Hacettepe Journal of Mathematics and Statistics
Volume 43 (3) (2014), 529 – 542

Generalized modified linear systematic sampling scheme for finite populations

J. Subramani*[†] and Sat N. Gupta[‡]

Abstract

The present paper deals with a further modification on the selection of linear systematic sample, which leads to the introduction of a more generalized form of modified linear systematic sampling namely generalized modified linear systematic sampling (GMLSS) scheme, which is applicable for any sample size, irrespective of the population size whether it is a multiple of sample size or not. The performances of the proposed modified linear systematic sampling scheme are assessed with that of simple random sampling, circular systematic sampling for certain hypothetical populations as well as for some natural populations. As a result, it is observed that the proposed modified linear systematic sample means perform better than the simple random sample mean and circular systematic sample mean for estimating the population mean in the presence of linear trend among the population values. Further improvements on GMLSS are achieved by introducing Yates type end corrections.

Keywords: Linear Trend, Modified Systematic Sampling, Natural Population, Simple Random Sampling, Trend Free Sampling, Yates End Corrections, Yates Type End Corrections.

2000 AMS Classification: 62D05

1. Introduction

It is well known that in the presence of linear trend among the population values the systematic sampling performs better than the simple random sampling without replacement for estimating the population mean. Later several modifications have been made to improve the efficiency of the systematic sampling by introducing changes on the method of selection which includes centered systematic sampling by Madow [12], balanced systematic sampling by Sethi [15], modified systematic

---

*Department of Statistics, Pondicherry University, R V Nagar, Kalapet, Puducherry 605014, India, Email: `drjsubramani@yahoo.co.in`

[†] Corresponding Author.

[‡] Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC 27402, USA, Email: `sngupta@uncg.edu`

sampling by Singh, Jindal, and Garg [16] and changes on the estimators itself like Yates end corrections [26]. In all these sampling schemes mentioned above and also in the case of linear systematic sampling (LSS), it is assumed that the population size N is a multiple of sample size n and there is no restriction on the part of sample size. For a detailed discussion on estimation of finite population means, one may refer to Bellhouse and Rao [2], Chang and Huang [3], Cochran [4], Fountain and Pathak [5], Gupta and Kabe [6], Kadilar and Cingi [8], Khan et al. [9] [11], Murthy [14], Singh S. [17], Subramani [18] [19] [20] [21], Subramani and Singh [24] and the references cited therein. The circular systematic sampling (CSS) is an alternative to LSS whenever the population size is not a multiple of sample size with certain restrictions to get distinct units in the sample. For a discussion on the choices for the sampling interval for the case of CSS one may refer to Bellhouse [1], Sudakar [25] and Khan et al. [10]. Chang and Huang [3] have suggested a modification on linear systematic sampling and introduced the Remainder Linear systematic sampling (RLSS). The RLSS can be used for population size is not a multiple of sample size, where $N = n\widehat{k} + r$, $\widehat{k}$ is the sampling interval for RLSS and depends upon the remainder r. When the remainder is zero, the RLSS reduces to the usual linear systematic sampling.

Recently Subramani [22] [23] has introduced a modification on the selection of a systematic sample in the linear systematic sampling by choosing two random starts and is called as modified linear systematic sampling (MLSS) scheme. However the problem is that the MLSS is also applicable only for the cases where the population size is a multiple of sample size and is not valid when the population size is not a multiple of sample size. In this paper, a more general form of the MLSS is introduced which is applicable for any sample size whether it is even or odd and for any population size N, where $N = nk'$ or $N \neq nk'$, where k' is the sampling interval. For the sake of convenience, it is assumed that, without loss of generality, that $N = n_1k_1 + n_2k_2$ such that $n = n_1 + n_2$ and the value of $n_1 \geq n_2$. Further, it is shown that the MLSS method discussed by Subramani [22] [23] and the usual LSS are the particular cases of the proposed methods of generalized modified systematic sampling. The explicit expressions for the GMLSS sample means, the bias and the mean squared error are obtained for certain hypothetical populations with a perfect linear trend among the population values and are compared with that of simple random sampling. Further the relative performances of GMLSS are assessed with that of the simple random sampling, circular systematic sampling and remainder linear systematic sampling for certain natural populations. From the numerical comparison, it has been shown that the GMLSS perform better than the simple random sampling and circular systematic sampling for estimating the finite population means in the presence of linear trend. The entire above are explained with the help of hypothetical as well as natural populations.

2. Proposed Modified Systematic Sampling Scheme for any sample size

The proposed modified systematic sampling scheme is explained here, firstly with the help of examples and later the generalized case. For the sake of simplicity and for the benefit of the readers, selecting a modified linear systematic sample of size 5 from a population of size $N = 12$ and $N = 16$ is explained with the help of following examples.

2.1. Example. Let $N = 12$, $n = 5$, $n_1 = 3$, $k_1 = 2$, $n_2 = 2$, $k_2 = 3$
Step 1: Arrange the 12 population units as given below:

Arrangement of the Population units

i		j		
1	2	1	2	3
1	2	3	4	5
6	7	8	9	10
11	12			

Step 2: Select two random numbers $1 \leq i \leq 2$ and $1 \leq j \leq 3$; include all the elements in the columns corresponding to i and j. The selected samples are given below:

GMLSS Samples

Sample No.	i	j	Sampled Units
1	1	1	1,3,6,8,11
2	1	2	1,4,6,9,11
3	1	3	1,5,6,10,11
4	2	1	2,3,7,8,12
5	2	2	2,4,7,9,12
6	2	3	2,5,7,10,12

2.2. Example. Let $N = 12$, $n = 5$, $n_1 = 3$, $k_1 = 4$, $n_2 = 2$, $k_2 = 2$
Step 1: Arrange the 16 population units as given below:

Arrangement of the Population units

i				j	
1	2	3	4	1	2
1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16		

Step 2: Select two random numbers $1 \leq i \leq 4$ and $1 \leq j \leq 2$; include all the elements in the columns corresponding to i and j. The selected samples are given below:

GMLSS Samples

Sample No.	i	j	Sampled Units
1	1	1	1,5,7,11,13
2	1	2	1,6,7,12,13
3	2	1	2,5,8,11,14
4	2	2	2,6,8,12,14
5	3	1	3,5,9,11,15
6	3	2	3,6,9,12,15
7	4	1	4,5,10,11,16
8	4	2	4,6,10,12,16

2.1. Generalized Modified Linear Systematic Sampling Scheme (GMLSS).

The steps involved in selecting a generalized modified linear systematic sample (GMLSS) of size n from a population of size $N = n_1k_1 + n_2k_2$, where $n = n_1 + n_2$, k_1 and k_2 such that $k = k_1 + k_2$ are positive integers, are as follows:

1. Firstly arrange the N population units (labels) in a matrix with $k = k_1 + k_2$ columns as given in the Fig. 1. That is, the first n_2k population units are arranged row wise in the first n_2 rows with k elements each and the remaining $(n_1 - n_2)k_1$ population units are arranged row wise in the next $(n_1 - n_2)$ rows with k_1 elements each as in the arrangement given below in the Fig. 1.

2. The first k_1 columns are assumed as Set 1 and the next k_2 columns are assumed as Set 2.

3. Select two random numbers, i in between 1 and k_1 and j in between 1 and k_2, then select all the n_1 units in the i^{th} column of Set 1 and all the n_2 units in the j^{th} column of Set 2, which together give the sample of size n.

4. The step 3 leads to $k_1 \times k_2$ samples of size n.

Figure 1. Arrangement of the population units

$$
\begin{array}{ccccccccccc}
1 & \cdots & i & \cdots & k_1 & k_1+1 & \cdots & k_1+j & \cdots & k \\
k+1 & \cdots & k+i & \cdots & k+k_1 & k+k_1+1 & \cdots & k+k_1+j & \cdots & 2k \\
2k+1 & \cdots & 2k+i & \cdots & 2k+k_1 & 2k+k_1+1 & \cdots & 2k+k_1+j & \cdots & 3k \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
(n_2-1)k+1 & \cdots & (n_2-1)k+i & \cdots & (n_2-1)k+k_1 & (n_2-1)k+k_1+1 & \cdots & (n_2-1)k+k_1+j & \cdots & n_2k \\
n_2k+1 & \cdots & n_2k+i & \cdots & n_2k+k_1 & & & & & \\
n_2k+k_1+1 & \cdots & n_2k+k_1+i & \cdots & n_2k+2k_1 & & & & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & & & & & \\
n_2k_2+(n_1-1)k_1+1 & \cdots & n_2k_2+(n_1-1)k_1+i & \cdots & n_2k_2+n_1k_1 & & & & &
\end{array}
$$

Since the generalized modified linear systematic sampling scheme has $k_1 \times k_2$ samples of size n and each unit in the Set-1 is included in k_1 samples and each unit in the Set-2 is included in k_2 samples, the first order and second order inclusion probabilities are obtained as given below:

$$\pi_i = \begin{cases} \frac{1}{k_1} & \text{if } i^{th} \text{ unit is from the Set-1,} \\ \frac{1}{k_2} & \text{if } i^{th} \text{ unit is from the Set-2.} \end{cases} \tag{2.1}$$

$$\pi_{ij} = \begin{cases} \frac{1}{k_1} & \text{if } i^{th} \text{ and } j^{th} \text{ units are from the same column of the Set-1,} \\ \frac{1}{k_2} & \text{if } i^{th} \text{ and } j^{th} \text{ units are from the same column of the Set-2,} \\ \frac{1}{k_1 \times k_2} & \text{if } i^{th} \text{ and } j^{th} \text{ units are from Set-1 and Set-2 respectively,} \\ 0 & \text{Otherwise.} \end{cases} \tag{2.2}$$

In general, for the given population size $N = n_1k_1 + n_2k_2$, the selected generalized modified linear systematic samples (labels of the population units) for the random starts i and j are given below:

$$S_{ij} = \begin{cases} i, i+k, i+2k, \cdots, i+(n_2-1)k, i+n_2k, i+n_2k+k_1, \cdots, i+n_2k+(n_1-n_2-1)k_1, \\ j+k_1, j+k_1+k, \cdots, j+k_1+(n_2-1)k \qquad (i = 1,2,3,\cdots,k_1 \text{ and } j = 1,2,3,\cdots,k_2) \end{cases} \tag{2.3}$$

The generalized modified linear systematic sample mean based on the random starts i and j is obtained as

$$\bar{y}_{gmlss} = \bar{y}_{ij} = \frac{1}{n}\left(\sum_{l=0}^{n_2-1} y_{i+kl} + \sum_{l=0}^{n_1-n_2-1} y_{i+n_2k+k_1l} + \sum_{l=0}^{n_2-1} y_{j+k_1+kl}\right) \tag{2.4}$$
$$(i = 1,2,3,\cdots,k_1 \text{ and } j = 1,2,3,\cdots,k_2)$$

Since the first order inclusion probabilities are not equal, the generalized modified linear systematic sample mean given above in Equation (2.4) is not an unbiased estimator. The mean squared error of the GMLSS mean can be obtained from the Equation (2.5) given below:

$$MSE\left(\bar{y}_{gmlss}\right) = \frac{1}{k_1k_2}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\left(\bar{y}_{ij} - \bar{Y}\right)^2 \tag{2.5}$$

3. Population with a Linear Trend

As stated earlier the linear systematic sampling (LSS) has less variance than the simple random sampling if the population consists solely of a linear trend among the population values. In this section, the relative efficiencies of the generalized modified systematic sampling schemes with that of simple random sampling for estimating the mean of finite populations with linear trend among the population values are assessed for certain hypothetical populations.

In this hypothetical population, the values of $N = n_1k_1 + n_2k_2$ population units are in arithmetic progression. That is,

$$Y_i = a + ib \qquad (i = 1,2,3,\cdots,N) \tag{3.1}$$

After a little algebra, one may obtain the generalized modified systematic sample means (GMLSS) with the random starts i and j and the population mean for the above hypothetical population are as given below:

$$\bar{y}_{gmlss} = \bar{y}_{ij} = a + \frac{1}{n}\left(n_1 i + n_2 j + n_2 k_1 + (k_1 + k_2) n_2 (n_1 - 1) + \frac{k_1(n_1 - n_2)(n_1 - n_2 - 1)}{2}\right) b \tag{3.2}$$

$$(i = 1, 2, 3, \cdots, k_1 \text{ and } j = 1, 2, 3, \cdots, k_2)$$

$$\bar{Y} = a + \left(\frac{k_1 n_1 + k_2 n_2 + 1}{2}\right) b \tag{3.3}$$

For the above population with a linear trend, variances of the simple random sample mean $V(\bar{y}_r)$ together with the Bias and Mean Squared Error of the generalized modified linear systematic sample means $Bias(\bar{y}_{gmlss})$ and $MSE(\bar{y}_{gmlss})$ are obtained as given below:

$$V(\bar{y}_r) = \frac{(N-n)(N+1)b^2}{12n} \tag{3.4}$$

$$Bias(\bar{y}_{gmlss}) = \frac{(k_1 - k_2) n_2 \left\{n_1 - (n_2 + 1)\right\}}{2n} b \tag{3.5}$$

$$MSE(\bar{y}_{gmlss}) = \frac{1}{12n^2}\left\{n_1^2(k_1^2 - 1) - n_2^2(k_2^2 - 1) + 3n_2^2\left\{n_1 - (n_2 + 1)\right\}^2 (k_1 - k_2)^2\right\} b^2 \tag{3.6}$$

Since the algebraic comparisons of the various expressions of the variance and the mean squared error given in Equations (3.4) and (3.6) are not possible due to the presence of several different parameters, we have compared them numerically and are presented in the Table 3.

3.1. Remark. If we put $k_1 = k_2$ in Equation (3.5), then $Bias(\bar{y}_{gmlss}) = 0$. That is, the Generalized modified linear systematic sample means become the modified linear systematic sample means of Subramani (2013a, b) and the resulting estimators are unbiased.

3.2. Remark. Even when $k_1 \neq k_2$ if we put $n_1 = n_2 + 1$ in Equation (3.5), the $Bias(\bar{y}_{gmlss}) = 0$. That is, the GMLSS estimators are unbiased estimators of the population mean.

3.3. Remark. When $n_2 = 0$ in Equation (3.5), the $Bias(\bar{y}_{gmlss}) = 0$. When $n_2 = 0$ which means $n_1 = n$, Equation (3.6) reduces to the variance of linear systematic sample mean. That is, the GMLSS estimators are unbiased estimators of the population mean and the resulting generalized modified linear systematic sample means become the linear systematic sample means.

3.4. Remark. The mean squared error of the Remainder Linear Systematic sample mean is given below:

$$MSE(\bar{y}_{rlss}) = \frac{1}{\widehat{k}(\widehat{k}+1)} \sum_{i=1}^{\widehat{k}} \sum_{j=1}^{\widehat{k}+1} (\bar{y}_{ij} - \bar{Y})^2 \tag{3.7}$$

where $\widehat{k}$ is the sampling inverval for RLSS.

3.5. Remark. Under the Remainder Linear systematic sampling procedure, the variance of Horvitz-Thompson estimator obtained by Chang and Huang [3] is given below:

$$V_{rlss}\left(\bar{y}_{HT}\right) = \frac{1}{N^2}\left\{(n-r)^2\widehat{k}^2\left\{\frac{1}{\widehat{k}}\sum_{i=1}^{\widehat{k}}\left(\bar{y}_{1i}-\bar{Y}_1\right)^2\right\} + r^2(\widehat{k}+1)^2\left\{\frac{1}{\widehat{k}+1}\sum_{i=1}^{\widehat{k}+1}\left(\bar{y}_{2i}-\bar{Y}_2\right)^2\right\}\right\} \tag{3.8}$$

where $\bar{y}_{1i}$ is the i^{th} sample mean of the Set-1
$\bar{y}_{2i}$ is the i^{th} sample mean of the Set-2
$\bar{Y}_1$ is the population mean of the first stratum
$\bar{Y}_2$ is the population mean of the second stratum

3.6. Remark. Under the Generalized modified linear systematic sampling, the variance of Horvitz-Thompson estimator is given below:

$$V_{gmlss}\left(\bar{y}_{HT}\right) = \frac{1}{N^2}\left\{n_1^2k_1^2\left\{\frac{1}{k_1}\sum_{i=1}^{k_1}\left(\bar{y}_{1i}-\bar{Y}_1\right)^2\right\} + n_2^2k_2^2\left\{\frac{1}{k_2}\sum_{i=1}^{k_2}\left(\bar{y}_{2i}-\bar{Y}_2\right)^2\right\}\right\} \tag{3.9}$$

where $\bar{y}_{1i}$ is the i^{th} sample mean of the Set-1
$\bar{y}_{2i}$ is the i^{th} sample mean of the Set-2
$\bar{Y}_1$ is the population mean of the Set-1
$\bar{Y}_2$ is the population mean of the Set-2

3.7. Remark. When $n_1 = (n-r)$, $k_1 = \widehat{k}$, $n_2 = r$, and $k_2 = \widehat{k}+1$, the generalized modified linear systematic sampling reduces to Remainder Linear systematic sampling for the population with a perfect linear trend. This shows that RLSS is the particular case of GMLSS for the population with a perfect linear trend. However the arrangement of the population units are different and hence the Set 1(2) and Stratum 1(2) are not one and the same.

4. Some Modifications on Generalized Modified Linear Systematic Sample Means

It has been shown in Section 3 that the estimators based on the generalized modified linear systematic sampling schemes (GMLSS) are, in general, not unbiased estimators. However, further improvements can be achieved by modifying the generalized modified linear systematic sample means as done by Subramani [21] in the case of modified linear systematic sample mean by introducing Yates type end corrections [26]. Consequently the proposed sampling scheme becomes completely trend free sampling (Mukerjee and Sengupta [13]).

4.1. Yates Type End Corrections for GMLSS Means. The modification involves the usual generalized modified linear systematic sampling, but the modified sample mean is defined as

$$\bar{y}^*_{gmlss} = \bar{y}_{gmlss} + \beta(y_1 - y_n) = \bar{Y} \tag{4.1}$$

That is, the units selected first and last are given the weights $\frac{1}{n}+\beta$ and $\frac{1}{n}-\beta$ respectively, whereas the remaining units will get the equal weight of $\frac{1}{n}$, so as to make the proposed estimator is equal to the population mean. That is, the value of β is obtained as

$$\beta = \frac{\bar{Y} - \bar{y}_{gmlss}}{(y_1 - y_n)} \tag{4.2}$$

For the hypothetical population defined in Section 3, after a little algebra, we have obtained the value of β for the two random starts i and j as given below:

$$\beta = \frac{(n_1 - n_2)\left\{k_1(n_2+1) - k_2 n_2\right\} + 2\left\{k_2 n_2 - n_1(i+1) - n_2(j+1)\right\}}{2n\left\{(i-j) - (k_1 + k(n_2 - 1))\right\}} \tag{4.3}$$

In the similar manner one can propose Yates type end corrections, alternate to Yates end corrections by giving different weights to two different units other than the first and last units. For example one may give different weights to two successive units; the first units of the two subsamples and so on.

If we give different weights to the two successive units between 1 and n_2 in Set 1 or successive units in Set 2 then the revised estimator $\bar{y}^{**}_{gmlss}$ for the case of generalized modified linear systematic sample mean $\bar{y}^{**}_{gmlss} = \bar{y}_{gmlss} + \beta_1(y_l - y_{l+1}) = \bar{Y}$, which yields the value of β_1 as is given below:

$$\beta_1 = \frac{\bar{Y} - \bar{y}_{gmlss}}{(y_l - y_{l+1})} = \frac{(n_2 - n_1)\left\{k_1(n_2+1) - k_2 n_2\right\} - 2\left\{k_2 n_2 - n_1(i+1) - n_2(j+1)\right\}}{2nk} \tag{4.4}$$

If we give different weights to the two successive units between $n_2 + 1$ and n_1 in Set 1 or the first units of the two sets then the revised estimator $\bar{y}^{***}_{gmlss}$ for the case of generalized modified linear systematic sample mean $\bar{y}^{***}_{gmlss} = \bar{y}_{gmlss} + \beta_2(y_{l'} - y_{l'+1}) = \bar{Y}$, which yields the value of β_2 as is given below:

$$\beta_2 = \frac{\bar{Y} - \bar{y}_{gmlss}}{(y_{l'} - y_{l'+1})} = \frac{(n_2 - n_1)\left\{k_1(n_2+1) - k_2 n_2\right\} - 2\left\{k_2 n_2 - n_1(i+1) - n_2(j+1)\right\}}{2nk_1} \tag{4.5}$$

If we give different weights to the first and the last units in the Set 1 then the revised estimator $\bar{y}^{****}_{gmlss}$ for the case of generalized modified linear systematic sample mean $\bar{y}^{****}_{gmlss} = \bar{y}_{gmlss} + \beta_3(y_1 - y_{n_1}) = \bar{Y}$, which yields the value of β_3 as is given below.

$$\beta_3 = \frac{\bar{Y} - \bar{y}_{gmlss}}{(y_1 - y_{n_1})} = \frac{(n_2 - n_1)\left\{k_1(n_2+1) - k_2 n_2\right\} - 2\left\{k_2 n_2 - n_1(i+1) - n_2(j+1)\right\}}{2n\left\{k_2 n_2 + (n_1 - 1)k_1\right\}} \tag{4.6}$$

If we give different weights to the units $n_1 + 1$ and n in Set 2 (i.e. the first and the last units in Set 2), then the revised estimator $\bar{y}^{*****}_{gmlss}$ for the case of generalized modified linear systematic sample mean $\bar{y}^{*****}_{gmlss} = \bar{y}_{gmlss} + \beta_4(y_{l'} - y_{l'+1}) = \bar{Y}$, which yields the value of β_4 as is given below:

$$\beta_4 = \frac{\bar{Y} - \bar{y}_{gmlss}}{(y_1 - y_{n_2})} = \frac{2(n_1 i + n_2 j) - N + n + n_1 n_2(k_2 - k_1)}{2nk(n_2 - 1)} \tag{4.7}$$

4.1. Remark. In the presence of a perfect linear trend, the revised generalized modified linear systematic sampling estimators $\bar{y}^{*}_{gmlss} = \bar{y}^{**}_{gmlss} = \bar{y}^{***}_{gmlss} = \bar{y}^{****}_{gmlss} = \bar{y}^{*****}_{gmlss} = \bar{Y}$ and hence the variances are zero. That is, GMLSS becomes a trend free sampling (Mukerjee and Sengupta [13]).

5. Relative Performance of Modified Linear Systematic Sampling for Certain Natural Populations

The Generalized proposed modified linear systematic sampling schemes were introduced in Section 2, where as its means, bias and the mean squared error were derived in Section 3 for the hypothetical populations with a perfect linear trend among the population values. Further it has been shown (Remarks 3.1 to 3.3) that the usual linear systematic sampling (LSS) and the modified linear systematic sampling (MLSS) schemes are the particular cases of the proposed GMLSS scheme. Further it has been shown (Remark 3.7) that the RLSS is the particular case of the GMLSS for the population with a perfect linear trend. Since the expression of the mean squared error of GMLSS means is involved several parameters compared to simple random sampling without replacement, it is not feasible to make an algebraic comparison. Hence we have assessed the performances of GMLSS means with that of SRSWOR means, CSS means and RLSS means for a hypothetical population together with some natural populations considered by Murthy, p.228 [14]. The data were collected for estimating the output of 80 factories in a region. The data pertaining to the number of workers, fixed capital and the output are respectively denoted as population 2, population 3 and population 4, where as its labels (presuming a hypothetical population with a perfect linear trend, as considered as population 1. It is already established in Subramani [22] [23] that whenever the population size is a multiple of sample size the modified linear systematic sampling performs well compared to simple random sampling as well as linear systematic sampling. Hence we have obtained the variance of the simple random sample mean $V\left(\bar{y}_{r}\right)$, the variance of the circular systematic sample mean $V\left(\bar{y}_{css}\right)$, the mean squared error of the remainder linear systematic sample mean, $MSE\left(\bar{y}_{rlss}\right)$, the mean squared error of the generalized modified systematic sample mean, $MSE\left(\bar{y}_{gmlss}\right)$, the variance of Horvitz Thompson estimator under RLSS, $V_{rlss}\left(\bar{y}_{HT}\right)$ and the variance of Horvitz Thompson estimator under GMLSS, for the odd sample sizes from 7 to 25 so as the population size 80 is not a multiple of the sample size with various possible combinations of n_1 and n_2 such that $n_1 > n_2$ and are presented in Table 3. From the table values, it is seen that, out of the 40 cases considered the generalized modified linear systematic sample (GMLSS) means perform better than simple random sample mean and circular systematic sample mean in all the 40 cases. Further it is observed from Equation (3.6) that whenever the differences between k_1 and k_2 and n_1 and n_2 approaches zero simultaneously, then the efficiency of the GMLSS improves.

It is to be noted that for the fixed population size $N = 80$ there are 10 different choices for the sample of size $n = 11$. The mean squared error of the generalized modified linear systematic sample mean under the possible combinations of k_1, n_1, k_2 and n_2 for population 1 (population with a perfect linear trend) is given in the following Table 1:

Table 1. $MSE\left(\bar{y}_{gmlss}\right)$ under the possible combinations of k_1, n_1, k_2 and n_2 for population 1

N	n	k_1	n_1	k_2	n_2	$MSE\left(\bar{y}_{gmlss}\right)$
80	11	7	10	10	1	4.56
		4	9	22	2	98.56
		6	9	13	2	16.99
		8	9	4	2	8.32
		4	8	16	3	45.08
		7	8	8	3	2.80
		4	7	13	4	13.07
		8	7	6	4	3.04
		10	6	4	5	2.71
		5	6	10	5	2.30

From the possible combinations of k_1, n_1, k_2 and n_2, one can choose the optimum value of k_1, n_1, k_2 and n_2 such that the differences between k_1 and k_2, and n_1 and n_2 approaches zero simultaneously.

The different sample sizes with the population size $N = 80$ together with the sampling interval of Circular systematic sampling, Remainder linear systematic sampling and Generalized modified linear systematic sampling considered for the numerical comparisons of these sampling schemes are given below:

Table 2. Sample size and sampling interval of Circular systematic sampling (CSS), Remainder linear systematic sampling (RLSS) and Generalized modified linear systematic sampling (GMLSS)

N	n	CSS	RLSS				GMLSS			
		k^*	$\hat{k}$	$n-r$	r	$\hat{k}+1$	k_1	n_1	k_2	n_2
80	7	11	11	4	3	12	11	4	12	3
	9	9	8	1	8	9	8	5	10	4
	11	7	7	8	3	8	5	6	10	5
	13	6	6	11	2	7	8	7	4	6
	15	-	5	10	5	6	3	8	8	7
	17	5	4	5	12	5	4	11	6	6
	19	4	4	15	4	5	4	15	5	4
	21	-	3	4	17	4	4	17	3	4
	23	3	3	12	11	4	3	12	4	11
	25	3	3	20	5	4	3	20	4	5

where k^* is Sampling interval for Circular Systematic Sampling
and $\widehat{k}$ is Sampling interval for Remainder Linear Systematic Sampling

Table 3. Comparison of simple random sample, Circular systematic sample, Remainder Linear systematic sample, Horvitz-Thompson estimator based on RLSS, Generalized modified linear systematic sample means and Horvitz-Thompson estimator based on GMLSS for the four natural populations considered by Murthy (1967), page 228. The four populations are Population 1-Labels (presuming a hypothetical population with a perfect linear trend), Population 2-Number of workers, Population 3-Fixed capital and Population 4-The output

Population Number	n	$V(\bar{y}_r)$	$V(\bar{y}_{css})$	$MSE(\bar{y}_{rlss})$	$V_{rlss}(\bar{y}_{HT})$	$MSE(\bar{y}_{gmlss})$	$V_{gmlss}(\bar{y}_{HT})$
1	7	70.39	11.53	5.45	5.44	5.45	5.44
	9	53.25	6.59	5.33	5.45	3.25	3.38
	11	42.34	5.07	2.51	2.43	2.30	3.50
	13	34.79	3.41	2.18	2.11	1.79	2.69
	15	29.25	–	1.21	1.19	1.33	1.19
	17	25.01	3.84	1.10	1.20	0.89	0.97
	19	21.67	2.73	0.87	0.83	0.87	0.83
	21	18.96	–	0.84	0.92	0.84	0.92
	23	16.73	10.99	0.47	0.51	0.47	0.51
	25	14.85	2.85	0.48	0.45	0.48	0.45
2	7	9533.59	1908.46	1490.64	1638.59	1075.56	1024.33
	9	7211.86	1127.89	1089.76	1117.15	570.51	555.67
	11	5734.4	856.99	403.87	470.99	403.80	569.45
	13	4711.54	585.35	284.44	322.85	331.71	511.43
	15	3961.44	–	282.34	352.48	229.31	253.62
	17	3387.84	605.59	294.84	332.73	191.55	158.40
	19	2935.00	428.32	136.84	174.17	231.43	209.19
	21	2568.41	–	200.99	221.55	196.86	216.94
	23	2265.58	1458.65	150.12	197.91	95.33	95.01
	25	2011.20	420.57	75.82	105.19	138.26	121.73
3	7	93212.63	19244.93	12379.70	13481.46	10706.31	10267.41
	9	70512.45	11648.10	9916.21	10147.90	5716.93	5809.58
	11	56066.87	8139.36	2841.66	3074.49	6724.62	8492.78
	13	46066.09	7247.69	3141.28	3280.75	4227.09	6596.81
	15	38732.19	–	3759	4551.23	2479.96	3725.77
	17	33123.91	8457.46	4372.18	4901.98	1568.98	1529.61
	19	28696.32	4029.56	2373.67	2864.38	1934.37	1759.56
	21	25112.08	–	1336.18	1470.22	1534.75	1682.77
	23	22151.18	15277.04	820.73	1070.23	797.58	739.63
	25	19664.03	4967.74	326.32	411.64	2785.79	2459.51
4	7	439256.84	125737.19	71902.22	71792.05	80223.8	75178.72
	9	332284.09	67808.68	37887.33	37515.29	32907.60	30070.34
	11	264210.53	53569.57	20689.43	20220.72	25476.05	28633.15
	13	217082.67	37048.46	16479.84	16355.28	24579.86	39846.35
	15	182522.25	–	12708.17	13113.80	13021.36	22414.21
	17	156093.69	33489.79	10350.21	10625.27	8622.55	9824.10
	19	135229.03	23203.85	10372.55	10353.09	14168.90	13446.96
	21	118338.60	–	5886.70	6320.10	10217.66	11206.30
	23	104385.63	64036.67	3632.85	4219.70	6324.86	6336.19
	25	92665.14	19367.25	2507.05	2663.48	8598.95	7640.17

Table 4. Efficiency of the Generalized modified linear systematic sample means and Horvitz-Thompson estimator based on GMLSS for the 4 natural populations

Population Number	n	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9
1	7	12.92	2.12	1.00	1.00	12.94	2.12	1.00	1.00	1.00
	9	16.38	2.03	1.64	1.68	15.75	1.95	1.58	1.61	1.04
	11	18.41	2.20	1.09	1.06	12.10	1.45	0.72	0.69	1.52
	13	19.44	1.91	1.22	1.18	12.93	1.27	0.81	0.78	1.50
	15	21.99	–	0.91	0.89	24.58	–	1.02	1.00	0.89
	17	28.10	4.31	1.24	1.35	25.78	3.96	1.13	1.24	1.09
	19	24.91	3.14	1.00	0.95	26.11	3.29	1.05	1.00	0.95
	21	22.57	–	1.00	1.10	20.61	–	0.91	1.00	1.10
	23	35.60	23.38	1.00	1.09	32.80	21.55	0.92	1.00	1.09
	25	30.94	5.94	1.00	0.94	33.00	6.33	1.07	1.00	0.94
2	7	8.86	1.77	1.39	1.52	9.31	1.86	1.46	1.60	0.95
	9	12.64	1.98	1.91	1.96	12.98	2.03	1.96	2.01	0.97
	11	14.20	2.12	1.00	1.17	10.07	1.50	0.71	0.83	1.41
	13	14.20	1.76	0.86	0.97	9.21	1.14	0.56	0.63	1.54
	15	17.28	–	1.23	1.54	15.62	–	1.11	1.39	1.11
	17	17.69	3.16	1.54	1.74	21.39	3.82	1.86	2.10	0.83
	19	12.68	1.85	0.59	0.75	14.03	2.05	0.65	0.83	0.90
	21	13.05	–	1.02	1.13	11.84	–	0.93	1.02	1.10
	23	23.77	15.30	1.57	2.08	23.85	15.35	1.58	2.08	1.00
	25	14.55	3.04	0.55	0.76	16.52	3.45	0.62	0.86	0.88
3	7	8.71	1.80	1.16	1.26	9.08	1.87	1.21	1.31	0.96
	9	12.33	2.04	1.73	1.78	12.14	2.00	1.71	1.75	1.02
	11	8.34	1.21	0.42	0.46	6.60	0.96	0.33	0.36	1.26
	13	10.90	1.71	0.74	0.78	6.98	1.10	0.48	0.50	1.56
	15	15.62	–	1.52	1.84	10.40	–	1.01	1.22	1.50
	17	21.11	5.39	2.79	3.12	21.66	5.53	2.86	3.2	0.97
	19	14.83	2.08	1.23	1.48	16.31	2.29	1.35	1.63	0.91
	21	16.36	–	0.87	0.96	14.92	–	0.79	0.87	1.10
	23	27.77	19.15	1.03	1.34	29.95	20.65	1.11	1.45	0.93
	25	7.06	1.78	0.12	0.15	8.00	2.02	0.13	0.17	0.88
4	7	5.48	1.57	0.90	0.89	5.84	1.67	0.96	0.95	0.94
	9	10.10	2.06	1.15	1.14	11.05	2.26	1.26	1.25	0.91
	11	10.37	2.10	0.81	0.79	9.23	1.87	0.72	0.71	1.12
	13	8.83	1.51	0.67	0.67	5.45	0.93	0.41	0.41	1.62
	15	14.02	–	0.98	1.01	8.14	–	0.57	0.59	1.72
	17	18.10	3.88	1.20	1.23	15.89	3.41	1.05	1.08	1.14
	19	9.54	1.64	0.73	0.73	10.06	1.73	0.77	0.77	0.95
	21	11.58	–	0.58	0.62	10.56	–	0.53	0.56	1.10
	23	16.50	10.12	0.57	0.67	16.47	10.11	0.57	0.67	1.00
	25	10.78	2.25	0.29	0.31	12.13	2.53	0.33	0.35	0.89

where $R_1 = \frac{V(\bar{y}_r)}{MSE(\bar{y}_{gmlss})}$, $R_2 = \frac{V(\bar{y}_{css})}{MSE(\bar{y}_{gmlss})}$, $R_3 = \frac{MSE(\bar{y}_{rlss})}{MSE(\bar{y}_{gmlss})}$, $R_4 = \frac{V_{rlss}(\bar{y}_{HT})}{MSE(\bar{y}_{gmlss})}$,
$R_5 = \frac{V(\bar{y}_r)}{V_{gmlss}(\bar{y}_{HT})}$, $R_6 = \frac{V(\bar{y}_{css})}{V_{gmlss}(\bar{y}_{HT})}$, $R_7 = \frac{MSE(\bar{y}_{rlss})}{V_{gmlss}(\bar{y}_{HT})}$, $R_8 = \frac{V_{rlss}(\bar{y}_{HT})}{V_{gmlss}(\bar{y}_{HT})}$,
and $R_9 = \frac{V_{gmlss}(\bar{y}_{HT})}{MSE(\bar{y}_{gmlss})}$

6. Conclusion

In this paper a generalized version of modified linear systematic sampling (GMLSS) scheme is introduced irrespective of the sample size, whether it is odd or even and the population size is not a multiple of the sample size. The explicit expressions for the sample means and the mean squared error of the GMLSS estimators are derived for certain hypothetical populations with a perfect linear trend among the population values. The performances of the proposed sampling scheme are assessed with that of the simple random sampling without replacement, Circular systematic sampling and Remainder linear systematic sampling for a hypothetical population together with some natural populations considered by Murthy [14]. The comparative studies reveal that whenever there is a linear trend among the population values the generalized modified linear systematic sampling (GMLSS) performs well compared to simple random sampling. In the case of circular systematic sampling, the sample size and the population size must be a co-prime to get all distinct units in the sample. Hence the case with sample size 15 is not considered. Further it is observed that when $N = 80$ and $n = 21$ with $k = 4$ the first unit and the last unit are the same and hence this case is also not considered. From the above points, it is concluded that the proposed GMLSS is applicable for all possible combinations of population size and sample size even where the CSS fails. In comparing with the RLSS, the GMLSS performs better than the RLSS in majority of the cases.

Acknowledgements. The authors are thankful to the editor and the reviewers for their constructive comments, which have improved the presentation of the paper. The first author wishes to record his gratitude and thanks to University Grants Commission for the financial assistance through UGC–MRP.

References

[1] Bellhouse, D. R. *On the choice of sampling interval in circular systematic sampling*, Sankhya B, 247-248, 1984.

[2] Bellhouse, D. R. and Rao, J. N. K. *Systematic sampling in the presence of linear trends*, Biometrika **62**, 694-697, 1975.

[3] Chang, H. J and Huang, K. C. *Remainder linear systematic sampling*, Sankhya B **62**, 249-256, 2000.

[4] Cochran, W. G. *Sampling Techniques* (3rd ed.), New York: John Wiley and Sons, 1977.

[5] Fountain, R. L. and Pathak, P. L. *Systematic and non-random sampling in the presence of linear trends*, Communications in Statistics-Theory and Methods **18**, 2511-2526, 1989.

[6] Gupta, A. K. and Kabe, D. G. *Theory of Sample Surveys*, World Scientific Publishing Company, 2011.

[7] Horvitz, D. G. and Thompson, D. J. *A generalization of sampling without replacement from a finite universe*, Journal of the American Statistical Association **47**, 663-685, 1952.

[8] Kadilar, C. and Cingi, H. *Advances in Sampling Theory-Ratio Method of Estimation*, Bentham Science Publishers, 2009.

[9] Khan, Z., Gupta, S. N. and Shabbir, J. *A new sampling design for systematic sampling*, Communications in Statistics-Theory and Methods **42** (18), 2659-2670, 2013.

[10] Khan, Z., Gupta, S. N. and Shabbir, J. *A Note on Diagonal circular systematic sampling*, Forthcoming in Journal of Statistical Theory and Practice DOI: 10.1080/15598608.2013.801805, 2013.

[11] Khan, Z., Gupta, S. N. and Shabbir, J. *Generalized Systematic Sampling*, Forthcoming in Communications in Statistics-Simulation and Computation, 2013.

[12] Madow, W. G. *On the theory of systematic sampling III-comparison of centered and random start systematic sampling*, The Annals of Mathematical Statistics **24**, 101-106, 1953.
[13] Mukerjee, R. and Sengupta, S. *Optimal estimation of a finite population means in the presence of linear trend*, Biometrika **77**, 625-630, 1990.
[14] Murthy, M. N. *Sampling Theory and Methods*, Calcutta, India: Statistical Publishing House, 1967.
[15] Sethi, V. K. *On optimum pairing of units*, Sankhya B **27**, 315-320, 1965.
[16] Singh, D., Jindal, K. K. and Garg, J. N. *On modified systematic sampling*, Biometrika **55**, 541-546, 1968.
[17] Singh, S. *Advanced Sampling Theory with Applications*, Springer, 2003.
[18] Subramani, J. *Diagonal systematic sampling scheme for finite populations*, Journal of the Indian Society of Agricultural Statistics **52** (2), 187-195, 2000.
[19] Subramani, J. *Further results on diagonal systematic sampling for finite populations*, Journal of the Indian Society of Agricultural Statistics **63** (3), 277-282, 2009.
[20] Subramani, J. *Generalization of diagonal systematic sampling scheme for finite populations*, Model Assisted Statistics and Applications **5**, 117-128, 2010.
[21] Subramani, J. *A modification on linear systematic sampling for odd sample size*, Bonfring International Journal of Data Mining **2** (2), 32-36, 2012.
[22] Subramani, J. *A modification on linear systematic sampling*, Model Assisted Statistics and Applications **8** (3), 215-227, 2013.
[23] Subramani, J. *A further modification on linear systematic sampling for finite populations*, Journal of Statistical Theory and Practice **7** (3), 471-479, 2013.
[24] Subramani, J. and Singh, S. *Estimation of Population Mean in the Presence of Linear Trend*, Forthcoming in Communications in Statistics-Theory and Methods, 2013.
[25] Sudakar, K. *A note on circular systematic sampling*, Sankhya C **40**, 72-73, 1978.
[26] Yates, F. *Systematic Sampling*, Philosophical Transactions of the Royal Society A **241**, 345-377, 1948.