Efficient Hierarchic Predictive Multivariate Product Estimator Based On Harmonic Mean

K.B.Panda[1] and P.Das[2]

Department of Statistics, Utkal University, Bhubaneswar, Odisha, India

***Abstract*−*In this paper, extending the work of hierarchic estimation proposed by Agrawal and Sthapit(1997), we define a multivariate product estimator using harmonic means of multi-auxiliary variables which conforms to predictive character. Furthermore, it has been shown that the proposed multivariate product estimator of order k, when k is determined optimally, fares better than its competitors both in terms of bias and mean square error under some practical conditions. Empirical investigations in support of the theoretical findings have been carried out.*

Keywords−Hierarchic multivariate product estimator, auxiliary information, harmonic mean

I. INTRODUCTION

In the literature of survey sampling, the use of auxiliary information at estimation stage often results in considerable gain in efficiency of the proposed estimators for population parameters under study. Singh(1967) has utilized multi-auxiliary variables negatively correlated with the study variable to propose the customary multivariate product estimator. Adhering to the method of hierarchic estimation introduced by Agrawal and Sthapit(1997) and carried forward by Panda and Sahoo(2015), this paper develops a new multivariate product estimator of order k using harmonic means of multi-auxiliary variables.

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 with simple random sampling without replacement. 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 $\bar{x}_i$. 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 traditional multivariate product estimator due to Agrawal and Panda (1993) is given by

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

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

$$B(\bar{y}_{MPH}) = \theta \bar{Y} \left[\sum_{i=1}^{p} w_i C_i^2 + \sum_{i=1}^{p} w_i C_{0i} \right] \quad (1.2)$$

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

$$= \boldsymbol{wBw}^T,$$

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$\neq j = 1, \dots \dots, p$) and $\theta = \frac{1}{n} - \frac{1}{N}$. The superscript T refers to transpose. Minimization of the mean square error of$\bar{y}_{MPH}$ yields the following optimal weight vector :

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

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

Comparing the minimum MSE of $\bar{y}_{MPH}$ given in (1.1) with the variance of simple mean $\bar{y}$, we find that $\bar{y}_{MPH}$ fares better than $\bar{y}$ if condition

$\frac{1}{2} \leq \frac{-\sum_{i=1}^{p} W_i C_{0i}}{\sum_{i=1}^{p} W_i^2 C_i^2 + \sum\sum_{i \neq j=1}^{p} W_i W_j C_{ij}}$ (1.5) holds.

II. THE NEWLY PROPOSED MULTIVARIATE PRODUCT ESTIMATOR

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

$$Y = \sum_{l \epsilon s} y_l + \sum_{l \epsilon \bar{s}} y_l \,, \tag{2.1}$$

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 have 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

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

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(1993) given in (1.1) as an intuitive predictor of $y_l (l \in \bar{s})$, then we arrive at

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

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

where $\bar{y}_{MPH}^{(1)} = \phi_1 \bar{z}_{MPH} + \bar{y}_{MPH}$, with $\phi_1 = 1 + \lambda \phi_0$, $\phi_0 = 0$, $\lambda = 1 - \frac{n}{N}$

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

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

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

where $\bar{y}_{MPH}^{(2)} = \phi_2 \bar{z}_{MPH} + \bar{y}_{MPH}$ and $\phi_2 = 1 + \lambda \phi_1$. Proceeding in this manner, we would, at the kth iteration, reach

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

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

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

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

We have, thus, arrived at the newly proposed multivariate product estimator of order k. It is worth mentioning that when $k = 0$, the proposed estimator is same as the customary multivariate product estimator $\bar{y}_{MPH}$ &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 $\bar{y}_{MPH}^{(k)}$ will be no different from $\bar{y}_{MPH}$ as $\lambda = 1$. Now, with a view to examining the proposed estimator from the standpoint of predictive character , we use the following expression

$$(1-\lambda^k)\bar{y}+\lambda^k\sum_{i=1}^{p}W_i\frac{\bar{y}}{x_{ij}}\tilde{x}_i$$

as the intuitive predictor of $y_l(l \in \bar{s})$ in (2.2) and can easily conclude that

$$\hat{\hat{Y}} = \bar{y}_{MPH}^{(k)} ,$$

which shows that the proposed multivariate product estimator of order k $(k \geq 1)$ is endowed with the predictive character.

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

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

$$B(\bar{y}_{MPH}^{(k)})= \lambda^k\theta\bar{Y}\left[\sum_{i=1}^{p} w_iC_i^2+\sum_{i=1}^{p} W_iC_{0i}\right]. \tag{3.1}$$

It is evident from (3.1) 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 $\bar{y}_{MPH}^{(k)}$,to $o(n^{-1})$, can be worked out as

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

$$= \boldsymbol{WBW}^T,$$

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

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

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

$$\lambda^k=\frac{-\sum_{i=1}^{p} w_iC_{0i}}{\sum_{i=1}^{p} W_i^2C_i^2+\sum\sum_{i\neq j=1}^{p} W_iW_jC_{ij}}. \tag{3.3}$$

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

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

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

$$\frac{1}{2}\lambda^k \leq \frac{-\sum_{i=1}^{p} WiC_{0i}}{\sum_{i=1}^{p} W_i^2C_i^2+\sum\sum_{i\neq j=1}^{p} W_iW_jC_{ij}}. \tag{3.5}$$

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

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

a condition which holds good in real-life situations quite often. Under optimality of k,i.e., when (3.3) holds, the above condition reduces to

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

which is invariably true as $\lambda < 1$ and $k \geq 1$, revealing 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 $\bar{y}_{MPH}^{(k)}$ (for various values of k) will be more efficient than $\bar{y}_{MPH}$ 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 can serve as an aid to locate a suitable value of k for given values of the pivotal quantityand 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 $\bar{y}_{MPH}^{(k)}$ vis-à-vis $\bar{y}_{MPH}$ 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 $\bar{y}_{MPH}^{(k)}$ superior to $\bar{y}_{MPH}$ and $\bar{y}$ might still be found from the above table.

Here attention is drawn 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 Panda(1993) and its mean square error, under optimality of k, remains same as that of the linear regression estimator.

IV. EMPIRICAL INVESTIGATIONS

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$Y$.

Example1

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 speed limit(X_1) and the federal and interstate highway (X_2). Here X_1 and X_2 are being considered as the auxiliary variables:

N=39 and $$C_{ij} = \begin{bmatrix} 0.2616 & -0.0892 & -0.2730 \\ -0.0892 & 0.2003 & 0.4779 \\ -0.2730 & 0.4779 & 7.1768 \end{bmatrix} (i, j = 0, 1, 2)$$

Making use of these quantities, we have found the optimum weights W_1 , W_2 and the pivotal quantity given in (3.6) as 1.0146, -0.0146 and 0.4493, respectively. For assessing the performance of the proposed estimator $\bar{y}_{MPH}^{(k)}$ over $\bar{y}_{MPH}$ and $\bar{y}$, we have prepared the following table:

Table 1: Bias and Mean Square error of Competing Estimators

Estimator	$\lvert$ Bias/ $\theta\bar{Y}\rvert$	MSE/ $\theta\bar{Y}^2$
$\bar{y}$	0.0000	0.2616
$\bar{y}_{MPH}$	0.0119	0.2809
$\bar{y}_{MPH}^{(k)}$	0.0054	0.2227

From the above table, it is observed thatgain in efficiency of the proposed estimator $\bar{y}_{MPH}^{(k)}$ with respect to $\bar{y}_{MPH}$ and $\bar{y}$ are 26% and 17.46%, respectively, implying thereby that there is an appreciable 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 based on harmonic mean.

Example2

We consider a hypothetical population of size 25 with

$$C_{ij} = \begin{bmatrix} 0.0682 & -0.0116 & -0.0100 \\ -0.0116 & 0.0266 & 0.0360 \\ -0.0100 & 0.0360 & 0.1687 \end{bmatrix} (i, j = 0, 1, 2)$$

These quantities yield the optimum weights W_1 , W_2 and the pivotal quantity given in (3.6) as 1.0892, -0.0892 and 0.2964, respectively.

Table 2: Bias and Mean Square error of Competing Estimators

Estimator	$\lvert$ Bias/ $\theta\bar{Y}\rvert$	MSE/ $\theta\bar{Y}^2$
$\bar{y}$	0.0000	0.0682
$\bar{y}_{MPH}$	0.0021	0.0706
$\bar{y}_{MPH}^{(k)}$	0.0006	0.0635

From the above table, it is seen thatgain in efficiency of the proposed estimator $\bar{y}_{MPH}^{(k)}$ with respect to $\bar{y}_{MPH}$ and $\bar{y}$ are 11.18% and 7.40%, respectively.

V. CONCLUSION

The proposed multivariate product estimator of order k, as evident from the theoretical findings coupled with numerical illustrations, is superior to the one due to Agrawal and Panda(1993) and the simple mean under condition which holds good in practice very often. It is clearly found from both the numerical illustrations that while the estimator due to Agrawal and Panda is found to less efficient than the simple mean, our newly proposed estimator performs better.

REFERENCES

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

[2]Agrawal, M.C. & Panda, K.B.(1993): Multivariate product estimators, Jour. Ind. Soc. Ag. Statistics, 45(3).359-371.

[3]Basu, D.(1971): 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.

[4]Panda, K.B. and Sahoo, N.(2015): Systems of exponential ratio-based and exponential product-based estimators with their efficiency. ISOR Journal of Mathematics, Volume 11, Issue 3 Ver.I(May-June),PP 73-77.

[5]Weisberg, S. (1980): Applied linear regression analysis, John Wiley & Sons, Inc., New York.