REVSTAT – Statistical Journal
Volume 16, Number 2, April 2018, 187–212

MODIFIED SYSTEMATIC SAMPLING WITH MULTIPLE RANDOM STARTS

Authors: Sat Gupta
– Department of Mathematics and Statistics, University of North Carolina, Greensboro, USA
sngupta@uncg.edu

Zaheen Khan
– Department of Mathematics and Statistics, Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan
zkurdu@gmail.com

Javid Shabbir
– Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan
javidshabbir@gmail.com

Received: February 2017 Revised: July 2017 Accepted: August 2017

Abstract:

- Systematic sampling has been facing two problems since its beginning; situational complications, e.g., population size N not being a multiple of the sample size n, and unavailability of unbiased estimators of population variance for all possible combinations of N and n. These problems demand a sampling design that may solve the said problems in a practicable way. In this paper, therefore, a new sampling design is introduced and named as, "Modified Systematic Sampling with Multiple Random Starts". Linear systematic sampling and simple random sampling are the two extreme cases of the proposed design. The proposed design is analyzed in detail and various expressions have been derived. It is found that the expressions for linear systematic sampling and simple random sampling may be extracted from these expressions. Finally, a detailed efficiency comparison is also carried out in this paper.

Key-Words:

- *Modified Systematic Sampling; Linear Systematic Sampling; Simple Random Sampling; Circular Systematic Sampling; Modified Systematic Sampling: Linear Trend.*

AMS Subject Classification:

- 62DO5.

1. INTRODUCTION

Systematic sampling is generally more efficient than Simple Random Sampling (SRS) because SRS may include bulk of units from high density or low density parts of the region, whereas the systematic sampling ensures even coverage of the entire region for all units. In many situations, systematic sampling is also more precise than stratified random sampling. Due to this, researchers and field workers are often inclined towards systematic sampling.

On the other hand, in Linear Systematic Sampling (LSS), we may obtain sample sizes that vary when the population size N is not a multiple of the sample size n, i.e., $N \neq nk$, where k is the sampling interval. However, this problem can be dealt by Circular Systematic Sampling (CSS), Modified Systematic Sampling (MSS) proposed by Khan *et al.* (2013), Remainder Linear Systematic Sampling (RLSS) proposed by Chang and Huang (2000) and Generalized Modified Linear Systematic Sampling Scheme (GMLSS) proposed by Subramani and Gupta (2014). Another well-known and long-standing problem in systematic sampling design is an absence of a design based variance estimator that is theoretically justified and generally applicable. The main reason behind this problem lies in the second-order inclusion probabilities which are not positive for all pairs of units under systematic sampling scheme. It is also obvious that population variance can be unbiasedly estimated if and only if the second-order inclusion probabilities are positive for all pairs of units. To overcome this problem, several alternatives have been proposed by different researchers. However, the simplest one is the use of multiple random starts in systematic sampling. This procedure was adopted by Gautschi (1957) in case of LSS. Later on, Sampath (2009) has considered LSS with two random starts and developed an unbiased estimator for finite-population variance. Sampath and Ammani (2012) further studied the other versions (balanced and centered systematic sampling schemes) of LSS for estimating the finite-population variance. They also discussed the question of determination of the number of random starts. Besides these attempts, the other approaches proposed by different researchers in the past are not much beneficial due to the considerable loss of simplicity.

From the attempts of Gautschi (1957), Sampath (2009), Sampath and Ammani (2012) and Naidoo *et al.* (2016), unbiased estimation of population variance becomes possible just for the case in which $N = nk$. Therefore, to avoid the difficulty in estimation of population variance for the case $N \neq nk$, practitioners are unwillingly inclined towards SRS instead of systematic sampling. Such limitations demand a more generalized sampling design which can play wide-ranging role in the theory of systematic sampling. Thus, in this paper we propose Modified Systematic Sampling with Multiple Random Starts (MSSM). The MSSM ensures unbiased estimation of population variance for the situation where $N \neq nk$.

As one can see, MSS proposed by Khan *et al.* (2013) nicely arranges the population units into k_1 systematic groups each containing s number of units. In MSS, initially a group is selected at random and other $(m-1)$ groups are systematically selected. In this way, a sample of size n consisting of m groups of size s is achieved. Whereas in MSSM, we propose to select all m systematic groups at random to get a sample of size n. Such selection enables us to derive the unbiased variance estimator in systematic sampling. It is interesting to note that LSS and SRS become the extreme cases of MSSM. The MSSM becomes LSS in a situation when N itself is the least common multiple (lcm) of N and n or equivalently $N = nk$, and becomes SRS if lcm is the product of N and n. Because in the case when $N = nk$ we are selecting $m = 1$ group at random which resembles LSS. Whereas, if lcm is the product of N and n we have N groups each containing only one unit from which we are selecting n groups at random in MSSM, which is similar to SRS. In case of LSS, variance estimation can be easily dealt with by Gautschi (1957), Sampath (2009), Sampath and Ammani (2012) and Naidoo *et al.* (2016); whereas the worst case of MSSM is SRS, where unbiased variance estimation can be done using SRS approach.

2. MODIFIED SYSTEMATIC SAMPLING WITH MULTIPLE RANDOM STARTS

Suppose, we have a population of size N, the units of which are denoted by $\{U_1, U_2, U_3, ..., U_N\}$. To select a sample of size n from this population, we will arrange N units into $k_1 = L/n$ (where L is the least common multiple of N and n) groups, each containing $s = N/k_1$ elements. The partitioning of groups is shown in Table 1. A set of $m = L/N$ groups from these k_1 groups are selected using simple random sampling without replacement to get a sample of size $ms = n$.

Table 1: Labels of population units arranged in MSSM.

		Labels of Sample units				
Groups	G_1	U_1	U_{k_1+1}	.	.	$U_{(s-1)k_1+1}$
	G_2	U_2	U_{k_1+2}	.	.	$U_{(s-1)k_1+2}$
	G_3	U_3	U_{k_1+3}	.	.	$U_{(s-1)k_1+3}$
	G_i	U_i	U_{k_1+i}	.	.	$U_{(s-1)k_1+i}$
	G_{k_1}	U_{k_1}	U_{2k_1}	.	.	$U_{sk_1=N}$

Thus sample units with random starts $r_i (i = 1, 2, ..., m)$ selected from 1 to k_1 correspond to the following labels:

$$r_i + (j-1)k_1, \qquad i = 1, 2, ..., m \quad \text{and} \quad j = 1, 2, ..., s. \tag{2.1}$$

2.1. Estimation of Population Mean and its Variance in MSSM

Consider the mean estimator

$$\bar{y}_{MSSM} = \frac{1}{ms}\sum_{i=1}^{m}\sum_{j=1}^{s} y_{r_ij} = \frac{1}{m}\sum_{i=1}^{m}\Big(\frac{1}{s}\sum_{j=1}^{s} y_{r_ij}\Big).$$

where y_{r_ij} is the value of the jth unit of the ith random group.

Taking expectation on both sides, we get:

$$\begin{aligned}E\left(\bar{y}_{MSSM}\right) &= \frac{1}{m}\sum_{i=1}^{m} E\Big(\frac{1}{s}\sum_{j=1}^{s} y_{r_ij}\Big)\\ &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{k_1}\sum_{i=1}^{k_1}\Big(\frac{1}{s}\sum_{j=1}^{s} y_{ij}\Big) = \frac{1}{sk_1}\sum_{i=1}^{k_1}\sum_{j=1}^{s} y_{ij} = \mu,\end{aligned}$$

where y_{ij} is the value of the jth unit of the ith group and μ is the population mean.

The variance of $\bar{y}_{MSSM}$ is given by

$$V\left(\bar{y}_{MSSM}\right) = E\left(\bar{y}_{MSSM} - \mu\right)^2 = \frac{1}{m^2}E\left[\sum_{i=1}^{m}\left(\bar{y}_{r_i.} - \mu\right)\right]^2,$$

where $\bar{y}_{r_i.}$ is the mean of ith random group.

After simplification, we have:

$$V\left(\bar{y}_{MSSM}\right) = \frac{1}{mk_1}\frac{(k_1 - m)}{(k_1 - 1)}\sum_{i=1}^{k_1}\left(\bar{y}_{i.} - \mu\right)^2, \tag{2.2}$$

where $\bar{y}_{i.}$ is the mean of ith group.

Further, it can be observed that in a situation when MSSM becomes LSS, the variance expression given in Equation (2.2) reduces to variance of LSS, i.e., $V\left(\bar{y}_{MSSM}\right) = \frac{1}{k}\sum_{i=1}^{k}\big(\bar{y}_{i.} - \mu\big)^2 = V\left(\bar{y}_{LSS}\right)$.

Similarly, in the case when MSSM becomes SRS, $V\left(\bar{y}_{MSSM}\right)$ reduces to variance of SRS without replacement, i.e.,

$$V\left(\bar{y}_{MSSM}\right) = \frac{(N-n)}{nN}\frac{1}{(N-1)}\sum_{i=1}^{N}\big(y_i - \mu\big)^2 = V\left(\bar{y}_{SRSWOR}\right).$$

The alternative expressions for $V\left(\bar{y}_{MSSM}\right)$ have been presented in Theorems 2.1, 2.2 and 2.3:

Theorem 2.1. *The variance of sample mean under MSSM is:*

$$V\left(\bar{y}_{MSSM}\right) = \frac{1}{mN}\,\frac{(k_1-m)}{(k_1-1)}\Big[(N-1)S^2 - k_1(s-1)S^2_{wg}\Big],$$

where $S^2 = \frac{1}{N-1}\sum_{i=1}^{k_1}\sum_{j=1}^{s}(y_{ij}-\mu)^2$, *and* $S^2_{wg} = \frac{1}{k_1(s-1)}\sum_{i=1}^{k_1}\sum_{j=1}^{s}(y_{ij}-\bar{y}_i)^2$ *is the variance among the units that lie within the same group.*

Proof: From analysis of variance, we have:

$$\sum_{i=1}^{N}(y_i-\mu)^2 = s\sum_{i=1}^{k_1}(\bar{y}_i-\mu)^2 + \sum_{i=1}^{k_1}\sum_{j=1}^{s}(y_{ij}-\bar{y}_i)^2, \quad \text{or}$$

$$(N-1)S^2 = s\sum_{i=1}^{k_1}(\bar{y}_i-\mu)^2 + k_1(s-1)S^2_{wg}.$$

Thus

$$V\left(\bar{y}_{MSSM}\right) = \frac{1}{mN}\,\frac{(k_1-m)}{(k_1-1)}\Big[(N-1)S^2 - k_1(s-1)S^2_{wg}\Big]. \tag{2.3}$$

□

Theorem 2.2. *The variance of sample mean under MSSM is:*

$$V\left(\bar{y}_{MSSM}\right) = \frac{1}{n}\left(\frac{k_1-m}{k_1-1}\right)\left(\frac{N-1}{N}\right)S^2\Big[1+(s-1)\rho_w\Big],$$

where

$$\rho_w = \frac{\sum\limits_{i=1}^{k_1}\sum\limits_{j=1}^{s}\sum\limits_{\substack{j'=1\\ j'\neq j}}^{s}(y_{ij}-\mu)(y_{ij'}-\mu)/s(s-1)k_1}{\sum\limits_{i=1}^{k_1}\sum\limits_{j=1}^{s}(y_{ij}-\mu)^2/sk_1}.$$

Proof: Note that

$$\begin{aligned}
V\left(\bar{y}_{MSSM}\right) &= \frac{1}{mk_1}\,\frac{(k_1-m)}{(k_1-1)}\sum_{i=1}^{k_1}(\bar{y}_i-\mu)^2\\
&= \frac{1}{s^2mk_1}\,\frac{(k_1-m)}{(k_1-1)}\sum_{i=1}^{k_1}\Big[\sum_{j=1}^{s}(y_{ij}-\mu)\Big]^2\\
&= \frac{1}{s^2mk_1}\,\frac{(k_1-m)}{(k_1-1)}\Big[\sum_{i=1}^{k_1}\sum_{j=1}^{s}(y_{ij}-\mu)^2 + \sum_{i=1}^{k_1}\sum_{j\neq 1}^{s}(y_{ij}-\mu)(y_{iu}-\mu)\Big]\\
&= \frac{1}{s^2mk_1}\,\frac{(k_1-m)}{(k_1-1)}\Big[(sk_1-1)S^2 + (sk_1-1)(s-1)S^2\rho_w\Big].
\end{aligned}$$

Hence

$$V(\bar{y}_{MSSM}) = \frac{1}{n}\frac{(k_1-m)}{(k_1-1)}\frac{(N-1)}{N}S^2\Big[1+(s-1)\rho_w\Big], \tag{2.4}$$

where ρ_w is the intraclass correlation between the pairs of units that are in the same group. □

Theorem 2.3. *The variance of $\bar{y}_{MSSM}$ is:*

$$V(\bar{y}_{MSSM}) = \frac{(k_1-m)}{mN}S^2_{wst}\Big[1+(s-1)\rho_{wst}\Big],$$

where

$$S^2_{wst} = \frac{1}{s(k_1-1)}\sum_{j=1}^{s}\sum_{i=1}^{k_1}(y_{ij}-\bar{y}_{.j})^2$$

and

$$\rho_{wst} = \frac{\sum\limits_{i=1}^{k_1}\sum\limits_{j=1}^{s}\sum\limits_{\substack{j'=1\\ j'\neq j}}^{s}(y_{ij}-\bar{y}_j)(y_{ij'}-\bar{y}_{j'})}{s(s-1)(k_1-1)S^2_{wst}}.$$

Proof: Note that

$$\begin{aligned}
V(\bar{y}_{MSSM}) &= \frac{1}{mk_1}\frac{(k_1-m)}{(k_1-1)}\sum_{i=1}^{k_1}(\bar{y}_i-\mu)^2\\
&= \frac{1}{mk_1}\frac{(k_1-m)}{(k_1-1)}\sum_{i=1}^{k_1}\left[\frac{1}{s}\sum_{j=1}^{s}y_{ij}-\frac{1}{s}\sum_{j=1}^{s}\bar{y}_j\right]^2\\
&= \frac{1}{s^2mk_1}\frac{(k_1-m)}{(k_1-1)}\sum_{i=1}^{k_1}\left[\sum_{j=1}^{s}(y_{ij}-\bar{y}_j)\right]^2\\
&= \frac{1}{smN}\frac{(k_1-m)}{(k_1-1)}\left[\sum_{j=1}^{s}\sum_{i=1}^{k_1}(y_{ij}-\bar{y}_j)^2+\sum_{i=1}^{k_1}\sum_{j=1}^{s}\sum_{\substack{j'=1\\ j'\neq j}}^{s}(y_{ij}-\bar{y}_j)(y_{ij'}-\bar{y}_{j'})\right]\\
&= \frac{1}{smN}\frac{(k_1-m)}{(k_1-1)}s(k_1-1)S^2_{wst}\Big[1+(s-1)\rho_{wst}\Big].
\end{aligned}$$

Hence

$$V(\bar{y}_{MSSM}) = \left(\frac{k_1-m}{mN}\right)S^2_{wst}\Big[1+(s-1)\rho_{wst}\Big]. \tag{2.5}$$ □

3. MEAN, VARIANCE AND EFFICIENCY COMPARISON OF MSSM FOR POPULATIONS EXHIBITING LINEAR TREND

Generally the efficiency of every new systematic sampling design is evaluated for populations having linear trend. Therefore, consider the following linear model for the hypothetical population

$$Y_t = \alpha + \beta t, \qquad t = 1, 2, 3, ..., N, \tag{3.1}$$

where α and β respectively are the intercept and slope terms in the model.

3.1. Sample Mean under MSSM

$$\bar{y}_{MSSM} = \alpha + \frac{\beta}{ms} \sum_{i=1}^{m} \sum_{j=1}^{s} \Big\{ r_i + (j-1)k_1 \Big\}, \quad \text{or}$$

$$\bar{y}_{MSSM} = \alpha + \frac{\beta}{m} \Big\{ \sum_{i=1}^{m} r_i + \frac{m}{2}(s-1)k_1 \Big\}. \tag{3.2}$$

$$E\left(\bar{y}_{MSSM}\right) = \alpha + \beta \frac{(N+1)}{2} = \mu. \tag{3.3}$$

$$V\left(\bar{y}_{MSSM}\right) = E\left\{\bar{y}_{MSSM} - E(\bar{y}_{MSSM})\right\}^2 = \beta^2 E\Big[\frac{1}{m}\sum_{i=1}^{m} r_i - \frac{(k_1+1)}{2}\Big]^2.$$

Hence

$$V\left(\bar{y}_{MSSM}\right) = \beta^2 \frac{(k_1+1)(k_1-m)}{12m}. \tag{3.4}$$

Note that $m = 1$ and $k_1 = k$ in situations when MSSM is LSS; therefore

$$V\left(\bar{y}_{MSSM}\right) = \beta^2 \frac{(k^2-1)}{12} = V\left(\bar{y}_{LSS}\right). \tag{3.5}$$

Similarly, $m = n$ and $k_1 = N$ in situations when MSSM is SRS, so

$$V\left(\bar{y}_{MSSM}\right) = \beta^2 \frac{(N+1)(N-n)}{12n} = V\left(\bar{y}_{SRS}\right). \tag{3.6}$$

The efficiency of MSSM with respect to SRS can be calculated as below:

$$\textit{Efficiency} = \frac{V\left(\bar{y}_{SRS}\right)}{V\left(\bar{y}_{MSSM}\right)} = \frac{m(N+1)(N-n)}{(k_1+1)(k_1-m)n} = \frac{(sk_1+1)}{(k_1+1)} \geq 1, \tag{3.7}$$

as $s \geq 1$. One can see that MSSM is always more efficient than SRS if $s > 1$ and is equally efficient if $s = 1$.

4. ESTIMATION OF VARIANCE

Sampath and Ammani (2012) have considered LSS, Balanced Systematic Sampling (BSS) proposed by Sethi (1965), and Modified Systematic Sampling (MS) proposed by Singh *et al.* (1968) using multiple random starts. They have derived excellent expressions of unbiased variance estimators and their variances for these schemes. However, these schemes are not applicable if $N \neq nk$. Fortunately, MSSM nicely handles this by producing unbiased variance estimator and its variance for the case, where $N \neq nk$. Adopting the procedure mentioned in Sampath and Ammani (2012), we can get an unbiased variance estimator and its variance in MSSM for the case where $N \neq nk$.

In MSSM, the probability that the i^{th} unit will be included in the sample is just the probability of including the group containing the specific unit in the sample. Hence, the first-order inclusion probability that corresponds to the population unit with label i, is given by

$$\pi_i = \frac{m}{k_1} = \frac{ms}{sk_1} = \frac{n}{N}, \qquad i = 1,\, 2,\, 3,\, ...,\, N.$$

In the second-order inclusion probabilities, the pairs of units may belong to the same or the different groups. The pairs of units belong to the same group only if the respective group is included in the sample. Thus, the second-order inclusion probabilities for pairs of units belonging to the same group are equivalent to the first-order inclusion probabilities, i.e.,

$$\pi_{ij} = \frac{m}{k_1} = \frac{ms}{sk_1} = \frac{n}{N}, \qquad i,\, j \in\, s_{r_u} \text{ for some } r_u \quad (r_u = 1, 2, ..., k_1).$$

On the other hand, pairs of units belonging to two different groups occurs only when the corresponding pair of groups is included in the sample. Hence, the second-order inclusion probability is given by

$$\pi_{ij} = \frac{m(m-1)}{k_1(k_1-1)}, \qquad \text{if } i \in\, s_{r_u} \text{ and } j \in s_{r_v} \text{ for some } u \neq v.$$

Thus

$$\pi_{ij} = \frac{m(m-1)}{k_1(k_1-1)} = \frac{ms(ms-s)}{sk_1(sk_1-s)} = \frac{n(n-s)}{N(N-s)}.$$

Since the second-order inclusion probabilities are positive for all pairs of units in the population, an unbiased estimator of population variance can be established. To accomplish this, the population variance

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \mu)^2$$

can be written as

$$S^2 = \frac{1}{2N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} (Y_i - Y_j)^2.$$

By using second-order inclusion probabilities, an unbiased estimator of the population variance can be obtained as

$$\hat{S}^2_{MSSM} = \frac{1}{2N(N-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{(y_i - y_j)^2}{\pi_{ij}}.$$

As $n = ms$, it means that there are m random sets each containing s units. Therefore, taking r_u $(u = 1, 2, ...m)$ as the random start for the u^{th} set, the expression for $\hat{S}^2_{MSSM}$ can be rewritten as:

$$\begin{aligned}
\hat{S}^2_{MSSM} &= \frac{1}{2N(N-1)} \left[\sum_{u=1}^{m} \left\{ \sum_{i=1}^{s} \sum_{\substack{j=1 \\ j \neq i}}^{s} \frac{(y_{r_u i} - y_{r_u j})^2}{\pi_{ij}} \right\} \right. \\
&\quad \left. + \sum_{\substack{u=1 \\ v=1 \\ u \neq v}}^{m} \left\{ \sum_{i=1}^{s} \sum_{j=1}^{s} \frac{(y_{r_u i} - y_{r_v j})^2}{\pi_{ij}} \right\} \right] \\
&= \frac{1}{2N(N-1)} \left[\frac{N}{n} \sum_{u=1}^{m} \left\{ \sum_{i=1}^{s} \sum_{\substack{j=1 \\ j \neq i}}^{s} (y_{r_u i} - y_{r_u j})^2 \right\} \right. \\
&\quad \left. + \frac{N(N-s)}{n(n-s)} \sum_{\substack{u=1 \\ v=1 \\ u \neq v}}^{m} \left\{ \sum_{i=1}^{s} \sum_{j=1}^{s} (y_{r_u i} - y_{r_v j})^2 \right\} \right] \\
&= \frac{1}{2N(N-1)} \left[\frac{N}{n} \sum_{u=1}^{m} \left\{ 2s \sum_{i=1}^{s} (y_{r_u i} - \bar{y}_{r_u})^2 \right\} \right. \\
&\quad \left. + \frac{N(N-s)}{n(n-s)} \sum_{\substack{u=1 \\ v=1 \\ u \neq v}}^{m} \left\{ \sum_{i=1}^{s} \sum_{j=1}^{s} \left((y_{r_u i} - \bar{y}_u)^2 + (y_{r_v j} - \bar{y}_{r_v})^2 + (\bar{y}_{r_u} - \bar{y}_{r_v})^2 \right) \right\} \right] \\
&= \frac{1}{2N(N-1)} \left[\frac{N}{n} \sum_{u=1}^{m} \left\{ 2s^2 \hat{\sigma}^2_{r_u} \right\} \right. \\
&\quad \left. + \frac{N(N-s)}{n(n-s)} \sum_{\substack{u=1 \\ v=1 \\ u \neq v}}^{m} \left\{ \left(s^2 \hat{\sigma}^2_{r_u} + s^2 \hat{\sigma}^2_{r_v} + s^2 (\bar{y}_{r_u} - \bar{y}_{r_v})^2 \right) \right\} \right],
\end{aligned}$$

where $\bar{y}_{r_u}$ and $\hat{\sigma}^2_{r_u} = \frac{1}{s} \sum_{i=1}^{s} (y_{r_u i} - \bar{y}_{r_u})^2$ are the mean and variance of the u^{th} group

$(u = 1, 2, ..., m)$. Further,

$$\begin{aligned}
\hat{S}^2_{MSSM} &= \frac{1}{2N(N-1)}\left[\frac{N}{n}\Big\{2s^2\sum_{u=1}^{m}\hat{\sigma}^2_{r_u}\Big\}\right. \\
&\quad \left. + \frac{N(N-s)}{n(n-s)}\left\{\Big(2(m-1)s^2\sum_{u=1}^{m}\hat{\sigma}^2_{r_u} + s^2\sum_{u=1}^{m}\sum_{\substack{v=1\\u\neq v}}^{m}(\bar{y}_{r_u}-\bar{y}_{r_v})^2\Big)\right\}\right] \\
&= \frac{1}{2N(N-1)}\left[\frac{N}{n}\Big\{2s^2\sum_{u=1}^{m}\hat{\sigma}^2_{r_u}\Big\}\right. \\
&\quad \left. + \frac{N(N-s)}{n(n-s)}\left\{\Big(2(m-1)s^2\sum_{u=1}^{m}\hat{\sigma}^2_{r_u} + s^2 2\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}(\bar{y}_{r_u}-\bar{y}_{r_v})^2\Big)\right\}\right] \\
&= \frac{s^2}{ms(N-1)}\left[\sum_{u=1}^{m}\hat{\sigma}^2_{r_u}\left\{1+\frac{(N-s)}{(ms-s)}(m-1)\right\}\right. \\
&\quad \left. + \frac{(N-s)}{(ms-s)}\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}(\bar{y}_{r_u}-\bar{y}_{r_v})^2\right].
\end{aligned}$$

Hence

$$\hat{S}^2_{MSSM} = \frac{1}{(N-1)}\left[\sum_{u=1}^{m}\hat{\sigma}^2_{r_u}\frac{N}{m} + \frac{(N-s)}{m(m-1)}\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}(\bar{y}_{r_u}-\bar{y}_{r_v})^2\right]. \tag{4.1}$$

For simplicity, Equation (4.1) can be written as

$$\hat{S}^2_{MSSM} = \frac{1}{(N-1)}\left[\frac{N}{m}\sum_{u=1}^{m}\hat{\sigma}^2_{r_u} + \frac{(N-s)}{(m-1)}\sum_{u=1}^{m}(\bar{y}_{r_u}-\bar{y}_{MSSM})^2\right].$$

The resulting estimator obtained in Equation (4.1) is an unbiased estimator of population variance S^2. It is mentioned in Section 2, if lcm of N and n is the product of N and n, i.e., $L = N \times n$, then MSSM becomes SRS.

Consequently, $\hat{\sigma}^2_{r_u} = 0 (u = 1, 2, ..., m)$ and

$$\hat{S}^2_{MSSM} = \hat{S}^2_{SRS} = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i-\bar{y})^2,$$

which is a well-known unbiased estimator of S^2 in SRS without replacement.

4.1. Variance of $\hat{S}^2_{MSSM}$

The variance of $\hat{S}^2_{MSSM}$ is given by

(4.2)

$$V\left(\hat{S}^2_{MSSM}\right) = \frac{1}{m(N-1)^2}\left[\frac{N^2(k_1-m)}{(k_1-1)}\sigma_0^2 + \frac{(N-s)^2 k_1}{(m-1)}\left[\left\{\frac{(m-1)}{(k_1-1)} - \frac{(m-2)(m-3)}{(k_1-2)(k_1-3)}\right\}\mu_4 + \left\{\frac{(k_1-3)-(m-2)(k_1+3)}{(k_1-1)^2} + \frac{(m-2)(m-3)\left(k_1^2-3\right)}{(k_1-1)^2(k_1-2)(k_1-3)}\right\}\mu_2^2\right] + 2\frac{N(N-s)(k_1-m)}{(k_1-1)(k_1-2)}\left\{\sum_{r=1}^{k_1}\hat{\sigma}_r^2\left(\bar{y}_r-\bar{Y}\right)^2 - k_1\bar{\sigma}^2\mu_2\right\}\right]$$

(see details in Appendix A).

Note that, if $L = N$, then MSSM becomes LSS and the above formula is not valid in this case. Fortunately, in LSS, due to Gautschi (1957), the population is divided into $m'k$ groups of n/m' elements, and m' of these groups will randomly be selected to get a sample of size n. Thus, one can easily modify the above formula by just putting $m = m'$, $k_1 = m'k$ and $s = n/m'$ in Equation (A.9) and get $V\left(\hat{S}^2_{LSS}\right)$ as below:

$$V\left(\hat{S}^2_{LSS}\right) = \frac{1}{m'(N-1)^2}\left[\frac{N^2 m'(k-1)}{(m'k-1)}\sigma_0^2 + \frac{(m'N-n)^2 k}{m'(m'-1)}\left[\left\{\frac{(m'-1)}{(m'k-1)} - \frac{(m'-2)(m'-3)}{(m'k-2)(m'k-3)}\right\}\mu_4 + \left\{\frac{(m'k-3)-(m'-2)(m'k+3)}{(m'k-1)^2} + \frac{(m'-2)(m'-3)\left(m'^2k^2-3\right)}{(m'k-1)^2(m'k-2)(m'k-3)}\right\}\mu_2^2\right] + 2\frac{N(m'N-n)(k-1)}{(m'k-1)(m'k-2)}\left\{\sum_{r=1}^{m'k}\hat{\sigma}_r^2(\bar{y}_r-\mu)^2 - m'k\bar{\sigma}^2\mu_2\right\}\right]. \tag{4.3}$$

This is the general formula for the variance of unbiased variance estimator with m' random starts for LSS. Further, one can also easily deduce the following formula of $V\left(\hat{S}^2_{SRS}\right)$ by putting $k_1 = N$, $m = n$ and $s = 1$ in Equation (A.9):

$$V\left(\hat{S}^2_{SRS}\right) = \frac{N}{n(n-1)}\left[\left\{\frac{(n-1)}{(N-1)} - \frac{(n-2)(n-3)}{(N-2)(N-3)}\right\}\mu_4 + \left\{\frac{(N-3)-(n-2)(N+3)}{(N-1)^2} + \frac{\left(N^2-3\right)(n-2)(n-3)}{(N-1)^2(N-2)(N-3)}\right\}\mu_2^2\right]. \tag{4.4}$$

5. EFFICIENCY COMPARISON OF VARIANCE ESTIMATORS

In this section, we compare $\hat{S}^2_{MSSM}$ with $\hat{S}^2_{SRS}$ by using natural and simulated populations. Furthermore, this study is carried out for those choices of sample sizes in which the condition "$N < L < (N \times n)$" is satisfied. It has already been mentioned that MSSM becomes LSS when $L = N$. On the other hand, MSSM becomes SRS when $L = (N \times n)$.

5.1. Natural Populations

In Population 1 (see Murthy, 1967, p. 131–132), the data on volume of timber of 176 forest strips have been considered. In this data, the volume of timber has been arranged with respect to its length. In Population 2 (see Murthy, 1967, p. 228), the data of output along with the fixed capital of 80 factories have been considered. Here, output is arranged with respect to fixed capital. It is observed that the data considered in Population 1 and Population 2 approximately follow a linear trend. In this empirical study, the variances of $\hat{S}^2_{MSSM}$ and $\hat{S}^2_{SRS}$ are computed for various sample sizes and efficiency is computed using the expression:

$$\textit{Efficiency} = \frac{V\left(\hat{S}^2_{SRS}\right)}{V\left(\hat{S}^2_{MSSM}\right)}.$$

The population size N, sample size n, number of random starts m, number of elements in each group s, the number of groups k_1 containing the N units of the population and efficiency of MSSM over SRS are respectively presented in Columns 1 to 6 for Population 1 and Columns 7 to 12 for Population 2 in Table 2. From the efficiency comparison presented in Table 2, it has been observed that MSSM is more efficient than SRS. Moreover, one can also see that as the number of elements s in each group are increased, the efficiency of MSSM also increases. Such increase in efficiency is due to the fact that in MSSM, the units within the groups are arranged in a systematic pattern. So, more number of units with systematic pattern will cause increase in efficiency.

Table 2: Efficiency comparison of $\hat{S}^2_{MSSM}$ and $\hat{S}^2_{SRS}$ in both natural populations.

Population 1						Population 2					
N	n	m	s	k_1	Efficiency	N	n	m	s	k_1	Efficiency
176	10	5	2	88	1.41	80	6	3	2	40	2.31
	12	3	4	44	3.69		12	3	4	20	3.56
	14	7	2	88	2.04		14	7	2	40	2.29
	18	9	2	88	2.03		15	3	5	16	5.91
	20	5	4	44	3.64		18	9	2	40	2.28
	24	3	8	22	5.79		22	11	2	40	2.27
	26	13	2	88	2.01		24	3	8	10	14.11
	28	7	4	44	3.61		25	5	5	16	5.91
	30	15	2	88	2.00		26	13	2	40	2.27
	32	2	16	11	6.22		28	7	4	20	3.50
	34	17	2	88	2.00		30	3	10	8	10.85
	36	9	4	44	3.59		32	2	16	5	15.14
	38	19	2	88	2.00		34	17	2	40	2.26
	40	5	8	22	5.70		35	7	5	16	5.89
	42	21	2	88	1.99		36	9	4	20	3.49
	46	23	2	88	1.99		38	19	2	40	2.26
	50	25	2	88	1.99						

5.2. Simulated Populations

The simulation study, two populations of sizes 160 and 280 are generated for the following distribution with variety of parameters by using R-packages:

(i) *Uniform distribution:* Here only three sets of the parametric values are considered, i.e., (10, 20), (10, 30) and (10, 50).

(ii) *Normal distribution:* In this case, six sets of parametric values are considered with means 20, 40 and 60 and standard deviations 5 and 8.

(iii) *Gamma distribution:* Eight sets of parametric values are considered in this case. Here, 1, 3, 5 and 10 are considered as the values of scale parameter with 2 and 4 as the values of shape parameter.

In each distribution, using each combination of the parametric values for each choice of the sample size, each population with and without order is replicated 1000 times. $V\left(\hat{S}^2_{MSSM}\right)$ and $V\left(\hat{S}^2_{SRS}\right)$ are computed for each population (with and without order) for the various choices of sample sizes. The average of 1000 values of the variances of $\hat{S}^2_{MSSM}$ and $\hat{S}^2_{SRS}$ is then computed for each population.

The efficiencies, *Eff* 1 and *Eff* 2 of MSSM compared to SRS are computed using the following expressions:

$$Eff1 = \frac{Average\{V(\hat{S}^2_{SRSWOR})\}}{Average\{V(\hat{S}^2_{MSSM})\}} \qquad \text{without ordered population}$$

and

$$Eff2 = \frac{Average\{V(\hat{S}^2_{SRSWOR})\}}{Average\{V(\hat{S}^2_{MSSM})\}} \qquad \text{with ordered population.}$$

The efficiencies, *Eff* 1 and *Eff* 2 for Uniform distribution, Normal distribution and Gamma distribution are presented in Tables 3, 4 and 5 respectively.

It is observed that *Eff* 1 is approximately equal to 1 for almost all choices of parametric values and sample sizes. This mean that MSSM and SRS are equally efficient in case of random populations. Thus, for such populations, MSSM can be preferred over SRS due to the qualities that there are no more issues of unbiased estimation of population variance.

Furthermore, it is also observed from Tables 3, 4 and 5 that *Eff* 2 is greater than 1 in all cases. It indicates that MSSM is more efficient than SRS in ordered populations. The discussion of *Eff* 2 in Tables 3, 4 and 5 is as follows:

In Table 3, the efficiency (*Eff* 2) is not effected much by the different combinations of parametric values of the uniform distribution and changes are caused by the number of groups k_1. It is also observed that MSSM is much more efficient for the ordered populations of uniform distribution as compared to the normal and gamma distributions.

In Table 4, the efficiency *Eff* 2 is also not much changed like uniform distribution for different combinations of parametric values of the normal distribution. However, *Eff* 2 is mainly changed due to the formation of number of groups k_1 of the population units in MSSM. Efficiency will increase with the decrease in the number of groups k_1, and it will decrease with the increase in the number of groups k_1.

In Table 5, the efficiency *Eff* 2 is effected by the number of groups k_1 along with the shape parameter of the Gamma distribution. However, change in scale parameter has no significant effect on efficiency of MSSM. Here also the efficiency increases with decrease in the number of groups k_1.

From the above discussion, it is obvious that MSSM performs better than SRS for the populations that follow uniform and parabolic trends. However, such populations must be ordered with certain characteristics. To know further about the performance of MSSM, it would be interesting to study the variances of $\hat{S}^2_{MSSM}$ and $\hat{S}^2_{SRS}$ in the presence of linear trend. This study has been carried out in the following section.

Table 3: Efficiency of MSSM over SRS using uniform distribution.

Uniform Distribution										
N	n	m	s	k_1	$a=10,\ b=20$		$a=10,\ b=30$		$a=10,\ b=50$	
					Eff 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2
160	12	3	4	40	0.94	24.17	0.94	24.67	0.94	24.33
	14	7	2	80	1.00	5.99	0.99	6.04	0.98	6.05
	15	3	5	32	0.95	39.72	0.96	38.39	0.95	39.85
	18	9	2	80	0.99	6.19	0.99	6.16	1.00	6.15
	22	11	2	80	1.00	6.17	1.00	6.26	1.00	6.25
	24	3	8	20	0.94	91.20	0.95	90.56	0.98	89.22
	25	5	5	32	0.99	44.66	0.99	44.24	0.98	45.01
	26	13	2	80	1.00	6.30	1.00	6.37	1.00	6.26
	28	7	4	40	0.99	29.47	1.00	30.36	0.99	28.90
	30	3	10	16	0.98	125.89	0.98	125.98	0.96	120.79
	34	17	2	80	1.00	6.33	1.00	6.50	1.00	6.44
	35	7	5	32	1.00	43.94	0.99	46.98	0.99	45.59
	36	9	4	40	0.99	30.19	1.00	30.60	0.99	30.23
	38	19	2	80	1.00	6.44	1.00	6.45	0.99	6.37
280	12	3	4	70	0.94	28.07	0.94	28.44	0.93	28.48
	15	3	5	56	0.94	47.65	0.94	47.37	0.94	47.68
	16	2	8	35	0.89	102.47	0.89	104.25	0.90	103.83
	18	9	2	140	0.99	6.41	0.99	6.50	0.99	6.51
	22	11	2	140	0.99	6.61	0.99	6.48	1.00	6.60
	24	3	8	35	0.96	123.18	0.96	121.88	0.96	124.69
	25	5	5	56	0.98	55.48	0.98	56.05	0.99	54.97
	26	13	2	140	0.99	6.74	0.99	6.77	1.00	6.62
	30	3	10	28	0.96	189.18	0.97	182.58	0.97	184.39
	32	4	8	35	0.97	135.94	0.98	131.74	0.99	134.95
	34	17	2	140	1.00	6.75	1.00	6.88	1.00	6.84
	36	9	4	70	0.99	38.04	1.00	36.47	0.99	36.59
	38	19	2	140	1.00	6.82	1.00	6.85	1.00	6.83
	42	3	14	20	0.99	292.91	0.98	320.06	0.96	310.45
	44	11	4	70	1.00	38.00	0.99	37.56	1.00	37.28
	45	9	5	56	1.00	61.45	1.00	60.35	1.00	59.46
	46	23	2	140	1.00	7.02	1.00	6.86	1.00	6.95
	48	6	8	35	0.99	144.63	0.99	148.64	1.01	141.89
	49	7	7	40	1.00	111.72	1.00	118.32	1.00	114.09
	50	5	10	28	0.99	195.80	0.99	199.00	0.99	207.92

Table 4: Efficiency of MSSM over SRS using normal distribution.

Normal distribution																
					$\sigma = 5$						$\sigma = 10$					
N	n	m	s	k_1	$\mu = 20$		$\mu = 40$		$\mu = 60$		$\mu = 20$		$\mu = 40$		$\mu = 60$	
					Eff 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff1*	*Eff* 2	*Eff* 1	*Eff* 2
160	12	3	4	40	0.97	3.33	0.97	3.34	0.98	3.35	0.96	3.34	0.97	3.31	0.97	3.28
	14	7	2	80	0.99	1.79	1.00	1.80	1.00	1.79	1.01	1.79	0.99	1.80	1.00	1.80
	15	3	5	32	0.98	4.01	0.97	4.11	0.98	4.11	0.97	4.00	0.97	4.02	0.99	4.06
	18	9	2	80	1.00	1.78	0.99	1.78	0.99	1.80	0.99	1.79	1.00	1.78	1.00	1.79
	22	11	2	80	1.00	1.79	0.99	1.80	0.99	1.78	1.00	1.79	1.00	1.78	0.99	1.77
	24	3	8	20	0.98	6.04	0.98	6.33	0.98	6.19	0.98	6.23	1.00	6.19	0.97	6.16
	25	5	5	32	0.99	3.98	0.99	4.04	0.99	4.05	0.98	4.06	0.99	4.07	1.00	4.03
	26	13	2	80	1.00	1.79	1.00	1.79	1.00	1.77	1.00	1.78	1.00	1.79	0.99	1.79
	28	7	4	40	0.99	3.30	0.99	3.28	1.00	3.27	1.00	3.32	0.99	3.30	0.99	3.32
	30	3	10	16	0.98	7.49	0.98	7.67	0.99	7.72	0.99	7.54	1.00	7.41	0.99	7.77
	34	17	2	80	1.00	1.78	1.00	1.78	1.00	1.79	0.99	1.79	1.00	1.79	1.00	1.78
	35	7	5	32	1.00	4.00	1.00	4.03	0.99	4.03	0.99	4.04	0.99	3.99	1.01	4.01
	36	9	4	40	1.00	3.34	1.00	3.25	1.00	3.28	1.01	3.32	0.99	3.30	1.00	3.29
	38	19	2	80	0.99	1.81	1.00	1.79	1.00	1.77	1.00	1.78	1.00	1.78	1.00	1.79
280	12	3	4	70	0.96	3.34	0.97	3.31	0.96	3.33	0.97	3.33	0.97	3.35	0.97	3.33
	15	3	5	56	0.96	4.12	0.98	4.03	0.97	4.05	0.97	4.14	0.99	4.09	0.97	4.01
	16	2	8	35	0.95	6.30	0.95	6.16	0.94	6.26	0.95	6.31	0.95	6.29	0.94	6.35
	18	9	2	140	1.00	1.79	0.99	1.79	1.00	1.80	0.99	1.79	1.00	1.79	1.00	1.79
	22	11	2	140	1.00	1.79	1.00	1.78	1.00	1.80	0.99	1.79	1.00	1.80	1.00	1.80
	24	3	8	35	0.98	6.17	0.99	6.35	0.99	6.06	0.98	6.25	0.98	6.14	0.98	6.26
	25	5	5	56	0.99	4.01	1.00	4.11	1.00	4.05	1.00	4.07	0.99	4.04	0.99	4.02
	26	13	2	140	1.00	1.80	0.99	1.79	1.00	1.80	1.00	1.78	1.00	1.81	1.00	1.79
	30	3	10	28	0.98	7.53	0.98	7.73	0.99	7.72	0.99	7.98	1.00	7.78	0.99	7.71
	32	4	8	35	0.99	6.17	1.00	6.38	1.00	6.16	1.00	6.35	0.99	6.16	0.99	6.37
	34	17	2	140	1.00	1.78	1.00	1.79	0.99	1.78	1.00	1.78	1.00	1.79	1.00	1.79
	36	9	4	70	1.00	3.33	0.99	3.33	1.00	3.33	0.99	3.28	1.00	3.34	0.99	3.38
	38	19	2	140	1.00	1.78	1.00	1.78	1.00	1.79	1.00	1.79	1.00	1.79	1.00	1.80
	42	3	14	20	0.98	10.32	0.98	10.12	0.99	10.35	1.00	10.55	1.00	10.36	0.99	10.53
	44	11	4	70	1.00	3.30	1.00	3.31	1.00	3.36	1.00	3.28	0.99	3.30	1.00	3.30
	45	9	5	56	0.99	4.07	0.99	4.08	1.01	4.01	1.00	4.07	0.99	4.03	1.00	4.10
	46	23	2	140	1.00	1.78	1.00	1.79	0.99	1.79	1.00	1.79	1.00	1.79	0.99	1.78
	48	6	8	35	0.99	6.22	0.98	6.24	1.00	6.16	0.99	6.19	1.01	6.39	1.00	6.21
	49	7	7	40	1.00	5.66	1.00	5.52	1.00	5.49	0.99	5.49	0.99	5.48	1.00	5.50
	50	5	10	28	1.00	7.89	0.99	7.54	1.00	7.82	0.99	7.59	0.99	7.54	1.01	7.72

Table 5: Efficiency of MSSM over SRS using gamma distribution.

Gamma distribution																				
N	n	m	s	k_1	$shape=2$								$shape=4$							
					$scale=1$		$scale=3$		$scale=5$		$scale=10$		$scale=1$		$scale=3$		$scale=5$		$scale=10$	
					Eff 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2	*Eff* 1	*Eff* 2
160	12	3	4	40	1.00	1.50	0.98	1.48	0.99	1.45	0.98	1.42	0.98	1.77	0.99	1.75	0.97	1.77	0.99	1.74
	14	7	2	80	0.99	1.21	0.99	1.20	1.00	1.21	1.00	1.21	1.00	1.34	0.99	1.34	1.00	1.33	1.00	1.34
	15	3	5	32	1.01	1.54	0.98	1.57	0.99	1.54	1.00	1.59	0.99	1.89	1.00	1.89	0.97	1.90	0.98	1.89
	18	9	2	80	0.99	1.21	0.99	1.20	1.00	1.21	1.00	1.20	1.00	1.33	1.00	1.34	1.00	1.33	1.00	1.34
	22	11	2	80	1.01	1.20	1.01	1.21	1.00	1.21	1.00	1.20	0.99	1.32	0.99	1.33	1.00	1.32	1.00	1.33
	24	3	8	20	0.99	1.81	1.00	1.79	1.00	1.77	1.00	1.86	0.99	2.32	1.00	2.27	1.00	2.25	1.00	2.26
	25	5	5	32	1.01	1.55	0.99	1.57	1.00	1.53	0.99	1.56	0.99	1.92	1.00	1.87	0.99	1.89	1.00	1.87
	26	13	2	80	1.00	1.20	1.00	1.19	1.00	1.20	1.00	1.20	0.99	1.33	1.00	1.32	1.00	1.33	1.00	1.32
	28	7	4	40	1.00	1.45	1.00	1.44	1.00	1.46	0.99	1.44	0.99	1.76	1.00	1.72	1.00	1.73	0.99	1.72
	30	3	10	16	0.99	1.91	1.00	1.98	1.00	1.90	0.98	1.98	0.98	2.53	0.98	2.44	0.99	2.51	1.01	2.44
	34	17	2	80	1.00	1.20	1.00	1.21	1.00	1.20	1.00	1.19	1.00	1.33	1.00	1.33	1.00	1.33	1.00	1.31
	35	7	5	32	0.99	1.55	0.99	1.56	1.00	1.53	1.01	1.56	1.00	1.86	1.00	1.89	0.98	1.89	1.00	1.89
	36	9	4	40	0.99	1.43	1.00	1.45	1.00	1.47	1.01	1.45	1.00	1.75	0.99	1.76	1.00	1.76	0.99	1.76
	38	19	2	80	0.99	1.20	1.00	1.21	0.99	1.20	1.00	1.20	1.01	1.32	1.00	1.32	1.00	1.32	1.00	1.33
280	12	3	4	70	0.98	1.48	0.99	1.46	0.99	1.46	0.99	1.46	0.98	1.76	0.98	1.78	0.99	1.76	0.99	1.76
	15	3	5	56	0.99	1.57	0.98	1.58	0.99	1.56	1.00	1.57	0.98	1.92	0.98	1.95	0.99	1.89	0.99	1.94
	16	2	8	35	0.99	1.82	0.98	1.80	0.98	1.77	0.97	1.85	0.98	2.29	0.98	2.26	0.98	2.32	0.96	2.29
	18	9	2	140	1.00	1.21	1.00	1.20	1.00	1.21	1.00	1.20	1.00	1.33	1.00	1.33	1.00	1.34	1.00	1.34
	22	11	2	140	1.00	1.21	0.99	1.21	1.00	1.20	1.00	1.21	1.00	1.33	1.00	1.32	1.00	1.32	1.00	1.34
	24	3	8	35	0.98	1.84	0.99	1.79	0.99	1.82	1.00	1.82	0.99	2.31	0.98	2.27	0.99	2.29	0.99	2.24
	25	5	5	56	0.99	1.55	1.01	1.56	1.00	1.53	1.00	1.55	0.99	1.92	1.00	1.88	1.00	1.89	1.00	1.87
	26	13	2	140	1.00	1.20	0.99	1.20	1.00	1.20	1.00	1.21	1.00	1.33	1.00	1.32	1.00	1.32	1.00	1.32
	30	3	10	28	1.00	1.93	0.99	1.91	0.99	1.89	1.00	1.96	0.99	2.48	0.98	2.53	0.98	2.52	0.99	2.47
	32	4	8	35	1.00	1.81	0.99	1.79	1.00	1.80	1.00	1.80	1.00	2.24	1.00	2.27	1.01	2.28	0.98	2.26
	34	17	2	140	1.00	1.20	1.00	1.20	1.00	1.21	1.00	1.19	1.00	1.33	1.00	1.33	1.00	1.33	1.00	1.32
	36	9	4	70	0.99	1.44	1.00	1.44	1.00	1.44	0.99	1.44	0.99	1.71	1.00	1.75	1.00	1.75	1.00	1.70
	38	19	2	140	1.00	1.20	1.00	1.20	1.00	1.19	1.00	1.20	1.00	1.34	1.00	1.32	1.01	1.32	1.00	1.33
	42	3	14	20	0.97	2.13	0.98	2.19	0.99	2.19	0.98	2.13	1.00	2.81	0.98	2.93	1.00	2.83	0.99	2.80
	44	11	4	70	1.00	1.46	1.00	1.46	1.00	1.47	0.99	1.46	0.99	1.71	1.00	1.74	1.00	1.72	1.01	1.74
	45	9	5	56	1.01	1.56	1.00	1.55	1.00	1.54	1.00	1.53	0.99	1.90	1.01	1.91	0.99	1.88	0.98	1.90
	46	23	2	140	1.00	1.20	1.00	1.19	1.00	1.20	1.00	1.20	1.00	1.33	1.00	1.33	1.00	1.32	1.00	1.32
	48	6	8	35	1.01	1.79	1.01	1.82	0.98	1.76	1.01	1.78	1.00	2.27	1.00	2.30	1.00	2.29	1.00	2.26
	49	7	7	40	1.00	1.72	0.99	1.70	0.99	1.70	0.99	1.74	1.00	2.13	1.00	2.16	0.99	2.11	1.00	2.12
	50	5	10	28	1.00	1.92	0.99	1.96	1.00	1.96	0.98	1.94	1.01	2.47	0.99	2.48	1.00	2.46	1.00	2.42

6. VARIANCE OF $\hat{S}^2_{MSSM}$ IN THE PRESENCE OF LINEAR TREND

The variance of $\hat{S}^2_{MSSM}$ under the linear Model (3.1) is given by

(6.1)
$$\begin{aligned} V\left(\hat{S}^2_{MSSM}\right) = & \frac{\beta^4\left(k_1^2-1\right)}{m\left(N-1\right)^2}\frac{\left(N-s\right)^2k_1}{\left(m-1\right)} \\ & \times\left[\frac{\left(3k_1^2-7\right)}{240}\left\{\frac{\left(m-1\right)}{\left(k_1-1\right)}-\frac{\left(m-2\right)\left(m-3\right)}{\left(k_1-2\right)\left(k_1-3\right)}\right\}+\frac{1}{144}\left(k_1^2-1\right)\right. \\ & \left.\times\left\{\frac{\left(k_1-3\right)-\left(m-2\right)\left(k_1+3\right)}{\left(k_1-1\right)^2}\frac{\left(m-2\right)\left(m-3\right)\left(k_1^2-3\right)}{\left(k_1-1\right)^2\left(k_1-2\right)\left(k_1-3\right)}\right\}\right] \end{aligned}$$

(see details in Appendix B).

Substituting $m=n$, $s=1$ and $k_1=N$ in (B.7), the variance of $\hat{S}^2_{SRS}$ can be obtained in the presence of linear trend, i.e.,

$$\begin{aligned} V\left(\hat{S}^2_{SRS}\right) = & \frac{\beta^4\left(N^2-1\right)N}{n(n-1)}\left[\frac{\left(3N^2-7\right)}{240}\left\{\frac{\left(n-1\right)}{\left(N-1\right)}-\frac{\left(n-2\right)\left(n-3\right)}{\left(N-2\right)\left(N-3\right)}\right\}\right. \\ & +\frac{1}{144}\left(N^2-1\right)\left\{\frac{\left(N-3\right)-\left(n-2\right)\left(N+3\right)}{\left(N-1\right)^2}\right. \\ & \left.\left.+\frac{\left(n-2\right)\left(n-3\right)\left(N^2-3\right)}{\left(N-1\right)^2\left(N-2\right)\left(N-3\right)}\right\}\right]. \end{aligned} \tag{6.2}$$

Similarly, substituting $m=m'$, $k_1=m'k$ and $s=n/m'$ in Equation (B.7), one can get the following formula of variance of unbiased variance estimator with m' random starts for LSS in the presence of linear trend.

$$\begin{aligned} V\left(\hat{S}^2_{LSS}\right) = & \frac{\beta^4\left(m'^2k^2-1\right)}{\left(N-1\right)^2}\frac{\left(m'N-n\right)^2k}{m'^2\left(m'-1\right)} \\ & \times\left[\frac{\left(3m'^2k^2-7\right)}{240}\left\{\frac{\left(m'-1\right)}{\left(m'k-1\right)}-\frac{\left(m'-2\right)\left(m'-3\right)}{\left(m'k-2\right)\left(m'k-3\right)}\right\}\right. \\ & +\frac{1}{144}\left(m'^2k^2-1\right)\left\{\frac{\left(m'k-3\right)-\left(m'-2\right)\left(m'k+3\right)}{\left(m'k-1\right)^2}\right. \\ & \left.\left.+\frac{\left(m'-2\right)\left(m'-3\right)\left(m'^2k^2-3\right)}{\left(m'k-1\right)^2\left(m'k-2\right)\left(m'k-3\right)}\right\}\right]. \end{aligned} \tag{6.3}$$

6.1. Efficiency Comparison of $\hat{S}^2_{MSSM}$ and $\hat{S}^2_{SRS}$ in the Presence of Linear Trend

Due to complicated expressions given in Equation (B.7) and (6.2), theoretical comparison of $\hat{S}^2_{MSSM}$ and $\hat{S}^2_{SRS}$ is not easy. Therefore, a numerical comparison is carried out by considering the linear Model (3.1) and results are presented in Table 6.

Table 6: Efficiency of MSSM over SRS using linear model.

N	n	m	s	k_1	Efficiency	N	n	m	s	k_1	Efficiency
160	12	3	4	40	34.76	280	12	3	4	70	34.81
	14	7	2	80	6.72		15	3	5	56	65.24
	15	3	5	32	65.13		16	2	8	35	169.87
	18	9	2	80	6.98		18	9	2	140	6.98
	22	11	2	80	7.15		22	11	2	140	7.15
	24	3	8	20	250.30		24	3	8	35	250.80
	25	5	5	32	84.36		25	5	5	56	84.56
	26	13	2	80	7.27		26	13	2	140	7.27
	28	7	4	40	49.07		30	3	10	28	479.29
	30	3	10	16	478.57		32	4	8	35	299.77
	34	17	2	80	7.42		34	17	2	140	7.43
	35	7	5	32	94.01		36	9	4	70	52.04
	36	9	4	40	51.92		38	19	2	140	7.49
	38	19	2	80	7.48		42	3	14	20	1282.26
							44	11	4	70	53.96
							45	9	5	56	100.13
							46	23	2	140	7.57
							48	6	8	35	356.73
							49	7	7	40	252.94
							50	5	10	28	641.04

In Table 6, one can easily see that the lower the number of groups k_1, the higher is the efficiency, and vice versa. Note that different choices of α and β do not have any effect on the efficiencies as the parameters α and β will drop out from variance and efficiency expressions respectively.

7. CONCLUSION

The proposed MSSM design is based on adjusting the population units in groups. Thus, except the two extreme cases of this design, MSSM is neither completely systematic nor random but displaying the amalgamation of systematic and simple random sampling. In the two extreme cases, one of them becomes LSS and other SRS. The MSSM makes it possible to develop the modified expressions of all the results that relates to the LSS. A few such modifications are reported in Sections 2 and 3. A theoretical efficiency comparison of MSSM and SRS using the variances of mean in the presence of linear trend is carried out and is shown in Equation (3.1). This comparison clearly indicates that MSSM is more efficient than SRS.

In this study, population variance is unbiasedly estimated in MSSM for all possible combinations of N and n. An explicit expression for variance of unbiased variance estimator is also obtained in the proposed design. Moreover, it enables us to deduce the expressions for variance of unbiased variance estimator for LSS and SRS. Due to the complex nature of these expressions, theoretical comparison is not an easy task. Therefore, numerical comparison of MSSM and SRS is carried out in Sections 5 and 6. This numerical efficiency comparison is done for natural population, simulated population and linear model having a perfect linear trend. The results show that if populations (with linear or parabolic trend) are arranged with certain characteristics then MSSM is more efficient than SRS. However, in simulated populations, MSSM is almost equally efficient to SRS as units are not arranged in specific order. In this case, one can benefit from MSSM due to its simplicity and economical status. Furthermore, the findings reveal that the efficiency of MSSM is quite high for those combinations of N and n in which all population units are arranged in minimum number of groups.

APPENDIX A — Variance of $\hat{S}^2_{MSSM}$

The variance of $\hat{S}^2_{MSSM}$ can be written as

(A.1)
$$V\left(\hat{S}^2_{MSSM}\right) = \frac{1}{(N-1)^2}\Bigg[\Big(\frac{N}{m}\Big)^2 V\Big(\sum_{u=1}^{m}\hat{\sigma}^2_{r_u}\Big) + \Big(\frac{(N-s)}{m(m-1)}\Big)^2 V\Big(\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}(\bar{y}_{r_u}-\bar{y}_{r_v})^2\Big) + 2\,\frac{N}{m}\,\frac{(N-s)}{m(m-1)}\; Cov\Big(\sum_{u=1}^{m}\hat{\sigma}^2_{r_u}, \sum_{u=1}^{m-1}\sum_{v=u+1}^{m}(\bar{y}_{r_u}-\bar{y}_{r_v})^2\Big)\Bigg].$$

Note that

$$V\Big(\sum_{u=1}^{m}\hat{\sigma}^2_{r_u}\Big) = \sum_{u=1}^{m} V\left(\hat{\sigma}^2_{r_u}\right) + \sum_{u=1}^{m}\sum_{\substack{v=1\\ v\neq u}}^{m} Cov\left(\hat{\sigma}^2_{r_u}, \hat{\sigma}^2_{r_v}\right),$$

where

$$V(\hat{\sigma}^2_{r_u}) = \frac{1}{k_1}\sum_{u=1}^{k_1}\left(\hat{\sigma}^2_{r_u}-\bar{\sigma}^2\right)^2 = \frac{1}{k_1}\sum_{r=1}^{k_1}\left(\hat{\sigma}^2_{r}-\bar{\sigma}^2\right)^2 = \sigma_0^2 \qquad \text{(say)}$$

such that

$$\bar{\sigma}^2 = \frac{1}{k_1}\sum_{u=1}^{k_1}\hat{\sigma}^2_{r_u} = \frac{1}{k_1}\sum_{r=1}^{k_1}\hat{\sigma}^2_{r} \qquad \text{and} \qquad Cov\left(\hat{\sigma}^2_{r_u}, \hat{\sigma}^2_{r_v}\right) = -\frac{\sigma_0^2}{(k_1-1)}.$$

Thus

(A.2)
$$V\Big(\sum_{u=1}^{m}\hat{\sigma}^2_{r_u}\Big) = m\sigma_0^2\left(\frac{k_1-m}{k_1-1}\right).$$

Now consider

(A.3)
$$V\Big[\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}(\bar{y}_{r_u}-\bar{y}_{r_v})^2\Big] = \sum_{u=1}^{m-1}\sum_{v=u+1}^{m} V\left\{(\bar{y}_{r_u}-\bar{y}_{r_v})^2\right\} + 2\Bigg[\sum_{u=1}^{m}\sum_{\substack{v=1\\ v\neq u}}^{m}\sum_{\substack{u'=1\\ u'\neq u,v}}^{m} Cov\Big\{(\bar{y}_{r_u}-\bar{y}_{r_v})^2, (\bar{y}_{r_u}-\bar{y}_{r_{u'}})^2\Big\} + \sum_{u=1}^{m}\sum_{\substack{v=1\\ v\neq u}}^{m}\sum_{\substack{u'=1\\ u'\neq u,v}}^{m}\sum_{\substack{v'=1\\ v'\neq u,v,u'}}^{m} Cov\Big\{(\bar{y}_{r_u}-\bar{y}_{r_v})^2, (\bar{y}_{r_{u'}}-\bar{y}_{r_{v'}})^2\Big\}\Bigg],$$

where

$$\text{(A.4)} \qquad V\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2 = \frac{2k_1}{(k_1-1)}\left\{\mu_4+\frac{k_1-3}{(k_1-1)}\mu_2^2\right\},$$

such that

$$\mu_2 = \frac{1}{k_1}\sum_{u=1}^{k_1}\left(\bar{y}_{r_u}-\mu\right)^2 = \frac{1}{k_1}\sum_{r=1}^{k_1}\left(\bar{y}_r-\mu\right)^2 \quad \text{and} \quad \mu_4 = \frac{1}{k_1}\sum_{r=1}^{k_1}\left(\bar{y}_r-\mu\right)^4.$$

$$\text{(A.5)} \qquad Cov\left\{\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2, \left(\bar{y}_{r_u}-\bar{y}_{r_{u'}}\right)^2\right\} = \frac{k_1}{(k_1-1)}\left[\mu_4-\frac{k_1+3}{(k_1-1)}\mu_2^2\right].$$

(A.6)

$$Cov\left\{\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2, \left(\bar{y}_{r_{u'}}-\bar{y}_{r_{v'}}\right)^2\right\} = \frac{-4k_1}{(k_1-2)(k_1-3)}\left[\mu_4-\frac{\left(k_1^2-3\right)}{(k_1-1)^2}\mu_2^2\right].$$

Putting (A.4), (A.5) and (A.6) in (A.3), we have

$$\begin{aligned} V\left[\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2\right] = &\binom{m}{2}\left[\frac{2k_1}{(k_1-1)}\left\{\mu_4+\frac{k_1-3}{(k_1-1)}\mu_2^2\right\}\right] \\ &+ 2\left[m\binom{m-1}{2}\left\{\frac{k_1}{(k_1-1)}\left(\mu_4-\frac{k_1+3}{(k_1-1)}\mu_2^2\right)\right\}\right. \\ &+ \left\{\binom{\frac{m(m-1)}{2}}{2} - m\binom{m-1}{2}\right\} \\ &\left.\times\left\{\frac{-4k_1}{(k_1-2)(k_1-3)}\left(\mu_4-\frac{\left(k_1^2-3\right)}{(k_1-1)^2}\mu_2^2\right)\right\}\right], \end{aligned}$$

or

(A.7)

$$\begin{aligned} V\left[\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2\right] = \; & m(m-1)k_1\left[\left\{\frac{(m-1)}{(k_1-1)}-\frac{(m-2)(m-3)}{(k_1-2)(k_1-3)}\right\}\mu_4\right. \\ &+\left\{\frac{(k_1-3)-(m-2)(k_1+3)}{(k_1-1)^2}\right. \\ &\left.\left.+\frac{(m-2)(m-3)\left(k_1^2-3\right)}{(k_1-1)^2(k_1-2)(k_1-3)}\right\}\mu_2^2\right]. \end{aligned}$$

Also consider

$$\begin{aligned} Cov&\left\{\sum_{u=1}^{m}\hat{\sigma}_{r_u}^2, \sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2\right\} = \\ &= E\left\{\sum_{u=1}^{m}\hat{\sigma}_{r_u}^2\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2\right\} \\ &\quad - E\left(\sum_{u=1}^{m}\hat{\sigma}_{r_u}^2\right)E\left\{\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2\right\}, \end{aligned}$$

where

$$E\left(\sum_{u=1}^{m}\hat{\sigma}_{r_u}^2\right) = \sum_{u=1}^{m} E\left(\hat{\sigma}_{r_u}^2\right) = m\frac{1}{k_1}\sum_{u=1}^{k_1}\hat{\sigma}_{r_u}^2 = m\frac{1}{k_1}\sum_{r=1}^{k_1}\hat{\sigma}_{r}^2 = m\bar{\sigma}^2,$$

$$E\left\{\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2\right\} = \sum_{u=1}^{m-1}\sum_{v=u+1}^{m} E\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2 = \binom{m}{2}\frac{2k_1}{(k_1-1)}\,\mu_2$$

and

$$E\left\{\sum_{u=1}^{m}\hat{\sigma}_{r_u}^2\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2\right\} = \frac{m\,(m-1)}{(k_1-1)}\left[\left\{1+\frac{(m-2)\,(k_1-1)}{(k_1-2)}\right\}k_1\bar{\sigma}^2\mu_2 + \left\{1-\frac{(m-2)}{(k_1-2)}\right\}\sum_{r=1}^{k_1}\hat{\sigma}_r^2\left(\bar{y}_r-\mu\right)^2\right],$$

or

(A.8)

$$Cov\left\{\sum_{u=1}^{m}\hat{\sigma}_{r_u}^2,\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left(\bar{y}_{r_u}-\bar{y}_{r_v}\right)^2\right\} = \\ = \frac{m\,(m-1)\,(k_1-m)}{(k_1-1)\,(k_1-2)}\left\{\sum_{r=1}^{k_1}\hat{\sigma}_r^2\left(\bar{y}_r-\mu\right)^2 - k_1\bar{\sigma}^2\mu_2\right\}.$$

Putting (A.1), (A.7) and (A.8) in (A.1) and then simplifying, we have
(A.9)

$$V\left(\hat{S}_{MSSM}^2\right) = \frac{1}{m\,(N-1)^2}\left[\frac{N^2\,(k_1-m)}{(k_1-1)}\,\sigma_0^2 + \frac{(N-s)^2\,k_1}{(m-1)} \times\left[\left\{\frac{(m-1)}{(k_1-1)}-\frac{(m-2)\,(m-3)}{(k_1-2)\,(k_1-3)}\right\}\mu_4 + \left\{\frac{(k_1-3)-(m-2)\,(k_1+3)}{(k_1-1)^2}+\frac{(m-2)\,(m-3)\left(k_1^2-3\right)}{(k_1-1)^2\,(k_1-2)\,(k_1-3)}\right\}\mu_2^2\right] + 2\,\frac{N\,(N-s)\,(k_1-m)}{(k_1-1)\,(k_1-2)}\left\{\sum_{r=1}^{k_1}\hat{\sigma}_r^2\left(\bar{y}_r-\bar{Y}\right)^2 - k_1\bar{\sigma}^2\mu_2\right\}\right].$$

APPENDIX B — Variance of $\hat{S}^2_{MSSM}$

Assuming the linear Model (3.1), the mean of the r^{th} $(r = 1, 2, ..., k_1)$ group can be written as

$$\bar{y}_r = \frac{1}{s}\sum_{i=1}^{s}\Big\{\alpha + \beta\big(r + (i-1)k_1\big)\Big\},$$

$$\bar{y}_r = \alpha + \beta\Big(r + \frac{1}{2}(s-1)k_1\Big), \tag{B.1}$$

$$\begin{aligned}\hat{\sigma}_r^2 &= \frac{1}{s}\sum_{i=1}^{s}\left\{\alpha + \beta\big(r + (i-1)k_1\big) - \alpha - \beta\Big(r + \frac{1}{2}(s-1)k_1\Big)\right\}^2 \\ &= \frac{1}{s}\sum_{i=1}^{s}\left\{\beta(i-1)k_1 - \beta\Big(\frac{1}{2}(s-1)k_1\Big)\right\}^2 \\ &= \frac{1}{12}\beta^2 k_1^2 (s^2-1),\end{aligned} \tag{B.2}$$

$$\bar{\sigma}_r^2 = \frac{1}{12}\beta^2 k_1^2 (s^2-1), \tag{B.3}$$

$$\sigma_0^2 = 0, \tag{B.4}$$

$$\mu_2 = \frac{1}{k_1}\sum_{r=1}^{k_1}(\bar{y}_r - \mu)^2 = \frac{\beta^2}{12}\left(k_1^2 - 1\right) \tag{B.5}$$

and

$$\mu_4 = \frac{1}{k_1}\sum_{r=1}^{k_1}(\bar{y}_r - \mu)^4 = \beta^4\left(\frac{k_1^4}{80} - \frac{k_1^2}{24} + \frac{7}{240}\right), \tag{B.6}$$

where

$$\mu = \alpha + \beta\frac{N+1}{2}.$$

Putting Equations (B.1)–(B.6) in (A.9), we have

(B.7)

$$\begin{aligned}V\left(\hat{S}^2_{MSSM}\right) &= \frac{\beta^4\left(k_1^2-1\right)}{m\,(N-1)^2}\,\frac{(N-s)^2\,k_1}{(m-1)} \\ &\quad\times\left[\frac{\left(3k_1^2-7\right)}{240}\left\{\frac{(m-1)}{(k_1-1)} - \frac{(m-2)\,(m-3)}{(k_1-2)\,(k_1-3)}\right\} + \frac{1}{144}\left(k_1^2-1\right)\right. \\ &\quad\left.\times\left\{\frac{(k_1-3)-(m-2)\,(k_1+3)}{(k_1-1)^2}\,\frac{(m-2)\,(m-3)\left(k_1^2-3\right)}{(k_1-1)^2\,(k_1-2)\,(k_1-3)}\right\}\right].\end{aligned}$$

ACKNOWLEDGMENTS

The authors offer their sincere thanks to the two reviewers for their careful reading of the paper and their helpful suggestions.

REFERENCES

[1] Chang, H.J. and Huang, K.C. (2000). Remainder linear systematic sampling, *Sankhya*, **62**(B), 249–256.

[2] Gautschi, W. (1957). Some remarks on systematic sampling, *The Annals of Mathematical Statistics*, **28**(2), 385–394.

[3] Khan, Z.; Shabbir, J. and Gupta, S.N. (2013). A new sampling design for systematic sampling, *Communications in Statistics — Theory and Methods*, **42**(18), 3359–3370.

[4] Murthy, M.N. (1967). *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, India.

[5] Naidoo, L.R.; North, D.; Zewotir, T. and Arnab, R. (2016). Multiple-start balanced modified systematic sampling in the presence of linear trend, *Communications in Statistics — Theory and Methods*, **45**(14), 4307–4324.

[6] Sampath, S. (2009). Finite population variance estimation under lss with multiple random starts, *Communication in Statistics — Theory and Methods*, **38**, 3596–3607.

[7] Sampath, S. and Ammani, S. (2012). Finite-population variance estimation under systematic sampling schemes with multiple random starts, *Journal of Statistical Computation and Simulation*, **82**(8), 1207–1221.

[8] Sethi, V.K. (1965). On optimum paring of units, *Sankhya*, **27**(B), 315–320.

[9] Singh, D.; Jindal, K.K. and Garg, J.N. (1968). On modified systematic sampling, *Biometrika*, **55**, 541–546.

[10] Subramani, J. and Gupta, S.N. (2014). Generalized modified linear systematic sampling scheme, *Hacettepe Journal of Mathematics and Statistics*, **43**(3), 529–542.