Vol.7.Issue.1.2019 (Jan-Mar.)
©KY PUBLICATIONS

# BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

*A Peer Reviewed International Research Journal*

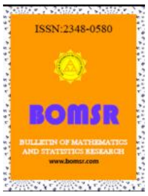
http://www.bomsr.com
Email:editorbomsr@gmail.com

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

## SOME REGRESSION-TYPE ESTIMATORS IN TWO-PHASE SAMPLING USING TWO AUXILIARY VARIABLES

**N.R. DAS[1], L.N. SAHOO[2]**
[1]Research Scholar, Department of Statistics, Utkal University, Bhubaneswar 751004, India
E-mail: nihardas41@gmail.com
[2]Visiting Faculty, Institute of Mathematics and Applications, Bhubaneswar 751003, India
E-mail: lnsahoostatuu@rediffmail.com
https://doi.org/10.33329/bomsr.71.1




**ABSTRACT**
Using two-phase sampling mechanism, the problem of estimating population mean is considered availing information on an additional auxiliary variable$z$when the population mean of the main auxiliary variable$x$ is unknown. To address this issue, although a variety of estimation methods are possible, we are confined to the regression-based methods only. We developed an alternative estimation technique by suggesting certain modifications over Kiregyera's (1984) regression-type estimator and formulated two new regression-type estimators. Considering generalization of the concept developed, we also focused attention on the creation of two generalized estimators constituting two separate families/classes of estimators. We derived two regression-type estimators as the minimum variance bound estimators of the proposed generalized estimators. An empirical study has been included to give support to our theoretical findings as well as to demonstrate performances of the competing regression-type estimators numerically.
**AMS Subject Classification: 62 D 05**
**Keywords:** Auxiliary variable, minimum variance bound estimator, regression estimator, two-phase sampling.

### 1.INTRODUCTION AND TWO-PHASE SAMPLING SET-UP

Consider a finite population$U = \{1,2,..,i,\dots,N\}$. Let $y$ and $x$ be the study variable and anauxiliary variable, taking values $y_i$ and $x_i$ respectively for the $i$th unit. When the two variables are strongly related but no information is available on the population mean$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ of$x$,regression method of estimation can be engaged to estimate the unknown population mean$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$of $y$from a sample obtained through a two phase selection. Allowing simple random sampling without replacement (SRSWOR) in each phase, the two-phase sampling scheme isdescribed in the following manner:

(a) A first phase sample $s_1(s_1 \subset U)$ of size $n_1$ is drawn from $U$ to observe only $x$ in order to find an estimate of $\bar{X}$,

(b) Given$s_1$, a second phase sample$s_2(s_2 \subset s_1)$ of size $n_2$ is drawn from $s_1$to observe $y$ only.

The two-phase sampling classical regression estimator for $\bar{Y}$is then defined by

$$t_{RG} = \bar{y}_2 - \beta_{yx}(\bar{x}_2 - \bar{x}_1),$$

where$\bar{x}_1 = \frac{1}{n_1}\sum_{i\in s_1} x_i$ and $\bar{x}_2 = \frac{1}{n_2}\sum_{i\in s_2} x_i$ are the first and second phase sample means of $x$,$\bar{y}_2 = \frac{1}{n_1}\sum_{i\in s_1} y_i$ is the second phase sample mean of $y$, and $\beta_{yx}$ is the regression coefficient of $y$ on $x$. Thevariance of $t_{RG}$ is given by

$$V(t_{RG}) = S_y^2\left[\theta_2\left(1-\rho_{yx}^2\right) + \theta_1\rho_{yx}^2\right], \quad (1)$$

where$\theta_1 = \frac{1}{n_1} - \frac{1}{N}$,$\theta_2 = \frac{1}{n_2} - \frac{1}{N}$,$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$ and $\rho_{yx}$ is the correlation coefficient between $y$ and $x$.

Ordinarily, $\beta_{yx}$ remains as an unknown quantity. Estimation process is therefore carried out with the replacement of this unknown parameter by a$s_2$ −based sample regression coefficient of $y$ on$x$. However, under this adjustment, the resulting regression estimator although consistent for$\bar{Y}$cannot retain its unbiasedness property. Butin spite of this, equation (1) serves as an asymptotic expression for its mean square error or variance to a first order of approximation. This, of course, does not affect our basic theory of regression method of estimation. In our subsequent discussions, we shall therefore assume that various regression coefficients under consideration are known prior to survey operation [*cf*., Cochran (1977, p. 190)].Sampling theory of regression estimates when regression coefficients are pre-assigned is both simple and informative.

In this paper, our purpose is to consider some modified estimation techniquesthough the utilization of an additional auxiliary variable$z$ assuming value $z_i$ for the $i$th unit of $U$with known population mean$\bar{Z} = \frac{1}{N}\sum_{i=1}^{N} z_i$, in order to develop some new regression-type estimators. It has also been shown that acceptable precision gains compared to $t_{RG}$ and other existing regression/regression-type estimators can be achieved by the developed estimators.

## 2. USE OF AN ADDITIONAL AUXILIARY VARIABLE

Consider a practical situation where$\bar{X}$ is unknown but information on a cheaply ascertainable auxiliary variable $z$is readily available on all units of the population. For example, suppose $U$ is a population of villages in a district such that $y$ equals the number of female agricultural laborers, $x$ equals the total number of laborers in a village and $z$ equals the cultivated area of the village. The value of $\bar{X}$*i.e.*, the average number of laborers per village may not be available. But, information on $z$ can be readily available from the district records so that the value of$\bar{Z}$ can be computed accurately. As a second example, if the population units are hospitals, and $y_i, x_i$ and $z_i$ are respectively the number of deaths, number of patients admitted and number of available beds, relating to the $i$th hospital, then information on $z_i'$s can be collected easily from the official records of the Health Department to find$\bar{Z}$.

In this context, we name $x$ as the main auxiliary variable and $z$ as an additional auxiliary variable. The first phase sample $s_1$ is drawn to observe $x$ and $z$ to furnish a good estimate of $\bar{X}$ by taking advantage of the correlation between $x$ and$z$. Then, the second phase sample $s_2$ is drawn from $s_1$ to observe $y$ only.We define$\bar{z}_1 = \frac{1}{n_1}\sum_{i\in s_1} z_i$and$\bar{z}_2 = \frac{1}{n_2}\sum_{i\in s_2} z_i$.

In the above scenarios, the basic work on estimation was initiated by Chand (1975) and subsequently studied by several authors producing a huge stock estimators in the survey sampling literature. The technique adopted by Chand to construct an estimator is the replacement of $\bar{x}_1$ by an improved estimator of $\bar{X}$ (usually defined in term of the additional auxiliary variable$z$ utilizing data on$s_1$) in the standard two-phase ratio, product or regression estimators. But, here our discussion is confined to regression estimators only.

Motivated by Chand (1975), Kiregyera (1984) recommended the use of a regression estimator$\bar{x}_1 - \beta_{xz}(\bar{z}_1 - \bar{Z})$ in place of $\bar{x}_1$ to produce a regression-in-regression estimator

$$t_{KRG} = \bar{y}_2 - \beta_{yx}[\bar{x}_2 - \{\bar{x}_1 - \beta_{xz}(\bar{z}_1 - \bar{Z})\}]$$

from$t_{RG}$. The variance of this estimator is

$$V(t_{KRG}) = S_y^2[\theta_2(1-\rho_{yx}^2) + \theta_1(\rho_{yx}^2 + \rho_{yx}^2\rho_{xz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz})], \quad (2)$$

where$\rho_{yz}$ and $\rho_{xz}$ are respectively correlation coefficients between $y$, $z$ and $x$, $z$.

Mukherjee *et al.* (1987) proposed three regression-type estimators as simple as the Kiregyera's estimator for practical use. The authors used a simple regression method to propose their estimators based upon information on two auxiliary variables. The estimators proposed by them are

$$t_{MRG}^{(1)} = \bar{y}_2 - \beta_{yx}(\bar{x}_2 - \bar{x}_1) - \beta_{yz}(\bar{z}_2 - \bar{z}_1),$$
$$t_{MRG}^{(2)} = \bar{y}_2 - \beta_{yx}(\bar{x}_2 - \bar{x}_1) - \beta_{yz}(\bar{z}_2 - \bar{Z}),$$
$$t_{MRG}^{(3)} = \bar{y}_2 - \beta_{yx}(\bar{x}_2 - \bar{x}_1) - \beta_{yx}\beta_{xz}(\bar{z}_1 - \bar{Z}) - \beta_{yz}(\bar{z}_2 - \bar{Z}),$$

where$\beta_{yz}$ and $\beta_{xz}$ are respectively regression coefficients of $y$ on $z$ and $x$ on $z$.The variances of the estimators are given by

$$V\left(t_{MRG}^{(1)}\right) = S_y^2[\theta_2 - (\theta_2-\theta_1)(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz})], \quad (3)$$
$$V\left(t_{MRG}^{(2)}\right) = S_y^2[\theta_2 - \theta_1\rho_{yz}^2 - (\theta_2-\theta_1)(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz})], \quad (4)$$
$$V\left(t_{MRG}^{(3)}\right) = S_y^2[\theta_1(\rho_{yz} - \rho_{yx}\rho_{xz})^2 + \theta_2(1 - \rho_{yx}^2 - \rho_{yz}^2 + 2\rho_{yx}\rho_{yz}\rho_{xz})]. \quad (5)$$

The estimator proposed by Sahoo *et al.* (1993) is defined by

$$t_{SRG} = \bar{y}_2 - \beta_{yx}(\bar{x}_2 - \bar{x}_1) - \beta_{yz}(\bar{z}_1 - \bar{Z})$$

whose variance is given by

$$V(t_{SRG}) = S_y^2[\theta_2(1-\rho_{yx}^2) + \theta_1(\rho_{yx}^2 - \rho_{yz}^2)]. \quad (6)$$

Roy (2003) suggested the use of a regression-type estimator of the form

$$t_{RRG} = \bar{y}_2 - \beta_{yx.z}\left[\left\{\bar{x}_2 - \left(\beta_{xz} - \frac{\beta_{yz}}{\beta_{yx.z}}\right)(\bar{z}_2 - \bar{Z})\right\} - \{\bar{x}_1 - \beta_{xz}(\bar{z}_1 - \bar{Z})\}\right],$$

where$\beta_{yx.z}$ is the partial regression coefficient of $y$ on $x$.The variance of $t_{RRG}$ is given by

$$V(t_{RRG}) = S_y^2[\theta_2(1-\rho_{y.xz}^2) + \theta_1(\rho_{y.xz}^2 - \rho_{yz}^2)], \quad (7)$$

where$\rho_{y.xz}$ is the multiple correlation coefficient of $y$ on $x$and $z$.

Tripathi and Ahmed (1995), and Samiuddin and Hanif (2007)considered regression-type estimators

$$t_{TRG} = \bar{y}_2 - \beta_{yx.z}(\bar{x}_2 - \bar{x}_1) - \beta_{yz.x}(\bar{z}_2 - \bar{z}_1) - \beta_{yz}(\bar{z}_1 - \bar{Z})$$

and $$t_{SHRG} = \bar{y}_2 - \beta_{yx.z}(\bar{x}_2 - \bar{x}_1) - \beta_{yz.x}(\bar{z}_2 - \bar{Z}) - (\beta_{yz} - \beta_{yz.x})(\bar{z}_1 - \bar{Z}),$$

respectively, where$\beta_{yz.x}$ is the partial regression coefficient of $y$ on $z$. However, the estimators$t_{TRG}$ and $t_{SHRG}$ are structurally similar and therefore possess the same mathematical properties. The variance expressions of these two estimators are the same and equal to that of Roy (2003) as given by (7).

**3. AN ALTERNATIVE ESTIMATION MECHANISM**

From the discussions of the preceding section, it should be clear that some authors simply forwarded regression-type estimators without providing proper justifications on their formulation [*cf.*, Sahoo *et al.* (1993), Tripathi and Ahmed (1995), Mukerjee*et al.* (1987)]. But, here our objective is how to achieve improvements over the standard regression estimator with the involvement of an additional auxiliary variable$z$. We are also motivated by the approach undertaken by Chand (1975) and Kiregyera (1980, 1984).

It should be noted here that Kiregyera (1984) simply replaced$\bar{x}_1$ by the regression estimator $\bar{x}_1 - \beta_{xz}(\bar{z}_1 - \bar{Z})$ in $t_{RG}$ in order to compose his regression-in-regression estimator $t_{KRG}$ considering later as a better estimator of the unknown mean $\bar{X}$ than the latter. But in practice, we realize that$\bar{x}_2$ provides a less efficient estimate of $\bar{X}$than$\bar{x}_1$. Therefore, one may also think of utilizing $z$ to find better estimates of $\bar{X}$than$\bar{x}_2$. For this we have two options: use of either $\bar{x}_2 - \beta_{xz}(\bar{z}_2 - \bar{z}_1)$ or $\bar{x}_2 - \beta_{xz}(\bar{z}_2 - \bar{Z})$ in place of$\bar{x}_2$. At the same time we also think of using $\bar{x}_1 - \beta_{xz}(\bar{z}_1 - \bar{Z})$ in place of$\bar{x}_1$. This mechanism leads to develop the following new regression-type estimators:

$$t_{RG}^{(1)} = \bar{y}_2 - \beta_{yx}[\{\bar{x}_2 - \beta_{xz}(\bar{z}_2 - \bar{z}_1)\} - \{\bar{x}_1 - \beta_{xz}(\bar{z}_1 - \bar{Z})\}]$$
$$t_{RG}^{(2)} = \bar{y}_2 - \beta_{yx}[\{\bar{x}_2 - \beta_{xz}(\bar{z}_2 - \bar{Z})\} - \{\bar{x}_1 - \beta_{xz}(\bar{z}_1 - \bar{Z})\}].$$

The variances of these estimators can be easily derived as

$$V\left(t_{RG}^{(1)}\right) = S_y^2\left[\theta_2(1 - \rho_{yx}^2) + \theta_1\rho_{yx}^2 - (\theta_2 - 2\theta_1)(\rho_{yx}^2\rho_{xz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz})\right] \quad (8)$$
$$V\left(t_{RG}^{(2)}\right) = S_y^2\left[\theta_2(1 - \rho_{yx}^2) + \theta_1\rho_{yx}^2 - (\theta_2 - \theta_1)(\rho_{yx}^2\rho_{xz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz})\right]. \quad (9)$$

**4. TWO GENERALIZED ESTIMATORS**

Instead of considering two-phase regression estimator$t_{RG}$ as the base estimator, let us now consider more generally a two-phase difference estimator. As in previous section, here two cases are also taken into account. In the first case, we consider a two-phase difference estimator of the form $\bar{y}_2 - \eta(\bar{x}_2 - \bar{x}_1)$ and create a generalized estimator replacing sample mean estimators $\bar{x}_2$ and $\bar{x}_1$ respectively by two difference estimators $\bar{x}_2 - \eta_2(\bar{z}_2 - \bar{z}_1)$and$\bar{x}_1 - \eta_1(\bar{z}_1 - \bar{Z})$. In the second case, considering the difference estimator $\bar{y}_2 - \omega(\bar{x}_2 - \bar{x}_1)$ as a base estimator,a generalized estimator has been formulated substituting difference estimators$\bar{x}_2 - \omega_2(\bar{z}_2 - \bar{Z})$ and $\bar{x}_1 - \omega_1(\bar{z}_1 - \bar{Z})$ for $\bar{x}_2$ and $\bar{x}_1$ respectively. After going through the said estimation procedures, we now arrive at the following generalized estimators:

$$t_D^{(g1)} = \bar{y}_2 - \eta[\{\bar{x}_2 - \eta_2(\bar{z}_2 - \bar{z}_1)\} - \{\bar{x}_1 - \eta_1(\bar{z}_1 - \bar{Z})\}]$$
$$t_D^{(g2)} = \bar{y}_2 - \omega[\{\bar{x}_2 - \omega_2(\bar{z}_2 - \bar{Z})\} - \{\bar{x}_1 - \omega_1(\bar{z}_1 - \bar{Z})\}].$$

The coefficients $\eta$, $\eta_1$, $\eta_2$, $\omega$, $\omega_1$ and $\omega_2$ appearing in $t_D^{(g1)}$ and $t_D^{(g2)}$, are constants, which in particular may be random variables converging in probability to some known or unknown constants which may depend on the population characteristics.

**5. SOME MATHEMATICAL PROPERTIES OF THE GENERALIZEDESTIMATORS**

Alternatively, the generalized estimators$t_D^{(g1)}$ and $t_D^{(g2)}$ can also be rewritten as

$$t_D^{(g1)} = \bar{y}_2 - \eta[\{\bar{x}_2 - \eta_2((\bar{z}_2 - \bar{Z}) - (\bar{z}_1 - \bar{Z}))\} - \{\bar{x}_1 - \eta_1(\bar{z}_1 - \bar{Z})\}]$$
$$= \bar{y}_2 - \eta[\{\bar{x}_2 - \eta_2(\bar{z}_2 - \bar{Z})\} - \{\bar{x}_1 - (\eta_1 + \eta_2)(\bar{z}_1 - \bar{Z})\}],$$

and $$t_D^{(g2)} = \bar{y}_2 - \omega\left[\{\bar{x}_2 - \omega_2\left((\bar{z}_2 - \bar{z}_1) + (\bar{z}_1 - \bar{Z})\right)\} - \{\bar{x}_1 - \omega_1(\bar{z}_1 - \bar{Z})\}\right]$$
$$= \bar{y}_2 - \omega[\{\bar{x}_2 - \omega_2(\bar{z}_2 - \bar{z}_1)\} - \{\bar{x}_1 - (\omega_1 - \omega_2)(\bar{z}_1 - \bar{Z})\}].$$

These formulas clearly show that $t_D^{(g1)}$ arises as a special case of$t_D^{(g2)}$ if $\omega = \eta, \omega_2 = \eta_2$and$\omega_1 = \eta_1 + \eta_2$; and $t_D^{(g2)}$ arises as a special case of $t_D^{(g1)}$ if $\eta = \omega, \eta_2 = \omega_2$ and$\eta_1 = \omega_1 - \omega_2$. Although, these intimate connections between the coefficients imply that one estimator is a particular case of the other, in our discussions we shall treat$t_D^{(g1)}$ and $t_D^{(g2)}$ as separate generalized estimators.

The estimators $t_D^{(g1)}$ and $t_D^{(g2)}$ are quite capable of producing classes/families of estimators. Because, for proper selections of their coefficients, they can be transferred to large number estimators (ratio-type, product-type and regression-type) using one or two auxiliary variables. However, it should be noted here that the two classes of estimators corresponding to $t_D^{(g1)}$ and $t_D^{(g2)}$ are not necessarily mutually exclusive. Because, for the simplest case when$\eta = \omega = 0$, we see that $t_D^{(g1)} = t_D^{(g2)} = \bar{y}_2$ which is the mean per unit estimator based on the second-phase sample using no auxiliary information. Similarly, if$\eta_1 = \eta_2 = \omega_1 = \omega_2 = 0$, this corresponds to the case of the use of single auxiliary variable $x$ and $t_D^{(g1)} = \bar{y}_2 - \eta(\bar{x}_2 - \bar{x}_1)$ and $\bar{y}_2 - \omega(\bar{x}_2 - \bar{x}_1)$ which implies that$t_D^{(g1)} = t_D^{(g2)}$. Further, as is expected, when$\eta = \omega = \frac{\bar{y}_2}{\bar{x}_2}$, $t_D^{(g1)} = t_D^{(g2)} = \bar{y}_2 \frac{\bar{x}_1}{\bar{x}_2}$, the classical two-phase ratio estimator; when $\eta = \omega = -\frac{\bar{y}_2}{\bar{x}_1}$, $t_D^{(g1)} = t_D^{(g2)} = \bar{y}_2 \frac{\bar{x}_2}{\bar{x}_1}$, the classical two-phase product estimator; and when$\eta = \omega = \beta_{yx}$, $t_D^{(g1)} = t_D^{(g2)} = t_{RG}$ which is our considered classical two-phase regression estimator.

We would also like to remark here that the series of regression-type estimatorsdefined earlier can be derived as special cases of either $t_D^{(g1)}$ or $t_D^{(g2)}$ or both. In table1, different choices of the coefficients are suggested for which these estimators can be converted to the said regression-type estimators.

**Table–1: Choices of Coefficients forSpecial Cases of $t_D^{(g1)}$ and $t_D^{(g2)}$**

| Estimator | Special case of $t_D^{(g1)}$ when | Special case of $t_D^{(g2)}$ when | Special cases of both $t_D^{(g1)}$ and $t_D^{(g2)}$ when |
|---|---|---|---|
| $t_{KRG}$ | - | - | $\eta = \omega = \beta_{yx}, \eta_1 = \omega_1 = \beta_{xz}, \eta_2 = \omega_2 = 0$ |
| $t_{MRG}^{(1)}$ | $\eta = \beta_{yx}, \eta_1 = 0,$ $\eta_2 = -\frac{\beta_{yz}}{\beta_{yx}}$ | $\omega = \beta_{yx}, \omega_1 = \omega_2 = -\frac{\beta_{yz}}{\beta_{yx}}$ | - |
| $t_{MRG}^{(2)}$ | $\eta = \beta_{yx}, \eta_1 = \frac{\beta_{yz}}{\beta_{yx}},$ $\eta_2 = -\frac{\beta_{yz}}{\beta_{yx}}$ | $\omega = \beta_{yx}, \omega_1 = 0,$ $\omega_2 = -\frac{\beta_{yz}}{\beta_{yx}}$ | - |
| $t_{MRG}^{(3)}$ | $\eta = \beta_{yx}, \eta_1 = \beta_{xz} + \frac{\beta_{yz}}{\beta_{yx}},$ | $\omega = \beta_{yx}, \omega_1 = \beta_{xz},$ $\omega_2 = -\frac{\beta_{yz}}{\beta_{yx}}$ | - |

| | | | |
|---|---|---|---|
| | $\eta_2 = -\frac{\beta_{yz}}{\beta_{yx}}$ | | |
| $t_{SRG}$ | $\eta = \beta_{yx}, \eta_1 = \eta_2 = \frac{\beta_{yz}}{\beta_{yx}}$ | $\omega = \beta_{yx}, \omega_1 = 2\frac{\beta_{yz}}{\beta_{yx}}, \omega_2 = \frac{\beta_{yz}}{\beta_{yx}}$ | $\eta = \omega = \beta_{yx}, \eta_1 = \omega_1 = \frac{\beta_{yz}}{\beta_{yx}}, \eta_2 = \omega_2 = 0$ |
| $t_{RRG}$ | $\eta = \beta_{yx.z}, \eta_1 = \frac{\beta_{yz}}{\beta_{yx.z}}$, $\eta_2 = \beta_{xz} - \frac{\beta_{yz}}{\beta_{yx.z}}$ | $\omega = \beta_{yx.z}, \omega_1 = \beta_{xz}$, $\omega_2 = \beta_{xz} - \frac{\beta_{yz}}{\beta_{yx.z}}$ | - |
| $t_{TRG}$ | $\eta = \beta_{yx.z}, \eta_1 = \frac{\beta_{yz}}{\beta_{yx.z}}$, $\eta_2 = -\frac{\beta_{yz.x}}{\beta_{yx.z}}$ | $\omega = \beta_{yx.z}, \omega_1 = \frac{\beta_{yz}-\beta_{yz.x}}{\beta_{yx.z}}, \omega_2 = -\frac{\beta_{yz.x}}{\beta_{yx.z}}$ | - |
| $t_{SHRG}$ | $\eta = \beta_{yx.z}, \eta_1 = \frac{\beta_{yz}}{\beta_{yx.z}}$, $\eta_2 = -\frac{\beta_{yz.x}}{\beta_{yx.z}}$ | $\omega = \beta_{yx.z}, \omega_1 = \frac{\beta_{yz}-\beta_{yz.x}}{\beta_{yx.z}}, \omega_2 = -\frac{\beta_{yz.x}}{\beta_{yx.z}}$ | - |
| $t_{RG}^{(1)}$ | $\eta = \beta_{yx}, \eta_1 = \eta_2 = \beta_{xz}$ | $\omega = \beta_{yx}, \omega_1 = \beta_{yx} + \beta_{xz}$, $\omega_2 = \beta_{xz}$ | - |
| $t_{RG}^{(2)}$ | $\eta = \beta_{yx}, \eta_1 = 0, \eta_2 = \beta_{xz}$ | $\omega = \beta_{yx}, \omega_1 = \omega_2 = \beta_{xz}$ | - |

## 6. OPTIMAL COEFFICIENTS FOR THE GENERALIZED ESTIMATORS

Although we presented some specific cases of $t_D^{(g1)}$ and $t_D^{(g2)}$ for some specific selections of their coefficients, we did not discuss the optimal choices of these constants. Rather, we require that these constants be independent of sample data and determined, in principle, before sampling. We shall now find the best choices of $\eta, \eta_1, \eta_2, \omega, \omega_1$ and $\omega_2$, that is, choices that minimizes the variance of the concerned estimator. The optimal coefficient-values usually depend on the unknown population characteristics, so the optimal estimators cannot be useful. In order to meet this challenge, we have two options. One is to replace the optimal-coefficient values by sample-based estimates. Anotheris to use regression estimation approach on the basis of strong a priori or theoretical expectations (or expectations based on some previous empirical work) which of course plays a central role in our present work.

Noting that $E\left(t_D^{(g1)}\right) = E\left(t_D^{(g2)}\right) = \bar{Y}$ and for simplicity using $H_1 = \eta\eta_1, H_2 = \eta\eta_2$, $K_1 = \omega\omega_1$ and $K_2 = \omega\omega_2$, after some straightforward algebra the errors of the estimators are brought into the following forms:

$$t_D^{(g1)} - \bar{Y} = (\bar{y}_2 - \bar{Y}) - \eta(\bar{x}_2 - \bar{X}) + \eta(\bar{x}_1 - \bar{X}) - H_1(\bar{z}_1 - \bar{Z}) + H_2(\bar{z}_2 - \bar{Z}) - H_2(\bar{z}_1 - \bar{Z}),$$

$$t_D^{(g2)} - \bar{Y} = (\bar{y}_2 - \bar{Y}) - \omega(\bar{x}_2 - \bar{X}) + \omega(\bar{x}_1 - \bar{X}) - K_1(\bar{z}_1 - \bar{Z}) + K_2(\bar{z}_2 - \bar{Z}).$$

Hence, after considerable algebraic treatments, variances of $t_D^{(g1)}$ and $t_D^{(g2)}$ are obtained as the following compact expressions:

$$V\left(t_D^{(g1)}\right) = \theta_2 S_y^2 + \theta_1\left(H_1^2 S_z^2 - 2H_1 S_{yz}\right) + (\theta_2 - \theta_1)\left(\eta^2 S_x^2 - 2\eta S_{yx} + H_2^2 S_z^2 + 2H_2 S_{yz} - 2\eta H_2 S_{xz}\right) \quad (10)$$

$$V\left(t_D^{(g2)}\right) = \theta_2 S_y^2 + \theta_2\left(K_2^2 S_z^2 + 2K_2 S_{yz}\right) + \theta_1\left[K_1(K_1 - 2K_2)S_z^2 - 2K_1 S_{yz}\right] + (\theta_2 - \theta_1)\left(\omega^2 S_x^2 - 2\omega S_{yx} - 2\omega K_2 S_{xz}\right). \quad (11)$$

In order to minimize $V\left(t_D^{(g1)}\right)$ with respect to $H_1, H_2$ and$\eta$, we differentiate (10) partially with respect to the three unknown constants and then setting the resulting equations to zero, we obtain the followingnormal equations:

$$H_1 - \beta_{yz} = 0, \quad (12)$$

$$H_2 + \beta_{yz} - \eta\beta_{xz} = 0, \quad (13)$$

and $$\eta - \beta_{yx} - H_2\beta_{zx} = 0. \quad (14)$$

Solving these normal equations, optimum values of $H_1, H_2$ and $\eta$are obtained as

$$\hat{H}_1 = \beta_{yz}, \quad (15)$$

$$\hat{H}_2 = -(\beta_{yz} - \hat{\eta}\beta_{xz}) = -\beta_{yz.x}, \quad (16)$$

and $$\hat{\eta} = \beta_{yx.z}. \quad (17)$$

respectively. Finally, we also find the optimum values of $\eta_1$ and $\eta_2$ as

$$\hat{\eta}_1 = \frac{\beta_{yz}}{\beta_{yx.z}} \text{and} \hat{\eta}_2 = -\frac{\beta_{yz.x}}{\beta_{yx.z}}. \quad (18)$$

Use of these optimal values of the coefficients, we obtain the minimum value of $V(t_D^{(g1)})$ as

$$V_{min}\left(t_D^{(g1)}\right) = S_y^2\left[\theta_2\left(1 - \rho_{y.xz}^2\right) + \theta_1(\rho_{y.xz}^2 - \rho_{yz}^2)\right], \quad (19)$$

which may be called as the minimum variance bound (MVB) of$t_D^{(g1)}$. It should be noted here that this minimum variance is alsoequation (7) which is the variance of Roy's (2003) estimator$t_{RRG}$.

The optimum estimator corresponding to$V_{min}\left(t_D^{(g1)}\right)$, obtained after substituting optimal values $\hat{\eta}, \hat{\eta}_1$ and $\hat{\eta}_2$ in$t_D^{(g1)}$, is a regression-type estimator defined by

$$\hat{t}_{RG}^{(g1)} = \bar{y}_2 - \beta_{yx.z}\left[\left\{\bar{x}_2 + \frac{\beta_{yz.x}}{\beta_{yx.z}}(\bar{z}_2 - \bar{z}_1)\right\} - \left\{\bar{x}_1 - \frac{\beta_{yz}}{\beta_{yx.z}}(\bar{z}_1 - \bar{Z})\right\}\right].$$

Here we also see that the optimum estimator $\hat{t}_{RG}^{(g1)}$ [may be called as the minimum variance bound estimator (MVB estimator) of$t_D^{(g1)}$] is the same as theTripathi and Ahmed's (1995) regression-type estimator$t_{TRG}$.

In a similar manner, differentiating $V\left(t_D^{(g2)}\right)$ in (11) with respect to $K_1, K_2$ and $\omega$ partially and then solving the derived normal equations, optimum values of$\omega$, $\omega_1$ and $\omega_2$ are respectively obtained as

$$\hat{\omega} = \beta_{yx.z}, \hat{\omega}_1 = \frac{\beta_{yz} - \beta_{yz.x}}{\beta_{yx.z}} \text{and} \hat{\omega}_2 = -\frac{\beta_{yz.x}}{\beta_{yx.z}}. \quad (20)$$

In the light of these derived optimal values, the minimum variance and the corresponding optimum estimator of $t_D^{(g2)}$are obtained as

$$V_{min}\left(t_D^{(g2)}\right) = S_y^2\left[\theta_2\left(1 - \rho_{y.xz}^2\right) + \theta_1(\rho_{y.xz}^2 - \rho_{yz}^2)\right], \quad (21)$$

$$\hat{t}_{RG}^{(g2)} = \bar{y}_2 - \beta_{yx.z}\left[\left\{\bar{x}_2 + \frac{\beta_{yz.x}}{\beta_{yx.z}}(\bar{z}_2 - \bar{Z})\right\} - \left\{\bar{x}_1 - \frac{\beta_{yz} - \beta_{yz.x}}{\beta_{yx.z}}(\bar{z}_1 - \bar{Z})\right\}\right].$$

Here the important points to note is that$V_{min}\left(t_D^{(g1)}\right) = V_{min}\left(t_D^{(g2)}\right)$, which implies that the MVBs of $t_D^{(g1)}$ and $t_D^{(g2)}$ are equal, and on the other hand, the MVB estimator $\hat{t}_{RG}^{(g2)}$ is theSamiuddin and Hanif's (2007) regression-type estimator$t_{SHRG}$.

To sum up our preceding findings,we state here that the classes of estimators defined by $t_D^{(g1)}$ and $t_D^{(g2)}$although possess the same MVB, their MVB estimators are different.

**7. SOME REMARKS ON THE EFFICIENCIES OF THE NEW REGRESSION-TYPEESTIMATORS**

Our earlier results show that the series of regression/regression-type estimators*viz.*,$t_{RG}$, $t_{KRG}$, *tMRG(1), tMRG(2), tMRG(3),tSRG, tRRG, tTRG, tSHRG, tRG(1)* and *tRG(2)*considered in this work are particular cases of either $t_D^{(g1)}$ or $t_D^{(g2)}$ or both.These estimators therefore should beexpected to be less efficient than $\hat{t}_{RG}^{(g1)}$ and $\hat{t}_{RG}^{(g2)}$ as the latters are the MVB estimators of $t_D^{(g1)}$and$t_D^{(g2)}$. But, as we seethat$\hat{t}_{RG}^{(g1)} = t_{TRG}$,$\hat{t}_{RG}^{(g2)} = t_{SHRG}$ and$V(t_{RRG}) = V(t_{TRG}) = V(t_{SHRG}) = V\left(\hat{t}_{RG}^{(g1)}\right) = V\left(\hat{t}_{RG}^{(g2)}\right)$, comparing equations (1) to (6), 8 and 9 with (19) or (21),we can directly and easilyverify that$t_{RG}$, $t_{KRG}$, $t_{MRG}^{(1)}$, $t_{MRG}^{(2)}$, $t_{MRG}^{(3)}$,$t_{SRG}$, $t_{RG}^{(1)}$ and $t_{RG}^{(2)}$are unconditionally less efficient than either$\hat{t}_{RG}^{(g1)}$ or $\hat{t}_{RG}^{(g2)}$.

In order to have an idea on the performances of the proposed new regression estimators $t_{RG}^{(1)}$ and $t_{RG}^{(2)}$ over$t_{RG}$, $t_{KRG}$, $t_{MRG}^{(1)}$, $t_{MRG}^{(2)}$, $t_{MRG}^{(3)}$, $t_{SRG}$ and $t_{RRG}$, we compare variances of $t_{RG}^{(1)}$ and $t_{RG}^{(2)}$ with those of others and derived various sufficient conditions in terms of the correlation coefficients $\rho_{yz}$, $\rho_{yz}$, $\rho_{xz}$, and sample sizes $n_1$, $n_2$ to show when the new estimators are better ones. Although detailed derivations and results are suppressed here to save space, we just point out here that some of the derived conditions are very complicated in nature to be verified in a specific situation. This of course makes the job of identifying $t_{RG}^{(1)}$ and $t_{RG}^{(2)}$as better estimators than their competitors a difficult one.However, to relax this job to some extent, we do carry out an empirical study in the next section.

**8. EMPIRICAL STUDY**

In order to check the worthiness of different estimators quantitatively and to give support to our theoretical findings, we carry out an empirical study to compare the performances of different estimators on the ground of their variance.We considered 10 natural populations as described in table 2. Out of the 10 considered populations, six populations *i.e.*, populations 1, 3, 6, 7, 9 and 10 satisfy the condition$\rho_{yz} < \frac{1}{2}\rho_{yx}\rho_{xz}$, a derived sufficient condition for variance comparison of different estimators. We mark these populations as Group–A populations and rest as Group–B. For comparison purposes, out of the five estimators $t_{RRG}$, $t_{TRG}$, $t_{SHRG}$, $\hat{t}_{RG}^{(g1)}$ and $\hat{t}_{RG}^{(g2)}$ whose variances are equal, only the MVB estimator $\hat{t}_{RG}^{(g2)}$ was taken into consideration.

Assuming SRSWOR at each phase, percentage relative efficiencies of different estimators for selected values of $n_1$ and $n_2$ when compared to the direct estimator$\bar{y}_2$, are displayed in table 3. Sample sizes for each population (as shown in table 3) are decided according to the condition$n_2 < \frac{n_1}{2}$ which is also a derived sufficient condition for efficiency comparison.

A close scrutiny of the entries of table 3 highlights the following empirical findings:

(i) The estimator $\hat{t}_{RG}^{(g2)}$ attains the maximum precision gain amongst all for all populations under consideration in conformity with our theoretical findings.

(ii) For the populations in Group A where the conditions$\rho_{yz} < \frac{1}{2}\rho_{yx}\rho_{xz}$ and $n_2 < \frac{n_1}{2}$ are met, $t_{RG}^{(1)}$ is less efficient than$t_{RG}^{(2)}$, $t_{KRG}$ is less efficient than$t_{RG}$, and $t_{RG}$ is less efficient than both $t_{RG}^{(1)}$ and$t_{RG}^{(2)}$. At the same time, for the populations in Group B although$t_{RG}^{(2)}$ is inferior to$t_{RG}^{(1)}$, in two cases both are inferior to $t_{KRG}$ which is also superior to$t_{RG}$.

(iii) For 5 populations in Group A and 2 populations in Group B, $t_{SRG}$ is inferior to both $t_{RG}^{(1)}$ and $t_{RG}^{(2)}$. Out of the remaining 3 populations where the former one is better than the later ones, comes out as the second best performer among the comparable estimators in two cases.

(iv) Among the estimators $t_{MRG}^{(1)}, t_{MRG}^{(2)}$ and $t_{MRG}^{(3)}$, $t_{MRG}^{(1)}$ comes out as the worst one whereas $t_{MRG}^{(3)}$ being better than other two for 8 populations, is worse than $t_{RG}^{(1)}$ and $t_{RG}^{(2)}$ for 7 populations.

**Table–2: Description of the Populations**

| Pop. No. | Source | Size | $y$ | $x$ | $z$ |
|---|---|---|---|---|---|
| 1 | Srivastava *et al.* (1988) | 21 plants | pods/plant | no. of primary branches | seeds/pod |
| 2 | Srivastava *et al.* (1988) | 21 plants | pods/plant | no. of secondary branches | seeds/pod |
| 3 | Srivastava *et al.* (1988) | 55 persons | mid arm circumference | body weight | skull circumference |
| 4 | Sahoo and Swain (1980) | 50 plants | yield/plant | no. of tillers | percentage of sterility |
| 5 | Sukhatme and Chand (1977) | 120 trees | bushels of apples harvested in 1964 | apple trees of bearing age in 1964 | bushels of apples harvested in 1959 |
| 6 | Snedecor and Cochran (1967), p.113 | 18 soil samples | estimated plant available phosphorus | inorganic phosphorus | organic phosphorus |
| 7 | Shukla (1966) | 50 plants | fiber yield/plant | plant green weight | base diameter |
| 8 | Srivastava (1971) | 50 plants | yield/plant | height of the plant | base diameter |
| 9 | Tripathi (1980) | 225 households | persons in service | educated persons | size of households |
| 10 | Murthy (1977), p.228 | 80 industries | output | no. of workers | fixed capital |

After further analyzing our empirical results, we now tentatively conclude that the method of estimation used to compose $t_D^{(g1)}$ and $t_D^{(g2)}$ can be effectively used in many practical situations whereas the proposed estimators $t_{RG}^{(1)}$ and $t_{RG}^{(2)}$ can perform well under the favorable conditions $\rho_{yz} < \frac{1}{2}\rho_{yx}\rho_{xz}$ and $n_2 < \frac{n_1}{2}$. These empirical findings of course agree with our analytical findings.

**Table–3: Relative Efficiencies of Different Estimators w.r.t. $\bar{y}_2$(in %)**

| Pop. No. | Sample Sizes | | Estimators | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n_1$ | $n_2$ | $t_{RG}$ | $t_{KRG}$ | $t_{MRG}^{(1)}$ | $t_{MRG}^{(2)}$ | $t_{MRG}^{(3)}$ | $t_{SRG}$ | $t_{RG}^{(1)}$ | $t_{RG}^{(2)}$ | $\hat{t}_{RG}^{(g2)}$ |
| 1 | 12 | 5 | 159.21 | 159.08 | 156.80 | 156.95 | 157.30 | 159.41 | 159.91 | 160.05 | 160.95 |
| 2 | 12 | 5 | 135.91 | 136.42 | 135.95 | 136.07 | 136.29 | 136.05 | 136.68 | 136.56 | 137.97 |
| 3 | 25 | 10 | 127.02 | 122.09 | 134.54 | 134.81 | 135.52 | 127.37 | 136.79 | 143.00 | 147.22 |
| 4 | 20 | 8 | 156.05 | 156.48 | 170.73 | 176.17 | 167.53 | 158.07 | 159.05 | 158.37 | 196.14 |
| 5 | 50 | 20 | 265.61 | 548.68 | 135.83 | 185.83 | 149.73 | 556.00 | 146.31 | 114.25 | 651.56 |
| 6 | 10 | 4 | 156.02 | 153.50 | 139.55 | 145.29 | 157.92 | 160.95 | 161.10 | 161.70 | 163.95 |
| 7 | 30 | 12 | 122.21 | 121.93 | 108.33 | 109.33 | 112.03 | 122.86 | 123.38 | 124.83 | 126.60 |
| 8 | 20 | 8 | 164.84 | 177.36 | 181.35 | 199.62 | 245.32 | 194.08 | 148.87 | 139.98 | 256.09 |
| 9 | 60 | 25 | 152.14 | 151.24 | 149.55 | 150.85 | 151.32 | 151.06 | 152.71 | 152.78 | 154.02 |
| 10 | 25 | 10 | 201.33 | 152.98 | 101.74 | 142.05 | 205.46 | 267.93 | 132.17 | 296.68 | 411.45 |

## REFERENCES


[1]. Chand, L. (1975). *Some ratio-type estimators based on two or more auxiliary variables*. Unpublished Ph.D. Dissertation, Iowa State University, Ames, Iowa.

[2]. Cochran, W.G. (1977). *Sampling Techniques* (3rd Edition). Wiley Eastern Limited.

[3]. Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika*, 27, 217-223.

[4]. Kiregyera, B. (1984). Regression-type estimators using two auxiliary variables and the model of double sampling. *Metrika*, 31, 215-226.

[5]. Mukerjee, R., Rao, T.J. and Vijayan, K. (1987). Regression-type estimators using multiple auxiliary information. *Australian Journal of Statistics*, 29, 244-254.

[6]. Murthy, M.N. (1977). *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta.

[7]. Roy, D.C. (2003). A regression-type estimate in two-phase sampling using two auxiliary variables. *Pakistan Journal of Statistics*, 19, 281-290.

[8]. Sahoo, J., Sahoo, L.N. and Mohanty, S. (1993). A regression approach to estimation in two-phase sampling using two auxiliary variables. *Current Science*, 65, 73-75.

[9]. Sahoo, L.N. and Swain, A.K.P.C. (1980). Unbiased ratio-cum-product estimator. *Sankhyā*, C42, 56-62.

[10]. Samiuddin, M. and Hanif, M. (2007). Estimation of population mean in single and two phase sampling with or without additional information. *Pakistan Journal of Statistics*, 23, 99-118.

[11]. Shukla, G.K. (1966). An alternative multivariate ratio estimate for finite population. *Calcutta Statistical Association Bulletin*, 15, 127-134.

[12]. Snedecor, G.W. and Cochran, W.G. (1967). *Statistical Methods* (6th Edition). Iowa State University Press, Ames, Iowa.

[13]. Srivastava (1971). Generalized estimator for the mean of a finite population using multi-auxiliary information. *Journal of the American Statistical Association*, 66, 404-407.

[14]. Srivastava, S. Rani, Khare, B.B. and Srivastava, S.R. (1988). On generalized chain estimators for ratio and product of two population means using auxiliary characters. *Assam Statistical Review*, 2, 21-29.

[15]. Sukhatme, B.V. and Chand, L. (1977). Multivariate ratio type estimator. *Proceedings, Social Statistics Section, American Statistical Association*, 927-931.

[16]. Tripathi, T.P. (1980). A general class of estimators for population ratio. *Sankhyā*, C42, 63-75.

[17]. Tripathi, T.P. and Ahmed, M.S. (1995). A class of estimators for a finite population mean based on multivariate information and double sampling. *Calcutta Statistical Association Bulletin*, 45, 203-218.