JOURNAL OF STATISTICS, OPTIMIZATION AND DATA SCIENCE

VOL: 1, ISSUE: 2, (DECEMBER 2023), PP:36-48 , ISSN: 2583-9462

A Note on Variance Estimation Using Multi-Auxiliary Information

Reena[a] and Vyas Dubey[b]

[ab]*School of Studies in Statistics, Pt. Ravishankar Shukla University, Raipur, Chhattisgarh-492010, India*

ARTICLE HISTORY
Compiled January 15, 2024

ABSTRACT
In this paper we explore the problem of estimation of finite population variance in simple random sampling without replacement by utilizing information of multi-auxiliary variables. We propose an almost unbiased multivariate estimator that has a smaller mean squared error than the conventional biased multivariate estimators. In addition, we support these theoretical result with the aid of a numerical investigation and simulation study into the performance of the estimator has been made.

KEYWORDS
Auxiliary information, Bias, Mean Squared Error, Simple random sampling without replacement (SRSWOR), Relative efficiency, simulation technique.

1. Introduction

In sample surveys the information on auxiliary character X is used to achieve higher precision in the estimates of some population parameters such as the mean or the variance of the study variable. It is well established that when the auxiliary information is to be used at the estimation stage, the ratio, product and regression methods of estimation are widely used in many situations. When correlation between study variable Y and auxiliary variable X are positive, ratio method of estimation is used. If correlation between Y and X are negative, product method of estimation is preferred. Further if the correlation between Y and X are linearly related, regression method of estimation is used.
Let Y_i and X_i be the measurement in respect of the study variable Y and the auxiliary variable X respectively, on the ith unit of the population of size N from which a random sample of size n is drawn. Further let s_y^2 and s_x^2 be unbiased estimators of population variance S_y^2 and S_x^2 of variables Y and X. Now assume that the problem is to estimate the population variance

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2,$$

CONTACT Reena[a]. Email: sonkarreena011@gmail.com

it is assumed that

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{X})^2,$$

is known. Assume that population size N is large so that the finite population correction terms are ignored. The usual unbiased estimator of S_y^2 is defined by

$$t_0 = s_y^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \overline{y})^2, \tag{1.1}$$

where $\overline{y} = n^{-1}\sum y_i$ is the sample mean of study variable y. For increasing efficiency of usual estimators Das and Tripathi (1978) suggested class of estimators

$$t_{\alpha_1} = s_y^2\left(\frac{S_x^2}{s_x^2}\right)^{\theta} \tag{1.2}$$

and

$$t_{\alpha_2} = s_y^2[S_x^2/\{S_x^2 + \theta(s_x^2 - S_x^2)\}] \tag{1.3}$$

where $\theta = \lambda_{22}^*/\beta_2^*(x)$ being suitable chosen scalar. Upto the terms of order n^{-1} both estimators are biased and is equally efficient as t_{re}. Utilizing single auxiliary variable for estimation S_y^2 Isaki (1983) suggested a ratio and regression estimators for S_y^2 as

$$t_r = s_y^2\left(\frac{S_x^2}{s_x^2}\right) \tag{1.4}$$

and

$$t_{re} = s_y^2 + B_{yx}(S_x^2 - s_x^2) \tag{1.5}$$

where $s_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ and $B_{yx} = (S_y^2\lambda_{yx}^*/S_x^2\beta_2^*(x))$ the usual regression coefficient of S_y^2 and S_x^2. In general difference estimators is known to be more precise than ratio and product estimators.

Utilizing p-auxiliary Information a random sample without replacement of size n is selected $(Y_i, X_{1i}), ..., X_{pi},\ i = 1, 2, ..., n,$ from the population is observed. Isaki (1983) generalized multivariate ratio and regression estimator

$$t_{\delta r} = \sum_{i=1}^{p}\delta_i\hat{r}_i S_{x_i}^2, \sum\delta_i = 1 \tag{1.6}$$

and

$$t_{bre} = s_y^2 + \sum_{i=1}^{p}B_{yx_i}(S_{x_i}^2 - s_{x_i}^2) \tag{1.7}$$

where δ_i are suitable chosen constant, $\hat{r}_i = (s_y^2/s_{x_i}^2)$, $B_{yx_i} = (S_y^2\lambda^*_{yx_i}/S_{x_I}^2\beta_2^*(x_i))$ the usual regression coefficient of S_y^2 and $S_{x_i}^2$, $s_{x_i}^2 = (n-1)^{-1}\sum_{i=1}^{p}(x_i-\bar{x})^2$ are the sample variance of Y and X. In general regression estimator is known to be more precise than ratio estimator. Further improving efficiency of estimators Arcos Ceberian et al. (1997) proposed multivariate estimator

$$t_\alpha = s_y^2\prod_{i=1}^{p}\left(\frac{S_{x_i}}{s_{x_i}}\right)^{\theta_i} \tag{1.8}$$

where $\theta_i = \lambda^*_{yx_i}/\beta_2^*(x_i)$ are suitable constants, respectively, up to the terms of order n^{-1} the estimator is biased and is equally efficient as t_{bre}.

It can be see that all the above multivariate estimators are biased. Therefore in section 2 we define an estimator which is unbiased up to first order approximation and is more efficient than all the above estimators under certain conditions.

2. Proposed Estimator

Motivated by Das and Tripathi (1978), we assume that above value of θ_i is known, then the proposed estimator of S_y^2 using multi-auxiliary information X_i, $i = 1, 2, ..., p$

$$t = s_y^2\sum_{i=1}^{p}W_i\left[\frac{S_{x_i}^2}{S_{x_i}^2+\theta(s_{x_i}^2-S_{x_i}^2)}\right] \tag{2.1}$$

where, W_i are suitable chosen constant so that $\sum_{i=1}^{p}W_i = 1$.
In order to obtain approximations for Bias and MSE of estimators we considered

Let $s_y^2 = S_y^2 + e$ and $s_{x_i}^2 = S_{x_i}^2 + e_i$

$E(e_i) = 0, \forall\ i = 0, 1, 2, ...p.$

$E(e_0^2) = fS_y^4\beta_2^*(y)$, $E(e_i^2) = fS_{x_i}^4\beta_2^*(x_i)$,

and $E(e_0e_i) = fS_y^2S_{x_i}^2\lambda^*_{yx_i}$, $E(e_ie_j) = fS_{x_i}^2S_{x_j}^2\lambda^*_{x_ix_j}$,$\forall\ i = 1, 2, ...p.$

We followed by Biradar and Singh (1998), we have

$\beta_2^*(y) = (A\beta_2(y) - M)$, $\beta_2^*(x_i) = (A\beta_2(x_i) - M)$,

$\lambda^*_{yx_i} = (A\lambda_{yx_i} + 2B\rho^2_{yx_i} - D)$, $\lambda^*_{x_i,x_j} = (A\lambda_{x_i,x_j} + 2B\rho^2_{x_i,x_j} - D)$,

where, $\beta_2(y) = \frac{\mu_{40}(y,y)}{\mu_{20}^2(y,y)}$, $\beta_2(x_i) = \frac{\mu_{04}(x_i,x_i)}{\mu_{02}^2(x_i,x_i)}$,

$\lambda_{yx_i} = \frac{\mu_{22}(y,x_i)}{\mu_{20}(y,x_i)\mu_{02}(y,x_i)}$, $\lambda_{x_i,x_j} = \frac{\mu_{22}(x_i,x_j)}{\mu_{20}(x_i,x_j)\mu_{02}(y,x_i)}$

$\mu_{pqr} = \frac{1}{N}\sum(y_i - \overline{Y})^p(x_i - \overline{X})^q(z_i - \overline{Z})^r$ and for pq and $r = 0$ to 4.

(here p, q and r are being non-negative integers)

$f = \frac{N-n}{n(N-1)}$, $A = \frac{(N-1)(nN-N-n-1)}{N(n-1)(N-3)}$, $M = \frac{N^2n-3N^2+6N-2n-3}{N(n-1)(N-3)}$

$B = \frac{(N-1)(N-n-1}{N(n-1)(N-3)}$, $D = \frac{(N^2n-2Nn-N^2+2N-n-1)}{N(n-1)(N-3)}$

Up to order n^{-1} it can easily be seen that the proposed estimator is unbiased i.e.

$$B(t) = 0 \tag{2.2}$$

The MSE of t

$$M(t) = \sum_{i,j=1}^{p} W_i W_j Cov(t_i, t_j) \tag{2.3}$$

where $i \neq j = 1, 2, ...p$;

From equation (2.3) co-variances as follows

$$Cov(t_i, t_j) = f S_y^4 [\beta_2^*(y) - \theta_i \lambda_{yx_i}^* - \theta_j \lambda_{yx_j}^* + \theta_i \theta_j \lambda_{x_i xj}^*]$$

From (2.3)

$$M(t) = S_y^4 \sum_{i,j}^{p} W_i W_j \left(\frac{a_{ij}}{n} \right) \tag{2.4}$$

where

$$a_{ij} = \frac{(N-n)}{(N-1)} [\beta_2^*(y) - \theta_i \lambda_{yx_i}^* - \theta_j \lambda_{yx_j}^* + \theta_i \theta_j \lambda_{x_i xj}^*]$$

Now we find the optimum value $W_i, ..., W_p$, we define the vectors $a = (a_1, ...a_p)$ and $W'(1 \times p) = (W_i, ..., W_p)$, the matrix $A = |a_{ij}|_{p \times p}$.
Then (2.4) can be written as

$$M(t) = n^{-1} S_y^4 W A W' \tag{2.5}$$

where n is the sample size and $W'$is the transpose of W.

For determination of optimum weights, we follow the technique used by Olkin (1958) it can easily be established that i.e.

$\hat{W} = \frac{eA^{-1}}{eA^{-1}e'}$

where, $e'(1 \times p) = (1, ..., 1)$ and A^-1 is the matrix inverse of A. Assuming that the weights will be uniform if and only if the column sums of A are equal i.e. $Ae = eD$, where $D \neq 0$ is a scalar.

Hence $eA^{-1} = e/D$ and $eA^{-1}e' = p/D$

so that $\hat{W} = e/p$

From equation (2.5)

$$M(t) = n^{-1} S_y^4 \left(\frac{D}{p} \right) \tag{2.6}$$

From equation (2.6)

$$M(t) = \frac{(N-n)}{n(N-1)p} S_y^4 \left[\beta_2^*(y) - \theta_i \lambda_{yx_i}^* - \theta_j \lambda_{yx_j}^* + \theta_i \theta_j \lambda_{x_i xj}^* \right] \tag{2.7}$$

After putting θ_i and θ_j values for obtain minimum $M(t)$

$$M(t) = \frac{f S_y^4}{p} [\beta_2^*(y) - \frac{\lambda_{yx_i}^{*2}}{\beta_2^*(x_i)} - \frac{\lambda_{yx_j}^{*2}}{\beta_2^*(x_j)} + \frac{\lambda_{yx_i}^* \lambda_{yx_j}^* \lambda_{x_i xj}^*}{\beta_2^*(x_i) \beta_2^*(x_j)}] \tag{2.8}$$

Above we can seen that it is difficult to comparisons with all the existing estimators if multi-auxiliary information are available. Then we use for comparison and it is seen that the reduction in MSE of the suggested estimator is high as compared to all the existing estimators.

3. Special case

For sample units selection, If $p = 2$ then equation (2.1) we have,

$$t_1 = s_y^2 [(1 - W_1) \psi_1 + W_1 \psi_2] \tag{3.1}$$

where $\psi_1 = \left(\frac{S_{x_1}^2}{S_{x_1}^2 + \theta(s_{x_1}^2 - S_{x_1}^2)} \right)$
and $\psi_2 = \left(\frac{S_{x_2}^2}{S_{x_2}^2 + \theta(s_{x_2}^2 - S_{x_2}^2)} \right)$

From (3.1) for Bias expression we have,

$$E(t_1 - \overline{Y}) = [(1 - W_1) B(\psi_1) + W_1 B(\psi_2)] \tag{3.2}$$

where, $B(\psi_1) = f S_y^2 [\theta_1^2 \beta_2^*(x_1) - \theta_1 \lambda_{220}^*]$

and $B(\psi_2) = f S_y^2 [\theta_2^2 \beta_2^*(x_2) - \theta_2 \lambda_{202}^*]$

Again from (3.1) for MSE expression we have,

$$E(t_1 - \overline{Y})^2 = [(1 - W_1)^2 M(\psi_1) + W_1^2 M(\psi_2) + W_1(1 - W_1) Cov(\psi_1, \psi_2)] \tag{3.3}$$

where, $M(\psi_1) = f S_y^4 [\beta_2^*(y) + \theta_1^2 \beta_2^*(x_1) - 2\theta_1 \lambda_{220}^*]$

and $M(\psi_2) = f S_y^4 [\beta_2^*(y) + \theta_2^2 \beta_2^*(x_2) - 2\theta_2 \lambda_{202}^*]$

Using the optimal value θ_1 and θ_2 to minimize the Bias and MSE of can easily be shown as:

$\theta_1 = \lambda^*_{220}/\beta^*_2(x_1)$ and $\theta_2 = \lambda^*_{202}/\beta^*_2(x_2)$

From (3.2) the Bias of $B(t_1)$

$$B(t_1) = 0 \tag{3.4}$$

From (3.3) the $M(t_1)$

$$M(t_1) = fS_y^4\left[\left\{\beta^*_2(y) - \frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)}\right\} + W_1^2\left\{\frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)} + \frac{\lambda^{*2}_{202}}{\beta^*_2(x_2)} - \frac{2\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\}\right.$$

$$\left. - 2W_1\left\{\frac{\lambda^{*2}_{202}}{\beta^*_2(x_2)} - \frac{2\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\}\right] \tag{3.5}$$

For getting optimum value of W_1, we differentiate the equation (3.5) with respect to W_1

i.e. $\frac{dM(t_1)}{dW_1} = 0$

we have,

$$W_1\left\{\frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)} + \frac{\lambda^{*2}_{202}}{\beta^*_2(x_2)} - \frac{2\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\} - \left\{\frac{\lambda^{*2}_{202}}{\beta^*_2(x_2)} - \frac{2\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\} = 0 \tag{3.6}$$

The optimum value of W_1, which is minimizes $M(t_1)$can easily be found as:

$$W_1 = \frac{\left\{\frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)} - \frac{\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\}}{\left\{\frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)} + \frac{\lambda^{*2}_{202}}{\beta^*_2(x_2)} - \frac{2\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\}}$$

From (3.6) multiplying by W_1 and substituting with (3.5), we have,

$$M(t_1) = fS_y^4\left[\left\{\beta^*_2(y) - \frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)}\right\} - W_1\left\{\frac{\lambda^{*2}_{202}}{\beta^*_2(x_2)} - \frac{2\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\}\right]$$

After putting optimum value of W_1 to minimize the $M(t_1)$

$$M(t_1) = fS_y^4\left[\left\{\beta^*_2(y) - \frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)}\right\} - \frac{\left\{\frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)} - \frac{\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\}^2}{\left\{\frac{\lambda^{*2}_{220}}{\beta^*_2(x_1)} + \frac{\lambda^{*2}_{202}}{\beta^*_2(x_2)} - \frac{2\lambda^*_{220}\lambda^*_{202}\lambda^*_{022}}{\beta^*_2(x_1)\beta^*_2(x_2)}\right\}}\right] \tag{3.7}$$

We consider if $p = 2$, forgetting the expectation of bias and mean square error of all the existing estimators is considered up to terms of order n^{-1}, we have

$$B(t_{\delta r}) = fS_y^2[(1-\delta_1)\{\beta_2^*(x_1) - \lambda_{220}^{*2}\} + \delta_1\{\beta_2^*(x_1) - \lambda_{220}^{*2}\}] \tag{3.8}$$

$$B(t_\alpha) = fS_y^2\left[\frac{1}{2}\left\{\lambda_{220}^* + \frac{\lambda_{220}^2}{\beta_2^*(x1)} + \lambda_{202}^* + \frac{\lambda_{202}^2}{\beta_2^*(x2)}\right\} - \frac{\lambda_{220}^2}{\beta_2^*(x1)} - \frac{\lambda_{202}^2}{\beta_2^*(x2)} + \frac{\lambda_{220}^*\lambda_{202}^*\lambda_{022}^*}{\beta_2^*(x_1)\beta_2^*(x_2)}\right] \tag{3.9}$$

and minimum MSE of $t_{\delta r}$, t_{bre} and t_α can be shown to be

$$M(t_{\delta r}) = fS_y^4[\{\beta_2^*(y) + \beta_2^*(x_1) - 2\lambda_{220}^*\} - \delta_1\{\beta_2^*(x_1) - \lambda_{220}^* + \lambda_{202}^* - \lambda_{022}^*\}] \tag{3.10}$$

where the optimum value of δ_1

$$\delta 1 = \frac{\{\beta_2^*(x_1) - \lambda_{220}^* + \lambda_{202}^* - \lambda_{022}^*\}}{\{\beta_2^*(x_1) + \beta_2^*(x_1) - 2\lambda_{022}^*\}}$$

$$M(t_{bre}) = f[S_y^4\beta_2^*(y) + S_{x1}^4B_{yx_1}^2\beta_2^*(x_1) + S_{x2}^4B_{yx_2}^2\beta_2^*(x_2) - 2S_y^2S_{x_1}^2B_{yx_1}\lambda_{220}^* - 2S_y^2S_{x_2}^2B_{yx_2}\lambda_{202}^*$$

$$+2S_{x_1}^2S_{x_2}^2B_{yx_1}B_{yx_2}\lambda_{022}^*] \tag{3.11}$$

where the optimum value of B_{yx_1} and B_{yx_2}

$B_{yx_1} = \frac{S_y^2\lambda_{220}^*}{S_{x_1}^2\beta_2^*(x_1)}$ and $B_{yx_2} = \frac{S_y^2\lambda_{202}^*}{S_{x_2}^2\beta_2^*(x_2)}$

$$M(t_\theta) = \overline{Y}^2 f[\beta_2^*(y) + \theta_1^2\beta_2^*(x_1) + \theta_2^2\beta_2^*(x_2) - 2\theta_1\lambda_{220}^* - 2\theta_2\lambda_{202}^* + 2\theta_1\theta_2\lambda_{022}^*] \tag{3.12}$$

From $eq^n s$. (3.10), (3.11) and (3.12) the minimum MSE can be shown to be

$$M(t_{\delta r}) = fS_y^4\left[\{\beta_2^*(y) + \beta_2^*(x_1) - 2\lambda_{220}^*\} - \frac{\{\beta_2^*(x_1) - \lambda_{220}^* + \lambda_{202}^* - \lambda_{022}^*\}}{\{\beta_2^*(x_1) + \beta_2^*(x_1) - 2\lambda_{022}^*\}}\right] \tag{3.13}$$

$$M(t_{bre}) = fS_y^4\left[\{\beta_2^*(y) - \frac{\lambda_{220}^{*2}}{\beta_2^*(x1)} - \frac{\lambda_{202}^{*2}}{\beta_2^*(x2)}\} + \frac{2\lambda_{220}^*\lambda_{202}^*\lambda_{022}^*}{\beta_2^*(x_1)\beta_2^*(x_2)}\right] \tag{3.14}$$

It can be seen that

$$M(t_{bre})_{min}. = M(t_{\alpha})_{min}. \tag{3.15}$$

4. Efficiency Comparison

For efficiency comparison we assume

$\beta_2^*(x1) = \beta_2^*(x2)$ and $\lambda_{220}^* = \lambda_{202}^*$ then

Comparing equation (3.7) and (3.13), we have

$M(t_1) < M(t_{\delta r}) \quad if$

$$\beta_2^*(x_1) - 4\lambda_{220}^* + \lambda_{022}^* - \frac{\lambda_{220}^{*2}}{\beta_2^*(x_1)}\left(\frac{\lambda_{022}^*}{\beta_2^*(x_1)} - 3\right) > 0 \tag{4.1}$$

Comparing equation (3.7) and (3.14), we have
$M(t_1) < M(t_{bre}) \quad if$

$$\left\{\frac{3\lambda_{022}^*}{\beta_2^*(x_1)} - 1\right\} > 0 \tag{4.2}$$

Further comparing equation (3.13) and (3.14), we have

$M(t_{bre}) < M(t_{\delta r}) \quad if$

$$\beta_2^*(x_1) - 4\lambda_{220}^* + \lambda_{022}^* + \frac{4\lambda_{220}^{*2}}{\beta_2^*(x_1)}\left(1 - \frac{\lambda_{022}^*}{\beta_2^*(x_1)}\right) > 0 \tag{4.3}$$

Above we can seen that it is difficult to comparison, then we following bivariate symmetric populations by Sukhatme (1954), we have

$\beta_2(y) = \beta_2(x_i) = 3$

$\lambda_{220} = \lambda_{202} = (1 + 2\rho_{yx_i}^2)$, for all $i = 1, 2$

and $\lambda_{022} = (1 + 2\rho_{x_1x_2}^2)$

where $\rho_{ij} = S_{ij}/S_iS_j$, $(i \neq j = y, x_1, x_2)$

If we assuming that N is large, so that f.p.c. (n/N) factor can be ignored than we have

$f = \frac{1}{n}$, $A = 1$, $B = \frac{1}{(n-1)}$, $M = \frac{n-3}{(n-1)}$

In this case we find values
$\beta_2^*(y) = \beta_2^*(x_i) = \frac{2n}{(n-1)}$

$\lambda^*_{220} = \lambda^*_{202} = \rho^2_{yx_i} \frac{2n}{(n-1)}$

and $\lambda^*_{022} = \rho^2_{x_i x_j} \frac{2n}{(n-1)}$

Then we consider minimum MSE attained by the above estimators we have,

$$M(t_1) = \frac{2S_y^4}{(n-1)}\left[(1-\rho^4_{yx_1}) - \frac{(\rho^4_{yx_2} - \rho^2_{yx_1}\rho^2_{yx_2}\rho^2_{x_1x_2})^2}{(\rho^4_{yx_1} + \rho^4_{yx_2} - 2\rho^2_{yx_1}\rho^2_{yx_2}\rho^2_{x_1x_2})}\right] \tag{4.4}$$

$$M(t_{\delta r}) = \frac{2S_y^4}{(n-1)}\left[2(1-\rho^2_{yx_1}) - \frac{(1-\rho^2_{yx_1} + \rho^2_{yx_2} - \rho^2_{x_1x_2})^2}{2(1-\rho^2_{x_1x_2})}\right] \tag{4.5}$$

$$M(t_{bre}) = M(t_\alpha) = \frac{2S_y^4}{(n-1)}\left[1 - \rho^4_{yx_1} - \rho^4_{yx_2} + 2\rho^2_{yx_1}\rho^2_{yx_2}\rho^2_{x_1x_2}\right] \tag{4.6}$$

For comparison we assuming that

$\rho_{yx_1} = \rho_{yx_2} = \rho$

and $\rho_{x_1x_2} = \rho_0$

Then (4.1), (4.2) and (4.3) we have,

$$M(t_1) = \frac{2S_y^4}{(n-1)}\left[(1-\rho^4) - \frac{\rho^4(1-\rho_0^2)}{2}\right] \tag{4.7}$$

$$M(t_{\delta r}) = \frac{2S_y^4}{(n-1)}\left[2(1-\rho^2) - \frac{(1-\rho_0^2)}{2}\right] \tag{4.8}$$

$$M(t_{bre}) = \frac{2S_y^4}{(n-1)}\left[1 - 2\rho^4(1-\rho_0^2)\right] \tag{4.9}$$

Comparing (4.7) and (4.8), we have

$$M(t_1) < M(t_{\delta r}) \qquad if \qquad (1-\rho_0^2) > 0 \tag{4.10}$$

which will always true.

Comparing (4.7) and (4.9), we have

$$M(t_1) < M(t_{bre}) \quad if \quad \rho_0^2 > \frac{1}{3} \tag{4.11}$$

which is hold.
Further we comparing (4.8) and (4.9), we have

$$M(t_{bre}) < M(t_{\delta r}) \quad if \quad (1-\rho^2) > 0 \tag{4.12}$$

which will always true.
We combining equation (4.10), (4.11) and (4.12), we have

$M(t_1) < M(t_{bre}) < M(t_{\delta r})$

Which is show that the proposed estimator is more precise than all the other estimators.

5. Empirical Study

We use the following data sets for the numerical comparison of the estimators

Data I- [Source: Anderson (1958)]
$\bar{Y}$: Head length of second son.
$\bar{X_1}$: Head length of first son.
$\bar{X_2}$: Head breadth of first son.

Data II- [Source: Khare and Sinha (2007)]
$\bar{Y}$: Weights (in kg) of children.
$\bar{X_1}$: Skull circumference (in cm) of the children and.
$\bar{X_2}$: Chest circumference (in cm) of the children.

Table 1. Descriptions of the population parameters.

Data	N	n	$\overline{Y}$	S_y^2	$\rho_{yx_1}^2$	$\rho_{yx_2}^2$	$\rho_{x_1x_2}^2$
1	25	25	183.84	100.755	0.711	0.693	0.735
2	95	95	19.4968	9.266	0.328	0.846	0.297

We have computed the percentage relative efficiency (PRE) with respect to s_y^2 which is given in the table 2 defined by

$$PRE = \frac{Var(t_0)}{MSE(.)} \times 100$$

In table 2 we present the percentage relative efficiency for each of the estimators. Based on these results we can see that the estimator t_α is equally efficient then estimator of t_{bre}, but estimator t_1 is highest efficient then all the existing estimators.

Table 2. The percentage relative efficiency of the estimators, we have

	Percentage Relative Efficiency	
Estimators	**Data** 1	**Data** 2
t_0	100	100
$t_{\delta r}$	127.650	193.050
t_{bre}	128.890	204.181
t_α	128.890	204.181
t_1	142.813	205.042

6. Simulation Study

To evaluate the efficiency of the proposed estimator, first we have generated three groups of population size $N = 5{,}000$ with population means and different covariance matrices, following multivariate normal distribution using R software. Randomly we select 15,000 samples without replacement of size 300, 500 and 700 are drawn from the whole population. For each of the sample, we computed the MSE of all the estimators as follows:

$M(.) = \frac{1}{15{,}000}\sum_{i=1}^{15{,}000}(t_{ij} - S_y^2)^2$, where, $j = 0, \delta r, bre, \alpha$ and 1

where s_y^2 denote the estimation of sample variance for $i = 1, 2, ..., 15000$ and S_y^2 represents the known population variance of the study variably. For three the population means and different covariance matrices, are given below:

Data I

$$\mu = \begin{bmatrix}400\\300\\500\end{bmatrix}, \Sigma = \begin{bmatrix}900 & 245 & 400\\245 & 1500 & 980\\400 & 980 & 1000\end{bmatrix} \text{ and } \rho_{yx_1} = 0.2,\ \rho_{yx_2} = 0.4,\ \rho_{x_1x_2} = 0.8$$

Data II

$$\mu = \begin{bmatrix}200\\600\\100\end{bmatrix}, \Sigma = \begin{bmatrix}100 & 52 & 75\\52 & 300 & 150\\75 & 150 & 200\end{bmatrix} \text{ and } \rho_{yx_1} = 0.3,\ \rho_{yx_2} = 0.5,\ \rho_{x_1x_2} = 0.6$$

Data III

$$\mu = \begin{bmatrix}100\\250\\400\end{bmatrix}, \Sigma = \begin{bmatrix}1200 & 730 & 220\\730 & 850 & 280\\220 & 280 & 990\end{bmatrix} \text{ and } \rho_{yx_1} = 0.7,\ \rho_{yx_2} = 0.2,\ \rho_{x_1x_2} = 0.3$$

Table 3 present the percentage relative efficiency for each of the estimators, Based on these results we can seen that for all data set the estimator is more efficient then all the existing estimators.

Table 3. The percentage relative efficiency of the estimators for generated data

Data	Sample Size n	Percentage Relative Efficiency Estimators				
		t_0	$t_{\delta r}$	t_{bre}	t_α	t_1
I	300	100	102.1905	110.1418	112.1403	**197.084**
	500	100	101.4044	115.6808	114.4205	**196.9417**
	700	100	103.0980	119.3504	117.7151	**197.8424**
II	300	100	105.9237	125.8871	113.8743	**214.8272**
	500	100	109.2975	118.8317	117.1416	**208.5441**
	700	100	108.9371	116.9689	116.8258	**207.7972**
III	300	100	105.4748	120.421	111.8979	**265.1738**
	500	100	106.6913	113.0217	109.1734	**264.3548**
	700	100	109.8154	115.4483	114.4826	**264.4997**

7. Conclusion

From section 4 the result of the theoretical discussions, it is inferred that the proposed estimator for estimating the population variance of the study variable under the certain condition performs better than all the existing estimators. Also it is clear from table 2 and 3 the proposed estimator is more precise then all the existing estimators for all data sets.

Acknowledgement

We thank the learned reviewers for their thorough review and highly appreciate the comments.

References

[1] Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York.

[2] Biradar, R. S. and Singh, H. P. (1998). Predictive Estimation of Finite Population Variance. Calcutta Statistical Association Bulletin, 48, 229-235.
.

[3] Cebnan, A. and Gracia, M. R. (1997). Variance Estimation Using Auxiliary Information: An Almost Unbiased Multivariate Ratio Estimator. Metrika, 45, 171-178.

[4] Das, A. K. and Tripathi, T. P. (1978). Use of Auxiliary Information in Estimating the Finite Population Variance. Sankhya, 40, 139-148.

[5] Olkin, I. (1958). Multivariate Ratio Estimator for Finite Populations. Biometrika, 45, 154-165.

[6] Isaki, T. (1983). Variance Estimation Using Auxiliary Information. Journal of the American Statistical Association, 78, 117-123.

[7] Khare, B. B. and Sinha, R. R. (2007). Estimation of ratio of the two population means using multi-auxiliary characters in the presence of non-response. Narosa Publishing House, 10, 3–14.

[8] Sukhatme, P. V. (1954). Sampling Theory of Surveys with Applications. Iowa: Iowa

State College New Delhi, India.