

Some Systematic Sampling Strategies using Multiple Random Starts

S. Sampath
Department of Statistics
University of Madras, Chennai 600005
India
sampath1959@yahoo.com

S. Ammani
Department of Mathematics
KRMM College, Chennai 600020
India
sammani_1967@rediffmail.com

Abstract

This paper deals with some linear trend free sampling – estimating strategies. Two different systematic sampling schemes which are multiple random starts analogues of Balanced Systematic Sampling (BSS) due to Sethi (1965) and Modified Systematic Sampling (MSS) of Singh, Jindal, and Garg (1968) have been considered. Further, Yates corrected estimator has been developed for Linear Systematic Sampling (LSS) with multiple random starts. A detailed numerical comparative study has been carried out using appropriate super population models with the help of R package for statistical computing.

Keywords: Multiple random starts, Systematic sampling, Yates corrected estimator, Average variance.

AMS classification: DMS605.

1. Introduction

Consider a finite population of N units. The values of the population units with respect to the characteristic under study will be denoted by Y_1, Y_2, \dots, Y_N .

Let n be the sample size and k a positive integer such that $k = \frac{N}{n}$. A Linear

systematic sample (LSS) is drawn by choosing at random an integer from 1 to k , say r . Starting from the r^{th} unit in the population, every k^{th} unit is selected until a sample of size n is obtained. This procedure is equivalent to dividing the population of N units into k mutually exclusive and exhaustive groups $\{S_1, S_2, \dots, S_k\}$ of n units each and choosing one of them at random where the units in the r^{th} group are given by

$$S_r = \{r, r+k, \dots, r+(n-1)k\}, \quad r = 1, 2, \dots, k$$

Even though the LSS described above is operationally more convenient, it has two limitations, namely, (i) when the population size is not divisible by sample size, the sample mean becomes biased and (ii) the second order inclusion

probabilities are zero for several pairs of units which makes variance estimation difficult. To overcome this problem, several alternatives are discussed in the literature by many including Leu and Tsui (1996), Suresh Chandra, Sampath and Balasubramani (1992), Sampath and Uthayakumaran (1998), Chang and Huang (2000) and Kuo (2004). Although these new sampling methods solve some limitations of LSS, they lose the operational simplicity of LSS. Preserving the inherent characteristics of LSS, like operational simplicity, Tukey (1950) and later on Gautschi (1957) suggested LSS with multiple random starts as an alternative to LSS with single random start which is explained below.

In LSS with multiple random starts, instead of choosing only one number at random from 1 to k , a simple random sample of size s is drawn from the first k elements and then every k^{th} element is chosen. If r_1, r_2, \dots, r_s are the labels of the s units sampled from the first k units, then the sample contains units with labels $r_1, r_1 + k, \dots, r_1 + (n-1)k, r_2, r_2 + k, \dots, r_2 + (n-1)k, \dots, r_s, r_s + k, \dots, r_s + (n-1)k$.

It may be noted that ultimately the sample contains a total of ns units.

As competitors of LSS many systematic sampling schemes have been proposed in literature each one having an edge over LSS in some sense. Among them, two linear trend free sampling methods (methods in which the estimated value coincides with the population value) are the Balanced Systematic Sampling due to Sethi (1965) and Modified Systematic Sampling due to Singh, Jindal and Garg (1968). Even though the properties of these two schemes have been widely studied by several researchers, their performances have not been studied so far when multiple random starts are employed. In the following section, we consider two new systematic sampling schemes with multiple random starts, which inherit the characteristics of BSS and MSS. Their behaviour for populations with linear trend is also studied.

2. New Sampling Schemes

(i) Balanced Systematic Sampling with multiple random starts

Under the BSS, the population units are divided into $\frac{n}{2}$ groups (assuming the sample size n is even) of $2k$ units each and a pair of units equidistant from the end points are selected from each group. This method is explained in detail as follows:

A random number r is selected from 1 to k and units with labels r and $2k - r + 1$ will be selected from first group and thereafter from the remaining $\frac{n}{2} - 1$ groups, the corresponding pairs of elements will be selected in a balanced manner. Thus,

the BSS of size n corresponding to the random start r is given by the units with labels

$$s_r = \left\{ r + 2jk, 2(j+1)k - r + 1, j = 0, 1, \dots, \frac{n}{2} - 1 \right\}$$

Now let us consider an analogue of BSS involving multiple random starts. When a BSS with t random starts is desired, the population is divided into $\frac{n}{2t}$ groups of $2tk$ units where the labels of the elements of the i^{th} group, $i = 1, 2, \dots, \frac{n}{2t}$, namely G_i are as given in the following description.

		Sampled units			
Groups	G_1	1	2	...	$2tk$
	G_2	$(2tk + 1)$	$(2tk + 2)$...	$4tk$
	G_i	$2(i-1)tk + 1$	$2(i-1)tk + 2$...	$2itk$
	$G_{\frac{n}{2t}}$	$2\left(\frac{n}{2t} - 1\right)tk + 1$	$2\left(\frac{n}{2t} - 1\right)tk + 2$...	$2\left(\frac{n}{2t}\right)tk$

We choose randomly t numbers from 1 to tk . Corresponding to every random number drawn we take pairs of units equidistant from the group ends. Note that there are $\frac{n}{2t}$ groups and hence a random number contributes $\frac{n}{t}$ units to the sample. Thus, the above method of selection leads to a sample of size n .

Under balanced systematic sampling with multiple random starts the sample

mean \hat{Y}_{BSSM} is defined by
$$\hat{Y}_{BSSM} = \frac{1}{n} \sum_{j=1}^{\frac{n}{2t}} \left[\sum_{i=1}^t (Y_{r_i+2(j-1)tk} + Y_{2tjk-r_i+1}) \right] \tag{1}$$

Theorem 1 The sample mean \hat{Y}_{BSSM} of the Balanced Systematic Sampling with multiple random starts is unbiased for the population mean.

Proof: The sample mean corresponding to t random starts r_1, r_2, \dots, r_t can be

written as
$$\hat{Y}_{BSSM}(r_1, r_2, \dots, r_t) = \frac{1}{n} \sum_{j=1}^{\frac{n}{2t}} \left[\sum_{i=1}^t (Y_{r_i+2(j-1)tk} + Y_{2tjk-r_i+1}) \right]$$

Taking expectations on both sides we get,

$$\left(\hat{Y}_{BSSM} \right) = \frac{1}{n} \sum_{i=1}^{nk} (T_i + U_i) \frac{t}{tk} \quad \text{where } T_i = \sum_{j=1}^{\frac{n}{2t}} Y_{r_i+2t(j-1)k} \text{ and } U_i = \sum_{j=1}^{\frac{n}{2t}} Y_{2tjk-r_i+1}$$

$$\begin{aligned}
 &= \frac{1}{nk} \sum_{i=1}^{tk} Y_i \\
 &= \bar{Y}
 \end{aligned}$$

Therefore, the sample mean is unbiased for the population mean under BSSM.

Remark It can be easily shown that,

$$V\left(\hat{Y}_{BSSM}\right) = \frac{tk-t}{tk(t)} S^2, \tag{2}$$

where, $S^2 = \frac{1}{tk-1} \sum_{i=1}^{tk} (Z_i - \bar{Z})^2$, $Z_i = (T_i + T_{2tk-i+1}), i = 1, 2, \dots, 2tk$ and

$$\bar{Z} = \frac{1}{tk} \sum_{i=1}^{tk} Z_i$$

The following theorem proves that the sample mean \hat{Y}_{BSSM} under balanced systematic sampling with multiple random starts coincides with the population mean when the population values Y_1, Y_2, \dots, Y_N satisfy the model

$$Y_u = \alpha + \beta u, u = 1, 2, \dots, N \tag{3}$$

Theorem 2 Under the model $Y_u = \alpha + \beta u, u = 1, 2, \dots, N$, the sample mean \hat{Y}_{BSSM} coincides with the population mean.

Proof In the presence of linear trend, the sample mean under balanced systematic sampling with multiple random starts, may be written as

$$\begin{aligned}
 \hat{Y}_{BSSM}(r_1, r_2, \dots, r_t) &= \frac{1}{n} \left\{ \sum_{j=1}^{\frac{n}{2t}} [(\alpha + \beta(r_1 + 2t(j-1)k)) + (\alpha + \beta(r_2 + 2t(j-1)k)) + \dots + \right. \\
 &\quad \left. (\alpha + \beta(r_t + 2t(j-1)k))] + \sum_{j=1}^{\frac{n}{2t}} [(\alpha + \beta(2tjk - r_1 + 1)) + \right. \\
 &\quad \left. (\alpha + \beta(2tjk - r_2 + 1)) + \dots + (\alpha + \beta(2tjk - r_t + 1))] \right\} \\
 &= \frac{1}{n} \left[n\alpha + \frac{n\beta}{2} + 2kt^2 \frac{\left(\frac{n}{2t}\right)\left(\frac{n}{2t}-1\right)}{2} + 2kt^2 \frac{\left(\frac{n}{2t}\right)\left(\frac{n}{2t}+1\right)}{2} \right] \\
 &= \alpha + \beta \left(\frac{nk+1}{2} \right)
 \end{aligned}$$

$$= \alpha + \beta \frac{(N+1)}{2}$$

Hence the proof.

(ii) Modified Systematic Sampling with multiple random starts

This method of sampling developed by Singh, Jindal, Garg (1968) is another scheme meant for populations exhibiting linear trend. Under this method, a sample of size n is drawn by selecting a pair of units equidistant from both the ends of the population in a systematic manner. The details are furnished below.

As in the case of linear and balanced systematic sampling here also a random number r is selected from 1 to k . When the sample size n is even, the sample corresponding to the random start $r(r = 1, 2, \dots, k)$ is given by the set of units with labels

$$s_r = \left\{ r + jk, N - r - jk + 1, j = 0, 1, \dots, \frac{n}{2} - 1 \right\}$$

We shall describe MSS with multiple random starts as follows. If a MSS with t random starts is desired, we choose at random t numbers from 1 to tk . Each random number will contribute $\frac{n}{t}$ units to the sample, thus leading to an ultimate sample of size n . The sample corresponding to the random start $r_i (i = 1, 2, \dots, t)$ is given by the set of units with labels $\left\{ r_i + t(j-1)k, N - r_i - t(j-1)k + 1, j = 1, 2, \dots, \frac{n}{2t} \right\}$.

The sample mean \hat{Y}_{MSSM} of modified systematic sampling under multiple random starts, corresponding to the random starts r_1, r_2, \dots, r_t , can be written as

$$\hat{Y}_{MSSM} = \frac{1}{n} \sum_{j=1}^{\frac{n}{2t}} \sum_{i=1}^t (Y_{r_i + t(j-1)k} + Y_{N - r_i - t(j-1)k + 1}) \tag{4}$$

Theorem 3 The sample mean \hat{Y}_{MSSM} under modified systematic sampling with multiple random starts, for a choice of t random starts r_1, r_2, \dots, r_t , is unbiased for the population mean.

Proof Corresponding to a choice of t random starts r_1, r_2, \dots, r_t the sample mean \hat{Y}_{MSSM} under modified systematic sampling with multiple random starts can be

$$\text{written as } \hat{Y}_{MSSM}(r_1, r_2, \dots, r_t) = \frac{1}{n} \sum_{j=1}^{\frac{n}{2t}} \sum_{i=1}^t (Y_{r_i + t(j-1)k} + Y_{N - r_i - t(j-1)k + 1})$$

Taking expectations on both sides we get,

$$\begin{aligned}
 E\left(\widehat{Y}_{MSSM}\right) &= \frac{1}{n} \sum_{i=1}^{tk} (T'_i + U'_i) \text{ where, } T'_i = \sum_{j=1}^{\frac{n}{2t}} Y_{r_i+t(j-1)k} \text{ and } U'_i = Y_{N-r_i-t(j-1)k} \\
 &= \frac{1}{nk} \sum_{i=1}^{tk} Y_i \\
 &= \bar{Y}
 \end{aligned}$$

Hence the proof.

Theorem 4 When the population units are modelled according to $Y_u = \alpha + \beta u, u = 1, 2, \dots, N$, the sample mean \widehat{Y}_{MSSM} under modified systematic sampling with multiple random starts coincides with the population mean.

Proof Choose t random numbers r_1, r_2, \dots, r_t from 1 to tk . The sample mean \widehat{Y}_{MSSM} in the presence of linear trend can be expressed as

$$\begin{aligned}
 \widehat{Y}_{MSSM}(r_1, r_2, \dots, r_t) &= \frac{1}{n} \left\{ \sum_{j=1}^{\frac{n}{2t}} [(\alpha + \beta(r_1 + t(j-1)k)) + (\alpha + \beta(r_2 + t(j-1)k)) + \dots + \right. \\
 &\quad \left. (\alpha + \beta(r_t + t(j-1)k))] + \sum_{j=1}^{\frac{n}{2t}} [(N - r_1 - t(j-1)k + 1) + \right. \\
 &\quad \left. + (N - r_2 - t(j-1)k + 1) + \dots + (N - r_t - t(j-1)k + 1)] \right\} \\
 &= \frac{1}{n} \left[t \frac{n}{2t} \alpha + t \frac{n}{2t} \alpha + tN\beta \frac{n}{2t} + \frac{n}{t} \frac{\beta}{2} t \right] \\
 &= \alpha + \beta \left(\frac{N+1}{2} \right) \tag{5}
 \end{aligned}$$

Hence it can be concluded that the sample mean coincides with the population mean under MSSM in the presence of linear trend.

3. Yates Corrected Estimator

In the previous two sections of this paper, we have seen that the sample mean coincides with the population mean in the presence of linear trend, under balanced and modified systematic sampling with multiple random starts, which are developed in this work. This property is comparable with that of the BSS and MSS with single random start. In this section, it is intended to develop an estimator corresponding to LSS with multiple random starts which is an analogue of Yates Corrected estimator (Yates (1948)) with single random start. The Yates corrected estimator coincides with the population mean in the presence of linear trend. To maintain the readability of the paper, it is briefly given below.

When the r^{th} group S_r is drawn as sample, the first and last units in the sample are corrected by the weights λ_1 and λ_2 respectively and the corresponding sample mean is taken as an estimator for the population mean where the weights λ_1 and λ_2 are selected so that the sample mean coincides with the population mean in the presence of linear trend. That is, the corrected mean

$$\bar{y}_c = \frac{1}{n} \left[\lambda_1 Y_r + \sum_{j=2}^{n-1} Y_{r+(j-1)k} + \lambda_2 Y_{r+(n-1)k} \right]$$

is equated to the population mean \bar{Y} after substituting $Y_u = \alpha + \beta u$, $u = 1, 2, \dots, N$ and on comparing corresponding coefficients λ_1 and λ_2 are obtained. We shall now proceed to extend this for LSS with multiple random starts.

For LSS with t random numbers r_1, r_2, \dots, r_t , we take

$$\hat{Y}_c(r_1, r_2, \dots, r_t) = \frac{1}{tn} \left[\lambda_1 \sum_{v=1}^s Y_{r_v} + \sum_{v=1}^s \sum_{j=2}^{n-1} (Y_{r_v+(j-1)k}) + \lambda_2 \sum_{v=1}^s (Y_{r_v+(n-1)k}) \right] \quad (6)$$

as an estimator, where the weights λ_1 and λ_2 are given respectively to the first and last elements of the groups corresponding to the random starts selected. The remaining units are given unit weights. The weights λ_1 and λ_2 are chosen in such a manner that the estimator coincides with the population mean when the population units are modelled by (3). The following theorem gives the expressions for the weights λ_1 and λ_2 occurring in expression (6).

Theorem 5 In the presence of linear trend, the weights λ_1 and λ_2 in the Yates corrected estimator $\hat{Y}_c(r_1, r_2, \dots, r_s)$ defined by (6) are given by

$$\lambda_1 = \frac{1}{2} + \left[\frac{2n(r_1 + r_2 + \dots + r_s) - s(k+n)}{2s(n-1)k} \right] \text{ and } \lambda_2 = 1 + \left[\frac{ns(1+k) - 2(r_1 + r_2 + \dots + r_s)}{2s(n-1)k} \right]$$

Proof Under the model $Y_u = \alpha + \beta u$, $u = 1, 2, \dots, N$, the expression for the corrected estimator with random starts r_1, r_2, \dots, r_t can be expressed as

$$\begin{aligned} \hat{Y}_c(r_1, r_2, \dots, r_s) = \frac{1}{sn} \left\{ \lambda_1 [(\alpha + \beta r_1) + (\alpha + \beta r_2) + \dots + (\alpha + \beta r_s)] + \right. \\ \left. + \sum_{j=2}^{n-1} [(\alpha + \beta(r_1 + (j-1)k)) + (\alpha + \beta(r_2 + (j-1)k)) + \dots + \right. \\ \left. (\alpha + \beta(r_s + (j-1)k))] + \lambda_2 [(\alpha + \beta(r_1 + (n-1)k)) + \right. \end{aligned}$$

$$\begin{aligned}
 & \left. (\alpha + \beta(r_2 + (n-1)k)) + \dots + (\alpha + \beta(r_s + (n-1)k)) \right] \left. \right\} \\
 & = \frac{1}{sn} \left\{ (\lambda_1 + \lambda_2 + n - 2)(s\alpha + \beta(r_1 + r_2 + \dots + r_s)) + \right. \\
 & \quad \left. \beta \left[\lambda_2 s(n-1)k + sk \frac{(n-2)(n-1)}{2} \right] \right\}
 \end{aligned}$$

On equating the above expression to the population mean namely, $\alpha + \beta \frac{(N+1)}{2}$, we get

$$\begin{aligned}
 & \frac{1}{sn} \left\{ (\lambda_1 + \lambda_2 + n - 2)(s\alpha + \beta(r_1 + r_2 + \dots + r_s)) + \beta \left[\lambda_2 s(n-1)k + sk \frac{(n-2)(n-1)}{2} \right] \right\} \\
 & = \alpha + \beta \frac{(N+1)}{2}
 \end{aligned}$$

Comparing the coefficients of α and β we get,

$$\lambda_1 + \lambda_2 = 2 \text{ and}$$

$$\frac{1}{sn} \left[(r_1 + r_2 + \dots + r_s)(\lambda_1 + \lambda_2 + n - 2) + \lambda_2 s(n-1)k + sk \frac{(n-2)(n-1)}{2} \right] = \frac{(N+1)}{2}$$

Solving for λ_1 and λ_2 , we get

$$\lambda_2 = 1 + \left[\frac{ns(1+k) - 2(r_1 + r_2 + \dots + r_s)}{2s(n-1)k} \right] \tag{7}$$

and

$$\begin{aligned}
 \lambda_1 & = 2 - \lambda_2 \\
 & = \frac{1}{2} + \left[\frac{2n(r_1 + r_2 + \dots + r_s) - s(k+n)}{2s(n-1)k} \right] \tag{8}
 \end{aligned}$$

It may be noted that the weights do not require any prior knowledge of α & β .

4. Average Variance

In conventional finite sampling, values assumed by survey variables over various population units are treated as constants, and the main aim of a survey statistician is to develop a sampling strategy that minimizes the variance or mean-squared error of an estimator that is proposed to estimate the unknown parametric values. The super population model adopts a different approach to study this problem. The values assumed by the population units with respect to

the survey variables are assumed to be the realized values of random variables having a known probability distribution which involves both known and unknown parametric values. In this approach called the ‘super population model approach’ one tries to identify a sampling-estimating strategy that minimizes the average mean square error or the variance where averaging is done with respect to the probability distribution of the underlying variables, namely, Y_1, Y_2, \dots, Y_N . In order to compare the efficiency of the sampling strategies proposed and discussed in detail in the previous sections we need to find the average variance of the estimators under the super population model suitable for populations with linear trend. The model is described as follows:

Y_1, Y_2, \dots, Y_n are random variables modeled by

$$Y_u = \alpha + \beta u + e_u,$$

where $E_M(e_u) = 0, V_M(e_u) = \sigma^2 u^g$ and $\text{cov}_M(e_u, e_v) = 0 \quad u \neq v$ (9)

Under the assumed model,

$$\begin{aligned} \bar{Y} &= \frac{1}{N} \sum_{u=1}^N Y_u \\ &= \frac{1}{N} \sum_{u=1}^N (\alpha + \beta u + e_u) \\ &= \frac{1}{N} \left[N\alpha + \beta \frac{N(N+1)}{2} + \sum_{u=1}^N e_u \right] \\ &= \alpha + \beta \frac{(nk+1)}{2} + \sum_{u=1}^N e_u \end{aligned}$$

(i) Average Variance under LSS with multiple random starts

Denoting the Yates estimator described in Section 3 by \hat{Y}_{LSSM} for the case of linear systematic sampling with multiple random starts, we express its average variance under the super population model given in (9) as

$$\begin{aligned} E_M V(\hat{Y}_{LSSM}) &= E_M E(\hat{Y}_{LSSM} - \bar{Y})^2 \\ &= \frac{1}{\binom{k}{t}} \sum_{r_1=1}^{k-t+1} \sum_{r_2=r_1+1}^{k-t+2} \dots \sum_{r_t=r_{t-1}+1}^k \left\{ \frac{\sigma^2}{t^2 n^2} [\lambda_1^2 (r_1^g + r_2^g + \dots + r_t^g) + \right. \\ &\quad \left. \lambda_2^2 ((r_1 + (n-1)k)^g + (r_2 + (n-1)k)^g + \dots + (r_t + (n-1)k)^g) \right\} \end{aligned}$$

$$\begin{aligned}
 & + \sum_{j=2}^{n-1} ((r_1 + (j-1)k)^g + (r_2 + (j-1)k)^g + \dots + (r_t + (j-1)k)^g)] + \\
 & \frac{1}{N^2} \sum_{u=1}^N \sigma^2 u^g - \frac{2\sigma^2}{nN} \left(\lambda_1 (r_1^g + r_2^g + \dots + r_t^g) + \lambda_2 ((r_1 + (n-1)k)^g \right. \\
 & \qquad \qquad \qquad \left. + (r_2 + (n-1)k)^g + \dots + (r_t + (n-1)k)^g) + \right. \\
 & \left. \sum_{j=2}^{n-1} [(r_1 + (j-1)k)^g + (r_2 + (j-1)k)^g + \dots + (r_t + (j-1)k)^g] \right) \} \quad (9)
 \end{aligned}$$

(ii) Average Variance under BSS with multiple random starts

The average variance of the estimator \hat{Y}_{BSSM} under balanced systematic sampling with multiple random under the super population model given in (9) is

$$\begin{aligned}
 E_M V(\hat{Y}_{BSSM}) &= E_M E(\hat{Y}_{BSSM} - \bar{Y})^2 \\
 &= \frac{1}{\binom{tk}{t}} \sum_{r_1=1}^{tk-t+1} \sum_{r_2=r_1+1}^{tk-t+2} \dots \sum_{r_t=r_{t-1}+1}^{tk} \left\{ \frac{\sigma^2}{n^2} \sum_{j=1}^n ((r_1 + 2t(j-1)k)^g + \right. \\
 & \qquad \qquad \qquad (r_2 + 2t(j-1)k)^g + \dots + (r_t + 2t(j-1)k)^g + \\
 & \qquad \qquad \qquad (2tjk - r_1 + 1)^g + (2tjk - r_2 + 1)^g + \dots + (2tjk - r_t + 1)^g) \\
 & \qquad \qquad \qquad \left. + \frac{1}{N^2} \sum_{u=1}^N \sigma^2 u^g - \frac{2\sigma^2}{nN} \left[\sum_{j=1}^n ((r_1 + 2t(j-1)k)^g + \right. \right. \\
 & \qquad \qquad \qquad (r_2 + 2t(j-1)k)^g + \dots + (r_t + 2t(j-1)k)^g + (2tjk - r_1 + 1)^g + \\
 & \qquad \qquad \qquad \left. \left. (2tjk - r_2 + 1)^g + \dots + (2tjk - r_t + 1)^g) \right] \right\} \quad (10)
 \end{aligned}$$

In the particular case of $t = 2$, we obtain from (9),

$$E_M V(\hat{Y}_{BSSM}) = \begin{cases} \sigma^2 & \text{if } g = 0 \\ \sigma^2 \frac{(N+1)}{2} & \text{if } g = 1 \end{cases} \quad (11)$$

(iii) Average Variance under MSS with multiple random starts

Under modified systematic sampling with multiple random starts the average variance of the estimator \hat{Y}_{MSSM} under the super population model given by (9) is given by

$$\begin{aligned}
 E_M V(\hat{Y}_{MSSM}) &= E_M E(\hat{Y}_{MSSM} - \bar{Y})^2 \\
 E_M V(\hat{Y}_{MSSM}) &= \frac{1}{\binom{tk}{t}} \sum_{r_1=1}^{tk-t+1} \sum_{r_2=r_1+1}^{tk-t+2} \dots \sum_{r_t=r_{t-1}+1}^{tk} \left\{ \frac{\sigma^2}{n^2} \left[\sum_{j=1}^{\frac{n}{2t}} ((r_1 + t(j-1)k)^g + \right. \right. \\
 &\quad (r_2 + t(j-1)k)^g + \dots + (r_t + t(j-1)k)^g + (N - r_1 - t(j-1)k + 1)^g + \\
 &\quad \left. \left. (N - r_2 - t(j-1)k + 1)^g + \dots + (N - r_t - t(j-1)k + 1)^g \right) \right. \\
 &\quad \left. + \frac{1}{N^2} \sum_{u=1}^N \sigma^2 u^g - \frac{2\sigma^2}{nN} \sum_{j=1}^{\frac{n}{2t}} ((r_1 + t(j-1)k)^g + (r_2 + t(j-1)k)^g \right. \\
 &\quad \left. + \dots + (r_t + t(j-1)k)^g + (N - r_1 - t(j-1)k + 1)^g + \right. \\
 &\quad \left. (N - r_2 - t(j-1)k + 1)^g + \dots + (N - r_t - t(j-1)k + 1)^g \right) \right\} \quad (12)
 \end{aligned}$$

As in the case of BSSM, we observe from (12) that when $g = 0, t = 2$,

In the particular case of $t = 2$, we obtain from (11),

$$E_M V(\hat{Y}_{BSSM}) = \begin{cases} \sigma^2 & \text{if } g = 0 \\ \sigma^2 \frac{(N+1)}{2} & \text{if } g = 1 \end{cases} \quad (13)$$

Hence we may conclude that for the case of two random starts, average variances of the sample mean under BSSM and MSSM are both equal when $g = 0$ and $g = 1$.

5. Comparative Study

In the previous section of this paper, the average variances of Yates corrected estimator under LSS with multiple random starts and sample means of the two sampling schemes namely, BSS with multiple random starts and MSS with multiple random starts have been derived. Even though the proposed methods

perform with the same efficiency when $g = 0$ for all choices of the number of random starts, their performance cannot be mathematically studied in view of their complex nature for other cases of g . Hence a numerical study has been carried out to compare their performances for different choices of $N, n, g,$ and t . The results of the study are given in Table 1. The entries in the table give the values of ratios of average of variance of Yates corrected estimator under LSS with multiple random starts to the average variance of sample mean under BSSM as well as MSSM. The contents of the table clearly support the use of BSSM and MSSM instead of Yates corrected estimator under LSSM for all choices of g and sample sizes.

Table 1: Ratios of the average variance of Yates Corrected Estimator (\hat{Y}_{LSSM} to the average variances of the estimators \hat{Y}_{BSSM} and \hat{Y}_{MSSM}

g	t	$\frac{E_M V(\hat{Y}_{LSSM})}{E_M V(\hat{Y}_{BSSM})}$			$\frac{E_M V(\hat{Y}_{LSSM})}{E_M V(\hat{Y}_{MSSM})}$		
		N=120 n=24	N=180 n=36	N=240 n=60	N=120 n=24	N=180 n=36	N=240 n=60
		0	2	4.01	4.01	3.00	4.01
	3	4.00	4.01	3.00	4.00	4.01	3.00
	4	4.00	4.00	3.00	4.00	4.01	3.00
1	2	4.01	4.01	3.00	4.01	4.01	3.00
	3	3.99	4.00	3.00	4.00	4.01	3.00
	4	3.96	3.99	3.00	4.00	4.00	3.00
2	2	4.01	4.01	3.01	4.01	4.01	3.01
	3	3.96	3.99	3.00	4.00	4.01	3.01
	4	3.87	3.95	3.00	4.00	4.00	3.00
3	2	4.01	4.02	3.01	4.01	4.02	3.01
	3	3.92	3.98	3.00	4.01	4.01	3.01
	4	3.71	3.88	3.00	4.00	4.00	3.01

6. Conclusion

Thus in this paper the following three sampling-estimating strategies involving multiple random starts have been compared for populations possessing linear trend.

Strategy – 1 Balanced systematic sampling with multiple random starts – Mean estimator

Strategy – 2 Modified systematic sampling with multiple random starts – Mean estimator

Strategy – 3 Linear Systematic Sampling with multiple random starts – Yates corrected estimator

The following are the findings of the study.

- (i) The multiple random start systematic sampling schemes, namely BSSM and MSSM estimate the population mean without any error in the presence of linear trend.
- (ii) In the case of BSSM and MSSM, the average variance of the estimator coincides for all choices of random starts and sample sizes for $g = 0$.
- (iii) The average variances of the estimators under BSSM and MSSM are equal when we use two random starts for all choices of g and sample sizes.
- (iv) It was also observed that BSSM and MSSM were more efficient when compared to LSSM for all choices of g and sample sizes.
- (v) Regarding the number of random starts to be used, it is recommended to use two random starts instead of more number of random starts since the numerical study indicates that average variances tend to increase as the number of random starts is increased. This pattern was observed for both BSSM and MSSM. While this is observed in the case of BSSM and MSSM, the average variance under LSSM (with Yates corrected estimator) tends to decrease as the number of random starts increase, for the cases $g = 2$ and $g = 3$.

Acknowledgement

The authors wish to thank the referees for their constructive suggestions which have helped in improving the organization of the paper to a great extent.

References

1. C.H. Leu, and K.W. Tsui, (1996). New partially systematic sampling, *Statistica Sinica*, 6:617-630.
2. D. Singh, K.K. Jindal, J.N. Garg, (1968). On modified systematic sampling, *Biometrika* 55: 541-546.
3. F. Yates, (1948). Systematic Sampling, *Phil. Trans. Roy. Soc., London*, A241: 345-377.
4. H.J. Chang, K.C. Huang, (2000). Remainder linear systematic sampling, *Sankhya B*, 62:376-384.
5. J.W. Tukey, (1950). Some sampling simplified, *Jour. Amer. Stat. Assoc.*, 45:501-519.
6. K.C. Huang, (2004). Mixed random systematic sampling designs, *Metrika*, 59:1-11.
7. S. Sampath, and N. Uthayakumaran, (1998). Markov systematic sampling, *Biometrical Journal*, 40:883-895.

8. K. Suresh Chandra, S. Sampath, G.K. Balasubramani, (1992). Markov sampling for finite populations, *Biometrika*, 79:210-213.
9. V.K. Sethi, (1965). On optimum pairing of units, *Sankhya B*, 27: 315-320.
10. W. Gautschi, (1957). Some remarks on systematic sampling, *Ann. Math. Statist.*, 28: 385-394.