi_k \bar{z}_{MP} + \bar{y}_{MP} \,,$$

where $\phi_k = 1 + \lambda\, \phi_{k-1} = \frac{1 - \lambda^k}{1 - \lambda}$.

With ϕ_k as stated above, $\bar{y}_{MP}^{(k)}$ can be rewritten as

$$\bar{y}_{MP}^{(k)} = (1 - \lambda^k)\bar{y} + \lambda^k \bar{y}_{MP} \qquad (2.4)$$

We have, thus, arrived at the newly proposed multivariate product-based estimator of order k. It may be noted here that when $k = 0$, the proposed estimator is same as the multivariate product estimator $\bar{y}_{MP}$ & when $k \to \infty$, this becomes $\bar{y}$. It is apt to mention here that sampling is carried out from a finite population, i.e., when $N < \infty$, for if we draw samples of fixed sizes from an infinite population, then the proposed estimator $\bar{y}_{MP}^{(k)}$ will be no different from $\bar{y}_{MP}$ as $\lambda = 1$.

III. COMPARISON OF BIAS AND MEAN SQUARE ERROR OF THE PROPOSED ESTIMATOR VIS-À-VIS THE COMPETING ESTIMATOR

The bias of the estimator $\bar{y}_{MP}^{(k)}$, to $o(n^{-1})$, can be found as

$$B(\bar{y}_{MP}^{(k)}) = \lambda^k \theta \bar{Y} \left[\sum_{i=1}^{p} W_i C_{0i}\right] \qquad (3.1)$$

It can easily be examined that the absolute value of the bias obtained above is, for $k \geq 1$, invariably less than that of the customary multivariate product estimator given in (1.2). The mean square error of $\bar{y}_{MP}^{(k)}$, to $o(n^{-1})$, can be worked out as

$$M\left(\bar{y}_{MP}^{(k)}\right) = \theta \bar{Y}^2 \left(C_0^2 + \lambda^{2k} \sum_{i=1}^{P} W_i{}^2 C_i^2 + \lambda^{2k} \sum \sum_{i \neq j=1}^{P} W_i W_j C_{ij} + 2\lambda^k \sum_{i=1}^{P} W_i C_{0i}\right) \qquad (3.2)$$

$$= \boldsymbol{W B W}^T,$$

where $\boldsymbol{W}$ is the p-vector as defined in the foregoing section, $B = (b_{ij})$ and

$$b_{ij} = \theta \bar{Y}^2 \left[C_0^2 + \lambda^k C_{0i} + \lambda^k C_{0j} + \lambda^{2k} C_{ij}\right].$$

When k is determined optimally in order to minimize (3.2) , we get

$$\lambda^k = \frac{-\sum_{i=1}^{p} W_i C_{0i}}{\sum_{i=1}^{p} W_i^2 C_i^2 + \sum \sum_{i \neq j=1}^{p} W_i W_j C_{ij}}. \qquad (3.3)$$

Comparing the minimum mean square error of the multivariate product estimator $\bar{y}_{MP}$(using optimum weights in (1.3)) with that of the proposed multivariate product estimator $\bar{y}_{MP}^{(k)}$ (using optimum k in (3.2)), we find that the estimator $\bar{y}_{MP}^{(k)}$ fares better than the estimator $\bar{y}_{MP}$ if

$$\frac{1}{2}(1+\lambda^k) \geq \frac{-\sum_{i=1}^{p} W_i C_{0i}}{\sum_{i=1}^{p} W_i^2 C_i^2 + \sum\sum_{i\neq j=1}^{p} W_i W_j C_{ij}}, \quad (3.4)$$

and it fares better than $\bar{y}$ if

$$\frac{1}{2}\lambda^k \leq \frac{-\sum_{i=1}^{p} W i C_{0i}}{\sum_{i=1}^{p} W_i^2 C_i^2 + \sum\sum_{i\neq j=1}^{p} W_i W_j C_{ij}}. \quad (3.5)$$

Thus, $\bar{y}_{MP}^{(k)}$ will perform better than both $\bar{y}_{MP}$ and $\bar{y}$ when

$$\frac{1}{2}\lambda^k \leq \frac{-\sum_{i=1}^{p} W_i C_{0i}}{\sum_{i=1}^{p} W_i^2 C_i^2 + \sum\sum_{i\neq j=1}^{p} W_i W_j C_{ij}} \leq \frac{1}{2}(1+\lambda^k), \quad (3.6)$$

a condition which holds good in real-life situations many a time. Under optimality of k, the above condition reduces to

$$\frac{1}{2}\lambda^k \leq \lambda^k \leq \frac{1}{2}(1+\lambda^k), \quad (3.7)$$

which is invariably true as $\lambda < 1$ and $k \geq 1$, indicating the supremacy of the proposed estimator over its competitors. The bounds given in (3.6) are called the efficiency bounds, the term in the middle of (3.6) being treated as a pivotal quantity. By choosing values of the sampling fraction $f(=\frac{n}{N})$ and hence $\lambda(=1-f)$, we have computed the following table which gives the bounds of $\frac{-\sum_{i=1}^{p} W_i C_{0i}}{\sum_{i=1}^{p} W_i^2 C_i^2 + \sum\sum_{i\neq j=1}^{p} W_i W_j C_{ij}}$ within which $\bar{y}_{MP}^{(k)}$ (for various values of k) will be more efficient than $\bar{y}_{MP}$ and $\bar{y}$.

Table 1: Efficiency bounds of $\frac{-\sum_{i=1}^{p} W_i C_{0i}}{\sum_{i=1}^{p} W_i^2 C_i^2 + \sum\sum_{i\neq j=1}^{p} W_i W_j C_{ij}}$ for various values of f and k

	K					
F	1	2	5	8	10	50
0.05	(0.475,0.975)	(0.451,0.951)	(0.387,0.587)	(0.332,0.532)	(0.299,0.799)	(0.038,0.538)
0.10	(0.450,0.950)	(0.405,0.905)	(0.295,0.795)	(0.215,0.715)	(0.174,0.674)	(0.003,0.503)
0.20	(0.400,0.900)	(0.320,0.820)	(0.164,0.664)	(0.084,0.584)	(0.054,0.554)	(0.000,0.500)
0.25	(0.375,0.875)	(0.281,0.781)	(0.118,0.618)	(0.050,0.550)	(0.028,0.528)	(0.000,0.500)
0.30	(0.350,0.850)	(0.245,0.745)	(0.084,0.584)	(0.028,0.528)	(0.014,0.514)	(0.000,0.500)
0.40	(0.300,0.800)	(0.180,0.680)	(0.038,0.538)	(0.008,0.508)	(0.003,0.503)	(0.000,0.500)
0.50	(0.250,0.750)	(0.125,0.625)	(0.016,0.516)	(0.002,0.502)	(0.001,0.501)	(0.000,0.500)
0.60	(0.200,0.700)	(0.080,0.580)	(0.005,0.505)	(0.000,0.500)	(0.000,0.500)	(0.000,0.500)
0.70	(0.150,0.650)	(0.045,0.545)	(0.001,0.501)	(0.000,0.500)	(0.000,0.500)	(0.000,0.500)
0.80	(0.100,0.600)	(0.020,0.520)	(0.000,0.500)	(0.000,0.500)	(0.000,0.500)	(0.000,0.500)

Table 1 may be referred with a view to locating a suitable value of k for given values of the pivotal quantity and f. Knowledge of the pivotal quantity consisting of various population parameters such as the population correlation coefficients and coefficients of variation, as they remain stable over a period of time, can be gathered from past survey, pilot survey, educated guess etc. For a specified value of the pivotal quantity, Table 1 provides more than one value of k which ensures better performance of $\bar{y}_{MP}^{(k)}$ vis-à-vis $\bar{y}_{MR}$ and $\bar{y}$. However the optimal value of k can be arrived at from equation (3.3) provided $\frac{-\sum_{i=1}^{p} W_i C_{0i}}{\sum_{i=1}^{p} W_i^2 C_i^2 + \sum\sum_{i\neq j=1}^{p} W_i W_j C_{ij}} < 1$. When an optimum value of k

is not obtainable, a suitable value of k that renders $\bar{y}_{MP}^{(k)}$ superior to $\bar{y}_{MP}$ and $\bar{y}$ might still be found from the above table.

Here attention must be paid to the fact that if any one of the p-weights becomes 1 and the rest are zero each, then the proposed estimator of order k will be no different from the one due to Agrawal and Sthapit(1997) and its mean square error, under optimality of k, remains same as that of the linear regression estimator.

IV. PERFORMANCE OF THE PROPOSED ESTIMATOR IN TWO –PHASE SAMPLING

There exist cases when $\bar{X}_i$'s $(i = 1,2, \dots \dots, p)$ are unknown. To get rid of such cases, two-phase sampling or double sampling procedure comes into play, wherein we replace $\bar{X}_i$ by $\bar{x}_i^{/}$ $(i = 1,2, \dots \dots, p)$, the sample mean based on large preliminary sample of size n' drawn with simple random sampling without replacement from the population of size N, corresponding to ith auxiliary variable. Thus, the multivariate product estimator due to Agrawal and Panda(1963) and the proposed estimator of order k can be expressed, respectively, as

$$\bar{y}_{MPd} = \bar{y}\sum_{i=1}^{p}\frac{W_i\bar{x}_i}{\bar{x}_i'} \qquad (4.1)$$

and $\bar{y}_{MPd}^{(k)} = (1-\lambda^k)\bar{y} + \lambda^k\bar{y}\sum_{i=1}^{p}\frac{W_i\bar{x}_i}{\bar{x}_i'}$, (4.2)

their biases and mean square errors, to the first degree of approximation, i.e., to $o(n^{-1})$ being expressed, respectively, as

$$B(\bar{y}_{MPd}) = (\theta-\theta')\,\bar{Y}\left[\sum_{i=1}^{p} w_i C_{0i}\right], \qquad (4.3)$$

$$M(\bar{y}_{MPd}) = \theta\bar{Y}^2C_0^2+(\theta-\theta')\bar{Y}^2\left(\sum_{i=1}^{P} {w_i}^2C_i^2 + \sum\sum_{i\neq j}^{P} w_i\,w_jC_{ij} + 2\sum_{i=1}^{P} w_i\,C_{0i}\right) \qquad (4.4)$$

and

$$B(\bar{y}_{MPd}^{(k)}) = \lambda^k(\theta-\theta')\,\bar{Y}\left[\sum_{i=1}^{p} w_iC_{0i}\right], \qquad (4.5)$$

$$M\left(\bar{y}_{MPd}^{(k)}\right) = \theta\bar{Y}^2C_0^2 + (\theta-\theta')\bar{Y}^2\left(\lambda^{2k}\sum_{i=1}^{P} {w_i}^2C_i^2 + \lambda^{2k}\sum\sum_{i\neq j=1}^{P} w_i\,w_jC_{ij} + 2\lambda^k\sum_{i=1}^{P} w_i\,C_{0i}\right), \qquad (4.6)$$

where $\theta' = \frac{1}{n'} - \frac{1}{N}$. For the purpose of comparison of bias and mean square error of the estimators given in (4.1) and (4.2), the weights used in the expressions (4.3), (4.4), (4.5) and (4.6) should be replaced by the optimum weights $W_i's$ $(i = 1,2, \dots \dots \dots, p)$, say, obtained by minimizing the mean square error of $\bar{y}_{MPd}$ given in (4.4).

It can easily be seen that, in two-phase sampling, the performance of the proposed estimator as measured in terms of bias and mean square error remains the same as in the case of one-phase sampling.

V. EMPIRICAL INVESTIGATION

For the purpose of empirical investigation, we have considered two auxiliary variables X_1 and X_2 each being negatively correlated with the study variable Y.

Example 1

We have computed the following population quantities from the information given in Weisberg(1980, p.179), wherein accident rates per million vehicle miles is considered as the study variable (Y) which is negatively correlated with the length of the segment in miles (X_1) and the minor arterial highway (X_2).

N=39 and $C_{ij} = \begin{bmatrix} 0.2616 & -0.1438 & -0.2503 \\ -0.1438 & 0.3582 & 0.1120 \\ -0.2503 & 0.1120 & 2.10 \end{bmatrix}$

$(i,j = 0,1,2)$

Making use of these quantities, we have found the optimum weights W_1 , W_2 and the pivotal quantity given in (3.6) as 0.8421, 0.1579 and 0.4779, respectively. For assessing the performance of the proposed estimator $\bar{y}_{MP}^{(k)}$ over $\bar{y}_{MP}$ and $\bar{y}$, we have prepared the following table:

Table 2: Bias and Mean Square error of Competing Estimators

Estimator	Bias/ $\theta\bar{Y}$	MSE/ $\theta\bar{Y}^2$
$\bar{y}$	0.0000	0.2616
$\bar{y}_{MP}$	−0.1606	0.2764
$\bar{y}_{MP}^{(k)}$	−0.0767	0.1848

From the above table, it is observed that gains in efficiency of the proposed estimator $\bar{y}_{MR}^{(k)}$ with respect to $\bar{y}_{MR}$ and $\bar{y}$ are 49.57% and 41.55%, respectively, implying thereby that there is a substantial increase of gain in efficiency of the proposed estimator over its competing estimators. As regards bias of the proposed estimator, it is also much less

than that of the customary multivariate product estimator due to Agrawal and Panda(1993).

Example 2

Another data set from the same source as of example 1 is incorporated wherein accident rates per million vehicle miles is considered as the study variable (Y) which is negatively correlated with the speed limit(X_1) and the federal aid interstate highway(X_2).The required population quantities are computed as follows

N=39 and

$$C_{ij} = \begin{bmatrix} 0.2616 & -0.0892 & -0.2730 \\ -0.0892 & 0.2003 & 0.4779 \\ -0.2730 & 0.4779 & 7.1768 \end{bmatrix} \quad (i, j = 0, 1, 2)$$

which yields, W_1 , W_2 and the pivotal quantity given in (3.6) as 0.8421, 0.1579 and 0.4779, respectively. The performance of the proposed estimator in terms of bias and mean square error is depicted in the following table:

Table 3: Bias and Mean Square error of Competing Estimators

Estimator	Bias/ $\theta\bar{Y}$	MSE/ $\theta\bar{Y}^2$
$\bar{y}$	0.0000	0.2616
$\bar{y}_{MP}$	−0.0865	0.2809
$\bar{y}_{MP}^{(k)}$	−0.0388	0.2227

It is evident from the above table that gains in efficiency of the proposed estimator against $\bar{y}_{MR}$ and $\bar{y}$ are 26% and 17.46% respectively.

VI. CONCLUSION:

Making use of auxiliary variables each being negatively correlated with the study variable , a multivariate product estimator of order k is suggested which performs better than its compititors under conditions that hold good in practice.The superiority of the estimator has been empirically established. In view of this, the proposed estimator is recommended for use in practice.

REFERENCES

[1] M.C. Agrawal, K.B. Panda- *Multivariate Product Estimators*, Jour. Ind. Soc. Ag. Statistics, 45(3), 359-371, 1993.

[2] M.C. Agrawal, A. B. Sthapit- *Hierarchic predictive ratio-based & product-based estimators and their efficiencies.* Journal of Applied Statistics, Vol. 24, No. 1, 97- 104. 1997.

[3] D. Basu- *An essay on the logical foundations of statistical inference*, Part I, Foundations of Statistical Inference, Ed. By V.P. Godambe and D.A. Sportt, New York, 203-233, 1971.

[4] K.B. Panda- *Some contributions to the Theory of Survey Sampling.* Ph.D. thesis submitted to University of Delhi,Delhi-11007, 1994.

[4] T.M.F. Smith- *The foundations of survey sampling, a review*, Jour. R. Statist. -Soc., Series A, 139, 183-204, 1976.

[5] G.K. Vishwakarma, M. Kumar- *An improved class of chain ratio-product type estimators in two-phase sampling using two auxiliary variables*, Jour. of Prob. and Stat., Vol. 2014

[6] S. Weisberg,- Applied Linear Regression, John Wiley & Sons, Inc., New York, 1980.

