Asian Journal of Statistical Sciences
Vol. 5, No-1, (2025), pp. 13-29

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

ESTIMATING POPULATION MEAN THROUGH HIERARCHIC PREDICTIVE WEIGHTED DIFFERENCE ESTIMATOR

K. B. Panda[a] and P. P. Mohanty [b]

[a] *Dept. of Statistics, Central University of Jharkhand, Ranchi.*
[b] *Dept. of Statistics, Utkal University, Bhubaneswar, Odisha.*

ARTICLE HISTORY
Compiled September 24, 2025

Received 18 August 2024; Accepted 22 August 2025

ABSTRACT
For estimating the finite population mean, we have proposed a new difference estimator as a weighted estimator of the customary difference estimator and the difference estimator due to "Sahoo et. al. (2007)" when the preassigned constants used therein are the corresponding population regression coefficients. The proposed difference estimator is unbiased. Employing predictive approach advocated by "Basu (1971)", the estimator is found to be endowed with the predictive character. Using the hierarchic estimation technique due to "Agrawal and Sthapit (1997)", a sequence of estimators is generated from the proposed estimator. The estimator of order k, when k is optimum, is found to be more efficient than its competing estimators under conditions that hold good in practice. While simulation study reveals supremacy of the proposed difference estimator over existing difference estimators from the standpoint of efficiency, empirical investigation based on three real populations has been carried out to demonstrate the gain in efficiency of the suggested estimator over the competing estimators.

KEYWORDS
Hierarchic estimation; Predictive estimation; Weighted difference estimator; Simulation study

1. INTRODUCTION

In survey sampling literature, use of auxiliary information at estimation stage has been contributing significantly to the improvement of precision of the estimates under study. "Royall (1970)", in the context of superpopulation set-up and "Basu (1971)" in the context of fixed population set-up were concerned with the prediction of the non-surveyed part of the population. "Basu (1971)" has suggested "predictive approach" for examining plausibility and face-validity of an estimator. This approach deals with the estimation of population total or population mean (i.e., surveyed and nonsurveyed components). "Royall (1970)" dealt with an optimal predictor of the unknown component under a specified superpopulation model. From the predictive point of view, "Agrawal and Jain (1988)" have examined ratio, ratio-type and regression estimators in double sampling procedures. "Agrawal and Kulldorff (1987)"

CONTACT: K. B. Panda Email: kbpanda@cuj.ac.in

have exploited the predictive aspect as well as the precision of ratiotype estimators. As the customary product estimator due to "Murthy (1964)" lacks in the predictive character, "Agrawal and Jain (1989)" came up with a product estimator using harmonic mean, which is predictive in form. Along the lines of hierarchic estimation introduced by "Agrawal and Sthapit (1997)", "Panda and Sahoo (2015)" have developed a system of ratio-based and product-based estimators. While "Panda and Das (2018)" have carried forward the study of hierarchic estimation to multivariate product estimator based on harmonic mean, "Panda and Parida (2021)" have considered estimators of order k , using variable transformation.

The linear regression estimator, which is of primary importance in survey sampling literature, is originated from the difference estimator given by

$$\bar{y}_d = \bar{y} + \beta_1(\bar{X} - \bar{x}), \tag{1}$$

where $\bar{y}$ is the sample mean of y-variable and $\bar{X}$ and $\bar{x}$ are, respectively the population mean and the sample mean of x-variable and β_1 is a preassigned constant. The estimator $\bar{y}_d$ is unbiased for the population mean $\bar{Y}$ with minimum variance

$$V(\bar{y}_d) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left(1 - \rho_{yx}^2\right), \tag{2}$$

when β_1 is equal to the population regression coefficient of y on x, and S_y^2 and ρ_{yx} being the population mean square of y and the population correlation coefficient between y and x, respectively.
Following "Agrawal and Jain (1989)", "Sahoo et. al. (2007)" have introduced a new difference estimator

$$\bar{y}_d^* = \bar{y} + \beta_2(\bar{Z} - \bar{z}), \tag{3}$$

where $z = x^{-1}(x > 0)$ and β_2 is a preassigned constant. The estimator $\bar{y}_d^*$ is unbiased for the population mean $\bar{Y}$ and its minimum variance is

$$V(\bar{y}_d^*) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left(1 - \rho_{yz}^2\right). \tag{4}$$

Here β_2, as before, is no different from the population regression coefficient of y on $z = x^{-1}(x > 0)$ and ρ_{yz} is the population correlation coefficient between y and z.

2. PROPOSED ESTIMATOR

The above two difference estimators are combined through weights, α and $1 - \alpha$, to form the following weighted difference estimator:

$$\begin{aligned}
\bar{y}'_d &= \alpha\bar{y}_d + (1-\alpha)\bar{y}^*_d \\
&= \alpha\left\{\bar{y} + \beta_1(\bar{X} - \bar{x})\right\} + (1-\alpha)\left\{\bar{y} + \beta_2(\bar{Z} - \bar{z})\right\} \\
\Rightarrow \bar{y}'_d &= \bar{y} + \alpha\beta_1(\bar{X} - \bar{x}) + (1-\alpha)\beta_2(\bar{Z} - \bar{z}).
\end{aligned} \tag{5}$$

Here, the preassigned constats β_1 and β_2 have been taken as the population regression coefficient of y on x and the population regression coefficient of y on z, respectively.

The estimator is unbiased since

$$\begin{aligned}
E\left(\bar{y}'_d\right) &= E\left(\bar{y} + \alpha\beta_1(\bar{X} - \bar{x}) + (1-\alpha)\beta_2(\bar{Z} - \bar{z})\right) \\
&= \bar{Y}.
\end{aligned}$$

For the purpose of examining the predictive character of the proposed estimator we invoke, "Basu (1971)" in expressing the population total Y as

$$Y = \sum_{i\in S} y_i + \sum_{i\in\bar{S}} y_i, \tag{6}$$

where s stands for the observed segment and $\bar{s}$ is its complement, i.e., the unobserved segment. To estimate the population total Y, we have to predict the second component of the right-hand side of equation (6), which is unknown, the first component being known.
The usual predictive format for estimating Y, the population total, is

$$\widehat{Y} = \sum_{i\in s} y_i + \sum_{i\in\bar{s}} \hat{y}_i, \tag{7}$$

where $\hat{y}_i$ is the implied predictor of $y_i, (i \in \bar{s})$.

Thus,

$$\begin{aligned}
\hat{\bar{Y}} &= \frac{1}{N}\sum_{i\in S} y_i + \frac{1}{N}\sum_{i\in\bar{s}} \hat{y}_i \\
&= \frac{n}{N}\bar{y} + \frac{1}{N}\sum_{i\in\bar{s}}\left[\bar{y} + \alpha\beta_1\left(x_i - \bar{x}\right) + (1-\alpha)\beta_2\left(z_i - \bar{z}\right)\right] \\
&= \bar{y} + \frac{1}{N}\left[\alpha\beta_1\{(N\bar{X} - n\bar{x}) - (N-n)\bar{x}\} + (1-\alpha)\beta_2\{(N\bar{Z} - n\bar{z}) - (N-n)\bar{z}\}\right] \\
&= \bar{y} + \alpha\beta_1(\bar{X} - \bar{x}) + (1-\alpha)\beta_2(\bar{Z} - \bar{z}) \\
\Rightarrow \hat{\bar{Y}} &= \bar{y}'_d.
\end{aligned}$$

So, $\bar{y}'_d$ is predictive in form. It is apt to mention here that the two existing estimators viz. $\bar{y}_d$ and $\bar{y}^*_d$ conform to predictive character too. For the details of the proof, the reader may see the Appendix-1.

The expression for variance of the proposed estimator can be obtained as

$$V\left(\bar{y}_d'\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2\left[1+\left\{\alpha^2\rho_{yx}^2+(1-\alpha)^2\rho_{yz}^2+2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{xz}\right\}\right.$$
$$\left.-2\left\{\alpha\rho_{yx}^2+(1-\alpha)\rho_{yz}^2\right\}\right]. \tag{8}$$

On simplification, we get

$$V\left(\bar{y}_d'\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2\left[\left(1-\rho_{yz}^2\right)+\alpha^2\left(\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}\right)\right.$$
$$\left.-2\alpha\left(\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}\right)\right]. \tag{9}$$

With a view to minimizing $V\left(\bar{y}_d'\right)$ subject to variation in α, we proceed as follows:

$$\frac{\partial V\left(\bar{y}_d'\right)}{\partial\alpha}=0$$

$$\Rightarrow\alpha_{opt}=\frac{\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}}{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}. \tag{10}$$

Replacing α by α_{opt} in (9), we obtain

$$V\left(\bar{y}_d'\right)_{\min}=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2\left[\left(1-\rho_{yz}^2\right)-\left\{\frac{\left(\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}\right\}\right]. \tag{11}$$

The details of derivation of the variance of the proposed estimator are given in Appendix-2.

3. EFFICIENCY COMPARISON

We know that, in simple random sampling without replacement, the mean per unit estimator $\bar{y}$ is unbiased for the population mean $\bar{Y}$ and its variance is given by

$$V(\bar{y})=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2.$$

The proposed estimator performs better than

(I) the mean per unit estimator if and only if

$$
\begin{aligned}
& V\left(\bar{y}_d'\right)_{\min} \leq V(\bar{y}) \\
& \Rightarrow \left(\frac{1}{n}-\frac{1}{N}\right) S_y^2\left[\left(1-\rho_{yz}^2\right)-\left\{\frac{\left(\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}\right\}\right] \leq \left(\frac{1}{n}-\frac{1}{N}\right) S_y^2 \\
& \Rightarrow \left[\frac{\left(\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}\right] \geq -\rho_{yz}^2 \\
& \Rightarrow 2\rho_{yx}\rho_{yz}\rho_{xz} \leq \rho_{yx}^2+\frac{\rho_{yz}^2\left(\rho_{yz}^2+\rho_{yx}^2\rho_{xz}^2\right)}{\rho_{yx}^2+\rho_{yz}^2}. \qquad (12)
\end{aligned}
$$

(II) the usual difference estimator if and only if

$$
\begin{aligned}
& V\left(\bar{y}_d'\right)_{\min} \leq V\left(\bar{y}_d\right) \\
& \Rightarrow \left(\frac{1}{n}-\frac{1}{N}\right) S_y^2\left[\left(1-\rho_{yz}^2\right)-\left\{\frac{\left(\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}\right)\right] \leq \left(\frac{1}{n}-\frac{1}{N}\right) S_y^2\left(1-\rho_{yx}^2\right) \\
& \Rightarrow \left[\left(1-\rho_{yz}^2\right)-\left\{\frac{\left(\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}\right\}\right] \leq \left(1-\rho_{yx}^2\right) \\
& \Rightarrow 2\rho_{yx}\rho_{yz}\rho_{xz} \leq \rho_{yz}^2+\rho_{yx}^2\rho_{xz}^2. \qquad (13)
\end{aligned}
$$

(III) the difference estimator due to “Sahoo et. al. (2007)” if and only if

$$
\begin{aligned}
& V\left(\bar{y}_d'\right)_{\min} \leq V\left(\bar{y}_d^*\right) \\
& \Rightarrow \left(\frac{1}{n}-\frac{1}{N}\right) S_y^2\left[\left(1-\rho_{yz}^2\right)-\left\{\frac{\left(\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}\right\}\right] \leq \left(\frac{1}{n}-\frac{1}{N}\right) S_y^2\left(1-\rho_{yz}^2\right) \\
& \Rightarrow \left[\frac{\left(\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}\right] \geq 0 \\
& \Rightarrow 2\rho_{yx}\rho_{yz}\rho_{xz} \geq \rho_{yz}^2+\rho_{yx}^2\left[1-\left(\rho_{yx}-\rho_{yz}\rho_{xz}\right)^2\right]. \qquad (14)
\end{aligned}
$$

The conditions (12), (13) and (14) hold good in practice, a fact that can be evidenced from the simulation study and empirical illustrations undertaken in Section 6.

4. A SEQUENCE OF PREDICTIVE DIFFERENCE ESTIMATORS AND THE RELATED PERFORMANCE

Using the proposed estimator $\bar{y}_d'$ as an intuitive predictor of $y_i, (i \in \bar{s})$, we reach

$$\begin{aligned}\hat{Y} =& \sum_{i\in s} y_i + \sum_{i\in \bar{S}} \bar{y}'_d \\ &\Rightarrow \hat{\bar{Y}} = \bar{y}_d'^{(1)},\ \text{say,}\end{aligned} \tag{15}$$

where

$$\bar{y}_d'^{(1)} = \emptyset_1 \bar{z}'_d + \bar{y}'_d, \tag{16}$$

with $\emptyset_1 = 1 + \lambda\emptyset_0, \emptyset_0 = 0, \lambda = 1 - \frac{n}{N}$ and $\bar{z}'_d = -\frac{n}{N}\left[\alpha\beta_1(\bar{X} - \bar{x}) + (1-\alpha)\beta_2(\bar{Z} - \bar{z})\right]$.

In second iteration, $\bar{y}_d'^{(1)}$ is used as an intuitive predictor of $y_i, (i \in \bar{s})$ with a view to arriving at $\bar{y}_d'^{(2)}$ given by

$\bar{y}_d'^{(2)} = \emptyset_2 \bar{z}'_d + \bar{y}'_d$ with $\emptyset_2 = 1 + \lambda\emptyset_1$.

Proceeding exactly in the similar manner, we would, at kth iteration, find

$$\bar{y}_d'^{(k)} = \emptyset_k \bar{z}'_d + \bar{y}'_d, \tag{17}$$

where, $\emptyset_k = 1 + \lambda\emptyset_{k-1}$

$$\begin{aligned}&=1 + \lambda\left(1 + \lambda\emptyset_{k-2}\right)\\ &=1 + \lambda + \lambda^2\left(1 + \lambda\emptyset_{k-3}\right)\\ &=\left(1 + \lambda + \lambda^2 + \cdots + \lambda^{k-1}\right) - \lambda^k\left(1 + \lambda + \lambda^2 + \cdots + \lambda^{k-1}\right)\\ &=\frac{\left(1-\lambda^k\right)}{(1-\lambda)}.\end{aligned}$$

Putting the value of $\emptyset_k$ in equation (17), we get

$$\begin{aligned}\bar{y}_d'^{(k)} &= \frac{\left(1-\lambda^k\right)}{(1-\lambda)}\bar{z}'_d + \bar{y}'_d\\ &= \frac{\left(1-\lambda^k\right)}{(1-\lambda)}\frac{n}{N}\left(\bar{y} - \bar{y}'_d\right) + \bar{y}'_d\\ &\Rightarrow \bar{y}_d'^{(k)} = \left(1-\lambda^k\right)\bar{y} + \lambda^k\bar{y}'_d,\end{aligned} \tag{18}$$

which can be easily shown to be unbiased for $\bar{Y}$ and predictive in form.

Particular Case

I. When $k = 0$, we have $\lambda^k \to 1$

$$\Rightarrow \bar{y}_d^{\prime(k)} = \bar{y}_d^{\prime}.$$

II. When $k \to \infty$, we have $\lambda^k \to 0$

$$\Rightarrow \bar{y}_d^{\prime(k)} = \bar{y}.$$

The variance of the estimator, up to $0\left(\frac{1}{n}\right)$, can be obtained as

$$V\left(\bar{y}_d^{\prime(k)}\right) = E\left[\bar{y}_d^{\prime(k)} - \bar{Y}\right]^2$$

$$\begin{aligned}\Rightarrow V\left(\bar{y}_d^{\prime(k)}\right) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \Big\{ & 1 + \lambda^{2k}\left(\alpha^2\rho_{yx}^2 + \rho_{yz}^2(1-\alpha)^2 + 2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}\right) \\ & -2\lambda^k\left(\alpha\rho_{yx}^2 + (1-\alpha)\rho_{yz}^2\right)\Big\}. \end{aligned} \tag{19}$$

Optimum k can be found by minimizing $V\left(\bar{y}_d^{\prime(k)}\right)$ subject to variation in k, i.e.,

$$\frac{\partial V\left(\bar{y}_d^{\prime}(k)\right)}{\partial k} = 0$$

$$\Rightarrow \lambda^{\mathrm{k}} = \frac{\left[\alpha\rho_{yx}^2 + (1-\alpha)\rho_{yz}^2\right]}{\left[\alpha^2\rho_{yx}^2 + \rho_{yz}^2(1-\alpha)^2 + 2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}\right]}. \tag{20}$$

Putting λ^{k} in equation (19), we have

$$V\left(\bar{y}_d^{\prime(k)}\right)_{\min} = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left[1 - \frac{\left[\alpha{\rho_{yx}}^2 + (1-\alpha){\rho_{yz}}^2\right]^2}{\left[\alpha^2{\rho_{yx}}^2 + {\rho_{yz}}^2(1-\alpha)^2 + 2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}\right]}\right]. \tag{21}$$

5. COMPARISON OF EFFICIENCY OF $\bar{y}_d^{(k)}$ VIS-À-VIS $\bar{y}_d^{\prime}$ AND $\bar{y}$

From equation (19) and (8), $\bar{y}_d^{\prime(k)}$ will be more efficient than $\bar{y}_d^{\prime}$, if and only if

$$\frac{\left[\alpha\rho_{yx}^2 + (1-\alpha)\rho_{yz}^2\right]}{\left[\alpha^2\rho_{yx}^2 + \rho_{yz}^2(1-\alpha)^2 + 2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}\right]} < \frac{1+\lambda^{\mathrm{k}}}{2}. \tag{22}$$

Between $\bar{y}_d'^{(k)}$ and $\bar{\mathrm{y}}$, the former is more efficient than the later, if and only if

$$\frac{\left[\alpha\rho_{yx}^2 + (1-\alpha)\rho_{yz}^2\right]}{\left[\alpha^2\rho_{yx}^2 + \rho_{yz}^2(1-\alpha)^2 + 2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}\right]} < \frac{\lambda^{\mathrm{k}}}{2}. \tag{23}$$

Thus, combining equation (21) and (22) we get that $\bar{y}_{d1}^{(k)}$ will be more efficient than $\bar{y}_d'$ & $\bar{y}$, if and only if

$$\frac{\lambda^{\mathrm{k}}}{2} < \frac{\left[\alpha\rho_{yx}^2 + (1-\alpha)\rho_{yz}^2\right]}{\left[\alpha^2\rho_{yx}^2 + \rho_{yz}^2(1-\alpha)^2 + 2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}\right]} < \frac{1+\lambda^{\mathrm{k}}}{2}. \tag{24}$$

The bounds on $\frac{[\alpha\rho_{yx}{}^2+(1-\alpha)\rho_{yz}{}^2]}{[\alpha^2\rho_{yx}{}^2+\rho_{yz}{}^2(1-\alpha)^2+2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}]}$ given in equation (24) are called the efficiency bounds. By choosing values of the sampling fraction $f\left(=\frac{n}{N}\right)$ and hence $\lambda(=1-\mathrm{f})$, we have prepared and presented in the Appendix-4 a Table which presents the bounds on $\frac{[\alpha\rho_{yx}{}^2+(1-\alpha)\rho_{yz}{}^2]}{[\alpha^2\rho_{yx}{}^2+\rho_{yz}{}^2(1-\alpha)^2+2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}]}$ as per equation (24).

In order to compute the percentage gain in efficiency of the competing estimators with respect to $\bar{y}$, when k is optimally determined, we use the following formula:

$$G = \left[\frac{V(\bar{y})}{V(.)} - 1\right] \times 100,$$

where $V(.)$ denotes the variance of the competing estimator.

6. NUMERICAL ILLUSTRATION

(A) Simulation Study

For the purpose of simulation study, first we generate a finite population from a bivariate normal distribution (size N), ensuring $x > 0$ and creating a derived variable $z = 1/x$. In the second step, we draw a sample of size n simple random sampling without replacement. Using these values, we compute the variances of the five estimators and, based thereon, the percentage gain in efficiency of the competing estimators. The process is repeated 1000 times and the results are presented below:
$N = 100, n = 10, \bar{Y} = 100, \bar{X} = 50, \sigma_X = 10, \sigma_Y = 20, \rho_{yx} = 0.90$

Table 1. *Percentage Gain in Efficiency of the Competing Estimators with respect to* $\overline{y}$

Sl. No.	Estimator	Variance	Percentage gain in efficiency
1	$\bar{y}$	29.758	000.000
2	$\bar{y}_d$	06.071	391.154
3	$\bar{y}_d^*$	08.586	246.595
4	$\bar{y}_d'$	06.001	395.895
5	$\bar{y}_d'^{(k)}$	05.993	396.506

Here, y and x are generated data set. The details of the data set for conducting simulation study are given in Appendix-3.

The simulation study demonstrates that the conventional sample mean exhibits the highest variance 29.758 among all the competing estimators. In contrast, estimators that exploit auxiliary information show substantially lower variances which are presented in the Table 1 above. In terms of relative efficiency, the gain over the usual estimator is remarkable. While the simple mean is taken as the baseline (0% efficiency gain), the auxiliary-based estimators achieve efficiency gains of about 246% - 396%. Notably, the estimator, $\bar{y}_d'^{(k)}$, yields the minimum variance with the highest efficiency gain, making it the most efficient estimator among all the considered estimators.

(B) Examples Based on Natural Populations

We consider three different natural populations from various sources detailed below:

Population. 1

Gujarati [1995](PP.92)

y: Telephone ownership in Singapore, 1960-1981,

x: Per Capita GDP in Singapore, 1960-1981,

$N = 22, n = 7, \bar{Y} = 116.8182, \bar{X} = 2812.045, \bar{Z} = 0.0004416, S_y^2 = 7829.299, S_z^2 = 4.125384e - 08, S_x^2 = 1750400, \rho_{yx} = 0.973, \rho_{yz} = -0.850, \rho_{xz} = -0.943.$

Table 2. *Percentage Gain in Efficiency of the Competing Estimators with respect to $\overline{y}$*

Sl. No.	Estimator	Variance	Percentage gain in efficiency
1	$\bar{y}$	762.594	0.000
2	$\bar{y}_d$	41.147	1753.336
3	$\bar{y}_d^*$	211.264	260.967
4	$\bar{y}_d'$	18.445	4034.396
5	$\bar{y}_d^{\prime(k)}$	13.451	5569.509

Population. 2

Maddala [2012](PP.294)

y: Imports in France (Millions of New France at 1959 Prices),

x: Consumption in France (Millions of New France at 1959 Prices),

$N = 18, n = 4, \bar{Y} = 30.07778, \bar{X} = 167.3778, \bar{Z} = 0.006329841, S_y^2 = 155.783, S_x^2 = 1728.984$, $S_z^2 = 2.414329e-06, \rho_{yx} = 0.985, \rho_{yz} = -0.927, \rho_{xz} = -0.975$.

Table 3. *Percentage Gain in Efficiency of the Competing Estimators with respect to $\overline{y}$*

Sl. No.	Estimator	Variance	Percentage gain in efficiency
1	$\bar{y}$	30.291	0000.000
2	$\bar{y}_d$	0.916	3207.004
3	$\bar{y}_d^*$	4.282	00607.454
4	$\bar{y}_d'$	0.338	08863.932
5	$\bar{y}_d^{\prime(k)}$	0.269	11169.690

Population. 3

Murthy [1964](PP.399)

y : Total area under wheat in1964 for 34 sample villages

x: Total cultivated area in 1961 for 34 sample villages

$N = 34, n = 6, \bar{Y} = 199.4412, \bar{X} = 747.5882, \bar{Z} = 0.002161105, S_y^2 = 22564.56, S_x^2 = 197095.3$, $S_z^2 = 3.941534e-06, \rho_{yx} = 0.904, \rho_{yz} = -0.600, \rho_{xz} = -0.719$.

Table 4. *Percentage Gain in Efficiency of the Competing Estimators with respect to $\bar{y}$*

Sl. No.	Estimator	Variance	Percentage gain in efficiency
1	$\bar{y}$	3097.096	000.000
2	$\bar{y}_d$	0564.628	448.519
3	$\bar{y}_d^*$	1982.105	056.253
4	$\bar{y}_d'$	0557.400	455.632
5	$\bar{y}_d'^{(k)}$	0553.646	459.400

7. Conclusion

Besides being unbiased and predictive in character, the proposed weighted difference estimator is more efficient than the simple mean and the existing difference estimators. Based on the proposed difference estimator, a sequence of difference estimators is proposed. Under optimality of k, the difference estimator has been found to perform better than the simple mean, the customary difference estimator, the difference estimator due to Sahoo et. al. and the proposed weighted difference estimator. Simulation study based on a random population consisting of 100 units and empirical investigation based on real populations reveal that the proposed estimator of order k is more efficient than any other competing estimator.

Acknowledgement

The authors are highly grateful to the reviewer for his valuable suggestions leading to substantial improvement over the original manuscript.

References

[1] Agrawal, M.C. and Kulldorff, G. (1987). Precision of Ratio-type Estimators and a Predictive Counter- Part, Statistical Research Report No. 1987-3, Institute of Mathematical Statistics, University of Umea(Sweden).

[2] Agrawal, M.C. and Jain, N. (1988). Predictive Estimation in Double Sampling Procedures, The American Statistician , Vol. 42, No. 3, pp. 184-186.

[3] Agrawal, M.C. and Jain, N (1989). A new predictive product estimator, Biometrika, 76, 822-823.

[4] Agrawal, M.C. and Sthapit, A.B. (1997). Hierarchic Predictive Ratio-based and Product-based Estimators and their Efficiencies. Journal of Applied Statistics, 24(1), 97-104.

[5] 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.

[6] Gujarati, D. (1995). Basic Econometrics, Mc-Graw Hill, India, International Editions.

[7] Maddala, G.S. and Lahiri, K. (2012). Introduction to Econometrics, Wiley India(P) Ltd, Fourth Edition.

[8] Murthy, M.N. (1964). Product Method of Estimation. Sankhya, Series A, Vol. 26, 69-74.

[9] Panda, K.B. and Sahoo, N. (2015). Systems of Exponential Ratio-based and Exponential Product-based Estimators with their Efficiencies. ISOR Journal of Mathematics, Vol 11, Issue 3, PP 73-77.

[10] Panda, K.B., Das, P. (2018). Efficient Hierarchic Predictive Multivariate Product Estimator Based on Harmonic Mean. International Journal of Mathematics Trends and Technology, Vol 56, Issue-6, pp.437-442.
[11] Panda, K.B. and Parida, S.S. (2021). Efficient Hierarchic Ratio-Based and Product-Based Estimators of Population Mean Using Variable Transformation. International Journal of Agricultural and Statistical Sciences, Volume. 17, No. 2, pp.741-745.
[12] Panda, K.B. and Mohanty, P.P. (2022). Efficient Hierarchic Predictive Weighted Linear Regression Estimator with Efficiency. Asian Journal of Statistical Sciences. Vol. 02, No. 2, pp.141-149.
[13] Royall, R.M. (1970). On finite population sampling under certain linear regression models, Biometrika, 57, 377-387.
[14] Sahoo, L. N., Dalabehra, M., Mangaraj A. K. (2007). A regression estimator using harmonic mean of the auxiliary variable, The Philippine Statistician, Vol. 56, Nos. 3-4, pp. 31-36.
[15] Smith, T.M.F. (1976). The foundations in survey sampling, a review, Journal of the Royal Statistical Society, Series A, 139, pp.183-204.
[16] Sukhatme, P. V., Sukhatme, B. V., Sukhatmai, S. Asok. C (1984). Sampling Theory of Surveys with Applications, 3rd revised edition. Ames: Iowa State University Press.

APPENDIX-1

The usual predictive format for estimating Y, the population total, is

$$\hat{Y} = \sum_{i\in S} y_i + \sum_{i\in \bar{S}} \hat{y}_i,$$

where $\hat{y}_i$ is the implied predictor of $y_i, (i \in \bar{s})$.

Thus, for usual difference estimator

$$\begin{aligned}
\hat{\bar{Y}} &= \frac{1}{N}\sum_{i\in s} y_i + \frac{1}{N}\sum_{i\in \bar{s}} \hat{y}_i \\
&= \frac{n}{N}\bar{y} + \frac{1}{N}\sum_{i\in \bar{s}} \left[\bar{y} + \beta_1\left(\overline{x_s} - \bar{x}\right)\right] \\
&= \frac{n}{N}\bar{y} + \frac{N-n}{N}\bar{y} + \frac{N-n}{N}\beta_1\left(\frac{(N\bar{X} - n\bar{x})}{(N-n)} - \bar{x}\right) \\
&= \bar{y} + \frac{1}{N}\beta_1(N\bar{X} - N\bar{x}) \\
&= \bar{y} + \beta_1(\bar{X} - \bar{x}) \\
\Rightarrow \hat{Y} &= \bar{y}_d.
\end{aligned}$$

So, $\bar{y}_d$ is predictive in form.

For the estimator due to Sahoo et. al.,

$$
\begin{aligned}
\hat{\hat{Y}} &= \frac{1}{N}\sum_{i\in s} y_i + \frac{1}{N}\sum_{i\in \bar{s}} \hat{y}_i \\
&= \frac{n}{N}\bar{y} + \frac{1}{N}\sum_{i\in \bar{s}} \left[\bar{y} + \beta_2\left(\overline{z_s} - \bar{z}\right)\right] \quad \left[\text{ where } z = x^{-1}(x>0)\right] \\
&= \frac{n}{N}\bar{y} + \frac{N-n}{N}\bar{y} + \frac{N-n}{N}\beta_2\left(\frac{(N\bar{Z} - n\bar{Z})}{(N-n)} - \bar{x}\right) \\
&= \bar{y} + \frac{1}{N}\beta_2(N\bar{Z} - N\bar{z}) \\
&= \bar{y} + \beta_2(\bar{Z} - \bar{Z}) \\
\Rightarrow \hat{\hat{Y}} &= \bar{y}_d^*.
\end{aligned}
$$

So, $\bar{y}_d^*$ is also predictive in form.

APPENDIX-2

$$
\begin{aligned}
V\left(\bar{y}_d'\right) &= E\left(\bar{y}_d' - \bar{Y}\right)^2 \\
&= E\left(\bar{y} + \alpha\beta_1(\bar{X} - \bar{x}) + (1-\alpha)\beta_2(\bar{Z} - \bar{z}) - \bar{Y}\right)^2 \\
&= E\left((\bar{y} - \bar{Y}) - \alpha\beta_1(\bar{x} - \bar{X}) - (1-\alpha)\beta_2(\bar{z} - \bar{Z})\right)^2 \\
&= E(\bar{y} - \bar{Y})^2 + \alpha^2\beta_1^2 E(\bar{x} - \bar{X})^2 + (1-\alpha)^2\beta_2^2 E(\bar{z} - \bar{Z})^2 - 2\alpha\beta_1 E\{(\bar{y} - \bar{Y})(\bar{x} - \bar{X})\} \\
&+ 2\alpha(1-\alpha)\beta_2\beta_1 E\{(\bar{z} - \bar{Z})(\bar{x} - \bar{X})\} - 2(1-\alpha)\beta_2 E\{(\bar{y} - \bar{Y})(\bar{z} - \bar{Z})\} \\
&= V(\bar{y}) + \alpha^2\beta_1^2 V(\bar{x}) + (1-\alpha)^2\beta_2^2 V(\bar{z}) - 2\alpha\beta_1 \operatorname{Cov}(\bar{y},\bar{x}) + 2\alpha(1-\alpha)\beta_2\beta_1 \operatorname{Cov}(\bar{z},\bar{x}) \\
&- 2(1-\alpha)\beta_2 \operatorname{Cov}(\bar{y},\bar{z}) \\
&= \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2 + \alpha^2\beta_1^2 S_x^2 + (1-\alpha)^2\beta_2^2 S_z^2 - 2\alpha\beta_1 S_{yx} + 2\alpha(1-\alpha)\beta_2\beta_1 S_{zx}\right. \\
&\left.-2(1-\alpha)\beta_2 S_{yz}\right] \\
&= \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2 + \alpha^2\frac{S_{yx}^2}{S_x^4}S_x^2 + (1-\alpha)^2\frac{S_{yz}^2}{S_z^4}S_z^2 - 2\alpha\frac{S_{yx}}{S_x^2}S_{yx} + 2\alpha(1-\alpha)\frac{S_{yx}S_{yx}}{S_x^2}\frac{S_{yz}^2}{S_{zx}}\right. \\
&\left.-2(1-\alpha)\frac{S_z^2}{S_z^2}S_{yz}\right] \\
&= \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2 + \alpha^2\rho_{yx}^2 S_y^2 + (1-\alpha)^2\rho_{yz}^2 S_y^2 - 2\alpha\rho_{yx}^2 S_y^2 + 2\alpha(1-\alpha)S_{yx}S_{yz}S_{zx}S_y^2\right. \\
&\left.-2(1-\alpha)\rho_{yz}^2 S_y^2\right] \\
&= \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2\left[1 + \left\{\alpha^2\rho_{yx}^2 + (1-\alpha)^2\rho_{yz}^2 + 2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{xz}\right\}\right. \\
&\left.-2\left\{\alpha\rho_{yx}^2 + (1-\alpha)\rho_{yz}^2\right\}\right]
\end{aligned}
$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left[1 + \alpha^2 \rho_{yx}^2 + \alpha^2 \rho_{yz}^2 + \rho_{yz}^2 - 2\alpha\rho_{yz} + 2\alpha\rho_{yx}\rho_{yz}\rho_{xz} - 2\alpha^2 \rho_{yx}\rho_{yz}\rho_{xz} - 2\alpha\rho_{yx}^2 \right.$$
$$\left. -2\rho_{yz}^2 + 2\alpha\rho_{yz}^2\right]$$
$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left[1 + \alpha^2 \left(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}\right) - 2\alpha \left(\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}\right) - \rho_{yz}^2\right]$$
$$\Rightarrow V\left(\bar{y}_d'\right) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left[\left(1 - \rho_{yz}^2\right) + \alpha^2 \left(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}\right)\right.$$
$$\left. -2\alpha \left(\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}\right)\right]. \tag{$*$}$$

With a view to minimizing $V\left(\bar{y}_d'\right)$ subject to variation in α, we proceed as follows:

$$\frac{\partial V\left(\bar{y}_d'\right)}{\partial \alpha} = 0$$
$$\Rightarrow 2\alpha \left(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}\right) = 2\left(\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}\right)$$
$$\Rightarrow \alpha_{opt} = \frac{\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}}{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}.$$

Replacing α by α_{opt} in (*), we obtain

$$V\left(\bar{y}_d'\right)_{\min} = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left[\left(1 - \rho_{yz}^2\right) + \left(\frac{\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}}{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}\right)^2 \left(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}\right)\right.$$
$$\left. -2\left(\frac{\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}}{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}\right)\left(\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}\right)\right]$$
$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left[\left(1 - \rho_{yz}^2\right) + \left\{\frac{\left(\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}} - 2\frac{\left(\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}\right\}\right]$$
$$\Rightarrow V\left(\bar{y}_d'\right)_{\min} = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left[\left(1 - \rho_{yz}^2\right) - \left\{\frac{\left(\rho_{yx}^2 - \rho_{yx}\rho_{yz}\rho_{xz}\right)^2}{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}\right\}\right].$$

APPENDIX-3

SL. NO	Y	X	Z
1	87.63646	47.31681	0.02113414
2	95.84435	46.9299	0.02130838
3	130.64946	65.43456	0.01528244
4	100.82123	51.90492	0.019266
5	100.97767	54.62064	0.0183081
6	134.10438	66.1706	0.01511245
7	107.86691	57.11812	0.01750758
8	71.98568	44.17498	0.02263725
9	85.67264	44.95267	0.02224562
10	92.66241	42.54323	0.0235055
11	123.42837	63.49004	0.0157505
12	108.19236	51.17764	0.01953978
13	105.26865	59.54492	0.01679404
14	102.11247	51.23248	0.01951887
15	89.79521	42.94825	0.02328383
16	136.11793	65.59938	0.01524405
17	110.09937	54.26803	0.01842706
18	59.73017	33.9413	0.02946263
19	112.54982	59.59494	0.01677995
20	88.85682	49.26072	0.02030015
21	78.9167	39.61048	0.02524584
22	94.06393	51.36429	0.01946878
23	78.7282	42.1827	0.0237064
24	85.04342	44,11495	0.02266805
25	90.64154	37.55101	0.02663044
26	65.28996	36.59455	0.02732647
27	117.09201	56.9758	0.01755131
28	103.18761	51.15197	0.01954959
29	75.70158	42.82574	0.0233499
30	124.86783	61.9594	0.0161396
31	110.92632	48.79863	0.02049238
32	94.87807	45.62597	0.02191734
33	117.9085	58.20753	0.01717991
34	116.79069	59.71249	0.01674691
35	112.92358	65.03473	0.0153764
36	115.62503	52.36949	0.01909509
37	108.58495	60.41035	0.01555345
38	100.00943	46.76762	0.02138232
39	97.11013	40.29534	0.02481676
40	89.99139	51.62833	0.01936921
41	87.33415	40.99777	0.02439157
42	95.4158	48.99982	0.02040824
43	72.13996	43.82819	0.02281637
44	140.68076	75.67955	0.01321361
45	121.38287	67.0209	0.01492072
46	76.72519	41.42104	0.02414232
47	89.51469	51.48319	0.01942381

48	91.85575	43.17624	0.02316089
49	119.07284	49.7466	0.02010188
50	96.17599	53.83887	0.01857394
51	106.37205	49.53865	0.02018626
52	100.72326	46.97469	0.02128806
53	99.7038	48.40809	0.0206577
54	125.58096	66.39249	0.01506194
55	95.29983	48.32109	0.0206949
56	129.75106	65.1612	0.01534656
57	70.08041	33.52369	0.02982965
58	111.02515	56.79309	0.01760778
59	104.10988	47.65133	0.02098577
60	103.67418	53.35945	0.01874082
61	109.33475	49.76722	0.02009355
62	88.22618	49.07492	0.02037701
63	91.24205	51.41037	0.01945133
64	85.14623	28.86538	0.03464357
65	77.93954	41.49092	0.02410166
66	106.55021	51.7634	0.01931867
67	110,00209	51.90014	0.01926777
68	100.24347	52.23027	0.01914599
69	119.24859	56.75462	0.01761971
70	141.47669	67.81266	0.01474651
71	89.8522	46.18918	0.02165009
72	54.0896	28.2111	0.03544704
73	119.98642	59.51016	0.01680385
74	89.44244	35,74448	0.02797635
75	85.04285	46.23732	0.02162755
76	118.59738	63.50483	0.01574683
77	94.38817	47.20626	0.02118363
78	76.19364	37.49154	0.02667268
79	104.34671	50.12638	0.01994958
80	96.46183	50.34787	0.01986181
81	98.32823	53.86836	0.01856377
82	109.80082	49.06483	0.0203812
83	92.02551	47.79445	0.02092293
84	111.3877	59.11978	0.01691481
85	95.20885	48.78952	0.02049621
86	106.28088	53.80432	0.01858587
87	123.72415	56.25561	0.01777569
88	108.81524	53.75817	0.01860183
89	94.7714	44.25255	0.02259757
90	122.056	62.51456	0.01599627
91	120.16015	58.50445	0.01709272
92	110.38361	56.28372	0.01776713
93	104.91668	51.88911	0.01927187
94	85.98185	47.35089	0.02111893
95	124.91438	67.4032	0.01483609

84	59.11978	111.3877	0.01691481
85	48.78952	95.20885	0.02049621
96	91.39305	37.23218	0.02685849
97	144.60134	68.26243	0.01464935
98	128.44141	68.79476	0.01453599
99	94.27575	49.99237	0.02000305
100	77.55226	44.67176	0.02238551

APPENDIX-4

Efficiency bounds of $\frac{[\alpha\rho_{yx}{}^2+(1-\alpha)\rho_{yz}{}^2]}{[\alpha^2\rho_{yx}{}^2+\rho_{yz}{}^2(1-\alpha)^2+2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}]}$ for various values of f and k.

	k					
f	1	2	5	10	15	20
0.05	(0.475, 0.975)	(0.451, 0.951)	(0.387, 0.887)	(0.299, 0.799)	(0.232, 0.732)	(0.179, 0.679)
0.10	(0.450, 0.950)	(0.405, 0.905)	(0.295, 0.795)	(0.174, 0.674)	(0.103, 0.603)	(0.061, 0.561)
0.15	(0.425, 0.925)	(0.361, 0.861)	(0.221, 0.721)	(0.098, 0.598)	(0.044, 0.544)	(0.019, 0.519)
0.20	(0.400, 0.900)	(0.320, 0.820)	(0.164, 0.664)	(0.054, 0.554)	(0.017, 0.517)	(0.005, 0.505)
0.25	(0.375, 0.875)	(0.281, 0.781)	(0.118, 0.618)	(0.028, 0.528)	(0.006, 0.506)	(0.001, 0.501)
0.30	(0.350, 0.850)	(0.245, 0.745)	(0.084, 0.584)	(0.014, 0.514)	(0.002, 0.502)	(0.000, 0.500)
0.40	(0.300, 0.800)	(0.180, 0.680)	(0.039, 0.539)	(0.003, 0.503)	(0.000, 0.500)	(0.000, 0.500)
0.50	(0.250, 0.750)	(0.125, 0.625)	(0.015, 0.515)	(0.000, 0.500)	(0.000, 0.500)	(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)
0.75	(0.125, 0.625)	(0.031, 0.531)	(0.000, 0.500)	(0.000, 0.500)	(0.000, 0.500)	(0.000, 0.500)

The above Table will be of help in the determination of optimum value of k for given values of the pivotal quantity $\frac{[\alpha\rho_{yx}{}^2+(1-\alpha)\rho_{yz}{}^2]}{[\alpha^2\rho_{yx}{}^2+\rho_{yz}{}^2(1-\alpha)^2+2\alpha(1-\alpha)\rho_{yx}\rho_{yz}\rho_{zx}]}$ and f. It may be noted here that the maximum of all values of k for which the value of the pivotal quantity is found to lie within the efficiency bounds for a given f is its optimum value. However, the Table can be extended further by considering other values of f and k depending on one's need.