A general class of estimators for the population mean using multi-phase sampling with the non-respondents

Saba Riaz[*][†], Giancarlo Diana[‡] and Javid Shabbir[§]

Abstract

This paper addresses the problem of non-response when estimating the population mean of the variable of interest. A general class of estimators is suggested for the unknown population mean of the study variable. Two vectors of the auxiliary information under multi-phase sampling scheme are used in presence of non response. The asymptotic variance of the proposed class is determined and compared with some other existing estimators theoretically and numerically. It is shown that the proposed class of estimators is more efficient than [13] and [17] classes of estimators.

Keywords: Auxiliary information, Multi-phase sampling, Non-response, Asymptotic variance, Efficiency.

2000 AMS Classification: 62D05

1. Introduction

While collecting information through the sample surveys there may arise several problems, one of the common problems is non-response. This happens especially in the surveys conducted through mails. When research participants can not be approached directly, they may refuse to acknowledge survey questionnaires sent to them through mails. The estimates obtained from such incomplete surveys are often biased. The data obtained from respondents group differs from that of non-respondents group and thus affects data reliability. [6] suggested non-respondents sub sampling scheme to handle this problem. In this scheme, initially the information is collected from the respondent's group through mail survey then non-respondents are re-contacted by using sub sampling process and complete

---

[*]Department of Statistics, Quaid-i-Azam University, Islamabad (Pakistan), Email: sabaqau@gmail.com

[†]Corresponding Author.

[‡]Department of Statistical Sciences, University of Padova (Italy), Email: diana@stat.unipd.it

[§]Department of Statistics, Quaid-i-Azam University, Islamabad (Pakistan), Email: javidshabbir@gmail.com

information is retrieved through personal interviews. The main objective in sampling theory is to estimate the unknown population parameter of interest, and the information about population characteristic of the auxiliary variables may or may not be available. Of course the use of the auxiliary information increases the efficiency of the estimates of the population parameter. Many authors have proposed different estimators to estimate the population mean using the auxiliary information with or without considering non-response see [19], [7], [14], [12], [9], [13], [20], [8], [4], [16, 17], [10], [3] and references cited therein.

When information about the population characteristic of the auxiliary variable (X) is unavailable, in such situation it is estimated from a large first phase sample drawn from the population and then a smaller second phase sample is taken from the first phase sample to estimate the population parameter of the variable of interest (Y). [10] has considered regression-cum-ratio estimator using single auxiliary variable in two phase sampling scheme in presence of non-response. [17] have proposed regression-cum-ratio estimators considering different situations using two auxiliary variables in two-phase sampling scheme in presence of non-response.

Multi-phase sampling is very useful scheme in a situation when the variable of interest is very expensive and, connected with other cheaper auxiliary variables. Limited literature is available about this technique see for instance [18], [11], [1], [2] and [15].

This paper is inspired by the previous studies and the main objective is to propose a general class of estimators to estimate the unknown population mean $\bar{Y}$ under multi-phase sampling scheme using a vector $\underline{X} = (x_1, \ldots, x_k)^t$ with unknown population mean vector $\bar{\underline{X}} = (\bar{X}_1, \ldots, \bar{X}_k)^t$ and another vector $\underline{Z} = (z_1, \ldots, z_k)^t$ with known or unknown population mean vector $\bar{\underline{Z}} = (\bar{Z}_1, \ldots, \bar{Z}_k)^t$ in the presence of non-response. This research will also explore that the optimal estimators proposed in the generalized class are regression type estimators.

2. Notations and Background

Let us assume that P be a finite population of N distinct units. Let Y and X be the study and the auxiliary variables having values y_i and x_i, $i = (1, \ldots, N)$. Let X is correlated with Y and is used to estimate the unknown population mean $\bar{Y}$. When the mean of the auxiliary variable X is available, the ratio, product and regression estimators are used to increase the efficiency of the estimates of $\bar{Y}$. When $\bar{X}$ is unknown, two-phase sampling scheme is used, at first phase only $\bar{X}$ is estimated and the second phase is devoted for the estimation of $\bar{Y}$.

Let a large sample s' of size m_1 $(m_1 < N)$ is drawn by simple random sampling without replacement (SRSWOR) to collect information on the auxiliary variable X. It is assumed that all m_1 units provide complete information on X. In the second phase, a smaller sample s of size m_2 from m_1 units $(m_2 < m_1)$ is drawn by SRSWOR for obtaining information of the study variable Y. Suppose that non-response is present in second phase, in this situation, a subset s_1 of size m_2' supplies information on Y and the remaining $m_2^* = m_2 - m_2'$ units are non-respondents. Therefore, following the familiar technique of [6], a sub-sample s_{2r} of size $r = m_2^*/b$, $b > 1$ is selected from the m_2^* non-response units where r would be an integer otherwise must be rounded. Assuming that all r selected units show

full response on second call. Consequently, the whole population is said to be stratified into two strata P_1 and P_2, where P_1 is the stratum of respondents of size N_1 that would give response on first call at second phase whereas P_2 is the stratum of non-respondents of size N_2 which would not respond on first call at second phase but will cooperate on the second call. Obviously N_1 and N_2 are not known in advance.

Now we can define a dummy variable $u = (y, x, z)$ in the following presentation

$$\bar{u}^* = d_1 \bar{u}_1 + d_2 \bar{u}_{2r},$$

where

$$\bar{u}' = \frac{\sum_{i=1}^{m_1} u_i}{m_1}, \quad \bar{u} = \frac{\sum_{i=1}^{m_2} u_i}{m_2}, \quad \bar{u}_1 = \frac{\sum_{i=1}^{m_2'} u_i}{m_2'}, \quad \bar{u}_{2r} = \frac{\sum_{i=1}^{r} u_i}{r}$$

and $u_i = (y_i, x_i, z_i)$ having i^{th} value. Also

$$d_1 = \frac{m_2'}{m_2} \quad \text{and} \quad d_2 = \frac{m_2^*}{m_2}.$$

Similarly we have

$$\bar{U} = D_1 \bar{U}_1 + D_2 \bar{U}_2,$$

where

$$\bar{U}_1 = \frac{\sum_{i=1}^{N_1} u_i}{N_1}, \quad \bar{U}_2 = \frac{\sum_{i=1}^{N_2} u_i}{N_2}, \quad D_1 = \frac{N_1}{N} \quad \text{and} \quad D_2 = \frac{N_2}{N}.$$

The variance of $\bar{u}^*$ is given by

(2.1) $\quad \mathrm{Var}(\bar{u}^*) = \theta_2 S_u^2 + \omega_2 S_{u(2)}^2 = \widetilde{S}_u^2,$

where

$$S_u^2 = \frac{\sum_i^N (u_i - \bar{U})^2}{N-1}, \quad S_{u(2)}^2 = \frac{\sum_i^{N_2} (u_i - \bar{U}_2)^2}{N_2 - 1},$$

$$\theta_1 = \left(\frac{1}{m_1} - \frac{1}{N}\right), \quad \theta_2 = \left(\frac{1}{m_2} - \frac{1}{N}\right) \quad \text{and} \quad \omega_2 = \frac{N_2(b-1)}{m_2 N}.$$

One can define the covariance as

(2.2) $\quad \mathrm{Cov}(\bar{u}^*, \bar{v}^*) = \theta_2 S_{uv} + \omega_2 S_{uv(2)} = \widetilde{S}_{uv},$

where

$$S_{uv} = \frac{\sum_i^N (u_i - \bar{U})(v_i - \bar{V})}{N-1} \quad \text{and} \quad S_{uv(2)} = \frac{\sum_i^{N_2} (u_i - \bar{U}_2)(v_i - \bar{V}_2)}{N_2 - 1}, \quad (u = y, v = (x, z)).$$

From now on, we shall consider that the Asymptotic Variance (AV) of the considered estimators obtained by using a first order Taylor series [21].

[6] proposed an estimator $\bar{y}^*$ for the population mean $\bar{Y}$ when non response occurs

(2.3) $\quad \bar{y}^* = d_1 \bar{y}_1 + d_2 \bar{y}_{2r}.$

[13] proposed the following regression estimator using double sampling scheme when non-response occurs on Y and X both

(2.4) $\quad t_{\mathrm{OL}} = \bar{y}^* + \hat{\beta}_{yx}^* \left(\bar{x}' - \bar{x}^*\right),$

where $\hat{\beta}^*_{yx} = \dfrac{s^*_{yx}}{s^{*2}_x}$ is a sample estimate of the population regression coefficient $\beta_{yx} =$

$\dfrac{S_{yx}}{S^2_x}$ with $s^*_{yx} = \dfrac{\sum_{i=1}^{m'_2} y_i x_i + b \sum_{i=1}^{r} y_i x_i - m_2 \bar{y}^* \bar{x}^*}{m_2 - 1}$ and $s^{*2}_x = \dfrac{\sum_{i=1}^{m'_2} x_i^2 + b \sum_{i=1}^{r} x_i^2 - m_2 \bar{x}^{*2}}{m_2 - 1}$.

The asymptotic variance of t_{OL} is obtained by using the Taylor linearization and is given by

$$\mathrm{AV}(t_{\mathrm{OL}}) \cong \theta_1 S^2_y + (\theta_2 - \theta_1) S^2_y \left(1 - \rho^2_{yx}\right) + \omega_2 \left\{ S^2_{y(2)} + \beta^2_{yx} S^2_{x(2)} - 2\beta_{yx} S_{yx(2)} \right\},$$

where $\rho^2_{yx} = \dfrac{S^2_{yx}}{S^2_y S^2_x}$.

[10] proposed the following regression-cum-ratio estimator using [13] regression estimator

$$(2.5) \quad t_{\mathrm{K}} = t_{\mathrm{OL}} \left(\dfrac{a\bar{x}^* + b}{a\bar{x}' + b} \right)^{\alpha} \left(\dfrac{a\bar{x}' + b}{a\bar{x} + b} \right)^{\beta},$$

where (a, b) are known constants and (α, β) are suitably chosen constants.

The minimum asymptotic variance of t_{K} is given by

$$\mathrm{minAV}(t_{\mathrm{K}}) \cong \theta_1 S^2_y + (\theta_2 - \theta_1) S^2_y \left(1 - \rho^2_{yx}\right) + \omega_2 S^2_{y(2)} \left(1 - \rho^2_{yx(2)}\right),$$

where $\rho^2_{yx(2)} = \dfrac{S^2_{yx(2)}}{S^2_{y(2)} S^2_{x(2)}}$.

[17] considered two auxiliary variables x and z in four different situations. We consider only two of those situations which are in our opinion the most interesting and related to our proposed class.

Situation I: $- \bar{X}$ unknown and $\bar{Z}$ known

When the population mean of the first auxiliary variable x is unknown and the population mean of the second auxiliary variable z is known. Also non-response occurs on both the study and the auxiliary variables. The proposed estimator is given by

$$(2.6) \quad t_{\mathrm{SK}(1)} = \left[\bar{y}^* + \hat{\beta}^*_{yx}(\bar{x}' - \bar{x}^*) \right] \dfrac{\bar{Z}}{\bar{Z} + \alpha(\bar{z}^* - \bar{Z})},$$

where α is a suitably chosen constant.

The minimum asymptotic variance of the estimator $t_{\mathrm{SK}(1)}$ is given by

$$\mathrm{minAV}(t_{\mathrm{SK}(1)}) \cong \mathrm{AV}(t_{\mathrm{OL}}) - \dfrac{M^{*2}}{D^*},$$

where

$$M^* = \{\theta_2 S^2_z \beta_{yz} + \omega_2 S^2_{z(2)} \beta_{yz(2)}\} - \beta_{yx}\{(\theta_2 - \theta_1) S^2_z \beta_{xz} + \omega_2 S^2_{z(2)} \beta_{xz(2)}\},$$

$$D^* = \theta_2 S^2_z + \omega_2 S^2_{z(2)}.$$

Situation II: $- \bar{X}$ and $\bar{Z}$ both unknown

When the population means $\bar{X}$ and $\bar{Z}$ are unknown. Also non-response occurs on the study and the auxiliary variables. The estimator is given by

$$(2.7) \quad t_{\mathrm{SK}(2)} = \left[\bar{y}^* + \hat{\beta}_{yx}^*(\bar{x}' - \bar{x}^*)\right] \frac{\bar{z}'}{\bar{z}' + \gamma(\bar{z}^* - \bar{z}')},$$

where γ is a suitably chosen constant.

The minimum asymptotic variance of $t_{\mathrm{SK}(2)}$ is given by

$$\mathrm{minAV}(t_{\mathrm{SK}(2)}) \cong \mathrm{AV}(t_{\mathrm{OL}}) - \frac{N^{*2}}{D^*},$$

where

$$N^* = \left\{(\theta_2 - \theta_1)\, S_z^2 \beta_{yz} + \omega_2 S_{z(2)}^2 \beta_{yz(2)}\right\} - \beta_{yx} \left\{(\theta_2 - \theta_1)\, S_z^2 \beta_{xz} + \omega_2 S_{z(2)}^2 \beta_{xz(2)}\right\}.$$

It is important to note that our proposed general class is an extension of regression estimator. We consider specifically regression estimators because our consideration is to show that regression estimator(s) perform better than regression-cum-ratio estimator(s). Therefore, we expect that our proposed class of estimators perform better than [17] class.

3. Multi-Phase Scheme

By generalizing the multi-phase sampling scheme proposed by [2], we construct a sampling design with two vectors $\underline{X} = (x_1, \ldots, x_k)^t$ and $\underline{Z} = (z_1, \ldots, z_k)^t$ of auxiliary variables in such a way that at each ith and $(i+1)^{th}$ phase, the samples s_i and s_{i+1} of sizes $m_i\,(m_i < N)$ and $m_{i+1}\,(m_{i+1} < m_i)$ are drawn by SRSWOR. At the ith phase, the variables x_i and z_i are observed while at last phase, the auxiliary variables as well as y are measured, according to Table 1.

Table 1. Suggested multi-phase sampling design

1	2	3	. . .	i	. . .	k	$(k+1)$
m_1	m_2	m_3	. . .	m_i	. . .	m_k	m_{k+1}
(X_1, Z_1)	(X_1, Z_1)						
	(X_2, Z_2)	(X_2, Z_2)					
		(X_3, Z_3)	$\ddots$				
				(X_{i-1}, Z_{i-1})			
				(X_i, Z_i)	$\ddots$		
						(X_{k-1}, Z_{k-1})	
						(X_k, Z_k)	(X_k, Z_k)
							Y

4. General Class of Estimators

Using this multi phase sampling design, we propose a general class of estimators for the population mean $\bar{Y}$ when non response can occur on the study variable as well as on the auxiliary vectors.

$$(4.1) \quad t_{kk} = g\left(\bar{y}^*, \mathbf{w}^t\right),$$

where

$$\mathbf{w} = \left(\bar{x}_1^{(1)}, \bar{z}_1^{(1)}, \widetilde{x}_1^{(2)}, \widetilde{z}_1^{(2)}, \dots, \bar{x}_k^{(k)}, \bar{z}_k^{(k)}, \widetilde{x}_k^{(k+1)}, \widetilde{z}_k^{(k+1)}\right)^t$$

and g is a function satisfying the following regularity conditions

C1: $g : \mathcal{S} \to \mathbb{R}$ where $\mathcal{S} \subseteq \mathbb{R}^{4k+1}$ is a convex and bounded set containing the point $\left(\bar{Y}, \mathbf{W}^t\right)$, with $\mathbf{W} = \mathbb{E}\left(\mathbf{w}\right) = \left(\bar{X}_1, \bar{X}_1, \bar{Z}_1, \bar{Z}_1, \dots, \bar{X}_k, \bar{X}_k \bar{Z}_k, \bar{Z}_k\right)^t$;

C2: it is continuous and bounded in $\mathcal{S}$;

C3: its first and second partial derivatives exist and are continuous and bounded in $\mathcal{S}$;

C4: $g\left(\bar{y}^*, \mathbf{W}^t\right) = \bar{y}^*$.

In order to determine the minimum asymptotic variance (AV) of the class t_{kk}, let us indicate

$$g_0 = \left.\frac{\partial g\left(\bar{y}^*, \mathbf{w}^t\right)}{\partial \bar{y}^*}\right|_{\left(\bar{y}^*, \mathbf{w}^t\right) = \left(\bar{Y}, \mathbf{W}^t\right)}$$

and

$$g_i = \left.\frac{\partial g\left(\bar{y}^*, \mathbf{w}^t\right)}{\partial w_i}\right|_{\left(\bar{y}^*, \mathbf{w}^t\right) = \left(\bar{Y}, \mathbf{W}^t\right)},$$

the first partial derivatives of g with respect to the component $\bar{y}^*$ and w_i, $i = 1, 2, \dots, k$, of the vector $\mathbf{w}$.

Expanding t_{kk} at the point $\left(\bar{Y}, \mathbf{W}^t\right)$ in a first order Taylor's series, we get

$$
\begin{aligned}
(4.2) \quad t_{kk} \cong{} & \bar{y}^* + \sum_{i=1}^{k} \eta_i \left(\bar{X}_i - \bar{x}_i^{(i)}\right) + \sum_{i=1}^{k} \xi_i \left(\bar{X}_i - \widetilde{x}_i^{(i+1)}\right) \\
& + \sum_{i=1}^{k} \phi_i \left(\bar{Z}_i - \bar{z}_i^{(i)}\right) + \sum_{i=1}^{k} \varphi_i \left(\bar{Z}_i - \widetilde{z}_i^{(i+1)}\right),
\end{aligned}
$$

where first k indicates the general class and second k for having k auxiliary variables in the estimator t_{kk}.

Further, consider $\widetilde{u}_i = \delta_u \bar{u}_i^* + \bar{\delta}_u \bar{u}_i$ with $\bar{\delta}_u = 1 - \delta_u$ where δ_u is an indicator function taking value $\delta_u = 1$ when non-response occurs and 0 otherwise, and $u_i = (x_i, z_i)$, $i = 1, \dots, k$.

4.1. Situation 1: $\bar{X}$ unknown and $\bar{Z}$ known. In this situation it is assumed that the population mean vector $\bar{X}$ is unknown but the mean vector $\bar{Z}$ is known. Since $\bar{X}$ is unknown so it is necessary to impose a constraint $\eta_i = -\xi_i$ and for computational reasons it may be useful to set $(\phi_i + \varphi_i) = \psi_i$. We can write Eq.(4.2) as

$$(4.3) \quad t_{1k} \cong \bar{y}^* + \sum_{i=1}^{k} \xi_i \left(\bar{x}_i^{(i)} - \widetilde{x}_i^{(i+1)} \right) + \sum_{i=1}^{k} \varphi_i \left(\bar{z}_i^{(i)} - \widetilde{z}_i^{(i+1)} \right) + \sum_{i=1}^{k} \psi_i \left(\bar{Z}_i - \bar{z}_i^{(i)} \right).$$

We can write (4.3) in a generalized vector form as

$$(4.4) \quad t_{1k} \cong \bar{y}^* + \left(\underline{\bar{\nu}}' - \underline{\widetilde{\nu}} \right)^t \underline{\xi} + \left(\underline{\bar{Z}} - \underline{\bar{z}}' \right) \underline{\psi},$$

where

$$\underline{\bar{\nu}}' = \left(\bar{x}_1^{(1)}, \bar{z}_1^{(1)}, \dots, \bar{x}_k^{(k)}, \bar{z}_k^{(k)} \right)^t,$$

$$\underline{\widetilde{\nu}} = \left(\widetilde{x}_1^{(2)}, \widetilde{z}_1^{(2)}, \dots, \widetilde{x}_k^{(k+1)}, \widetilde{z}_k^{(k+1)} \right)^t,$$

$$\underline{\xi} = \left(\xi_1, \varphi_1, \dots, \xi_k, \varphi_k \right)^t,$$

$$\underline{\bar{z}}' = \left(\bar{z}_1^{(1)}, \dots, \bar{z}_k^{(k)} \right)^t$$

and

$$\underline{\psi} = \left(\psi_1, \dots, \psi_k \right)^t.$$

The asymptotic variance of the proposed class t_{1k}

$$(4.5) \quad \text{AV}(t_{1k}) \cong \text{Var}(\bar{y}^*) + \underline{\xi}^t \underline{\widetilde{\mathbf{S}}}_{xz} \underline{\xi} + \underline{\psi}^t \underline{\bar{S}}_{\bar{z}\bar{z}} \underline{\psi} - 2 \left(\underline{\xi}^t \underline{\widetilde{\mathbf{S}}}_{yxz} + \underline{\psi}^t \underline{\bar{S}}_{y\bar{z}} - \underline{\xi}^t \underline{\bar{S}}_{xz\bar{z}} \underline{\psi} \right),$$

where

$$\text{Var}(\bar{y}^*) = \widetilde{S}_y^2 = \theta_{k+1} S_y^2 + \omega_{k+1} S_{y(2)}^2,$$

$$\underline{\widetilde{\mathbf{S}}}_{xz} = E\left[(\underline{\bar{\nu}}' - \underline{\widetilde{\nu}})(\underline{\bar{\nu}}' - \underline{\widetilde{\nu}}^t \right] = \begin{bmatrix} \underline{\widetilde{S}}_{xx} & \underline{\widetilde{S}}_{xz} \\ \underline{\widetilde{S}}_{xz} & \underline{\widetilde{S}}_{zz} \end{bmatrix},$$

$$\underline{\widetilde{\mathbf{S}}}_{yxz} = E\left[(\bar{y}^* - \bar{Y})(\underline{\bar{\nu}}' - \underline{\widetilde{\nu}}) \right] = \begin{bmatrix} \underline{\widetilde{S}}_{yx} & \underline{\widetilde{S}}_{yz} \end{bmatrix}^t,$$

$$\underline{\bar{S}}_{\bar{z}\bar{z}} = E\left[(\underline{\bar{z}}' - \underline{\bar{Z}})(\underline{\bar{z}}' - \underline{\bar{Z}})^t \right],$$

$$\underline{\bar{S}}_{y\bar{z}} = E\left[(\bar{y}^* - \bar{Y})(\underline{\bar{z}}' - \underline{\bar{Z}}) \right] = \begin{bmatrix} \theta_1 S_{yz_1} & \theta_2 S_{yz_2} & \dots & \theta_k S_{yz_k} \end{bmatrix}^t,$$

and

$$\underline{\bar{S}}_{xz\bar{z}} = \begin{bmatrix} \underline{S}_{x\bar{z}} & \underline{S}_{z\bar{z}} \end{bmatrix}^t.$$

The details of above mentioned matrices are given in Appendix A.

Minimizing $\text{AV}(t_{1k})$ w.r.t $\underline{\xi}$ and $\underline{\psi}$ leads to optimum vector

$$\underline{\psi} = \left\{ \underline{\bar{S}}_{\bar{z}\bar{z}} - \underline{\bar{S}}_{xz\bar{z}}^t \underline{\widetilde{\mathbf{S}}}_{xz}^{-1} \underline{\bar{S}}_{xz\bar{z}} \right\}^{-1} \left\{ \underline{\bar{S}}_{y\bar{z}} - \underline{\bar{S}}_{xz\bar{z}} \underline{\widetilde{\mathbf{S}}}_{xz}^{-1} \underline{\widetilde{\mathbf{S}}}_{yxz} \right\} = \underline{\psi}^{(o)} \text{(say)},$$

and

$$\underline{\xi} = \underline{\widetilde{\mathbf{S}}}_{xz}^{-1} \left(\underline{\widetilde{\mathbf{S}}}_{yxz} - \underline{\bar{S}}_{xz\bar{z}} \left\{ \underline{\bar{S}}_{\bar{z}\bar{z}} - \underline{\bar{S}}_{xz\bar{z}}^t \underline{\widetilde{\mathbf{S}}}_{xz}^{-1} \underline{\bar{S}}_{xz\bar{z}} \right\}^{-1} \left\{ \underline{\bar{S}}_{y\bar{z}} - \underline{\bar{S}}_{xz\bar{z}} \underline{\widetilde{\mathbf{S}}}_{xz}^{-1} \underline{\widetilde{\mathbf{S}}}_{yxz} \right\} \right) = \underline{\xi}^{(o)} \text{(say)}.$$

Since it is not easy to express minimum AV (t_{1k}) in close form, in the following sub sections, we can express the class t_{1k} in two sub classes $t_{1k(1)}$ and $t_{1k(2)}$ with two and three phases.

4.1.1. *Two Phase.* Using Eq.(4.3), let consider (X_1, Z_1) auxiliary variables under two-phase sampling scheme assuming $\bar{X}_1$ unknown and $\bar{Z}_1$ known

$$(4.6) \quad t_{1k(1)} \cong \bar{y}^* + \xi_1 \left(\bar{x}_1^{(1)} - \widetilde{x}_1^{(2)} \right) + \varphi_1 \left(\bar{z}_1^{(1)} - \widetilde{z}_1^{(2)} \right) + \psi_1 \left(\bar{Z}_1 - \bar{z}_1^{(1)} \right).$$

The asymptotic variance of $t_{1k(1)}$ can be written as

(4.7)
$$\text{AV}\,(t_{1k(1)}) \cong \widetilde{S}_y^2 + \xi_1^2 \widetilde{S}_{x_1}^2 + \varphi_1^2 \widetilde{S}_{z_1}^2 + \psi_1^2 \bar{S}_{z_1}^2 - 2\xi_1 \widetilde{S}_{yx_1} - 2\varphi_1 \widetilde{S}_{yz_1} - 2\psi_1 \bar{S}_{yz_1} + 2\xi_1 \varphi_1 \widetilde{S}_{x_1 z_1}.$$

Minimizing Eq.(4.7) w.r.t $(\xi_1,\ \varphi_1,\ \psi)$, one can get

$$\xi_1 = \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2 \right)^{-1} \left(\widetilde{S}_{z_1}^2 \widetilde{S}_{yx_1} - \widetilde{S}_{x_1 z_1} \widetilde{S}_{yz_1} \right) = \xi_1^o (\text{say}),$$

$$\varphi_1 = \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2 \right)^{-1} \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{yz_1} - \widetilde{S}_{x_1 z_1} \widetilde{S}_{yx_1} \right) = \varphi_1^o (\text{say})$$

and

$$\psi_1 = \left(\bar{S}_{z_1}^2 \right)^{-1} \bar{S}_{yz_1} = \psi_1^o (\text{say}).$$

The minimum asymptotic variance of $t_{1k(1)}$ is given by

$$(4.8) \quad \text{minAV}(t_{1k(1)}) \cong \widetilde{S}_y^2 - \left[\frac{\widetilde{S}_{z_1}^2 \widetilde{S}_{yx_1}^2 + \widetilde{S}_{x_1}^2 \widetilde{S}_{yz_1}^2 - 2\widetilde{S}_{yx_1} \widetilde{S}_{yz_1} \widetilde{S}_{x_1 z_1}}{\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2} \right] - \frac{\bar{S}_{yz_1}^2}{\bar{S}_{z_1}^2},$$

$$(4.9) \quad \text{minAV}(t_{1k(1)}) \cong \widetilde{S}_y^2 \left[1 - \left(\frac{\widetilde{\rho}_{yx_1}^2 + \widetilde{\rho}_{yz_1}^2 - 2\widetilde{\rho}_{yx_1} \widetilde{\rho}_{yz_1} \widetilde{\rho}_{x_1 z_1}}{1 - \widetilde{\rho}_{x_1 z_1}^2} \right) - \bar{\rho}_{yz_1}^2 \right],$$

where $\widetilde{\rho}_{yu_1}^2 = \dfrac{\widetilde{S}_{yu_1}^2}{\widetilde{S}_y^2 \widetilde{S}_{u_1}^2}$, $u_1 = (x_1,\ z_1)$ and $\widetilde{\rho}_{x_1 z_1}^2 = \dfrac{\widetilde{S}_{x_1 z_1}^2}{\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2}$.

$$(4.10) \quad \text{minAV}(t_{1k(1)}) \cong \widetilde{S}_y^2 \left[1 - \widetilde{R}_{y.x_1 z_1}^2 - \bar{\rho}_{yz_1}^2 \right]$$

where $\widetilde{R}_{y.x_1 z_1}$ is a multiple correlation coefficient.

4.1.2. *Three Phase.* In this case, we take (X_1, X_2, Z_1, Z_2) auxiliary variables using three-phase sampling, assuming population means $(\bar{X}_1,\ \bar{X}_2)$ unknown and $(\bar{Z}_1,\ \bar{Z}_2)$ known

$$t_{1k(2)} \cong \bar{y}^* + \xi_1 \left(\bar{x}_1^{(1)} - \widetilde{x}_1^{(2)} \right) + \xi_2 \left(\bar{x}_2^{(2)} - \widetilde{x}_2^{(3)} \right) + \varphi_1 \left(\bar{z}_1^{(1)} - \widetilde{z}_1^{(2)} \right)$$

$$(4.11) \qquad + \varphi_2 \left(\bar{z}_2^{(2)} - \widetilde{z}_2^{(3)} \right) + \psi_1 \left(\bar{Z}_1 - \bar{z}_1^{(1)} \right) + \psi_2 \left(\bar{Z}_2 - \bar{z}_2^{(2)} \right).$$

The asymptotic variance of $t_{1k(2)}$

$$AV(t_{1k(2)}) \cong \widetilde{S}_y^2 + \sum_{j=1}^{2} \xi_i^2 \widetilde{S}_{x_j}^2 + \sum_{j=1}^{2} \varphi_j^2 \widetilde{S}_{z_j}^2 - 2 \left(\sum_{j=1}^{2} \xi_j \widetilde{S}_{yx_j} + \sum_{j=1}^{2} \varphi_j \widetilde{S}_{yz_j} + \sum_{j=1}^{2} \psi_j \bar{S}_{yz_j} \right)$$

$$(4.12) \quad + \sum_{j=1}^{2} \psi_j^2 \bar{S}_{z_j}^2 + 2 \sum_{j=1}^{2} \xi_j \varphi_j \widetilde{S}_{x_j z_j} + 2 \left(\psi_1 \psi_2 \bar{S}_{z_1 z_2} + \xi_1 \psi_2 \bar{\bar{S}}_{x_1 z_2} + \varphi_1 \psi_2 \bar{\bar{S}}_{z_1 z_2} \right).$$

The AV of $t_{1k(2)}$ is minimum when

$$\xi_2 = \left(\widetilde{S}_{x_2}^2 \widetilde{S}_{z_2}^2 - \widetilde{S}_{x_2 z_2}^2 \right)^{-1} \left(\widetilde{S}_{z_2}^2 \widetilde{S}_{yx_2} - \widetilde{S}_{x_2 z_2} \widetilde{S}_{yz_2} \right) = \xi_2^o (\text{say}),$$

$$\varphi_2 = \left(\widetilde{S}_{x_2}^2 \widetilde{S}_{z_2}^2 - \widetilde{S}_{x_2 z_2}^2 \right)^{-1} \left(\widetilde{S}_{x_2}^2 \widetilde{S}_{yz_2} - \widetilde{S}_{x_2 z_2} \widetilde{S}_{yx_2} \right) = \varphi_2^o (\text{say}),$$

$$\xi_1 = E^{-1} A = \xi_1^o (\text{say}),$$

$$\varphi_1 = E^{-1} B = \varphi_1^o (\text{say}),$$

$$\psi_1 = E^{-1} C = \psi_1^o (\text{say})$$

and

$$\psi_2 = E^{-1} D = \psi_2^o (\text{say}).$$

The details of (A, B, C, D, E) are given in Appendix B.

The resulting minimum asymptotic variance of $t_{1k(2)}$ is given by

$$(4.13) \quad \min AV(t_{1k(2)}) = \widetilde{S}_y^2 \left(1 - \bar{R}_{y.x_1 z_1 \bar{z}_1 \bar{z}_2}^2 - \widetilde{R}_{y.x_2 z_2}^2 \right),$$

where $\bar{R}_{y.x_1 z_1 \bar{z}_1 \bar{z}_2}$ and $\widetilde{R}_{y.x_2 z_2}$ are the multiple correlation coefficients.

4.2. Situation 2: $\underline{\bar{X}}$ and $\underline{\bar{Z}}$ both unknown. In this situation it is assumed that the population mean vectors $\underline{\bar{X}}$ and $\underline{\bar{Z}}$ are unknown. From the expression (4.3), we have

$$(4.14) \quad t_{2k} \cong \bar{y}^* + \sum_{i=1}^{k} \xi_i \left(\bar{x}_i^{(i)} - \widetilde{x}_i^{(i+1)} \right) + \sum_{i+1}^{k} \varphi_i \left(\bar{z}_i^{(i)} - \widetilde{z}_i^{(i+1)} \right).$$

We can write Eq.(4.14) in a generalized vector form as

$$(4.15) \quad t_{2k} \cong \bar{y}^* + (\underline{\bar{\nu}}' - \underline{\widetilde{\nu}}) \underline{\xi}.$$

The asymptotic variance of t_{2k} is

$$(4.16) \quad AV(t_{2k}) \cong Var(\bar{y}^*) + \underline{\xi}^t \widetilde{\mathbf{S}}_{xz} \underline{\xi} - 2\underline{\xi}^t \widetilde{\mathbf{S}}_{yxz},$$

where $(\underline{\xi}, \widetilde{\mathbf{S}}_{xz}, \widetilde{\mathbf{S}}_{yxz})$ are defined earlier in Situation 1.

Minimization of $AV(t_{2k})$ w.r.t g_{xz} leads to optimum vector

$$\underline{\xi}^o = \begin{bmatrix} \xi^o \\ \varphi^o \end{bmatrix} = \begin{bmatrix} (\underline{\widetilde{S}}_{xx} \underline{\widetilde{S}}_{zz} - \underline{\widetilde{S}}_{xz}^t \underline{\widetilde{S}}_{xz})^{-1} (\underline{\widetilde{S}}_{zz} \underline{\widetilde{S}}_{yx} - \underline{\widetilde{S}}_{xz} \underline{\widetilde{S}}_{yz}) \\ (\underline{\widetilde{S}}_{xx} \underline{\widetilde{S}}_{zz} - \underline{\widetilde{S}}_{xz}^t \underline{\widetilde{S}}_{xz})^{-1} (\underline{\widetilde{S}}_{xx} \underline{\widetilde{S}}_{yz} - \underline{\widetilde{S}}_{xz} \underline{\widetilde{S}}_{yx}) \end{bmatrix}.$$

Replacing $\underline{\xi}^o$ in Eq.(4.16), one can get the optimal estimator in the class which attains the minimum asymptotic variance bound given by

$$(4.17) \quad \text{minAV}(t_{2k}) \cong \widetilde{S}_y^2 \left[1 - \underline{\widetilde{R}}_{y.xz}^2\right],$$

where $\underline{\widetilde{R}}_{y.xz}^2$ is a vector of the squared multiple correlation coefficients of Y on $\underline{X}$ and $\underline{Z}$ and is explained in detail in Appendix C.

4.2.1. *Two Phase.* Now suppose that both X_1 and Z_1 have unknown means, in this case expression (4.14) can be expressed as

$$(4.18) \quad t_{2k(1)} \cong \bar{y}^* + \xi_1 \left(\bar{x}_1^{(1)} - \widetilde{x}_1^{(2)}\right) + \varphi_1 \left(\bar{z}_1^{(1)} - \widetilde{z}_1^{(2)}\right).$$

The asymptotic variance of $t_{2k(1)}$ can be written as

$$(4.19) \quad \text{AV}(t_{2k(1)}) \cong \widetilde{S}_y^2 + \xi_1^2 \widetilde{S}_{x_1}^2 + \varphi_1^2 \widetilde{S}_{z_1}^2 - 2\xi_1 \widetilde{S}_{yx_1} - 2\varphi_1 \widetilde{S}_{yz_1} + 2\xi_1 \varphi_1 \widetilde{S}_{x_1 z_1}.$$

Now minimizing (4.19) w.r.t (ξ_1, φ_1), we have

$$\xi_1 = \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2\right)^{-1} \left(\widetilde{S}_{z_1}^2 \widetilde{S}_{yx_1} - \widetilde{S}_{x_1 z_1} \widetilde{S}_{yz_1}\right) = \xi_1^o (\text{say})$$

and

$$\varphi_1 = \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2\right)^{-1} \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{yz_1} - \widetilde{S}_{x_1 z_1} \widetilde{S}_{yx_1}\right) = \varphi_1^o (\text{say}).$$

One can write minimum AV of $t_{2k(1)}$ as

$$(4.20) \quad \text{minAV}(t_{2k(1)}) \cong \text{Var}(\bar{y}^*) - \left[\frac{\widetilde{S}_{z_1}^2 \widetilde{S}_{yx_1}^2 + \widetilde{S}_{x_1}^2 \widetilde{S}_{yz_1}^2 - 2\widetilde{S}_{yx_1} \widetilde{S}_{yz_1} \widetilde{S}_{x_1 z_1}}{\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2}\right],$$

$$(4.21) \quad \text{minAV} (t_{2k(1)}) = \widetilde{S}_y^2 \left(1 - \widetilde{R}_{y.x_1 z_1}^2\right),$$

where $\widetilde{R}_{y.x_1 z_1}^2$ is explained earlier in the Section (4.1.1).

4.2.2. *Three Phase.* Now consider the same auxiliary variables as taken earlier in Section (4.1.2) with unknown means

$$(4.22) \quad \begin{aligned} t_{2k(2)} &\cong \bar{y}^* + \xi_1 \left(\bar{x}_1^{(1)} - \widetilde{x}_1^{(2)}\right) + \xi_2 \left(\bar{x}_2^{(2)} - \widetilde{x}_2^{(3)}\right) \\ &\quad + \varphi_1 \left(\bar{z}_1^{(1)} - \widetilde{z}_1^{(2)}\right) + \varphi_2 \left(\bar{z}_2^{(2)} - \widetilde{z}_2^{(3)}\right). \end{aligned}$$

The asymptotic variance of $t_{2k(2)}$ can be expressed as

(4.23)

$$\text{AV}(t_{2k(2)}) \cong \widetilde{S}_y^2 + \sum_{j=1}^{2} \xi_j^2 \widetilde{S}_{x_j}^2 + \sum_{j=1}^{2} \varphi_j^2 \widetilde{S}_{z_j}^2 - 2 \left(\sum_{j=1}^{2} \xi_j \widetilde{S}_{yx_j} + \sum_{j=1}^{2} \varphi_j \widetilde{S}_{yz_j}\right) + 2 \sum_{j=1}^{2} \xi_j \varphi_j \widetilde{S}_{x_j z_j}.$$

Now minimizing (4.23) w.r.t $(\xi_1, \ \xi_2, \ \varphi_1, \ \varphi_2)$, we have

$$\underline{\xi}^o = \begin{bmatrix} \xi_1^o \\ \xi_2^o \\ \varphi_1^o \\ \varphi_2^o \end{bmatrix} = \begin{bmatrix} \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2\right)^{-1} \left(\widetilde{S}_{z_1}^2 \widetilde{S}_{yx_1} - \widetilde{S}_{x_1 z_1}\widetilde{S}_{yz_1}\right) \\ \left(\widetilde{S}_{x_2}^2 \widetilde{S}_{z_2}^2 - \widetilde{S}_{x_2 z_2}^2\right)^{-1} \widetilde{S}_{z_2}^2 \widetilde{S}_{yx_2} - \widetilde{S}_{x_2 z_2}\widetilde{S}_{yz_2} \\ \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2\right)^{-1} \left(\widetilde{S}_{x_1}^2 \widetilde{S}_{yz_1} - \widetilde{S}_{x_1 z_1}\widetilde{S}_{yx_1}\right) \\ \left(\widetilde{S}_{x_2}^2 \widetilde{S}_{z_2}^2 - \widetilde{S}_{x_2 z_2}^2\right)^{-1} \left(\widetilde{S}_{x_2}^2 \widetilde{S}_{yz_2} - \widetilde{S}_{x_2 z_2}\widetilde{S}_{yx_2}\right) \end{bmatrix}.$$

By replacing $(\xi_1^o, \xi_2^o, \varphi_1^o, \ \varphi_2^o)$ in (4.23), one can get the optimal estimator in the class which attains the minimum asymptotic variance bound given by:

(4.24)
$$\text{minAV}(t_{2k(2)}) \cong \widetilde{S}_y^2 - \left(\frac{\widetilde{S}_{z_1}^2 \widetilde{S}_{yx_1}^2 + \widetilde{S}_{x_1}^2 \widetilde{S}_{yz_1}^2 - 2\widetilde{S}_{yx_1}\widetilde{S}_{yz_1}\widetilde{S}_{x_1 z_1}}{\widetilde{S}_{x_1}^2 \widetilde{S}_{z_1}^2 - \widetilde{S}_{x_1 z_1}^2}\right)$$
$$- \left(\frac{\widetilde{S}_{z_2}^2 \widetilde{S}_{yx_2}^2 + \widetilde{S}_{x_2}^2 \widetilde{S}_{yz_2}^2 - 2\widetilde{S}_{yx_2}\widetilde{S}_{yz_2}\widetilde{S}_{x_2 z_2}}{\widetilde{S}_{x_2}^2 \widetilde{S}_{z_2}^2 - \widetilde{S}_{x_2 z_2}^2}\right),$$

(4.25) $\quad \text{minAV}(t_{2k(2)}) \cong \widetilde{S}_y^2 \left(1 - \widetilde{R}_{y.x_1 z_1}^2 - \widetilde{R}_{y.x_2 z_2}^2\right),$

where $\widetilde{R}_{y.x_1 z_1}^2$ and $\widetilde{R}_{y.x_2 z_2}^2$ are the squared multiple correlation coefficients. From Eq.(4.10), Eq.(4.13), Eq.(4.21) and Eq.(4.25), we can accomplish that $\left(t_{1k(1)}, \ t_{1k(2)}, \ t_{2k(1)} \text{ and } t_{2k(2)}\right)$ are regression type estimators in their optimal cases.

5. Numerical Comparison

In order to illustrate the gain in efficiency for the best estimator in the two situations discussed in previous section, we carried out a numerical study using the Population Census Report of Sialkot District (1998), Pakistan and this data is earlier used by [5]. Each variable is taken from rural locality. The description of variables is given below

Variable	Description
Y	Literacy Ratio
X_1	Population of primary but below matric
X_2	Population of matric and above
Z_1	Population of 18 years old and above
Z_2	Population of women 15-49 years old

We compute the Percent Relative Efficiency (PRE) of the considered estimators with respect to the [6] estimator $\bar{y}^*$, for different values of b, as

$$\text{PRE}(t_\bullet) = \frac{\text{Var}(\bar{y}^*)}{\text{AV}(t_\bullet)} \times 100,$$

where $t_\bullet = \left(t_{\text{OL}}, t_{\text{K}}, t_{\text{SK}(1)}, t_{\text{SK}(2)}, t_{1k(1)}, t_{1k(2)}, t_{2k(1)}, t_{2k(2)}\right).$

Table 2. PREs of the considered classes with respect to $\bar{y}^*$
for different values of b

m_1	m_2	Situation	b			
			2	3	4	5
160	100	t_{OL}	119.26	121.72	123.22	124.24
		t_{K}	120.52	123.69	125.65	126.98
		$t_{\text{SK}(1)}$	119.56	123.05	125.48	127.24
		$t_{\text{SK}(2)}$	121.34	124.98	127.31	128.91
		$t_{1k(1)}$	*128.95*	*131.98*	*133.91*	*135.22*
		$t_{2k(1)}$	124.78	128.71	131.20	132.92
	120	t_{OL}	116.96	120.11	122.00	123.26
		t_{K}	118.27	122.15	124.49	126.06
		$t_{\text{SK}(1)}$	117.02	121.04	123.85	125.88
		$t_{\text{SK}(2)}$	118.86	123.26	126.01	127.87
		$t_{1k(1)}$	*127.18*	*130.87*	*133.11*	*134.62*
		$t_{2k(1)}$	121.95	126.79	129.77	131.79
180	100	t_{OL}	121.50	123.45	124.64	125.44
		t_{K}	122.81	125.48	127.11	128.22
		$t_{\text{SK}(1)}$	122.29	125.43	127.55	129.05
		$t_{\text{SK}(2)}$	123.90	127.02	129.00	130.36
		$t_{1k(1)}$	*130.82*	*133.42*	*135.07*	*136.20*
		$t_{2k(1)}$	127.69	130.98	133.06	134.50
	120	t_{OL}	119.78	122.27	123.74	124.72
		t_{K}	121.16	124.38	126.31	127.59
		$t_{\text{SK}(1)}$	120.24	123.90	126.35	128.08
		$t_{\text{SK}(2)}$	122.07	125.80	128.10	129.66
		$t_{1k(1)}$	*129.54*	*132.65*	*134.55*	*135.82*
		$t_{2k(1)}$	125.57	129.59	132.06	133.72
200	100	t_{OL}	123.36	124.87	125.79	126.40
		t_{K}	124.71	126.95	128.31	129.24
		$t_{\text{SK}(1)}$	124.72	127.49	129.31	130.58
		$t_{\text{SK}(2)}$	126.03	128.71	130.39	131.54
		$t_{1k(1)}$	*132.37*	*134.59*	*136.01*	*136.99*
		$t_{2k(1)}$	130.12	132.85	134.59	135.80
	120	t_{OL}	122.14	124.05	125.18	125.92
		t_{K}	123.58	126.23	127.80	128.85
		$t_{\text{SK}(1)}$	123.19	126.41	128.50	129.94
		$t_{\text{SK}(2)}$	124.77	127.91	129.83	131.12
		$t_{1k(1)}$	*131.51*	*134.12*	*135.72*	*136.80*
		$t_{2k(1)}$	128.64	131.93	133.95	135.31

Table 3. PREs of the proposed class with respect to $\bar{y}^*$ for different values of b

m_1	m_2	m_3	Situation	b 2	3	4	5
		100	$t_{1k(2)}$	136.32	141.78	145.44	148.04
			$t_{2k(2)}$	133.54	139.50	143.46	146.26
200	140	110	$t_{1k(2)}$	136.50	142.56	146.57	149.40
			$t_{2k(2)}$	133.32	139.95	144.32	147.36
		120	$t_{1k(2)}$	136.72	143.41	147.79	150.86
			$t_{2k(2)}$	133.11	140.46	145.24	148.56
		100	$t_{1k(2)}$	137.16	141.93	145.11	147.36
			$t_{2k(2)}$	135.34	140.44	143.81	146.17
220	160	110	$t_{1k(2)}$	137.45	142.72	146.19	148.62
			$t_{2k(2)}$	135.36	141.02	144.70	147.27
		120	$t_{1k(2)}$	137.78	143.58	147.34	149.97
			$t_{2k(2)}$	135.40	141.64	145.66	148.43

In Table 2, we assume 40% non-response rate of the total population $N = 268$ (assuming last 107 units as non-respondents for second phase) and the numerical results of the estimators $t_{1k(1)}$ and $t_{2k(1)}$ with different combinations of sample sizes m_1 and m_2 are given. To compute PREs of the estimators $t_{1k(2)}$ and $t_{2k(2)}$, we consider 40% and 30% non-response rates of the total population at first and second phase (assuming last 80 units as non-respondents for third phase). The results are provided in Table 3 for different choices of sample sizes m_1, m_2 and m_3.

In Tables 2 and 3, it is seen that the PREs of all considered estimators are higher than $\bar{y}^*$ due to inclusion of the auxiliary information. Also, the PREs of all estimators increase with the increase of inverse sampling rate b. In Table 2, the estimators $(t_{OL},\ t_K)$ with single auxiliary variable and $(t_{SK(1)},\ t_{SK(2)})$ with two auxiliary variables perform almost similar but the proposed estimators $(t_{1k(1)},\ t_{2k(1)})$ show higher efficiency. It is also seen that the regression estimators $(t_{1k(1)},\ t_{2k(1)})$ perform better than the regression-cum-ratio estimators $(t_{SK(1)},\ t_{SK(2)})$.

When we consider more auxiliary variables and phases for t_{1k} and t_{2k} in Table 3, it is observed that they have considerable increase in efficiency. Moreover, it is observed that when we increase sample sizes, the PREs of all considered estimators also increase which confirms the large sample theory aspect.

It is also expected that the first proposed class t_{1k} shows always higher efficiency than the second class t_{2k}, because we are using more auxiliary information in Situation 1. Hence, we can conclude that the estimator t_{1k} is the best choice if $\bar{\underline{Z}}$ is known and, of course, t_{2k} is to use when $\bar{\underline{Z}}$ is unknown.

6. Conclusions

In this paper we propose a general class of estimators for the estimation of population mean $\bar{Y}$. For this we consider multi phase sampling scheme when auxiliary information is available. The effects of non-response on the study and on the auxiliary variables are discussed in detail. We determine the asymptotic variance for the proposed classes. To compare the efficiency of the suggested ones with other existing estimators in the literature, [6] estimator is used. It is noted that the performance of the proposed estimator t_{1k} is better than the other considered estimators. From the numerical analysis one can draw the conclusion that regression type estimators (t_{1k}, t_{2k}) always perform better if compared to regression-cum-ratio estimators $\left(t_{SK(1)}, t_{SK(2)}\right)$. However, our proposed generalized class of estimators is more efficient, in terms of asymptotic variance, if compared to all previous estimators available in the literature.

Acknowledgement

Appendix A

The details of matrices in Eq.(4.5) are described here

$$\underline{\bar{S}}_{\bar{z}\bar{z}} = E\left[(\underline{\bar{z}}' - \underline{\bar{Z}})(\underline{\bar{z}}' - \underline{\bar{Z}})^t\right] = \begin{bmatrix} \theta_1 S_{z_1}^2 & \theta_1 S_{z_1 z_2} & \dots & \theta_1 S_{z_1 z_k} \\ \theta_1 S_{z_1 z_2} & \theta_2 S_{z_2}^2 & \dots & \theta_2 S_{z_2 z_k} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_1 S_{z_1 z_k} & \theta_2 S_{z_2 z_k} & \dots & \theta_k S_{z_k}^2 \end{bmatrix},$$

$$\underline{\bar{S}}_{xz\bar{z}} = \begin{bmatrix} \underline{S}_{x\bar{z}} & \underline{S}_{z\bar{z}} \end{bmatrix}^t,$$

where

$$\underline{S}_{x\bar{z}} = \begin{bmatrix} 0 & (\theta_2 - \theta_1)S_{x_1 z_2} & 0 & \dots & 0 \\ 0 & 0 & (\theta_3 - \theta_2)S_{x_2 z_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (\theta_k - \theta_{k-1})S_{x_{k-1} z_k} \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix},$$

$$\underline{S}_{z\bar{z}} = \begin{bmatrix} 0 & (\theta_2 - \theta_1)S_{z_1 z_2} & 0 & \dots & 0 \\ 0 & 0 & (\theta_3 - \theta_2)S_{z_2 z_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (\theta_k - \theta_{k-1})S_{z_{k-1} z_k} \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

$$\underline{\widetilde{\mathbf{S}}}_{xz} = E\left[(\underline{\nu}' - \underline{\widetilde{\nu}})(\underline{\nu}' - \underline{\widetilde{\nu}}^t\right] = \begin{bmatrix} \underline{\widetilde{S}}_{xx} & \underline{\widetilde{S}}_{xz} \\ \underline{\widetilde{S}}_{xz} & \underline{\widetilde{S}}_{zz} \end{bmatrix},$$

Now using the dummy variable u for the following representations

$$\underline{\widetilde{S}}_{uu} = \mathrm{diag}(\widetilde{S}^2_{u_i}),$$

where

$$\widetilde{S}^2_{u_i} = (\theta_{i+1} - \theta_i)\, S^2_{u_i} + \omega_{i+1} S^2_{u_i(2)},$$

and

$$\underline{\widetilde{S}}_{yu} = \begin{bmatrix} \widetilde{S}_{yu_1} & \widetilde{S}_{yu_2} & \dots & \widetilde{S}_{yu_k} \end{bmatrix}^t,$$

where

$$\widetilde{S}_{yu_i} = (\theta_{i+1} - \theta_i) S_{yu_i} + \omega_{i+1} S_{yu_i(2)},$$

with

$$u_i = (x_i, z_i), \qquad i = 1, \dots, k$$

$$\theta_i = \left(\frac{1}{m_i} - \frac{1}{N}\right), \quad \theta_{i+1} = \left(\frac{1}{m_{i+1}} - \frac{1}{N}\right), \quad \omega_{i+1} = \frac{N_{i+1}(b-1)}{m_{i+1}N}, \quad b > 1.$$

$$\underline{\widetilde{S}}_{uv} = \mathrm{diag}(\widetilde{S}_{u_i v_i}), \qquad (u_i = x_i, v_i = z_i).$$

where

$$\widetilde{S}_{u_i v_i} = (\theta_{i+1} - \theta_i) S_{u_i v_i} + \omega_{i+1} S_{u_i v_i(2)}.$$

Appendix B

Minimizing Eq.(4.12), the following terms are obtained

$$\begin{aligned}
A =\ & \widetilde{S}^2_{z_1}\bar{S}^2_{z_1}\bar{S}^2_{z_2}\widetilde{S}_{yx_1} - \bar{S}^2_{z_1}\bar{S}^2_{z_2}\widetilde{S}_{yz_1}\widetilde{S}_{x_1z_1} - \bar{S}^2_{z_1}\bar{S}^2_{z_1}\widetilde{S}_{yz_2}\bar{\bar{S}}_{x_1z_2} - \bar{S}^2_{z_1}\widetilde{S}_{yx_1}\bar{S}^2_{z_1z_2} \\
& - \bar{S}^2_{z_1}\widetilde{S}_{yx_1}\bar{\bar{S}}^2_{z_1z_2} + \bar{S}^2_{z_1}\widetilde{S}_{yz_1}\bar{\bar{S}}_{x_1z_2}\bar{\bar{S}}_{z_1z_2} + \widetilde{S}^2_{z_1}\widetilde{S}_{yz_1}\bar{\bar{S}}_{x_1z_2}\bar{S}_{z_1z_2} \\
& + \bar{S}^2_{z_1}\bar{S}_{yz_2}\widetilde{S}_{x_1z_1}\bar{\bar{S}}_{z_1z_2} + \widetilde{S}_{yz_1}\widetilde{S}_{x_1z_1}\bar{S}^2_{z_1z_2} - \bar{S}_{yz_1}\widetilde{S}_{x_1z_1}\bar{S}_{z_1z_2}\bar{\bar{S}}_{z_1z_2},
\end{aligned}$$

$$\begin{aligned}
B =\ & \widetilde{S}^2_{x_1}\bar{S}^2_{z_1}\bar{S}^2_{z_2}\widetilde{S}_{yz_1} - \bar{S}^2_{z_1}\bar{S}^2_{z_2}\widetilde{S}_{x_1z_1}\widetilde{S}_{yx_1} - \bar{S}^2_{x_1}\bar{S}^2_{z_1}\bar{\bar{S}}_{z_1z_2}\bar{S}_{yz_2} - \widetilde{S}^2_{x_1}\bar{S}^2_{z_1z_2}\widetilde{S}_{yz_1} \\
& - \bar{S}^2_{z_1}\bar{\bar{S}}^2_{x_1z_2}\widetilde{S}_{yz_1} + \bar{S}^2_{z_1}\bar{\bar{S}}_{x_1z_2}\bar{\bar{S}}_{z_1z_2}\widetilde{S}_{yx_1} + \widetilde{S}^2_{x_1}\bar{S}_{z_1z_2}\bar{\bar{S}}_{z_1z_2}\bar{S}_{yz_1} \\
& + \bar{S}^2_{z_1}\widetilde{S}_{x_1z_1}\bar{\bar{S}}_{x_1z_2}\bar{S}_{yz_2} + \widetilde{S}_{x_1z_1}\bar{S}^2_{z_1z_2}\widetilde{S}_{yx_1} - \widetilde{S}_{x_1z_1}\bar{S}_{z_1z_2}\bar{\bar{S}}_{x_1z_2}\bar{S}_{yz_1},
\end{aligned}$$

$$\begin{aligned}
C =\ & \widetilde{S}^2_{x_1}\widetilde{S}^2_{z_1}\bar{S}^2_{z_2}\bar{S}_{yz_1} - \widetilde{S}^2_{x_1}\widetilde{S}^2_{z_1}\bar{S}_{z_1z_2}\bar{S}_{yz_2} - \widetilde{S}^2_{x_1}\bar{\bar{S}}^2_{z_1z_2}\bar{S}_{yz_1} - \widetilde{S}^2_{z_1}\bar{S}^2_{x_1z_2}\bar{S}_{yz_1} \\
& - \bar{S}^2_{z_2}\widetilde{S}^2_{x_1z_1}\bar{S}_{yz_1} + \widetilde{S}^2_{x_1z_1}\bar{S}_{z_1z_2}\bar{S}_{yz_2} + \widetilde{S}^2_{z_1}\bar{\bar{S}}_{x_1z_2}\bar{S}_{z_1z_2}\widetilde{S}_{yx_1} + \widetilde{S}^2_{x_1}\bar{S}_{z_1z_2}\bar{\bar{S}}_{z_1z_2}\widetilde{S}_{yz_1} \\
& - \widetilde{S}^2_{x_1z_1}\bar{S}_{z_1z_2}\bar{\bar{S}}_{z_1z_2}\widetilde{S}_{yx_1} - \widetilde{S}^2_{x_1z_1}\bar{S}_{z_1z_2}\bar{\bar{S}}_{x_1z_2}\widetilde{S}_{yz_1} + 2\widetilde{S}^2_{x_1z_1}\bar{\bar{S}}_{z_1z_2}\bar{\bar{S}}_{x_1z_2}\bar{S}_{yz_1},
\end{aligned}$$

$$\begin{aligned}
D =\ & \widetilde{S}^2_{x_1}\widetilde{S}^2_{z_1}\bar{S}^2_{z_1}\bar{S}_{yz_2} - \widetilde{S}^2_{x_1}\widetilde{S}^2_{z_1}\bar{S}_{z_1z_2}\bar{S}_{yz_1} - \widetilde{S}^2_{x_1}\bar{S}^2_{z_1}\bar{\bar{S}}_{z_1z_2}\widetilde{S}_{yz_1} - \widetilde{S}^2_{z_1}\bar{S}^2_{z_1}\bar{\bar{S}}_{x_1z_2}\widetilde{S}_{yx_1} \\
& - \bar{S}^2_{z_1}\widetilde{S}^2_{x_1z_1}\bar{S}_{yz_2} + \widetilde{S}^2_{x_1z_1}\bar{S}_{z_1z_2}\bar{S}_{yz_1} + \bar{S}^2_{z_1}\widetilde{S}_{x_1z_1}\bar{\bar{S}}_{z_1z_2}\widetilde{S}_{yx_1} + \bar{S}^2_{z_1}\widetilde{S}_{x_1z_1}\bar{\bar{S}}_{x_1z_2}\widetilde{S}_{yz_1},
\end{aligned}$$

$$\begin{aligned}
E =\ & \widetilde{S}^2_{x_1}\widetilde{S}^2_{z_1}\bar{S}^2_{z_1}\bar{S}^2_{z_2} - (\widetilde{S}^2_{x_1}\widetilde{S}^2_{z_1}\bar{S}^2_{z_1z_2} + \widetilde{S}^2_{x_1}\bar{S}^2_{z_1}\bar{\bar{S}}^2_{z_1z_2} + \widetilde{S}^2_{z_1}\bar{S}^2_{z_1}\bar{\bar{S}}^2_{x_1z_2} \\
& + \bar{S}^2_{z_1}\bar{S}^2_{z_2}\widetilde{S}^2_{x_1z_1}) + \widetilde{S}^2_{x_1z_1}\bar{S}^2_{z_1z_2} + 2\bar{S}^2_{z_1}\widetilde{S}_{x_1z_1}\bar{\bar{S}}_{x_1z_2}\bar{\bar{S}}_{z_1z_2}.
\end{aligned}$$

Appendix C

From (4.17), the vector of the squared multiple correlation coefficient $\widetilde{\underline{R}}^2_{y.xz}$ is explained as

$$\widetilde{\underline{R}}^2_{y.xz} = \widetilde{R}^2_{y.x_1 z_1} + \widetilde{R}^2_{y.x_2 z_2} + \cdots + \widetilde{R}^2_{y.x_\kappa z_\kappa},$$

where

$$\widetilde{R}^2_{y.x_i z_i} = \left(\frac{\widetilde{\rho}^2_{yx_i} + \widetilde{\rho}^2_{yz_i} - 2\widetilde{\rho}_{yx_i}\widetilde{\rho}_{yz_i}\widetilde{\rho}_{x_i z_i}}{1 - \widetilde{\rho}^2_{x_i z_i}} \right), \quad (i = 1, \ldots, \kappa),$$

$$\widetilde{\rho}^2_{yu_i} = \frac{\widetilde{S}^2_{yu_i}}{\widetilde{S}^2_y \widetilde{S}^2_{u_i}}, \qquad u_i = (x_i, z_i),$$

and

$$\widetilde{\rho}^2_{x_i z_i} = \frac{\widetilde{S}^2_{x_i z_i}}{\widetilde{S}^2_{x_i} \widetilde{S}^2_{z_i}}.$$

References

[1] Ahmed, M. S. *Generalized chain estimators under multi phase sampling*, Journal of Applied Statistics **12**, 243–250, 2003.

[2] Diana, G. and Perri, P. F. *Estimation of finite population mean using multi-auxiliary information*, Metron **LXV**, 99–112, 2007.

[3] Diana, G. and Perri, P. F. *A class of estimators in two-phase sampling with subsampling the non-respondents*, Applied Mathematics and Computation **219**, 10033–10043, 2013.

[4] Gupta, S. and Shabbir, J. *On improvement in estimating the population mean in simple random sampling*, Journal of Applied Statistics **35**, 559–566, 2008.

[5] Hanif, M., Ahmed, Z. and Ahmad, M. *Generalized multivariate ratio estimators using multi-auxiliary variables for multi-phase sampling*, Pakistan Journal of Statistics **25** (4), 615–629, 2009.

[6] Hansen, M. H. and Hurwitz, W. N. *The problems of non-response in sample surveys*, Journal of American Statistical Association **41**, 517–529, 1946.

[7] Jain, R. K. *Properties of estimators in simple random sampling using auxiliary variable*, Metron **XLV**, 265–271, 1987.

[8] Kadilar, C. and Cingi, H. *Improvement in estimating the population mean in simple random sampling*, Applied Mathematics Letters **19**, 75–79, 2006.

[9] Khare, B. B. and Srivastava, S. *Transformed ratio type estimtors for the population mean in the presence of non-response*, Communication in Statistics, Theory and Methods **26**, 1779–1791, 1997.

[10] Kumar, S. *Ratio cum regression estimator for estimating a population mean with a sub sampling of non respondents*, Communications of the Korean Statistical Society **19** (5), 663–671, 2012.

[11] Mukerjee, R., Rao, T. J. and Vijayan, K. *Regression type estimators using multi auxiliary information*, The Australian Journal of Statistics **29** (3), 244–254, 1987.

[12] Naik, V. D. and Gupta, P.C. *A general class of estimators for estimating population mean using auxiliary information*, Metrika **38**, 11–17, 1991.

[13] Okafor, F. C. and Lee, H. *Double sampling for ratio and regression estimation with subsampling the non-respondents*, Survey Methodology **26** (2), 183–188, 2000.

[14] Rao, P. S. R. S. *Ratio estimation with sub sampling the non-respondents*, Survey Methodology **12**, 217–230, 1986.

[15] Riaz, S. and Shabbir, J. *An improved class of regression estimators for the population mean under multi-phase sampling scheme in the presence of non-response*, Proceedings of the 11th Islamic Countries Conference on Statistical Sciences **21**, 563–576, 2011.

[16] Singh, H. P. and Kumar, S. *A general family of estimators of finite population ratio, product and mean using two phase sampling scheme in the presence of non-response*, Journal of Statistical Theory and Practice **2** (4), 677–692, 2008.

[17] Singh, H. P. and Kumar, S. *Improved estimation of population mean under two phase sampling with subsampling the non-respondents*, Journal of Statistical Planning and Inference **140**, 2536–2550, 2010.

[18] Srinath, K. P. *Multiphase sampling in nonresponse problems*, Journal of the American Statistical Association **66** (335), 583–586, 1971.

[19] Srivastava, S. K. *A generalized estimator for the mean of a finite population using multi-auxiliary information*, Journal of the American Statistical Association **66** (334), 404–407, 1971.

[20] Tabasum, R. and Khan, I. A. *Double sampling for ratio estimator for the population mean in presence of non-response*, Assam Statistical Review **20**, 73–83, 2006.

[21] Wolter, K. M. *Introduction to Variance Estimation*, (Springer, New York, 1985).