**Far East Journal of Applied Mathematics**

https://pphmjopenaccess.com
http://dx.doi.org/10.17654/0972096024003
Volume 117, Number 1, 2024, Pages 49-66 **P-ISSN: 0972-0960**

# COMPOSITE RATIO ESTIMATORS IN A TWO-PHASE SAMPLING USING MULTIPLE ADDITIONAL SUPPLEMENTARY VARIABLES

**Ajit Kumar Sabat, R. K. Sahoo and L. N. Sahoo**

Asian School of Business Management University
Shiksha Vihar, Bhola, Chandaka
Bhubaneswar-754012, India
e-mail: ajitsabat.1992@gmail.com

Department of Statistics
Central University of Haryana
Jant Pali, Mahendragarh-123031, India
e-mail: ranjansahoostatistics@gmail.com

Utkal University
Bhubaneswar-751004, India
e-mail: lnsahoostatuu@rediffmail.com

Received: October 25, 2023; Accepted: March 13, 2024

2020 Mathematics Subject Classification: 62D05.

Keywords and phrases: supplementary variable, modified approach, ratio estimator, two-phase sampling.

How to cite this article: Ajit Kumar Sabat, R. K. Sahoo and L. N. Sahoo, Composite ratio estimators in a two-phase sampling using multiple additional supplementary variables, Far East Journal of Applied Mathematics 117(1) (2024), 49-66.
http://dx.doi.org/10.17654/0972096024003



Published Online: March 28, 2024

**Abstract**

In this paper, we focus on the formulation of two multi-variate composite (generalized) ratio-type estimators for the population mean in the presence of $(p+1)$ supplementary variables. Estimation mechanism has been carried out in the framework of a two-phase sampling procedure when no information is sought on the population mean of the main supplementary variable but the population means of the rest $p$ supplementary variables (called as the additional supplementary variables) are known accurately.

## 1. Introduction and the Configuration of Two-phase Sampling

We explore two positively correlated variables $y$ and $x$, respectively, as the survey variable and a supplementary variable assuming paired values $(y_i, x_i)$, $i = 1, 2, ..., N$, on the $i$th unit of $U$, the finite population under consideration. In sample surveys, two-phase sampling technique is ordinarily expended when mean $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is unknown and there is a need to estimate mean $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. Here this scheme of sampling consists of taking a large initial sample (first phase sample) $s_1 \, (s_1 \subset U)$ of $n_1$ units by simple random sampling without replacement (SRSWOR) to obtain a good estimate of $\overline{X}$ by collecting measurements on $x$ for all the $n_1$ sampled units. Then, a sub-sample (second phase sample) $s_2 \, (s_2 \subset s_1)$ of $n_2$ units is selected from $s_1$ by SRSWOR to measure $y$ for each of these $n_2$ units. Define $\bar{x}_1 = \frac{1}{n_1}\sum_{i \in s_1} x_i$ and $\bar{x}_2 = \frac{1}{n_2}\sum_{i \in s_2} x_i$ as the sample means of $x$ based on $s_1$ and $s_2$, respectively; and $\bar{y}_2 = \frac{1}{n_2}\sum_{i \in s_2} y_i$ as the sample mean of $y$ based on $s_2$. In this context, ratio method of estimation can be recommended as an error reducing technique for which the classical two-phase ratio estimator of $\overline{Y}$ is defined by

$$t_R = \bar{y}_2 \frac{\bar{x}_1}{\bar{x}_2}.$$

The asymptotic mean square error (MSE) of $t_R$ is given by

$$M(t_R) = \theta_2 S_y^2 + (\theta_2 - \theta_1)(R^2 S_x^2 - 2RS_{yx}), \tag{1}$$

where $\theta_1 = \frac{1}{n_1} - \frac{1}{N}$, $\theta_2 = \frac{1}{n_2} - \frac{1}{N}$, $S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/(N-1)$, $S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2/(N-1)$, $S_{yx} = \sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})/(N-1)$ and $R = \bar{Y}/\bar{X}$.

As is already known (cf., Cochran [3, p. 158]), $t_R$ performs well when $R = \beta_{yx} = S_{yx}/S_x^2$, i.e., when the regression line of $y_i$ on $x_i$ is a straight line through the origin and the variance of $y_i$ about this line is proportional to $x_i$. Improvements over $t_R$ can also be made either by modifying the sampling scheme or the estimator to bring considerable variance reduction compared to $t_R$. However, one of the ways for this achievement is the engagement of one or more additional supplementary variables to recompose the basic estimators, originally grown by Chand [2], Sukhatme and Chand [20] and consequently by Kiregyera [7, 8] with more clarifications. The followed technique (we call as *Chand-Kiregyera Approach*) is to estimate $\bar{X}$ exploiting information on an additional supplementary variable and then to use such an estimate in place of $\bar{x}_1$ in $t_R$ (may be called as the *base estimator*) for $\bar{Y}$. In due course of time, several authors also followed the idea of using an additional supplementary variable to compose large varieties of estimators (see, for example, Kumar and Vishwakarma [10], Patel and Shah [13], Kumar and Tiwari [9], Sahoo et al. [15], Dubey et al. [4, 5]). But in many cases, it is seen that the estimators are simply forwarded without providing any explanation on their formulation technique (cf., Mukerjee et al. [11], Srivenkataramana and Tracy [19], Srivastava et al. [17, 18]). Our present paper considers certain modification techniques which have been applied to $t_R$ on the availability of $p$ additional supplementary variables to bring about some new composite or generalized ratio-type estimators.

Consider a survey scenario where prior details on $p$ cheaply ascertainable additional supplementary variables $z_1, z_2, ..., z_p$ (may be called as $z$-variables), having strong association with the survey variable $y$ and the main supplementary variable $x$, are readily available. Define $\boldsymbol{z}' = (z_1, z_2, ..., z_p)$ as the vector of $z$-variables and $\overline{\boldsymbol{Z}}' = (\overline{Z}_1, \overline{Z}_2, ..., \overline{Z}_p)$ as the vector of their population means, where $\overline{Z}_j = \frac{1}{N}\sum_{i=1}^{N} z_{ji}$ such that $z_{ji}$ is the observation for $j$th $z$-variable on the $i$th population unit, $j = 1, 2, ..., p$. It is also assumed that population means of all $z$-variables are known precisely.

In the considered two-phase sampling framework, $s_1$ is used to gather information on $x$ and all $z$-variables whereas $s_2$ on $y$ only. Of course, plausibility of an estimate of $\overline{X}$ based on $s_1$ depends on the correlation strengths between $x$ and $z$-variables. Further, we define

$$\overline{\boldsymbol{z}}_{\boldsymbol{1}}' = (\bar{z}_{11}, \bar{z}_{12}, ..., \bar{z}_{1p}) \text{ and } \overline{\boldsymbol{z}}_{\boldsymbol{2}}' = (\bar{z}_{21}, \bar{z}_{22}, ..., \bar{z}_{2p}),$$

where $\bar{z}_{1j} = \frac{1}{n_1}\sum_{i \in s_1} z_{ji}$ and $\bar{z}_{2j} = \frac{1}{n_2}\sum_{i \in s_2} z_{ji}$, $j = 1, 2, ..., p$.

As an example of the foregoing circumstances, we may think of the estimation of mean value of the earning/profit $(y)$ of a group of companies for a given time period taking sales $(x)$ as the main supplementary variable; and working capital $(z_1)$, total assets $(z_2)$, total liabilities $(z_3)$ and market value equity $(z_4)$ as additional supplementary variables. Here it is easily acceptable that accurate value of $\overline{X}$ may not be available but those values for $\overline{Z}_1$, $\overline{Z}_2$, $\overline{Z}_3$ and $\overline{Z}_4$ may be readily available or can be computed easily from the available records.

In the presence of one $z$-variable $z_1$, replacing $\bar{x}_1$ by the ratio estimator $\bar{x}_1\bar{Z}_1/\bar{z}_{11}$ and regression estimator $\bar{x}_1 - b_{xz(1)}(\bar{z}_{11} - \bar{Z}_1)$ of $\bar{X}$ in $t_R$, Chand [2] and Kiregyera [7], respectively, suggested ratio-in-ratio and regression-in-ratio estimators for $\bar{Y}$ defined by $t_{RR} = \bar{y}_2 \frac{\bar{x}_1}{\bar{x}_2}\frac{\bar{z}_1}{\bar{z}_{11}}$ and

$$t_{RGR} = \bar{y}_2 \frac{[\bar{x}_1 - b_{xz_1(1)}(\bar{z}_{11} - \bar{Z}_1)]}{\bar{x}_2},$$

where $b_{xz_1(1)} = \frac{\sum_{i\in s_1}(x_i - \bar{x}_1)(z_{1i} - \bar{z}_{11})}{\sum_{i\in s_1}(z_{1i} - \bar{z}_{11})^2}$ is the sample regression coefficient of $x$ on $z$ for $s_1$. Similarly, expecting the product estimator $\bar{x}_1\bar{z}_{11}/\bar{Z}_1$ as an alternative to $\bar{x}_1$, a product-in-ratio estimator $t_{PR} = \bar{y}_2 \frac{\bar{x}_1}{\bar{x}_2}\frac{\bar{z}_{11}}{\bar{Z}_1}$ can also be considered. But, more generally, replacement of $\bar{x}_1$ by a difference estimator of the form $\bar{x}_1 - d(\bar{z}_{11} - \bar{Z}_1)$ yields a generalized estimator for $\bar{Y}$ defined by

$$t = \bar{y}_2 \frac{[\bar{x}_1 - d(\bar{z}_{11} - \bar{Z}_1)]}{\bar{x}_2},$$

that gives rise to a family or a system of estimators so that $t_R$, $t_{RR}$, $t_{PR}$ and $t_{RGR}$ are its specific cases when the coefficient $d = 0, \bar{x}_1/\bar{z}_{11}, -\bar{x}_1/\bar{Z}_1$ and $b_{xz_1(1)}$, respectively.

Sahoo et al. [15] developed a general framework for the adequate use of available information on $z_1$. The intention was to achieve increased precision over $t_R$ considering certain modification over Chand-Kiregyera approach. The theory developed behind this philosophy may be called as a *Modified Approach* explained below:

As said above, the main reason behind selections of alternative estimators of $\overline{X}$ over $\bar{x}_1$ in the construction of $t_{RR}$, $t_{PR}$, $t_{RGR}$ and $t$ is that the estimators are better than $\bar{x}_1$ from the efficiency point of view under certain situations. But, as argued in Sahoo et al. [15], $\bar{x}_2$ is less efficient estimate than $\bar{x}_1$ for estimating $\overline{X}$. This idea encouraged the authors to formulate a modified approach over Chand-Kiregyera approach that selects an alternative estimator for $\bar{x}_2$ engaging additional covariate $z_1$. But, to generalize the estimation methodology, a difference estimator $\bar{x}_2 - \omega(\bar{z}_{12} - \overline{Z}_1)$ was preferred in place of $\bar{x}_2$, and at the same time motivated by Chand-Kiregyera approach another difference estimator $\bar{x}_1 - \eta(\bar{z}_{11} - \overline{Z}_1)$ in place of $\bar{x}_1$ to define a generalized ratio-type estimator of the form

$$t^{(G)} = \bar{y}_2 \frac{\bar{x}_1 - \eta(\bar{z}_{11} - \overline{Z}_1)}{\bar{x}_2 - \omega(\bar{z}_{12} - \overline{Z}_1)}.$$

Here, as our primary objective is to develop more improved estimators than $t_R$ based on the measured values of $(x_i, z_{ji}, j = 1, 2, ..., p)$, $i \in s_1$, it would be more productive to think of multi-variate extension of $t^{(G)}$. At the same time, it would also be more effective to give priority towards multi-variate extension of $t$ defined above. This will promote for analyzing profitability of the proposed modified approach over Chand-Kiregyera approach in the presence of multiple additional auxiliary variables. In what follows, following additional notations have been used.

Let $\boldsymbol{S}_{zz} = \begin{pmatrix} S_{z_1}^2 & \cdots & S_{z_1 z_p} \\ \vdots & \ddots & \vdots \\ S_{z_p z_1} & \cdots & S_{z_p}^2 \end{pmatrix}$ be the variance-covariance matrix of $\boldsymbol{z}$, where

$$S_{z_j}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (z_{ji} - \overline{Z}_j)^2$$

and

$$S_{z_j z_k} = \frac{1}{N-1}\sum_{i=1}^{N}(z_{ji} - \bar{Z}_j)(z_{ki} - \bar{Z}_k), \;\; j \neq k = 1, 2, ..., p.$$

Here $\boldsymbol{S}_{zz}$ is assumed to be symmetric, non-singular and positive definite. Further, let $\boldsymbol{S}'_{yz} = (S_{yz_1}, S_{yz_2}, ..., S_{yz_p})$ and $\boldsymbol{S}'_{xz} = (S_{xz_1}, S_{xz_2}, ..., S_{xz_p})$, where

$$S_{yz_j} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})(z_{ji} - \bar{Z}_j)$$

and

$$S_{xz_j} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})(z_{ji} - \bar{Z}_j), \;\; j = 1, 2, ..., p,$$

are the vectors of covariances between $y$ or $x$ and members of $\boldsymbol{z}$.

Further define $\rho_{yz_j} = S_{yz_j}/S_y S_{z_j}$ and $\rho_{xz_j} = S_{xz_j}/S_x S_{z_j}$, $j = 1, 2, ..., p$ as the simple correlation coefficients between $(y, z_j)$ and $(x, z_j)$, respectively; $\beta_{yz} = \boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{yz}$ and $\beta_{xz} = \boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{xz}$ as the vectors of regression coefficients of $y$ and $x$, respectively, on $\boldsymbol{z}$; and $\rho_{y.z}^2 = \dfrac{\boldsymbol{S}'_{yz}\boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{yz}}{S_y^2}$ as the squared multiple correlation coefficient of $y$ with the members of $\boldsymbol{z}$.

### 2. Formulation of the Generalized Ratio Estimators

Motivated by the above said two approaches, consider following composite estimators of $\bar{Y}$ considering $t_R$ as the base:

$$t_{MR} = \bar{y}_2 \frac{[\bar{x}_1 - \boldsymbol{A}'(\bar{\boldsymbol{z}}_1 - \bar{\boldsymbol{Z}})]}{\bar{x}_2}$$

and

$$t_{MR}^{(G)} = \bar{y}_2 \frac{\bar{x}_1 - \boldsymbol{C}'_1(\bar{\boldsymbol{z}}_1 - \bar{\boldsymbol{Z}})}{\bar{x}_2 - \boldsymbol{C}'_2(\bar{\boldsymbol{z}}_2 - \bar{\boldsymbol{Z}})},$$

where $\boldsymbol{A}' = (A_1, A_2, ..., A_p)$, $\boldsymbol{C_1'} = (C_1, C_2, ..., C_p)$ and $\boldsymbol{C_2'} = (C_{21}, C_{22}, ..., C_{2p})$ are vectors of constants (coefficients). These coefficient vectors are either known constants or random variables. But, normally these quantities are decided keeping in mind that the MSEs of the estimators are minimum. See that for the construction of $t_{MR}$ $\bar{x}_1$ is succeeded by a multi-variate difference estimator $\bar{x}_1 - \boldsymbol{A}'(\bar{\boldsymbol{z}}_{\boldsymbol{1}} - \overline{\boldsymbol{Z}})$ whereas for $t_{MR}^{(G)}$ both $\bar{x}_1$ and $\bar{x}_2$ are simultaneously succeeded by the multi-variate difference estimators $\bar{x}_1 - \boldsymbol{C_1'}(\bar{\boldsymbol{z}}_{\boldsymbol{1}} - \overline{\boldsymbol{Z}})$ and $\bar{x}_2 - \boldsymbol{C_2'}(\bar{\boldsymbol{z}}_{\boldsymbol{2}} - \overline{\boldsymbol{Z}})$. This means that the estimators $t_{MR}$ and $t_{MR}^{(G)}$ are direct multivariate extensions of *t* and $t^{(G)}$, respectively.

Employing Taylor linearization technique, the following asymptotic MSE expressions are obtained:

$$M(t_{MR}) = \theta_2 S_y^2 + (\theta_2 - \theta_1)(R^2 S_x^2 - 2RS_{yx}) + \theta_1(R^2 \boldsymbol{A}'\boldsymbol{S}_{zz}\boldsymbol{A} - 2R\boldsymbol{A}'\boldsymbol{S}_{yz}) \quad (2)$$

and

$$M(t_{MR}^{(G)}) = \theta_2 S_y^2 + (\theta_2 - \theta_1)(R^2 S_x^2 - 2RS_{yx} - 2R^2 \boldsymbol{C_2'}\boldsymbol{S}_{xz})$$
$$+ \theta_2(R^2 \boldsymbol{C_2'}\boldsymbol{S}_{zz}\boldsymbol{C_2} + 2R\boldsymbol{C_2'}\boldsymbol{S}_{yz})$$
$$+ \theta_1(R^2 \boldsymbol{C_1'}\boldsymbol{S}_{zz}\boldsymbol{C_1} - 2R\boldsymbol{C_1'}\boldsymbol{S}_{yz} - 2R^2 \boldsymbol{C_2'}\boldsymbol{S}_{zz}\boldsymbol{C}_1). \quad (3)$$

**2.1. Precision gained by the proposed estimators**

Now, we obtain some useful guidelines for achieving higher precision/efficiency towards estimation of $\overline{Y}$ by (a) $t_{MR}$ and $t_{MR}^{(G)}$ over $t_R$, and (b) $t_{MR}^{(G)}$ over $t_{MR}$. To meet these requirements, the formulas for $M(t_{MR})$ and $M(t_{MR}^{(G)})$ are expressed in the following alternative patterns:

$$M(t_{MR}) = M(t_R) + \theta_1 R\boldsymbol{A}'\boldsymbol{S}_{zz}(R\boldsymbol{A} - 2\beta_{yz}), \quad (4)$$

$$M(t_{MR}^{(G)}) = M(t_R) + \theta_2 R\boldsymbol{C_2'}\boldsymbol{S}_{zz}(R\boldsymbol{C_2} + 2\beta_{yz} - 2R\beta_{xz})$$

$$+ \theta_1 R \boldsymbol{C}_1' \boldsymbol{S}_{zz}(R\boldsymbol{C}_1 - 2\beta_{yz}) + 2\theta_1 R^2 \boldsymbol{C}_2' \boldsymbol{S}_{zz}(\beta_{xz} - \boldsymbol{C}_1). \quad (5)$$

From (4), $M(t_{MR}) < M(t_R)$, i.e., $t_{MR}$ is a competent for giving a noticeable increase of precision over $t_R$ when

$$\boldsymbol{A} < 2\frac{\beta_{yz}}{R}. \quad (6)$$

But, when $R = \beta_{y\boldsymbol{x}}$, an increase in precision over $t_R$ is feasible if

$$A < 2\frac{\beta_{yz}}{\beta_{yx}}. \quad (7)$$

From (5) it seems impossible to extract both necessary and sufficient conditions for which $M(t_{MR}^{(G)}) < M(t_R)$. But, omitting details, the following sufficient conditions are obtained to achieve a significant efficiency gain for $t_{MR}^{(G)}$ over $t_R$:

$$\boldsymbol{C}_2 < \frac{2(R\beta_{xz} - \beta_{yz})}{R} \text{ and } \beta_{xz} < \boldsymbol{C}_1 < 2\frac{\beta_{yz}}{R}. \quad (8)$$

However, for the case $R = \beta_{yx}$, the conditions are simplified to

$$\boldsymbol{C}_2 < \frac{2(\beta_{yx}\beta_{xz} - \beta_{yz})}{\beta_{yx}} \text{ and } \beta_{xz} < \boldsymbol{C}_1 < 2\frac{\beta_{yz}}{\beta_{yx}}. \quad (9)$$

It is also highly essential to examine the situation for which $M(t_{MR}^{(G)}) < M(t_{MR})$. But to make this task more meaningful, assume that $\boldsymbol{C}_1 = \boldsymbol{A}$. Then combining (2) and (3), we have

$$M(t_{MR}^{(G)}) = M(t_{MR}) + \theta_2 R \boldsymbol{C}_2' \boldsymbol{S}_{zz}(R\boldsymbol{C}_2 + 2\beta_{yz} - 2R\beta_{xz})$$

$$+ 2\theta_1 R^2 \boldsymbol{C}_2' \boldsymbol{S}_{zz}(\beta_{xz} - \boldsymbol{A}). \quad (10)$$

This means that $t_{MR}^{(G)}$ would be more efficient than $t_{MR}$, when

$$\boldsymbol{C_2} < \frac{2(R\beta_{xz} - \beta_{yz})}{R} \text{ and } \boldsymbol{A} < \beta_{xz}. \tag{11}$$

However, we may also consider $R = \beta_{yx}$ in the first sufficient condition of (11) for hoping $t_{MR}^{(G)}$ to be more appropriate than $t_{MR}$.

Various superiority conditions found out in this work are of course very arduous for their verification in a specific situation. But, the conditions unveil the fact that there are certain real life situations where our new methodology can be fruitfully launched.

## 3. Optimum Selections of Coefficients for the Proposed Estimators

The optimum value of $\boldsymbol{A}$ computed in the usual manner to minimize $M(t_{MR})$ is

$$\hat{\boldsymbol{A}} = R^{-1}\boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{yz} = R^{-1}\beta_{yz}. \tag{12}$$

Utilization of this optimum value gives the minimum value of $M(t_{MR})$, i.e., the minimum MSE bound of $t_{MR}$ as

$$M_{\min}(t_{MR}) = M(t_R) - \theta_1\boldsymbol{S}'_{yz}\boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{yz} = M(t_R) - \theta_1 S_y^2\rho_{y.\boldsymbol{z}}^2, \tag{13}$$

and the optimum, i.e., minimum MSE bound estimator of $t_{MR}$ as

$$\hat{t}_{MR} = \bar{y}_2 \frac{[\bar{x}_1 - R^{-1}\beta'_{yz}(\bar{\boldsymbol{z}}_{\mathbf{1}} - \bar{\boldsymbol{Z}})]}{\bar{x}_2}.$$

For the case of single additional auxiliary variable $z_1$, as is expected,

$$\hat{t}_{MR} \to \hat{t} = \bar{y}_2 \frac{\left[\bar{x}_1 - \frac{\beta_{yz_1}}{R}(\bar{z}_{11} - \bar{Z}_1)\right]}{\bar{x}_2},$$

and

$$M_{\min}(t_{MR}) \to M_{\min}(t) = M(\hat{t}) = M(t_R) - \bar{Y}^2\theta_1 C_y^2\rho_{yz_1}^2. \tag{14}$$

Following conventional optimization procedure, optimum values of $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ for minimizing $M(t_{MR}^{(G)})$ given in (5) are determined as

$$\hat{\boldsymbol{C}}_1 = \boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{xz} = \boldsymbol{\beta}_{xz} \tag{15}$$

and

$$\hat{\boldsymbol{C}}_2 = \boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{xz} - R^{-1}\boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{yz} = \beta_{xz} - R^{-1}\beta_{yz}. \tag{16}$$

After considerable simplification, the minimum MSE bound of $t_{MR}^{(G)}$ is obtained as

$$M_{\min}(t_{MR}^{(G)}) = M(t_R) - \theta_2\boldsymbol{S}'_{yz}\boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{yz} - (\theta_2 - \theta_1)R(R\boldsymbol{S}'_{xz}\boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{xz} - 2\boldsymbol{S}'_{xz}\boldsymbol{S}_{zz}^{-1}\boldsymbol{S}_{yz}), \tag{17}$$

$$\Rightarrow M_{\min}(t_{MR}^{(G)}) = M(t_R) - \theta_1 S_y^2\rho_{y.z}^2 - (\theta_2 - \theta_1)(\boldsymbol{S}_{yz} - R\boldsymbol{S}_{xz})'\boldsymbol{S}_{zz}^{-1}(\boldsymbol{S}_{yz} - R\boldsymbol{S}_{xz}). \tag{18}$$

It is noteworthy to point out that the minimum MSE bound of $t_{MR}^{(G)}$ is dependent on the partial correlation coefficient between $y$ and $x$, and multiple correlation coefficient between $y$ and $z$-variables. This is of course not pronounced in the case of using one $z$-variable.

Conclusively, the minimum MSE bound estimator of $t_{MR}^{(G)}$ corresponding to expression (17) or (18) is given by

$$\hat{t}_{MR}^{(G)} = \bar{y}_2 \frac{\bar{x}_1 - \beta'_{xz}(\bar{\boldsymbol{z}}_1 - \bar{\boldsymbol{Z}})}{\bar{x}_2 - (\beta'_{xz} - R^{-1}\beta'_{yz})(\bar{\boldsymbol{z}}_2 - \bar{\boldsymbol{Z}})}.$$

For the use of one additional auxiliary variable $z_1$,

$$\hat{t}_{MR}^{(G)} \rightarrow \hat{t}^{(G)} = \bar{y}_2 \frac{\bar{x}_1 - \beta_{xz_1}(\bar{z}_{11} - \bar{Z}_1)}{\bar{x}_2 - (\beta_{xz_1} - \frac{\beta_{yz_1}}{R})(\bar{z}_{12} - \bar{Z}_1)}$$

and

$$M_{\min}(t_{MR}^{(G)}) = M_{\min}(t^{(G)}) = M(\hat{t}^{(G)})$$

$$= M(t_R) - S_y^2\left[(\theta_2 - \theta_1)\left(\rho_{yz_1} - R\frac{S_x}{S_y}\rho_{xz_1}\right)^2 + \theta_1\rho_{yz_1}^2\right] \quad (19)$$

as given in Sahoo et al. [15].

## 4. Some Technical Dimensions on the Minimum MSE

One of the main causes of deducing expressions for the minimum MSE is to utilize them to compare efficiency of $t_{MR}^{(G)}$ with that of $t_{MR}$, and also efficiencies of both $t_{MR}$ and $t_{MR}^{(G)}$ with that of $t_R$. Keeping this finer point in mind, the following results are extrapolated:

**4.1.** Considering equations (13) and (14), and noting that the multiple correlation coefficient $\rho_{y,z}$ is not less than any of the simple correlation coefficients $\rho_{yz_j}$, $j = 1, 2, ..., p$,

$$M_{\min}(t_{MR}) < M_{\min}(t) < M(t_R).$$

This authenticates that in the fold of Chand-Kiregyera estimation mechanism, choice of $t_{MR}$ that makes use of multiple $z$-variables definitely improves precision in estimation compared to $t$ and $t_R$ using one $z$-variable and no $z$-variable, respectively.

**4.2.** Write expression (18) as

$$M_{\min}(t_{MR}^{(G)}) = M(t_R) - \theta_1 S_y^2 \rho_{y.z}^2 - (\theta_2 - \theta_1)\boldsymbol{B}'\boldsymbol{S}_{zz}^{-1}\boldsymbol{B}, \quad (20)$$

where $\boldsymbol{B}' = (B_1\ B_2 \ldots B_p)$ such that $B_j = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - Rx_i)(z_{ji} - \overline{Z}_j)$, $j = 1, 2, ..., p$. Since $\boldsymbol{S}_{zz}$ is positive definite, $\boldsymbol{S}_{zz}^{-1}$ is also positive definite. Hence, according to Theorem A.1.1 (Anderson [1, p. 628]), the quadratic form $\boldsymbol{B}'\boldsymbol{S}_{zz}^{-1}\boldsymbol{B}$ is positive definite, i.e., $\boldsymbol{B}'\boldsymbol{S}_{zz}^{-1}\boldsymbol{B} \geq 0$. Then, from (13) and (20), it follows that

$$M_{\min}(t_{MR}^{(G)}) < M_{\min}(t_{MR}) < M(t_R),$$

showing that $\hat{t}_{MR}^{(G)}$ is more efficient than both $\hat{t}_{MR}$ and $t_R$. This means that our modified method has more scope for the advancement of estimation compared to Chand-Kiregyera method in the presence of $p$ $z$-variables and classical ratio method with no $z$-variable. This further indicates that the technique worked in the formulation of $t_{MR}^{(G)}$ is definitely more powerful than that worked for $t_{MR}$ under the minimum MSE bound criterion.

**4.3.** One of the important aspects of the study is also comparison of $M_{\min}(t_{MR}^{(G)})$ with $M_{\min}(t^{(G)})$. Because, this will facilitate to examine how the proposed approach on the consideration of $p$ $z$-variables is more profitable than that considers only one $z$-variable. However, to make this unmanageable job smooth, assume that $\rho_{yz_j} = \rho_{yz}$ and $\rho_{xz_j} = \rho_{xz}$, $j = 1, 2, \ldots, p$. Then, (17) simplifies to

$$M_{\min}(t_{MR}^{(G)}) = M(t_R)$$

$$- S_y^2 \boldsymbol{S}_z' \boldsymbol{S}_{zz}^{-1} \boldsymbol{S}_z \left[ (\theta_2 - \theta_1) \left( \rho_{yz} - R \frac{S_x}{S_y} \rho_{xz} \right)^2 + \theta_1 \rho_{yz}^2 \right], \quad (21)$$

where $\boldsymbol{S}_z' = (S_{z_1}, S_{z_2}, \ldots, S_{z_p})$ and $S_{z_j} = +\sqrt{S_{z_j}^2}$.

Comparing (19) and (21), it appears that $\hat{t}_{MR}^{(G)}$ is more precise than $\hat{t}^{(G)}$, i.e., the proposed method with $p$ additional supplementary variables would be better than that with only one additional supplementary variable provided

$$\boldsymbol{S}_z' \boldsymbol{S}_{zz}^{-1} \boldsymbol{S}_z > 1. \quad (22)$$

In survey sampling situations, we may not have any prior information or idea to directly scrutinize this condition. But, in order to gain an idea on this situation, we consider $p = 2$, i.e., involvement of two additional supplementary variables $z_1$ and $z_2$. Then, after simplification, $\boldsymbol{S}_z' \boldsymbol{S}_{zz}^{-1} \boldsymbol{S}_z$ = $2/(1+\rho_{z_1 z_2})$ which is greater than 1 for $-1 < \rho_{z_1 z_2} < 1$, where $\rho_{z_1 z_2}$ is the correlation coefficient between $z_1$ and $z_2$. Of course, when $\rho_{z_1 z_2} = \pm 1$, i.e., relation between $z_1$ and $z_2$ is perfect linear, the said two minimum MSE bound estimators may be equally efficient.

In the light of the preceding theoretical results, it may be concluded that when the $p$ additional supplementary variables are perfectly linearly related, i.e., when there exists perfect multi-collinearity among the $z$-variables, use of $t_{MR}^{(G)}$ in place of $t^{(G)}$ may not be enormously profitable. But, this is a rare phenomenon in the context of a sample survey situation.

## 5. Empirical Study

To validate preceding theoretical findings on the proposed composite estimators, consider data sets of 5 populations each one being comprised of two additional supplementary variables $z_1$ and $z_2$. Table 1 gives brief description of these populations with regard to their source, size and definitions of the variables.

In order to assess and compare relative efficiencies of various generalized estimators, their minimum MSEs only were taken into account because various algebraic/statistical limitations make this operation strenuous to concentrate on any particular estimators of their families. Hence, this empirical study is confined to the MSE bound estimators $\hat{t}$, $\hat{t}^{(G)}$, $\hat{t}_{MR}$ and $\hat{t}_{MR}^{(G)}$ so that the minimum MSE bound is considered as the benchmark of efficiency comparison. Further, to make the study more meaningful, the traditional two-phase ratio estimator $t_R$ is also incorporated in the investigation. Relative efficiencies of the said estimators compared to the single-phase mean per unit estimator $\bar{y}_2$ for 40% and 20% sampling fractions in the first and second phases, respectively, are compiled in Table 2.

**Table 1.** Details of the populations

| Population | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Source | Sarndal et al. [16, p. 662] | Murthy [12, p. 127] | Gujarati [6, p. 794] | Rencher [14, p. 279] | Rencher [14, p. 109] |
| $N$ | 124 countries | 128 villages | 88 observations | 90 persons | 51 workers |
| $Y$ | 1983 exports | no. of cultivators in 1961 | profit | front-to-back measurement at eye level | hemoglobin concentration |
| $x$ | 1982 gross national product | cultivated area in 1951 | personal disposable income | head circumference | packed cell volume |
| $z_1$ | 1983 population | no. of persons in 1961 | gross domestic product | head width at widest dimension | white blood cell count |
| $z_2$ | 1980 population | area in 1951 | dividend | ear-to-top-of-head measurement | neutrophil count |

**Table 2.** Relative efficiencies of the comparable estimators compared to $\bar{y}_2$ (in %)

| Estimator | Supplementary variable(s) used | Population | | | | |
|---|---|---|---|---|---|---|
| | | 1<br>$n_1 = 50$<br>$n_2 = 25$ | 2<br>$n_1 = 51$<br>$n_2 = 25$ | 3<br>$n_1 = 35$<br>$n_2 = 18$ | 4<br>$n_1 = 36$<br>$n_2 = 18$ | 5<br>$n_1 = 20$<br>$n_2 = 10$ |
| $t_R$ | $x$ | 131.54 | 124.18 | 122.68 | 117.45 | 121.18 |
| $\hat{t}$ | $x, z_1$ | 163.68 | 154.32 | 326.55 | 164.58 | 158.09 |
| | $x, z_2$ | 163.82 | 166.77 | 251.91 | 175.62 | 151.45 |
| $\hat{t}^{(G)}$ | $x, z_1$ | 185.66 | 193.92 | 397.97 | 165.25 | 168.11 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | $x$, $z_2$ | 183.20 | 196.94 | 277.44 | 176.02 | 161.48 |
| $\hat{t}_{MR}$ | $x$, $z_1$, $z_2$ | 175.45 | 184.46 | 413.44 | 182.24 | 166.98 |
| $\hat{t}_{MR}^{(G)}$ | $x$, $z_1$, $z_2$ | 204.71 | 222.87 | 456.09 | 195.65 | 189.33 |

It is seen from Table 2 that all minimum MSE bound estimators achieve considerably high efficiency gains over $t_R$. Amongst minimum MSE bound estimators, $\hat{t}_{MR}^{(G)}$ comes out as the most efficient, either $\hat{t}_{MR}$ or $\hat{t}^{(G)}$ turns out as the second most efficient, and $\hat{t}$ appears to be the worst performer. Although in many cases $\hat{t}_{MR}$ stands as the second best performer, in some cases, it is less efficient compared to $\hat{t}^{(G)}$. This discouraging result for $\hat{t}_{MR}$ shows that Chand-Kiregyera method with more than one supplementary variable may not always compete with our suggested modified method with only one additional supplementary variable. On the whole, the empirical findings confirm superiority of $t_{MR}^{(G)}$ over other comparable estimators.

## 6. A Useful Remark

It also appears to appreciate and consider a composite estimator of the form

$$t_{MR}^{(AG)} = \bar{y}_2 \frac{\bar{x}_1 - \boldsymbol{G}_1'(\bar{\boldsymbol{z}}_1 - \bar{\boldsymbol{Z}})}{\bar{x}_2 - \boldsymbol{G}_2'(\bar{\boldsymbol{z}}_2 - \bar{\boldsymbol{z}}_1)}$$

alternative to $t_{MR}^{(G)}$ originated after substitutions of $\bar{x}_2 - \boldsymbol{G}_2'(\bar{\boldsymbol{z}}_2 - \bar{\boldsymbol{z}}_1)$ and $\bar{x}_1 - \boldsymbol{G}_1'(\bar{\boldsymbol{z}}_1 - \bar{\boldsymbol{Z}})$ for $\bar{x}_2$ and $\bar{x}_1$ in $t_R$. Note that $t_{MR}^{(G)}$ and $t_{MR}^{(AG)}$ make two classes of estimators for $\bar{Y}$ which are not necessarily disjoint. But, their minimum MSE bounds are equal whereas their minimum MSE bound estimators are different. Because of these similarities, the estimators $t_{MR}^{(G)}$ and $t_{MR}^{(AG)}$ are not considered as good competitors.

## References

[1] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, 3rd ed., Wiley-Interscience, 2003.

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

[3] W. G. Cochran, Sampling Techniques, 3rd ed., Wiley Eastern Limited, 1977.

[4] S. K. Dubey, M. Kumar, V. N. Rai and A. Tiwari, A composite class of estimators for a finite population mean using two auxiliary variables in two-phase sampling, The Pharma Innovation Journal 11(4) (2022), 44-49.

[5] S. K. Dubey, B. V. S. Sisodia and M. K. Sharma, Some transformed and composite chain ratio-type estimators using two auxiliary variables, Journal of the Indian Society of Agricultural Statistics 74(1) (2020), 17-22.

[6] D. N. Gujarati, Basic Econometrics, 4th ed., McGraw-Hill, 2003.

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

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

[9] M. Kumar and A. Tiwari, A generalized class of estimators of population mean in two-phase sampling using two auxiliary variables, International Journal of Mathematics and Statistics 22(3) (2021), 18-27.

[10] M. Kumar and G. K. Vishwakarma, Estimation of mean in double sampling using exponential technique on multi-auxiliary variates, Communications in Mathematics and Statistics 5 (2017), 429-445.

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

[12] M. N. Murthy, Sampling Theory and Methods, Statistical Publishing Society, Calcutta, 1977.

[13] P. A. Patel and F. H. Shah, Two-phase ratio-type estimator of a finite population mean, International Journal of Scientific Research in Mathematical and Statistical Sciences 5(5) (2018), 199-203.

[14] A. C. Rencher, Methods of Multivariate Analysis, Second ed., Wiley-Interscience, 2002.

[15] R. K. Sahoo, Ajit Kumar Sabat and L. N. Sahoo, Generalized ratio estimators in two-phase sampling with two auxiliary variables, International Journal of Applied Mathematics and Statistical Sciences 11(2) (2022), 13-30.

[16] C. E. Sarndal, B. Swensson and J. Wretman, Model Assisted Survey Sampling, Springer-Verlag, 2003.

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

[18] S. Rani Srivastava, B. B. Khare and S. R. Srivastava, A generalized chain ratio estimator for mean of finite population, Journal of the Indian Society of Agricultural Statistics 42 (1990), 108-117.

[19] T. Srivenkataramana and D. S. Tracy, Two-phase sampling for selection with probability proportional to size in sample surveys, Biometrika 76 (1989), 818-821.

[20] B. V. Sukhatme and L. Chand, Multivariate ratio type estimator, Proceedings of the Social Statistics Section, American Statistical Association, 1977, pp. 927-931.