Hacettepe Journal of Mathematics and Statistics
Volume 46 (4) (2017), 743 – 765

Modified classes of estimators in circular systematic sampling

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

Abstract

In this paper we consider the problem of estimation of population mean in circular systematic sampling design along with the non-response problem. For the population mean using auxiliary information three generalized classes of estimators are suggested. The biases and the mean square errors of the suggested classes of estimators are obtained and compared with sample mean, linear regression estimators, [23] estimator and [21] estimators. A numerical study is provided to show that the proposed classes of estimators based on circular systematic design can be more efficient than the estimators based on simple random sampling. Moreover, a simulation study is accomplished when some population parameters are assumed to be unknown.

Keywords: Bias, Mean square error, Ratio function, Exponential function, Two-phase sampling, Non-response, Efficiency.

2000 AMS Classification: 62D05

Received : 21.03.2014 *Accepted* : 25.08.2015 *Doi* : 10.15672/HJMS.20158412762

[*]Department of Mathematics and Statistics, Riphah International 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

1. Introduction

Systematic sampling is a simple scheme in which the sample selection is based on one random start i.e. from a finite population of N units, the first unit is selected randomly from the first k units and every kth unit thereafter, assuming that the sample size n satisfies $N = nk$, for some integer k. Generally most of the systematic sampling methods considered in the literature are based on this assumption to obtain estimators for estimating the population mean of the variable of interest, see for instance [30], [13], [18], [22], [24] and references therein. In various situations this systematic sampling, called linear systematic sampling, provides more efficient estimates than simple random sampling and/or stratified random sampling for certain types of population [see [2], [7] etc.]. Nowadays, systematic sampling design is becoming more popular than simple random sampling due to its simplicity and effective accessibility. Recently, [20], [23], [19], [27] and [28] suggested some classes of estimators for estimation of the population mean using this design.

However this sampling design has some drawbacks, such as the impossibility to get an unbiased estimator of the variance of the sample mean on the basis of a single sample. Moreover, if k is not an integer then there will be dissimilarity between the actual sample size and the specified one. As a consequence, the resulting sample mean will be a biased estimator of the population mean. To tackle these problems, [5] suggested a multi-start systematic sampling. [12] proposed different partially systematic sampling procedures. Afterwards, [10] developed a circular systematic sampling method for obtaining unbiased estimators for the population mean of the study variable when the population size is not a multiple of the sample size. [26] proposed the additional circular systematic sampling when the populations reveal linear and parabolic trends. Later, [11] have proposed estimators under balanced circular systematic sampling and centered circular systematic sampling when population trend is linear or parabolic. Circular systematic sampling can be used in both cases, whether k is an integer or not.

The present study aims to give some contribution on this subject. For this purpose, taking motivation from [3] and [21], three classes of estimators are modified for the estimation of the population mean of the variable of interest, using the auxiliary information in circular systematic sampling design, considering the possibility that non-response may (or may not) present in the study variable.

2. Notations and background

Let us assume that U be a finite population consists of N distinct units labelled from 1 to N in some order and n be a fixed sample size. Let Y and X be the study and the auxiliary variables having values y_{ij} and x_{ij}, $(i = 1, \dots, N \quad \text{and} \quad j = 1, \dots, n)$. As the aim of the present paper is the estimation of the unknown population mean $\bar{Y}$, we consider the circular systematic sampling (CSS), proposed by [10], to collect information of the variables Y and X. In this design, let the sample interval k is defined as follows

$$k = \begin{cases} \mathrm{INT}\left(\frac{N}{n}\right), & \text{if } \left(\frac{N}{\mathrm{INT}\left(\frac{N}{n}\right)+1}\right) \text{ is an integer,} \\ \mathrm{INT}\left(\frac{N}{n}+\frac{1}{2}\right), & \text{otherwise,} \end{cases}$$

where INT(.) denotes the truncated integer of the mentioned quantity [17].

As well known, this choice of k ensures good properties:

i- the sample size is the same for all samples;

ii- the sample mean is an unbiased estimator of the population mean;

iii- the first order probabilities of inclusion are the same for all units;

iv- each sample is without replacement.

Let s_i be the ith possible sample of size n with a start i, randomly selected from 1 to N, consists of units selected by the following procedure

$$\text{Label} = \begin{cases} i+(j-1)k, & 1 \leq i+(j-1)k \leq N, \\ i+(j-1)k-N, & N < i+(j-1)k. \end{cases}$$

for $i = (1, \ldots, N)$ and $j = (1, \ldots, n)$.

In this way, we may draw N circular systematic samples, each of size n, as displayed in the following table

Table 1. Possible Circular Systematic Samples

sample number	1	2	...	i	...	N
	u_1	u_2		u_i		u_N
	u_{k+1}	u_{k+2}		u_{k+i}		u_{2N}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	$u_{(n-1)k+1}$	$u_{(n-1)k+2}$		$u_{(n-1)k+i}$		u_{nN}

From this N possible CSS, a sample of size n is selected randomly to observe Y and X. The sample means $\bar{y}_{\text{css}} = \dfrac{\sum_{j=1}^{n} y_{ij}}{n}$ and $\bar{x}_{\text{css}} = \dfrac{\sum_{j=1}^{n} x_{ij}}{n}$ are unbiased estimators of the population means $\bar{Y} = \dfrac{\sum_{i}^{N} y_{ij}}{N}$ and $\bar{X} = \dfrac{\sum_{i}^{N} x_{ij}}{N}$ respectively. The variance of $\bar{y}_{\text{css}}$ and $\bar{x}_{\text{css}}$ can be expressed as

$$\text{(2.1)} \qquad \mathrm{V}(\bar{y}_{\text{css}}) = \left(\frac{N-1}{N}\right)\left[1+(n-1)\rho_y\right]\frac{S_y^2}{n} = \widetilde{S}_y^2$$

and

$$\text{(2.2)} \qquad \mathrm{V}(\bar{x}_{\text{css}}) = \left(\frac{N-1}{N}\right)\left[1+(n-1)\rho_x\right]\frac{S_x^2}{n} = \widetilde{S}_x^2,$$

where

$$S_y^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij}-\bar{Y})^2}{n(N-1)}, \qquad S_x^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n}(x_{ij}-\bar{X})^2}{n(N-1)},$$

with

$$\rho_y = \frac{2}{n(n-1)(N-1)S_y^2}\sum_{i}^{N}\sum_{j<u}(y_{ij}-\bar{Y})(y_{iu}-\bar{Y})$$

and

$$\rho_x = \frac{2}{n(n-1)(N-1)S_x^2}\sum_{i}^{N}\sum_{j<u}(x_{ij}-\bar{X})(x_{iu}-\bar{X}),$$

where $(\rho_y,\ \rho_x)$ are intraclass correlation coefficients between pairs of units within the CSS for Y and X, respectively.

We can define covariance as

$$\text{(2.3)} \qquad \mathrm{Cov}(\bar{y}_{\text{css}}, \bar{x}_{\text{css}}) = \left(\frac{N-1}{N}\right)\left[1+(n-1)\rho_y\right]^{\frac{1}{2}}\left[1+(n-1)\rho_x\right]^{\frac{1}{2}}\frac{S_{yx}}{n} = \widetilde{S}_{yx},$$

where

$$S_{yx} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij}-\bar{Y})(x_{ij}-\bar{X})}{n(N-1)}.$$

We can also represent (2.1), (2.2) and (2.3) as

$$\mathrm{V}(\bar{y}_{\mathrm{css}}) = \widetilde{S}_y^2 = \bar{Y}^2\widetilde{C}_y^2, \quad \mathrm{V}(\bar{x}_{\mathrm{css}}) = \widetilde{S}_x^2 = \bar{X}^2\widetilde{C}_x^2$$

and

$$\mathrm{Cov}(\bar{y}_{\mathrm{css}}, \bar{x}_{\mathrm{css}}) = \widetilde{S}_{yx} = \bar{Y}\bar{X}\widetilde{C}_{yx}.$$

Considering that regression estimators are always better than ratio estimators, at least asymptotically, we can consider linear regression estimators based on CSS as benchmark for making comparison with our suggested classes of estimators.

2.1. Regression estimator with known $\bar{X}$. The linear regression estimator of the population mean $\bar{Y}$ based on CSS with known $\bar{X}$ can be defined as

$$\bar{y}_{\mathrm{lr(1)c}} = \bar{y}_{\mathrm{css}} + \hat{\beta}_{yx}(\bar{X} - \bar{x}_{\mathrm{css}}), \tag{2.4}$$

where $\hat{\beta}_{yx} = \dfrac{s_{yx}}{s_x^2}$ is an estimator of the population regression coefficient β_{yx} with $s_{yx} = \dfrac{\sum_j^n (y_{ij} - \bar{y}_{\mathrm{css}})(x_{ij} - \bar{x}_{\mathrm{css}})}{n-1}$ and $s_x^2 = \dfrac{\sum_j^n (x_{ij} - \bar{x}_{\mathrm{css}})^2}{n-1}$.

The mean square error of $\bar{y}_{\mathrm{lr(1)c}}$ is given by

$$\mathrm{MSE}(\bar{y}_{\mathrm{lr(1)c}}) = \widetilde{S}_y^2\left(1 - \widetilde{\rho}_{yx}^2\right), \tag{2.5}$$

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

2.2. Regression estimator with unknown $\bar{X}$. When the population mean $\bar{X}$ is unknown, a two-phase sampling design is used. In the first phase, the population is divided into N clusters of size n, each according to CSS, and select randomly m distinct clusters $(1 < m < k)$ to estimate $\bar{X}$ only. In the second phase, a cluster is selected randomly from m circular systematic samples to estimate $\bar{Y}$. Hence, the analogue of $\bar{y}_{\mathrm{lr(1)c}}$, with unknown $\bar{X}$, can be

$$\bar{y}_{\mathrm{lr(2)c}} = \bar{y}_{\mathrm{css}} + \hat{\beta}_{yx}(\bar{x}'_{\mathrm{css}} - \bar{x}_{\mathrm{css}}), \tag{2.6}$$

where $\bar{x}'_{\mathrm{css}} = \dfrac{\sum_i^m \sum_j^n x_{ij}}{mn}$.

We can define

$$\mathrm{V}(\bar{x}'_{\mathrm{css}}) = \mathrm{Cov}(\bar{x}'_{\mathrm{css}}, \bar{x}_{\mathrm{css}}) = \left(\frac{N-1}{N}\right)[1 + (n-1)\rho_x]\,\frac{S_x^2}{nm} = \frac{\widetilde{S}_x^2}{m} \tag{2.7}$$

and

$$\mathrm{Cov}(\bar{y}_{\mathrm{css}}, \bar{x}'_{\mathrm{css}}) = \left(\frac{N-1}{N}\right)[1 + (n-1)\rho_y]^{\frac{1}{2}}\,[1 + (n-1)\rho_x]^{\frac{1}{2}}\,\frac{S_{yx}}{nm} = \frac{\widetilde{S}_{yx}}{m}. \tag{2.8}$$

The mean square error of $\bar{y}_{\mathrm{lr(2)c}}$ is given by

$$\mathrm{MSE}(\bar{y}_{\mathrm{lr(2)c}}) = \widetilde{S}_y^2\left(1 - \frac{\mathbf{S}_{yx}^2}{\widetilde{S}_y^2\mathbf{S}_x^2}\right), \tag{2.9}$$

where

$$\mathbf{S}_x^2 = \left(\frac{m-1}{m}\right)\widetilde{S}_x^2 \quad \text{and} \quad \mathbf{S}_{yx} = \left(\frac{m-1}{m}\right)\widetilde{S}_{yx}.$$

3. Modified classes of estimators

In this Section, we introduce three general classes of estimators for estimating the population mean $\bar{Y}$ using the auxiliary variable X, based on the CSS. The first class is modified by taking motivation from [3], while the second class is motivated from [21] (hereafter SS). We consider Diana et al. and SS classes of estimators because these are more efficient than the regression estimator in simple random sampling without replacement (SRSWOR).

3.1. First class.

3.1.1. *Assuming $\bar{X}$ known.* Diana et al. [3] have proposed a general class of biased estimators of the population mean $\bar{Y}$ using known mean $\bar{X}$ in SRSWOR. We take motivation and propose a class of estimators in CSS assuming $\bar{X}$ is known

$$\mathrm{T}_1 = \left[w_1\bar{y}_{\mathrm{css}} + w_2(\bar{X} - \bar{x}_{\mathrm{css}})\right] \exp\left(\frac{\lambda\left(\bar{X} - \bar{x}_{\mathrm{css}}\right)}{\bar{X} + \bar{x}_{\mathrm{css}}}\right), \tag{3.1}$$

where λ being a constant takes values (0, 1, -1) for designing different estimators and $(w_1,\, w_2)$ are suitably chosen constants.
Expressing the class T_1 in terms of δ's, we have

$$\mathrm{T}_1 = \left[w_1\bar{Y}(1+\delta_y) - w_2\bar{X}\delta_x\right] \exp\left\{-\frac{\lambda\delta_x}{2}\left(1+\frac{\delta_x}{2}\right)^{-1}\right\}, \tag{3.2}$$

where

$$\delta_y = \frac{(\bar{y}_{\mathrm{css}} - \bar{Y})}{\bar{Y}} \quad \text{and} \quad \delta_x = \frac{(\bar{x}_{\mathrm{css}} - \bar{X})}{\bar{X}}.$$

Now expanding T_1 in a first order Taylor's series, we get

$$\mathrm{T}_1 \cong \left[w_1\bar{Y} + w_1\bar{Y}\delta_y - w_2\bar{X}\delta_x - \frac{w_1\bar{Y}\lambda\delta_y\delta_x}{2} + \frac{w_2\bar{X}\lambda\delta_x^2}{2} + w_1\left(\frac{\lambda}{4}+\frac{\lambda^2}{8}\right)\delta_x^2\right]. \tag{3.3}$$

To obtain the bias and the mean square error, let us define

$$E(\delta_y) = E(\delta_x) = 0,$$
$$E(\delta_y^2) = \widetilde{C}_y^2, \quad E(\delta_x^2) = \widetilde{C}_x^2 \quad \text{and} \quad E(\delta_y\delta_x) = \widetilde{C}_{yx}.$$

From now onward, we consider the bias and the mean square error (MSE) of the considered estimators by using a first order Taylor series [[29]].

The bias and the mean square error of T_1 are given by

$$\mathrm{Bias}(\mathrm{T}_1) = \bar{Y}(w_1 - 1) + \left[\frac{w_2\lambda\bar{X}}{2} + w_1\bar{Y}\left(\frac{\lambda}{4}+\frac{\lambda^2}{8}\right)\right]\widetilde{C}_x^2 - \frac{w_1\lambda\bar{Y}}{2}\widetilde{C}_{yx} \tag{3.4}$$

and

$$\begin{aligned}\mathrm{MSE}(\mathrm{T}_1) &= \bar{Y}^2(w_1-1)^2 + w_1^2\bar{Y}^2\widetilde{C}_y^2 - w_1\bar{Y}\left(2w_2\bar{X} + (2w_1-1)\lambda\bar{Y}\right)\widetilde{C}_{yx} \\ &\quad + \left[\frac{4w_2^2\bar{X}^2 + 4w_2\bar{Y}\bar{X}\lambda(2w_1-1) + w_1\bar{Y}^2\lambda\left(2w_1(\lambda+1)-\lambda-2\right)}{4}\right]\widetilde{C}_x^2.\end{aligned} \tag{3.5}$$

Now we can minimize the mean square error of T_1 to get the optimum values of the constants w_1 and w_2

$$w_1 = \frac{\widetilde{C}_x^2\left[\lambda(3\lambda-2)\widetilde{C}_x^2 - 8\right]}{4\left[\lambda(\lambda-1)\widetilde{C}_x^4 - 2\widetilde{C}_x^2(\widetilde{C}_y^2+1) + 2\widetilde{C}_{yx}^2\right]} = w_1^o(\mathrm{say})$$

and

$$w_2 = -\frac{\bar{Y}\left[\lambda^3\widetilde{C}_x^4 + \lambda\widetilde{C}_x^2\left(4\widetilde{C}_y^2 + (2-3\lambda)\widetilde{C}_{yx} - 4\right) - 4\widetilde{C}_{yx}(\lambda\widetilde{C}_{yx} - 2)\right]}{4\bar{X}\left[\lambda(\lambda-1)\widetilde{C}_x^4 - 2\widetilde{C}_x^2(\widetilde{C}_y^2 + 1) + 2\widetilde{C}_{yx}^2\right]} = w_2^o(\text{say}).$$

The minimum mean square error of T_1 can be written as

$$(3.6)\quad \operatorname{minMSE}(\mathrm{T}_1) = \frac{\bar{Y}^2\left[\lambda^2(\lambda-2)^2\widetilde{C}_x^6 + 16\lambda^2\widetilde{C}_y^2\widetilde{C}_x^4 - 16\widetilde{C}_x^2(4\widetilde{C}_y^2 + \lambda^2\widetilde{C}_{yx}^2) + 64\widetilde{C}_{yx}^2\right]}{32\left[\lambda(\lambda-1)\widetilde{C}_x^4 - 2\widetilde{C}_x^2(\widetilde{C}_y^2 + 1) + 2\widetilde{C}_{yx}^2\right]},$$

or, using (5)

$$(3.7)\quad \operatorname{minMSE}(\mathrm{T}_1) = \frac{\bar{Y}^2\lambda^2(\lambda-2)^2\widetilde{C}_x^4 + 16(\lambda\widetilde{C}_x^2 - 1)\operatorname{MSE}(\bar{y}_{\mathrm{lr(1)c}})}{\lambda(\lambda-1)\widetilde{C}_x^2 - 2\left(1 + \dfrac{\operatorname{MSE}(\bar{y}_{\mathrm{lr(1)c}})}{\bar{Y}^2}\right)}.$$

3.1.2. *Assuming $\bar{X}$ unknown.* Now suppose $\bar{X}$ unknown, the analogue of T_1 becomes

$$(3.8)\quad \mathrm{T}_{1(2)} = \left[w_1\bar{y}_{\mathrm{css}} + w_2(\bar{x}'_{\mathrm{css}} - \bar{x}_{\mathrm{css}})\right]\exp\left(\frac{\lambda\left(\bar{x}'_{\mathrm{css}} - \bar{x}_{\mathrm{css}}\right)}{\bar{x}'_{\mathrm{css}} + \bar{x}_{\mathrm{css}}}\right)$$

$$(3.9)\quad = \left[w_1\bar{Y}(1+\delta_y) + w_2\bar{X}(\delta'_x - \delta_x)\right]\exp\left\{\frac{\lambda(\delta'_x - \delta_x)}{2}\left(1 + \frac{(\delta'_x + \delta_x)}{2}\right)^{-1}\right\},$$

where

$$\delta'_x = \frac{(\bar{x}'_{\mathrm{css}} - \bar{X})}{\bar{X}},\quad E(\delta'_x) = 0,$$
$$E(\delta'^2_x) = E(\delta'_x\delta_x) = \frac{\widetilde{C}_x^2}{m}\quad\text{and}\quad E(\delta_y\delta'_x) = \frac{\widetilde{C}_{yx}}{m}.$$

The bias and the mean square error of $\mathrm{T}_{1(2)}$ can be written as

$$(3.10)\quad \operatorname{Bias}(\mathrm{T}_{1(2)}) = \bar{Y}(w_1 - 1) + \left[\frac{w_2\lambda\bar{X}}{2} + w_1\bar{Y}\left(\frac{\lambda}{4} + \frac{\lambda^2}{8}\right)\right]\widetilde{\mathbf{C}}_x^2 - \frac{w_1\lambda\bar{Y}}{2}\widetilde{\mathbf{C}}_{yx}$$

and

$$(3.11)\quad \begin{aligned}\operatorname{MSE}(\mathrm{T}_{1(2)}) &= \bar{Y}^2(w_1-1)^2 + w_1^2\bar{Y}^2\widetilde{C}_y^2 - w_1\bar{Y}\left(2w_2\bar{X} + (2w_1-1)\lambda\bar{Y}\right)\widetilde{\mathbf{C}}_{yx}\\ &+ \left[\frac{4w_2^2\bar{X}^2 + 4w_2\bar{Y}\bar{X}\lambda(2w_1-1) + w_1\bar{Y}^2\lambda\left(2w_1(\lambda+1) - \lambda - 2\right)}{4}\right]\widetilde{\mathbf{C}}_x^2,\end{aligned}$$

with

$$\widetilde{\mathbf{C}}_x^2 = \frac{\widetilde{\mathbf{S}}_x^2}{\bar{X}^2}\quad\text{and}\quad \widetilde{\mathbf{C}}_{yx} = \frac{\widetilde{\mathbf{S}}_{yx}}{\bar{Y}\bar{X}}.$$

We can obtain the minimum mean square error of $\mathrm{T}_{1(2)}$ when

$$w_1 = \frac{\widetilde{\mathbf{C}}_x^2\left[\lambda(3\lambda-2)\widetilde{\mathbf{C}}_x^2 - 8\right]}{4\left[\lambda(\lambda-1)\widetilde{\mathbf{C}}_x^4 - 2\widetilde{\mathbf{C}}_x^2(\widetilde{C}_y^2 + 1) + 2\widetilde{\mathbf{C}}_{yx}^2\right]} = w_1^o(\text{say})$$

and

$$w_2 = -\frac{\bar{Y}\left[\lambda^3\widetilde{\mathbf{C}}_x^4 + \lambda\widetilde{\mathbf{C}}_x^2\left(4\widetilde{C}_y^2 + (2-3\lambda)\widetilde{\mathbf{C}}_{yx} - 4\right) - 4\widetilde{\mathbf{C}}_{yx}(\lambda\widetilde{\mathbf{C}}_{yx} - 2)\right]}{4\bar{X}\left[\lambda(\lambda-1)\widetilde{\mathbf{C}}_x^4 - 2\widetilde{\mathbf{C}}_x^2(\widetilde{C}_y^2 + 1) + 2\widetilde{\mathbf{C}}_{yx}^2\right]} = w_2^o(\text{say}).$$

The minimum mean square error of $\mathrm{T}_{1(2)}$ can be written as

$$\begin{aligned}
&\operatorname{minMSE}(\mathrm{T}_{1(2)}) \\
&= \frac{\bar{Y}^2\left[\lambda^2(\lambda-2)^2\widetilde{\mathbf{C}}_x^6+16\lambda^2\widetilde{C}_y^2\widetilde{\mathbf{C}}_x^4-16\widetilde{\mathbf{C}}_x^2(4\widetilde{C}_y^2+\lambda^2\widetilde{\mathbf{C}}_{yx}^2)+64\widetilde{\mathbf{C}}_{yx}^2\right]}{32\left[\lambda(\lambda-1)\widetilde{\mathbf{C}}_x^4-2\widetilde{\mathbf{C}}_x^2(\widetilde{C}_y^2+1)+2\widetilde{\mathbf{C}}_{yx}^2\right]}.
\end{aligned} \tag{3.12}$$

By using (2.9), minimum MSE of $\mathrm{T}_{1(2)}$ can be expressed as

$$\operatorname{minMSE}(\mathrm{T}_{1(2)}) = \frac{\bar{Y}^2\lambda^2(\lambda-2)^2\widetilde{\mathbf{C}}_x^4+16(\lambda\widetilde{\mathbf{C}}_x^2-1)\operatorname{MSE}(\bar{y}_{\mathrm{lr}(2)\mathrm{c}})}{\lambda(\lambda-1)\widetilde{\mathbf{C}}_x^2-2\left(1+\dfrac{\operatorname{MSE}(\bar{y}_{\mathrm{lr}(2)\mathrm{c}})}{\bar{Y}^2}\right)}. \tag{3.13}$$

From [(3.4), (3.1)] and [(3.7), (3.13)], it can be observed that the bias and the minimum mean square error of T_1 and $\mathrm{T}_{1(2)}$ look similar. However, the dissimilarity exists only in terms ($\widetilde{C}_x^2$, $\widetilde{C}_{yx}$) and ($\widetilde{\mathbf{C}}_x^2$, $\widetilde{\mathbf{C}}_{yx}$) due to single and two-phase sampling.

3.2. Second class.

3.2.1. *Assuming $\bar{X}$ known.* Motivated by SS [21], the following class of estimators has been defined for the population mean $\bar{Y}$ assuming that $\bar{X}$ is known

$$\mathrm{T}_2 = \bar{y}_{\mathrm{css}}\left[w_1\left(\frac{\bar{X}}{\bar{x}_{\mathrm{css}}}\right)^{\eta}+w_2\exp\left(\frac{\lambda(\bar{X}-\bar{x}_{\mathrm{css}})}{\bar{X}+\bar{x}_{\mathrm{css}}}\right)\right], \tag{3.14}$$

where (λ, η) being constants take values (0, 1, -1) for designing different estimators and (w_1, w_2) are suitably chosen constants.

$$\mathrm{T}_2 = \bar{Y}(1+\delta_y)\left[w_1(1+\delta_x)^{-\eta}+w_2\exp\left\{-\frac{\lambda\delta_x}{2}\left(1+\frac{\delta_x}{2}\right)^{-1}\right\}\right]. \tag{3.15}$$

Expanding T_2 in a first order Taylor's series, we get

$$\begin{aligned}
\mathrm{T}_2 \cong{}& w_1\bar{Y}\left(1+\delta_y-\eta(\delta_x+\delta_y\delta_x)+\frac{\eta(\eta+1)}{2}\delta_x^2\right)\\
&+w_2\bar{Y}\left(1+\delta_y-\frac{\lambda}{2}(\delta_x+\delta_y\delta_x)+\frac{\lambda(\lambda+2)\delta_x^2}{8}\right).
\end{aligned} \tag{3.16}$$

The bias and the mean square error of T_2 are given by

$$\begin{aligned}
\operatorname{Bias}(\mathrm{T}_2) ={}& \bar{Y}\left[w_1\left\{1+\frac{\eta}{2}\left((\eta+1)\widetilde{C}_x^2-2\widetilde{C}_{yx}\right)\right\}\right]\\
&+\bar{Y}\left[w_2\left\{1+\frac{\lambda}{8}\left((\lambda+2)\widetilde{C}_x^2-4\widetilde{C}_{yx}\right)\right\}-1\right]
\end{aligned} \tag{3.17}$$

and

$$\operatorname{MSE}(\mathrm{T}_2) = \bar{Y}^2\left[1+w_1^2A+w_2^2C+2w_1w_2D-2w_1B-2w_2E\right], \tag{3.18}$$

where

$$A = 1+\widetilde{C}_y^2+\left(2\eta^2+\eta\right)\widetilde{C}_x^2-4\eta\widetilde{C}_{yx},$$

$$B = 1+\frac{1}{2}\left(\eta^2+\eta\right)\widetilde{C}_x^2-\eta\widetilde{C}_{yx},$$

$$C = 1+\widetilde{C}_y^2+\frac{1}{2}\left(\lambda^2+\lambda\right)\widetilde{C}_x^2-2\lambda\widetilde{C}_{yx},$$

$$D = 1+\widetilde{C}_y^2+\frac{1}{8}\left((2\eta+\lambda)^2+2(2\eta+\lambda)\right)\widetilde{C}_x^2-(2\eta+\lambda)\,\widetilde{C}_{yx}$$

and

$$E = 1 + \frac{1}{8}\left(\lambda^2 + 2\lambda\right)\widetilde{C}_x^2 - \frac{1}{2}\lambda\widetilde{C}_{yx}.$$

The mean square error of T_2 will be minimum when

$$w_1 = \frac{BC - DE}{AC - D^2} = w_1^o(\text{say}) \quad \text{and} \quad w_2 = \frac{AE - BD}{AC - D^2} = w_2^o(\text{say}).$$

We can write the minimum MSE of T_2 as

$$\text{(3.19)} \qquad \min\mathrm{MSE}(\mathrm{T}_2) = \bar{Y}^2\left[1 - \frac{B^2C - 2BDE + AE^2}{AC - D^2}\right].$$

3.2.2. *Assuming $\bar{X}$ unknown.* Now assuming that $\bar{X}$ is unknown, then the analogue of T_2 becomes

$$\text{(3.20)} \qquad \mathrm{T}_{2(2)} = \bar{y}_{\text{css}}\left[w_1\left(\frac{\bar{x}'_{\text{css}}}{\bar{x}_{\text{css}}}\right)^{\eta} + w_2 \exp\left(\frac{\lambda(\bar{x}'_{\text{css}} - \bar{x}_{\text{css}})}{\bar{x}'_{\text{css}} + \bar{x}_{\text{css}}}\right)\right].$$

The bias and the mean square error of $\mathrm{T}_{2(2)}$ will be almost similar of T_2. The difference between the bias and MSE of T_2 and $\mathrm{T}_{2(2)}$ will be the same like the difference between T_1 and $\mathrm{T}_{1(2)}$ explained in Section 3.1. Therefore, replacing the terms $\left(\widetilde{C}_x^2,\ \widetilde{C}_{yx}\right)$ in (3.17) and (3.19) by $\left(\widetilde{\mathbf{C}}_x^2,\ \widetilde{\mathbf{C}}_{yx}\right)$, we can get the bias and the minimum mean square error of $\mathrm{T}_{2(2)}$.

3.3. Third class.

3.3.1. *Assuming $\bar{X}$ known.* Taking motivation from the classes T_1 and T_2, the following class of estimators has been proposed with known $\bar{X}$

$$\text{(3.21)} \qquad \mathrm{T}_3 = \bar{y}_{\text{css}}\left[w_1 + w_2(\bar{X} - \bar{x}_{\text{css}})\right]\exp\left(\frac{\lambda\left(\bar{X} - \bar{x}_{\text{css}}\right)}{\bar{X} + \bar{x}_{\text{css}}}\right),$$

where $(\lambda,\ w_1,\ w_2)$ are defined earlier.
Now using the first order Taylor series to expand T_3 as

$$\text{(3.22)} \qquad \begin{aligned} T_3 \cong \bar{Y}&\left[w_1 + w_1\delta_y - \frac{(w_1\lambda + 2w_2\bar{X})}{2}\delta_x - \frac{(w_1\lambda + 2w_2\bar{X})}{2}\delta_y\delta_x\right] \\ &+ \bar{Y}\left[\frac{w_2\bar{X}\lambda}{2}\delta_x^2 + w_1\left(\frac{\lambda}{4} + \frac{\lambda^2}{8}\right)\delta_x^2\right]. \end{aligned}$$

The bias and the mean square error of T_3 are given by

$$\text{(3.23)} \qquad \begin{aligned} \mathrm{Bias}(\mathrm{T}_3) = \bar{Y}&\left[(w_1 - 1) + \left\{\frac{w_2\lambda\bar{X}}{2} + w_1\left(\frac{\lambda}{4} + \frac{\lambda^2}{8}\right)\right\}\widetilde{C}_x^2\right] \\ &- \bar{Y}\left[\frac{(w_1\lambda + 2w_2\bar{X})}{2}\widetilde{C}_{yx}\right] \end{aligned}$$

and

$$\text{(3.24)} \qquad \begin{aligned} \mathrm{MSE}(T_3) = {}&\bar{Y}^2(w_1 - 1)^2 + w_1^2\bar{Y}^2\widetilde{C}_y^2 + \bar{Y}^2(2w_2\bar{X} + w_1\lambda)(1 - 2w_1)\widetilde{C}_{yx} \\ &+ \bar{Y}^2\left\{\frac{4w_2^2\bar{X}^2 + 4w_2\bar{X}\lambda(2w_1 - 1) + w_1\lambda\left(2w_1(\lambda + 1) - \lambda - 2\right)}{4}\right\}\widetilde{C}_x^2. \end{aligned}$$

Now we can obtain the minimum mean square error of T_3 when

$$w_1 = \frac{\lambda(3\lambda-2)\widetilde{C}_x^4 - 4\widetilde{C}_x^2(3\lambda\widetilde{C}_{yx}+2) + 16\widetilde{C}_{yx}^2}{4\left[\lambda(\lambda-1)\widetilde{C}_x^4 - 2\widetilde{C}_x^2(\widetilde{C}_y^2 + 2\lambda\widetilde{C}_{yx}+1) + 8\widetilde{C}_{yx}^2\right]} = w_1^o(\text{say})$$

and

$$w_2 = \frac{(2\widetilde{C}_{yx} - \lambda\widetilde{C}_x^2)\left(\lambda^2\widetilde{C}_x^2 + 4(\widetilde{C}_y^2 - \lambda\widetilde{C}_{yx} - 1)\right)}{4\bar{X}\left[\lambda(\lambda-1)\widetilde{C}_x^4 - 2\widetilde{C}_x^2(\widetilde{C}_y^2 + 2\lambda\widetilde{C}_{yx}+1) + 8\widetilde{C}_{yx}^2\right]} = w_2^o(\text{say}).$$

The minimum mean square error of T_3 can be written as

$$\min\mathrm{MSE}(\mathrm{T}_3) = \frac{\bar{Y}^2 L_1}{L_2}, \tag{3.25}$$

where

$$L_1 = \lambda^2(\lambda-2)^2\widetilde{C}_x^6 + 8\lambda^2\widetilde{C}_x^4\left(2\widetilde{C}_y^2 + (2-\lambda)\widetilde{C}_{yx}\right) - 16\widetilde{C}_x^2\left(4\widetilde{C}_y^2(1+\lambda\widetilde{C}_{yx}) + \lambda(2-\lambda)\widetilde{C}_{yx}^2\right) + 64\widetilde{C}_{yx}^2(\widetilde{C}_y^2+1)$$

and

$$L_2 = 32\left[\lambda(\lambda-1)\widetilde{C}_x^4 - 2\widetilde{C}_x^2\left(\widetilde{C}_y^2 + 2\lambda\widetilde{C}_{yx}+1\right) + 8\widetilde{C}_{yx}^2\right].$$

Table 2. Some members of the classes of estimators T_*

Estimators	η	λ
$\mathrm{T}_{1(\text{reg})} = \left[w_1\widetilde{y}_{\text{css}} + w_2(\bar{X} - \bar{x}_{\text{css}})\right]$	-	0
$\mathrm{T}_{1(\text{regre})} = \left[w_1\widetilde{y}_{\text{css}} + w_2(\bar{X} - \bar{x}_{\text{css}})\right]\exp\left(\frac{\bar{X} - \bar{x}_{\text{css}}}{\bar{X} + \bar{x}_{\text{css}}}\right)$	-	1
$\mathrm{T}_{1(\text{regpe})} = \left[w_1\widetilde{y}_{\text{css}} + w_2(\bar{X} - \bar{x}_{\text{css}})\right]\exp\left(\frac{\bar{x}_{\text{css}} - \bar{X}}{\bar{X} + \bar{x}_{\text{css}}}\right)$	-	-1
$\mathrm{T}_{2(\text{r})} = \widetilde{y}_{\text{css}}\left[w_1 + w_2\left(\frac{\bar{X}}{\bar{x}_{\text{css}}}\right)\right]$	1	0
$\mathrm{T}_{2(\text{p})} = \widetilde{y}_{\text{css}}\left[w_1 + w_2\left(\frac{\bar{x}_{sy}}{\bar{X}}\right)\right]$	-1	0
$\mathrm{T}_{2(\text{re})} = \widetilde{y}_{\text{css}}\left[w_1 + w_2\exp\left(\frac{\bar{X} - \bar{x}_{css}}{\bar{X} + \bar{x}_{css}}\right)\right]$	0	1
$\mathrm{T}_{2(\text{pe})} = \widetilde{y}_{\text{css}}\left[w_1 + w_2\exp\left(\frac{\bar{x}_{\text{css}} - \bar{X}}{\bar{X} + \bar{x}_{\text{css}}}\right)\right]$	0	-1
$\mathrm{T}_{2(\text{rre})} = \widetilde{y}_{\text{css}}\left[w_1\left(\frac{\bar{X}}{\bar{x}_{\text{css}}}\right) + w_2\exp\left(\frac{\bar{X} - \bar{x}_{\text{css}}}{\bar{X} + \bar{x}_{\text{css}}}\right)\right]$	1	1
$\mathrm{T}_{2(\text{rpe})} = \widetilde{y}_{\text{css}}\left[w_1\left(\frac{\bar{X}}{\bar{x}_{\text{css}}}\right) + w_2\exp\left(\frac{\bar{x}_{\text{css}} - \bar{X}}{\bar{X} + \bar{x}_{\text{css}}}\right)\right]$	1	-1
$\mathrm{T}_{2(\text{pre})} = \widetilde{y}_{\text{css}}\left[w_1\left(\frac{\bar{x}_{\text{css}}}{\bar{X}}\right) + w_2\exp\left(\frac{\bar{X} - \bar{x}_{\text{css}}}{\bar{X} + \bar{x}_{\text{css}}}\right)\right]$	-1	1
$\mathrm{T}_{2(\text{ppe})} = \widetilde{y}_{\text{css}}\left[w_1\left(\frac{\bar{x}_{\text{css}}}{\bar{X}}\right) + w_2\exp\left(\frac{\bar{x}_{\text{css}} - \bar{X}}{\bar{X} + \bar{x}_{\text{css}}}\right)\right]$	-1	-1
$\mathrm{T}_{3(\text{r})} = \widetilde{y}_{\text{css}}\left[w_1 + w_2(\bar{X} - \bar{x}_{\text{css}})\right]$	-	0
$\mathrm{T}_{3(\text{rre})} = \widetilde{y}_{\text{css}}\left[w_1 + w_2(\bar{X} - \bar{x}_{\text{css}})\right]\exp\left(\frac{\bar{X} - \bar{x}_{\text{css}}}{\bar{X} + \bar{x}_{\text{css}}}\right)$	-	1
$\mathrm{T}_{3(\text{rpe})} = \widetilde{y}_{\text{css}}\left[w_1 + w_2(\bar{X} - \bar{x}_{\text{css}})\right]\exp\left(\frac{\bar{x}_{\text{css}} - \bar{X}}{\bar{X} + \bar{x}_{\text{css}}}\right)$	-	-1

3.3.2. *Assuming $\bar{X}$ unknown.* In case of unknown $\bar{X}$, the analogue of T_3 becomes

$$\mathrm{T}_{3(2)} = \bar{y}_{\text{css}}\left[w_1 + w_2(\bar{x}'_{\text{css}} - \bar{x}_{\text{css}})\right]\exp\left(\frac{\lambda\left(\bar{x}'_{\text{css}} - \bar{x}_{\text{css}}\right)}{\bar{x}'_{\text{css}} + \bar{x}_{\text{css}}}\right). \tag{3.26}$$

Then the bias and the mean square error of $\mathrm{T}_{3(2)}$ can be the same like T_3. Because the difference between T_3 and $\mathrm{T}_{3(2)}$ will be the same like the difference which is in T_1 and $\mathrm{T}_{1(2)}$ i.e. the presence of $\left(\widetilde{\mathbf{C}}_x^2,\ \widetilde{\mathbf{C}}_{yx}\right)$ instead of $\left(\widetilde{C}_x^2,\ \widetilde{C}_{yx}\right)$.
There are many ways to construct classes of estimators under the class $(\mathrm{T}_1,\ \mathrm{T}_2,\ \mathrm{T}_3)$. SS have discussed numerous estimators members of their proposed class in SRSWOR. It is observed that use of known population parameters of the auxiliary variable give very less contribution for increasing efficiency of estimators. Due to this reason, we consider only those parameters which are giving support not only in designing the estimators but also in increasing the efficiency of the estimators. In Table 2, some estimators are considered that are members of the classes $(\mathrm{T}_1,\ \mathrm{T}_2,\ \mathrm{T}_3)$.

4. Efficiency comparison

As stated in [2], [page-208], the mean of a circular systematic sample is more precise than the mean of a simple random sample if and only if $S^2_{wsy} > S^2_y$, or equivalently $\rho_y < -\dfrac{1}{N-1}$, where $S^2_{wsy} = \dfrac{\sum_i^N \sum_j^n (y_{ij} - \bar{y}_{i.})^2}{N(n-1)}$ is the variance of within units of the same circular systematic sample. The same conditions hold also for CSS with respect to SRSWOR.

Accordingly, to compare the performance of the proposed classes of estimators based on CSS with the estimators based on SRSWOR, we can use the following estimators $\left(\bar{y},\ \bar{y}_{\text{lr}(1)},\ \bar{y}_{\text{lr}(2)}\right)$ in SRSWOR

$$\bar{y} = \frac{\sum_{j=1}^n y_j}{n}, \quad \bar{x} = \frac{\sum_{j=1}^n x_j}{n}, \quad \bar{x}' = \frac{\sum_{j=1}^{n'} x_j}{n'},$$

$$\bar{y}_{\text{lr}(1)} = \bar{y} + \hat{\beta}(\bar{X} - \bar{x}),$$

where $\hat{\beta} = \dfrac{s_{yx}}{s_x^2}$ is an estimator of the population regression coefficient β with $s_{yx} = \dfrac{\sum_{j=1}^n (y_j - \bar{y})(x_j - \bar{x})}{n-1}$ and $s_x^2 = \dfrac{\sum_{j=1}^n (x_j - \bar{x})^2}{n-1}$.

$$\bar{y}_{\text{lr}(2)} = \bar{y} + \hat{\beta}(\bar{x}' - \bar{x}),$$

$$\mathrm{V}(\bar{y}) = \theta S_y^2,$$

$$\mathrm{MSE}(\bar{y}_{\text{lr}(1)}) = \theta S_y^2\left(1 - \rho_{yx}^2\right),$$

and

$$\mathrm{MSE}(\bar{y}_{\text{lr}(2)}) = \theta' S_y^2 + \theta^* S_y^2\left(1 - \rho_{yx}^2\right)$$

where

$$\rho_{yx} = \frac{S_{yx}}{S_y S_x}, \quad \theta = \left(\frac{1}{n} - \frac{1}{N}\right), \quad \theta' = \left(\frac{1}{n'} - \frac{1}{N}\right), \quad \theta^* = \theta - \theta'.$$

Note: Here using SRSWOR design, at first phase a large sample s' of size n' $(n' < N)$ is selected randomly to estimate $\bar{X}$ only. In second phase, a sub-sample s of size n from n' units $(n < n')$ is drawn randomly to estimate $\bar{Y}$ where $n' = mn$.

Remark: From eq. (3.13), (3.19) and (3.25), it is not easy to make analytical comparison of the proposed classes of estimators with respect to regression estimators. To get numerical results of the minimum MSE of the considered estimators in CSS along with the minimum MSE of the estimators in SRSWOR, we use population data set as earlier considered by [9] and [21]. The data concerns primary and secondary schools of 923 districts of Turkey in 2007. The description of variables is given below

$$y = \text{number of teachers in both primary and secondary school;}$$

$$x = \text{number of students in both primary and secondary school.}$$

Table 3. The minimum MSE of the considered estimators

Estimators in SRSWOR	MSE	Estimators in CSS	MSE
$\bar{y}$	2515.17	$\bar{y}_{\text{css}}$	1698.75
$\bar{y}_{\text{lr}(1)}$	224.62	$\bar{y}_{\text{lr}(1)\text{c}}$	151.71
$\text{t}_{1(\text{reg})}$	224.35	$\text{T}_{1(\text{reg})}$	151.59
$\text{t}_{1(\text{regre})}$	222.76	$\text{T}_{1(\text{regre})}$	151.06
$\text{t}_{1(\text{regpe})}$	220.44	$\text{T}_{1(\text{regpe})}$	**150.66**
$\text{t}_{2(\text{r})}$	223.10	$\text{T}_{2(\text{r})}$	151.55
$\text{t}_{2(\text{p})}$	201.40	$\text{T}_{2(\text{p})}$	141.24
$\text{t}_{2(\text{re})}$	224.55	$\text{T}_{2(\text{re})}$	146.54
$\text{t}_{2(\text{pe})}$	213.87	$\text{T}_{2(\text{pe})}$	146.34
$\text{t}_{2(\text{rre})}$	223.56	$\text{T}_{2(\text{rre})}$	151.56
$\text{t}_{2(\text{ppe})}$	158.05	$\text{T}_{2(\text{ppe})}$	**125.66**
$\text{t}_{2(\text{rpe})}$	222.56	$\text{T}_{2(\text{rpe})}$	151.54
$\text{t}_{2(\text{pre})}$	222.20	$\text{T}_{2(\text{pre})}$	149.77
$\text{t}_{3(\text{r})}$	201.40	$\text{T}_{3(\text{r})}$	141.24
$\text{t}_{3(\text{rre})}$	223.80	$\text{T}_{3(\text{rre})}$	150.82
$\text{t}_{3(\text{rpe})}$	110.43	$\text{T}_{3(\text{rpe})}$	**109.12***

$$N = 923, \quad n' = 360, \quad n = 180, \quad m = 2, \quad \bar{Y} = 436.43, \quad \bar{X} = 11440.50,$$

$$S_y = 749.94, \quad S_x = 21331.13, \quad \rho_{yx} = 0.9543, \quad \rho_y = -0.00255, \quad \rho_x = -0.00316.$$

For two-phases, one can select $1 < m < 5$ (as we mentioned earlier $1 < m < k$). All possible values of m are considered and numerical results are provided only for $m = 2$. Because it is observed that for $m = 2$, all the considered estimators are more efficient in CSS than SRSWOR. For $m = 3$ and $m = 4$, the estimators under SRSWOR perform a little better than CSS. So in this numerical example $m = 2$ can be the best choice among others.

Following is the description of the considered estimators in Tables 3 and 4. The estimators $\left(\bar{y},\ \bar{y}_{\text{lr}(1)},\ t_*\right)$ are based on SRSWOR and $\left(\bar{y}_{\text{css}},\ \bar{y}_{\text{lr}(1)\text{c}},\ T_*\right)$ on CSS, the estimators $\left(\bar{y}_{\text{lr}(2)},\ \bar{y}_{\text{lr}(2)\text{c}},\ t_{*(2)},\ T_{*(2)}\right)$ are considered for two-phase sampling. Moreover, to highlight the numerical quantities in tables, we use "bold" to indicate the minimum MSE of the estimator in the same class and "bold" with "∗" to show the minimum MSE in all considered classes of estimators.

In Table 3, it can be seen that the variance of $\bar{y}_{\mathrm{css}}$ is smaller than the variance of $\bar{y}$. Also, the mean square error of $\bar{y}_{\mathrm{lr}(1)\mathrm{c}}$ is smaller than the mean square error of $\bar{y}_{\mathrm{lr}(1)}$. Hence, we can conclude that the estimators based on CSS are more efficient than the estimators based on SRSWOR. Note that ρ_y and ρ_x both are less than $-\frac{1}{(N-1)}$. It is also observed that all the considered estimators $\left(\mathrm{T}_{1(.)},\ \mathrm{T}_{2(.)},\ \mathrm{T}_{3(.)}\right)$ are more efficient than the regression estimator $\bar{y}_{\mathrm{lr}(1)}c$. Furthermore, the estimators $\mathrm{T}_{1(\mathrm{regpe})}$ in class T_1, $\mathrm{T}_{2(\mathrm{ppe})}$ in class T_2 and $\mathrm{T}_{3(\mathrm{rpe})}$ in class T_3 provide the minimum mean square error among others. Henceforth, the estimator $T_{3(\mathrm{rpe})}$ results the best one in terms of efficiency among all considered estimators.

In Table 4, it is observed that the mean square error of the regression estimator $\bar{y}_{\mathrm{lr}(2)\mathrm{c}}$ is higher, as expected, than of $\bar{y}_{\mathrm{lr}(1)\mathrm{c}}$ in Table 3. Also, all the considered estimators $\left(\mathrm{T}_{1(.)(2)},\ \mathrm{T}_{2(.)(2)},\ \mathrm{T}_{3(.)(2)}\right)$ are more efficient than the regression estimator $\bar{y}_{\mathrm{lr}(2)\mathrm{c}}$. Moreover, the mean square error of the estimator $T_{1(\mathrm{regre})(2)}$ is minimum as compared to all other considered estimators.

Hence, from Tables 3 and 4, we can conclude that the class T_3 in case of single-phase and T_1 for two-phase may be the best choice among others.

Table 4. The minimum MSE of the considered estimators in two phase

Estimators in SRSWOR	MSE	Estimators in CSS	MSE
$\bar{y}_{\mathrm{lr}(2)}$	1092.44	$\bar{y}_{\mathrm{lr}(2)\mathrm{c}}$	925.23
$\mathrm{t}_{1(\mathrm{reg})(2)}$	1086.21	$\mathrm{T}_{1(\mathrm{reg})(2)}$	920.76
$\mathrm{t}_{1(\mathrm{regre})(2)}$	1083.31	$\mathrm{T}_{1(\mathrm{regre})(2)}$	**919.74**∗
$\mathrm{t}_{1(\mathrm{regpe})}$	1091.58	$\mathrm{T}_{1(\mathrm{regpe})(2)}$	923.17
$\mathrm{t}_{2(\mathrm{r})(2)}$	1083.80	$\mathrm{T}_{2(\mathrm{r})(2)}$	920.64
$\mathrm{t}_{2(\mathrm{p})(2)}$	1091.86	$\mathrm{T}_{2(\mathrm{p})(2)}$	925.11
$\mathrm{t}_{2(\mathrm{re})(2)}$	1088.36	$\mathrm{T}_{2(\mathrm{re})(2)}$	922.36
$\mathrm{t}_{2(\mathrm{pe})(2)}$	1092.41	$\mathrm{T}_{2(\mathrm{pe})(2)}$	924.59
$\mathrm{t}_{2(\mathrm{rre})(2)}$	1084.52	$\mathrm{T}_{2(\mathrm{rre})(2)}$	920.67
$\mathrm{t}_{2(\mathrm{ppe})(2)}$	1084.71	$\mathrm{T}_{2(\mathrm{ppe})(2)}$	924.91
$\mathrm{t}_{2(\mathrm{rpe})(2)}$	1083.05	$\mathrm{T}_{2(\mathrm{rpe})(2)}$	**920.61**
$\mathrm{t}_{2(\mathrm{pre})(2)}$	1090.94	$\mathrm{T}_{2(\mathrm{pre})(2)}$	923.66
$\mathrm{t}_{3(\mathrm{r})(2)}$	1091.86	$\mathrm{T}_{3(\mathrm{r})(2)}$	925.11
$\mathrm{t}_{3(\mathrm{rre})(2)}$	1089.80	$\mathrm{T}_{3(\mathrm{rre})(2)}$	**923.05**
$\mathrm{t}_{3(\mathrm{rpe})(2)}$	1074.08	$\mathrm{T}_{3(\mathrm{rpe})(2)}$	923.67

5. Non-response problem in CSS

When a sample of size n is selected from N circular systematic samples to collect information of Y, then incomplete or missing information might be present. The reasons non-response problem occurrence may vary in different situations. For instance, the reasons of non-response in the data set considered in previous the section may be due to strikes, holidays etc.

When non-response occurs in a CSS, we can follow the well-known [8] non-respondents sub-sampling technique. Suppose that n_1 units out of n can supply information on Y and remaining $n_2 = n - n_1$ units are taken as non-respondents. Following the technique of [8], a sub sample of size $n_r = \dfrac{n_2}{l}$, $(l > 1)$ is selected by SRSWOR from n_2 non-respondent units. Assume that all n_r units show full response on second call (of course n_r must be an integer and if it isn't so, it is necessary to round). The population is said to be divided into two groups U_1 and U_2 of sizes N_1 and N_2, where U_1 is a group of respondents that would give response on the first call and U_2 is non-respondents group which could respond on the second call. Obviously N_1 and N_2 are unknown quantities.

One can define the unbiased estimator for the population mean $\bar{Y}$ in CSS assuming non-response in Y

$$\bar{y}^*_{\text{css}} = d_1 \bar{y}_1 + d_2 \bar{y}_{2r}, \tag{5.1}$$

where

$$\bar{y}_1 = \frac{\sum_{j=1}^{n_1} y_{ij}}{n_1}, \quad \bar{y}_{2r} = \frac{\sum_{j=1}^{n_r} y_{ij}}{n_r}, \quad d_1 = \frac{n_1}{n} \quad \text{and} \quad d_2 = \frac{n_2}{n}.$$

The variance of $\bar{y}^*_{\text{css}}$ can be written as

$$\mathrm{V}(\bar{y}^*_{\text{css}}) = \widetilde{S}^2_y + \omega S^2_{y(2)} = \widetilde{\mathbf{S}}^2_y = \bar{Y}^2 \widetilde{\mathbf{C}}^2_y, \tag{5.2}$$

where

$$S^2_{y(2)} = \frac{\sum_{i=1}^{N_2} \sum_{j=1}^{n_2} (y_{ij} - \bar{Y}_2)^2}{n_2(N_2 - 1)}, \quad \bar{Y}_2 = \frac{\sum_i^{N_2} y_{ij}}{N_2} \quad \text{and} \quad \omega = \frac{N_2(l-1)}{nN}.$$

The linear regression estimator defined in (2.4), in case of non-response in Y, can be written as

$$\bar{y}^*_{\text{lr}(1)\text{c}} = \bar{y}^*_{\text{css}} + \hat{\beta}^*_{yx}(\bar{X} - \bar{x}_{\text{css}}), \tag{5.3}$$

where $\hat{\beta}^*_{yx} = \dfrac{s^*_{yx}}{s^2_x}$ is an estimator of the population regression coefficient β_{yx} with

$$s^*_{yx} = \frac{\sum_{j=1}^{n_1} y_{ij} x_{ij} + l \sum_{j=1}^{n_r} y_{ij} x_{ij} - n \bar{x}_{\text{css}} \bar{y}^*_{\text{css}}}{n-1}$$

The mean square error of $\bar{y}^*_{\text{lr}(1)\text{c}}$ is given by

$$\mathrm{MSE}(\bar{y}^*_{\text{lr}(1)\text{c}}) = \widetilde{S}^2_y \left(1 - \widetilde{\rho}^2_{yx}\right) + \omega S^2_{y(2)}. \tag{5.4}$$

When $\bar{X}$ unknown and non-response in Y, then (2.6) becomes

$$\bar{y}^*_{\text{lr}(2)\text{c}} = \bar{y}^*_{\text{css}} + \hat{\beta}^*_{yx}(\bar{x}'_{\text{css}} - \bar{x}_{\text{css}}). \tag{5.5}$$

The mean square error of $\bar{y}^*_{\text{lr}(2)\text{c}}$ is given by

$$\mathrm{MSE}(\bar{y}^*_{\text{lr}(2)\text{c}}) = \widetilde{S}^2_y \left(1 - \frac{\widetilde{\mathbf{S}}^2_{yx}}{\widetilde{S}^2_y \widetilde{\mathbf{S}}^2_x}\right) + \omega S^2_{y(2)}. \tag{5.6}$$

Recently [23] suggested a class of estimators using linear systematic sampling design assuming non-response in Y and known $\bar{X}$. We can define same class using CSS as

$$t_{\mathrm{S}}^{*}=\left[w_1\bar{y}_{\mathrm{css}}^{*}+w_2(\bar{X}-\bar{x}_{\mathrm{css}})\right]\left(\frac{\bar{X}}{\bar{x}_{\mathrm{css}}}\right)^{p}, \tag{5.7}$$

where w_1, w_2 and p are constants.

For $p=1$, the minimum mean square error of t_{S}^{*} is given by

$$\operatorname{minMSE}(t_{\mathrm{S}}^{*})=\frac{(1-\widetilde{C}_x^2)\operatorname{MSE}(\bar{y}_{\mathrm{lr(1)c}}^{*})}{(1-\widetilde{C}_x^2)+\dfrac{\operatorname{MSE}(\bar{y}_{\mathrm{lr(1)c}}^{*})}{\bar{Y}^2}} \tag{5.8}$$

5.1. Classes under non-response. Now we can express suggested classes $(\mathrm{T}_1,\ \mathrm{T}_2,\ \mathrm{T}_3)$ in presence of non-response in Y.

The analogue of the class of estimators T_1 becomes

$$\mathrm{T}_1^{*}=\left[w_1\bar{y}_{\mathrm{css}}^{*}+w_2(\bar{X}-\bar{x}_{\mathrm{css}})\right]\exp\left(\frac{\lambda\left(\bar{X}-\bar{x}_{\mathrm{css}}\right)}{\bar{X}+\bar{x}_{\mathrm{css}}}\right). \tag{5.9}$$

The bias of T_1^{*} will be same of T_1. The minimum mean square error of T_1^{*} can be written as

$$\operatorname{minMSE}(\mathrm{T}_1^{*})=\frac{\bar{Y}^2\left[\lambda^2(\lambda-2)^2\widetilde{C}_x^6+16\lambda^2\widetilde{\mathbf{C}}_y^2\widetilde{C}_x^4-16\widetilde{C}_x^2(4\widetilde{\mathbf{C}}_y^2+\lambda^2\widetilde{C}_{yx}^2)+64\widetilde{C}_{yx}^2\right]}{32\left[\lambda(\lambda-1)\widetilde{C}_x^4-2\widetilde{C}_x^2(\widetilde{\mathbf{C}}_y^2+1)+2\widetilde{C}_{yx}^2\right]}. \tag{5.10}$$

The analogue of T_2 becomes

$$\mathrm{T}_2^{*}=\bar{y}_{\mathrm{css}}^{*}\left[w_1\left(\frac{\bar{X}}{\bar{x}_{\mathrm{css}}}\right)^{\eta}+w_2\exp\left(\frac{\lambda(\bar{X}-\bar{x}_{\mathrm{css}})}{\bar{X}+\bar{x}_{\mathrm{css}}}\right)\right] \tag{5.11}$$

and the minimum mean square error of T_2^{*} can be expressed as

$$\operatorname{minMSE}(\mathrm{T}_2^{*})=\bar{Y}^2\left[1-\frac{B^2C^{*}-2BD^{*}E+A^{*}E^2}{A^{*}C^{*}-D^{*2}}\right], \tag{5.12}$$

where

$$A^{*}=1+\widetilde{\mathbf{C}}_y^2+\left(2\eta^2+\eta\right)\widetilde{C}_x^2-4\eta\widetilde{C}_{yx},$$
$$C^{*}=1+\widetilde{\mathbf{C}}_y^2+\frac{1}{2}\left(\lambda^2+\lambda\right)\widetilde{C}_x^2-2\lambda\widetilde{C}_{yx}$$

and

$$D^{*}=1+\widetilde{\mathbf{C}}_y^2+\frac{1}{8}\left((2\eta+\lambda)^2+2(2\eta+\lambda)\right)\widetilde{C}_x^2-(2\eta+\lambda)\,\widetilde{C}_{yx}.$$

The analogue of T_3 becomes

$$\mathrm{T}_3^{*}=\bar{y}_{\mathrm{css}}^{*}\left[w_1+w_2(\bar{X}-\bar{x}_{\mathrm{css}})\right]\exp\left(\frac{\lambda\left(\bar{X}-\bar{x}_{\mathrm{css}}\right)}{\bar{X}+\bar{x}_{\mathrm{css}}}\right). \tag{5.13}$$

The minimum mean square error of T_3^{*} can be written as

$$\operatorname{minMSE}(\mathrm{T}_3^{*})=\frac{\bar{Y}^2L_1^{*}}{L_2^{*}}, \tag{5.14}$$

where

$$\begin{aligned}L_1^{*}&=\lambda^2(\lambda-2)^2\widetilde{C}_x^6+8\lambda^2\widetilde{C}_x^4\left(2\widetilde{\mathbf{C}}_y^2+(2-\lambda)\widetilde{C}_{yx}\right)\\&\quad-16\widetilde{C}_x^2\left(4\widetilde{\mathbf{C}}_y^2(1+\lambda\widetilde{C}_{yx})+\lambda(2-\lambda)\widetilde{C}_{yx}^2\right)+64\widetilde{C}_{yx}^2(\widetilde{\mathbf{C}}_y^2+1)\end{aligned}$$

and

$$L_2^* = 32\left[\lambda(\lambda-1)\widetilde{C}_x^4 - 2\widetilde{C}_x^2\left(\widetilde{\mathbf{C}}_y^2 + 2\lambda\widetilde{C}_{yx} + 1\right) + 8\widetilde{C}_{yx}^2\right].$$

Recently [25] proposed a class of estimators under SRSWOR sampling when non-response is present in the study variable. If we consider the same class in CSS design, then it becomes member of class T_1^* for $\lambda = 1$

$$\mathrm{T}_{\mathrm{RDS}}^* = \left[w_1\bar{y}_{\mathrm{css}}^* + w_2(\bar{X} - \bar{x}_{\mathrm{css}})\right]\exp\left(\frac{\bar{X} - \bar{x}_{\mathrm{css}}}{\bar{X} + \bar{x}_{\mathrm{css}}}\right). \tag{5.15}$$

When $\bar{X}$ is unknown, the biases of classes $(\mathrm{T}_{1(2)}^*,\ \mathrm{T}_{2(2)}^*,\ \mathrm{T}_{3(2)}^*)$ will be same of $(\mathrm{T}_{1(2)},\ \mathrm{T}_{2(2)},\ \mathrm{T}_{3(2)})$. For the minimum mean square errors of these classes, only replacing the terms $\left(\widetilde{C}_x^2,\ \widetilde{C}_{yx}\right)$ in (5.10), (5.12) and (5.14) by $\left(\widetilde{\mathbf{C}}_x^2,\ \widetilde{\mathbf{C}}_{yx}\right)$.

5.2. Numerical illustration. To make efficiency comparison of classes $(\mathrm{T}_1^*,\ \mathrm{T}_2^*,\ \mathrm{T}_3^*)$, we can use the estimators $\bar{y}^*$, $\bar{y}_{\mathrm{lr}(1)}^*$ and $\bar{y}_{\mathrm{lr}(2)}^*$ in SRSWOR

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

where

$$\bar{y}_1 = \frac{\sum_{j=1}^{n_1} y_j}{n_1}, \quad \bar{y}_{2r} = \frac{\sum_{j=1}^{n_r} y_j}{n_r}, \quad d_1 = \frac{n_1}{n} \quad \text{and} \quad d_2 = \frac{n_2}{n},$$

$$\mathrm{V}(\bar{y}^*) = \theta S_y^2 + \omega S_{y(2)}^2,$$

$$\bar{y}_{\mathrm{lr}(1)}^* = \bar{y}^* + \hat{\beta}_{yx}^{**}(\bar{X} - \bar{x}),$$

where $\hat{\beta}_{yx}^{**} = \dfrac{s_{yx}^{**}}{s_x^2}$ is an estimator of the population regression coefficient β_{yx} with

$$s_{yx}^{**} = \frac{\sum_{j=1}^{n_1} y_j x_j + l\sum_{j=1}^{n_r} y_j x_j - n\bar{x}\bar{y}^*}{n-1}.$$

$$\mathrm{MSE}(\bar{y}_{\mathrm{lr}(1)}^*) = \theta S_y^2\left(1 - \rho_{yx}^2\right) + \omega S_{y(2)}^2,$$

$$\bar{y}_{\mathrm{lr}(2)}^* = \bar{y}^* + \hat{\beta}_{yx}^{**}(\bar{x}' - \bar{x}),$$

and

$$\mathrm{MSE}(\bar{y}_{\mathrm{lr}(2)}^*) = \theta' S_y^2 + \theta^* S_y^2\left(1 - \rho_{yx}^2\right) + \omega S_{y(2)}^2.$$

Table 5. The minimum MSE of the considered estimators

Estimators in SRSWOR	MSE	Estimators in CSS	MSE
$\bar{y}^*$	2940.51	$\bar{y}^*_{\mathrm{css}}$	2124.09
$\bar{y}^*_{\mathrm{lr}(1)}$	649.96	$\bar{y}^*_{\mathrm{lr}(1)\mathrm{c}}$	577.05
t_{S}	647.71	t^*_{S}	575.29
$\mathrm{t}^*_{1(\mathrm{reg})}$	647.75	$\mathrm{T}^*_{1(\mathrm{reg})}$	575.30
$\mathrm{t}^*_{1(\mathrm{regre})} = \mathrm{T}_{\mathrm{RDS}}$	644.51	$\mathrm{T}^*_{1(\mathrm{regre})} = \mathrm{T}^*_{\mathrm{RDS}}$	**573.89**
$\mathrm{t}^*_{1(\mathrm{regpe})}$	648.83	$\mathrm{T}^*_{1(\mathrm{regpe})}$	577.04
$\mathrm{t}^*_{2(\mathrm{r})}$	261.76	$\mathrm{T}^*_{2(\mathrm{r})}$	189.69
$\mathrm{t}^*_{2(\mathrm{p})}$	230.68	$\mathrm{T}^*_{2(\mathrm{p})}$	173.23
$\mathrm{t}^*_{2(\mathrm{re})}$	262.12	$\mathrm{T}^*_{2(\mathrm{re})}$	188.73
$\mathrm{t}^*_{2(\mathrm{pe})}$	246.80	$\mathrm{T}^*_{2(\mathrm{pe})}$	180.55
$\mathrm{t}^*_{2(\mathrm{rre})}$	261.91	$\mathrm{T}^*_{2(\mathrm{rre})}$	189.68
$\mathrm{t}^*_{2(\mathrm{ppe})}$	179.02	$\mathrm{T}^*_{2(\mathrm{ppe})}$	**153.10***
$\mathrm{t}^*_{2(\mathrm{rpe})}$	261.60	$\mathrm{T}^*_{2(\mathrm{rpe})}$	189.69
$\mathrm{t}^*_{2(\mathrm{pre})}$	257.56	$\mathrm{T}^*_{2(\mathrm{pre})}$	185.30
$\mathrm{t}^*_{3(\mathrm{r})}$	635.35	$\mathrm{T}^*_{3(\mathrm{r})}$	572.00
$\mathrm{t}^*_{3(\mathrm{rre})}$	649.95	$\mathrm{T}^*_{3(\mathrm{rre})}$	577.05
$\mathrm{t}^*_{3(\mathrm{rpe})}$	556.65	$\mathrm{T}^*_{3(\mathrm{rpe})}$	**546.57**

We can have different choices for weights of the missing values (10%, 20%, 30%, 40%) etc. We take all these possibilities and observe that the relative efficiency of the considered estimators is not affected by different weights of missing values. Although numerical results are different for different weights, the behavior of results is similar in all cases. Hence, numerical results are provided only for 10% weight of missing values and consider last 92 values as non-respondents.

$$\bar{Y}_2 = 522.80, \qquad S_{y(2)} = 876.42, \qquad N_2 = 92, \qquad l = 2.$$

Remarks: Due to the inclusion of non-response problem, extra variability is introduced in the estimators. As expected, the variability of all considered estimators with incomplete information (see Tables 5 and 6) are higher than the estimators with complete response (Tables 3 and 4). Additionally, as expected, for $l > 2$, the mean square errors of estimators become higher, so we show results only for $l = 2$. Again the estimators based on CSS are more efficient than the estimators based on SRSWOR. In Table 5, the estimators $\left(t^*_{\mathrm{S}},\ \mathrm{T}^*_{1(.)},\ \mathrm{T}^*_{2(.)},\ \mathrm{T}^*_{3(.)}\right)$ are more efficient than the regression estimator $\bar{y}^*_{\mathrm{lr}(1)\mathrm{c}}$. The estimator $\mathrm{T}^*_{1(\mathrm{regre})}$ is more efficient than [23] estimator t^*_{S}. Furthermore, for efficiency, the estimators $\mathrm{T}^*_{2(\mathrm{ppe})}$ for single-phase in Table 5 and $\mathrm{T}^*_{2(\mathrm{rre})}$ for two-phase in Table 6 are observed the best ones among others.

Table 6. The minimum MSE of the considered estimators in two-phase

Estimators in SRSWOR	MSE	Estimators in CSS	MSE
$\bar{y}^*_{\mathrm{lr}(2)}$	1517.78	$\bar{y}^*_{\mathrm{lr}(2)\mathrm{c}}$	1350.57
$\mathrm{t}^*_{1(\mathrm{reg})(2)}$	1505.78	$\mathrm{T}^*_{1(\mathrm{reg})(2)}$	1341.06
$\mathrm{t}^*_{1(\mathrm{regre})(2)}$	1501.87	$\mathrm{T}^*_{1(\mathrm{regre})(2)}$	**1339.61**
$\mathrm{t}^*_{1(\mathrm{regpe})(2)}$	1514.17	$\mathrm{T}^*_{1(\mathrm{regpe})(2)}$	1344.78
$\mathrm{t}^*_{2(\mathrm{r})(2)}$	1267.50	$\mathrm{T}^*_{2(\mathrm{r})(2)}$	1150.94
$\mathrm{t}^*_{2(\mathrm{p})(2)}$	1276.38	$\mathrm{T}^*_{2(\mathrm{p})(2)}$	1156.70
$\mathrm{t}^*_{2(\mathrm{re})(2)}$	1272.69	$\mathrm{T}^*_{2(\mathrm{re})(2)}$	1153.12
$\mathrm{t}^*_{2(\mathrm{pe})(2)}$	1277.17	$\mathrm{T}^*_{2(\mathrm{pe})(2)}$	1156.01
$\mathrm{t}^*_{2(\mathrm{rre})(2)}$	1267.81	$\mathrm{T}^*_{2(\mathrm{rre})(2)}$	**1150.71***
$\mathrm{t}^*_{2(\mathrm{ppe})(2)}$	1268.23	$\mathrm{T}^*_{2(\mathrm{ppe})(2)}$	1156.61
$\mathrm{t}^*_{2(\mathrm{rpe})(2)}$	1267.18	$\mathrm{T}^*_{2(\mathrm{rpe})(2)}$	1151.16
$\mathrm{t}^*_{2(\mathrm{pre})(2)}$	1275.80	$\mathrm{T}^*_{2(\mathrm{pre})(2)}$	1154.94
$\mathrm{t}^*_{3(\mathrm{r})(2)}$	1517.73	$\mathrm{T}^*_{3(\mathrm{r})(2)}$	1348.81
$\mathrm{t}^*_{3(\mathrm{rre})(2)}$	1511.05	$\mathrm{T}^*_{3(\mathrm{rre})(2)}$	**1344.59**
$\mathrm{t}^*_{3(\mathrm{rpe})(2)}$	1506.99	$\mathrm{T}^*_{3(\mathrm{rpe})(2)}$	1350.50

6. Simulation study

In Section 3, we can see that all the minimum mean square errors consist on the population parameters e.g means, variances and covariances. In Section 4, the efficiency comparisons were performed with the assumption that all these population parameters are known. But in many real situations, these parameters are generally unknown and can not be guessed on the basis of previous data or a pilot survey. Hence they need to be estimated. In such situations, an extra source of variability is introduced in the estimates that could invalid the theoretical comparisons. In this section, we are concentrating our attention to the efficiency comparisons when unknown population parameters are estimated from the selected sample. The empirical performance of the estimators is analyzed by using a Monte Carlo simulation.

The simulation design is arranged as follows: we run a numerical study by considering a population of N=100,000 values. A variable $X \sim G(a,b)$ is generated from gamma distribution with parameters (a=2.2, b=3.5) and a variable Y which is related with X is defined by a model as $y_i = Rx_i + \varepsilon x_i^g$ where $\varepsilon \sim N(0,1)$, R=(1.0, 1.5, 2.0) and g=1.5. This model is earlier considered by [4] for SRSWOR sampling design. Here, Circular systematic sampling is considered for sample sizes n=(100, 500). The sampling has been replicated $B = 1,000$ times.

We investigate the behavior of the following estimators for different values of ρ_{yx} $\left(\bar{y}_{\mathrm{lr}(1)\mathrm{c}},\ \mathrm{T}_{1(.)},\ \mathrm{T}_{2(.)},\ \mathrm{T}_{3(.)}\right)$. For each considered estimator, the simulated mean square

Table 7. Simulated results when population parameters are estimated

Estimators	$\rho = 0.70$		$\rho = 0.59$		$\rho = 0.43$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
Simulated PRE						
$\bar{y}_{\mathrm{lr(1)c}}$	100	100	100	100	100	100
$\mathrm{T}_{1(\mathrm{reg})}$	99.89	99.99	99.90	99.99	99.90	99.99
$\mathrm{T}_{1(\mathrm{regre})}$	76.55	92.75	72.77	91.52	69.04	90.26
$\mathrm{T}_{1(\mathrm{regpe})}$	96.52	99.01	95.25	98.92	93.69	98.77
$\mathrm{T}_{2(\mathrm{r})}$	**100.98**	**100.33**	**101.33**	**100.39**	**101.65**	**100.44**
$\mathrm{T}_{2(\mathrm{p})}$	95.98	97.95	98.03	98.60	99.88	99.23
$\mathrm{T}_{2(\mathrm{re})}$	**100.06**	99.78	**100.72**	99.97	**101.33**	**100.16**
$\mathrm{T}_{2(\mathrm{pe})}$	97.61	98.59	99.10	99.08	**100.47**	99.55
$\mathrm{T}_{2(\mathrm{rre})}$	89.16	96.19	87.64	95.69	86.03	95.16
$\mathrm{T}_{2(\mathrm{rpe})}$	**104.40**	**104.04**	**104.41**	**104.54**	**104.19**	**105.01**
$\mathrm{T}_{2(\mathrm{pre})}$	**104.62**	103.16	**104.76**	**103.92**	**104.39**	**104.65**
$\mathrm{T}_{2(\mathrm{ppe})}$	75.72	91.81	77.42	92.23	79.11	92.63
$\mathrm{T}_{3(\mathrm{r})}$	99.41	99.87	99.57	99.91	99.70	99.94
$\mathrm{T}_{3(\mathrm{rre})}$	76.02	92.60	72.45	91.43	68.96	90.25
$\mathrm{T}_{3(\mathrm{rpe})}$	92.92	97.98	92.33	98.09	91.46	98.14
Simulated RB (%)						
$\bar{y}_{\mathrm{lr(1)c}}$	0.000	0.005	0.000	0.004	0.000	0.004
$\mathrm{T}_{1(\mathrm{reg})}$	-0.006	0.003	-0.005	0.003	-0.004	0.003
$\mathrm{T}_{1(\mathrm{regre})}$	-0.216	-0.032	-0.224	-0.034	-0.231	-0.035
$\mathrm{T}_{1(\mathrm{regpe})}$	0.197	0.041	0.207	0.042	0.216	0.043
$\mathrm{T}_{2(\mathrm{r})}$	-0.052	0.005	-0.038	-0.003	-0.025	-0.001
$\mathrm{T}_{2(\mathrm{p})}$	-0.057	-0.004	-0.041	-0.002	-0.027	0.000
$\mathrm{T}_{2(\mathrm{re})}$	-0.053	-0.005	-0.038	-0.003	-0.025	-0.001
$\mathrm{T}_{2(\mathrm{pe})}$	-0.055	-0.005	-0.040	-0.003	-0.026	-0.001
$\mathrm{T}_{2(\mathrm{rre})}$	-0.060	-0.003	-0.046	-0.001	-0.033	0.001
$\mathrm{T}_{2(\mathrm{rpe})}$	-0.043	-0.007	-0.029	-0.005	-0.016	-0.003
$\mathrm{T}_{2(\mathrm{pre})}$	-0.045	-0.007	-0.030	-0.004	-0.016	-0.002
$\mathrm{T}_{2(\mathrm{ppe})}$	-0.072	-0.003	-0.056	-0.001	-0.040	0.001
$\mathrm{T}_{3(\mathrm{r})}$	0.001	-0.005	-0.002	0.004	-0.003	0.003
$\mathrm{T}_{3(\mathrm{rre})}$	-0.236	-0.036	-0.239	-0.036	-0.241	-0.037
$\mathrm{T}_{3(\mathrm{rpe})}$	0.230	0.047	0.228	0.046	0.227	0.046

error and the simulated bias are calculated

$$\widehat{\mathrm{Bias}(\bar{y}_{\mathrm{lr(1)c}})} = \frac{\sum_{i=1}^{B}\left(\bar{y}_{\mathrm{lr(1)c}}^{(i)} - \bar{Y}\right)}{B},$$

$$\widehat{\mathrm{Bias}(T_{*(.)})} = \frac{\sum_{i=1}^{B}\left(T_{*(.)}^{(i)} - \bar{Y}\right)}{B},$$

$$\widehat{\mathrm{MSE}(\bar{y}_{\mathrm{lr(1)c}})} = \frac{\sum_{i=1}^{B}\left(\bar{y}_{\mathrm{lr(1)c}}^{(i)} - \bar{Y}\right)^2}{B}$$

and

$$\widehat{\mathrm{MSE}(T_{*(.)})} = \frac{\sum_{i=1}^{B}\left(T_{*(.)}^{(i)} - \bar{Y}\right)^2}{B}.$$

[16], [6], [14] and [1] have introduced the empirical relative bias (RB) and the empirical relative root mean square error (RRMSE) to measure the efficiency of their suggested estimators. Following this, the empirical relative mean square error (RMSE) and the empirical relative bias (RB) of each considered estimator is calculated. The performance of each estimator is computed with respect to the regression estimator by means of the simulated PRE (percent relative efficiency)

$$\widehat{\mathrm{RB}(\bar{y}_{\mathrm{lr}(1)\mathrm{c}})} = \frac{\widehat{\mathrm{Bias}(\bar{y}_{\mathrm{lr}(1)\mathrm{c}})}}{\bar{Y}}\times 100, \qquad \widehat{\mathrm{RB}(T_{*(.)})} = \frac{\widehat{\mathrm{Bias}(T_{*(.)})}}{\bar{Y}}\times 100,$$

$$\widehat{\mathrm{RMSE}(\bar{y}_{\mathrm{lr}(1)\mathrm{c}})} = \frac{\widehat{\mathrm{MSE}(\bar{y}_{\mathrm{lr}(1)\mathrm{c}})}}{\bar{Y}^2}, \qquad \widehat{\mathrm{RMSE}(T_{*(.)})} = \frac{\widehat{\mathrm{MSE}(T_{*(.)})}}{\bar{Y}^2}$$

and

$$\widehat{\mathrm{PRE}(T_{*(.)})} = \frac{\widehat{\mathrm{RMSE}(\bar{y}_{\mathrm{lr}(1)\mathrm{c}})}}{\widehat{\mathrm{RMSE}(\mathrm{T}_{*(.)})}}\times 100.$$

The results are shown in Table 7. To highlight the performance of the considered estimators in the Table 7, we use the "bold" sign to indicate the more efficient estimators than regression estimator. We can see that the estimators ($\mathrm{T}_{2(\mathrm{r})}$, $\mathrm{T}_{2(\mathrm{pe})}$, $\mathrm{T}_{2(\mathrm{rpe})}$, $\mathrm{T}_{2(\mathrm{pre})}$) are more efficient than the regression estimator.

Finally, for each sample, we determine simulated confidence interval for the mean of the suggested estimators at 95% nominal confidence level by assuming normality

$$\frac{\sum_{i=1}^{B}\bar{y}_{\mathrm{lr}(1)\mathrm{c}}^{(i)}}{B} \pm z_{\alpha/2}\sqrt{\widehat{\mathrm{RMSE}(\bar{y}_{\mathrm{lr}(1)\mathrm{c}})}}$$

and

$$\frac{\sum_{i=1}^{B}T_{*(.)}^{(i)}}{B} \pm z_{\alpha/2}\sqrt{\widehat{\mathrm{RMSE}(T_{*(.)})}}$$

The results are shown in Table 8, where LL denotes lower limit and UL denotes upper limit. In Table 8, the results are displayed only for $n = 100$ due to less space. In Table 9, coverage rates of 95 percent rescaled bootstrap confidence interval for the mean are shown. It is observed that all the coverages are close to the nominal 95 percent.

Moreover, to establish the stability of results, we also consider the theoretical results for the minimum mean square errors and the biases, assuming that the parameters of the simulated population are known. When population parameters are assumed to be known, the good performance of our proposed classes of estimators is achieved. In Table 10, it can be seen that all the estimators are more efficient than the linear regression estimator.

7. Conclusions

To our knowledge, this is the first time that CSS design has been considered for sample selection when the problem of non-response is present. Considering estimators based on CSS, three general classes of estimators have been suggested for estimation of the population mean $\bar{Y}$ with the auxiliary information using single and two-phase sampling. The existence of non-response problem in the study variable has been well deliberated in CSS. To tackle this problem, we have considered the well-known [8] non-respondents sub-sampling technique and also have determined the minimum mean square

Table 8. Confidence Interval results when population parameters are estimated

Estimators	$\rho = 0.70,\ n = 100$		$\rho = 0.43,\ n = 100$	
	LL	Ul	LL	UL
$\bar{y}_{\text{lr}(1)\text{c}}$	10.45	10.77	12.20	12.52
$\text{T}_{1(\text{reg})}$	10.45	10.77	12.20	12.52
$\text{T}_{1(\text{regre})}$	10.40	10.78	12.14	12.53
$\text{T}_{1(\text{regpe})}$	10.47	10.80	12.22	12.56
$\text{T}_{2(\text{r})}$	10.45	10.77	12.20	12.52
$\text{T}_{2(\text{p})}$	10.44	10.77	12.20	12.52
$\text{T}_{2(\text{re})}$	10.45	10.77	12.20	12.52
$\text{T}_{2(\text{pe})}$	10.44	10.77	12.20	12.52
$\text{T}_{2(\text{rre})}$	10.43	10.78	12.18	12.53
$\text{T}_{2(\text{rpe})}$	10.45	10.77	12.20	12.52
$\text{T}_{2(\text{pre})}$	10.45	10.77	12.20	12.52
$\text{T}_{2(\text{ppe})}$	10.42	10.79	12.17	12.54
$\text{T}_{3(\text{r})}$	10.45	10.78	12.20	12.52
$\text{T}_{3(\text{rre})}$	10.40	10.77	12.14	12.53
$\text{T}_{3(\text{rpe})}$	10.47	10.80	12.22	12.56

Table 9. Coverage rates (%) of the 95 percent rescaled bootstrap confidence interval for the mean

Estimators	$\rho = 0.70$		$\rho = 0.59$		$\rho = 0.43$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
$\bar{y}_{\text{lr}(1)\text{c}}$	0.905	0.981	0.905	0.981	0.905	0.981
$\text{T}_{1(\text{reg})}$	0.921	0.981	0.921	0.981	0.921	0.981
$\text{T}_{1(\text{regre})}$	0.917	0.985	0.928	0.985	0.917	0.984
$\text{T}_{1(\text{regpe})}$	0.864	0.963	0.864	0.96	0.864	0.959
$\text{T}_{2(\text{r})}$	0.935	0.986	0.935	0.987	0.921	0.982
$\text{T}_{2(\text{p})}$	0.935	0.980	0.935	0.980	0.935	0.981
$\text{T}_{2(\text{re})}$	0.935	0.986	0.935	0.981	0.935	0.982
$\text{T}_{2(\text{pe})}$	0.935	0.981	0.935	0.981	0.935	0.982
$\text{T}_{2(\text{rre})}$	0.908	0.974	0.908	0.973	0.908	0.970
$\text{T}_{2(\text{rpe})}$	0.921	0.991	0.906	0.991	0.898	0.991
$\text{T}_{2(\text{pre})}$	0.921	0.990	0.921	0.991	0.921	0.991
$\text{T}_{2(\text{ppe})}$	0.901	0.965	0.908	0.964	0.908	0.964
$\text{T}_{3(\text{r})}$	0.905	0.981	0.905	0.981	0.921	0.981
$\text{T}_{3(\text{rre})}$	0.918	0.987	0.928	0.987	0.912	0.985
$\text{T}_{3(\text{rpe})}$	0.850	0.959	0.850	0.959	0.850	0.957

errors of the suggested classes of estimators in both cases. The linear regression estimators based on CSS has been considered as a benchmark for making comparison with the suggested classes. To see the performance of the classes of estimators, we have provided numerical results of the considered estimators not only based on CSS but also based on SRSWOR. It has been observed that all considered estimators are more efficient in CSS

Table 10. Theoretical results when population parameters are known for $n = 100$

Estimators	$\rho = 0.70$	$\rho = 0.59$	$\rho = 0.43$
PRE			
$\bar{y}_{\mathrm{lr(1)c}}$	100	100	100
$\mathrm{T}_{1(\mathrm{reg})}$	100.01	100.00	100.00
$\mathrm{T}_{1(\mathrm{regre})}$	100.72	100.84	100.97
$\mathrm{T}_{1(\mathrm{regpe})}$	105.26	106.42	107.75
$\mathrm{T}_{2(\mathrm{r})}$	100.46	100.28	100.14
$\mathrm{T}_{2(\mathrm{p})}$	100	100.00	100.00
$\mathrm{T}_{2(\mathrm{re})}$	100.26	100.17	100.09
$\mathrm{T}_{2(\mathrm{pe})}$	100.03	100.02	100.02
$\mathrm{T}_{2(\mathrm{rre})}$	100.49	101.07	101.94
$\mathrm{T}_{2(\mathrm{ppe})}$	103.00	103.25	103.58
$\mathrm{T}_{2(\mathrm{rpe})}$	104.36	104.55	104.70
$\mathrm{T}_{2(\mathrm{pre})}$	103.00	103.49	103.97
$\mathrm{T}_{3(\mathrm{r})}$	100	100.00	100.00
$\mathrm{T}_{3(\mathrm{rre})}$	101.24	101.28	101.28
$\mathrm{T}_{3(\mathrm{rpe})}$	107.54	108.14	108.89
Bias			
$\bar{y}_{\mathrm{lr(1)c}}$	0	0	0
$\mathrm{T}_{1(\mathrm{reg})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{1(\mathrm{regre})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{1(\mathrm{regpe})}$	-0.001	0.001	0
$\mathrm{T}_{2(\mathrm{r})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{2(\mathrm{p})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{2(\mathrm{re})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{2(\mathrm{pe})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{2(\mathrm{rre})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{2(\mathrm{ppe})}$	-0.001	-0.001	0
$\mathrm{T}_{2(\mathrm{rpe})}$	-0.001	-0.001	0
$\mathrm{T}_{2(\mathrm{pre})}$	-0.001	-0.001	0
$\mathrm{T}_{3(\mathrm{r})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{3(\mathrm{rre})}$	-0.001	-0.001	-0.001
$\mathrm{T}_{3(\mathrm{rpe})}$	-0.001	-0.001	0

than SRSWOR for single-phase. In case of two-phases, the efficiency of estimators under CSS and SRSWOR is dependent on m. Further, it is also important to note that our efficiency results have dependent on the considered population and particularly from ρ_y and ρ_x values. But from a general point of view, it is expected that CSS can be preferred for the greater simplicity. It is noted that the first and third classes can be preferable in case of complete response on Y, while second class may be a better choice in non-response case.

Furthermore, we have analyzed the numerical comparison on the real population by a Monte Carlo study with the intention to comprehend the validness of certain results when extra estimates are needed in CSS design. In addition, the simulated confidence intervals for the suggested estimators are also analyzed. When the information of the population parameters is unavailable, some of the proposed estimators are shown more efficient than linear regression estimators. However, the results are ambiguous for some estimators investigated in the simulation study. Apart from this, all the suggested estimators are more efficient than regression estimator while the information of the population parameters is accessible.

Acknowledgement

The authors wish to thank to anonymous referees for their careful reading and constructive suggestions which led to improvement over an earlier version of the paper. Thanks to the Department of Statistical Sciences, University of Padova, Italy and the Higher Education Commission (HEC), Islamabad, Pakistan for their logistic and financial support for this research.

References

[1] Berger, Y. G. and Muñoz, J. F. *On estimating quantiles using auxiliary information*, Journal of Official Statistics **31(1)**, 101–119, 2015.

[2] Cochran, W. G. *Sampling Techniques*, (New York, Wiley, 1977).

[3] Diana, G. and Giordan, M. and Perri, P. F. *An improved class of estimators for the population mean*, Statistical Methods and Applications **20**, 123–140, 2011.

[4] Diana, G. and Riaz, S. and Shabbir, J. *Hansen and Hurwitz estimator with scrambled response on the second call*, Journal of Applied Statistics **41(3)**, 596–611, 2014.

[5] Gautschi, W. *Some remarks on systematic sampling*, The Annals of Mathematical Statistics **28**, 385–394, 1957.

[6] Harms, T. and Duchesne, P. *On Calibration estimation for quantiles*, Survey Methdology **32**, 37–52, 2006.

[7] Hajeck, J. *Optimum strategy and other problems in probability sampling*, Cosopis pro Pestovani Mathematiky **84**, 387–423, 1959.

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

[9] Koyuncu, N. and Kadilar, C. *Efficient estimators for the population mean*, Hacettepe Journal of Mathematics and Statistics **38**, 217–233, 2009.

[10] Lahiri, D. B. *A method for selection providing unbiased estimates*, Bulletin of the International Statistical Institute **33(2)**, 133–140, 1951.

[11] Leu, C.H. and Kao, F. F. *Modified balanced circular systematic sampling*, Statistics and Probability Letters **76**, 373–383, 2006.

[12] Leu, C.H. and Tsui, K. W. *New partially systematic sampling*, Statistica Sinica **6**, 617–630, 1996.

[13] Madow, W. G. *On the theory of systematic sampling*, The Annals of Mathematical Statistics **24**, 101–106, 1953.

[14] Muñoz, J. F., Álvarez–Verdejo, E., García–Fernández, R. M. and Barroso, L. J. *Efficient Estimation of the Headcount Index*, Social Indicators Research DOI 10.1007/s11205-014-0757-9, 2014.

[15] 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.

[16] Rao, J. N. K, Kovar, J. G. and Mantel, H. J. *On estimating distribution functions and quantiles from survey data using auxilairy information*, Bimetrika **77**, 365–375, 1990.

[17] Sengupta, S. and Chattophyadhyay, S. *A note on circular systematic sampling*, Sankhya B **49**, 186–187, 1987.

[18] Sethi, V. K. *On optimum pairing of units*, Sankhya B **27**, 315–320.

[19] Singh, H. P. and Jatwa, N. K. *A class of exponential-type estimators in systematic sampling*, Economic Quality Control **27**, 195–208, 2012.

[20] Singh, H. P. and Solanki, R. S. *An efficient class of estimators for the population mean using auxiliary information in systematic sampling*, Journal of Statistical Theory and Practice **6**, 274–285, 2012.

[21] Singh, H. P. and Solanki, R. S. *An efficient class of estimators for the population mean using auxiliary information*, Communication in Statistics-Theory and Methods **42**, 145–163, 2013.

[22] Singh, P., Jindal, K. K. and Garg, J. N. *On modified systematic sampling*, Biometrika **55**, 541–546, 1968.

[23] Singh, R., Malik, S., Chaudry, M. K., Verma, H. K. and Adewara, A. A. *A general family of ratio-type estimators in systematic sampling*, Journal of Reliability and Statistical Studies **5(1)**, 73–82, 2012.

[24] Singh, R. and Singh, H. P. *Almost unbiased ratio and product-type estimators in systematic sampling*, QUESTIIO **22(3)**, 403–416, 1998.

[25] Riaz, S., Diana, G. and Shabbir, J. *Improved classes of estimators for population mean in presence of non-response*, Pakistan Journal of Statistics **30(1)**, 83–100, 2014.

[26] Uthayakumaran, N. *Additional circular systematic sampling methods*, Biometrical Journal **40(4)**, 467–474, 1998.

[27] Verma, H., Singh, R. D. and Singh, R. *A general class of regression type estimators in systematic sampling under non-response*, Octogon Mathematical Magazine **20(2)**, 542–550, 2012.

[28] Verma, H., Singh, R. D. and Singh, R. *Some improved estimators in systematic sampling under non-response*, National Academy Science Letters **37(1)**, 91–95, 2014.

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

[30] Yates, F. *Systematic sampling*, Philosophical Transactions Royal Society, London, A **241**, 345–377, 1948.