# An Assessment of Modified Systematic Sampling Designs in the Presence of Linear Trend


by

**Llewellyn Reeve**

**Naidoo**


Submitted in fulfilment of the academic requirements for the degree of Doctor of Philosophy in the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban

March 9, 2017

As the candidate's supervisor I have/have not approved this thesis/dissertation for submission.

Signed: ________________ Name: ___________________________

Date: ________________

**Abstract**

Sampling is used to estimate population parameters, as it is usually impossible to study a whole population, due to time and budget restrictions. There are various sampling designs to address this issue and this thesis is related with a particular probability sampling design, known as systematic sampling.

Systematic sampling is operationally convenient and efficient and hence is used extensively in most practical situations. The shortcomings associated with systematic sampling include: (*i*) it is impossible to obtain an unbiased estimate of the sampling variance when conducting systematic sampling with a single random start; (*iii*) if the population size is not a multiple of the sample size, then conducting conventional systematic sampling, also known as linear systematic sampling, may result in variable sample sizes. In this thesis, I would like to provide some contribution to the current body of knowledge, by proposing modifications to the systematic sampling design, so as to address these shortcomings.

Firstly, a discussion on the measures used to compare the various probability sampling designs is provided, before reviewing the general theory of systematic sampling. The performance of systematic sampling is dependent on the population structure. Hence, this thesis concentrates on a specific and common population structure, namely, linear trend. A discussion on the performance of linear systematic sampling and all relative modifications, including a new proposed modification, is then presented under the assumption of linear trend among the population units. For each of the above-mentioned problems, a brief review of all the associated sampling designs from existing literature, along with my proposed modified design, will then be explored. Thereafter, I will introduce a modified sampling design that addresses the above-mentioned problems in tandem, before providing a comprehensive report on the thesis. The aim of this thesis is to provide solutions to the above-mentioned disadvantages, by proposing modified systematic sampling designs and/or estimators that are favourable over its existing literature counterparts.

*Keywords*: systematic sampling; super-population model; Horvitz-Thompson estimator; Yates' end corrections method; balanced modified systematic sampling; multiple-start balanced modified systematic sampling; remainder modified systematic sampling; balanced centered random sampling.

# Contents

# List of acronyms

| | |
|---|---|
| BCRS | balanced centered random sampling |
| BMSS | balanced modified systematic sampling |
| BMSSEC | balanced modified systematic sampling with end corrections |
| BRS | balanced random sampling |
| BSS | balanced systematic sampling |
| CESS | centered systematic sampling |
| CSS | circular systematic sampling |
| DSS | diagonal systematic sampling |
| DSSEC | diagonal systematic sampling with end corrections |
| ICC | intra-class correlation coefficient |
| LSS | linear systematic sampling |
| MBCSS | modified balanced circular systematic sampling |
| MBCSSEC | modified balanced circular systematic sampling with end corrections |
| MBMSS | multiple-start balanced modified systematic sampling |
| MBMSSEC | multiple-start balanced modified systematic sampling with end corrections |
| MBSS | multiple-start balanced systematic sampling |
| MCCSS | modified centered circular systematic sampling |
| MCCSSEC | modified centered circular systematic sampling with end corrections |
| MLSS | multiple-start linear systematic sampling |
| MLSSEC | multiple-start linear systematic sampling with end corrections |
| MMSS | multiple-start modified systematic sampling |
| MSE | mean square error |
| MSS | modified systematic sampling |
| MYEC | multiple-start Yates' end corrections |
| RLSS | remainder linear systematic sampling |
| RLSSEC | remainder linear systematic sampling with end corrections |
| RMSS | remainder modified systematic sampling |
| RMSSEC | remainder modified systematic sampling with end corrections |
| SRS | simple random sampling without replacement |
| STR | stratified random sampling |

# List of Tables

# Preface and Declarations

The work described by this thesis was carried out in the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, from **August 2013** to **November 2016**, under the supervision of **Professor Delia North** and the co-supervision of **Professor Temesgen Zewotir** and **Professor Raghunath Arnab** .

This thesis is entirely, unless specifically contradicted in the text, the work of the candidate, Llewellyn Reeve Naidoo, and has not been previously submitted, in whole or in part, to any other tertiary institution. Where use has been made of the work of others, it is duly acknowledged in the text.

## Declaration - Plagiarism

I, ______________________________ declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been rewritten but the general information attributed to them has been referenced.
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This thesis does not contain, text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed: ____________________

## Declaration - Publications

Details of publications that form part and/or include research presented in this thesis.

Publication 1.
Naidoo, L. R., North, D., Zewotir, T. & Arnab, R. (2015), 'Balanced modified systematic sampling in the presence of linear trend', *South African Statistical Journal* **49**, 187-203.
Author Contributions: All the work was carried out by L. R. Naidoo, while all the co-authors have proof-read the article.

Publication 2.
Naidoo, L. R., North, D., Zewotir, T. & 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.
Author Contributions: All the work was carried out by L. R. Naidoo, while all the co-authors have proof-read the article.

Publication 3.
Naidoo, L. R., North, D., Zewotir, T. & Arnab, R., 'Remainder modified systematic sampling in the presence of linear trend', *Communications in Statistics - Theory and Methods* (Revised Submission).
Author Contributions: All the work was carried out by L. R. Naidoo, while all the co-authors have proof-read the article.

Publication 4.
Naidoo, L. R., North, D., Zewotir, T. & Arnab, R., 'Centric random sampling in the presence of linear trend', (Prepared for Submission).
Author Contributions: All the work was carried out by L. R. Naidoo. Prof. R. Arnab has assisted with the simplification of a variance expression.

Signed: ____________________

# Acknowledgements

I would like to express my sincere appreciation to the NRF (National Research Foundation) for the financial support I received during the preparation of this thesis. I am also thankful for the moral support that I received from my parents, Marge Naidoo and Tony Naidoo. Furthermore, I would like to extend my gratitude to Prof. R. Arnab and Prof. T. Zewotir for their assistance and guidance. Finally, I'd like to thank my supervisor for her patience, support and particularly for advising me to pursue this research topic.

# Chapter 1

# INTRODUCTION

## 1.1 Overview of Sampling

Statistics entails the collection, organization, analysis, interpretation, explanation and presentation of data. Governments, clients, medical companies, institutions and organizations, frequently use statistics to effect decision-making, e.g. choosing between different options, executing new policies, assessing current policy situations, etc.

When a problem is identified and/or presented to a statistician, he/she must then create a workable objective before planning the research approach. Collection of suitable data is then proceeded by the relevant analysis of the corresponding data and finally the results are presented thereon. These phases of the research cycle are related and interconnected. Failure to impose a solid research design leads to inefficient data collection methods, which in turn contributes to inaccurate data analysis and finally results in a flawed research report. Clearly each stage of this research cycle is crucial. We now take a more focused look at the data collection stage.

There are three basic types of statistical studies, namely, surveys, experiments and observational studies. Each of these are associated with different data collection methods, e.g. questionnaires, case studies, behaviour observation checklists, performance tests, etc. More often than not, collecting data from the whole population is difficult, due to time and money restrictions, as well as the regular problem of not being able to reach the whole population at any given point in time. As a result, we commonly opt to select a sample/subset from the population. Generally, the population and sample sizes are respectively denoted as $N$ and $n$, where $N > n$. Generalizations about a population, based on results from the sample, are made by ensuring that the characteristics of the sample accurately mirrors the corresponding characteristics of the population, i.e. we need to select a representative sample. Two fundamental conditions for selecting a representative sample are that (*i*) the sample must be sufficiently large, such that all aspects of the population are captured, and (*ii*) should be selected in a way that minimizes bias, which is given by definition as the misrepresentation of sample characteristics from the relative population characteristics. Condition (*i*) is commonly related to a trade-off between sampling error and cost, i.e. larger sample sizes are subject to greater costs, while more often than not reducing the associated sampling error. Condition (*ii*) is related to the method of selecting a sample from the population.

A specific sampling design is generally implemented to select a sample which produces an estimate of the population parameter. This estimator is also known as a sample statistic. A single numerical value, used to estimate a specific population parameter, is computed from the sample and is defined as a point estimator.

A point estimator is said to be *unbiased* if the expected value of this estimator is equivalent to the population parameter being estimated, otherwise it is known to be *biased*. When referring to a biased point estimator, the degree of bias is calculated as the difference between the expectation of the point estimator and the real value of the population parameter. Accuracy is related to bias, where unbiased estimators are generally expected to be equivalent to the corresponding population parameter and are therefore regarded as perfectly accurate estimates of the associated population parameter.

A point estimator differs from sample to sample and is thus a random variable which has a distribution. The variance of this estimator, which is a measure of precision, is known as the sampling variance and reflects the extent by which the point estimator differs from sample to sample. There exists many estimators of a specific population parameter and the estimator which is related to the smallest sampling variance, is known as the most precise estimator. Therefore, the most preferred point estimator(s) will be unbiased and display minimum variance. Under these circumstances, we obtain optimum sampling results, with respect to obtaining maximum information about the population parameter.

Note that a biased estimator may offer more information about a population parameter than that of an unbiased estimator, since the biased estimator, which corresponds to a low degree of bias, may exhibit a much higher degree of precision than the unbiased estimator. There is thus a trade-off between accuracy and precision and the appropriate measure that represents this trade-off is the mean square error (MSE) of a point estimator. The

MSE of a point estimator is calculated as the sum of the corresponding sampling variance and the squared bias of the associated estimator, whereby the estimator which exhibits a minimum MSE is considered to be the most desirable. Note that the MSE of an *unbiased* estimator is equivalent to the variance of that estimator. Additionally, the relative MSE between two point estimators is given as the ratio of their MSEs (expressed in percentage terms), i.e. the relative MSE of point estimator $a$, with respect to point estimator $b$, is given as the MSE of estimator $b$ divided by the MSE of estimator $a$ and then multiplied by 100%. If this percentage is less than 100%, then we conclude that estimator $b$ is to be preferred over estimator $a$, while a percentage that is greater than 100% would suggest that estimator $a$ is to be preferred over estimator $b$.

When aiming to provide an estimate of a population parameter, it is thus of the utmost importance that one explores the ramifications of the different sampling designs on the simplicity of implementation, degree of bias, variance and MSE of the corresponding point estimators, in addition to their capability to generate an unbiased estimate of the associated sampling variance. We next focus our attention on some well-known probability sampling designs which can be implemented to estimate population parameters.

There are several probability sampling designs that can be employed to select a representative sample, e.g. simple random sampling, stratified random sampling (STR), cluster sampling, systematic sampling etc. Simple random sampling entails a random selection of each sampling unit from the population, where the probability of selection of each possible unit is equivalent within each of the $n$ phases of the random selection. If simple random sampling with replacement is to be achieved, then each unit is replaced into the population after being selected for the sample and is thus eligible for each of the following phases of the random selection. We therefore have a possibility of duplicate sampling units, since the units which are randomly selected for the sample are replaced into the population, thus having a possibility of being chosen again in the next phases of the random selections. If simple random sampling without replacement (SRS) is to be achieved, then we randomly select each sampling unit, but now we do not replace these units into the population before the next phases of the random selections, thus ensuring a sample of distinct sampling units. Hence, to avoid duplicate sampling units and improve results, preference will be given to SRS over simple random sampling with replacement (Lohr 2010). Stratified sampling entails dividing the whole population into subgroups (or strata) based on some characteristic, before applying a specific random selection within each subgroup (or stratum), such that the selected units for each of the strata collectively represent the stratified sample. Note that if SRS is applied within each stratum, then this sampling design is referred to as STR and the selected units for each of the strata collectively represent a stratified random sample. Cluster sampling involves dividing the whole population into groups (or clusters), before randomly selecting entire clusters, such that the units within each of the randomly selected clusters collectively represent a cluster sample. Systematic sampling entails a random selection of a unit from the population and subsequent units at equally spaced intervals thereafter, such that the selected units collectively represent a systematic sample.

We next focus our study on systematic sampling, while making comparisons to the other above-mentioned probability sampling designs.

## 1.2 Systematic Sampling

A detailed discussion on systematic sampling was originally given by Madow & Madow (1944), Cochran (1946) and Yates (1948). Systematic sampling is frequently applied in forestry, land use/cover area frame surveys, census, record sampling and for household and establishment surveys (Murthy & Rao 1988). Applications on systematic sampling for forestry are given by Hasel (1938), Finney (1948) and Zinger (1964), while applications on systematic sampling for land use/cover area frames are provided by Osborne (1942), Dunn & Harrison (1993) and D'Orazio (2003). Some examples of systematic sampling are given in the areas of soil sampling (Mason 1994, Jacobsen 1998) and nature studies (McArthur 1987, Pawley 2006). Comprehensive reviews on systematic sampling are provided by Murthy (1967), Cochran (1977), Iachan (1982), Bellhouse (1988) and Murthy & Rao (1988).

The fundamental process of systematic sampling is given as follows: To select sample of size $n$ from a population of size $N$ using systematic sampling, we randomly select a unit from the first $k = N/n$ population units and every subsequent $k$th unit, until the required sample size is achieved. This design is referred to as linear systematic sampling (LSS), provided that the sampling interval $k$ is an integer (Cochran 1977). The random start is given by $i$, where $i \in \{1, ...., k\}$. LSS is advantageous over SRS and STR, owing to its convenience and operational simplicity when implemented.

Let us consider a finite population $U = (U_1, ..., U_N)$ of size $N$ and let $y_q$ be the value of the study variable of the $q$th unit of population $U$, for $q \in \{1, ..., N\}$. All possible values of the random start $i$, as well as the corresponding sample outcomes and sample means, when conducting LSS, are presented in Table 1.1. Note that $y_{ij}$ represents the value of the study variable related to the $j$th unit of the $i$th linear systematic sample, i.e. $y_{ij} = y_{i+(j-1)k}$, for $i \in \{1, .., k\}$ and $j \in \{1, ..., n\}$. From Table 1.1, we see that the sample is automatically determined by the choice of the first sampling unit/random start. Furthermore, the whole population of size $N$ is divided into $k$ groups, each of size $n$. The methodology of LSS is therefore equivalent to the random selection of one of these $k$ groups. LSS is thus considered as special case of cluster sampling, since each individual sample is regarded as a cluster and one cluster is then randomly selected.

Table 1.1: Samples and sample means for possible values of $i$ using LSS

| Possible values of $i$ | Sample | Sample Mean ($\overline{y}_{LSS}$) |
|---|---|---|
| $i = 1$ | $S_1$=$\{U_1,\ U_{1+k},\ U_{1+2k},\ ...,\ U_{1+(n-1)k}\}$ | $\sum_{j=1}^{n} y_{1,1+(j-1)k}/n$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $i = h$ | $S_h$=$\{U_h,\ U_{h+k},\ U_{h+2k},\ ...,\ U_{h+(n-1)k}\}$ | $\sum_{j=1}^{n} y_{h,h+(j-1)k}/n$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $i = k$ | $S_k$=$\{U_k,\ U_{2k},\ ...,\ U_{nk}\}$ | $\sum_{j=1}^{n} y_{k,jk}/n$ |

From Table 1.1, we can easily verify that the linear systematic sample mean, denoted as $\overline{y}_{LSS}$, is an unbiased estimate of the population mean $\overline{Y} = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}/nk = \mathrm{E}(\overline{y}_{LSS})$. The corresponding sampling variance can be written in terms of the intra-class correlation coefficient (ICC). The ICC between pairs of sampling units that are located within the same linear systematic sample is denoted as

$$\rho = \mathrm{Cov}(y_{ij}, y_{il})/\sigma^2, \qquad j, l = 1, ..., n, (j \neq l) \text{ and } i = 1, ..., k, \tag{1.1}$$

where

$$\mathrm{Cov}(y_{ij}, y_{il}) = \frac{1}{nk(n-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l \neq j}}^{n} (y_{ij} - \overline{Y})(y_{il} - \overline{Y}); \tag{1.2}$$

$y_{ij}$ and $y_{il}$ denote random variables which represent two different units in the $i$th linear systematic sample; and the population variance is given by definition as

$$\sigma^2 \triangleq \frac{1}{N} \sum_{q=1}^{N} \left(y_q - \overline{Y}\right)^2 = \frac{1}{nk} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \overline{Y})^2.$$

Applying the above notation, the sampling variance is expressed as

$$\mathrm{V}(\overline{y}_{LSS}) = \frac{S_Y^2}{n}\left(\frac{N-1}{N}\right)\Big[1+(n-1)\rho\Big], \tag{1.3}$$

where

$$S_Y^2 \triangleq \frac{1}{N-1}\sum_{q=1}^{N}\left(y_q-\overline{Y}\right)^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij}-\overline{Y}\right)^2 \tag{1.4}$$

denotes the adjusted population variance (Cochran 1977). When referring to (1.3), we note that the sampling variance is dependent on $n$ and $\rho$, as $S_Y^2$ and $N$ are fixed. We further note that it is not guaranteed that a larger sample size will result in a smaller sampling variance, unlike SRS and STR which exhibit an inversely proportional relationship between the sample size and the associated sampling variance. Empirical results provided by Madow (1946) indicate an inconsistent behaviour of the sampling variance in relation to the sample size, when conducting LSS. Thus, the only component which is proportionately related to the sampling variance is $\rho$, which is dependent on (*i*) the arrangement of the population units, (*ii*) the degree of correlation between consecutive population units and (*iii*) the sample size $n$ (Murthy & Rao 1988).

When conducting SRS, the sample mean, denoted as $\overline{y}_{SRS}$, provides an unbiased estimate of $\overline{Y}$ with an associated sampling variance written as

$$\mathrm{V}(\overline{y}_{SRS}) = \frac{S_Y^2}{n}\left(\frac{N-n}{N}\right) \tag{1.5}$$

(see Cochran (1977)). Comparing (1.3) to (1.5), we note that LSS is more efficient than SRS when $\rho < -1/(N-1)$. For large population sizes, we conclude through approximation that LSS is more efficient than SRS, if and only if $\rho < 0$. By referring to (1.1) and (1.2), this implies that the more heterogeneous the units which are located within the same linear systematic sample, the more substantial the efficiency gains when favouring LSS over SRS.

Consider STR, where each stratum is of size $k$ and one unit is selected within each stratum. The sample mean, denoted as $\overline{y}_{STR}$, provides an unbiased estimate of $\overline{Y}$ with the respective sampling variance written as

$$\mathrm{V}(\overline{y}_{STR}) = \frac{S_{wst}^2}{n}\left(\frac{N-n}{N}\right), \tag{1.6}$$

where the variance among units which are located within the same stratum is denoted as $S_{wst}^2 = \sum_{j=1}^{n}\sum_{i=1}^{k}\left(y_{ij}-\overline{y}_{\cdot j}\right)^2/n(k-1)$; the value of the study variable related to the $i$th unit of the $j$th stratum is denoted as $y_{ij}$; and the $j$th stratum mean is denoted as $\overline{y}_{\cdot j} = \sum_{i=1}^{k} y_{ij}/k$, for $j \in \{1, ..., n\}$ (Cochran 1977). Now, LSS can be defined as the division of the $N$ population units into $n$ strata of $k$ units each, before randomly selecting one unit from the first stratum and the unit selected from each of the other strata, is located in the same position as the randomly selected unit in the first strata. The usual LSS design, which was presented by Table 1.1, is therefore transposed and can now be compared to the above-mentioned STR design. Accordingly, the first $k$ population units are situated in the first stratum, the next $k$ population units are situated to the second stratum, and so forth. The corresponding variance expression is given by

$$\mathrm{V}\left(\overline{y}_{LSS}\right) = \frac{S_{wst}^2}{n}\left(\frac{N-n}{N}\right)\left[1 + (n-1)\rho_{wst}\right], \tag{1.7}$$

where the ICC between pairs of units which are located within the same linear systematic sample, with the deviations being calculated from their associated stratum means, is denoted as $\rho_{wst} = 2\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{l>j}^{n}\left(y_{ij} - \overline{y}_{\cdot j}\right)\left(y_{il} - \overline{y}_{\cdot l}\right)/\left[n\left(n-1\right)\left(k-1\right)S_{wst}^2\right]$ (Cochran 1977). Comparing (1.6) to (1.7), we note that LSS is more efficient than STR, if and only if $\rho_{wst} < 0$.

Thus, each different population structure will have an effect on $\rho$ and $\rho_{wst}$, which will then effect the efficiency of LSS. Consequently, for the remainder of this thesis, we narrow the study to focus on a specific type of population structure, i.e. populations exhibiting linear trend.

## 1.3 Disadvantages of Systematic Sampling

Now that we have a basic understanding of systematic sampling, we next consider the key disadvantages, which are given as follows:

(i) An unbiased estimate of the sampling variance is unobtainable when conducting LSS with a single random start, since it is impossible to select certain pairs of population units for the sample. This disadvantage will be further explained in Chapter 3.

(ii) If the population size is not a multiple of the sample size, then conducting LSS will either produce sample sizes that vary, or fixed sample sizes that are greater than required. Consequently, biased estimates of the population parameters are associated with the former situation, while the latter situation is undesirable since sample sizes are commonly fixed beforehand. These situations will be explored in Chapter 4.

The fundamental objectives of this thesis are to provide solutions to these disadvantages by proposing modified systematic sampling designs and/or estimators that are favourable over its existing literature counterparts. The theory presented thus far is a reflection of the authors' understanding of the existing literature, which can be found in Naidoo (2013), as well as any traditional sample survey book, e.g. Kalton (1983), Lehtonen & Pahkinen (2004), Lohr (2010), etc.

## 1.4 Scope of Thesis

This thesis is divided into six chapters. Chapter 2 entails a discussion of the performance of systematic sampling under the assumption of linear trend among the population units. All associated modified systematic sampling designs are explored, before providing a discussion on optimality. Thereafter, a proposed modified systematic sampling design is presented. Chapter 3 tackles the first disadvantage, i.e. estimation of the sampling variance. An overview of the problem at hand is presented, before exploring all relative modified systematic sampling designs. In the final section of this chapter, a modified systematic sampling design is suggested to tackle the corresponding shortcoming. Chapter 4 addresses the second disadvantage, i.e. if the population size is not a multiple of the sample size. An outline of this drawback is discussed, before briefly examining each associated modified systematic sampling design found in literature. Afterwards, a modified systematic sampling design is proposed to address the problem at hand. In Chapter 5, both shortcomings are solved in tandem by introducing a final modified sampling design. Finally, in Chapter 6, all the work from the previous chapters are integrated to provide a comprehensive report on the thesis, as well as future recommendations/studies. Note that Naidoo (2013) originally studied the proposed modified systematic sampling designs in Chapters 2 and 3, where approximate percentages of that work carried forward within this thesis are given as 90% and 40%, respectively.

# Chapter 2

# Systematic Sampling in the Presence of Linear Trend

Trend is described as the general path which is tracked by the $y_q$'s, as $q \in \{1, ..., N\}$ sequentially increases and/or as time moves forward, if $q$ represents points in time. If the $y_q$'s are inclined to increase as $q$ increases, then the population is said to exhibit positive trend. Conversely, if the $y_q$'s are inclined to decrease as $q$ increases, then the population is said to exhibit negative trend. Trend is either described as linear, or non-linear, i.e. parabolic, quadratic, exponential trend etc. The main focal point in this thesis is to only consider linear trends.

In this chapter, we first discuss a linear trend model and the performance of LSS, when compared to SRS and STR, under this model. Thereafter, we provide an overview of the various modified LSS designs found in literature, which aim to provide optimal results for populations exhibiting linear trend. We next consider some optimal sampling strategies/conditions, before proposing an associated modified systematic sampling design.

## 2.1 Linear trend model

The efficiency of LSS in the presence of linear trend was originally studied in a mathematical context by Madow & Madow (1944), and later discussed by Murthy (1967) and Cochran (1977). Let us consider a population that exhibits linear trend, represented by the model

$$y_q = a + bq + e_q, \qquad q = 1, ..., N, \tag{2.1}$$

where $a$ and $b$ are constants and the $e_q$'s denote the random errors which follow Cochran's (1946) super-population model, i.e.

$$\mathcal{E}\left(e_q\right)=0, \qquad \mathcal{E}\left(e_q^2\right)=\sigma^2, \qquad \mathcal{E}\left(e_q e_z\right)=0\left(q \neq z\right),$$

where the average of all potential finite populations that can be drawn from model (2.1) is denoted by the function $\mathcal{E}$. By referring to (2.1), we obtain

$$\overline{Y}=\frac{1}{N}\sum_{q=1}^{N} y_q=\frac{1}{N}\sum_{q=1}^{N} a+\frac{b}{N}\sum_{q=1}^{N} q+\frac{1}{N}\sum_{q=1}^{N} e_q=a+\frac{b\left(N+1\right)}{2}+\overline{\overline{e}},$$

where $\overline{\overline{e}}=\sum_{q=1}^{N} e_q/N$ denotes the average random error of the population. Thus, when estimating $\overline{Y}$, the expected MSEs of $\overline{y}_{LSS}$, $\overline{y}_{SRS}$, and $\overline{y}_{STR}$, are respectively given by

$$M_{LSS}=\sigma_e^2+\frac{b^2\left(k^2-1\right)}{12}, \tag{2.2}$$

$$M_{SRS}=\sigma_e^2+\frac{b^2\left(N+1\right)\left(k-1\right)}{12} \tag{2.3}$$

and

$$M_{STR}=\sigma_e^2+\frac{b^2\left(k^2-1\right)}{12n}, \tag{2.4}$$

where $\sigma_e^2=\sigma^2(1/n-1/N)$ represents the minimum expected error variance component, while the second terms on the right hand side represent the linear trend components (see Bellhouse (1988)). Now, comparing (2.2) through to (2.4), results in

$$M_{STR}\leq M_{LSS}\leq M_{SRS}. \tag{2.5}$$

Thus, for populations exhibiting linear trend, STR is more efficient than LSS, which in turn is more efficient than SRS. Note that equivalence occurs when $n=1$.

In the presence of linear trend, we obtain a high degree of variation between units that are located within the same linear systematic sample. The cross products for these pairs, with deviations calculated from the population mean, are thus inclined to be negative. Hence, LSS is more efficient than SRS as a result of $\rho$ being negative. Therefore, the greater the amount of linear trend, the more substantial the efficiency gains when favouring LSS over SRS.

Now, let us compare LSS to STR (as in Section 1.2) in the presence linear trend. Strata are thus predominantly internally homogeneous, where a deviation between any sampling unit and its corresponding stratum mean, is probable of having the same coefficient as the deviation between another sampling unit from their corresponding stratum mean. Both deviations between their corresponding stratum means are thus expected to be either positive, or negative, which in turn predominantly results in positive cross products. Accordingly, STR is more efficient than LSS as a result of $\rho_{wst}$ being positive. Note that the greater the amount of linear trend, the more substantial the efficiency loss when favouring LSS over STR.

## 2.2 Modified linear systematic sampling strategies

Many authors have addressed the above-mentioned scenario by suggesting modified LSS strategies. Most of these solutions remove the linear trend component in (2.2) and thus improve results. A review of such strategies, as well as their shortcomings, is presented below.

### 2.2.1 Yates' end corrections (Yates 1948)

This sampling design is equivalent to LSS; however, the sample mean (i.e. the Yates' end corrections (YEC) estimator) is corrected by employing appropriate weights on the first and last sampling units, given by

$$\overline{y}_{YEC} = \overline{y}_{LSS} + \frac{(2i-k-1)}{2\,(n-1)\,k}\left(y_i - y_{i+(n-1)k}\right).$$

Under the assumption of a perfect linear trend in the population (e.g. $y_q = a + bq$, for $q = 1, ..., N$), estimator $\overline{y}_{YEC}$ is equivalent to the population mean. If we consider model (2.1), then we can expect estimator $\overline{y}_{YEC}$ to be a slightly biased estimate of the population mean. Nevertheless, in the presence of a rough linear trend, estimator $\overline{y}_{YEC}$ is usually subject to less error than estimator $\overline{y}_{LSS}$ (Murthy & Rao 1988).

An expression for the expected MSE of $\overline{y}_{YEC}$, when estimating $\overline{Y}$ under model (2.1), is given by

$$M_{YEC} = \sigma_e^2 + \frac{\sigma^2\left(k^2-1\right)}{6(n-1)^2k^2} \tag{2.6}$$

(see Fountain & Pathak (1989)). The linear trend component is thus completely removed, but the resulting effect is a greater error variance component, owing to the uneven weighting of the sampling units.

### 2.2.2 Centered systematic sampling (Madow 1953)

Centered systematic sampling (CESS) adopts the usual LSS design; however, the centrally located linear systematic sample is selected and thus no randomization is required. The corresponding sample mean is given as

$$\overline{y}_{CESS} = \begin{cases} n^{-1} \sum\limits_{j=1}^{n} y_{[(2j-1)k+1]/2}, & \text{if } k \text{ is odd} \\ n^{-1} \sum\limits_{j=1}^{n} y_{(2j-1)k/2} \text{ or } n^{-1} \sum\limits_{j=1}^{n} y_{[(2j-1)k+2]/2}, & \text{if } k \text{ is even} \end{cases}$$

(Bellhouse & Rao 1975). If $k$ is odd, then the sample is selected with a predetermined start of $i = (k+1)/2$, while the predetermined start to select the sample is either $i = k/2$ or $i = (k+2)/2$, each with probability $1/2$, when $k$ is even (Bellhouse & Rao 1975). When estimating $\overline{Y}$ under model (2.1), the expected MSE of $\overline{y}_{CESS}$ is obtained as

$$M_{CESS} = \begin{cases} \sigma_e^2, & \text{if } k \text{ is odd} \\ \sigma_e^2 + b^2/4, & \text{if } k \text{ is even} \end{cases} \tag{2.7}$$

(Fountain & Pathak 1989). Hence, the linear trend component in $M_{CESS}$ is only removed when $k$ is odd. Moreover, certain population units have a zero probability of being included in the sample and thus $\overline{y}_{CESS}$ is subject to bias (Murthy 1967). However, under the assumption of a perfect linear trend in the population, estimator $\overline{y}_{CESS}$ is equivalent to the population mean when $k$ is odd.

### 2.2.3 Balanced systematic sampling (Sethi 1965, Murthy 1967)

In relation to the population unit indices, an arrangement associated with balanced systematic sampling (BSS) is such that the sequence of each alternative set of $k$ population units is reversed. LSS is then conducted on this balanced arrangement, so as to select a balanced systematic sample. Accordingly, the sample mean is given as

$$\overline{y}_{BSS} = \begin{cases} n^{-1} \sum_{j=0}^{(n-2)/2} (y_{i+2jk} + y_{2(j+1)k-i+1}), & \text{if } n \text{ is even} \\ n^{-1} \left[ \sum_{j=0}^{(n-3)/2} (y_{i+2jk} + y_{2(j+1)k-i+1}) + y_{i+(n-1)k} \right], & \text{if } n \text{ is odd.} \end{cases}$$

This estimator is design-unbiased, owing to each population unit having an equal probability, $1/k$, of selection. If we estimate $\overline{Y}$ under model (2.1), then the expected MSE of $\overline{y}_{BSS}$ is expressed as

$$M_{BSS} = \begin{cases} \sigma_e^2, & \text{if } n \text{ is even} \\ \sigma_e^2 + b^2(k^2-1)/12n^2, & \text{if } n \text{ is odd} \end{cases} \tag{2.8}$$

(Fountain & Pathak 1989).

### 2.2.4 Modified systematic sampling (Singh et al. 1968)

With respect to the population unit indices, an arrangement associated with modified systematic sampling (MSS) is such that the sequence of a subset of units, which occur at the end of the population, is reversed. If $n$ is even, then the last $N/2$ units are sequentially reversed, i.e. the population is now re-arranged as $U_1, ..., U_{N/2}, U_N, ..., U_{N/2+1}$. Alternatively, if $n$ is odd, then the last $(N-k)/2$ units are sequentially reversed, i.e. the population will be re-arranged as $U_1, ..., U_{(N+k)/2}, U_N, ..., U_{(N+k)/2+1}$. LSS is then conducted on this modified arrangement so as to select a modified systematic sample. Note that the resulting sample ensures an even spread over the population, except in the center. Consequently, the sample mean is given as

$$\overline{y}_{MSS} = \begin{cases} n^{-1} \sum_{j=0}^{(n-2)/2} (y_{i+jk} + y_{N-jk-i+1}), & \text{if } n \text{ is even} \\ n^{-1} [\sum_{j=0}^{(n-3)/2} (y_{i+jk} + y_{N-jk-i+1}) + y_{i+(n-1)k/2}], & \text{if } n \text{ is odd.} \end{cases}$$

Just as in the case of BSS, estimator $\overline{y}_{MSS}$ is design-unbiased. The expected MSE of $\overline{y}_{MSS}$, when estimating $\overline{Y}$ under model (2.1), is given as

$$M_{MSS} = \begin{cases} \sigma_e^2, & \text{if } n \text{ is even} \\ \sigma_e^2 + b^2(k^2-1)/12n^2, & \text{if } n \text{ is odd} \end{cases} \tag{2.9}$$

(Fountain & Pathak 1989). When comparing (2.8) to (2.9), we see that $M_{MSS} = M_{BSS}$, while further noting that the linear trend components are only eliminated when the sample size $n$ is even.

Good reviews pertaining to the modified LSS designs mentioned thus far, are given by Bellhouse & Rao (1975), Cochran (1977), Fountain & Pathak (1989), Gupta & Kabe (2011), etc.

### 2.2.5 Diagonal systematic sampling (Subramani 2000, 2009, 2010)

If we assume that $n \leq k$, then diagonal systematic sampling (DSS) is conducted as follows:

(i) Arrange the population units according to matrix $\mathbf{M}$, where

$$\mathbf{M} = \begin{bmatrix} U_1 & U_2 & \dots & U_k \\ U_{k+1} & U_{k+2} & \dots & U_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ U_{(n-1)k+1} & U_{(n-1)k+2} & \dots & U_{nk} \end{bmatrix}_{n \times k}.$$

(ii) Randomly select an integer between 1 and $k$, say $i$, where $1 \leq i \leq k$.

(iii) The sample, $S_i$, which is selected in a diagonal fashion from matrix $\mathbf{M}$, is obtained as follows:

(a) if $i \leq k-n+1$, then $S_i = \{U_i, U_{(k+1)+i}, ..., U_{(n-1)(k+1)+i}\}$;

(b) otherwise,

$$S_i = \{U_i, U_{(k+1)+i}, ..., U_{\gamma(k+1)+i}, U_{(\gamma+1)k+1}, U_{(\gamma+2)k+2}, ..., U_{(n-1)k+(n-\gamma-1)}\},$$

where $\gamma = 0, ..., n-2$ for $S_k, ...., S_{k-n+2}$, respectively.

The sample mean, denoted as $\overline{y}_{DSS}$, is a design unbiased estimator of the population mean. If we consider model (2.1), then the expected MSE of $\overline{y}_{DSS}$, when estimating $\overline{Y}$, is found as

$$M_{DSS} = \sigma_e^2 + \frac{b^2(k-n)[n(k-n)+2]}{12n}. \tag{2.10}$$

Clearly, the linear trend component is only removed when $n = k$. To remove the linear trend component for all other cases, Subramani (2000) proposed a DSS with end corrections (DSSEC) estimator, given by

$$\begin{aligned}
\overline{y}_{DSSEC} &= \overline{y}_{DSS} + \frac{(2i-k+n-2)}{N-k+n-1}(y_i - y_{(n-1)(k+1)+i}), && \text{if } i \leq k-n+1 \\
&= \overline{y}_{DSS} + \frac{(k-n)(2k-2i+2-n)}{2n(N-2k+n-1)}(y_i - y_{(n-1)k+(n-\gamma-1)}), && \text{otherwise.}
\end{aligned}$$

As is the case of the YEC estimator, if we assume a perfect linear trend in the population, then estimator $\overline{y}_{DSSEC}$ is equivalent to the population mean, while slightly biased under model (2.1).

Note that DSS is only applicable when $n \leq k$. Subramani (2009, 2010) later introduced generalized DSS (GDSS), which is applicable for all cases of $n$ and is given as follows:

(i) Apply steps (i) and (ii) from the DSS methodology above.

(ii) Select the $i$th population unit in the first row of $\mathbf{M}$ and from left to right, select all downward elements in a diagonal fashion until reaching the last column of $\mathbf{M}$.

(iii) Once the last column of $\mathbf{M}$ is reached, select the first population in the very next row and repeat the diagonal selection process until a sample of size $n$ is obtained.

An unbiased estimator of the population mean is given by the sample mean, which is denoted as $\overline{y}_{GDSS}$. If $n > k$, then the sample size can be expressed as $n = ck + d$, where $c$ and $d$ are non-negative integers. Accordingly, the expected MSE of $\overline{y}_{GDSS}$ under model (2.1) is given by

$$M_{GDSS} = \sigma_e^2 + \frac{b^2 d(k-d)[d(k-d)+2]}{12n^2}. \tag{2.11}$$

If we compare (2.10) to (2.11), while assuming $n = d$ (i.e. $c = 0$), then $M_{DSS} = M_{GDSS}$. Thus, DSS is a particular case of GDSS. Note that if $n = ck$ (i.e. if $d = 0$), then the linear trend component for $M_{GDSS}$ is eliminated.

## 2.3 Optimality in the presence of linear trend

When estimating the population mean in the presence of linear trend, an estimator is considered to provide optimal results if it exhibits minimum expected MSE, i.e. $(i)$ if the linear trend component in the expected MSE of the corresponding estimator is completely removed and $(ii)$ if the expected MSE of the associated estimator exhibits minimum expected error variance. Condition $(i)$ is satisfied for the YEC estimator, as well as for CESS, BSS, MSS, DSS and GDSS, as shown by equations (2.6), (2.7) (if $k$ is odd), (2.8) (if $n$ is even), (2.9) (if $n$ is even), (2.10) (if $n = k$) and (2.11) (if $d = 0$), respectively. Condition $(ii)$ is satisfied if an equal weighting is applied to all the sampling units, as seen for LSS, SRS, STR, CESS, BSS, MSS, DSS and GDSS, verified by equations (2.2) to (2.4) and (2.7) to (2.11), respectively. Thus, the most attractive sampling strategies in the presence of linear trend, are those that are related to estimators that satisfy both conditions, i.e. CESS ($k$ is odd), BSS ($n$ is even), MSS ($n$ is even), DSS ($n = k$) and GDSS ($d = 0$). Moreover, a linear trend free sampling design will have the sampling unit indices, for each and every possible sample, sum up to $n(N + 1)/2$, i.e. if $S$ denotes a sample selected when conducting any given sampling design, then this design is said to be linear trend free, if and only if for all $S$

$$\sum_{U_q \in S} q = \frac{1}{2} n(N+1)$$

(Mukerjee & Sengupta 1990). Using this definition, one can easily verify that CESS (if $k$ is odd), BSS (if $n$ is even), MSS (if $n$ is even), DSS (if $n = k$) and GDSS (if $d = 0$), are all linear trend free sampling designs. Another way of viewing this definition is that linear trend free sampling designs may only exist if $n(N+1) = n(nk+1)$ is even, i.e. if $n$ is even or if both $n$ and $k$ are odd. Thus, we cannot obtain a linear trend free sampling design for the case when $n$ is odd and $k$ is even. Alternatively, one may then opt to use the YEC estimator, as this estimator is usually subject to less error than all other estimators for this scenario. This drawback has motivated the study in the next section, where a modified LSS design, as well as a corresponding competitive end corrections estimator, are proposed.

Note that there are further modified LSS designs given by Subramani (2012, 2013*a*,*b*, 2014). These modifications will not be considered, as they do not provide optimality under any circumstance. One can refer to the modified LSS design discussed by Khan et al. (2015). This design is a generalization of either LSS, DSS and/or GDSS, under certain assumptions. Thus, this design will only exhibit optimum sampling results if the design reduces to either DSS or GDSS, while satisfying the linear trend free sampling conditions for DSS or GDSS, respectively.

## 2.4 Balanced modified systematic sampling

In this section, a modified LSS design, termed as *balanced modified systematic sampling* (BMSS), is proposed. In Section 2.4.1, a discussion on the methodology of BMSS is provided. For Section 2.4.2, the expected MSE of the BMSS sample mean, is compared to that of $M_{LSS}$, $M_{SRS}$, $M_{STR}$, $M_{YEC}$, $M_{CESS}$, $M_{BSS}$ and $M_{MSS}$. As a result, BMSS is only optimal for the case when $n/2$ is an even integer. A BMSS with end corrections (BMSSEC) estimator is thus constructed, so as to remove the linear trend component in the corresponding expected MSE for the other cases of $n$. A numerical example on a hypothetical population is then considered in Section 2.4.3, before carrying out a simulation study in Section 2.4.4. Note that $k$ is assumed to be an integer, i.e. assuming that $N$ is an exact multiple of $n$, so that sampling is conducted linearly.

### 2.4.1 Methodology

A modified arrangement used for BMSS is defined as follows: (a) if $n$ is even, then the order of every alternative set of $k$ population units is reversed, before reversing the order of the first/last $n/2$ sets of $k$ population units; and (b) if $n$ is odd, then the order of every alternative set of $k$ population units is reversed, before reversing the order of the last

$(n-1)/2$ sets of $k$ population units. LSS is then applied to this modified arrangement, so as to select the required sample. Note that different arrangements, before applying LSS, will result in different compositions of samples and this paper deals with a specific arrangement, as explained above. By reversing the order of $n/2$ (or $(n-1)/2$) sets of $k$ population units, a balancing effect is obtained which is optimal for populations exhibiting linear trend. Note that MSS reverses the order of the last $n/2$ (or $(n-1)/2$) sets of $k$ population units, without alternating the order of each set, while BSS alternates the order of each set, without reversing the order of the last $n/2$ (or $(n-1)/2$) sets of $k$ population units. Thus, the ordering of BMSS is a mixture of both, the MSS and BSS orderings. Moreover, BMSS reduces to LSS when $n = 2$ and we will thus assume that $n > 2$.

The above-mentioned design is equivalent to selecting sampling units according to the following indices:

(A) if $n/2$ is an even integer, then

$$i + 2jk, \qquad 2(j+1)k - i + 1, \qquad \text{for } j = 0, ..., (n-4)/4$$

and

$$N + i - k - 2jk, \qquad N - i - k - 2jk + 1, \qquad \text{for } j = 0, ..., (n-4)/4;$$

(B) if $n/2$ is an odd integer, then

$$i + 2jk, \qquad N + i - k - 2jk, \qquad \text{for } j = 0, ..., (n-2)/4$$

and

$$2(j+1)k - i + 1, \qquad N - i - k - 2jk + 1, \qquad \text{for } j = 0, ..., (n-6)/4;$$

(C) if $n = 3$, then

$$i, \qquad 2k - i + 1 \qquad \text{and} \qquad N - i + 1;$$

(D) if $n \neq 3$ and $(n+1)/2$ is an even integer, then

$$i + 2jk, \qquad 2(j+1)k - i + 1, \qquad N - i - 2jk + 1, \qquad \text{for } j = 0, ..., (n-3)/4$$

and

$$N + i - 2(j+1)k, \qquad \text{for } j = 0, ..., (n-7)/4;$$

(E) if $(n+1)/2$ is an odd integer, then

$$i+2jk, \qquad 2(j+1)k-i+1, \qquad N-i-2jk+1, \qquad N+i-2(j+1)k,$$
$$\text{for } j=0,...,(n-5)/4 \qquad \text{and} \qquad i+(n-1)k/2.$$

Note that Cases (A) and (B) are sub-cases of $n$ being even, while Cases (C) to (E) are sub-cases of $n>1$ being odd.

The $i$th ($i \in \{1,...,k\}$) sample mean, denoted by $\overline{y}_{BMSS}$, is obtained by using the above sampling unit indices for the respective cases, e.g. if we consider Case (A), then the sample mean is given as

$$\overline{y}_{BMSS} = \frac{1}{n}\sum_{j=0}^{(n-4)/4}\left(y_{i+2jk}+y_{2(j+1)k-i+1}+y_{N+i-k-2jk}+y_{N-i-k-2jk+1}\right).$$

Note that $\overline{y}_{BMSS}$ is design-unbiased, since BMSS is viewed as an arrangement of units before applying LSS.

### 2.4.2 Expected Mean Square Error Comparisons

To compare the expected MSE of the BMSS estimator, to that of $M_{LSS}$, $M_{SRS}$, $M_{STR}$, $M_{YEC}$, $M_{CESS}$, $M_{BSS}$ and $M_{MSS}$, we first need to consider the following theorem.

**Theorem 1:** *If we suppose model* B, *which is related to model* A, *i.e. model* (2.1), *given by*

$$y_q = a+bq, \qquad q=1,...,N \tag{2.12}$$

*such that*

$$\overline{Y}_B = \frac{1}{N}\sum_{q=1}^{N} y_q = \frac{1}{N}\left[(a+b)+...+(a+Nb)\right] = a+\frac{b\,(N+1)}{2},$$

*then by assuming equal weights* $(1/n)$ *applied to all the sampling units, the expected* MSE *of any sample mean, when estimating* $\overline{Y}$, *is given by*

$$M_A = \mathcal{E}\text{MSE}\left(\overline{y}_A\right) \triangleq \mathcal{E}\left\{\text{E}\left[\left(\overline{y}_A-\overline{Y}\right)^2\right]\right\} = \sigma_e^2+\text{Var}\left(\overline{y}_B\right), \tag{2.13}$$

*where* $\overline{y}_B$ *denotes a linear unbiased estimator of* $\overline{Y}_B$, *using the sampling design associated with* $\overline{y}_A$.

*Proof*: By using (2.1) and (2.12), we obtain $\overline{Y} = \overline{Y}_B + \overline{\overline{e}}$ and $\overline{y}_A = \overline{y}_B + \overline{e}_i$, where $\overline{e}_i = \sum e_i/n$

denotes the mean random error of the sample and $\sum$ denotes the sum over the sample. Using these expressions, it follows that

$$M_A \triangleq \mathcal{E}\left\{\mathrm{E}\left[\left(\overline{y}_A - \overline{Y}\right)^2\right]\right\}$$
$$= \mathcal{E}\left\{\mathrm{E}\left[\left(\overline{y}_B - \overline{Y}_B\right)^2 + \left(\overline{e}_i - \overline{\overline{e}}\right)^2\right]\right\} = \mathcal{E}\mathrm{Var}\left(\overline{y}_B\right) + \mathcal{E}\mathrm{Var}\left(\overline{e}_i\right) = \mathrm{Var}\left(\overline{y}_B\right) + \sigma_e^2.$$

If we let $P = 2i - k - 1$, then applying (2.12) to $\overline{y}_{BMSS}$ results in

$$\begin{aligned} \overline{y}_{BMSS} &= a + b(N+1)/2, && \text{for Case (A)} \\ &= a + b\left[N + 1 + 2P/n\right]/2, && \text{for Case (B)} \\ &= a + b\left[N + 1 - P/n\right]/2, && \text{for Cases (C) to (E).} \end{aligned}$$

Hence, the corresponding variance expression, when using $\overline{y}_{BMSS}$ to estimate $\overline{Y}_B$, is given by

$$\begin{aligned} \mathrm{Var}\left(\overline{y}_{BMSS}\right) &= 0, && \text{for Case (A)} \\ &= b^2(k^2-1)/3n^2, && \text{for Case (B)} \\ &= b^2(k^2-1)/12n^2, && \text{for Cases (C) to (E),} \end{aligned} \tag{2.14}$$

which follows since

$$\mathrm{E}\left(P^2\right) = \frac{1}{k}\sum_{i=1}^{k} P^2 = \frac{\left(k^2-1\right)}{3}.$$

Thus, if we assume model (2.1), then by substituting (2.14) into (2.13), we obtain

$$\begin{aligned} M_{BMSS} &= \sigma_e^2, && \text{for Case (A)} \\ &= \sigma_e^2 + b^2(k^2-1)/3n^2, && \text{for Case (B)} \\ &= \sigma_e^2 + b^2(k^2-1)/12n^2, && \text{for Cases (C) to (E).} \end{aligned} \tag{2.15}$$

By comparing (2.15) and (2.4), we note that $M_{BMSS} < M_{STR}$ for all the cases. Thus, by using (2.5), we conclude that BMSS is more efficient than LSS, SRS and STR. Also, by comparing (2.15) and (2.6), we see that $M_{BMSS} < M_{YEC}$, for (i) Case (A); (ii) Case (B) (if and only if $\sigma^2 > 2b^2(n-1)^2k^2/n^2$); and (iii) Cases (C) to (E) (if and only if $\sigma^2 > b^2(n-1)^2k^2/2n^2$). In addition, the comparison of (2.15) and (2.7) results in:

(i) $M_{BMSS} = M_{CESS}$ for Case (A) and if $k$ is odd;

(ii) $M_{BMSS} < M_{CESS}$ for Case (A) and if $k$ is even;

(iii) $M_{BMSS} > M_{CESS}$ for Cases (B) to (E) and if $k$ is odd;

(iv) $M_{BMSS} < M_{CESS}$ for Case (B), if $k$ is even and $4k^2 - 4 < 3n^2$;

(v) $M_{BMSS} < M_{CESS}$ for Cases (C) to (E), if $k$ is even and $k^2 - 1 < 3n^2$.

Finally, by comparing (2.15) and (2.8), we see that $M_{BMSS} > M_{BSS} = M_{MSS}$ for Case (B), while all other cases result in $M_{BMSS} = M_{BSS} = M_{MSS}$.

Clearly, we only obtain a complete removal of the linear trend component in (2.15) for Case (A). To remove the linear trend component for the other cases, we next consider the application of weights to the first and last sampling units. Accordingly, the resulting estimator and the corresponding expected MSE are respectively given in the next two theorems.

**Theorem 2:** *The* BMSSEC *estimator of* $\overline{Y}$ *with random start* $i$*, for* $i \in \{1, ..., k\}$*, is given as*

$$\begin{aligned}
\overline{y}_{BMSSEC} &= \overline{y}_{BMSS} + P(y_i - y_{N+i-k})/[n(N-k)], && \text{for Case (B)} \\
&= \overline{y}_{BMSS} - P(y_i - y_{N-i+1})/[2n(N-2i+1)], && \text{for Cases (C) and (D)} \\
&= \overline{y}_{BMSS} + P(y_i - y_{N-i+1})/[2n(N-2i+1)], && \text{for Case (E).}
\end{aligned}$$

*Proof:* See Appendix.

**Theorem 3:** *Under model* A*, the expected* MSE *of* $\overline{y}_{BMSSEC}$ *is given as*

$$\begin{aligned}
M_{BMSSEC} &= \sigma_e^2 + 2\sigma^2(k^2-1)/3n^2(N-k)^2, && \text{for Case (B)} \\
&= \sigma_e^2 + \sum_{i=1}^{k} \left\{P^2\sigma^2/[2(N-2i+1)^2 n^2 k]\right\}, && \text{for Cases (C) to (E).}
\end{aligned}$$

*Proof:* See Appendix.

If we compare $M_{BMSSEC}$ to all previous expected MSE expressions, then we note that simple theoretical comparisons are difficult to obtain and we will thus resort to some numerical comparisons in the next two sections. However, one can easily verify that $M_{BSS} = M_{MSS} < M_{BMSSEC} < M_{YEC}$ for Case (B), while $M_{CESS} < M_{BMSSEC}$ if $k$ is odd. Furthermore, just as in the case of the YEC estimator being slightly biased under the assumption of a rough linear trend, owing to the uneven weighting of the sampling units (Murthy, 1967), we obtain the same result for estimator $\overline{y}_{BMSSEC}$.

### 2.4.3 Numerical Example

Consider the hypothetical linear trend population given by Murthy and Rao (1988, p. 161), which is presented in Table 2.1. All the possible samples for various values of $n$ when conducting BMSS, which are obtained by using the sampling unit indices in Section 2.4.1 for the corresponding cases, are presented in Table 2.2. The associated MSEs for the various sampling designs mentioned in this paper are given in Table 2.3. The results suggest that BMSS offers a strict improvement over LSS, SRS and STR, regardless of the sample size. Moreover, if $n/2$ is not a even integer, then we obtain a reduction in estimation error by using the BMSSEC estimator, as opposed to the BMSS estimator. Comparisons amongst the modified LSS designs to either BMSS or the BMSSEC estimator requires further analysis, since we are only considering a single finite population, whereas our theoretical results obtained earlier are based on an infinite super-population. However, we note that in most cases, there is a significant reduction in error when applying any one of the modified LSS designs, as opposed to LSS, SRS and STR.

Table 2.1: A population of 40 units exhibiting a steady linear trend in the value of a variable $y$.

| $U_q$ | $y_q$ | $U_q$ | $y_q$ | $U_q$ | $y_q$ | $U_q$ | $y_q$ |
|---|---|---|---|---|---|---|---|
| $U_1$ | 0 | $U_{11}$ | 10 | $U_{21}$ | 23 | $U_{31}$ | 41 |
| $U_2$ | 1 | $U_{12}$ | 11 | $U_{22}$ | 25 | $U_{32}$ | 43 |
| $U_3$ | 2 | $U_{13}$ | 12 | $U_{23}$ | 29 | $U_{33}$ | 46 |
| $U_4$ | 3 | $U_{14}$ | 12 | $U_{24}$ | 30 | $U_{34}$ | 50 |
| $U_5$ | 4 | $U_{15}$ | 13 | $U_{25}$ | 32 | $U_{35}$ | 52 |
| $U_6$ | 5 | $U_{16}$ | 14 | $U_{26}$ | 33 | $U_{36}$ | 53 |
| $U_7$ | 7 | $U_{17}$ | 15 | $U_{27}$ | 35 | $U_{37}$ | 57 |
| $U_8$ | 7 | $U_{18}$ | 17 | $U_{28}$ | 38 | $U_{38}$ | 59 |
| $U_9$ | 8 | $U_{19}$ | 20 | $U_{29}$ | 39 | $U_{39}$ | 62 |
| $U_{10}$ | 9 | $U_{20}$ | 22 | $U_{30}$ | 40 | $U_{40}$ | 63 |

### 2.4.4 Empirical Comparisons

Three independent simulation studies will be carried out to further evaluate estimator $\overline{y}_{BMSSEC}$. Monte Carlo simulations are used with the statistical software package R, where

Table 2.2: For various values of $n$, the $k$ possible samples (for $i \in \{1, ..., k\}$) using BMSS.

| Case | $n$ | $k$ | Possible Samples |
|---|---|---|---|
| A | 4 | 10 | $S_i = \{U_i, U_{21-i}, U_{31-i}, U_{30+i}\}$ |
| E | 5 | 8 | $S_i = \{U_i, U_{17-i}, U_{41-i}, U_{24+i}\} \cup \{U_{16+i}\}$ |
| A | 8 | 5 | $S_i = \{U_i, U_{11-i}, U_{36-i}, U_{35+i}, U_{10+i}, U_{21-i}, U_{25+i}, U_{26-i}\}$ |
| B | 10 | 4 | $S_i = \{U_i, U_{36+i}, U_{8+i}, U_{28+i}, U_{16+i}, U_{20+i}, U_{9-i}, U_{37-i}, U_{17-i}, U_{29-i}\}$ |
| A | 20 | 2 | $S_i = \{U_i, U_{5-i}, U_{38+i}, U_{39-i}, U_{4+i}, U_{9-i}, U_{34+i}, U_{35-i}, U_{8+i}, U_{13-i}, U_{30+i}\}$ |
| | | | $\cup\{U_{31-i}, U_{12+i}, U_{17-i}, U_{26+i}, U_{27-i}, U_{16+i}, U_{21-i}, U_{22+i}, U_{23-i}\}$ |

Table 2.3: Mean square errors for a hypothetical population exhibiting a linear trend.

| | $n$ | | | | |
|---|---|---|---|---|---|
| | 4 | 5 | 8 | 10 | 20 |
| LSS | 23.1600 | 13.6475 | 6.3288 | 3.3825 | 0.4900 |
| SRS | 83.2264 | 64.7316 | 36.9895 | 27.7421 | 9.2474 |
| STR | 6.6350 | 3.1700 | 0.9625 | 0.4063 | 0.0350 |
| YEC | 0.4116 | 0.1887 | 0.1140 | 0.0240 | 0.0134 |
| CESS | 0.6400 | 0.4225 | 0.0400 | 0.9025 | 0.4900 |
| BSS | 0.4350 | 2.2475 | 0.0288 | 0.0275 | 0.0025 |
| MSS | 2.4725 | 0.0575 | 0.7538 | 0.2025 | 0.0400 |
| BMSS | 0.1475 | 0.5775 | 0.1788 | 0.2275 | 0.0025 |
| BMSSEC | N/A | 0.0730 | N/A | 0.0187 | N/A |

10 000 finite populations are simulated. The expected MSE of each estimator is obtained by averaging the MSEs over the 10 000 populations. The relative expected MSEs of each comparative estimator, with respect to that of estimator $\overline{y}_{BMSSEC}$, is denoted by $R_\alpha = 100 \times M_{BMSSEC}/M_\alpha(\%)$, where $\alpha \in \{\text{LSS}, \text{SRS}, \text{STR}, \text{YEC}, \text{CESS}, \text{BSS}, \text{MSS}, \text{BMSS}\}$. Without loss of generality, we suppose that the $e_q$'s are iid $N(0, 1)$ random variables and let $a = 5$.

In the first simulation study, Case (B) is examined and arbitrary values of $b = 0.5, 1, 2$ and 4, are assigned while varying $n$ and $k$. The associated relative expected MSEs are presented in Tables 2.4 to 2.7. From Tables 2.4 to 2.7, we note that only estimators $\overline{y}_{BSS}$ and $\overline{y}_{MSS}$, are marginally subjected to less error than that of estimator $\overline{y}_{BMSSEC}$. Also, estimator $\overline{y}_{BMSSEC}$ is always favourable over estimators $\overline{y}_{LSS}$, $\overline{y}_{SRS}$ and $\overline{y}_{CESS}$, with greater discrepancies as $n$, $k$ and/or $b$ increases. Similarly, we see that estimator $\overline{y}_{BMSSEC}$ is always preferred over estimator $\overline{y}_{STR}$, with greater discrepancies as $k$ and/or $b$ increases, while results remain constant as $n$ varies. Finally, we note that estimator $\overline{y}_{BMSSEC}$ always performs better than estimator $\overline{y}_{BMSS}$, with greater discrepancies as $k$ and/or $b$ increases and smaller discrepancies as $n$ increases. Thus, $M_{BMSS} \to M_{BMSSEC}$ as $n \to \infty$, provided that $k$ and $b$ are relatively small.

Table 2.4: Simulated relative expected mean square errors for populations exhibiting linear trend under Case (B) ($b = 0.5$).

| $k$ | $n$ | $R_{LSS}$ | $R_{SRS}$ | $R_{STR}$ | $R_{CESS}$ | $R_{BSS}$ | $R_{MSS}$ | $R_{BMSS}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 6 | 57.85 | 23.99 | 90.73 | 57.85 | 101.51 | 102.63 | 92.40 |
| 2 | 34 | 19.08 | 01.02 | 89.47 | 19.08 | 101.07 | 100.29 | 98.49 |
| 2 | 130 | 05.76 | 00.07 | 88.25 | 05.76 | 102.61 | 102.51 | 99.52 |
| 2 | 258 | 02.99 | 00.02 | 88.24 | 02.99 | 99.16 | 99.05 | 99.87 |
| 4 | 6 | 28.87 | 07.50 | 71.57 | 66.63 | 100.70 | 101.63 | 79.28 |
| 4 | 34 | 06.60 | 00.26 | 70.74 | 26.17 | 100.42 | 100.33 | 94.84 |
| 4 | 130 | 01.82 | 00.02 | 70.87 | 08.53 | 102.52 | 99.01 | 98.84 |
| 4 | 258 | 00.92 | $< 00.01$ | 70.81 | 04.45 | 100.31 | 99.04 | 99.18 |
| 8 | 6 | 10.06 | 02.01 | 40.24 | 70.84 | 100.66 | 100.10 | 50.14 |
| 8 | 34 | 01.93 | 00.06 | 40.08 | 29.27 | 99.81 | 99.85 | 85.14 |
| 8 | 130 | 00.51 | $< 00.01$ | 40.10 | 09.79 | 100.46 | 100.15 | 95.62 |
| 8 | 258 | 00.26 | $< 00.01$ | 39.82 | 05.15 | 100.27 | 100.55 | 97.43 |

For the second simulation study, Cases (C) to (E) (i.e. $n$ is odd) are considered and arbitrary values of $b = 0.5, 1, 2$ and 4, are assigned while varying $n$ and $k$. The corresponding relative expected MSEs are presented in Tables 2.8 to 2.11. From Tables 2.8 to 2.11, we note that estimator $\overline{y}_{BMSSEC}$ performs better than all the estimators considered in this study. In this simulation study, we obtain similar results as those obtained in the previous study. However, estimator $\overline{y}_{BMSSEC}$ now performs better than estimators $\overline{y}_{BSS}$ and $\overline{y}_{MSS}$. Moreover, we see that estimators $\overline{y}_{BSS}$, $\overline{y}_{MSS}$ and $\overline{y}_{BMSS}$, are relatively subject to the same amount of error. Thus, $M_{BSS}, M_{MSS}$ and $M_{BMSS} \to M_{BMSSEC}$ as $n \to \infty$, provided that $k$ and $b$ are relatively small.

Table 2.5: Simulated relative expected mean square errors for populations exhibiting linear trend under Case (B) ($b = 1$).

| $k$ | $n$ | $R_{LSS}$ | $R_{SRS}$ | $R_{STR}$ | $R_{CESS}$ | $R_{BSS}$ | $R_{MSS}$ | $R_{BMSS}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 6 | 24.95 | 07.16 | 66.55 | 24.95 | 99.00 | 99.53 | 75.38 |
| 2 | 34 | 05.59 | 00.25 | 66.74 | 05.59 | 100.84 | 100.51 | 94.65 |
| 2 | 130 | 01.54 | 00.02 | 67.78 | 01.54 | 100.91 | 102.35 | 98.60 |
| 2 | 258 | 00.80 | $<$ 00.01 | 69.26 | 00.80 | 105.18 | 105.60 | 99.44 |
| 4 | 6 | 09.25 | 01.99 | 38.14 | 33.49 | 100.93 | 101.08 | 48.34 |
| 4 | 34 | 01.72 | 00.06 | 37.23 | 08.08 | 98.99 | 99.99 | 83.69 |
| 4 | 130 | 00.46 | $<$ 00.01 | 37.46 | 02.25 | 99.82 | 101.26 | 95.29 |
| 4 | 258 | 00.23 | $<$ 00.01 | 37.49 | 01.15 | 100.23 | 101.06 | 97.11 |
| 8 | 6 | 02.70 | 00.51 | 14.28 | 36.91 | 100.38 | 100.58 | 19.93 |
| 8 | 34 | 00.49 | 00.02 | 14.26 | 09.31 | 100.27 | 100.73 | 58.78 |
| 8 | 130 | 00.13 | $<$ 00.01 | 14.16 | 02.61 | 99.12 | 99.41 | 84.22 |
| 8 | 258 | 00.06 | $<$ 00.01 | 14.18 | 01.33 | 102.09 | 100.64 | 91.56 |

Table 2.6: Simulated relative expected mean square errors for populations exhibiting linear trend under Case (B) ($b = 2$).

| $k$ | $n$ | $R_{LSS}$ | $R_{SRS}$ | $R_{STR}$ | $R_{CESS}$ | $R_{BSS}$ | $R_{MSS}$ | $R_{BMSS}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 6 | 07.68 | 01.90 | 33.33 | 07.68 | 99.73 | 101.68 | 42.20 |
| 2 | 34 | 01.44 | 00.06 | 32.99 | 01.44 | 99.04 | 99.59 | 80.93 |
| 2 | 130 | 00.38 | $<$ 00.01 | 33.50 | 00.38 | 99.90 | 101.72 | 94.94 |
| 2 | 258 | 00.19 | $<$ 00.01 | 33.54 | 00.19 | 101.66 | 101.58 | 96.97 |
| 4 | 6 | 02.43 | 00.50 | 13.03 | 11.09 | 100.52 | 99.60 | 18.31 |
| 4 | 34 | 00.44 | 00.02 | 13.21 | 02.18 | 100.84 | 99.90 | 56.22 |
| 4 | 130 | 00.11 | $<$ 00.01 | 12.95 | 00.57 | 101.01 | 99.83 | 83.06 |
| 4 | 258 | 00.06 | $<$ 00.01 | 13.07 | 00.29 | 99.54 | 99.92 | 90.80 |
| 8 | 6 | 00.68 | 00.13 | 03.99 | 12.77 | 99.65 | 99.86 | 05.86 |
| 8 | 34 | 00.12 | $<$ 00.01 | 04.00 | 02.51 | 100.08 | 100.40 | 25.97 |
| 8 | 130 | 00.03 | $<$ 00.01 | 04.01 | 00.67 | 99.76 | 99.37 | 57.89 |
| 8 | 258 | 00.02 | $<$ 00.01 | 03.99 | 00.34 | 99.41 | 99.41 | 72.91 |

Table 2.7: Simulated relative expected mean square errors for populations exhibiting linear trend under Case (B) ($b = 4$).

| $k$ | $n$ | $R_{LSS}$ | $R_{SRS}$ | $R_{STR}$ | $R_{CESS}$ | $R_{BSS}$ | $R_{MSS}$ | $R_{BMSS}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 6 | 02.08 | 00.49 | 11.30 | 02.08 | 103.38 | 102.87 | 16.23 |
| 2 | 34 | 00.36 | 00.02 | 11.05 | 00.36 | 102.27 | 99.00 | 51.55 |
| 2 | 130 | 00.10 | $<$ 00.01 | 11.16 | 00.10 | 101.55 | 99.06 | 80.80 |
| 2 | 258 | 00.05 | $<$ 00.01 | 10.83 | 00.05 | 99.02 | 100.08 | 88.48 |
| 4 | 6 | 00.62 | 00.12 | 03.60 | 03.02 | 100.68 | 99.61 | 05.31 |
| 4 | 34 | 00.11 | $<$ 00.01 | 03.59 | 00.54 | 99.15 | 99.43 | 23.98 |
| 4 | 130 | 00.03 | $<$ 00.01 | 03.61 | 00.14 | 98.94 | 99.51 | 55.27 |
| 4 | 258 | 00.01 | $<$ 00.01 | 03.59 | 00.07 | 99.13 | 100.16 | 70.01 |
| 8 | 6 | 00.17 | 00.03 | 01.03 | 03.54 | 100.02 | 100.92 | 01.54 |
| 8 | 34 | 00.03 | $<$ 00.01 | 01.04 | 00.65 | 100.75 | 100.82 | 08.21 |
| 8 | 130 | 00.01 | $<$ 00.01 | 01.03 | 00.17 | 100.02 | 100.40 | 25.13 |
| 8 | 258 | $<$ 00.01 | $<$ 00.01 | 01.03 | 00.08 | 100.62 | 99.91 | 40.06 |

Comparisons between estimators $\overline{y}_{BMSSEC}$ and $\overline{y}_{YEC}$ are evaluated in the third simulation study. Because there are no trend components in the expected MSEs of both estimators, an arbitrary value of $b = 4$ is assigned while varying $n$ and $k$. Also, only Cases (C) to (E) are explored, as it was theoretically shown previously that $M_{BMSSEC} < M_{YEC}$ for Case (B). The simulated relative expected MSEs are presented in Table 2.12. The results suggest that estimator $\overline{y}_{BMSSEC}$ is only preferred when $n$ and $k$ are small. Otherwise, there are marginal gains when choosing estimator $\overline{y}_{BMSSEC}$ over estimator $\overline{y}_{YEC}$.

### 2.4.5 Concluding Remarks

A modified LSS design (i.e. BMSS) that depends on an arrangement of population units before applying LSS, which results in the corresponding sample mean being design-unbiased, has been proposed. Results from Sections 2.4.2 to 2.4.4 indicate that BMSS is more efficient than LSS, SRS and STR, in the presence of linear trend. The optimal case of BMSS is when $n/2$ is an even integer, which results in linear trend free sampling and minimum expected MSE of the corresponding sample mean. For the other cases of BMSS, a modified end corrections estimator, i.e. estimator $\overline{y}_{BMSSEC}$, has been constructed.

Table 2.8: Simulated relative expected mean square errors for populations exhibiting linear trend under Cases (C) to (E) ($b = 0.5$).

| $k$ | $n$ | $R_{LSS}$ | $R_{SRS}$ | $R_{STR}$ | $R_{CESS}$ | $R_{BSS}$ | $R_{MSS}$ | $R_{BMSS}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 70.39 | 53.54 | 88.52 | 70.39 | 98.00 | 98.41 | 98.72 |
| 2 | 35 | 18.35 | 00.94 | 87.75 | 18.35 | 98.16 | 98.51 | 99.57 |
| 2 | 125 | 05.92 | 00.08 | 88.00 | 05.92 | 99.01 | 99.66 | 99.88 |
| 2 | 255 | 03.08 | 00.02 | 89.77 | 03.08 | 99.89 | 99.99 | 99.96 |
| 4 | 3 | 45.40 | 24.17 | 72.16 | 81.30 | 91.01 | 89.63 | 89.99 |
| 4 | 35 | 06.41 | 00.24 | 70.49 | 25.65 | 98.96 | 98.54 | 98.78 |
| 4 | 125 | 01.89 | 00.02 | 70.82 | 08.74 | 99.29 | 99.36 | 99.59 |
| 4 | 255 | 00.93 | $< 00.01$ | 70.67 | 04.50 | 99.84 | 99.99 | 99.84 |
| 8 | 3 | 18.50 | 07.55 | 40.69 | 83.34 | 68.31 | 67.80 | 68.28 |
| 8 | 35 | 01.86 | 00.06 | 39.89 | 28.77 | 95.34 | 96.24 | 95.97 |
| 8 | 125 | 00.54 | $< 00.01$ | 40.36 | 10.25 | 98.78 | 99.47 | 98.78 |
| 8 | 255 | 00.26 | $< 00.01$ | 40.13 | 05.26 | 99.40 | 99.88 | 99.32 |

Table 2.9: Simulated relative expected mean square errors for populations exhibiting linear trend under Cases (C) to (E) ($b = 1$).

| $k$ | $n$ | $R_{LSS}$ | $R_{SRS}$ | $R_{STR}$ | $R_{CESS}$ | $R_{BSS}$ | $R_{MSS}$ | $R_{BMSS}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 42.66 | 23.21 | 70.25 | 42.66 | 88.72 | 88.20 | 88.28 |
| 2 | 35 | 05.39 | 00.24 | 66.54 | 05.39 | 98.63 | 98.12 | 98.60 |
| 2 | 125 | 01.57 | 00.02 | 66.46 | 01.57 | 98.30 | 98.48 | 98.60 |
| 2 | 255 | 00.78 | $< 00.01$ | 66.48 | 00.78 | 99.51 | 99.97 | 99.79 |
| 4 | 3 | 17.03 | 07.28 | 38.28 | 50.89 | 65.39 | 66.23 | 65.88 |
| 4 | 35 | 01.68 | 00.06 | 37.29 | 07.88 | 94.72 | 94.44 | 95.06 |
| 4 | 125 | 00.48 | $< 00.01$ | 37.58 | 02.36 | 98.88 | 98.33 | 98.82 |
| 4 | 255 | 00.24 | $< 00.01$ | 38.00 | 01.18 | 99.99 | 98.97 | 99.51 |
| 8 | 3 | 05.42 | 02.02 | 14.71 | 54.98 | 34.36 | 34.26 | 34.36 |
| 8 | 35 | 00.48 | 00.02 | 14.39 | 09.24 | 85.56 | 85.21 | 85.64 |
| 8 | 125 | 00.13 | $< 00.01$ | 14.25 | 02.71 | 95.45 | 96.21 | 95.36 |
| 8 | 255 | 00.07 | $< 00.01$ | 14.37 | 01.36 | 96.95 | 97.72 | 97.57 |

Table 2.10: Simulated relative expected mean square errors for populations exhibiting linear trend under Cases (C) to (E) ($b = 2$).

| $k$ | $n$ | $R_{LSS}$ | $R_{SRS}$ | $R_{STR}$ | $R_{CESS}$ | $R_{BSS}$ | $R_{MSS}$ | $R_{BMSS}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 14.66 | 06.79 | 34.26 | 14.66 | 61.90 | 62.10 | 61.82 |
| 2 | 35 | 01.38 | 00.06 | 32.71 | 01.38 | 95.03 | 94.90 | 94.55 |
| 2 | 125 | 00.39 | $< 00.01$ | 32.60 | 00.39 | 98.22 | 98.37 | 98.43 |
| 2 | 255 | 00.19 | $< 00.01$ | 33.09 | 00.19 | 98.94 | 99.75 | 99.25 |
| 4 | 3 | 04.88 | 01.92 | 13.36 | 20.68 | 31.62 | 31.83 | 31.91 |
| 4 | 35 | 00.42 | 00.02 | 12.99 | 02.08 | 84.01 | 83.59 | 84.05 |
| 4 | 125 | 00.12 | $< 00.01$ | 12.98 | 00.59 | 94.68 | 93.93 | 94.98 |
| 4 | 255 | 00.06 | $< 00.01$ | 13.05 | 00.29 | 98.13 | 97.76 | 97.46 |
| 8 | 3 | 01.40 | 00.51 | 04.09 | 22.99 | 11.37 | 11.35 | 11.34 |
| 8 | 35 | 00.12 | $< 00.01$ | 03.98 | 02.43 | 59.25 | 59.13 | 59.01 |
| 8 | 125 | 00.03 | $< 00.01$ | 03.99 | 00.69 | 83.52 | 83.49 | 83.69 |
| 8 | 255 | 00.02 | $< 00.01$ | 04.03 | 00.34 | 91.83 | 90.98 | 91.54 |

Table 2.11: Simulated relative expected mean square errors for populations exhibiting linear trend under Cases (C) to (E) ($b = 4$).

| $k$ | $n$ | $R_{LSS}$ | $R_{SRS}$ | $R_{STR}$ | $R_{CESS}$ | $R_{BSS}$ | $R_{MSS}$ | $R_{BMSS}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 04.06 | 01.78 | 11.28 | 04.06 | 27.59 | 27.81 | 27.72 |
| 2 | 35 | 00.35 | 00.02 | 11.05 | 00.35 | 81.58 | 81.06 | 81.51 |
| 2 | 125 | 00.10 | $< 00.01$ | 11.16 | 00.10 | 92.83 | 93.56 | 93.80 |
| 2 | 255 | 00.05 | $< 00.01$ | 11.13 | 00.05 | 96.56 | 97.40 | 97.06 |
| 4 | 3 | 01.27 | 00.49 | 03.71 | 06.02 | 10.41 | 10.31 | 10.38 |
| 4 | 35 | 00.11 | $< 00.01$ | 03.62 | 00.53 | 56.26 | 56.53 | 56.25 |
| 4 | 125 | 00.03 | $< 00.01$ | 03.63 | 00.15 | 82.43 | 82.85 | 82.33 |
| 4 | 255 | 00.01 | $< 00.01$ | 03.64 | 00.07 | 90.81 | 90.58 | 90.72 |
| 8 | 3 | 00.36 | 00.13 | 01.06 | 07.02 | 03.12 | 03.11 | 03.11 |
| 8 | 35 | 00.03 | $< 00.01$ | 01.04 | 00.62 | 26.93 | 26.93 | 26.89 |
| 8 | 125 | 00.01 | $< 00.01$ | 01.03 | 00.17 | 56.48 | 57.07 | 56.55 |
| 8 | 255 | $< 00.01$ | $< 00.01$ | 01.03 | 00.09 | 73.05 | 73.11 | 72.63 |

Table 2.12: Simulated relative expected mean square errors of the YEC sample mean, with respect to that of the MBMSSEC sample mean, for populations exhibiting linear trend under Cases C to E.

| | $n$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 | 5 | 7 | 13 | 15 | 29 | 63 | 125 | 255 |
| $k = 2$ | 86.56 | 92.63 | 95.53 | 97.49 | 97.83 | 98.82 | 99.02 | 99.59 | 99.17 |
| $k = 4$ | 89.23 | 94.03 | 96.13 | 97.69 | 98.48 | 98.97 | 99.41 | 99.59 | 99.82 |
| $k = 8$ | 90.04 | 94.64 | 96.38 | 97.90 | 98.53 | 99.16 | 99.65 | 99.92 | 99.95 |

Populations exhibiting a rough linear trend result in estimator $\overline{y}_{BMSSEC}$ being a slightly biased estimate of $\overline{Y}$ as well as exhibiting an inflated error variance component in the corresponding expected MSE, owing to the uneven weighting of the sampling units.

If $n/2$ is an odd integer, then estimator $\overline{y}_{BMSSEC}$ is subject to less error than estimators $\overline{y}_{LSS}$, $\overline{y}_{SRS}$, $\overline{y}_{STR}$, $\overline{y}_{YEC}$ and $\overline{y}_{BMSS}$, while marginally susceptible to more error than estimators $\overline{y}_{BSS}$ and $\overline{y}_{MSS}$, as shown in Sections 2.4.2 and 2.4.4. In addition, if $n$ is odd, then estimator $\overline{y}_{BMSSEC}$ is subject to less error than all of the above-mentioned estimators. The simulation study in Section 2.4.4 indicates that estimator $\overline{y}_{BMSSEC}$ performs better than estimator $\overline{y}_{YEC}$ if $n$ is odd, provided that $n$ and $k$ are small. Otherwise, there are marginal gains when opting to use estimator $\overline{y}_{BMSSEC}$ over estimator $\overline{y}_{YEC}$. Under this circumstance, one may opt to use estimator $\overline{y}_{YEC}$, owing to simplicity.

Finally, we note that estimator $\overline{y}_{BMSSEC}$ performs better than estimator $\overline{y}_{CESS}$, provided that $k$ is even, as seen in the simulation study from Section 2.4.4. However, if $k$ is odd, then the theoretical results in Section 2.4.2 suggest that estimator $\overline{y}_{CESS}$ is to be the preferred, as CESS is an optimal sampling design for this scenario. Nevertheless, we can expect marginal gains when opting to use estimator $\overline{y}_{CESS}$ over estimator $\overline{y}_{BMSSEC}$ when $k$ is odd.

Recommendations for the most appropriate design(s) under various conditions are provided in Table 2.13. Note that the third column represents a trade-off between estimators $\overline{y}_{YEC}$ and $\overline{y}_{BMSSEC}$, where preference is either given to minimum MSE or simplicity.

Table 2.13: Recommended designs for populations exhibiting linear trend.

| Case(s) | Condition | Preference | Recommended Design(s) |
|---|---|---|---|
| A | $k$ is even | N/A | BSS, MSS or BMSS |
| A | $k$ is odd | N/A | CESS, BSS, MSS or BMSS |
| B | $k$ is even | N/A | BSS or MSS |
| B | $k$ is odd | N/A | CESS, BSS or MSS |
| C to E | $k$ is even; $n$ and $k$ are small | Minimum MSE | BMSSEC |
| C to E | $k$ is even; $n$ and $k$ are small | Simplicity | YEC |
| C to E | $k$ is even; $n$ and/or $k$ are not small | Minimum MSE | YEC or BMSSEC |
| C to E | $k$ is even; $n$ and/or $k$ are not small | Simplicity | YEC |
| C to E | $k$ is odd | N/A | CESS |

In this chapter, we have discussed systematic sampling, modifications of the usual systematic sampling design found in literature as well as a suggested modified systematic sampling design, all under the assumption of linear trend among the population units. We also included a section on conditions regarding the optimality of systematic sampling designs and modifications in the presence of linear trend. The results from this chapter suggest that values of the sample size and sampling interval needs to be considered, before selecting an appropriate modified systematic sampling design in the presence of linear trend, where Table 2.13 provides us with the most suitable modified systematic sampling design under various scenarios of $n$ and $k$. In the next chapter, we will investigate the first of the two shortcomings of systematic sampling, i.e. the impossibility of obtaining an unbiased estimate of the sampling variance when conducting systematic sampling with a single random start, under the assumption of linear trend among the population units.

# Chapter 3

# Estimation of the Sampling Variance

Remembering that $S_1$, ..., $S_k$ are the $k$ possible linear systematic samples that can be randomly selected (refer to Table 1.1), the first-order inclusion probability of the unit $U_q$ under LSS, is given by $\pi_q = \mathrm{P}(U_q \in S_i) = 1/k$, for all $i \in \{1, ..., k\}$ and $q \in \{1, ..., N\}$. This indicates that each population unit has an equal probability of inclusion for the linear systematic sample and the relative sample mean $\overline{y}_{LSS}$ is an unbiased estimator of the population mean, since $\pi_q > 0$ for all $q \in \{1, ..., N\}$. In addition, if for some $q, z \in \{1, ...., N\}$ $(q \neq z)$, the second-order inclusion probability is denoted as $\pi_{qz} = \mathrm{P}(U_q \text{ and } U_z \in S_i)$, then for all $i \in \{1, ..., k\}$ and $q, z \in \{1, ...., N\}$ $(q \neq z)$, the second-order inclusion probabilities for the pair of units $\{U_q, U_z\}$ are given by

$$\pi_{qz} = \begin{cases} 1/k, & \text{if } U_q \text{ and } U_z \in S_i \\ 0, & \text{otherwise.} \end{cases}$$

This indicates that some pairs of population units have a zero probability of inclusion for the linear systematic sample.

Now, an unbiased estimate of $\mathrm{V}(\overline{y}_{LSS}) = \mathrm{E}(\overline{y}^2_{LSS}) - \overline{Y}^2$ is given as $\overline{y}^2_{LSS} - Est(\overline{Y}^2)$, where an unbiased estimate of $\overline{Y}^2 = \sum_{q=1}^{N} y_q^2/N^2 + \sum_{q=1}^{N} \sum_{z \neq q}^{N} y_q y_z/N^2$ is denoted as $Est(\overline{Y}^2)$ (Murthy 1967). One can easily verify that an unbiased estimate of $\sum_{q=1}^{N} y_q^2/N^2$ is given by $\sum_{j=1}^{n} y^2_{i+(j-1)k}/n^2 k$, owing to each population unit having an equal probability of inclusion for the linear systematic sample. However, if we apply LSS with a single start, then it is impossible to obtain an unbiased estimate of $\sum_{q=1}^{N} \sum_{z \neq q}^{N} y_q y_z/N^2$, since some pairs of population units have a zero probability of inclusion for the linear systematic sample. Thus, it is impossible to obtain an unbiased estimate of $\overline{Y}^2$ when conducting LSS with a single random start, which in turn results in it being impossible to obtain an unbiased estimate of $\mathrm{V}(\overline{y}_{LSS})$. In light of this result, Wolter (1984, 2007) constructed a class of variance estimators and evaluated their performance under the assumption of various population structures. In the presence of linear trend, he found that two of these estimators provide a least biased estimate of $\mathrm{V}(\overline{y}_{LSS})$. In practice, samplers usually apply the LSS design and use an estimate of $\mathrm{V}(\overline{y}_{SRS})$ to estimate $\mathrm{V}(\overline{y}_{LSS})$. Under this circumstance, the estimator provides an overestimate of $\mathrm{V}(\overline{y}_{LSS})$ in the presence of linear trend. To overcome this shortcoming, such that unbiased estimates of the associated sampling variance are obtained, we next consider all relative modified LSS designs found in literature, before suggesting a corresponding modified LSS design.

## 3.1 Modified linear systematic sampling designs

### 3.1.1 Multiple-Start linear systematic sampling (Deming 1960, Gautschi 1957, Shiue 1960, Tornqvist 1963)

The theory of replicated sampling was originally suggested by Mahalanobis (1946) and Tukey (1950), and later as a variation of LSS by Deming (1960), Gautschi (1957), Shiue (1960) and Tornqvist (1963). Multiple-Start LSS (MLSS) entails conducting LSS with multiple random starts. If the required sample size is now $nm$ (i.e. we are assuming that the required sample size is a non-prime integer), where $m \in \{2, ..., k-1\}$, then the methodology of MLSS is given as follows:

(i) Select $m$ integers $(i1, ..., im)$ from the first $k$ integers using SRS.

(ii) The sampling unit indices are then given as

$$ih + (j-1)k, \qquad h = 1, ..., m \text{ and } j = 1, ..., n.$$

From Table 1.1, we note that the above methodology suggests that we are merely selecting $m$ samples of size $n$ from the $k$ possible linear systematic samples, using SRS. Thus, MLSS is a form of cluster sampling. Tornqvist (1963) proposed the use of simple random sampling with replacement for step (i). To avoid duplicate samples and obtain better estimation results, we will not consider this scenario. The usual case of obtaining more efficient results when favouring SRS over simple random sampling with replacement applies here, and we can thus expect better results when applying the above-mentioned procedure, as opposed to that which is considered by Tornqvist (1963).

The sample mean, denoted as $\overline{y}_{MLSS}$, is an unbiased estimate of the population mean and is given by

$$\overline{y}_{MLSS} = \frac{1}{m}\sum_{h=1}^{m}\overline{y}_{ih} = \frac{1}{nm}\sum_{h=1}^{m}\sum_{j=1}^{n} y_{ih+(j-1)k},$$

where $\overline{y}_{ih}$ denotes the LSS sample mean with random start $ih$. For this scenario, each sample mean is treated as a population unit, where we apply SRS to select $m$ sample means from a total of $k$ possible sample means. Thus, the adjusted population variance is obtained by respectively substituting $y_i$ and $N$ in (1.4), for $\overline{y}_{LSS}$ and $k$, i.e.

$$S_{\overline{y}}^2 = \frac{1}{k-1}\sum_{i=1}^{k}(\overline{y}_{LSS} - \overline{Y})^2,$$

where $\overline{Y} = \sum_{q=1}^{N} y_i/N = \sum_{i=1}^{k}\overline{y}_{LSS}/k$. Using this expression, the variance of $\overline{y}_{MLSS}$ is then obtained by respectively substituting $S_Y^2$, $N$ and $n$ in (1.5), for $S_{\overline{y}}^2$, $k$ and $m$, i.e.

$$\mathrm{V}\left(\overline{y}_{MLSS}\right) = \left(\frac{k-m}{mk}\right)\frac{1}{k-1}\sum_{i=1}^{k}\left(\overline{y}_{LSS} - \overline{Y}\right)^2 = \left(\frac{k-m}{k-1}\right)\frac{\mathrm{V}\left(\overline{y}_{LSS}\right)}{m}.$$

An unbiased estimate of $\mathrm{V}\left(\overline{y}_{MLSS}\right)$ is then given by

$$\hat{\mathrm{V}}\left(\overline{y}_{MLSS}\right) = \left(\frac{k-m}{mk}\right)\frac{1}{m-1}\sum_{h=1}^{m}\left(\overline{y}_{ih} - \overline{y}_{MLSS}\right)^2.$$

For random starts $m = 2$, Sampath (2009) provides a study that compares the efficiency of estimator $\hat{\mathrm{V}}\left(\overline{y}_{MLSS}\right)$ to that of an unbiased estimate of $\mathrm{V}(\overline{y}_{SRS})$, under various population structures. The results suggest that estimator $\hat{\mathrm{V}}\left(\overline{y}_{MLSS}\right)$ is favourable for all the population structures considered.

If we select a sample of size $nm$ from a population of size $N = nk = nml$, using either LSS, SRS, STR, or MLSS, then the expected MSEs of the corresponding sample means (i.e. $\overline{y}_{LSS}$, $\overline{y}_{SRS}$, $\overline{y}_{STR}$ and $\overline{y}_{MLSS}$), when estimating $\overline{Y}$ under model (2.1), are given as

$$M_{LSS} = \sigma_l^2 + \frac{b^2(l-1)(l+1)}{12}, \tag{3.1}$$

$$M_{SRS} = \sigma_l^2 + \frac{b^2(l-1)(N+1)}{12}, \tag{3.2}$$

$$M_{STR} = \sigma_l^2 + \frac{b^2(l-1)(l+1)}{12nm} \tag{3.3}$$

and

$$M_{MLSS} = \sigma_l^2 + \frac{b^2(l-1)(lm+1)}{12}, \tag{3.4}$$

where $\sigma_l^2 = \sigma^2(l-1)/N$ denotes the minimum expected error variance component (see Gautschi (1957)). Assuming that $n > 1$, the comparisons among equations (3.1) through to (3.4), results in

$$M_{STR} < M_{LSS} < M_{MLSS} < M_{SRS}.$$

We are thus presented with a trade-off when conducting MLSS, where obtaining an unbiased estimate of the sampling variance comes at a cost of reduced precision in estimating $\overline{Y}$. Note that it is impossible to obtain an unbiased estimate of the sampling variance when conducting STR with one unit selected per stratum, since certain pairs of units have a zero probability of inclusion for the sample. Moreover, when referring to (3.4), we see greater reductions in precision when estimating $\overline{Y}$, as $m$ increases. However, it is well-known that $m$ should be of sufficient size, so as to obtain a reasonably precise estimate of $\mathrm{V}(\overline{y}_{MLSS})$. This trade-off was studied by Kouijn (1973), where he introduced two modified MLSS designs to provide reasonable solutions. These designs will not be reviewed, as obtaining an efficient estimate of the sampling variance is not one of the key areas of concern in this thesis.

### 3.1.2 Balanced random sampling (Singh & Garg 1979)

So far we have assumed that the population size is a multiple of the sample size, i.e. $N = nk$. Under this assumption there exists three possible cases of balanced random sampling (BRS), namely, (A) $N$ and $n$ are both even; (B) $N$ is even and $n$ is odd; (C) $N$ and $n$ are both odd. The methodologies for these cases are given as follows:

Case (A):
- (i) Select a sample of size $n/2$ from the first $N/2$ population units using SRS, where the sampling unit indices are denoted as $f_g$, for $g = 1, ..., n/2$.
- (ii) The balanced random sample is then given as the sample in (i), as well as those population units with indices $N + 1 - f_g$, for $g = 1, ..., n/2$.

Case (B):
- (i) Select a sample of size $n - 1$ from the $N$ population units, using the procedure in Case (A).

(ii) The balanced random sample is then given as the sample in (i), as well as a randomly selected unit from the remaining $N - n + 1$ population units.

Case (C): (i) Randomly select a unit from the $N$ population units.

(ii) The balanced random sample is then given as the randomly selected unit in (i), as well as a further $(n - 1)$ units, which are selected from the remaining $(N - 1)$ units using the procedure in Case (A).

Note that in the presence of linear trend, only Case (A) provides linear trend free sampling results, i.e. $\sum_{U_q \in S} q = n(N+1)/2$, for all $S$ (refer to Section 2.3). This is also validated by Singh & Garg (1979), where a sampling variance expression is computed under the assumption of model (2.12). For Case (B), $(n - 1)$ units are paired optimally using the usual pairing procedure, while the other sampling unit is selected using SRS, which will thus contribute to a linear trend component in the expected MSE of the associated sample mean. Finally, under Case (C), some pairs of sampling units may not be optimally paired, as SRS is first applied before applying the usual pairing procedure.

### 3.1.3 Partially systematic sampling (Zinger 1963, 1964, 1980, Wu 1984)

Partially systematic sampling involves the supplementation of a linear systematic sample of size $n_1$ with a random sample of size $n_2$, where $n = n_1 + n_2$. The corresponding methodology is given as follows:

(i) Let $k' = N/n_1$ be an integer, such that LSS is applied to select $n_1$ sampling units.

(ii) The partially systematic sample is then given as the sample in (i), as well as $n_2$ units selected from the remaining $N - n_1$ population units, using SRS.

Let $\overline{y}_s$ and $\overline{y}_r$ denote the sample means corresponding to steps (i) and (ii), respectively. An unbiased estimate of the population mean, which is a weighted average of these sample means, as well as the variance of this estimator, are respectively given by

$$\overline{y}(\beta) = (1 - \beta)\overline{y}_s + \beta\overline{y}_r, \qquad 0 \leq \beta \leq 1 \tag{3.5}$$

and

$$\text{Var}(\overline{y}(\beta)) = \alpha_1(\beta)S_Y^2 + \alpha_2(\beta)\text{Var}(\overline{y}_s), \tag{3.6}$$

where

$$\alpha_1(\beta) = \frac{\beta^2(N-1)(N-n_1-n_2)}{n_2 N(N-n_1-1)},$$

$$\alpha_2(\beta) = \left[1 - \frac{\beta k}{(k-1)}\right]^2 - \frac{\beta^2(N-n_1-n_2)}{n_2(k-1)^2(N-n_1-1)},$$

$$\mathrm{Var}(\overline{y}_s) = \frac{S_Y^2}{n_1}\left(\frac{N-1}{N}\right)[1+(n_1-1)\rho']$$

and $\rho'$ is found by replacing $n$ and $k$ in (1.1) with $n_1$ and $k'$, respectively.

To obtain an unbiased estimate of $\mathrm{Var}(\overline{y}(\beta))$, which we will denote as $v(\overline{y}(\beta))$, Zinger (1980) independently derived unbiased estimates of $S_Y^2$ and $\mathrm{Var}(\overline{y}_s)$. He then showed that a value of $\beta$ which minimizes (3.6), as well as a natural weighted average, given by $\beta = n_2/(n_1+n_2)$, may result in $v(\overline{y}(\beta))$ assuming negative values. If we let $\beta = 1/2$, then $v(\overline{y}(\beta))$ will always assume non-negative values. Zinger (1980) studied this case and stated that $\beta = 1/2$ is optimal when minimizing (3.6), provided that $k'$ is large and $n_2 = n_1/(1+(n_1-1)\rho')$. This scenario is generally unrealistic, as $\rho'$ is usually unknown before sampling.

By letting $Q_s = \sum(y_i - \overline{y}_s)^2$ , $Q_r = \sum(y_i - \overline{y}_r)^2$ and $Q_b = \sum(\overline{y}_s - \overline{y}_r)^2$, Wu (1984) proposed an unbiased estimate of (3.6), given as

$$v'(\overline{y}(\beta)) = C(Q_s + \lambda Q_r) + DQ_b,$$

where

$$C = \frac{d_2\alpha_1(\beta) - d_1\alpha_2(\beta)}{d_2(n_1+\lambda c_1) + d_1(n_1+\lambda c_2)}, \qquad D = \frac{\alpha_1(\beta)[n_1+\lambda c_2] + \alpha_2(\beta)[n_1+\lambda c_1]}{d_2(n_1+\lambda c_1) + d_1(n_1+\lambda c_2)},$$

$$c_1 = \frac{(n_2-1)(N-n_1)}{(N-n_1-1)}, \qquad c_2 = \frac{n_1^2(n_2-1)}{(N-n_1)(N-n_1-1)},$$

$$d_1 = \frac{(N-n_1-n_2)}{n_2(N-n_1-1)}, \qquad d_2 = \frac{(n_2N^2 - n_2N - n_1^2 - n_1n_2)}{n_2(N-n_1)(N-n_1-1)}.$$

As noted by Wu (1984), $v'(\overline{y}(\beta))$ will always assume non-negative values if and only if

$$\text{(a) } \lambda \geq 0 \qquad \text{and} \qquad \text{(b) } \beta \geq (k-1)/2k. \tag{3.7}$$

In practice, we can usually assume that $k$ is large and $n_1 > n_2$, since if $n_1 \leq n_2$, then it would be more sensible to conduct MLSS (Wu 1984). Consequently, the value of $\beta$ that minimizes (3.6), which we denote as $\beta_{opt}$, is often less than $(k-1)/2k \approx 1/2$ (Wolter 2007). By referring to (3.7), we are therefore presented with a trade-off when selecting a suitable value of $\beta$, where we can either obtain the an efficient estimate the population mean or an unbiased estimate of $\text{Var}(\overline{y}(\beta))$ that always produces non-negative values. Accordingly, Wu (1984) suggested the following strategy to overcome this difficulty:

(i) Use $\overline{y}\,(\beta_{opt})$ and $v'\,(\overline{y}\,(\beta_{opt}))$ when $\beta_{opt} > (k-1)/2k$;

(ii) Use either $\overline{y}((k-1)/2k)$ and $v'(\overline{y}((k-1)/2k))$, or $\overline{y}(1/2)$ and $v'(\overline{y}(1/2))$, when $0.2 \leq \beta_{opt} \leq (k-1)/2k$;

(iii) Otherwise, use $\overline{y}(\beta_{opt})$ and $v'_{+}(\overline{y}(\beta_{opt})) = \max\{\overline{y}(\beta_{opt}), 0\}$.

This strategy seems sensible, except for case (iii), where the variance estimate, which may equal to zero, is just as undesirable as one that assumes negative values (Wolter 2007).

Rana & Singh (1989) suggested that $\beta = (k-1)/k$ and substituted this value into (3.5), before obtaining expressions for the variance of $\overline{y}(\beta = (k-1)/k)$ and an unbiased estimate of this variance. Rana & Singh (1989) noted that this value of $\beta$ is optimal, since the associated variance is minimized, provided that $n_1$ and $n_2$ are not too small. Furthermore, the corresponding variance estimate will always assume non-negative values. Ruiz Espejo (1997) discussed a generalization of Zinger's (1980) approach by considering the above-mentioned case of $\beta = (k-1)/k$.

### 3.1.4 Markov systematic sampling (Sampath & Uthayakumaran 1998)

Markov systematic sampling is a design that exhibits Markovian behaviour and is only applicable when $n$ is even. The associated methodology proceeds as follows:

(i) Divide the population into $n/2$ groups, each of size $2k$, where the $g$th group is given by $G_g = \{U_{p+\delta} | \delta = 1, ..., 2k\}$, with $p = 2(g-1)k$ and $g = 1, ..., n/2$.

(ii) For each $G_g$, assign a stochastic matrix $A_g$ with state space $\{p+\delta, \delta = 1, ..., 2k\}$ and zero diagonal elements, i.e.

$$A_g = \begin{bmatrix} 0 & a_{p+1,p+2} & a_{p+1,p+3} & \cdots & a_{p+1,2gk} \\ a_{p+2,p+1} & 0 & a_{p+2,p+3} & \cdots & a_{p+2,2gk} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{2gk,p+1} & a_{2gk,p+2} & a_{2gk,p+3} & \cdots & 0 \end{bmatrix}_{2k\times 2k}, \quad g = 1, ..., n/2,$$

where $a_{p+\delta,\kappa} = \text{P}(U_\kappa \text{ is selected} | U_{p+\delta} \text{ is selected})$ $(p+\delta \neq \kappa)$ and $\kappa = p+1, ...., p+2k$.

(iii) In a systematic fashion, select $n/2$ units from the $n/2$ groups, i.e. randomly select integer $i$ from 1 to $2k$ and define the sample as $S_i = \{U_i, U_{i+2k}, ..., U_{i+(n/2-1)2k}\}$. Thus, the location of the selected unit from each group $G_g$ is the same.

(iv) The markov systematic sample is then given as the sample in step (iii), as well as an additional sample of size $n/2$, which is obtained by randomly selecting a unit from each group independently using the conditional probabilities in $\mathbf{A}_g$.

Note that each sample will contain distinct sampling units, owing to the zero diagonal elements in matrix $\mathbf{A}_g$.

An unbiased estimate of the population mean is given by the Horvitz-Thompson estimator (Horvitz & Thompson 1952), while an expression for the variance of this estimate, as well as an estimate of the variance, can take the form of that proposed by Sen (1953) and Yates & Grundy (1953). The associated inclusion probabilities, derived by Sampath & Uthayakumaran (1998), can then be substituted into these expressions, so as to find simplified formulae. Note that the corresponding variance estimator may assume negative values.

### 3.1.5 Modified multiple-start linear systematic sampling strategies

We now consider analogues of the single-start strategies in Sections 2.2.1, 2.2.3, 2.2.4 and 2.2.5, which adopt the multiple-start approach, as in Section 3.1.1. Sampath & Ammani (2010) assumed that $n$ is even and the random errors in (2.1) follow a generalized super-population model. Thus, to provide completeness and consistency within this thesis, we will derive relative expressions.

#### Multiple-Start Yates' end corrections (Sampath & Ammani 2010)

By selecting $m$ linear systematic samples of size $n$, using SRS, before applying appropriate weights to the first and last sampling units of each selected sample, we can then remove the linear trend component given in (3.4). The corresponding estimate of the population mean is known as the multiple-start YEC (MYEC) estimator and is denoted as $\overline{y}_{MYEC}$. Accordingly, expressions for estimator $\overline{y}_{MYEC}$ and the expected MSE of this estimator under (2.1), are respectively derived in the next two theorems.

**Theorem 4:** *If we let $V = 2\sum_{h=1}^{m} ih - mk - m$, then the* MYEC *estimator of $\overline{Y}$ with* *random starts ih, for $h = 1, ..., m$ and $ih \in \{1, ..., k\}$, is given by*

$$\overline{y}_{MYEC} = \overline{y}_{MLSS} + \frac{V}{2m^2(n-1)k}\left[\sum_{h=1}^{m}(y_{ih} - y_{ih+(n-1)k})\right]. \tag{3.8}$$

*Proof*: See Appendix.

**Theorem 5:** *Under* (2.1), *the expected* MSE *of* $\overline{y}_{MYEC}$ *is given as*

$$M_{MYEC} = \sigma_l^2 + \frac{\sigma^2(l-1)(lm+1)}{6m^3(n-1)^2 l^2}. \tag{3.9}$$

*Proof*: See Appendix.

When comparing (3.4) to (3.9), we note that the removal of the linear trend component results in a larger error variance component in $M_{MYEC}$, owing to the uneven weighting of the sampling units. Moreover, in the presence of a rough linear trend, the usual YEC estimator is a slightly biased estimate of the population mean (Murthy 1967) and we thus conclude that $\overline{y}_{MYEC}$ is a slightly biased estimate of $\overline{Y}$ under model (2.1).

**Multiple-Start balanced systematic sampling and multiple-start modified systematic sampling (Sampath & Ammani 2010)**

For multiple-start BSS (MBSS) and multiple-start MSS (MMSS), we respectively use the arrangements in Sections 2.2.3 and 2.2.4, before selecting $m$ samples of size $n$ using the multiple-start approach in Section 3.1.1. As such, the MBSS and MMSS sample means are respectively given by

$$\overline{y}_{MBSS} = \frac{1}{nm}\sum_{h=1}^{m}\sum_{j=0}^{(n-2)/2}\left(y_{ih+2jk} + y_{2(j+1)k-ih+1}\right), \quad \text{if } n \text{ is even}$$
$$= \frac{1}{nm}\sum_{h=1}^{m}\left[\sum_{j=0}^{(n-3)/2}\left(y_{ih+2jk} + y_{2(j+1)k-ih+1}\right) + y_{ih+(n-1)k}\right], \quad \text{if } n \text{ is odd}$$

and

$$\overline{y}_{MMSS} = \frac{1}{nm}\sum_{h=1}^{m}\sum_{j=0}^{(n-2)/2}\left(y_{ih+jk} + y_{N-jk-ih+1}\right), \quad \text{if } n \text{ is even}$$
$$= \frac{1}{nm}\sum_{h=1}^{m}\left[\sum_{j=0}^{(n-3)/2}\left(y_{ih+jk} + y_{N-jk-ih+1}\right) + y_{ih+(n-1)k/2}\right], \quad \text{if } n \text{ is odd.}$$

**Theorem 6:** *The* MBSS *sample mean* $\overline{y}_{MBSS}$ *and the* MMSS *sample mean* $\overline{y}_{MMSS}$ *are unbiased estimators of* $\overline{Y}$.

*Proof*: See Appendix.

**Theorem 7:** *The variances of the* MBSS *sample mean* $\overline{y}_{MBSS}$ *and the* MMSS *sample mean* $\overline{y}_{MMSS}$*, along with unbiased estimates of these variances, are respectively given as*

$$\mathrm{V}(\overline{y}_{MBSS}) = \frac{S^2_{BSS}}{m}\left(\frac{k-m}{k}\right), \tag{3.10}$$

$$\mathrm{V}(\overline{y}_{MMSS}) = \frac{S^2_{MSS}}{m}\left(\frac{k-m}{k}\right), \tag{3.11}$$

$$\widehat{\mathrm{V}}(\overline{y}_{MBSS}) = \left(\frac{k-m}{mk}\right)\frac{1}{n^2(m-1)}\sum_{h=1}^{m}\left(BSST_{ih} - \overline{BSST}\right)^2 \tag{3.12}$$

*and*

$$\widehat{\mathrm{V}}(\overline{y}_{MMSS}) = \left(\frac{k-m}{mk}\right)\frac{1}{n^2(m-1)}\sum_{h=1}^{m}\left(MSST_{ih} - \overline{MSST}\right)^2, \tag{3.13}$$

*where* $S^2_{BSS} = \sum_{i=1}^{k}(BSST_i/n - \overline{Y})^2/(k-1)$, $S^2_{MSS} = \sum_{i=1}^{k}(MSST_i/n - \overline{Y})^2/(k-1)$,

$$BSST_i = \begin{cases} \sum_{j=0}^{(n-2)/2}(y_{i+2jk} + y_{2(j+1)k-i+1}), & \text{if } n \text{ is even} \\ \sum_{j=0}^{(n-3)/2}(y_{i+2jk} + y_{2(j+1)k-i+1}) + y_{i+(n-1)k}, & \text{if } n \text{ is odd,} \end{cases}$$

$$MSST_i = \begin{cases} \sum_{j=0}^{(n-2)/2}(y_{i+jk} + y_{N-jk-i+1}), & \text{if } n \text{ is even} \\ \sum_{j=0}^{(n-3)/2}(y_{i+jk} + y_{N-jk-i+1}) + y_{i+(n-1)k/2}, & \text{if } n \text{ is odd,} \end{cases}$$

$\overline{BSST} = \sum_{h=1}^{m} BSST_{ih}/m = n\overline{y}_{MBSS}$ *and* $\overline{MSST} = \sum_{h=1}^{m} MSST_{ih}/m = n\overline{y}_{MMSS}$.

*Proof*: See Appendix.

Sampath & Ammani (2012) compared the performance of estimators $\widehat{\mathrm{V}}(\overline{y}_{MLSS})$, $\widehat{\mathrm{V}}(\overline{y}_{MBSS})$ and $\widehat{\mathrm{V}}(\overline{y}_{MMSS})$, amongst each other. From their numerical study, they concluded that $\widehat{\mathrm{V}}(\overline{y}_{MLSS})$ is the preferred estimator, while the best choice for the number of random starts was $m = 2$.

**Theorem 8:** *Under* (2.1), *the expected* MSEs *of estimators* $\overline{y}_{MBSS}$ *and* $\overline{y}_{MMSS}$*, are given as*

$$M_{MBSS} = M_{MMSS} = \begin{cases} \sigma^2_l, & \text{if } n \text{ is even} \\ \sigma^2_l + b^2(l-1)(lm+1)/(12n^2), & \text{if } n \text{ is odd.} \end{cases} \tag{3.14}$$

*Proof*: See Appendix.

If we compare (3.4) to (3.14), then we note that $M_{MBSS} = M_{MMSS} < M_{MLSS}$, regardless of whether $n$ is even or odd. By referring to (3.14), we clearly see that optimal results are obtained for MBSS and MMSS, when $n = N/ml$ is even. Thus, a multiple-start approach that improves results, for the case when $n$ is odd, is the motivation for the study presented in Section 3.2.

**Multiple-Start diagonal systematic sampling (Subramani & Singh 2014)**

Multiple-Start diagonal systematic sampling (MDSS) is similar to the previous multiple-start designs, where we use the arrangement of matrix $\mathbf{M}$, before selecting $m$ samples of size $n$ using the multiple-start approach in Section 3.1.1. If we denote the corresponding sample mean as $\overline{y}_{MDSS}$, then under (2.1), the expected MSE of $\overline{y}_{MDSS}$ is given as

$$M_{MDSS} = \sigma_l^2 + \frac{b^2(l-1)(lm-n)[n(lm-n)+2]}{12n(lm-1)}.$$

It is difficult to obtain a simple theoretical comparison between $M_{MDSS}$ and $M_{MLSS}$, while Subramani & Singh (2014) notes that $M_{DSS} < M_{MDSS}$. Moreover, if $n = lm$, then this is a special case of optimality for MDSS, i.e. $M_{MDSS} = \sigma_l^2$.

## 3.2 Multiple-start balanced modified systematic sampling

In this section, we propose a modified multiple-start LSS design termed as multiple-start balanced modified systematic sampling (MBMSS). In Section 3.2.1, we discuss the methodology of MBMSS and obtain expressions for the corresponding sample mean, the variance of the sample mean and an unbiased estimate of the variance. For Section 3.2.2, we compare the expected MSE of the MBMSS sample mean, to that of LSS, SRS, STR, MLSS, MBSS, MMSS and the MYEC estimator, under the assumption of a linear trend model averaged over a super-population model. As a result, MBMSS is only optimal for one of the five possible cases of the design and we will thus introduce a linear trend free estimator for the other cases, i.e. MBMSS with end corrections (MBMSSEC) estimator. Finally, we provide some empirical results in Section 3.2.3. Throughout this section we will assume that the required sample size is $nm$, where integer $m$ denotes the number of random starts. Moreover, we will assume that the sampling intervals $k = N/n$ and $l = N/nm$ are integers, i.e. assuming that $N$ is an exact multiple of both $n$ or $nm$, so that we will be conducting sampling linearly.

### 3.2.1 Methodology

The five cases of MBMSS are given as follows:

(A) $n/2$ is an even integer;

(B) $n/2$ is an odd integer;

(C) $n = 3$;

(D) $n \neq 3$ and $(n+1)/2$ is an even integer;

(E) $n \neq 1$ and $(n+1)/2$ is an odd integer.

Note that Cases (A) and (B) correspond to all cases of $n$ being even, while Cases (C) to (E) are associated with all cases of $n > 1$ being odd. The method to select a sample of size $nm$ from a population of size $N$, using MBMSS, consists of the following steps:

(i) Randomly select $m$ integers $(i1, ..., im)$ from the first $k$ integers, using SRS, where $2 \leq m < k$.

(ii) For $h = 1, ..., m$, the sample units chosen for the respective cases will be those elements with indices given by

Case (A): $A1 = ih + 2jk$, $A2 = 2(j+1)k - ih + 1$, $A3 = N + ih - k - 2jk$,

$A4 = N - ih - k - 2jk + 1$, for $j = 0, ..., (n-4)/4$;

Case (B): $B1 = ih + 2jk$, $B2 = N + ih - k - 2jk$, for $j = 0, ..., (n-2)/4$,

$B3 = 2(j+1)k - ih + 1$, $B4 = N - ih - k - 2jk + 1$,

for $j = 0, ..., (n-6)/4$;

Case (C): $C1 = ih$, $C2 = 2k - ih + 1$, $C3 = N - ih + 1$;

Case (D): $D1 = ih + 2jk$, $D2 = 2(j+1)k - ih + 1$, $D3 = N - ih - 2jk + 1$,

for $j = 0, ..., (n-3)/4$,

$D4 = N + ih - 2(j+1)k$, for $j = 0, ..., (n-7)/4$;

Case (E): $E1 = ih + 2jk$, $E2 = 2(j+1)k - ih + 1$, $E3 = N - ih - 2jk + 1$,

$E4 = N + ih - 2(j+1)k$, for $j = 0, ..., (n-5)/4$,

$E5 = ih + (n-1)k/2$.

If $n = 2$, then single-start balanced modified systematic sampling is equivalent to LSS and hence MBMSS is equivalent to MLSS, i.e. the sample units for case (B) reduce to $ih$ and $ih + k$, for $h = 1, ..., m$. We will thus assume that $n > 2$ for the remainder of this section.

Using the above cases, the MBMSS sample mean is denoted as

$$\begin{aligned}
\overline{y}_{MBMSS} &= \frac{1}{nm}\sum_{h=1}^{m}\sum_{j=0}^{(n-4)/4}(y_{A1}+y_{A2}+y_{A3}+y_{A4}), && \text{for Case (A)}\\
&= \frac{1}{nm}\sum_{h=1}^{m}\left[\sum_{j=0}^{(n-2)/4}(y_{B1}+y_{B2})+\sum_{j=0}^{(n-6)/4}(y_{B3}+y_{B4})\right], && \text{for Case (B)}\\
&= \frac{1}{3m}\sum_{h=1}^{m}(y_{C1}+y_{C2}+y_{C3}), && \text{for Case (C)}\\
&= \frac{1}{nm}\sum_{h=1}^{m}\left[\sum_{j=0}^{(n-3)/4}(y_{D1}+y_{D2}+y_{D3})+\sum_{j=0}^{(n-7)/4}y_{D4}\right], && \text{for Case (D)}\\
&= \frac{1}{nm}\sum_{h=1}^{m}\left[\sum_{j=0}^{(n-5)/4}(y_{E1}+y_{E2}+y_{E3}+y_{E4})+y_{E5}\right], && \text{for Case (E).}
\end{aligned}$$

**Theorem 9:** *The* MBMSS *sample mean* $\overline{y}_{MBMSS}$ *is an unbiased estimator of* $\overline{Y}$.

*Proof*: For the respective cases of MBMSS, we denote the $i$th ($i \in \{1, ..., k\}$) balanced modified systematic sample totals by

$$TA_i = \sum_{j=0}^{(n-4)/4}(y_{i+2jk}+y_{2(j+1)k-i+1}+y_{N+i-k-2jk}+y_{N-i-k-2jk+1}),$$

$$TB_i = \sum_{j=0}^{(n-2)/4}(y_{i+2jk}+y_{N+i-k-2jk})+\sum_{j=0}^{(n-6)/4}(y_{2(j+1)k-i+1}+y_{N-i-k-2jk+1}),$$

$$TC_i = y_i + y_{2k-i+1} + y_{N-i+1},$$

$$TD_i = \sum_{j=0}^{(n-3)/4}(y_{i+2jk}+y_{2(j+1)k-i+1}+y_{N-i-2jk+1})+\sum_{j=0}^{(n-7)/4}y_{N+i-2(j+1)k},$$

and

$$TE_i = \sum_{j=0}^{(n-5)/4}(y_{i+2jk}+y_{2(j+1)k-i+1}+y_{N-i-2jk+1}+y_{N+i-2(j+1)k})+y_{i+(n-1)k/2}.$$

Let us assume an indicator variable, given by

$$I_i = \begin{cases} 1, & \text{if unit } y_{ih} \text{ is in the sample;}\\ 0, & \text{otherwise.}\end{cases}$$

For $j = A, ..., E$, if we assume that the $Tj_i$'s are fixed but unknown constants, then

$$\begin{aligned}
\mathrm{E}(\overline{y}_{MBMSS}) &= \mathrm{E}\left[\frac{1}{nm}\sum_{i=1}^{k}I_i(Tj_i)\right]\\
&= \frac{1}{nm}\sum_{i=1}^{k}\mathrm{E}(I_i)Tj_i = \frac{1}{nm}\sum_{i=1}^{k}\left(\frac{m}{k}\right)Tj_i = \frac{m}{nmk}\sum_{i=1}^{k}Tj_i = \frac{Y.}{nk} = \overline{Y},
\end{aligned}$$

since we are selecting $m$ samples from the $k$ possible samples, using SRS.

The single-start sample means ($Tj_i/n$, for $j \in \{A, ..., E\}$ and $i = 1, ..., k$) can now be viewed as population units. Remembering that SRS involves the random selection of $n$ sampling units from $N$ possible sampling units, the variance of the sample mean $\overline{y}_{SRS}$ is given by (1.5). Thus, by replacing $y_q$, $N$ and $n$ in (1.5) by $Tj_i/n$, $k$ and $m$ respectively, we obtain the variance of $\overline{y}_{MBMSS}$, which is written as

$$\mathrm{V}(\overline{y}_{MBMSS}) = \frac{S_T^2}{m}\left(\frac{k-m}{k}\right), \tag{3.15}$$

where $S_T^2 = \sum_{i=1}^{k}(Tj_i/n - \overline{Y})^2/(k-1)$, such that the replacement of $y_q$ and $N$ in $\overline{Y} = \sum_{q=1}^{N} y_q/N$, by $Tj_i/n$ and $k$ respectively, results in $\sum_{i=1}^{k} Tj_i/nk = \overline{Y}$.

**Theorem 10:** *An unbiased estimator of* (3.15) *is given by*

$$\widehat{\mathrm{V}}(\overline{y}_{MBMSS}) = \left(\frac{k-m}{mk}\right)\frac{1}{n^2(m-1)}\sum_{h=1}^{m}\left(Tj_{ih} - \overline{T}\right)^2, \tag{3.16}$$

*where $Tj_{ih} \in \{Tj_1, ..., Tj_k\}$ and $\overline{T} = \sum_{h=1}^{m} Tj_{ih}/m = n\overline{y}_{MBMSS}$ denotes the average of all the balanced modified systematic sample totals selected.*

*Proof*: See Appendix.

### 3.2.2 Expected Mean Square Error Comparisons

If we consider case (A) of MBMSS, such that

$$e_A = \sum_{j=0}^{(n-4)/4}\left(e_{i+2jk} + e_{2(j+1)k-i+1} + e_{N+i-k-2jk} + e_{N-i-k-2jk+1}\right)$$

denotes the random error associated with $i$th ($i \in \{1, ..., k\}$) balanced modified systematic sample, then substituting (2.1) into $TA_i$ results in

$$TA_i = \sum_{j=0}^{(n-4)/4}[4a + b(2N+2)] + e_A = an + bn\left(\frac{N+1}{2}\right) + e_A.$$

An expression for the expected MSE of $\overline{y}_{MBMSS}$ is obtained by taking the expectation of (3.15) and then substituting $TA_i$ and $\overline{Y}$ into that expression, given by

$$\begin{aligned} M_{MBMSS} &= \frac{(k-m)}{mk(k-1)}\sum_{i=1}^{k}\mathcal{E}\left(\frac{TA_i}{n} - \overline{Y}\right)^2 \\ &= \frac{(k-m)}{mk\,(k-1)}\sum_{i=1}^{k}\mathcal{E}\left[a + b\left(\frac{N+1}{2}\right) + \frac{e_A}{n} - \left(a + \frac{b\,(N+1)}{2} + \overline{\overline{e}}\right)\right]^2 \\ &= \frac{(k-m)}{mk(k-1)}\sum_{i=1}^{k}\mathcal{E}\left(\frac{e_A^2}{n^2} - \frac{2e_A\overline{\overline{e}}}{n} + \overline{\overline{e}}^2\right). \end{aligned} \tag{3.17}$$

Now, since there are $n$ terms in $e_A$ and $N$ terms in $\overline{\overline{e}}$, it follows that

$$\mathcal{E}\left(e_A^2\right) = n\sigma^2, \qquad \mathcal{E}\left(e_A\overline{\overline{e}}\right) = \frac{1}{N}\mathcal{E}[e_A(e_1 + ... + e_N)] = \frac{n\sigma^2}{N},$$
$$\mathcal{E}\left(\overline{\overline{e}}^2\right) = \frac{1}{N^2}\left[\sum_{q=1}^{N}\mathcal{E}\left(e_q^2\right) + \sum_{z=1}^{N}\sum_{q\neq z}^{N}\mathcal{E}\left(e_q e_z\right)\right] = \frac{\sigma^2}{N}.$$

Remembering that $k = lm$, we then substitute these expressions into (3.17) to obtain

$$M_{MBMSS} = \frac{(k-m)}{mk(k-1)}\sum_{i=1}^{k}\left(\frac{\sigma^2}{n} - \frac{2\sigma^2}{N} + \frac{\sigma^2}{N}\right) = \frac{(k-m)\sigma^2}{mN} = \frac{(l-1)\sigma^2}{N} = \sigma_l^2.$$

Similarly, we can use the above method for the other cases, such that

$$M_{MBMSS} = \begin{cases} \sigma_l^2, & \text{for Case (A)} \\ \sigma_l^2 + b^2(l-1)(lm+1)/(3n^2), & \text{for Case (B)} \\ \sigma_l^2 + b^2(l-1)(lm+1)/(12n^2), & \text{for Cases (C) to (E).} \end{cases} \tag{3.18}$$

By comparing (3.18) to (3.1), (3.2), (3.3) and (3.9) for Case (A), it follows that $M_{MBMSS}$ is less than $M_{LSS}$, $M_{SRS}$, $M_{STR}$ and $M_{MYEC}$, while simple theoretical comparisons are unobtainable for the other cases. Moreover, if we compare (3.18) to (3.4), then we note that $M_{MBMSS} < M_{MLSS}$ for all cases. Finally, the comparison of (3.18) to (3.14) results in $M_{MBMSS} = M_{MBSS} = M_{MMSS}$ for cases (A), (C), (D) and (E), while $M_{MBMSS} > M_{MBSS} = M_{MMSS}$ for case (B).

Clearly the linear trend component is only removed for case (A) in (3.18). To remove the linear trend component for the other cases, we can apply weights to the first and last sampling units of each of the $m$ selected samples of size $n$. Accordingly, the resulting estimator and the corresponding expected MSE are respectively given in the next two theorems.

**Theorem 11:** *The* MBMSSEC *estimator of $\overline{Y}$ with random starts $ih$, for $h = 1, ..., m$ and $ih \in \{1, ..., k\}$, is given by*

$$\overline{y}_{MBMSSEC} = \overline{y}_{MBMSS} + W$$

*where*

$$\begin{aligned} W &= \frac{V\sum_{h=1}^{m}\left(y_{ih} - y_{N+ih-k}\right)}{nm^2(N-k)}, && \text{for Case (B)} \\ &= \frac{-V\sum_{h=1}^{m}\left(y_{ih} - y_{N-ih+1}\right)}{2nm\sum_{h=1}^{m}\left(N - 2ih + 1\right)}, && \text{for Cases (C) and (D)} \\ &= \frac{V\sum_{h=1}^{m}\left(y_{ih} - y_{N-ih+1}\right)}{2nm\sum_{h=1}^{m}\left(N - 2ih + 1\right)}, && \text{for Case (E).} \end{aligned}$$

*Proof*: See Appendix.

**Theorem 12:** *Under* (2.1), *the expected* MSE *of* $\overline{y}_{MBMSSEC}$ *is given as*

$$M_{MBMSSEC} = \sigma_l^2 + \frac{2\sigma^2(l-1)(lm+1)}{3m^3n^2(n-1)^2l^2}, \qquad \text{for Case (B)}$$

$$= \sigma_l^2 + \mathrm{E}\left\{\frac{\sigma^2V^2}{2n^2m\left[\sum_{h=1}^{m}(N-2ih+1)\right]^2}\right\}, \qquad \text{for Cases (C) to (E).}$$

*Proof*: Similar to Theorem 5.

Again, by comparing $M_{MBMSSEC}$ to all the previous expected MSEs, we note that simple theoretical comparisons are difficult to obtain and we will thus resort to some numerical results in the next section. However, it is easily deduced that $M_{MBSS} = M_{MMSS} < M_{MBMSSEC} < M_{MYEC}$ for Case (B). Moreover, just as in the case of the MYEC estimator, we note that $\overline{y}_{MBMSSEC}$ is a slightly biased estimator of $\overline{Y}$ in the presence of a rough linear trend.

### 3.2.3 Empirical Comparisons

An evaluation of the performance of estimator $\overline{y}_{MBMSSEC}$ will now be carried out by means of three simulation studies. To obtain the expected MSE of each estimator, Monte Carlo simulations are applied by using the statistical software package R, where 10 000 finite populations are simulated, before averaging the MSEs over the 10 000 populations. The relative expected MSE of each comparative estimator, with respect to that of $\overline{y}_{MBMSSEC}$, is denoted by $R_\alpha = 100 \times M_{MBMSSEC}/M_\alpha(\%)$, where $\alpha =$ LSS, SRS, STR, MLSS, MYEC, MBSS, MMSS and MBMSS. Without loss of generality, we assume that the $e_q$'s are iid $N(0,1)$ random variables and let $a = 5$, as all expected MSE expressions previously derived are shown to be independent of $a$.

For the first simulation study, we consider Case (B) and assign arbitrary values of $b = 0.5, 1, 2$ and 4, while varying $n$, $m$ and $l$. The corresponding relative expected MSEs are presented in Tables 3.1 to 3.4. From Tables 3.1 to 3.4, we note that only estimators $\overline{y}_{MBSS}$ and $\overline{y}_{MMSS}$ are marginally subject to less error than that of $\overline{y}_{MBMSSEC}$. If we choose estimator $\overline{y}_{MBMSSEC}$ over estimator $\overline{y}_{LSS}$, $\overline{y}_{SRS}$ or $\overline{y}_{MLSS}$, then we obtain more favourable results as $n$, $m$, $l$ and/or $b$ increases. Likewise, if we choose estimator $\overline{y}_{MBMSSEC}$ over estimator $\overline{y}_{STR}$, then we see that the relative results are improved as $l$ and/or $b$ increases, while results seem to remain constant as $n$ or $m$ varies. Finally, by selecting estimator $\overline{y}_{MBMSSEC}$ over estimator $\overline{y}_{MBMSS}$, we note that relative results are better as $m$, $l$ and/or $b$ increase, while results deteriorate as $n$ increases. Thus, $M_{MBMSS} \to M_{MBMSSEC}$ as $n \to \infty$, provided that $m$, $l$ and $b$ are relatively small.

Table 3.1: Simulated relative expected mean square errors for populations exhibiting linear trend under Case B ($b$=0.5).

| | | $l = 2$ | | | | $l = 4$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $m$ | | $n = 6$ | $n = 34$ | $n = 130$ | $n = 258$ | $n = 6$ | $n = 34$ | $n = 130$ | $n = 258$ |
| | $R_{LSS}$ | 40.16 | 10.53 | 02.97 | 01.54 | 16.67 | 03.34 | 00.92 | 00.47 |
| | $R_{SRS}$ | 07.50 | 00.26 | 00.02 | $< 00.01$ | 02.00 | 00.06 | $< 00.01$ | $< 00.01$ |
| | $R_{STR}$ | 89.64 | 88.89 | 88.76 | 89.32 | 70.77 | 70.98 | 70.72 | 70.93 |
| 2 | $R_{MLSS}$ | 28.85 | 06.60 | 01.81 | 00.93 | 10.06 | 01.93 | 00.51 | 00.26 |
| | $R_{MBSS}$ | 100.99 | 101.13 | 100.10 | 100.18 | 100.25 | 101.43 | 100.14 | 99.78 |
| | $R_{MMSS}$ | 99.66 | 99.90 | 99.66 | 99.90 | 99.50 | 101.45 | 99.82 | 99.52 |
| | $R_{MBMSS}$ | 78.47 | 95.32 | 98.50 | 99.41 | 50.37 | 85.21 | 95.62 | 97.73 |
| | $R_{LSS}$ | 31.29 | 07.33 | 01.99 | 01.02 | 11.92 | 02.29 | 00.62 | 00.31 |
| | $R_{SRS}$ | 03.49 | 00.11 | 00.01 | $< 00.01$ | 00.91 | 00.03 | $< 00.01$ | $< 00.01$ |
| | $R_{STR}$ | 89.26 | 88.73 | 88.29 | 89.03 | 71.66 | 70.52 | 72.10 | 71.72 |
| 3 | $R_{MLSS}$ | 16.09 | 03.24 | 00.87 | 00.44 | 04.93 | 00.89 | 00.24 | 00.12 |
| | $R_{MBSS}$ | 100.53 | 99.61 | 99.70 | 100.09 | 100.72 | 101.07 | 101.46 | 101.87 |
| | $R_{MMSS}$ | 99.55 | 100.03 | 99.48 | 100.61 | 101.29 | 101.20 | 101.60 | 100.44 |
| | $R_{MBMSS}$ | 63.65 | 90.94 | 97.42 | 98.67 | 31.40 | 72.13 | 91.32 | 94.82 |

In the second simulation study, we consider odd values of $n$ (i.e. Cases (C) to (E)) and assign arbitrary values of $b = 0.5, 1, 2$ and 4, while varying $n$, $m$ and $l$. The corresponding relative expected MSEs are presented in Tables 3.5 to 3.8. From Tables 3.5 to 3.8, we can easily see that $\overline{y}_{MBMSSEC}$ is preferred over all other estimators in the study. As in the first simulation study, we obtain similar results. However, estimator $\overline{y}_{MBMSSEC}$ is now favourable over estimators $\overline{y}_{MBSS}$ and $\overline{y}_{MMSS}$. Furthermore, the performance of estimators $\overline{y}_{MBSS}$, $\overline{y}_{MMSS}$ and $\overline{y}_{MBMSS}$ are relatively similar. Hence, $M_{MBSS}$, $M_{MMSS}$ and $M_{MBMSS} \to M_{MBMSSEC}$ as $n \to \infty$, provided that $m$, $l$ and $b$ are relatively small.

In the final simulation study, we compare estimators $\overline{y}_{MBMSSEC}$ and $\overline{y}_{MYEC}$. Since the

Table 3.2: Simulated relative expected mean square errors for populations exhibiting linear trend under Case B ($b$=1).

| | | $l=2$ | | | | $l=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $m$ | | $n=6$ | $n=34$ | $n=130$ | $n=258$ | $n=6$ | $n=34$ | $n=130$ | $n=258$ |
| | $R_{LSS}$ | 14.64 | 02.89 | 00.74 | 00.39 | 04.73 | 00.85 | 00.23 | 00.12 |
| | $R_{SRS}$ | 01.97 | 00.07 | $<00.01$ | $<00.01$ | 00.50 | 00.02 | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 67.00 | 67.47 | 64.46 | 68.12 | 37.26 | 36.54 | 37.72 | 37.78 |
| 2 | $R_{MLSS}$ | 09.15 | 01.77 | 00.44 | 00.24 | 02.68 | 00.47 | 00.13 | 00.07 |
| | $R_{MBSS}$ | 102.81 | 101.02 | 102.04 | 103.64 | 102.13 | 99.66 | 101.46 | 100.95 |
| | $R_{MMSS}$ | 101.07 | 102.60 | 101.24 | 100.65 | 100.25 | 100.59 | 102.17 | 100.35 |
| | $R_{MBMSS}$ | 48.25 | 85.63 | 95.07 | 98.06 | 20.10 | 58.55 | 84.57 | 91.46 |
| | $R_{LSS}$ | 09.61 | 01.90 | 00.51 | 00.27 | 03.27 | 00.59 | 00.15 | 00.08 |
| | $R_{SRS}$ | 00.90 | 00.03 | $<00.01$ | $<00.01$ | 00.23 | 00.01 | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 66.21 | 66.31 | 66.56 | 68.79 | 37.99 | 37.93 | 36.36 | 37.47 |
| 3 | $R_{MLSS}$ | 04.52 | 00.84 | 00.22 | 00.11 | 01.28 | 00.23 | 00.06 | 00.03 |
| | $R_{MBSS}$ | 99.31 | 102.63 | 100.36 | 101.19 | 99.83 | 99.08 | 104.33 | 99.50 |
| | $R_{MMSS}$ | 103.60 | 102.37 | 101.03 | 99.95 | 99.52 | 103.29 | 106.36 | 100.51 |
| | $R_{MBMSS}$ | 30.06 | 70.66 | 89.80 | 94.07 | 10.52 | 40.08 | 70.59 | 82.68 |

expected MSEs of both estimators are trend free, we assign an arbitrary value of $b=4$, while varying $n$, $m$ and $l$. In addition, we will only consider Cases (C) to (E), as it can be theoretically shown that $M_{MBMSSEC} < M_{MYEC}$ for Case (B). The corresponding relative expected MSEs are presented in Table 3.9. From the results, we conclude that estimator $\overline{y}_{MBMSSEC}$ is only preferred when $m$, $n$ and $l$ are small, otherwise there is a very small reduction in estimation error when choosing estimator $\overline{y}_{MBMSSEC}$ over estimator $\overline{y}_{MYEC}$.

### 3.2.4 Concluding Remarks

We have proposed a modified multiple-start sampling design (MBMSS), which depends on five cases and provides an unbiased estimator of the sampling variance. For Case (A), theoretical expected MSE comparisons in Section 3.2.2 suggest that MBMSS is more efficient than LSS, SRS, STR, MLSS and MYEC, while equally efficient to MBSS and MMSS, i.e.

Table 3.3: Simulated relative expected mean square errors for populations exhibiting linear trend under Case B ($b$=2).

| | | $l=2$ | | | | $l=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $m$ | | $n=6$ | $n=34$ | $n=130$ | $n=258$ | $n=6$ | $n=34$ | $n=130$ | $n=258$ |
| | $R_{LSS}$ | 03.88 | 00.75 | 00.20 | 00.10 | 01.24 | 00.22 | 00.06 | 00.03 |
| | $R_{SRS}$ | 00.47 | 00.02 | $<00.01$ | $<00.01$ | 00.13 | $<00.01$ | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 32.33 | 34.11 | 34.11 | 33.31 | 13.13 | 13.06 | 13.07 | 12.88 |
| 2 | $R_{MLSS}$ | 02.33 | 00.45 | 00.12 | 00.06 | 00.69 | 00.12 | 00.03 | 00.02 |
| | $R_{MBSS}$ | 99.25 | 100.23 | 103.15 | 99.11 | 100.26 | 100.30 | 102.68 | 101.92 |
| | $R_{MMSS}$ | 99.96 | 102.67 | 103.15 | 99.11 | 100.26 | 100.30 | 102.68 | 101.92 |
| | $R_{MBMSS}$ | 17.94 | 55.77 | 82.72 | 90.47 | 05.91 | 26.44 | 57.62 | 72.02 |
| | $R_{LSS}$ | 02.67 | 00.49 | 00.13 | 00.06 | 00.82 | 00.15 | 00.04 | 00.02 |
| | $R_{SRS}$ | 00.22 | 00.01 | $<00.01$ | $<00.01$ | 00.06 | $<00.01$ | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 32.85 | 33.42 | 34.21 | 33.04 | 12.93 | 13.05 | 13.22 | 13.19 |
| 3 | $R_{MLSS}$ | 01.16 | 00.21 | 00.06 | 00.03 | 00.32 | 00.06 | 00.01 | 00.01 |
| | $R_{MBSS}$ | 103.19 | 103.70 | 104.64 | 99.41 | 100.74 | 102.61 | 101.63 | 101.09 |
| | $R_{MMSS}$ | 100.69 | 104.15 | 102.99 | 100.92 | 100.52 | 99.30 | 101.51 | 99.07 |
| | $R_{MBMSS}$ | 09.70 | 37.70 | 70.21 | 81.90 | 02.78 | 14.06 | 39.11 | 56.17 |

we obtain a complete removal of the linear trend component in the corresponding expected MSEs resulting in MBMSS, MBSS and MMSS being optimal.

To remove the linear trend component for the other cases, we proposed an estimator (MBMSSEC) which applies weights to the first and last sampling units of each selected single-start balanced modified systematic sample. The MBMSSEC estimator provides a slightly biased estimate of the population mean, as well as an inflated error variance component in the corresponding expected MSE, owing to the uneven weighting of the sampling units.

For Case (B), the comparisons in Sections 3.2.2 and 3.2.3 indicate that the MBMSSEC estimator is subject to less error when compared to that of LSS, SRS, STR, MLSS and MBMSS, while marginally subject to more error than that of MBSS and MMSS, i.e. for Case (B), MBSS and MMSS are linear trend free sampling designs that exhibit a minimum expected error variance component for the corresponding expected MSEs. Furthermore,

Table 3.4: Simulated relative expected mean square errors for populations exhibiting linear trend under Case B ($b$=4).

| $m$ | | $l=2$ | | | | $l=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $n=6$ | $n=34$ | $n=130$ | $n=258$ | $n=6$ | $n=34$ | $n=130$ | $n=258$ |
| | $R_{LSS}$ | 01.05 | 00.19 | 00.05 | 00.02 | 00.31 | 00.06 | 00.01 | 00.01 |
| | $R_{SRS}$ | 00.13 | $<00.01$ | $<00.01$ | $<00.01$ | 00.03 | $<00.01$ | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 11.31 | 11.22 | 11.63 | 11.14 | 03.58 | 03.68 | 03.58 | 03.60 |
| 2 | $R_{MLSS}$ | 00.64 | 00.11 | 00.03 | 00.01 | 00.17 | 00.03 | 00.01 | $<00.01$ |
| | $R_{MBSS}$ | 101.97 | 100.53 | 106.96 | 105.08 | 100.30 | 101.41 | 100.56 | 101.02 |
| | $R_{MMSS}$ | 105.20 | 99.96 | 104.45 | 103.33 | 100.44 | 100.01 | 101.53 | 100.37 |
| | $R_{MBMSS}$ | 05.47 | 24.85 | 55.71 | 71.91 | 01.52 | 08.27 | 24.93 | 40.26 |
| | $R_{LSS}$ | 00.69 | 00.12 | 00.03 | 00.02 | 00.21 | 00.04 | 00.01 | 00.01 |
| | $R_{SRS}$ | 00.06 | $<00.01$ | $<00.01$ | $<00.01$ | 00.01 | $<00.01$ | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 11.08 | 10.97 | 10.65 | 11.29 | 03.58 | 03.60 | 03.66 | 03.61 |
| 3 | $R_{MLSS}$ | 00.30 | 00.05 | 00.01 | $<00.01$ | 00.08 | 00.01 | $<00.01$ | $<00.01$ |
| | $R_{MBSS}$ | 102.75 | 104.57 | 101.14 | 100.46 | 100.23 | 101.03 | 100.61 | 104.98 |
| | $R_{MMSS}$ | 102.62 | 104.68 | 100.02 | 100.88 | 101.13 | 100.65 | 100.04 | 103.94 |
| | $R_{MBMSS}$ | 02.59 | 12.98 | 35.58 | 54.64 | 00.70 | 03.91 | 13.57 | 13.57 |

for Cases (C) to (E), the MBMSSEC estimator is subject to less error when compared to all the other estimators, apart from the MYEC estimator. The simulation study in Section 3.2.3 suggests that the expected MSE of the MBMSSEC estimator is smaller than that of the MYEC estimator, for Cases (C) to (E), provided that $n$, $m$ and $l$ are small. Otherwise, both estimators are shown to be approximately subject to the same amount of error and one may opt to use the MYEC estimator, owing to its simplicity. A summary of the suggested designs under various conditions is given in Table 3.10. Note that the third column in Table 3.10 represents a trade-off between the MBMSSEC and the MYEC estimators, where preference is either given to obtaining minimum expected MSE of the sample mean, or using a sample mean that is simpler to construct.

In this chapter, we have discussed the the first of two disadvantages of systematic sampling, i.e. the problem of estimating the sampling variance when conducting systematic sampling with a single random start. Also, we reviewed relevant modifications of the usual

Table 3.5: Simulated relative expected mean square errors for populations exhibiting linear trend under Cases C to E ($b$=0.5).

| | | $l=2$ | | | | $l=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $m$ | | $n=3$ | $n=35$ | $n=125$ | $n=255$ | $n=3$ | $n=35$ | $n=125$ | $n=255$ |
| | $R_{LSS}$ | 52.84 | 10.13 | 03.08 | 01.50 | 27.57 | 03.34 | 00.96 | 00.47 |
| | $R_{SRS}$ | 22.23 | 00.24 | 00.02 | $<00.01$ | 07.48 | 00.06 | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 85.45 | 86.67 | 88.52 | 86.33 | 70.08 | 70.72 | 71.09 | 70.21 |
| 2 | $R_{MLSS}$ | 39.36 | 06.33 | 01.88 | 00.90 | 18.07 | 01.87 | 00.53 | 00.26 |
| | $R_{MBSS}$ | 81.96 | 98.40 | 99.99 | 99.49 | 66.45 | 95.53 | 97.54 | 97.35 |
| | $R_{MMSS}$ | 81.92 | 99.43 | 99.05 | 97.76 | 66.60 | 94.78 | 98.15 | 97.37 |
| | $R_{MBMSS}$ | 88.19 | 99.09 | 99.72 | 99.87 | 67.89 | 96.31 | 98.90 | 99.46 |
| | $R_{LSS}$ | 49.17 | 07.30 | 02.09 | 01.03 | 21.78 | 02.26 | 00.63 | 00.31 |
| | $R_{SRS}$ | 12.42 | 00.11 | 00.01 | $<00.01$ | 03.38 | 00.03 | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 91.51 | 90.71 | 88.34 | 88.66 | 71.03 | 71.65 | 70.39 | 71.06 |
| 3 | $R_{MLSS}$ | 27.24 | 03.22 | 00.90 | 00.44 | 09.43 | 00.89 | 00.25 | 00.12 |
| | $R_{MBSS}$ | 79.87 | 96.84 | 99.55 | 98.12 | 47.59 | 92.91 | 97.64 | 96.96 |
| | $R_{MMSS}$ | 78.64 | 99.91 | 98.71 | 97.55 | 48.26 | 90.62 | 97.22 | 98.32 |
| | $R_{MBMSS}$ | 76.38 | 97.84 | 99.29 | 99.99 | 49.03 | 90.29 | 97.44 | 99.13 |

Table 3.6: Simulated relative expected mean square errors for populations exhibiting linear trend under Cases C to E ($b$=1).

| | | $l = 2$ | | | | $l = 4$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $m$ | | $n = 3$ | $n = 35$ | $n = 125$ | $n = 255$ | $n = 3$ | $n = 35$ | $n = 125$ | $n = 255$ |
| | $R_{LSS}$ | 24.90 | 02.66 | 00.78 | 00.37 | 08.85 | 00.86 | 00.24 | 00.19 |
| | $R_{SRS}$ | 07.18 | 00.06 | $< 00.01$ | $< 00.01$ | 01.96 | 00.02 | $< 00.01$ | $< 00.01$ |
| | $R_{STR}$ | 67.59 | 64.08 | 65.80 | 63.59 | 36.75 | 37.78 | 37.41 | 37.89 |
| 2 | $R_{MLSS}$ | 16.87 | 01.63 | 00.47 | 00.22 | 05.19 | 00.48 | 00.13 | 00.07 |
| | $R_{MBSS}$ | 66.24 | 92.74 | 97.32 | 96.91 | 32.88 | 85.25 | 94.24 | 96.91 |
| | $R_{MMSS}$ | 65.40 | 95.21 | 96.46 | 98.69 | 32.77 | 86.89 | 92.76 | 96.39 |
| | $R_{MBMSS}$ | 63.64 | 94.62 | 98.26 | 99.68 | 33.34 | 84.56 | 95.36 | 97.66 |
| | $R_{LSS}$ | 18.73 | 01.93 | 00.53 | 00.26 | 06.28 | 00.57 | 00.16 | 00.07 |
| | $R_{SRS}$ | 03.51 | 00.03 | $< 00.01$ | $< 00.01$ | 00.90 | 00.01 | $< 00.01$ | $< 00.01$ |
| | $R_{STR}$ | 69.09 | 68.25 | 67.10 | 65.61 | 37.85 | 37.69 | 37.76 | 37.56 |
| 3 | $R_{MLSS}$ | 08.99 | 00.83 | 00.23 | 00.11 | 02.52 | 00.22 | 00.06 | 00.03 |
| | $R_{MBSS}$ | 49.30 | 89.36 | 96.29 | 97.70 | 18.91 | 72.61 | 91.63 | 97.49 |
| | $R_{MMSS}$ | 46.18 | 90.41 | 96.88 | 96.75 | 18.86 | 73.35 | 91.06 | 96.09 |
| | $R_{MBMSS}$ | 47.15 | 90.60 | 97.38 | 98.20 | 18.90 | 73.56 | 91.15 | 95.79 |

systematic sampling design found in literature as well as a suggested modified systematic sampling design that address this problem, all under the assumption of linear trend among the population units. The results from this chapter suggest that values of the sample size, sampling interval and/or the number of random starts (if a multiple start sampling design is applied) needs to be considered, before selecting an appropriate modified systematic sampling design in the presence of linear trend to address the problem at hand, where Table 3.10 provides us with the most suitable modified systematic sampling design under various scenarios of $m$, $n$ and $l$. In the next chapter, we will investigate the second of the two shortcomings of systematic sampling, i.e. if the population size is not a multiple of the sample size, resulting in sample sizes that vary, or fixed sample sizes that are greater than required when conducting LSS.

Table 3.7: Simulated relative expected mean square errors for populations exhibiting linear trend under Cases C to E ($b$=2).

| | | $l=2$ | | | | $l=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $m$ | | $n=3$ | $n=35$ | $n=125$ | $n=255$ | $n=3$ | $n=35$ | $n=125$ | $n=255$ |
| | $R_{LSS}$ | 07.72 | 00.71 | 00.20 | 00.10 | 02.44 | 00.21 | 00.06 | 00.03 |
| | $R_{SRS}$ | 01.93 | 00.02 | $<$ 00.01 | $<$ 00.01 | 00.50 | $<$ 00.01 | $<$ 00.01 | $<$ 00.01 |
| | $R_{STR}$ | 33.45 | 33.42 | 33.58 | 32.49 | 13.06 | 12.90 | 12.77 | 12.95 |
| 2 | $R_{MLSS}$ | 04.85 | 00.43 | 00.12 | 00.06 | 01.38 | 00.12 | 00.03 | 00.02 |
| | $R_{MBSS}$ | 31.15 | 83.42 | 95.23 | 97.32 | 11.21 | 59.32 | 81.74 | 88.91 |
| | $R_{MMSS}$ | 31.57 | 82.94 | 95.81 | 96.63 | 11.32 | 59.12 | 82.97 | 92.55 |
| | $R_{MBMSS}$ | 32.06 | 83.52 | 94.61 | 98.22 | 11.03 | 58.29 | 83.03 | 91.21 |
| | $R_{LSS}$ | 05.23 | 00.47 | 00.13 | 00.06 | 01.67 | 00.14 | 00.04 | 00.02 |
| | $R_{SRS}$ | 00.87 | 00.01 | $<$ 00.01 | $<$ 00.01 | 00.23 | $<$ 00.01 | $<$ 00.01 | $<$ 00.01 |
| | $R_{STR}$ | 33.22 | 33.19 | 32.54 | 32.36 | 13.29 | 12.80 | 13.02 | 13.36 |
| 3 | $R_{MLSS}$ | 02.32 | 00.20 | 00.06 | 00.03 | 00.65 | 00.05 | 00.02 | 00.01 |
| | $R_{MBSS}$ | 17.55 | 72.04 | 89.76 | 92.40 | 05.53 | 39.87 | 69.63 | 86.17 |
| | $R_{MMSS}$ | 17.67 | 71.35 | 87.14 | 92.08 | 05.57 | 39.76 | 70.56 | 84.78 |
| | $R_{MBMSS}$ | 17.55 | 71.09 | 89.60 | 95.60 | 05.52 | 39.24 | 71.04 | 84.47 |

Table 3.8: Simulated relative expected mean square errors for populations exhibiting linear trend under Cases C to E ($b$=4).

| | | $l=2$ | | | | $l=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $m$ | | $n=3$ | $n=35$ | $n=125$ | $n=255$ | $n=3$ | $n=35$ | $n=125$ | $n=255$ |
| | $R_{LSS}$ | 02.11 | 00.18 | 00.05 | 00.02 | 00.62 | 00.05 | 00.15 | 00.01 |
| | $R_{SRS}$ | 00.50 | $<00.01$ | $<00.01$ | $<00.01$ | 00.12 | $<00.01$ | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 11.52 | 11.25 | 10.89 | 11.21 | 03.62 | 03.55 | 03.68 | 03.58 |
| 2 | $R_{MLSS}$ | 01.28 | 00.11 | 00.03 | 00.01 | 00.35 | 00.03 | 00.01 | $<00.01$ |
| | $R_{MBSS}$ | 10.25 | 56.32 | 81.60 | 90.40 | 03.05 | 26.18 | 57.09 | 73.16 |
| | $R_{MMSS}$ | 10.36 | 56.60 | 81.24 | 90.43 | 03.02 | 26.52 | 56.12 | 72.04 |
| | $R_{MBMSS}$ | 10.70 | 57.08 | 82.16 | 90.47 | 03.03 | 26.84 | 56.02 | 73.47 |
| | $R_{LSS}$ | 01.34 | 00.12 | 00.03 | 00.02 | 00.42 | 00.04 | 00.01 | $<00.01$ |
| | $R_{SRS}$ | 00.21 | $<00.01$ | $<00.01$ | $<00.01$ | 00.06 | $<00.01$ | $<00.01$ | $<00.01$ |
| | $R_{STR}$ | 10.88 | 11.19 | 11.43 | 11.33 | 03.63 | 03.66 | 03.66 | 03.60 |
| 3 | $R_{MLSS}$ | 00.58 | 00.05 | 00.01 | 00.01 | 00.16 | 00.01 | $<00.01$ | $<00.01$ |
| | $R_{MBSS}$ | 04.95 | 38.08 | 72.24 | 83.07 | 01.42 | 14.40 | 37.80 | 54.83 |
| | $R_{MMSS}$ | 04.96 | 37.98 | 71.63 | 81.86 | 01.42 | 14.57 | 37.76 | 55.54 |
| | $R_{MBMSS}$ | 04.96 | 38.59 | 69.58 | 82.08 | 01.42 | 14.50 | 37.72 | 54.74 |

Table 3.9: Simulated relative expected mean square errors of the MYEC sample mean, with respect to that of the MBMSSEC sample mean, for populations exhibiting linear trend under Cases C to E.

| | $m=2$ | | | $m=3$ | | |
|---|---|---|---|---|---|---|
| $n$ | $l=2$ | $l=4$ | $l=8$ | $l=2$ | $l=4$ | $l=8$ |
| 3 | 94.24 | 94.38 | 94.71 | 95.94 | 96.26 | 96.41 |
| 5 | 96.97 | 97.27 | 97.42 | 98.10 | 98.23 | 98.30 |
| 7 | 98.06 | 98.25 | 98.34 | 98.78 | 98.87 | 98.91 |
| 13 | 99.07 | 99.17 | 99.21 | 99.42 | 99.46 | 99.48 |
| 15 | 99.21 | 99.29 | 99.33 | 99.51 | 99.54 | 99.55 |
| 29 | 99.62 | 99.65 | 99.67 | 99.76 | 99.78 | 99.79 |
| 63 | 99.83 | 99.85 | 99.86 | 99.89 | 99.90 | 99.91 |
| 125 | 99.92 | 99.92 | 99.93 | 99.95 | 99.95 | 99.96 |
| 255 | 99.96 | 99.96 | 99.97 | 99.97 | 99.98 | 99.98 |

Table 3.10: Recommended designs for populations exhibiting linear trend.

| Case(s) | Condition | Preference | Recommended Design(s) |
|---|---|---|---|
| A | N/A | N/A | MBSS, MMSS or MBMSS |
| B | N/A | N/A | MBSS or MMSS |
| C to E | $m$, $n$ and $l$ are small | Minimum MSE | MBMSSEC |
| C to E | $m$, $n$ and $l$ are small | Simplicity | MYEC |
| C to E | $m$, $n$ or $l$ are not small | Minimum MSE | MBMSSEC or MYEC |
| C to E | $m$, $n$ or $l$ are not small | Simplicity | MYEC |

# Chapter 4

# Population Size is not a Multiple of the Sample Size

If the population size is not a multiple of the sample size, then this situation may be represented as $N = nk + r$, where $r \in \{1, ..., n-1\}$ and $k = \text{INT}(N/n)$, i.e. $\text{INT}(a)$ denotes the first integer before $a$. Now, if $r/k$ is an integer, then we obtain fixed sample sizes of $n + r/k$ when conducting LSS. The usual LSS design is now modified, such that we are selecting $n+r/k$ units with a sampling interval of $k$, as shown in the next example.

**Example 4.1:** If $N = 27$ and $n = 7$, then $k = 3$ and $r = 6$, which satisfies $N = nk + r$ and $r \in \{1, ..., n-1\}$. Each possible linear systematic sample is then given as:

(i) $S_1 = \{U_1, U_4, U_7, U_{10}, U_{13}, U_{16}, U_{19}, U_{22}, U_{25}\}$, for $i = 1$;

(ii) $S_2 = \{U_2, U_5, U_8, U_{11}, U_{14}, U_{17}, U_{20}, U_{23}, U_{26}\}$, for $i = 2$;

(iii) $S_3 = \{U_3, U_6, U_9, U_{12}, U_{15}, U_{18}, U_{21}, U_{24}, U_{27}\}$, for $i = 3$.

Clearly all possible samples are of fixed size $n + r/k = 7 + 6/3 = 9$, which is greater than the required size $n$. This situation is undesirable, as sample sizes are commonly fixed beforehand, owing to budget restrictions. Now, if $r/k$ is not an integer, then we obtain variable sample sizes of either $n$+INT$(r/k)$ or $n$+INT$(r/k)$+1, when conducting LSS, as shown in the following example.

**Example 4.2:** If $N = 19$ and $n = 5$, then $k = 3$ and $r = 4$, which satisfies $N = nk + r$

and $r \in \{1, ..., n-1\}$. Each possible linear systematic sample is then given as:

(i) $S_1 = \{U_1, U_4, U_7, U_{10}, U_{13}, U_{16}, U_{19}\}$, for $i = 1$;

(ii) $S_2 = \{U_2, U_5, U_8, U_{11}, U_{14}, U_{17}\}$, for $i = 2$;

(iii) $S_3 = \{U_3, U_6, U_9, U_{12}, U_{15}, U_{18}\}$, for $i = 3$.

Clearly we obtain samples of size $n + \text{INT}(r/k) = 6$ or $n + \text{INT}(r/k) + 1 = 7$. As a result, these samples of variable size may over-represent or under-represent the population, which in turn results in biased estimates of population parameters (Naidoo 2013). Consequently, many authors have proposed modified systematic sampling designs that generate fixed samples of size $n$, when the population size is not a multiple of the sample size. Reviews of each of these designs are given in the following section.

## 4.1 Modified systematic sampling designs

### 4.1.1 Circular systematic sampling (Lahiri 1951)

The ordering of population units associated with circular systematic sampling (CSS) is such that the units are arranged in a circular fashion, i.e. $U_{N+i} = U_i$. For this design, the sampling interval $k$ is now given as the closest integer to $N/n$, which ensures a more evenly spread sample over the population (Murthy 1967). Now, to select a sample of size $n$ from a population of size $N$ using CSS, we randomly select unit from the first $N$ population units and every subsequent $k$th unit, until the required sample size is achieved. The random start is given by $q$, where $q \in \{1, ...., N\}$.

If $N$ is a multiple of $k$, then sampling units will coincide when $N/n$ is rounded up (Sudhakar 1978). Consequently, Sudhakar (1978) proposed that $n$ distinct sampling units are obtainable if and only if the sampling interval $k$ is selected in advanced, were $N$ and $k$ are co-prime, i.e. $N \neq (n-1)k$. This approach, which assumes that $n$ is not fixed in advance, is undesirable as sample sizes are commonly predetermined and fixed, owing to budget restrictions. Bellhouse (1984) tackled this drawback by suggesting an alternative sampling interval, defined as

$$k^* = \begin{cases} \text{INT}(N/n), & \text{if } N = (n-1)k \\ \text{INT}(N/n + 1/2), & \text{if } N \neq (n-1)k. \end{cases} \tag{4.1}$$

Sengupta & Chattopadhyay (1987) claimed that sampling units may still coincide when $N/n$ is rounded up in (4.1). The authors then provided a theorem that states that $n$

distinct sampling units are obtainable, if and only if $\text{lcm}(N, k) \geq nk$ or, equivalently, if and only if $\text{gcd}(N, k) \leq N/n$, where $\text{lcm}(a, b)$ denotes the lowest common multiple and $\text{gcd}(a, b)$ denotes the greatest common divisor, for constants $a$ and $b$. This theorem does not contradict Sudhakar's (1978) results and may be applied as a supplement to Bellhouse's (1984) approach. It is common practice to apply a sampling interval of $k = \text{INT}(N/n)$, so as to ensure a sample of $n$ distinct sampling units. However, this sampling interval does not ensure an even spread of the sample over the population when $N/n$ is closer to integer $\text{INT}(N/n) + 1$. Subramani & Singh (2014) present an empirical study on the optimal choice of $k$, by assuming a perfect linear trend in the population and considering all the prime numbers from 7 to 37 as population sizes. Consequently, the authors suggested the following conjecture.

**Conjecture 1:** *The optimum choice for the sampling interval $k$ for selecting a circular systematic sample of size $n$ from the population of size $N$ is attained if and only if $kn \bmod N = \pm 1$, where $kn \bmod N = -1$ represents $kn \bmod N = N - 1$.*

*Proof*: Refer to Subramani et al. (2014).

Note that the proof of Conjecture 1 assumes that $N$ and $n$ are relatively prime numbers. Also, by choosing $k$ such that $kn \bmod N = \pm 1$, then one ensures that $N$ and $k$ are also relatively prime numbers. Conjecture 1 will always result in $n$ distinct sampling units.

**Theorem 13:** *Using Conjecture 1 under (2.1), the circular systematic sample mean $\overline{y}_{CSS}$ and the expected MSE of $\overline{y}_{CSS}$, when estimating the population mean, are respectively given as*

$$\overline{y}_{CSS} = a + \frac{b}{n}\left[q + \frac{(n-1)(N+1)}{2}\right], \qquad q \in \{1, ..., N\}$$

*and*

$$M_{CSS} = \sigma_e^2 + \frac{b^2(N^2-1)}{12n^2}. \tag{4.2}$$

*Proof*: Refer to Subramani et al. (2014).

By comparing (2.3) to (4.2), we see that $M_{CSS} < M_{SRS}$. A comparison between CSS and STR is unobtainable, as STR comparisons are only applicable when $N = nk$. If either $N$ and $n$ are not relatively prime numbers or $N$ and $k$ are not relatively prime numbers, then under model (2.1), an exact expression for the expected MSE of the sample mean when conducting CSS is difficult to obtain, owing to the circular nature of selecting the sample.

To remove the linear trend component in (4.2), Subramani et al. (2014) proposed an end corrections estimator, given by

$$\overline{y}^*_{CSS} = \overline{y}_{CSS} + \frac{(N+1-2i)}{2n(k \mp 1)}\left(y_q - y_{q+(n-1)k}\right), \qquad q \in \{1, ..., N\}.$$

### 4.1.2 Fractional interval method (Kish 1965, Murthy 1967)

The fractional interval method (FIM) is equivalent to the LSS design, where the sampling interval $k = N/n$ now takes on a fractional value. The random start is given as $i$, which is a randomly selected real number from a uniform distribution with interval $(0, k]$. The indices for the sampling units are given by $\alpha$, where

$$\alpha - 1 < i + (j-1)k \leq \alpha, \qquad j = 1, ..., n. \tag{4.3}$$

**Example 4.3:** Suppose that we want to select a sample of size 3 from a population of size 14, using the FIM. For this scenario, the sampling interval takes on the fractional value given by $k = 14/3$. Now, suppose that the random start is $i = 1/5$, which is a randomly selected real number from the uniform distribution with interval $(0, 14/3]$. Using (4.3), we obtain $\alpha = 1, 5, 10$. The sample is then given as $S_i = \{U_1, U_5, U_{10}\}$.

Naidoo (2013) shows that the FIM is equivalent to CSS, if and only if $2N/n$ is not an integer and $\text{lcm}(N, k) \geq nk$ (or $\gcd(N, k) \leq N/n$). Under these circumstances, the formulae obtained in Theorem 6 may then apply, which assumes that $N$ and $n$ are relatively prime numbers as well as $N$ and $k$ being relatively prime numbers. As a result, if we estimate the population mean under (2.1), then the sample mean and the expected MSE of the sample mean, associated with the FIM, are respectively given by $\overline{y}_{CSS}$ and $M_{CSS}$, if and only if $2N/n$ is not an integer, $\text{lcm}(N, k) \geq nk$ (or $\gcd(N, k) \leq N/n$), $N$ and $n$ are relatively prime numbers as well as $N$ and $k$ being relatively prime numbers.

### 4.1.3 New systematic sampling (Singh & Singh 1977)

Remembering that $U_{N+q} = U_q$, for $q \in \{1, ..., N\}$, new systematic sampling (NSS) is conducted as follows:

(i) Select a random integer $q$ on the interval $[1, N]$ and suppose that $l \leq n$ is an integer, such that $l$ consecutive units are selected starting with $U_q$, i.e. $\{U_q, ..., U_{q+l-1}\}$.

(ii) Suppose a sampling interval of $k'' = \text{INT}[(N-l)/(n-l)]$, such that the unit indices of the remaining $n-l$ sampling units are given by $q+l-1+jk''$, for $j = 1, ..., n-l$.

A sufficient and necessary requirement to obtain distinct sampling units is given by $(n-l)k'' \leq N-l$. Furthermore, if $k'' \leq l$ and $l+(n-l)k'' \geq N/2+1$, then the second-order inclusion probabilities for each possible pair of units will be non-zero, which will result in an unbiased estimate of the associated sampling variance. The proofs of these results are given by Naidoo (2013). Singh & Singh (1977) then showed that there is a restriction on the required sample size when non-zero second-order inclusion probabilities are obtained, i.e. $n \geq \sqrt{(2N+4)} - 1$. In the presence of linear trend, the variance of the sample mean, when conducting NSS, is complex. Singh & Singh (1977) thus provided an empirical study for populations exhibiting linear trend, which shows that NSS is less efficient than LSS and more efficient than SRS.

### 4.1.4 Balanced random sampling (Singh & Garg 1979)

In addition to the cases discussed for BRS in Section 3.1.2, we obtain a further case when population size is not a multiple of the sample size, i.e. Case (D) is when $N$ is odd and $n$ is even. The corresponding methodology is given as follows:

(i) Randomly select one unit from the population, before selecting a sample of size $(n-2)$ from the remaining $(N-1)$ population units, using the procedure in Case (A).

(ii) The balanced random sample is then given as the sampling units in (i), as well as a randomly selected unit from the remaining $(N-n+1)$ population units.

Note that some pairs of sampling units may not be optimally paired, as SRS is first applied before applying the usual pairing procedure, thus contributing to a linear trend component in the expected MSE of the associated sample mean.

### 4.1.5 New partially systematic sampling (Leu & Tsui 1996)

New partially systematic sampling (NPSS) is a modified NSS design, where the respective sampling procedure is given as follows:

(i) Select a random integer $q$ on the interval $[1, N]$.

(ii) Let $k$ be the nearest integer to $N/(n-1)$ and define $u$ as an integer, such that $2 \leq u \leq \text{INT}(n/2) + 1$. Also, let $s = N - (n-u)k$.

(iii) Select $u$ units from the sample space $S = \{U_q, U_{q+1}, ..., U_{q+s-1}\}$, using SRS.

(iv) The new partially systematic sample is then given as the sampling units obtained in (iii), as well as $n-u$ sampling units selected thereafter in a circular fashion, i.e. units $U_{q+s-1+ik}$, for $i = 1, ...., n-u$, where $U_{N+q} = U_q$.

By letting $s = N - (n-u)k$, Leu & Tsui (1996) ensures a sample of $n$ distinct sampling units, as there is only one circular transversal of unit indices. Moreover, all second-order inclusion probabilities are non-zero if (a) $u \geq 2$ and (b) $s \geq k$. Leu & Tsui (1996) then provided some recommendations on the choices of $u$ and $k$, so as to ensure an even spread of the sample over the population, while still ensuring a sample of $n$ distinct sampling units. These recommendations are based on the theory presented by Sudhakar (1978), Bellhouse (1984) and Sengupta & Chattopadhyay (1987), as reviewed in Section 4.1.1.

Note that NPSS is equivalent to NSS if $s = u$. Otherwise, NPSS may be regarded as the superior sampling design, as it depends on less stringent restrictions than that of NSS.

Since there are no optimal pairing of units for NPSS, we can expect the MSE of the associated sample mean to contain a linear trend component, owing to NPSS being a hybrid sampling design which combines SRS and CSS, where both SRS and CSS are not linear trend free sampling designs.

### 4.1.6 Modified circular systematic sampling designs (Uthayakumaran 1998, Leu & Kao 2006, Sampath & Varalakshmi 2009)

We now consider analogues of the linear trend free modified LSS designs in Sections 2.2.2, 2.2.3, 2.2.4 and 2.2.5, which adopt the CSS approach, as in Section 4.1.1.

#### Balanced and centered circular systematic sampling (Uthayakumaran 1998)

Let the sampling interval be defined as $k = \text{INT}(N/n)$. Now, if we assume that $N$ and $n$ are even, then the $q$th (for $q \in \{1, ..., N\}$) sample for balanced circular systematic sampling (BCSS) is given as

$$S_{BCSS} = \begin{cases} U_{q+2(j-1)k} | j = 1, ..., n/2, & \text{if } 1 \leq q + 2(j-1)k \leq N \\ U_{q+2(j-1)k-N} | j = 1, ..., n/2, & \text{if } q + 2(j-1)k > N \\ U_{2jk-q+1+N} | j = 1, ..., n/2, & \text{if } 2jk - q + 1 < 1 \\ U_{2jk-q+1} | j = 1, ..., n/2, & \text{if } 1 \leq 2jk - q + 1 \leq N. \end{cases}$$

Centered circular systematic sampling (CCSS) adopts the usual CSS design; however, the centrally located circular systematic sample is selected and thus no randomization is required. The sample for CCSS is thus given as

$$S_{CCSS} = \begin{cases} U_{(N+1)/2+(j-1)k} | j = 1, ..., n, & \text{if } 1 \leq (j-1)k \leq (N-1)/2 \text{ and } N \text{ is odd} \\ U_{(1-N)/2+(j-1)k} | j = 1, ..., n, & \text{if } (j-1)k > (N-1)/2 \text{ and } N \text{ is odd} \\ U_{N/2+(j-1)k} | j = 1, ..., n, & \text{if } 1 \leq (j-1)k \leq N/2 \text{ and } N \text{ is even} \\ U_{(j-1)k-N/2} | j = 1, ..., n, & \text{if } (j-1)k > N/2 \text{ and } N \text{ is even.} \end{cases}$$

Let $x_1, ..., x_n$ denote the indices of the sampling units, which are arranged in ascending order. Also, denote $\overline{y}_{BCSS}$ and $\overline{y}_{CCSS}$ as the sample means, which are unbiased estimates of the population mean, corresponding to BCSS and CCSS, respectively. Using this notation, the end corrections estimators associated with BCSS and CCSS, are respectively given as

$$\overline{y}_{BCSSEC} = \overline{y}_{BCSS} + \frac{[(x_n + x_1) - K]}{n(x_n - x_1)} (y_{x_1} - y_{x_n})$$

and

$$\overline{y}_{CCSSEC} = \overline{y}_{CCSS} + \frac{[(x_n + x_1) - K]}{n(x_n - x_1)} (y_{x_1} - y_{x_n}),$$

where $K = n(N+1)/2 - \sum_{j=2}^{n-1} x_j$. Under model (2.1), the expected MSEs of estimators $\overline{y}_{BCSSEC}$ and $\overline{y}_{CCSSEC}$, are given as

$$M_* = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{\sigma^2}{n^2} \left[ W_1^2 x_1 + \sum_{j=2}^{n-1} x_j + W_2^2 x_n \right] + \frac{\sigma^2}{N} - \frac{2\sigma^2}{Nn} \left[ W_1 x_1 + \sum_{j=2}^{n-1} x_j + W_2 x_n \right] \right\},$$

where $W_1 = (K - 2x_n)/(x_1 - x_n)$ and $W_2 = (2x_1 - K)/(x_1 - x_n)$. Clearly we do not obtain optimality under any circumstance. Thus, Leu & Kao (2006) introduced modifications to the above designs to tackle this problem, as shown in the next section.

**Modified balanced circular systematic sampling and modified centered circular systematic sampling (Leu & Kao 2006)**

Let the sampling interval be defined as $k = \text{INT}(N/n)$. Now, if we assume that $n$ is even, then the $q$th (for $q \in \{1, ..., N\}$) sample for modified BCSS (MBCSS) is given as

$$S_{MBCSS} = \begin{cases} U_{q+2(j-1)k} | j = 1, ..., n/2, & \text{if } 1 \leq q + 2(j-1)k \leq N \\ U_{q+2(j-1)k-N} | j = 1, ..., n/2, & \text{if } q + 2(j-1)k > N \\ U_{N-q-2(j-1)k+1} | j = 1, ..., n/2, & \text{if } 1 \leq q + 2(j-1)k - 1 < N \\ U_{2N-q-2(j-1)k+1} | j = 1, ..., n/2, & \text{if } q + 2(j-1)k - 1 \geq N. \end{cases}$$

Also, if we assume that $n$ is odd, then the $q$th sample for MBCSS consists of $(n-1)$ units given by

$$S_{MBCSS} = \begin{cases} U_{q+2(j-1)k} | j = 1, ..., (n-1)/2, & \text{if } 1 \leq q + 2(j-1)k \leq N \\ U_{q+2(j-1)k-N} | j = 1, ..., (n-1)/2, & \text{if } q + 2(j-1)k > N \\ U_{N-q-2(j-1)k+1} | j = 1, ..., n/2, & \text{if } 1 \leq q + 2(j-1)k - 1 < N \\ U_{2N-q-2(j-1)k+1} | j = 1, ..., n/2, & \text{if } q + 2(j-1)k - 1 \geq N, \end{cases}$$

as well as an $n$th sampling unit given as $U_{(N+1)/2}$ if $N$ is odd, or $U_{N/2}$ if $N$ is even. The sample for modified CCSS (MCCSS) is obtained by selecting the centrally located sample for MBCSS. If we assume that $n$ is even, then the sample for MCCSS contains $(n-1)$ units given by

$$S_{MCCSS} = \begin{cases} U_{(N+1)/2+(j-1)k}, U_{(N+1)/2-(j-1)k} | j = 1, ..., n/2, & \text{if } N \text{ is odd} \\ U_{N/2+(j-1)k}, U_{N/2-(j-1)k} | j = 1, ..., n/2, & \text{if } N \text{ is even,} \end{cases}$$

as well as an $n$th sampling unit given as $U_{(N+1+nk)/2}$ if $N$ is odd, or $U_{(N+nk)/2}$ if $N$ is even. Also, if we assume that $n$ is odd, then the sample for MCCSS is given as

$$S_{MCCSS} = \begin{cases} U_{(N+1)/2+(j-1)k}, U_{(N+1)/2-(j-1)k} | j = 1, ..., (n+1)/2, & \text{if } N \text{ is odd} \\ U_{N/2+(j-1)k}, U_{N/2-(j-1)k} | j = 1, ..., (n+1)/2, & \text{if } N \text{ is even.} \end{cases}$$

Let $\overline{y}_{MBCSS}$ and $\overline{y}_{MCCSS}$ denote the sample means, which are unbiased estimates of the population mean, corresponding to MBCSS and MCCSS, respectively. Using this notation, the end corrections estimators corresponding to MBCSS and MCCSS, are respectively given as

$$\overline{y}_{MBCSSEC} = \overline{y}_{MBCSS} + \frac{[(x_n + x_1) - K]}{n(x_n - x_1)} (y_{x_1} - y_{x_n})$$

and

$$\overline{y}_{MCCSSEC} = \overline{y}_{MCCSS} + \frac{[(x_n + x_1) - K]}{n(x_n - x_1)} (y_{x_1} - y_{x_n}) .$$

Under model (2.1), the expected MSEs of estimators $\overline{y}_{MBCSS}$ and $\overline{y}_{MCCSS}$ are given as $M_{MBCSS} = \sigma_e^2$ (if $N$ is odd or $n$ is even) and $M_{MCCSS} = \sigma_e^2$ (if both $N$ and $n$ are odd), respectively. For the other possible cases, the end corrections estimators are applied, i.e. the expected MSEs of estimators $\overline{y}_{MBCSSEC}$ and $\overline{y}_{MCCSSEC}$ are given as $M_{MBCSSEC} = M_*$ (if $N$ is even and $n$ is odd) and $M_{MCCSSEC} = M_*$ (if $N$ is even or $n$ is even), respectively. Thus, we only obtain optimality for MBCSS when $N$ is odd or $n$ is even, as well as for MCCSS when both $N$ and $n$ are odd.

**Diagonal circular systematic sampling (Sampath & Varalakshmi 2009)**

Define $k$ as the nearest integer to $N/n$. If $q + (j-1)(k+1) \leq N$, for all $j = 1, ..., n$ and $q \in \{1, ..., N\}$, then the $q$th sample for diagonal circular systematic sampling (DCSS) is given as $S_{DCSS} = \{U_{q+(j-1)(k+1)} | j = 1, ..., n\}$. Now, if $q + (j-1)(k+1) > N$ for some $j = 1, ..., n$, then the $q$th sample for DCSS is given as

$$S_{DCSS} = \begin{cases} U_{q+(j-1)(k+1)} | j = 1, 2, ..., n(q) \\ U_{q+(j-1)(k+1)-N} | j = n(q)+1, ..., n, \end{cases}$$

where $n(q)$ is the number of items in the state space $\{q + (j-1)(k+1) \leq N | j = 1, ..., n\}$ for a given $q$.

The corresponding sample mean, denoted as $\overline{y}_{DCSS}$, is an unbiased estimate of the population mean. Note that DCSS is not a linear trend free sampling design. If we obtain $q+(j-1)(k+1) \leq N$, for all $j = 1, ..., n$, then the DCSS with end corrections (DCSSEC) estimator is given as

$$\overline{y}_{DCSSEC} = \frac{N}{n}\left[\psi_1 y_q + \sum_{j=2}^{n-1} y_{q+(j-1)(k+1)} + \psi_2 y_{q+(n-1)(k+1)}\right],$$

where

$$\psi_1 = \frac{2nq + (k+1)(n-1)(n+2) - n(N+1)}{2(n-1)(k+1)}$$

and

$$\psi_2 = \frac{n(N+1) - 2nq - (k+1)(n-1)(n-2)}{2(n-1)(k+1)}.$$

Now, if $q+(j-1)(k+1) > N$ for some $j \in \{1, ..., n\}$, then the DCSS with end corrections (DCSSEC) estimator is given as

$$\overline{y}_{DCSSEC} = \frac{N}{n}\left[\psi_1^{'} y_q + \sum_{j=2}^{n(q)} y_{q+(j-1)(k+1)} + \sum_{j=n(q)+1}^{n-1} y_{q+(j-1)(k+1)-N} + \psi_2^{'} y_{q+(n-1)(k+1)-N}\right],$$

where

$$\psi_1^{'} = \frac{2\left[q + (n-1)(k+1) - N\right] - \eta}{(n-1)(k+1) - N},$$

$$\psi_2^{'} = \frac{\eta - 2q}{(n-1)(k+1) - N}$$

and

$$\eta = \frac{n(N+1)}{2} - (n-2)q - \frac{(k+1)(n-2)(n-1)}{2} + N\left[(n-1) - n(q)\right].$$

If we suppose that $J_1$ and $J_2$ are two subsets of $S = \{1, ..., N\}$, such that $J_2 = S - J_1$ and $J_1 = \{q | q + (j-1)(k+1) \leq N, \text{ for all } j = 1, ..., n\}$, then under model (2.1), the expected MSE of estimator $\overline{y}_{DCSSEC}$ is given as

$$M_{DCSSEC} = \sigma^2 \left[ \frac{N}{n^2} \sum_{q \in J_1} \{\psi_1^2 + (n-2) + \psi_2^2\} + \frac{N}{n^2} \sum_{q \in J_2} \left\{\psi_1'^2 + (n-2) + \psi_2'^2\right\} \right.$$
$$\left. + \frac{2}{n} \left[ \sum_{q \in J_1} \{\psi_1 + (n-2) + \psi_2\} + \sum_{q \in J_2} \left\{\psi_1' + (n-2) + \psi_2'\right\} \right] + N \right].$$

Sampath & Varalakshmi (2009) then provides some numerical comparisons which indicate that an end corrections estimator associated with CSS, i.e. $\overline{y}^*_{CSS}$, is preferred over estimator $\overline{y}_{DCSSEC}$.

Khan et al. (2014) noted that coincidence of sampling units are possible for certain cases of DCSS when $N/n$ is rounded up, just as in the case for CSS. The authors then solved this problem by adopting an approach which is similar to that of Sengupta & Chattopadhyay (1987), i.e. $n$ distinct sampling units are always obtainable under DCSS, if and only if $\text{lcm}(N, (k+1)) \geq n(k+1)$ or, equivalently, if and only if $\gcd(N, (k+1)) \leq N/n$.

### 4.1.7 Remainder linear systematic sampling (Chang & Huang 2000)

Define the sampling interval as $k = \text{INT}(N/n)$, such that the population size is represented as $N = nk + r = (n-r)k + r(k+1)$, where $r \in \{1, ..., n-1\}$. The methodology of remainder linear systematic sampling (RLSS) is then given as follows:

(i) Divide the population into two strata, where the first stratum, $ST_1$, contains the first $(n-r)k$ population units and the second stratum, $ST_2$, contains the remaining $r(k+1)$ units.

(ii) Select two random starts $k1$ and $k2$, where $k1 \in \{1, ..., k\}$ and $k2 \in \{1, ..., k+1\}$.

(iii) The samples selected from $ST_1$ and $ST_2$ are respectively given as

$$S_{k1} = \{U_{k1+(j-1)k} | j = 1, ..., (n-r)\}$$

and

$$S_{k2} = \{U_{(n-r)k+k2+(j-1)(k+1)} | j = 1, ..., r\}.$$

(iv) The final sample of size $n$ is then given as $S = S_{k1} \cup S_{k2}$.

If we denote the sample means from $ST_1$ and $ST_2$ as $\overline{y}_{k1}$ and $\overline{y}_{k2}$, respectively, then the RLSS sample mean, which is an unbiased estimate of the population mean, is given as $\overline{y}_{RLSS} = \left[(n-r)k\overline{y}_{k1} + r(k+1)\overline{y}_{k2}\right]/N$. Now, under model (2.1), the expected MSE of $\overline{y}_{RLSS}$ is given by

$$M_{RLSS} = \sigma_r^2 + \frac{b^2 k}{12N^2}\left[(n-r)^2 k(k^2-1) + r^2(k+1)^2(k+2)\right], \tag{4.4}$$

where $\sigma_r^2 = k\sigma^2\left(N-n+r\right)/N^2$ represents the minimum expected error variance when independently sampling from $ST_1$ and $ST_2$. Clearly there is a linear trend component in $M_{RLSS}$. To remove this component and improve estimation results, Chang & Huang (2000) suggested an end corrections estimator, i.e. RLSS with end corrections (RLSSEC) estimator, given by

$$\overline{y}_{RLSSEC} = \overline{y}_{RLSS} + Z(y_{k1} - y_{k1+(n-r-1)k}), \tag{4.5}$$

where

$$Z = \frac{2\left[(k1)(n-r)k + (k2)r(k+1)\right] - (N-r)(k+1)}{2(n-r-1)Nk}.$$

### 4.1.8 Mixed random systematic sampling (Huang 2004)

Represent the population size as $N = nk + r = (n-r)k + r(k+1)$, where $k = \text{INT}(N/n)$ and $r \in \{1, ..., n-1\}$. The methodology of mixed random systematic sampling (MRSS) is then given as follows:

(i) Select a random integer $q$ on the interval $[1, N]$.

(ii) Remembering that $U_{N+q} = U_q$, divide the population into two strata, where the first stratum, $ST_1'$, contains $(n-r)k$ units given by $\{U_q, U_{q+1}, U_{q+(n-r)k-1}\}$ and the second stratum, $ST_2'$, contains the remaining $r(k+1)$ units.

(iii) Randomly select $(n-r)$ sampling units from $ST_1'$ using SRS and represent this sample as $S_{q1}$.

(iv) The sample selected from $ST_2'$ is given as $S_{q2} = \{U_{q+(n-r)k+j(k+1)-1} | j = 1, ..., r\}$.

(v) The final sample of size $n$ is then given as $S_q = S_{q1} \cup S_{q2}$.

The associated sample mean is given by the Horvitz & Thompson (1952) estimator, i.e. $\hat{\overline{Y}}_{HT} = (1/N)\sum_{U_q \in S_q}(y_q/\pi_q)$, where $\pi_q = n/N$, for all $q = 1, ..., N$. Note that estimator $\hat{\overline{Y}}_{HT}$ is an unbiased estimate of the population mean. Under model (2.1), an exact expression for the expected MSE of estimator $\hat{\bar{Y}}_{HT}$ is difficult to obtain, owing to the circular nature of selecting the sample. However, Huang (2004) provides some numerical comparisons under the assumption of model $y_q = a + bq$, for $q = 1, ..., N$. The results indicate that the sample mean corresponding to MRSS is subject to more error than that which is associated with CSS.

### 4.1.9 Remainder Markov systematic sampling (Kao et al. 2011*a*)

Represent the population size as $N = nk + r = (n-r)k + r(k+1)$, where $k = \text{INT}(N/n)$ and $r \in \{1, ..., n-1\}$. The methodology of remainder Markov systematic sampling is then given as follows:

(i) Divide the population into two strata, as in step (i) of RLSS.

(ii) Apply Markov systematic sampling, as in Section 3.1.4, within each of these strata. If the units sampled are even then the stratification related to Markov systematic sampling within these strata are straightforward. On the other hand, if the units sampled are odd, then SRS is applied to select a unit from the remaining units after stratification. The first stratum corresponds to stochastic matrix $\mathbf{A}$ which is a $2k \times 2k$ matrix, while the second stratum corresponds to stochastic matrix $\mathbf{B}$ which is a $2(k+1) \times 2(k+1)$ matrix. Note that both matrices $\mathbf{A}$ and $\mathbf{B}$ are doubly stochastic matrices, with zero diagonal elements, so as to ensure distinct sampling units.

(iii) Remainder Markov systematic sampling is now classified into four cases, which are given as follows:

Case (A): If $n$ and $r$ are both even, then:

1. divide the units in the first stratum into $(n-r)/2$ groups, each containing $2k$ units, before applying Markov systematic sampling to select $(n-r)$ units using matrix $\mathbf{A}$;
2. divide the units in the second stratum into $r/2$ groups, of $2(k+1)$ units each, before applying Markov systematic sampling to select $r$ units using matrix $\mathbf{B}$.

Case (B): If $n$ is even and $r$ is odd, then:

1. divide the first $(n-r-1)k$ units in the first stratum into $(n-r-1)/2$ groups, each containing $2k$ units, before applying Markov systematic sampling to

select $(n-r)$ units using matrix $\mathbf{A}$;

2. divide the last $(r-1)(k+1)$ units in the second stratum into $(r-1)/2$ groups, each containing $2(k+1)$ units, before applying Markov systematic sampling to select $(r-1)$ units using matrix $\mathbf{B}$;
3. select two units from $ST_3 = \{U_{(n-r-1)k+1}, ..., U_{(n-r)k+(k+1)}\}$ using SRS.

Case (C): If $n$ is odd and $r$ is even, then:

1. divide the first $(n-r-1)k$ units in the first stratum into $(n-r-1)/2$ groups, each containing $2k$ units, before applying Markov systematic sampling to select $(n-r)$ units using matrix $\mathbf{A}$;
2. randomly select a unit from $ST_3 = \{U_{(n-r-1)k+1}, ..., U_{(n-r)k}\}$;
3. divide the units in the second stratum into $r/2$ groups, of $2(k+1)$ units each, before applying Markov systematic sampling to select $r$ units using matrix $\mathbf{B}$.

Case (D): If $n$ and $r$ are both odd:

1. divide the units in the first stratum into $(n-r)/2$ groups, each containing $2k$ units, before applying Markov systematic sampling to select $(n-r)$ units using matrix $\mathbf{A}$;
2. randomly select a unit from $ST_3 = \{U_{(n-r)k+1}, ..., U_{(n-r)k+(k+1)}\}$;
3. divide the last $(r-1)(k+1)$ units in the second stratum into $(r-1)/2$ groups, each containing $2(k+1)$ units, before applying Markov systematic sampling to select $(r-1)$ units using matrix $\mathbf{B}$.

Next, three types of stochastic matrices were considered and given as follows:

(i) To conduct remainder stratified systematic sampling (RSSS), the stochastic matrices corresponding to the first and second strata are respectively given as

$$\mathbf{H_1} = \begin{bmatrix} 0 & \frac{1}{2k-1} & 0 & \frac{1}{2k-1} & \frac{1}{2k-1} \\ \frac{1}{2k-1} & 0 & \cdots & \frac{1}{2k-1} & \frac{1}{2k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{2k-1} & \frac{1}{2k-1} & \cdots & 0 & \frac{1}{2k-1} \\ \frac{1}{2k-1} & \frac{1}{2k-1} & \cdots & \frac{1}{2k-1} & 0 \end{bmatrix}_{2k\times 2k}$$

and

$$\mathbf{H_2} = \begin{bmatrix} 0 & \frac{1}{2k+1} & 0 & \frac{1}{2k+1} & \frac{1}{2k+1} \\ \frac{1}{2k+1} & 0 & \dots & \frac{1}{2k+1} & \frac{1}{2k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{2k+1} & \frac{1}{2k+1} & \dots & 0 & \frac{1}{2k+1} \\ \frac{1}{2k+1} & \frac{1}{2k+1} & \dots & \frac{1}{2k+1} & 0 \end{bmatrix}_{2(k+1)\times 2(k+1)}.$$

(ii) To conduct remainder balanced systematic sampling (RBSS), the stochastic matrices corresponding to the first and second strata are respectively given as

$$\mathbf{J_1} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}_{2k\times 2k}$$

and

$$\mathbf{J_2} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}_{2(k+1)\times 2(k+1)}.$$

(ii) To conduct remainder balanced systematic-like sampling (RBSLS), the stochastic matrices corresponding to the first and second strata are respectively given as

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}_{2k\times 2k}$$

and

$$\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}_{2(k+1)\times 2(k+1)},$$

where

$$A_{11} = A_{22} = 0_{k\times k}, \qquad B_{11} = B_{22} = 0_{(k+1)\times(k+1)},$$

$$A_{12} = \begin{bmatrix} 0 & 0 & \dots & 0 & p_1 & 1-p_1 \\ 0 & 0 & \dots & p_1 & 1-p_1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_1 & 1-p_1 & \dots & 0 & 0 & 0 \\ 1-p_1 & 0 & \dots & 0 & 0 & p_1 \end{bmatrix}_{k\times k},$$

$$A_{21} = \begin{bmatrix} p_1 & 0 & 0 & \dots & 0 & 1-p_1 \\ 0 & 0 & 0 & \dots & 1-p_1 & p_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1-p_1 & p_1 & \dots & 0 & 0 \\ 1-p_1 & p_1 & 0 & \dots & 0 & 0 \end{bmatrix}_{k\times k},$$

$$B_{12} = \begin{bmatrix} 0 & 0 & \dots & 0 & p_2 & 1-p_2 \\ 0 & 0 & \dots & p_2 & 1-p_2 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_2 & 1-p_2 & \dots & 0 & 0 & 0 \\ 1-p_2 & 0 & \dots & 0 & 0 & p_2 \end{bmatrix}_{(k+1)\times(k+1)},$$

$$B_{21} = \begin{bmatrix} p_2 & 0 & 0 & \dots & 0 & 1-p_2 \\ 0 & 0 & 0 & \dots & 1-p_2 & p_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1-p_2 & p_2 & \dots & 0 & 0 \\ 1-p_2 & p_2 & 0 & \dots & 0 & 0 \end{bmatrix}_{(k+1)\times(k+1)},$$

$0 \le p_1 \le 1$ and $0 \le p_2 \le 1$.

The associated sample mean is given by the Horvitz & Thompson (1952) estimator,

$$\hat{\overline{Y}}_{HT} = \begin{cases} \left[(n-r)k\overline{y}_1 + r(k+1)\overline{y}_2\right]/N, & \text{for Case (A)} \\ \left[(n-r-1)k\overline{y}_1 + (r-1)(k+1)\overline{y}_2 + (2k+1)\overline{y}_3\right]/N, & \text{for Case (B)} \\ \left[(n-r-1)k\overline{y}_1 + r(k+1)\overline{y}_2 + ky_3\right]/N, & \text{for Case (C)} \\ \left[(n-r)k\overline{y}_1 + (r-1)(k+1)\overline{y}_2 + (k+1)y_3\right]/N, & \text{for Case (D),} \end{cases}$$

where $\overline{y}_1$, $\overline{y}_2$ and $\overline{y}_3$, are the sample means from $ST_1$, $ST_2$ and $ST_3$, respectively, and $y_3$ is the observed value from $ST_3$. Note that estimator $\hat{\overline{Y}}_{HT}$ is an unbiased estimate of the population mean. Kao et al. (2011*a*) then obtains values for the second-order inclusion probabilities, before claiming that is it possible to obtain an unbiased estimate of the variance of $\hat{\overline{Y}}_{HT}$, when adopting their design. However, by further inspection, we see that this claim is only correct for the stochastic matrices associated with the RSSS design.

Under model (2.1), if we apply remainder Markov systematic sampling for Cases (A), (C) and (D), then the expected MSE of $\hat{\overline{Y}}_{HT}$ is given by

$$\begin{aligned} M_{RM} = \sigma_r^2 + \frac{b^2}{N^2}\Bigg\{ & \sum_{i=1}^{(n-r)k} \sum_{j>i}^{(n-r)k} (i-j)^2(1-k^2\pi_{ij}) \\ & + \sum_{i=(n-r)k+1}^{N} \sum_{j>i}^{N} (i-j)^2\left[1-(k+1)^2\pi_{ij}\right] \Bigg\}. \end{aligned}$$

Similarly, if we consider Case (B), then the expected MSE of $\hat{\overline{Y}}_{HT}$ is given by

$$\begin{aligned} M_{RM} = & \sigma_r^2 - \frac{(2k+1)\sigma^2}{2N^2} + \frac{b^2}{N^2}\Bigg\{ \sum_{i=1}^{(n-r-1)k} \sum_{j>i}^{(n-r-1)k} (i-j)^2(1-k^2\pi_{ij}) \\ & + \sum_{i=(n-r-1)k+1}^{(n-r+1)k+1} \sum_{j>i}^{(n-r+1)k+1} (i-j)^2\left[1-\left(\frac{2k+1}{2}\right)^2\pi_{ij}\right] \\ & + \sum_{i=(n-r+1)k+2}^{N} \sum_{j>i}^{N} (i-j)^2\left[1-(k+1)^2\pi_{ij}\right] \Bigg\}. \end{aligned}$$

Finally, we note that $M_{RM}$ is only minimized when applying the stochastic matrices associated with RBSS for Case (A), i.e. all other scenarios result in a linear trend component in $M_{RM}$.

### 4.1.10 Remainder systematic Markov chain design (Kao et al. 2011*b*)

This approach is similar to the previous design, where the methodology is given as follows:

(i) Apply step (i) of the methodology of remainder Markov systematic sampling.

(ii) Apply Markov systematic sampling, as in Section 3.1.4, within each stratum. The first stratum corresponds to stochastic matrix $\mathbf{A}$ which is a $k \times k$ matrix, while the second stratum corresponds to stochastic matrix $\mathbf{B}$ which is a $(k+1) \times (k+1)$ matrix. Note that both matrices $\mathbf{A}$ and $\mathbf{B}$ are doubly stochastic matrices, with zero diagonal elements so as to ensure distinct sampling units.

(iii) In the first stratum, the two cases for selecting the $(n-r)$ sampling units are given as follows:

1. If $(n-r)$ is even, then divide the units in the first stratum into $(n-r)/2$ groups, of $2k$ units each, according to their unit indices. Randomly select a unit from the first $k$ units in the first group and every $2k$th units thereafter, until $(n-r)/2$ units are obtained, i.e. the unit selected from each group is located in the same position. Now, select units from the $(k+1)$th to the $2k$th unit of each group using the Markov chain design, such that the remaining $(n-r)/2$ units are obtained.

2. If $(n-r)$ is odd, then divide the units in the first stratum into $(n-r-1)/2$ groups, of $2k$ units each, and one group of $k$ units according to their unit indices. Randomly select a unit from the first $k$ units in the first group and every $2k$th units thereafter, until $(n-r+1)/2$ units are obtained. Now, select units from the $(k+1)$th to the $2k$th unit of each group (i.e. excluding the group containing $k$ units) using the Markov chain design, such that the remaining $(n-r-1)/2$ units are obtained.

(iv) In the second stratum, the two cases for selecting the $r$ sampling units are given as follows:

1. If $r$ is even, then divide the units in the second stratum into $r/2$ groups, of $2(k+1)$ units each, according to their unit indices. Randomly select a unit from the first $(k+1)$ units in the first group and every $2(k+1)$th units thereafter, until $r/2$ units are obtained, i.e. the unit selected from each group is located in the same position. Now, select units from the $(k+2)$th to the $2(k+1)$th unit of each group using the Markov chain design, such that the remaining $r/2$ units are obtained.

2. If $r$ is odd, then divide the units in the first stratum into $(r-1)/2$ groups, of $2(k+1)$ units each, and one group of $(k+1)$ units according to their unit indices.

Randomly select a unit from the first $(k+1)$ units in the first group and every $2(k+1)$th units thereafter, until $(r+1)/2$ units are obtained. Now, select units from the $(k+2)$th to the $2(k+1)$th unit of each group (i.e. excluding the group containing $(k+1)$ units) using the Markov chain design, such that the remaining $(r-1)/2$ units are obtained.

Next, four types of stochastic matrices, the first three of which were considered by Breidt (1995), are given as follows:

(i) To conduct remainder stratified systematic sampling (RSSS), the stochastic matrices corresponding to the first and second strata are respectively given as

$$\mathbf{H_1} = \left[\frac{1}{k}\right]_{k\times k}$$

and

$$\mathbf{H_2} = \left[\frac{1}{k+1}\right]_{(k+1)\times(k+1)}.$$

(ii) To conduct RLSS, the stochastic matrices corresponding to the first and second strata are given by identity matrices with dimensions $k \times k$ and $(k+1)\times(k+1)$, respectively.

(iii) To conduct RBSS, the stochastic matrices corresponding to the first and second strata are respectively given as

$$\mathbf{J_1} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}_{k\times k}$$

and

$$\mathbf{J_2} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}_{(k+1)\times(k+1)}.$$

(iv) To conduct RBSLS, the stochastic matrices corresponding to the first and second strata are respectively given as

$$\mathbf{P_1} = \begin{bmatrix} p_1 & 1-p_1 & 0 & \dots & 0 & 0 \\ 0 & p_1 & 1-p_1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & p_1 & 1-p_1 \\ 1-p_1 & 0 & 0 & \dots & 0 & p_1 \end{bmatrix}_{k\times k}$$

and

$$\mathbf{P_2} = \begin{bmatrix} p_2 & 1-p_2 & 0 & \dots & 0 & 0 \\ 0 & p_2 & 1-p_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & p_2 & 1-p_2 \\ 1-p_2 & 0 & 0 & \dots & 0 & p_2 \end{bmatrix}_{(k+1)\times(k+1)},$$

where $0 \leq p_1 \leq 1$ and $0 \leq p_2 \leq 1$.

Now, the corresponding sample mean, which is an unbiased estimate of the population mean, is given by the Horvitz & Thompson (1952) estimator, i.e. estimator $\hat{\overline{Y}}_{HT} = [(n-r)k\overline{y}_1 + r(k+1)\overline{y}_2]/N$, where $\overline{y}_1$ and $\overline{y}_2$ are the sample means from the first and second stratum, respectively. Kao et al. (2011*b*) then provides values for the second-order inclusion probabilities, which indicate that it is impossible to obtain an unbiased estimate of the variance of $\hat{\overline{Y}}_{HT}$ when adopting their design.

If we apply the remainder systematic Markov chain design under model (2.1), then the expected MSE of $\hat{\overline{Y}}_{HT}$ is given by

$$M_{RSMCD} = \sigma_r^2 + \frac{b^2}{N^2}\Bigg\{\sum_{i=1}^{(n-r)k}\sum_{j>i}^{(n-r)k}(i-j)^2(1-k^2\pi_{ij}) + \sum_{i=(n-r)k+1}^{N}\sum_{j>i}^{N}(i-j)^2\left[1-(k+1)^2\pi_{ij}\right]\Bigg\}.$$

By substituting the relevant values of $\pi_{ij}$, which are obtained by applying the corresponding stochastic matrices, we note that $M_{RSMCD}$ is only minimized for RBSS when $(n-r)$ and $r$ are both even, i.e. all other scenarios result in a linear trend component in $M_{RSMCD}$.

### 4.1.11 Modified systematic sampling (Khan et al. 2013)

Modified systematic sampling (MSYS) is a variation of circular systematic sampling and is given by the following methodology:

(i) Let $L = \text{lcm}(N, n)$, such that $v = L/N$, $k^* = L/n$, $w = N/k^*$ and $k = k^*/v$, or $k$ is taken as the closest integer to $(N/n)$. Thus, the required sample size is $n = vw$, which results in the selection of $v$ sets of size $w$.

(ii) Assume the population is arranged in a circular fashion.

(iii) Randomly select a unit from the first $k^*$ units in the population, say $U_i$, where $i \in \{1, ..., k^*\}$.

(iv) The first unit in each of the $v$ sets is given by unit $U_{i+(j-1)k}$, for $j = 1, ..., v$. The remaining $(w - 1)$ units for each set, are obtained by selecting every $k^*$th unit thereafter in a circular fashion. Thus, the sampling units obtained for the $j$th set is given by $S_{ij} = \{U_{i+jk}, U_{i+jk+k^*}, ..., U_{i+jk+(w-1)k^*}\}$.

Note that under MSYS, a necessary and sufficient condition for obtaining a sample of distinct sampling units is that $\text{lcm}(k^*, k)/k \geq v$ or, equivalently, if $(j - 1)k \neq k^*$, where $j \leq v - 1$. Thus, the design is susceptible to coincidence of sampling units. However, it is rarer for sampling units to coincide when conducting MSYS, as opposed to CSS.

The corresponding sample mean is an unbiased estimate of the population mean, while under (2.1), an exact expression for the expected MSE of this sample mean is difficult to obtain, owing to the circular nature of selection. Nevertheless, we can expect this expression to contain a linear trend component, as MSYS is merely a modification of CSS, with no optimum pairing of sampling units.

### 4.1.12 Generalized modified linear systematic sampling (Subramani & Gupta 2014)

Represent the population size as $N = n_1k_1 + n_2k_2$, where $n = n_1 + n_2$ and $\hat{k} = k_1 + k_2$, such that $n_1$, $n_2$, $k_1$ and $k_2$, are all positive integers and $n_1 > n_2$. The procedure of generalized modified linear systematic sampling (GMLSS) is then given as follows:

(i) Divide and arrange the $N$ population unit indices according to two matrices $\mathbf{K_1}$ and

$\mathbf{K_2}$, where

$$\mathbf{K_1} = \begin{bmatrix} 1 & \dots & i & \dots & k_1 \\ \hat{k}+1 & \dots & \hat{k}+i & \dots & \hat{k}+k_1 \\ 2\hat{k}+1 & \dots & 2\hat{k}+i & \dots & 2\hat{k}+k_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n_2\hat{k}+1 & \dots & n_2\hat{k}+i & \dots & n_2\hat{k}+k_1 \\ n_2\hat{k}+k_1+1 & \dots & n_2\hat{k}+k_1+i & \dots & n_2\hat{k}+2k_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n_2k_2+(n_1-1)k_1+1 & \dots & n_2k_2+(n_1-1)k_1+i & \dots & n_2k_2+n_1k_1 \end{bmatrix}_{n_1\times k_1}$$

and

$$\mathbf{K_2} = \begin{bmatrix} k_1+1 & \dots & k_1+j & \dots & \hat{k} \\ \hat{k}+k_1+1 & \dots & \hat{k}+k_1+j & \dots & 2\hat{k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (n_2-1)\hat{k}+k_1+1 & \dots & (n_2-1)\hat{k}+k_1+j & \dots & n_2\hat{k} \end{bmatrix}_{n_2\times k_2}.$$

(ii) The indices corresponding to the sampling units are then given by a randomly selected column in matrix $\mathbf{K_1}$, say the $i$th column, as well as a randomly selected column in matrix $\mathbf{K_2}$, say the $j$th column.

The first-order inclusion probabilities for the unit $U_q$ is given as

$$\pi_q = \begin{cases} 1/k_1, & \text{if } q \in \mathbf{K_1} \\ 1/k_2, & \text{if } q \in \mathbf{K_2}. \end{cases}$$

Subramani & Gupta (2014) then claimed that the sample mean, denoted as $\overline{y}_{GMLSS}$, is not an unbiased estimate of the population mean, as the first-order inclusion probabilities are unequal. Note that if sampling is achieved without replacement, then the only condition for obtaining an unbiased estimate of the population mean is that $\pi_q > 0$, for all $q \in \{1, ..., N\}$. We thus conclude that estimator $\overline{y}_{GMLSS}$ is an unbiased estimate of the population mean.

If $k_1 = k_2$, then GMLSS reduces to the designs discussed by Subramani (2013*b*,*a*). Moreover, if $n_2 = 0$, then GMLSS reduces to LSS. When comparing GMLSS to RLSS, one can easily verify that GMLSS reduces to RLSS for any population exhibiting a perfect linear trend, provided that $n_1 = n - r$, $n_2 = r$, $k_1 = k$ and $k_2 = k + 1$. However, the ordering of units in $ST_1$ and $ST_2$ are not equivalent to that in $\mathbf{K_1}$ and $\mathbf{K_2}$, respectively. Thus, GMLSS will not reduce to RLSS when other population structures are considered.

Under the assumption of model (2.12), the MSE of estimator $\overline{y}_{GMLSS}$ is given as

$$M_{GMLSS} = \frac{b^2}{12n^2}\left\{n_1^2(k_1^2-1) - n_2^2(k_2^2-1) + 3n_2^2\left[n_1-(n_2+1)\right]^2(k_1-k_2)^2\right\}.$$

Obtaining comparisons between GMLSS and either SRS, CSS or RLSS, is difficult, as $M_{GMLSS}$ depends on many variables. Subramani & Gupta (2014) thus provides a numerical study which shows that in the presence of linear trend, GMLSS always performs better than SRS and CSS, while in most cases GMLSS is preferred over RLSS. Note that GMLSS is not a linear trend free sampling design. Thus, to improve results for populations exhibiting linear trend, Subramani & Gupta (2014) derived some end corrections estimators, given by

$$\begin{aligned}
\overline{y}^{*}_{GMLSS} &= \overline{y}_{GMLSS} + \zeta_1(y_1 - y_n),\\
\overline{y}^{**}_{GMLSS} &= \overline{y}_{GMLSS} + \zeta_2(y_t - y_{t+1}),\\
\overline{y}^{***}_{GMLSS} &= \overline{y}_{GMLSS} + \zeta_3(y_{t'} - y_{t'+1}),\\
\overline{y}^{****}_{GMLSS} &= \overline{y}_{GMLSS} + \zeta_4(y_1 - y_{n_1})
\end{aligned}$$

and

$$\overline{y}^{*****}_{GMLSS} = \overline{y}_{GMLSS} + \zeta_5(y_{n_1+1} - y_n),$$

where

$$\zeta_1 = \frac{(n_1-n_2)\left[k_1(n_2+1)-k_2n_2\right] + 2\left[k_2n_2 - n_1(i+1) - n_2(j+1)\right]}{2n\left\{(i-j) - \left[k_1 - \hat{k}(n_2-1)\right]\right\}},$$

$$\zeta_2 = \frac{(n_1-n_2)\left[k_1(n_2+1)-k_2n_2\right] + 2\left[k_2n_2 - n_1(i+1) - n_2(j+1)\right]}{2n\hat{k}},$$

$$\zeta_3 = \frac{(n_1-n_2)\left[k_1(n_2+1)-k_2n_2\right] + 2\left[k_2n_2 - n_1(i+1) - n_2(j+1)\right]}{2nk_1},$$

$$\zeta_4 = \frac{(n_1-n_2)\left[k_1(n_2+1)-k_2n_2\right] + 2\left[k_2n_2 - n_1(i+1) - n_2(j+1)\right]}{2n\left[n_2k_2 + (n_1-1)k_1\right]},$$

and

$$\zeta_5 = \frac{2(n_1i + n_2j) - N + n + n_1n_2(k_2-k_1)}{2n\hat{k}(n_2-1)}.$$

Note that corrections are applied to: (i) the first and last sampling units, so as to obtain estimator $\overline{y}^{*}_{GMLSS}$; (ii) two successive sampling units from either $\mathbf{K_1}$ or $\mathbf{K_2}$, so as to construct $\overline{y}^{**}_{GMLSS}$; (iii) the first two sampling units derived from each matrix, or two successive sampling units between the $(n_2+1)$th and the $n_1$th sampling unit, derived from matrix $\mathbf{K_1}$, such that estimator $\overline{y}^{***}_{GMLSS}$ is obtained; (iv) the first and last sampling units derived from matrix $\mathbf{K_1}$, where estimator $\overline{y}^{****}_{GMLSS}$ is then constructed; (v) the first and last sampling units derived from matrix $\mathbf{K_2}$, such that estimator $\overline{y}^{*****}_{GMLSS}$ is obtained. Just as in the case of all previous end corrections estimators mentioned thus far, estimators $\overline{y}^{*}_{GMLSS}$, $\overline{y}^{**}_{GMLSS}$, $\overline{y}^{***}_{GMLSS}$, $\overline{y}^{****}_{GMLSS}$ and $\overline{y}^{*****}_{GMLSS}$, are all equivalent to the population mean when there is a perfect linear trend in the population.

### 4.1.13 Remainder linear systematic sampling with multiple random starts (Mostafa & Ahmad 2016)

This modified systematic sampling design uses multiple random starts, as in the previous chapter, on the RLSS design in Section 4.1.7. The methodology of remainder linear systematic sampling with multiple random starts (RLSSM) is given as follows:

(i) Divide the population into two strata, where the first stratum, $ST_1$, contains the first $(n-r)k$ population units and the second stratum, $ST_2$, contains the remaining $r(k+1)$ units.

(ii) For $ST_1$, select $1 \leq t_1 \leq (n-r)$ distinct integers from the first $t_1 k$ population units, where $(n-r)/t_1$ is an integer. Label these integers using the indices in $C = \{c_1, c_2, ..., c_{t_1}\}$, where $1 \leq c_i \leq t_1 k$ for $i = 1, 2, ..., t_1$. The $(n-r)$ randomly selected units from $ST_1$ are then given as

$S_c = \{c_i + (l'-1)t_1 k | i = 1, ..., t_1 \text{ and } l' = 1, 2, ..., (n-r)/t_1\}.$

(iii) For $ST_2$, select $1 \leq t_2 \leq r$ distinct integers from the set $\{(n-r)k+1, ..., (n-r)k + t_2(k+1)\}$, where $r/t_2$ is an integer. Label these integers using the indices in $D = \{d_1, d_2, ..., d_{t_2}\}$. The $r$ randomly selected units from $ST_2$ are then given as

$S_d = \{d_j + (l''-1)t_2(k+1) | j = 1, ..., t_2 \text{ and } l'' = 1, 2, ..., r/t_2\}.$

(iv) The final sample of size $n$ is then given as $S = S_c \cup S_d$.

If we denote the sample means from $ST_1$ and $ST_2$ as $\overline{y}_c$ and $\overline{y}_d$, respectively, then the RLSSM sample mean, which is a design-unbiased estimate of the population mean, is given as $\overline{y}_{RLSSM} = [(n-r)k\overline{y}_c + r(k+1)\overline{y}_d]/N$. Mostafa & Ahmad (2016) then showed that estimator $\overline{y}_{RLSSM}$ is a linear combination of Horvitz and Thompson (1952) estimators, i.e.

$$\overline{y}_{RLSSM} = N^{-1}\sum_{i\in S_c}\frac{y_i}{\pi_i} + N^{-1}\sum_{j\in S_d}\frac{y_j}{\pi_j}.$$

Using this expression, Mostafa & Ahmad (2016) computed expressions for the variance of estimator $\overline{y}_{RLSSM}$ and an unbiased estimator of the corresponding variance. Now, under model (2.1), the expected MSE of $\overline{y}_{RLSSM}$ is given by

$$M_{RLSSM} = \sigma_r^2 + \frac{b^2k}{12N^2}\{(n-r)^2k(k-1)(t_1k+1) + r^2(k+1)^2[t_2(k+1)+1]\},$$

Mostafa & Ahmad (2016) then presented an empirical study which shows that both RLSS and GMLSS are more efficient than RLSSM in the presence of linear trend.

Now that we have reviewed all the modified systematic sampling designs that tackle the problem of $N$ not being a multiple of $n$, we note that the only linear trend free sampling designs and/or estimators, are those given in Sections 4.1.4 (i.e. Case (A)), 4.1.6, 4.1.7, 4.1.9 (i.e. Case (A) for RBSS), 4.1.10 (i.e. Case (A) for RBSS) and 4.1.12. Moreover, we obtain minimum expected MSE of the associated sample means in Sections 4.1.4 (i.e. Case (A)), 4.1.6 (i.e. MBCSS if $N$ is odd or $n$ is even, and MCCSS if both $N$ and $n$ are odd), 4.1.9 (i.e. Case (A) for RBSS) and 4.1.10 (i.e. Case (A) for RBSS). This has motivated the study in the next section, where a linear trend free sampling design and/or estimator is proposed for the scenario when $N$ is not a multiple of $n$. Under certain cases, the proposed design is a linear trend free sampling design, while an appropriate end corrections estimator is proposed for the other cases.

## 4.2 Remainder modified systematic sampling

For this section, a modified systematic sampling design, termed as remainder modified systematic sampling (RMSS), is proposed. The proposed design extends Chang and Huangs' (2000) RLSS design, such that sampling units are selected according to a mixture of MSS and CESS. Section 4.2.1 contains a discussion of the methodology of RMSS. Under the assumption of a linear trend model, RMSS is compared SRS, RLSS and CSS, in Section 4.2.2. The results suggest that RMSS is a linear trend free sampling design for three of the seven cases of the design. For the other cases, a modified estimator, i.e. RMSS with end corrections (RMSSEC) estimator, is thus constructed in Section 4.2.3.

### 4.2.1 Methodology

The seven cases of RMSS are given in Table 4.1. As is the case with RLSS, we first divide the population into two strata, where the first stratum, $ST_1$, contains the first $(n-r)k$ units and the second stratum, $ST_2$, contains the remaining $r(k+1)$ units. We next need to select $(n-r)$ sampling units from $ST_1$ (note that RLSS applies LSS to select $(n-r)$ sampling units using the sampling interval $k$ in $ST_1$) and $r$ sampling units from $ST_2$ (note that RLSS applies LSS to select $r$ sampling units using the sampling interval $(k+1)$ in $ST_2$).

Table 4.1: Possible cases of RMSS.

| $n$ | $k$ | $r$ | $(n-r)$ | Case |
|---|---|---|---|---|
| even | even | even | even | A |
| even | odd | even | even | A |
| odd | even | odd | even | B |
| odd | odd | even | odd | C |
| odd | odd | odd | even | D |
| odd | even | even | odd | E |
| even | even | odd | odd | F |
| even | odd | odd | odd | G |

We know that MSS is a linear trend free sampling design if the sample size is even. Likewise, CESS is a linear trend free sampling design if the sampling interval is odd. Thus, if we consider Table 4.1, we can easily deduct that MSS should be applied in $ST_1$ for Cases A, B and D, as $(n-r)$ is even, as well as in $ST_2$ for Cases A, C and E, as $r$ is even. Similarly, CESS should be applied in $ST_1$ for Cases C and G, as $k$ is odd, as well as in $ST_2$ for Cases B and F, as $(k+1)$ is odd. Note that if both MSS and CESS offer linear trend free sampling, then MSS is preferred since MSS is a randomized design, unlike CESS which requires no randomization. Thus, we are left with Cases E and F for $ST_1$ and Cases D and G for $ST_2$, of which we will apply MSS. Hence, we can expect RMSS to be a linear trend free sampling design for Cases A, B and C. Note that each case of RMSS entails some form of randomized sampling. Selecting a sample of size $n$ from a population of size $N = nk + r$ using RMSS, consists of the following steps:

(i) Select two random starts $k1$ and $k2$ from 1 to $k$ and 1 to $k+1$, respectively.

(ii) The sample chosen from $ST_1$ is given by

$$\begin{aligned} s_{k1} &= \{U_{k1+jk}, U_{(n-r)k-jk-k1+1} | j = 0, ..., (n-r-2)/2\}, && \text{for A, B and D} \\ &= \{U_{[(2j-1)k+1]/2} | j = 1, ..., n-r\}, && \text{for C and G} \\ &= \{U_{k1+(n-r-1)k/2}\} \\ &\cup \{U_{k1+jk}, U_{(n-r)k-jk-k1+1} | j = 0, ..., (n-r-3)/2\}, && \text{for E and F.} \end{aligned}$$

(iii) The sample chosen from $ST_2$ is given by

$$\begin{aligned} s_{k2} &= \{U_{k2+j(k+1)+(n-r)k}, U_{N-j(k+1)-k2+1} | j = 0, ..., (r-2)/2\}, && \text{for A, C and E} \\ &= \{U_{j+k(j+n-r-1/2)} | j = 1, ..., r\}, && \text{for B and F} \\ &= \{U_{k2+N-(r+1)(k+1)/2}\} \\ &\cup \{U_{k2+j(k+1)+(n-r)k}, U_{N-j(k+1)-k2+1} | j = 0, ..., (r-3)/2\}, && \text{for D and G.} \end{aligned}$$

(iv) The final sample of size $n$ is given by $s = s_{k1} \cup s_{k2}$.

If $n - r = 1$, then the sampling unit for Cases E and F in $ST_1$, is obtained by randomly selecting a unit from the first $k$ units. Similarly, if $r = 1$, then the sampling unit for Cases D and G in $ST_2$, is obtained by randomly selecting a unit from the last $k+1$ units.

For RMSS, the first-order inclusion probability for unit $U_q$ is given by

$$\begin{aligned} \pi_q &= 1/k, && \text{if } q \in \{1, ..., (n-k)r\} \text{ for A, B, D, E and F} \\ &= 1/(k+1), && \text{if } q \in \{(n-k)r+1, ..., N\} \text{ for A, C, D, E and G} \\ &= 1, && \text{if } q \in \{[(2j-1)k+1]/2 | j = 1, ..., n-r\} \text{ for C and G} \\ &= 1, && \text{if } q \in \{j + k(j+n-r-1/2) | j = 1, ..., r\} \text{ for B and F} \\ &= 0, && \text{otherwise.} \end{aligned}$$

Note that $\pi_q = 0$ or 1, for Cases C and G in $ST_1$ and B and F in $ST_2$, as CESS is applied in these instances. Now, let us denote the $k1$th sample mean from $ST_1$ as $\overline{y}_{k1}$ and the $k2$th sample mean from $ST_2$ as $\overline{y}_{k2}$, which are estimates of the stratum means from $ST_1$ and $ST_2$ (denoted by $\overline{Y}_1$ and $\overline{Y}_2$), respectively. Thus, the sample mean for RMSS is obtained by using the first-order inclusion probabilities on the Horvitz-Thompson (1952) estimator, i.e.

$$\hat{\overline{Y}}_{HT} = \frac{(n-r)k\overline{y}_{k1} + r(k+1)\overline{y}_{k2}}{N} = \overline{y}_{RMSS}. \tag{4.6}$$

Note that estimator $\overline{y}_{RMSS}$ is an unbiased estimator of the population mean for Cases A, D and E, since $\pi_q \neq 0$ for all $q$, i.e. MSS is applied in both $ST_1$ and $ST_2$ for Cases A, D and E. Likewise, estimator $\overline{y}_{RMSS}$ is biased for Cases B, C, F and G, since $\pi_q = 0$ for some $q$, i.e. CESS is applied in either $ST_1$ or $ST_2$ for Cases B, C, F and G. However, if we consider Cases B, C, F and G, we can easily verify that under the assumption of a perfect linear trend in the population, estimator $\overline{y}_{RMSS}$ is unbiased for Cases B and C, as $\overline{y}_{k1} = \overline{Y}_1$ and $\overline{y}_{k2} = \overline{Y}_2$, i.e. for Case B, MSS is applied in $ST_1$ where $(n-r)$ is even and CESS is applied in $ST_2$ where $(k+1)$ is odd. Likewise, for Case C, CESS is applied in $ST_1$ where $k$ is odd and MSS is applied in $ST_2$ where $r$ is even. Thus, in each case, both designs offer linear trend free sampling and hence RMSS is a linear trend free sampling design.

Let,

$$\begin{aligned}
s_a &= \{k1 + jk, (n-r)k - jk - k1 + 1 | j = 0, ..., (n-r-2)/2\},\\
s_b &= \{[(2j-1)k+1]/2 | j = 1, ..., n-r\},\\
s_c &= \{k1 + (n-r-1)k/2\}\\
&\quad \cup \{k1 + jk, (n-r)k - jk - k1 + 1 | j = 0, ..., (n-r-3)/2\},\\
s_d &= \{k2 + j(k+1) + (n-r)k, N - j(k+1) - k2 + 1 | j = 0, ..., (r-2)/2\},\\
s_e &= \{j + k(j + n - r - 1/2) | j = 1, ..., r\},\\
s_f &= \{k2 + N - (r+1)(k+1)/2\}\\
&\quad \cup \{k2 + j(k+1) + (n-r)k, N - j(k+1) - k2 + 1 | j = 0, ..., (r-3)/2\}.
\end{aligned}$$

Thus, the second-order inclusion probabilities, $\pi_{qz}$, for the pair of units $(U_q, U_z)$, $q, z \in$

$\{1, ..., N\}(q \neq z)$, are given as follows:

$$
\begin{aligned}
\text{Case A: } \pi_{qz} &= 1/k, && \text{if } q \text{ and } z \in s_a \\
&= 1/k(k+1), && \text{if } q \in s_a \text{ and } z \in s_d\text{, or } q \in s_d \text{ and } z \in s_a \\
&= 1/(k+1), && \text{if } q \text{ and } z \in s_d \\
&= 0, && \text{otherwise.} \\
\text{Case B: } \pi_{qz} &= 1/k, && \text{if } q \text{ and } z \in s_a \\
&= 1/k, && \text{if } q \in s_a \text{ and } z \in s_e\text{, or } q \in s_e \text{ and } z \in s_a \\
&= 1, && \text{if } q \text{ and } z \in s_e \\
&= 0, && \text{otherwise.} \\
\text{Case C: } \pi_{qz} &= 1/(k+1), && \text{if } q \text{ and } z \in s_d \\
&= 1/(k+1), && \text{if } q \in s_b \text{ and } z \in s_d\text{, or } q \in s_d \text{ and } z \in s_b \\
&= 1, && \text{if } q \text{ and } z \in s_b \\
&= 0, && \text{otherwise.} \\
\text{Case D: } \pi_{qz} &= 1/k, && \text{if } q \text{ and } z \in s_a \\
&= 1/k(k+1), && \text{if } q \in s_a \text{ and } z \in s_f\text{, or } q \in s_f \text{ and } z \in s_a \\
&= 1/(k+1), && \text{if } q \text{ and } z \in s_f \\
&= 0, && \text{otherwise.}
\end{aligned}
$$

$$
\begin{aligned}
\text{Case E: } \pi_{qz} &= 1/k, && \text{if } q \text{ and } z \in s_c \\
&= 1/k(k+1), && \text{if } q \in s_c \text{ and } z \in s_d, \text{ or } q \in s_d \text{ and } z \in s_c \\
&= 1/(k+1), && \text{if } q \text{ and } z \in s_d \\
&= 0, && \text{otherwise.} \\
\text{Case F: } \pi_{qz} &= 1/k, && \text{if } q \text{ and } z \in s_c \\
&= 1/k, && \text{if } q \in s_c \text{ and } z \in s_e, \text{ or } q \in s_e \text{ and } z \in s_c \\
&= 1, && \text{if } q \text{ and } z \in s_e \\
&= 0, && \text{otherwise.} \\
\text{Case G: } \pi_{qz} &= 1/(k+1), && \text{if } q \text{ and } z \in s_f \\
&= 1/(k+1), && \text{if } q \in s_b \text{ and } z \in s_f, \text{ or } q \in s_f \text{ and } z \in s_b \\
&= 1, && \text{if } q \text{ and } z \in s_b \\
&= 0, && \text{otherwise.}
\end{aligned}
$$

Hence, it is impossible to obtain an unbiased estimate of the variance of estimator $\overline{y}_{RMSS}$, as certain second-order inclusion probabilities, $\pi_{qz}$, will be zero for each Case. Also, note that the variance of the sample mean is unobtainable when conducting CESS, as there is only one possible sample selected, thus the variance of estimator $\overline{y}_{RMSS}$ is undefined for Cases B, C, F and G.

### 4.2.2 Expected Mean Square Error Comparisons

Under model (2.1), the population mean and stratum means are given by

$$
\overline{Y} = \frac{1}{N}\sum_{q=1}^{N} y_q = a + \frac{b\,(N+1)}{2} + \overline{\overline{e}},
$$

$$
\overline{Y}_1 = \frac{1}{(n-r)k}\sum_{q=1}^{(n-r)k} y_q = a + \frac{b[(n-r)k+1]}{2} + \overline{e}_1
$$

and

$$
\overline{Y}_2 = \frac{1}{r(k+1)}\sum_{q=(n-r)k+1}^{N} y_q = a + \frac{b[(n-r)k+N+1]}{2} + \overline{e}_2,
$$

where

$$\bar{\bar{e}} = \sum_{q=1}^{N} e_q/N, \qquad \bar{e}_1 = \sum_{q=1}^{(n-r)k} e_q/[(n-r)k]$$

and

$$\bar{e}_2 = \sum_{q=(n-r)k+1}^{N} e_q/[(r(k+1)],$$

represent the mean random errors of the population, $ST_1$ and $ST_2$, respectively.

Now, if we consider Case A of RMSS, then by using (2.1) on $s_{k1}$ and $s_{k2}$, we obtain

$$\begin{aligned} \overline{y}_{k1} &= \frac{1}{n-r} \sum_{j=0}^{(n-r-2)/2} \left(y_{k1+jk} + y_{(n-r)k-jk-k1+1}\right) \\ &= a + \frac{b[(n-r)k+1]}{2} + \bar{e}_{k1} \end{aligned}$$

and

$$\begin{aligned} \overline{y}_{k2} &= \frac{1}{r} \sum_{j=0}^{(r-2)/2} \left(y_{k2+j(k+1)+(n-r)k} + y_{N-j(k+1)-k2+1}\right) \\ &= a + \frac{b[(n-r)k+N+1]}{2} + \bar{e}_{k2}, \end{aligned}$$

where $\bar{e}_{k1} = \sum_{j=0}^{(n-r-2)/2}(e_{k1+jk} + e_{(n-r)k-jk-k1+1})/[(n-r)]$ denotes the mean random error from $s_{k1}$ and $\bar{e}_{k2} = \sum_{j=0}^{(r-2)/2}(e_{k2+j(k+1)+(n-r)k} + e_{N-j(k+1)-k2+1})/r$ represents the mean random error from $s_{k2}$. Thus, the expected MSEs of $\overline{y}_{k1}$ and $\overline{y}_{k2}$ are given by

$$\begin{aligned} M_{k1} &= \mathcal{E}\left\{\mathrm{E}\left[(\overline{y}_{k1} - \overline{Y}_1)^2\right]\right\} \\ &= \mathrm{E}\left\{\mathcal{E}\left[(\bar{e}_{k1} - \bar{e}_1)^2\right]\right\} \\ &= \mathrm{E}\left[\mathcal{E}\left(\bar{e}_{k1}^2\right) + \mathcal{E}\left(\bar{e}_1^2\right) - 2\mathcal{E}\left(\bar{e}_{k1}\bar{e}_1\right)\right] \\ &= \frac{\sigma^2}{n-r} + \frac{\sigma^2}{(n-r)k} - \frac{2\sigma^2}{(n-r)k} = \frac{\sigma^2(k-1)}{(n-r)k} \end{aligned}$$

and

$$\begin{aligned} M_{k2} &= \mathcal{E}\left\{\mathrm{E}\left[(\overline{y}_{k2} - \overline{Y}_2)^2\right]\right\} \\ &= \mathrm{E}\left\{\mathcal{E}\left[(\bar{e}_{k2} - \bar{e}_2)^2\right]\right\} \\ &= \mathrm{E}\left[\mathcal{E}\left(\bar{e}_{k2}^2\right) + \mathcal{E}\left(\bar{e}_2^2\right) - 2\mathcal{E}\left(\bar{e}_{k2}\bar{e}_2\right)\right] \\ &= \frac{\sigma^2}{r} + \frac{\sigma^2}{r(k+1)} - \frac{2\sigma^2}{r(k+1)} = \frac{\sigma^2 k}{r(k+1)}, \end{aligned}$$

respectively. Finally, by using $M_{k1}$ and $M_{k2}$ as weighted components, the expected MSE of $\overline{y}_{RMSS}$ can be written as

$$\begin{aligned} M_{RMSS} &= \frac{1}{N^2}\left[(n-r)^2k^2M_{k1} + r^2(k+1)^2M_{k2}\right] \\ &= \frac{\sigma^2}{N^2}\left[(n-r)k(k-1) + r(k+1)k\right] = \frac{\sigma^2 k\,(N-n+r)}{N^2} = \sigma_r^2, \end{aligned}$$

where $\sigma_r^2$ denotes the minimum expected error variance when independently sampling from $ST_1$ and $ST_2$. Similarly, we can find $M_{RMSS}$ for the other cases of the design, such that

$$\begin{aligned} M_{RMSS} &= \sigma_r^2, && \text{for Cases A, B and C} \\ &= \sigma_r^2 + \frac{b^2k(k+2)(k+1)^2}{12N^2}, && \text{for Cases D and G} \\ &= \sigma_r^2 + \frac{b^2k^2(k^2-1)}{12N^2}, && \text{for Cases E and F.} \end{aligned} \tag{4.7}$$

If we compare equations (4.4) to (4.7), it can be shown that $M_{RMSS} < M_{RLSS}$, while simple theoretical comparisons between equations (2.3) and (4.7) are difficult to obtain. Also, the expected MSE of the sample mean when conducting CSS is difficult to obtain, owing to the circular selection procedure. Thus, we will resort to some empirical results below.

Without loss of generality, we consider the $e_i$'s in (2.1) to be iid $N(0,1)$ random variables and set $a = 5$, as expected MSE expressions have been shown to be independent of $a$. Monte Carlo simulations are then employed by means of the statistical software package R, whereby 10 000 finite linear trend populations are simulated. The expected MSE of each estimator is obtained as the mean of the MSEs over the 10 000 populations. The relative expected MSEs of each comparative estimator, with respect to that of estimator $\overline{y}_{RMSS}$, is denoted by $R_\alpha = 100 \times M_{RMSS}/M_\alpha(\%)$, where $\alpha$ = SRS, RLSS, or CSS. The results are presented in Tables 4.2 to 4.8, where we note that estimator $\overline{y}_{RMSS}$ is always subject to less error than estimators $\overline{y}_{SRS}$, $\overline{y}_{RLSS}$ and $\overline{y}_{CSS}$. Moreover, as $N$ and/or $b$ increases, we see further improvements when choosing RMSS over the other sampling designs.

### 4.2.3 Remainder modified systematic sampling with end corrections

To improve results for Cases D to G, an end corrections estimator is constructed to remove the linear trend components in (4.7). The corresponding estimator is given in the next theorem.

Table 4.2: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case A of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R_{RLSS}$ | $R_{SRS}$ | $R_{CSS}$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 10 | 4 | 2 | 2 | 70.71 | 32.97 | 48.13 |
| 0.5 | 28 | 8 | 3 | 4 | 40.74 | 05.70 | 13.98 |
| 0.5 | 38 | 16 | 2 | 6 | 41.81 | 03.39 | 04.44 |
| 0.5 | 104 | 32 | 3 | 8 | 16.92 | 00.45 | 01.44 |
| 1 | 10 | 4 | 2 | 2 | 36.73 | 10.57 | 17.84 |
| 1 | 28 | 8 | 3 | 4 | 14.73 | 01.50 | 03.93 |
| 1 | 38 | 16 | 2 | 6 | 14.96 | 00.86 | 01.13 |
| 1 | 104 | 32 | 3 | 8 | 04.80 | 00.11 | 00.36 |
| 2 | 10 | 4 | 2 | 2 | 12.52 | 02.84 | 05.09 |
| 2 | 28 | 8 | 3 | 4 | 04.13 | 00.38 | 01.01 |
| 2 | 38 | 16 | 2 | 6 | 04.22 | 00.22 | 00.28 |
| 2 | 104 | 32 | 3 | 8 | 01.25 | 00.03 | 00.09 |
| 4 | 10 | 4 | 2 | 2 | 03.47 | 00.73 | 01.32 |
| 4 | 28 | 8 | 3 | 4 | 01.07 | 00.09 | 00.25 |
| 4 | 38 | 16 | 2 | 6 | 01.09 | 00.05 | 00.07 |
| 4 | 104 | 32 | 3 | 8 | 00.31 | 00.01 | 00.02 |

**Theorem 14:** *The* RMSSEC *estimator of* $\overline{Y}$ *with random starts* $k1$ *and* $k2$*, where* $k1 \in \{1, ..., k\}$ *and* $k2 \in \{1, ..., k+1\}$*, is given by*

$$\begin{aligned}
\overline{y}_{RMSSEC} &= \overline{y}_{RMSS} + Z_1 \left( y_{k2+(n-r)k} - y_{N-k2+1} \right), \qquad \text{for Cases D and G } (r > 1), \\
&= \overline{y}_{RMSS} + Z_2 \left( y_{k1} - y_{(n-r)k-k1+1} \right), \qquad \text{for Cases E and F } (n - r > 1),
\end{aligned}$$

where

$$Z_1 = \frac{(k+1)(2k2 - k - 2)}{N[2N - 4k2 + 2 - 2(n-r)k]} \qquad \text{and} \qquad Z_2 = \frac{k(2k1 - k - 1)}{N[2(n-r)k - 4k1 + 2]}.$$

*Proof*: See Appendix.

One can easily verify that $\overline{y}_{RMSSEC} = \overline{Y}$, under the assumption of a perfect linear trend in the population. However, in the presence of a rough linear trend, estimator $\overline{y}_{RMSSEC}$ is biased, owing to the uneven weighting of the sampling units. Under model (2.1), an

Table 4.3: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case B of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R_{RLSS}$ | $R_{SRS}$ | $R_{CSS}$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 11 | 5 | 2 | 1 | 73.41 | 28.67 | 55.84 |
| 0.5 | 39 | 9 | 4 | 3 | 30.61 | 02.99 | 13.35 |
| 0.5 | 39 | 17 | 2 | 5 | 42.05 | 03.15 | 05.47 |
| 0.5 | 139 | 33 | 4 | 7 | 10.06 | 00.25 | 01.66 |
| 1 | 11 | 5 | 2 | 1 | 38.09 | 08.65 | 22.38 |
| 1 | 39 | 9 | 4 | 3 | 10.04 | 00.77 | 03.72 |
| 1 | 39 | 17 | 2 | 5 | 15.29 | 00.81 | 01.42 |
| 1 | 139 | 33 | 4 | 7 | 02.75 | 00.06 | 00.42 |
| 2 | 11 | 5 | 2 | 1 | 13.70 | 02.34 | 06.71 |
| 2 | 39 | 9 | 4 | 3 | 02.73 | 00.19 | 00.96 |
| 2 | 39 | 17 | 2 | 5 | 04.46 | 00.21 | 00.37 |
| 2 | 139 | 33 | 4 | 7 | 00.70 | 00.02 | 00.11 |
| 4 | 11 | 5 | 2 | 1 | 03.76 | 00.59 | 01.75 |
| 4 | 39 | 9 | 4 | 3 | 00.70 | 00.05 | 00.24 |
| 4 | 39 | 17 | 2 | 5 | 01.14 | 00.05 | 00.09 |
| 4 | 139 | 33 | 4 | 7 | 00.18 | $<$ 00.01 | 00.03 |

expression for the expected MSE of estimator $\overline{y}_{RMSSEC}$ (i.e. $M_{RMSSEC}$) is complex. We will thus consider a simulation study to evaluate the performance of estimator $\overline{y}_{RMSSEC}$.

We next consider a similar simulation study, as in the previous section. The relative expected MSEs of estimators $\overline{y}_{RMSS}$ and $\overline{y}_{RLSSEC}$, with respect to that of $\overline{y}_{RMSSEC}$, is denoted by $R'_\beta = 100 \times M_{RMSSEC}/M_\beta(\%)$, where $\beta$ = RMSS or RLSSEC. The simulation results are presented in Tables 4.9 and 4.10, where we see the superiority of estimator $\overline{y}_{RMSSEC}$ over $\overline{y}_{RMSS}$, with greater discrepancies as $b$ increases. Furthermore, we note that estimator $\overline{y}_{RMSSEC}$ is preferred over estimator $\overline{y}_{RLSSEC}$, with smaller discrepancies as the population size increases, i.e. there are marginal gains when selecting estimator $\overline{y}_{RMSSEC}$ over estimator $\overline{y}_{RLSSEC}$, for large scale sampling.

Table 4.4: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case C of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R_{RLSS}$ | $R_{SRS}$ | $R_{CSS}$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 11 | 3 | 3 | 2 | 57.83 | 27.21 | 43.24 |
| 0.5 | 39 | 7 | 5 | 4 | 25.29 | 03.02 | 11.08 |
| 0.5 | 51 | 15 | 3 | 6 | 29.38 | 01.83 | 04.76 |
| 0.5 | 163 | 31 | 5 | 8 | 07.91 | 00.18 | 01.40 |
| 1 | 11 | 3 | 3 | 2 | 25.74 | 08.46 | 15.55 |
| 1 | 39 | 7 | 5 | 4 | 07.71 | 00.76 | 02.98 |
| 1 | 51 | 15 | 3 | 6 | 09.39 | 00.46 | 01.23 |
| 1 | 163 | 31 | 5 | 8 | 02.07 | 00.05 | 00.35 |
| 2 | 11 | 3 | 3 | 2 | 08.02 | 02.26 | 04.42 |
| 2 | 39 | 7 | 5 | 4 | 02.05 | 00.19 | 00.76 |
| 2 | 51 | 15 | 3 | 6 | 02.57 | 00.12 | 00.32 |
| 2 | 163 | 31 | 5 | 8 | 00.53 | 00.01 | 00.09 |
| 4 | 11 | 3 | 3 | 2 | 02.14 | 00.58 | 01.15 |
| 4 | 39 | 7 | 5 | 4 | 00.52 | 00.05 | 00.19 |
| 4 | 51 | 15 | 3 | 6 | 00.65 | 00.03 | 00.08 |
| 4 | 163 | 31 | 5 | 8 | 00.13 | $< 00.01$ | 00.02 |

### 4.2.4 Concluding Remarks

A sampling design, namely RMSS, has been proposed by extending the RLSS design and selecting units according to a mixture of MSS and CESS. Thus, RMSS is applicable when the population size is not a multiple of the sample size and is appropriate for populations exhibiting linear trend. In the presence of linear trend, the RMSS sample mean is subject to less error than that of those provided by SRS, CSS and RLSS, as seen in Section 4.2.2. However, linear trend free sampling results are only obtained for three out of the seven cases of RMSS. In Section 4.2.3, an end corrections estimator is thus constructed for the other four cases. The simulation study conducted in Section 4.2.3, illustrates the superiority of this end corrections estimator over the RMSS sample mean. Further results from this study indicate that the proposed end corrections estimator is susceptible to less error than an end corrections estimator associated with RLSS. However, the expected

Table 4.5: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case D of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R_{RLSS}$ | $R_{SRS}$ | $R_{CSS}$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 18 | 5 | 3 | 3 | 53.50 | 13.90 | 28.48 |
| 0.5 | 32 | 9 | 3 | 5 | 38.12 | 04.70 | 10.29 |
| 0.5 | 82 | 15 | 5 | 7 | 16.04 | 00.76 | 03.51 |
| 0.5 | 226 | 31 | 7 | 9 | 04.72 | 00.10 | 01.06 |
| 1 | 18 | 5 | 3 | 3 | 27.02 | 04.94 | 11.32 |
| 1 | 32 | 9 | 3 | 5 | 15.15 | 01.42 | 03.24 |
| 1 | 82 | 15 | 5 | 7 | 05.46 | 00.23 | 01.09 |
| 1 | 226 | 31 | 7 | 9 | 01.43 | 00.03 | 00.31 |
| 2 | 18 | 5 | 3 | 3 | 14.73 | 02.37 | 05.60 |
| 2 | 32 | 9 | 3 | 5 | 06.66 | 00.57 | 01.31 |
| 2 | 82 | 15 | 5 | 7 | 02.33 | 00.09 | 00.45 |
| 2 | 226 | 31 | 7 | 9 | 00.57 | 00.01 | 00.12 |
| 4 | 18 | 5 | 3 | 3 | 11.13 | 01.37 | 04.12 |
| 4 | 32 | 9 | 3 | 5 | 04.18 | 00.35 | 00.81 |
| 4 | 82 | 15 | 5 | 7 | 01.53 | 00.06 | 00.29 |
| 4 | 226 | 31 | 7 | 9 | 00.35 | 00.01 | 00.07 |

MSEs of both end corrections estimators tend to converge as the population size increases. Thus, one can use either end corrections estimator for large-scale sampling applications. For this scenario, one may opt to use the end corrections estimator associated with RLSS, owing to its simplicity.

In this chapter, we have discussed the the second of two shortcomings of systematic sampling, i.e. if the population size is not a multiple of the sample size, resulting in sample sizes that vary, or fixed sample sizes that are greater than required when conducting LSS. Also, we reviewed relevant modifications of the usual systematic sampling design found in literature as well as a suggested modified systematic sampling design that address this problem, all under the assumption of linear trend among the population units. The results from this chapter suggest that values of the population size, sample size and/or the remainder needs to be considered for the relevant shortcoming at hand, before selecting

Table 4.6: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case E of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R_{RLSS}$ | $R_{SRS}$ | $R_{CSS}$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 12 | 5 | 2 | 2 | 68.99 | 25.39 | 41.73 |
| 0.5 | 42 | 9 | 4 | 6 | 24.77 | 02.71 | 07.25 |
| 0.5 | 98 | 15 | 6 | 8 | 11.54 | 00.52 | 02.64 |
| 0.5 | 258 | 31 | 8 | 10 | 03.79 | 00.08 | 00.84 |
| 1 | 12 | 5 | 2 | 2 | 37.12 | 08.08 | 15.79 |
| 1 | 42 | 9 | 4 | 6 | 08.26 | 00.75 | 02.08 |
| 1 | 98 | 15 | 6 | 8 | 03.57 | 00.15 | 00.77 |
| 1 | 258 | 31 | 8 | 10 | 01.10 | 00.02 | 00.24 |
| 2 | 12 | 5 | 2 | 2 | 14.77 | 02.49 | 05.15 |
| 2 | 42 | 9 | 4 | 6 | 02.95 | 00.26 | 00.71 |
| 2 | 98 | 15 | 6 | 8 | 01.36 | 00.06 | 00.29 |
| 2 | 258 | 31 | 8 | 10 | 00.40 | 00.01 | 00.09 |
| 4 | 12 | 5 | 2 | 2 | 06.30 | 00.98 | 02.07 |
| 4 | 42 | 9 | 4 | 6 | 01.50 | 00.13 | 00.36 |
| 4 | 98 | 15 | 6 | 8 | 00.78 | 00.03 | 00.16 |
| 4 | 258 | 31 | 8 | 10 | 00.22 | $< 00.01$ | 00.05 |

an appropriate modified systematic sampling design in the presence of linear trend. The sampling designs in Sections 4.1.3, 4.1.4 (i.e. BRS), 4.1.5, 4.1.8, 4.1.9 (i.e. RSSS), 4.1.10 (i.e. RSSS) and 4.1.13, also solve the problem of unbiased variance estimation, as discussed in the previous chapter. However, most of these designs do not offer favourable results in the presence of linear trend and are not simple to apply in real life situations. If we consider all the designs discussed in this chapter, then we note that BRS (for Case (A)) is the only design that allows us to obtain an unbiased estimate of the associated sampling variance, while providing linear trend free sampling results as well as offering simplicity in it's application. This has motivated the study in Chapter 5, where a modification to the BRS design is proposed, which addresses all the shortcomings of systematic sampling, while proving to be, suitable for populations exhibiting linear trend and simple to implement.

Table 4.7: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case F of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R_{RLSS}$ | $R_{SRS}$ | $R_{CSS}$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 9 | 4 | 2 | 1 | 78.03 | 37.66 | 61.65 |
| 0.5 | 35 | 8 | 4 | 3 | 35.30 | 03.92 | 15.66 |
| 0.5 | 101 | 16 | 6 | 5 | 11.38 | 00.49 | 03.99 |
| 0.5 | 263 | 32 | 8 | 7 | 03.22 | 00.07 | 01.17 |
| 1 | 9 | 4 | 2 | 1 | 49.31 | 13.82 | 29.38 |
| 1 | 35 | 8 | 4 | 3 | 13.09 | 01.12 | 04.88 |
| 1 | 101 | 16 | 6 | 5 | 03.55 | 00.14 | 01.18 |
| 1 | 263 | 32 | 8 | 7 | 00.94 | 00.02 | 00.33 |
| 2 | 9 | 4 | 2 | 1 | 22.44 | 04.61 | 10.97 |
| 2 | 35 | 8 | 4 | 3 | 05.11 | 00.40 | 01.80 |
| 2 | 101 | 16 | 6 | 5 | 01.35 | 00.05 | 00.44 |
| 2 | 263 | 32 | 8 | 7 | 00.34 | 00.01 | 00.12 |
| 4 | 9 | 4 | 2 | 1 | 11.20 | 02.08 | 05.12 |
| 4 | 35 | 8 | 4 | 3 | 02.87 | 00.22 | 00.99 |
| 4 | 101 | 16 | 6 | 5 | 00.79 | 00.03 | 00.26 |
| 4 | 263 | 32 | 8 | 7 | 00.19 | $< 00.01$ | 00.07 |

Table 4.8: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case G of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R_{RLSS}$ | $R_{SRS}$ | $R_{CSS}$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 15 | 4 | 3 | 3 | 52.83 | 18.78 | 32.67 |
| 0.5 | 43 | 8 | 5 | 3 | 29.71 | 02.87 | 13.59 |
| 0.5 | 117 | 16 | 7 | 5 | 09.31 | 00.38 | 03.57 |
| 0.5 | 295 | 32 | 9 | 7 | 02.74 | 00.06 | 01.09 |
| 1 | 15 | 4 | 3 | 3 | 27.44 | 07.16 | 14.01 |
| 1 | 43 | 8 | 5 | 3 | 12.53 | 01.00 | 05.11 |
| 1 | 117 | 16 | 7 | 5 | 03.28 | 00.13 | 01.21 |
| 1 | 295 | 32 | 9 | 7 | 00.86 | 00.02 | 00.34 |
| 2 | 15 | 4 | 3 | 3 | 15.75 | 03.71 | 07.49 |
| 2 | 43 | 8 | 5 | 3 | 06.80 | 00.51 | 02.67 |
| 2 | 117 | 16 | 7 | 5 | 01.62 | 00.06 | 00.59 |
| 2 | 295 | 32 | 9 | 7 | 00.38 | 00.01 | 00.15 |
| 4 | 15 | 4 | 3 | 3 | 12.02 | 02.74 | 05.59 |
| 4 | 43 | 8 | 5 | 3 | 05.23 | 00.39 | 02.06 |
| 4 | 117 | 16 | 7 | 5 | 01.19 | 00.04 | 00.43 |
| 4 | 295 | 32 | 9 | 7 | 00.19 | $< 00.01$ | 00.07 |

Table 4.9: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Cases D and G of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R'_{RMSS}$ | $R'_{RLSSEC}$ |
|---|---|---|---|---|---|---|
| 0.5 | 18 | 5 | 3 | 3 | 91.90 | 53.01 |
| 0.5 | 43 | 8 | 5 | 3 | 88.85 | 95.56 |
| 0.5 | 82 | 15 | 5 | 7 | 93.53 | 96.90 |
| 0.5 | 295 | 32 | 9 | 7 | 92.33 | 99.35 |
| 1 | 18 | 5 | 3 | 3 | 71.88 | 52.65 |
| 1 | 43 | 8 | 5 | 3 | 64.47 | 94.52 |
| 1 | 82 | 15 | 5 | 7 | 77.88 | 96.24 |
| 1 | 295 | 32 | 9 | 7 | 74.57 | 99.46 |
| 2 | 18 | 5 | 3 | 3 | 37.97 | 51.82 |
| 2 | 43 | 8 | 5 | 3 | 31.42 | 95.40 |
| 2 | 82 | 15 | 5 | 7 | 46.88 | 96.37 |
| 2 | 295 | 32 | 9 | 7 | 42.38 | 99.02 |
| 4 | 18 | 5 | 3 | 3 | 13.35 | 52.67 |
| 4 | 43 | 8 | 5 | 3 | 10.23 | 95.47 |
| 4 | 82 | 15 | 5 | 7 | 18.03 | 96.36 |
| 4 | 295 | 32 | 9 | 7 | 15.55 | 99.19 |

Table 4.10: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Cases E and F of RMSS.

| $b$ | $N$ | $n$ | $k$ | $r$ | $R'_{RMSS}$ | $R'_{RLSSEC}$ |
|---|---|---|---|---|---|---|
| 0.5 | 12 | 5 | 2 | 2 | 99.28 | 82.88 |
| 0.5 | 35 | 8 | 4 | 3 | 96.15 | 94.76 |
| 0.5 | 98 | 15 | 6 | 8 | 95.48 | 95.00 |
| 0.5 | 263 | 32 | 8 | 7 | 95.73 | 99.45 |
| 1 | 12 | 5 | 2 | 2 | 95.68 | 82.82 |
| 1 | 35 | 8 | 4 | 3 | 86.23 | 95.43 |
| 1 | 98 | 15 | 6 | 8 | 83.81 | 95.55 |
| 1 | 263 | 32 | 8 | 7 | 84.78 | 99.47 |
| 2 | 12 | 5 | 2 | 2 | 82.30 | 82.69 |
| 2 | 35 | 8 | 4 | 3 | 60.06 | 94.15 |
| 2 | 98 | 15 | 6 | 8 | 56.61 | 95.39 |
| 2 | 263 | 32 | 8 | 7 | 58.78 | 99.36 |
| 4 | 12 | 5 | 2 | 2 | 53.35 | 82.29 |
| 4 | 35 | 8 | 4 | 3 | 27.06 | 93.96 |
| 4 | 98 | 15 | 6 | 8 | 24.57 | 95.15 |
| 4 | 263 | 32 | 8 | 7 | 26.22 | 99.48 |

# Chapter 5

# Balanced Centered Random Sampling

In this chapter, a modification of BRS, termed as balanced centered random sampling (BCRS), is proposed. The methodologies for the various cases of the proposed design are explained in Section 5.1. For Section 5.2, a simulation study is carried out to compare the efficiency of BCRS, to that of SRS, LSS, CSS, STR, RLSS, RMSS, BRS, MBCSS, MCCSS, as well as the multiple-start modified LSS designs/estimators discussed in Sections 3.1.5 and 3.2 and end corrections estimators RLSSEC, RMSSEC, MBCSSEC and MCCSSEC. Finally, we provide a summary in Section 5.3.

## 5.1 Methodology

BCRS is an extension to Singh and Garg's (1979) BRS, which is divided into four cases. The sampling procedure for the respective cases is given as follows:

(A) if $N$ and $n > 6$ are both even and $N > n + 2$, then:

- (i) use SRS to select $(n-4)/2$ sampling units from the first $N/2-3$ population units, with unit indices $f_i$, for $i = 1, ..., (n-4)/2$;
- (ii) select two sampling units from the sample space $\{U_{N/2-2}, U_{N/2-1}, U_{N/2}\}$ using SRS and label these indices as $f_{(n-2)/2}$ and $f_{n/2}$;
- (iii) the unit indices for the remaining $n/2$ sampling units are then given as $N-f_i+1$, for $i = 1, ..., n/2$;

(B) if $N$ is even, $n > 5$ is odd and $N > n + 1$, then:

 (i) use SRS to select $(n-3)/2$ sampling units from the first $(N-4)/2$ population units, with unit indices $f_i$, for $i = 1, ..., (n-3)/2$;

 (ii) the unit indices for the next $(n-3)/2$ sampling units are given as $N - f_i + 1$, for $i = 1, ..., (n-3)/2$;

 (iii) use SRS to select the remaining three sampling units from the sample space $\{U_{N/2-1}, U_{N/2}, U_{N/2+1}, U_{N/2+2}\}$;

(C) if $N > 9$ and $n > 5$ are both odd and $N > n + 3$, then:

 (i) use SRS to select $(n-3)/2$ sampling units from the first $(N-5)/2$ population units, with unit indices $f_i$, for $i = 1, ..., (n-3)/2$;

 (ii) the unit indices for the next $(n-3)/2$ sampling units are given as $N - f_i + 1$, for $i = 1, ..., (n-3)/2$;

 (iii) use SRS to select the remaining three sampling units from the sample space $\{U_{(N-3)/2}, U_{(N-1)/2}, U_{(N+1)/2}, U_{(N+3)/2}, U_{(N+5)/2}\}$;

(D) if $N$ is odd, $n > 4$ is even and $N > n + 1$, then:

 (i) use SRS to select $(n-2)/2$ sampling units from the first $(N-3)/2$ population units, with unit indices $f_i$, for $i = 1, ..., (n-2)/2$;

 (ii) the unit indices for the next $(n-2)/2$ sampling units are given as $N - f_i + 1$, for $i = 1, ..., (n-2)/2$;

 (iii) use SRS to select the remaining two sampling units from the sample space $\{U_{(N-1)/2}, U_{(N+1)/2}, U_{(N+3)/2}\}$.

Note that only cases (A) to (C) are applicable when the population size is a multiple of the sample size, while all cases may be applicable when the population size is not a multiple of the sample size.

**Theorem 15:** *For the respective cases of* BCRS, *the first-order inclusion probabilities* $\pi_q$ *for the unit* $U_q$ *and the second-order inclusion probabilities* $\pi_{qz}$ *for the pair of units* $\{U_q, U_z\}$, $q, z \in \{1, ..., N\} (q \neq z)$, *are given as follows:*

(A) *if* $N$ *and* $n > 6$ *are both even, where* $N > n+2$ *and* $A = \{N/2-2, N/2-1, ..., N/2+$

3}, *then*:

$$\pi_q = 4/6, \qquad \text{if } q \in A$$
$$= (n-4)/(N-6), \qquad \text{otherwise}$$

*and*

$$\pi_{qz} = (n-4)/(N-6), \qquad \text{if } q+z = N+1 \text{ and } q,z \notin A$$
$$= (n-4)(n-6)/(N-6)(N-8), \qquad \text{if } q+z \neq N+1 \text{ and } q,z \notin A$$
$$= 4/6, \qquad \text{if } q+z = N+1 \text{ and } q,z \in A$$
$$= 1/3, \qquad \text{if } q+z \neq N+1 \text{ and } q,z \in A$$
$$= 4(n-4)/6(N-6), \qquad \text{otherwise;}$$

(B) *if* $N$ *is even and* $n > 5$ *is odd, where* $N > n+1$ *and* $B = \{N/2-1, ..., N/2+2\}$, *then*:

$$\pi_q = 3/4, \qquad \text{if } q \in B$$
$$= (n-3)/(N-4), \qquad \text{otherwise}$$

*and*

$$\pi_{qz} = (n-3)/(N-4), \qquad \text{if } q+z = N+1 \text{ and } q,z \notin B$$
$$= (n-3)(n-5)/(N-4)(N-6), \qquad \text{if } q+z \neq N+1 \text{ and } q,z \notin B$$
$$= 1/2, \qquad \text{if } q,z \in B$$
$$= 3(n-3)/4(N-4), \qquad \text{otherwise;}$$

(C) *if* $N > 9$ *and* $n > 5$ *is odd, where* $N > n+3$ *and* $C = \{(N-3)/2, ..., (N+5)/2\}$, *then*:

$$\pi_q = 3/5, \qquad \text{if } q \in C$$
$$= (n-3)/(N-5), \qquad \text{otherwise}$$

*and*

$$\pi_{qz} = (n-3)/(N-5), \qquad \text{if } q+z = N+1 \text{ and } q,z \notin C$$
$$= (n-3)(n-5)/(N-5)(N-7), \qquad \text{if } q+z \neq N+1 \text{ and } q,z \notin C$$
$$= 3/10, \qquad \text{if } q,z \in C$$
$$= 3(n-3)/5(N-5), \qquad \text{otherwise;}$$

(D) *if $N$ is odd and $n > 4$ is even, where $N > n+1$ and $D = \{(N-1)/2, (N+1)/2, (N+3)/2\}$, then*:

$$\begin{aligned} \pi_q &= 2/3, && \textit{if } q \in D \\ &= (n-2)/(N-3), && \textit{otherwise} \end{aligned}$$

*and*

$$\begin{aligned} \pi_{qz} &= (n-2)/(N-3), && \textit{if } q+z = N+1 \textit{ and } q,z \notin D \\ &= (n-2)(n-4)/(N-3)(N-5), && \textit{if } q+z \neq N+1 \textit{ and } q,z \notin D \\ &= 1/3, && \textit{if } q,z \in D \\ &= 2(n-2)/3(N-3), && \textit{otherwise.} \end{aligned}$$

*Proof*: One can easily prove the above theorem by using the basic derivation of the inclusion probabilities under SRS and first principles.

Using the above inclusion probabilities, the Horvitz-Thompson (1952) estimate of the population mean is given as

$$\overline{y}_{BCRS} = \hat{\overline{Y}}_{HT} = \frac{1}{n} \sum_{U_q \in s} \frac{y_q}{\pi_q},$$

the BCRS sample mean with a Yates-Grundy (1953) form variance

$$V(\overline{y}_{BCRS}) = \frac{1}{N^2} \sum_{q=1}^{N} \sum_{z>q}^{N} \left(1 - \frac{N^2}{n^2}\pi_{qz}\right)(y_q - y_z)^2, \tag{5.1}$$

which is estimated by

$$\hat{V}(\overline{y}_{BCRS}) = \frac{1}{2N^2} \sum_{U_q \in s} \sum_{\substack{U_z \in s \\ z \neq q}} \left(\frac{1}{\pi_{qz}} - \frac{N^2}{n^2}\right)(y_q - y_z)^2,$$

where $s$ is the sample selected.

Now, expression (5.1) can be written as

$$V(\overline{y}_{BCRS}) = \frac{1}{N^2} \sum_{q=1}^{N} \sum_{z>q}^{N} \left[1 - \frac{\pi_{qz}}{\pi_q^2}\right](y_q - y_z)^2 \tag{5.2}$$

Note that it is difficult to simplify (5.2) further, owing to the complex structure of the second-order inclusion probabilities for the proposed design, i.e. simple theoretical expected MSE comparisons between estimator $\overline{y}_{BCRS}$ and other comparative estimators are difficult to obtain. Thus, we will resort to a simulation study in the next section.

## 5.2 Empirical Comparisons

Without loss of generality, we consider the $e_i$'s in (2.1) to be iid $N(0,1)$ random variables and set $a = 5$. Monte Carlo simulations are then employed by means of the statistical software package R, whereby 10 000 finite linear trend populations are simulated. The expected MSE of each estimator is obtained as the mean of the MSEs over the 10 000 populations. The relative expected MSEs of each comparative estimator, with respect to that of estimator $\overline{y}_{BCRS}$, is denoted by $R_\alpha = 100 \times M_{BCRS}/M_\alpha(\%)$, where $M_{BCRS}$ denotes the expected MSE of the BCRS sample mean under model (2.1) and $\alpha$ = SRS, LSS, STR, CSS, MLSS, MYEC, MBSS, MMSS, MBMSS, MBMSSEC, RLSS, RMSS, RMSSEC, BRS, MBCSS, MBCSSEC, MCCSS and MCCSSEC. The results are presented in Tables 5.1 to 5.12. To compare BCRS to the multiple-start designs we let $n = n'm$ and $k = l$, i.e. we are selecting $m$ samples of size $n'$ from the $k = l$ possible samples. While we attempted for four variations of variables for each of the four Cases (for both $N = nk$ and $N \neq nk$ scenarios), we note that larger values of $N$ and $n$ were not possible for Cases B and C with the scenario of $N = nk$, as seen in Tables 5.5 and 5.8, respectively. This is due to the memory limitations when executing the relative R code on the computer.

From Tables 5.1 to 5.12, we note that BCRS is more efficient than SRS for all cases and more efficient than BRS for Cases B, C and D, while being equally efficient to BRS for Case A. Also, if we consider Tables 5.1, 5.2, 5.5 and 5.8, we see that BCRS is more efficient than LSS, STR and MLSS. By analysing the results in Tables 5.3, 5.4, 5.6, 5.7, 5.9, 5.10, 5.11 and 5.12, we note that BCRS is more efficient than CSS and RLSS. By choosing BCRS over LSS, SRS, STR (Cases B and C), MLSS, BRS (Cases B to D) and RLSS (Cases B to D), we see that there are more efficiency gains as $b$ and/or $N$ and/or $n$ increases. Also, if we opt for BCRS over STR (Case A), CSS and RLSS (Case A), we note that relative results are improved as $b$ increases.

Now, let us compare BCRS to BMSS and the BMSSEC estimator. In Tables 5.1 and 5.2 (i.e. Case A) we note that $n = 8$ and 12 are related to Case A of BMSS and $n = 10$ and 14 are related to Case B of BMSS. Hence, the BMSSEC estimator is not applicable for $n = 8$ and 12. Clearly BCRS is equally efficient to BMSS for Case A of BMSS for low values of $b$, while more efficient than BMSS as $b$ and/or $n$ increases. For Case B of BMSS, we see that BCRS is more efficient that BMSS. Also, BCRS is approximately equally efficient to BMSSEC for Case B of BMSS. If we consider Tables 5.5 and 5.8 (i.e.

Cases B and C), we note that $n = 9$ is related to Case E of BMSS and $n = 15$ is related to Case D of BMSS. For these scenarios, both BMSS and the BMMSSEC estimator are more efficient than BCRS with greater efficiency gains when selecting BMSS and the BMSSEC estimator over BCRS as $b$ increases.

Next, consider the comparisons of BCRS to the modified multiple-start designs, i.e. MBSS, MMSS, MYEC, MBMSS, and the MBMSSEC estimator. In Tables 5.1 and 5.2 we note that $n' = 4, 5, 6$ and 7 relate to Cases A, E, B and D of MBMSS, respectively. Thus, the MBMSSEC estimator is not applicable when $n' = 4$. The results are expected, as BCRS, which is a trend free sampling design for Case A, is equally efficient to MBSS ($n'$ is even), MMSS ($n'$ is even), MBMSS ($n' = 4$ or Case A of MBMSS) and the MBSSEC estimator. Furthermore, BCRS is more efficient than MBSS ($n'$ is odd), MMSS ($n'$ is odd), MYEC and MBMSS ($n' = 5, 6$ and 7 or Cases E, B and D of MBMSS). If we examine Tables 5.5 and 5.8 we note that $n' = 3$ and 5 relate to Cases C and E of MBMSS, respectively. Again, the results are expected, as BCRS is more efficient than MBSS, MMSS and MBMSS, since all these multiple-start designs are equally efficient when $n'$ is odd. Note that there are grater efficiency gains when choosing BCRS over MBSS, MMSS or MBMSS as $b$ increases. Moreover, BCRS is less efficient than estimators MYEC and MBMSSEC with greater efficiency gains when selecting estimators MYEC and MBMSSEC over BCRS as $b$ increases.

Next, let us compare BCRS to estimators RLSSEC and RMSSEC as well as RMSS. If we consider Tables 5.3 and 5.4 (i.e. Case A), then we note that Case A of RMSS is related. For this scenario, BCRS is more efficient than RMSS with greater efficiency gains as $b$ increases. Next, let us examine Tables 5.6 and 5.7 (i.e. Case B), where $N = 16$ and 20 represent Case E of RMSS, while $N = 24$ and 30 represent Case D of RMSS. Here, BCRS is more efficient than the RLSSEC estimator if $b = 0.5$, while less efficient if $b = 1, 2$ and 4. Similarly, BCRS is approximately equally efficient to the RMSSEC if $b = 0.5$, while less efficient if $b = 1, 2$ and 4. We see greater efficiency gains when selecting either the RLSSEC or RMSSEC estimator, over BCRS as $b$ increases. For Case E of RMSS, BCRS is approximately equally efficient to RMSS when $b = 0.5$, while less efficient if $b = 1, 2$ and 4. We note greater efficiency gains when opting for RMSS over BCRS as $b$ increases. For Case D of RMSS, BCRS is more efficient than RMSS with greater efficiency gains as $b$ increases. Now, let us analyse Tables 5.9 and 5.10 (i.e. Case C), where $N = 17$ and 21 relate to Case B of RMSS, while $N = 23$ and 29 relate to Case C of RMSS. Clearly RMSS is more

Table 5.1: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case A of BCRS ($N = nk$ and $b = 0.5, 1$).

| $b$ | 0.5 | 0.5 | 0.5 | 0.5 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 16 | 20 | 24 | 28 | 16 | 20 | 24 | 28 |
| $n, n'$ | 8, 4 | 10, 5 | 12, 6 | 14, 7 | 8, 4 | 10, 5 | 12, 6 | 14, 7 |
| $k = l$ | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| $m$ | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| $R_{LSS}$ | 48.86 | 44.27 | 40.04 | 35.53 | 19.74 | 16.41 | 14.26 | 12.53 |
| $R_{SRS}$ | 14.95 | 10.28 | 07.39 | 05.56 | 04.22 | 02.76 | 01.97 | 01.45 |
| $R_{STR}$ | 88.38 | 88.57 | 88.66 | 88.52 | 66.44 | 66.10 | 66.61 | 66.82 |
| $R_{BMSS}$ | 100.93 | 92.97 | 100.21 | 96.49 | 101.44 | 84.81 | 100.77 | 86.98 |
| $R_{BMSSEC}$ | N/A | 97.54 | N/A | 100.23 | N/A | 102.50 | N/A | 99.91 |
| $R_{MLSS}$ | 37.09 | 32.25 | 28.61 | 25.38 | 13.01 | 10.67 | 09.10 | 07.91 |
| $R_{MBSS}$ | 101.22 | 91.57 | 100.14 | 94.06 | 99.24 | 74.07 | 99.69 | 81.94 |
| $R_{MMSS}$ | 99.20 | 90.90 | 98.69 | 94.41 | 100.64 | 74.38 | 100.31 | 81.72 |
| $R_{MYEC}$ | 95.07 | 96.25 | 98.43 | 98.17 | 95.46 | 97.06 | 97.70 | 98.79 |
| $R_{MBMSS}$ | 99.22 | 91.84 | 77.65 | 94.87 | 99.25 | 74.70 | 47.82 | 81.59 |
| $R_{MBMSSEC}$ | N/A | 99.49 | 98.75 | 100.46 | N/A | 99.93 | 100.17 | 100.42 |
| $R_{BRS}$ | 99.99 | 99.94 | 99.94 | 99.98 | 100.04 | 100.09 | 100.08 | 99.95 |

efficient than BCRS with greater efficiency gains as $b$ increases. For the final scenario, we will evaluate Tables 5.11 and 5.12 (i.e. Case D), where $N = 15$ and 19 associate with Case F of RMSS, while $N = 21$ and 27 associate with Case G of RMSS. Here, BCRS is more often than not, more efficient than the RLSSEC estimator when $b = 0.5$ and 1, while BCRS is less efficient than the RLSSEC estimator if $b = 2$ and 4. Likewise, BCRS is more often than not, slightly more efficient than the RMSSEC estimator when $b = 0.5$, while BCRS is less efficient than the RMSSEC estimator if $b = 1, 2$ and 4. Note that there are greater efficiency gains when selecting either the RLSSEC or RMSSEC estimator, over BCRS as $b$ increases. For Case F of RMSS, BCRS is more efficient than RMSS when $b = 0.5$, while less efficient if $b = 1, 2$ and 4. Note that there are greater efficiency gains when opting for BCRS over RMSS, as $b$ increases. Conversely, BCRS is more efficient than RMSS for Case G of RMSS, with greater efficiency gains as $b$ increases.

Table 5.2: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case A of BCRS ($N = nk$ and $b = 2, 4$).

| $b$ | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 16 | 20 | 24 | 28 | 16 | 20 | 24 | 28 |
| $n, n'$ | 8, 4 | 10, 5 | 12, 6 | 14, 7 | 8, 4 | 10, 5 | 12, 6 | 14, 7 |
| $k = l$ | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| $m$ | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| $R_{LSS}$ | 05.93 | 04.72 | 04.00 | 03.45 | 01.55 | 01.23 | 01.02 | 00.89 |
| $R_{SRS}$ | 01.09 | 00.71 | 00.49 | 00.37 | 00.28 | 00.18 | 00.12 | 00.09 |
| $R_{STR}$ | 33.60 | 33.03 | 33.30 | 33.39 | 11.22 | 11.05 | 11.01 | 11.14 |
| $R_{BMSS}$ | 99.50 | 54.23 | 96.81 | 64.20 | 102.13 | 23.83 | 95.89 | 30.63 |
| $R_{BMSSEC}$ | N/A | 97.75 | N/A | 102.03 | N/A | 99.46 | N/A | 102.29 |
| $R_{MLSS}$ | 03.63 | 02.90 | 02.43 | 02.10 | 00.94 | 00.74 | 00.62 | 00.53 |
| $R_{MBSS}$ | 100.31 | 43.09 | 100.75 | 51.57 | 99.35 | 15.61 | 101.11 | 20.85 |
| $R_{MMSS}$ | 99.69 | 42.98 | 100.80 | 50.95 | 100.76 | 15.73 | 99.75 | 20.83 |
| $R_{MYEC}$ | 95.59 | 97.00 | 97.65 | 99.54 | 95.05 | 94.94 | 96.51 | 98.65 |
| $R_{MBMSS}$ | 101.58 | 42.74 | 18.36 | 51.03 | 100.41 | 15.70 | 05.26 | 20.90 |
| $R_{MBMSSEC}$ | N/A | 100.48 | 100.08 | 101.10 | N/A | 100.60 | 98.60 | 99.25 |
| $R_{BRS}$ | 100.17 | 100.09 | 99.96 | 100.04 | 100.12 | 100.04 | 99.99 | 100.11 |

Finally, we will compare BCRS to MBCSS, MCCSS as well as estimators MBCSSEC and MCCSSEC. If we examine Tables 5.3 and 5.4, we note that BCRS is equally efficient to MBCSS, while more efficient than MCCSS and estimator MCCSSEC. Note that there are greater efficiency gains when selecting BCRS over estimator MCCSSEC, as $b$ and/or $N$ increases. Next, consider Tables 5.6 and 5.7, where BCRS is more efficient than MBCSS and estimator MBCSSEC when $b$ is small and less efficient when $b$ is not small, i.e. there are greater efficiency losses when choosing BCRS over MBCSS (or estimator MBCSSEC) as $b$ and/or $N$ increases. We also note that BCRS is less efficient than estimator MCCSSEC, with greater efficiency losses as $b$ increases. Moreover, BCRS is more efficient than MCCSS, with greater efficiency gains as $b$ increases. Now, let us evaluate Tables 5.9 and 5.10. Here, BCRS is more efficient than MBCSS when $b$ is small and less efficient when $b$ is not small, i.e. there are greater efficiency losses when choosing BCRS over MBCSS as $b$ increases.

Also, BCRS is less efficient than MCCSS with greater efficiency losses when choosing BCRS over MBCSS as $b$ increases. Finally, we will examine Tables 5.11 and 5.12. We note that BCRS is more efficient than MCCSS, where we see greater gains in efficiency when selecting BCRS over MCCSS as $b$ and/or $N$ increases. Lastly for this scenario, we note that BCRS is more efficient than MBCSS and estimator MCCSSEC when $b$ is small and less efficient when $b$ is not small, where there are greater efficiency losses when selecting BCRS over MBCSS (or estimator MCCSSEC) as $b$ increases.

Table 5.3: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case A of BCRS ($N \neq nk$ and $b = 0.5, 1$).

| $b$ | 0.5 | 0.5 | 0.5 | 0.5 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 18 | 26 | 28 | 38 | 18 | 26 | 28 | 38 |
| $n$ | 8 | 12 | 8 | 12 | 8 | 12 | 8 | 12 |
| $k$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $r$ | 2 | 2 | 4 | 4 | 2 | 2 | 4 | 4 |
| $R_{CSS}$ | 29.53 | 21.87 | 13.73 | 17.39 | 09.31 | 06.57 | 03.80 | 04.99 |
| $R_{SRS}$ | 12.33 | 06.38 | 05.60 | 03.14 | 03.36 | 01.67 | 01.45 | 00.80 |
| $R_{BRS}$ | 99.93 | 99.95 | 100.11 | 100.11 | 100.01 | 99.92 | 99.71 | 99.92 |
| $R_{RLSS}$ | 02.48 | 02.10 | 02.79 | 03.01 | 00.96 | 00.73 | 01.03 | 00.95 |
| $R_{RMSS}$ | 02.52 | 02.14 | 02.91 | 03.20 | 00.98 | 00.75 | 01.10 | 01.04 |
| $R_{MBCSS}$ | 100.67 | 100.00 | 100.38 | 100.57 | 99.52 | 99.37 | 99.90 | 100.10 |
| $R_{MCCSS}$ | 53.14 | 41.46 | 26.58 | 18.54 | 21.47 | 15.34 | 08.20 | 05.40 |
| $R_{MCCSSEC}$ | 93.88 | 98.78 | 95.19 | 96.86 | 94.29 | 96.85 | 98.16 | 99.08 |

In this chapter, we suggested a modification to the BRS design, which addresses all the shortcomings of systematic sampling, while proving to be, suitable for populations exhibiting linear trend and simple to implement. The results from this chapter suggest that values of the population size, sample size and/or the degree of trend, needs to be considered, before determining if the proposed modified systematic sampling design is the most appropriate choice, when compared to alternative modified systematic sampling designs. Now that we have examined all modifications of LSS under the assumption of linear trend among the population units, we will next provide a comprehensive report on the thesis, which will include suggested designs and/or estimators under various scenarios,

Table 5.4: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case A of BCRS ($N \neq nk$ and $b = 2, 4$).

| $b$ | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 18 | 26 | 28 | 38 | 18 | 26 | 28 | 38 |
| $n$ | 8 | 12 | 8 | 12 | 8 | 12 | 8 | 12 |
| $k$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $r$ | 2 | 2 | 4 | 4 | 2 | 2 | 4 | 4 |
| $R_{CSS}$ | 02.55 | 01.72 | 00.98 | 01.31 | 00.65 | 00.44 | 00.25 | 00.33 |
| $R_{SRS}$ | 00.87 | 00.42 | 00.37 | 00.20 | 00.22 | 00.11 | 00.09 | 00.05 |
| $R_{BRS}$ | 99.99 | 99.92 | 100.15 | 100.21 | 99.97 | 100.02 | 99.94 | 100.17 |
| $R_{RLSS}$ | 00.31 | 00.22 | 00.33 | 00.27 | 00.09 | 00.06 | 00.09 | 00.07 |
| $R_{RMSS}$ | 00.33 | 00.23 | 00.35 | 00.30 | 00.09 | 00.06 | 00.10 | 00.08 |
| $R_{MBCSS}$ | 99.59 | 100.52 | 99.39 | 99.82 | 100.39 | 100.80 | 99.87 | 99.60 |
| $R_{MCCSS}$ | 06.47 | 04.32 | 02.12 | 01.40 | 01.69 | 01.10 | 00.56 | 00.36 |
| $R_{MCCSSEC}$ | 96.62 | 97.90 | 93.45 | 96.68 | 95.29 | 96.19 | 94.70 | 97.79 |

limitations of the current research as well as future recommendations/studies.

Table 5.5: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case B of BCRS ($N = nk$).

| $b$ | 0.5 | 0.5 | 1 | 1 | 2 | 2 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 18 | 30 | 18 | 30 | 18 | 30 | 18 | 30 |
| $n, n'$ | 9, 3 | 15, 5 | 9, 3 | 15, 5 | 9, 3 | 15, 5 | 9, 3 | 15, 5 |
| $k = l$ | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| $m$ | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| $R_{LSS}$ | 49.78 | 36.62 | 23.01 | 13.86 | 11.11 | 05.39 | 07.46 | 03.03 |
| $R_{SRS}$ | 13.15 | 05.13 | 04.32 | 01.49 | 01.83 | 00.54 | 01.19 | 00.30 |
| $R_{STR}$ | 95.01 | 92.92 | 84.77 | 78.03 | 70.42 | 55.63 | 60.57 | 40.74 |
| $R_{BMSS}$ | 106.52 | 101.48 | 119.44 | 114.50 | 173.88 | 143.67 | 288.64 | 242.83 |
| $R_{BMSSEC}$ | 108.21 | 102.17 | 125.67 | 118.32 | 211.90 | 162.72 | 550.19 | 370.12 |
| $R_{MLSS}$ | 29.44 | 19.41 | 11.09 | 06.33 | 04.88 | 02.35 | 03.21 | 01.31 |
| $R_{MBSS}$ | 83.32 | 88.98 | 58.94 | 68.54 | 37.14 | 43.95 | 27.51 | 30.02 |
| $R_{MMSS}$ | 82.40 | 88.22 | 59.32 | 68.33 | 37.07 | 43.83 | 27.63 | 30.03 |
| $R_{MYEC}$ | 101.71 | 102.49 | 121.97 | 113.73 | 201.50 | 163.95 | 519.35 | 366.10 |
| $R_{MBMSS}$ | 83.09 | 88.81 | 58.62 | 68.80 | 36.96 | 43.91 | 27.71 | 29.93 |
| $R_{MBMSSEC}$ | 106.57 | 104.16 | 126.63 | 116.64 | 208.79 | 166.59 | 542.82 | 369.50 |
| $R_{BRS}$ | 42.85 | 29.87 | 18.27 | 10.64 | 08.42 | 04.07 | 05.61 | 02.28 |

Table 5.6: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case B of BCRS ($N \neq nk$ and $b = 0.5, 1$).

| $b$ | 0.5 | 0.5 | 0.5 | 0.5 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 16 | 20 | 24 | 30 | 16 | 20 | 24 | 30 |
| $n$ | 7 | 9 | 7 | 9 | 7 | 9 | 7 | 9 |
| $k$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $r$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $R_{CSS}$ | 34.73 | 28.85 | 21.19 | 17.30 | 13.93 | 10.73 | 07.29 | 05.62 |
| $R_{SRS}$ | 16.14 | 10.94 | 07.89 | 05.15 | 05.57 | 03.50 | 02.49 | 01.53 |
| $R_{BRS}$ | 45.83 | 39.77 | 31.19 | 26.42 | 20.61 | 16.31 | 11.72 | 09.30 |
| $R_{RLSS}$ | 64.92 | 58.37 | 48.30 | 43.15 | 38.12 | 31.15 | 21.54 | 18.02 |
| $R_{RLSSEC}$ | 94.54 | 96.01 | 93.23 | 97.21 | 115.25 | 114.45 | 109.29 | 111.63 |
| $R_{RMSS}$ | 101.21 | 100.80 | 95.79 | 94.83 | 117.39 | 113.74 | 91.16 | 90.61 |
| $R_{RMSSEC}$ | 102.10 | 101.69 | 102.32 | 100.50 | 122.36 | 118.32 | 120.30 | 114.41 |
| $R_{MBCSS}$ | 89.52 | 90.65 | 94.67 | 95.56 | 104.72 | 102.39 | 109.19 | 107.62 |
| $R_{MBCSSEC}$ | 90.53 | 92.12 | 95.74 | 96.42 | 110.28 | 108.19 | 114.34 | 109.93 |
| $R_{MCCSS}$ | 60.26 | 53.17 | 65.42 | 58.81 | 32.50 | 24.61 | 36.14 | 28.56 |
| $R_{MCCSSEC}$ | 103.08 | 100.78 | 104.72 | 105.38 | 126.14 | 121.59 | 124.49 | 118.20 |

Table 5.7: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case B of BCRS ($N \neq nk$ and $b = 2, 4$).

| $b$ | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 16 | 20 | 24 | 30 | 16 | 20 | 24 | 30 |
| $n$ | 7 | 9 | 7 | 9 | 7 | 9 | 7 | 9 |
| $k$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $r$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $R_{CSS}$ | 06.53 | 04.58 | 03.08 | 02.17 | 04.47 | 02.91 | 01.94 | 01.28 |
| $R_{SRS}$ | 02.47 | 01.43 | 01.00 | 00.58 | 01.67 | 00.90 | 00.63 | 00.34 |
| $R_{BRS}$ | 10.04 | 07.24 | 05.08 | 03.68 | 06.96 | 04.65 | 03.24 | 02.19 |
| $R_{RLSS}$ | 21.43 | 15.50 | 10.07 | 07.57 | 15.55 | 10.30 | 06.55 | 04.58 |
| $R_{RLSSEC}$ | 196.77 | 181.57 | 175.43 | 166.47 | 532.19 | 457.23 | 439.03 | 390.52 |
| $R_{RMSS}$ | 177.91 | 161.19 | 84.25 | 82.03 | 323.11 | 294.51 | 77.23 | 74.47 |
| $R_{RMSSEC}$ | 208.95 | 184.60 | 193.39 | 171.35 | 562.71 | 468.22 | 483.16 | 398.99 |
| $R_{MBCSS}$ | 156.14 | 146.13 | 153.96 | 145.43 | 274.77 | 253.26 | 262.06 | 243.47 |
| $R_{MBCSSEC}$ | 190.16 | 172.38 | 182.81 | 164.32 | 516.23 | 436.18 | 460.14 | 383.67 |
| $R_{MCCSS}$ | 16.90 | 11.63 | 18.45 | 13.09 | 11.92 | 07.60 | 12.45 | 07.97 |
| $R_{MCCSSEC}$ | 217.99 | 193.51 | 198.79 | 176.06 | 579.41 | 487.04 | 498.99 | 411.68 |

Table 5.8: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case C of BCRS ($N = nk$).

| $b$ | 0.5 | 0.5 | 1 | 1 | 2 | 2 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 27 | 45 | 27 | 45 | 27 | 45 | 27 | 45 |
| $n, n'$ | 9, 3 | 15, 5 | 9, 3 | 15, 5 | 9, 3 | 15, 5 | 9, 3 | 15, 5 |
| $k = l$ | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| $m$ | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| $R_{LSS}$ | 34.85 | 21.94 | 14.93 | 08.02 | 08.14 | 03.53 | 06.20 | 02.40 |
| $R_{SRS}$ | 06.69 | 02.40 | 02.34 | 00.75 | 01.19 | 00.32 | 00.89 | 00.21 |
| $R_{STR}$ | 90.31 | 85.27 | 75.01 | 65.07 | 60.22 | 43.24 | 52.92 | 34.06 |
| $R_{BMSS}$ | 108.71 | 106.28 | 135.12 | 122.99 | 206.31 | 174.03 | 327.29 | 279.53 |
| $R_{BMSSEC}$ | 111.40 | 109.23 | 149.72 | 130.11 | 299.54 | 223.21 | 904.20 | 576.51 |
| $R_{MLSS}$ | 16.94 | 10.14 | 16.37 | 03.38 | 03.30 | 01.44 | 02.49 | 00.96 |
| $R_{MBSS}$ | 69.36 | 78.71 | 43.03 | 52.51 | 27.25 | 31.25 | 21.95 | 23.22 |
| $R_{MMSS}$ | 69.79 | 78.60 | 42.79 | 52.63 | 27.25 | 31.14 | 21.95 | 23.08 |
| $R_{MYEC}$ | 107.69 | 106.31 | 144.21 | 127.13 | 283.81 | 216.99 | 868.42 | 559.17 |
| $R_{MBMSS}$ | 69.38 | 79.70 | 42.73 | 52.63 | 27.27 | 31.06 | 21.95 | 23.12 |
| $R_{MBMSSEC}$ | 111.85 | 109.16 | 149.50 | 132.58 | 296.72 | 218.62 | 903.46 | 562.88 |
| $R_{BRS}$ | 50.32 | 35.63 | 25.31 | 14.63 | 14.51 | 06.67 | 11.26 | 04.53 |

Table 5.9: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case C of BCRS ($N \neq nk$ and $b = 0.5, 1$).

| $b$ | 0.5 | 0.5 | 0.5 | 0.5 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 17 | 21 | 23 | 29 | 17 | 21 | 23 | 29 |
| $n$ | 7 | 9 | 7 | 9 | 7 | 9 | 7 | 9 |
| $k$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $r$ | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 2 |
| $R_{CSS}$ | 26.77 | 21.13 | 29.73 | 24.41 | 11.85 | 08.56 | 12.98 | 09.57 |
| $R_{SRS}$ | 16.03 | 10.73 | 09.27 | 05.87 | 06.52 | 04.01 | 03.45 | 02.02 |
| $R_{BRS}$ | 74.56 | 68.67 | 52.30 | 45.68 | 51.41 | 43.38 | 27.67 | 21.74 |
| $R_{RLSS}$ | 65.73 | 60.16 | 53.99 | 45.49 | 43.93 | 37.13 | 29.90 | 21.72 |
| $R_{RMSS}$ | 108.79 | 107.24 | 112.18 | 109.20 | 161.07 | 149.65 | 155.09 | 146.37 |
| $R_{MBCSS}$ | 82.96 | 78.97 | 90.57 | 88.30 | 121.73 | 109.03 | 127.66 | 117.33 |
| $R_{MCCSS}$ | 121.56 | 114.60 | 115.33 | 111.54 | 175.18 | 158.96 | 163.55 | 148.93 |

Table 5.10: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case C of BCRS ($N \neq nk$ and $b = 2, 4$).

| $b$ | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 17 | 21 | 23 | 29 | 17 | 21 | 23 | 29 |
| $n$ | 7 | 9 | 7 | 9 | 7 | 9 | 7 | 9 |
| $k$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $r$ | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 2 |
| $R_{CSS}$ | 07.08 | 04.70 | 07.38 | 04.99 | 05.79 | 03.67 | 05.87 | 03.75 |
| $R_{SRS}$ | 03.81 | 02.15 | 01.87 | 01.01 | 03.09 | 01.67 | 01.47 | 00.75 |
| $R_{BRS}$ | 37.61 | 28.77 | 16.97 | 12.10 | 32.68 | 23.81 | 13.80 | 09.27 |
| $R_{RLSS}$ | 31.24 | 24.00 | 18.70 | 12.15 | 26.83 | 19.64 | 15.30 | 09.31 |
| $R_{RMSS}$ | 361.75 | 313.19 | 335.08 | 286.74 | 1191.79 | 954.76 | 1071.17 | 859.97 |
| $R_{MBCSS}$ | 276.35 | 232.60 | 273.70 | 232.65 | 888.61 | 712.96 | 851.24 | 687.06 |
| $R_{MCCSS}$ | 398.23 | 331.05 | 352.49 | 297.27 | 1248.78 | 1019.49 | 1070.99 | 869.78 |

Table 5.11: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case D of BCRS ($N \neq nk$ and $b = 0.5, 1$).

| $b$ | 0.5 | 0.5 | 0.5 | 0.5 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 15 | 19 | 21 | 27 | 15 | 19 | 21 | 27 |
| $n$ | 6 | 8 | 6 | 8 | 6 | 8 | 6 | 8 |
| $k$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $r$ | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| $R_{CSS}$ | 26.68 | 21.38 | 22.91 | 18.41 | 09.29 | 06.91 | 07.69 | 05.80 |
| $R_{SRS}$ | 17.45 | 11.74 | 09.73 | 06.12 | 05.59 | 03.49 | 02.91 | 01.74 |
| $R_{BRS}$ | 36.89 | 32.29 | 23.78 | 19.93 | 14.06 | 11.47 | 07.98 | 06.35 |
| $R_{RLSS}$ | 59.32 | 57.57 | 48.36 | 44.12 | 30.14 | 28.04 | 20.99 | 18.01 |
| $R_{RLSSEC}$ | 73.65 | 89.75 | 81.21 | 92.87 | 82.95 | 98.57 | 90.70 | 102.03 |
| $R_{RMSS}$ | 97.37 | 97.15 | 92.04 | 92.13 | 106.28 | 104.08 | 81.38 | 82.74 |
| $R_{RMSSEC}$ | 97.99 | 97.93 | 99.90 | 97.84 | 109.93 | 107.51 | 109.42 | 106.89 |
| $R_{MBCSS}$ | 79.04 | 75.41 | 86.99 | 85.12 | 89.51 | 83.86 | 96.29 | 92.70 |
| $R_{MCCSS}$ | 29.39 | 23.13 | 18.19 | 14.06 | 10.73 | 07.71 | 05.81 | 04.27 |
| $R_{MCCSSEC}$ | 88.72 | 92.18 | 88.04 | 94.44 | 97.01 | 100.05 | 97.87 | 102.00 |

Table 5.12: Simulated Relative Expected MSEs for Populations Exhibiting Linear Trend under Case D of BCRS ($N \neq nk$ and $b = 2, 4$).

| $b$ | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 15 | 19 | 21 | 27 | 15 | 19 | 21 | 27 |
| $n$ | 6 | 8 | 6 | 8 | 6 | 8 | 6 | 8 |
| $k$ | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 |
| $r$ | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| $R_{CSS}$ | 03.67 | 02.50 | 02.83 | 01.98 | 02.11 | 01.33 | 01.55 | 00.99 |
| $R_{SRS}$ | 02.16 | 01.24 | 01.04 | 00.58 | 01.23 | 00.65 | 00.57 | 00.29 |
| $R_{BRS}$ | 05.74 | 04.29 | 02.96 | 02.18 | 03.33 | 02.30 | 01.62 | 01.09 |
| $R_{RLSS}$ | 14.27 | 12.01 | 08.55 | 06.78 | 08.63 | 06.71 | 04.82 | 03.48 |
| $R_{RLSSEC}$ | 123.10 | 135.34 | 125.95 | 133.08 | 276.19 | 287.50 | 273.09 | 262.09 |
| $R_{RMSS}$ | 140.18 | 128.85 | 63.07 | 64.04 | 220.13 | 196.55 | 48.88 | 48.35 |
| $R_{RMSSEC}$ | 162.12 | 147.34 | 153.67 | 140.14 | 363.40 | 307.07 | 332.58 | 275.41 |
| $R_{MBCSS}$ | 130.79 | 114.84 | 134.75 | 121.03 | 298.42 | 241.80 | 289.61 | 241.31 |
| $R_{MCCSS}$ | 04.23 | 02.81 | 02.11 | 01.43 | 02.47 | 01.49 | 01.15 | 00.71 |
| $R_{MCCSSEC}$ | 141.34 | 139.19 | 138.08 | 131.98 | 318.73 | 286.99 | 292.42 | 266.86 |

# Chapter 6

# Summary

Conventional systematic sampling, also referred to as LSS, is commonly applied in practical survey sampling, owing to its simplicity and operational convenience when implemented. However, there exists two key disadvantages when conducting LSS, which are given as follows:

(i) It is impossible to obtain an unbiased estimate of the sampling variance when conducting LSS with a single random start.

(ii) If the population size is not a multiple of the sample size, then conducting LSS will either result in sample sizes that vary, or fixed sample sizes that are greater than the required sample size. The former situation leads to biased estimates of population parameters, while the latter scenario is undesirable, as sample sizes are usually fixed beforehand often owing to budget constraints.

In this study, we aim to solve both disadvantages independently and in tandem while assuming the linear trend population structure.

## 6.1 Conclusions

The results from this thesis are summarized as follows:

(i) In the presence of linear trend, LSS is more efficient than CSS, but less efficient than STR. Accordingly, many authors have suggested modified LSS designs and/or estimators, as seen in Chapter 2, which are briefly summarized as follows:

1: Yates (1948) proposed an end corrections estimator (i.e the YEC estimator) which corrects the usual LSS estimator by applying appropriate weights to the first and last sampling units, thus removing the linear trend component in the associated expected MSE of the sample mean. The uneven weighting of sampling units result in a larger error variance component in the expected MSE of the corresponding sample mean. Regardless, we can expect this estimator to be more efficient than that which is related to LSS, in the presence of a rough linear trend.

2: Madow (1953) suggested that the centrally located systematic sample be selected, thus requiring no randomization. As a result, certain population units have a zero probability of being selected for the sample, which results in the sample mean being biased. Nevertheless, the associated sample mean is equivalent to the population mean in the presence of a perfect linear trend in the population when $k$ is odd. Moreover, the expected MSE of the sample mean is minimized when $k$ is odd under the realistic linear trend model, given in (2.1).

3: Sethi (1965) and Murthy (1967) proposed a sampling design which reverses the order of every alternative set of $k$ population units, before applying LSS. As such, the corresponding sample mean is design-unbiased. Under model (2.1), the expected MSE of the sample mean is minimized when $n$ is even.

4: Singh et al. (1968) suggested an arrangement which reverses the order of a subset of population units which occurs at the end of the population. If $n$ is even, then the last $N/2$ population units are reversed, while the case of $n$ is odd results in the last $(N-k)/2$ population units being reversed. LSS is then applied to this modified arrangement, such that the sample mean is design-unbiased. The expected MSE of the sample mean is equivalent to that of BSS for model (2.1), where optimal results are obtained when $n$ is even.

5: Subramani (2000) proposed a sampling scheme, DSS, which arranges the population in a matrix, before selecting units in a diagonal fashion. This design is only applicable when $n \leq k$. The sample mean is a design-unbiased estimator of the population mean and the expected MSE of this sample mean, under model (2.1), is minimized when $n = k$. Consequently, an end corrections estimator is suggested to remove the linear trend component in the expected MSE of the sample mean. As with the YEC estimator, this adjusted estimator is equivalent to the population mean in the presence of a perfect linear trend, while being slightly biased under model (2.1). Subramani (2009, 2010) later introduced a generalization of Subramani's (2000) DSS which is applicable for all cases of $n$ and reduces to DSS under a certain condition on the sample size. Under model (2.1), the expected MSE of the associated sample mean is minimized if $n$ is a multiple of $k$.

(ii) Optimal sampling results in the presence of linear trend are unobtainable for the case of $n$ is odd and $k$ is even. A modified design is then suggested in Section 2.4, which is a mixture of BSS and MSS and is termed as BMSS. There are five cases of the design and one of the cases (i.e. $n/2$ is an even integer) results in optimum sampling results under model (2.1). For the other cases, an end corrections estimator was constructed, i.e. the BMSSEC estimator. Under model (2.1), BMSS is equally efficient to BSS and MSS for four of the five cases, while less efficient for one of the cases (i.e. $n/2$ is an odd integer). The results from this thesis suggest that BMSS, BSS or MSS is to be preferred for the case of $n/2$ being an even integer, while BSS or MSS is to be preferred when $n/2$ is an odd integer. For all other cases (i.e. $n$ is odd), CESS is to be preferred if $k$ is odd. The final scenario of $n$ is odd and $k$ is even, results in the BMSSEC estimator being preferred over the YEC estimator if $n$ and $k$ are small and preference is given to minimum MSE. Otherwise, one may opt to use the YEC estimator, owing to its simplistic preference over the BMSSEC estimator.

(iii) To solve the shortcoming of being unable to estimate the sample variance when conducting LSS with a single random start, various modified systematic sampling designs have been proposed, as seen in Section 3.1. Gautschi (1957) proposed that the usual LSS design be applied with multiple random starts. Thus, the second-order inclusion probabilities for every pair of population units are non-zero, ensuring that an unbiased estimate of the sampling variance is obtainable. The multiple-start approach assumes that the sample size is a non-prime integer. Sampath & Ammani (2010) applied this multiple-start approach to the YEC estimator as well as BSS and MSS, while Subramani & Singh (2014) applied the multiple-start approach to DSS. If we denote the required sample size as $n = n'm$ and let $N = n'mk$, where $m$ is the number of random starts, such that $m$ samples of size $n'$ are selected from the $k$ possible samples using SRS, then for model (2.1) we obtain optimal sampling results for MBSS and MMSS when $n'$ is even as well as for MDSS when $n' = mk$.

Another important modified design worth mentioning is BRS, proposed by Singh & Garg (1979), which offers linear trend free sampling if $N$ and $n$ are both even.

(iv) A modified design, MBMSS, which adopts the multiple-start approach on BMSS, was then proposed in Section 3.2. The results are similar to that in Section 2.4, where BMSS was suggested. There are five cases of the design and one of the cases (i.e. $n'/2$ is an even integer) results in optimum sampling results under model (2.1). For the other cases, an associated end corrections estimator was derived, i.e. the MBMSSEC estimator. Under model (2.1), MBMSS is equally efficient to MBSS and MMSS for four of the five cases, while less efficient for one of the cases (i.e. $n'/2$ is an odd integer). The results from this thesis suggest that MBMSS, MBSS or MMSS is to be preferred for the case of $n'/2$ being an even integer, while MBSS or MMSS is to be preferred when $n'/2$ is an odd integer. All other cases (i.e. $n'$ is odd), result in the MBMSSEC estimator being preferred over the MYEC estimator if $n'$, $m$ and $k$ are small and preference is given to minimum MSE. Otherwise, one may opt to use the MYEC estimator over the MBMSSEC estimator, when simplicity is to be preferred.

(v) To tackle to problem of the population size not being a multiple of the sample size, many modified designs have been presented in literature and summarized in Section 4.1. Lahiri (1951) considered CSS, whereby the population is arranged in a circular fashion and sampling units are selected systematically with respect to a sampling interval. BRS, which solves the problem of being unable to obtain an unbiased estimate of the sampling variance, also tackles the above-mentioned shortcoming. Uthayakumaran (1998) then adopted the CSS approach on BSS and CESS, which was later modified by Leu & Kao (2006) and known as MBCSS and MCCSS, respectively. Under model (2.1), optimum sampling results are obtained for MBCSS if $N$ is odd or $n$ is even, while MCCSS offers optimum sampling results when $N$ and $n$ are both odd. In addition, Sampath & Varalakshmi (2009) combined DSS and CSS, known as DCSS. DCSS is not a linear trend free sampling design and thus an associated end corrections estimator was constructed. Chang & Huang (2000) introduced RLSS, which divides the population into two strata before independently applying LSS within each strata, such the the selected sampling units from each strata collectively represent the sample. RLSS is not a linear trend free sampling design and hence a corresponding end corrections estimator, termed as the RLSSEC estimator, was derived by Chang & Huang (2000). Finally, we note that Mostafa & Ahmad (2016) adopted the multiple-start approach on RLSS, so as to solve both LSS shortcomings in tandem. This design is not a linear trend free sampling design. Also, the design requires that $(n - r)$ and $r$ are non-prime integers, which is often not the case.

(vi) A modified design, known as RMSS, which is a mixture of RLSS, MSS and CESS, was then proposed in Section 4.2, i.e. the population is divided into two strata, where either MSS or CESS is applied within each strata, such that the selected sampling units from each strata collectively represent the sample. There are seven cases of the design and three of these seven cases result in linear trend free sampling. For the other four cases, an ends corrections estimator, termed as the RLSSEC estimator, was derived. The results in this thesis indicate that RMSS is more efficient than SRS, CSS and RLSS, with greater efficiency gains as $N$ and/or $b$ increases. Also, we showed that the RMSSEC estimator is to be preferred over the RLSSEC estimator if we are not tackling a large-scale sampling scenario. For large-scale sampling, both estimators are approximately equally efficient, thus one may opt to use the MYEC estimator over the MBMSSEC estimator, owing to simplicity.

(vii) As noted previously, BRS is the only modified sampling design that tackles both LSS disadvantages in tandem without any loss of simplicity, while providing favourable results in the presence of linear trend. Therefore, a modification to the BRS design, termed as BCRS, was proposed in Chapter 5. BCRS is a mixture of BRS and SRS, which is applied on a centered subset of the population. Thus, we are able to maintain simplicity. Under model (2.1), BCRS is equally efficient to BRS for one of the four cases (i.e. if $N$ and $n > 6$ are both even, where $N > n + 2$), while more efficient than BRS for the other three cases, with greater efficiency gains as $N$, $n$ and/or $b$ increases. Various efficiency comparisons between BCRS and other modified sampling designs have been considered in Chapter 5. Earlier in this thesis, we noted that the most desirable sampling design(s) and/or estimator(s) are those that: (1) exhibit minimum MSE of the associated sample mean, (2) are simple to apply in practical situations and (3) offer the possibility of obtaining an unbiased estimate of the corresponding sampling variance. Using this notion, recommendations of sampling designs for various scenarios under model (2.1), which use the results from this thesis, have been presented in Tables 6.1 and 6.2. In Table 6.1, the first column represents the condition if the sample size is a prime integer or not. The second column represents the various cases of BCRS, while the third column represents two cases of MBMSS, i.e. if $n'/2$ is an even integer or not. Columns four to six indicate the three preferences when selecting a design, i.e. preference 1 is given more preference over preference 2, which in turn is given more preference over preference 3. In Table 6.2, the first column represents the various cases of BCRS, while columns three to five represent the various preferences, as in Table 6.1. From Tables 6.1 and 6.2, we note that the proposed designs and/or estimators are preferred for fourteen of the twenty possible scenarios and are unmatched for ten of the twenty possible cases.

Table 6.1: Recommended designs and/or estimators for populations exhibiting linear trend ($N = nk$).

| Cond 1 | Cond 2 | Cond 3 | Pref 1 | Pref 2 | Pref 3 | Recommended |
|---|---|---|---|---|---|---|
| $n$ non-prime | A | $n'/2$ even | N/A | N/A | N/A | MBSS, MMSS, MBMSS, BRS or BCRS |
| $n$ non-prime | A | $n'/2$ odd | N/A | N/A | N/A | MBSS, MMSS, BRS or BCRS |
| $n$ non-prime | B/C | N/A | Unbiased Sampling Variance | Simplicity | Minimum MSE | BCRS |
| $n$ non-prime | B/C | N/A | Simplicity | Minimum MSE | N/A | MYEC |
| $n$ non-prime | B/C | N/A | Minimum MSE | Simplicity | N/A | MBMSSEC |
| $n$ is prime | A | N/A | N/A | N/A | N/A | BRS or BCRS |
| $n$ is prime | B/C | N/A | N/A | N/A | N/A | BCRS |

Table 6.2: Recommended designs and/or estimators for populations exhibiting linear trend ($N \neq nk$).

| Cond 1 | Cond 2 | Pref 1 | Pref 2 | Pref 3 | Recommended |
|---|---|---|---|---|---|
| A | N/A | N/A | N/A | N/A | BRS or BCRS |
| B | $b$ small | Unbiased Sampling Variance | Simplicity | Minimum MSE | BCRS |
| B | $b$ small | Minimum MSE | Simplicity | N/A | MCCSSEC |
| B | $b$ small | Simplicity | Minimum MSE | N/A | RMSSEC |
| B | $b \geq 1$ | Unbiased Sampling Variance | Simplicity | Minimum MSE | BCRS |
| B | $b \geq 1$ | Simplicity | Minimum MSE | N/A | RLSSEC |
| B | $b \geq 1$ | Minimum MSE | Simplicity | N/A | MCCSSEC |
| C | N/A | Unbiased Sampling Variance | Simplicity | Minimum MSE | BCRS |
| C | N/A | Minimum MSE | Simplicity | N/A | MCCSS |
| D | $b$ small | N/A | N/A | N/A | BCRS |
| D | $b \geq 1$ | Unbiased Sampling Variance | Simplicity | Minimum MSE | BCRS |
| D | $b \geq 1$ | Simplicity | Minimum MSE | N/A | RLSSEC |
| D | $b \geq 1$ | Minimum MSE | Simplicity | N/A | RMSSEC |

## 6.2 Limitations

To conclude this thesis, we note that all the designs and estimators presented in this study are under the assumption of a linear trend among the population units. More often than not, one does not know the population structure prior to sampling. Applying the designs and/or estimators when the population does not exhibit linear trend may provide poor results, as the designs and estimators presented in this thesis are most suitable for the linear trend population structure. Prior to sampling, the onus thus lies on the statistician to acquire as much information as possible regarding the population, so as to estimate if the population exhibits linear trend. Alternatively, the sampler may arrange the population prior to sampling. Consequently, one may choose to arrange the population in increasing/decreasing order in accordance with an auxiliary variable (a variable that is correlated with the study variable, such that it is easier to acquire the values of this new variable, when compared to those of the study variable). As a result, we obtain an approximate trend in the population, where the higher the degree of correlation between the two variables, the greater the degree of linear trend in the rearranged population. Under these circumstances, the theory and results presented in this thesis may then apply.

## 6.3 Future Studies

To expand on the work carried out within this thesis, we will compare the proposed designs for various other population structures. Also, we only considered one-dimensional sampling. Systematic sampling is commonly used in spatial sampling scenarios. We will then look to extend the theory presented in this thesis to two-dimensional situations. Additionally, adaptive sampling was proposed by Thompson (1990) and Thompson & Seber (1996) and is used for spatial sampling. Sampling units could be selected using a systematic approach when applying adaptive sampling. Thus, as a topic for future studies, we will consider adopting the modified systematic sampling designs proposed in this thesis in conjunction with adaptive sampling. The use of the suggested designs will also be examined when sampling with unequal probabilities, e.g. if there exists a variable of size then we may adopt the proposed designs to sample with probabilities proportionate to size (pps).

## Appendix

***Proof of* Theorem 2**: An estimate of $\overline{Y}$ with random start $i$, for $i \in \{1, ..., k\}$, can be written as

$$\overline{y}_{BMSSEC} = \frac{1}{n}\left[\psi_1 y_{x_1} + \sum_{j=2}^{(n-1)} y_{x_j} + \psi_2 y_{x_n}\right], \tag{1}$$

where $\psi_1$ and $\psi_2$ are the weights applied to the first and the last sampling units respectively and $x_1, ..., x_n$ are the sampling unit indices, which are arranged in ascending order. By substituting (2.12) into (1) and then equating this result to $\overline{Y}_B$, we obtain

$$\overline{y}_{BMSSEC} = \frac{1}{n}\left[\psi_1 (a + bx_1) + \sum_{j=2}^{(n-1)} (a + bx_j) + \psi_2 (a + bx_n)\right] = a + \frac{b(N+1)}{2}. \tag{2}$$

By equating the coefficients of $a$ in (2), it follows that

$$\psi_1 = 2 - \psi_2. \tag{3}$$

Similarly, by equating the coefficients of $b$ in (2), we obtain

$$\frac{1}{n}\left[\psi_1 x_1 + \sum_{j=2}^{(n-1)} x_j + \psi_2 x_n\right] = \frac{N+1}{2}. \tag{4}$$

Substituting (3) into (4) results in

$$2\left[2x_1 - \psi_2 x_1 + \sum_{j=2}^{(n-1)} x_j + \psi_2 x_n\right] = n(N+1),$$

which simplifies to

$$\psi_2 = \frac{K - 2x_1}{x_n - x_1}, \tag{5}$$

where $K = n(N+1)/2 - \sum_{j=2}^{n-1} x_j$. The weight applied to the first sampling unit is thus obtained by substituting (5) into (3), i.e.

$$\psi_1 = \frac{2x_n - K}{x_n - x_1}.$$

Substituting $\psi_1$ and $\psi_2$ into (1) results in

$$\begin{aligned}
\overline{y}_{BMSSEC} &= \frac{1}{n}\left[\frac{(2x_n - K)}{(x_n - x_1)} y_{x_1} + \sum_{j=2}^{n-1} y_{x_j} + \frac{(K - 2x_1)}{(x_n - x_1)} y_{x_n}\right] \\
&= \overline{y}_{BMSS} + \frac{1}{n}\left[\frac{(2x_n - K)}{(x_n - x_1)} y_{x_1} + \frac{(K - 2x_1)}{(x_n - x_1)} y_{x_n} - y_{x_1} - y_{x_n}\right] \\
&= \overline{y}_{BMSS} + \frac{[(x_n + x_1) - K]}{n(x_n - x_1)} (y_{x_1} - y_{x_n}). \qquad (6)
\end{aligned}$$

Now, if we consider Case (B), then $x_1 = i$, $x_n = N + i - k$ and

$$
\begin{aligned}
K &= \frac{n(N+1)}{2} - \sum_{j=2}^{n-1} x_j \\
&= \frac{n(N+1)}{2} - \left[ \sum_{j=1}^{(n-2)/4} (2i + N - k) + \sum_{j=0}^{(n-6)/4} (k - 2i + 2 + N) \right] \\
&= \frac{n(N+1)}{2} - \sum_{j=1}^{(n-2)/4} (2i + N - k + k - 2i + 2 + N) = \frac{(N+1)[n-(n-2)]}{2} = N+1
\end{aligned}
$$

(refer to the sampling unit indices of Case (B) in Section 2.4.1). On substituting these values into (6), we obtain

$$
\overline{y}_{BMSSEC} = \overline{y}_{BMSS} + \frac{P}{n(N-k)} \left( y_i - y_{N+i-k} \right). \tag{7}
$$

We then conclude the proof by finding the values of $x_1$, $x_n$ and $K$ for the other cases, as shown above, and then substituting these values into (6).

***Proof of* Theorem 3**: The expected MSE of $\overline{y}_{BMSSEC}$ can be written as

$$
\begin{aligned}
M_{BMSSEC} &\triangleq \mathcal{E}\left[ \mathrm{E}\left( \{ \overline{y}_{BMSSEC} - \overline{Y} \}^2 \right) \right] \\
&= \mathrm{E}\left\{ \mathcal{E}\left[ \left( \overline{y}_{BMSSEC} - \overline{Y} \right)^2 \right] \right\} = \frac{1}{k} \sum_{i=1}^{k} \mathcal{E}\left[ \overline{y}_{BMSSEC} - \overline{Y} \right]^2.
\end{aligned} \tag{8}
$$

If we consider Case (B) for (2.1), then

$$
\overline{y}_{BMSS} - \overline{Y} = a + \frac{b}{2}\left[ N + 1 + \frac{2P}{n} \right] + \overline{e}_i - \left[ a + \frac{b(N+1)}{2} + \overline{\overline{e}} \right] = \frac{bP}{n} + \overline{e}_i - \overline{\overline{e}}. \tag{9}
$$

Moreover,

$$
\begin{aligned}
y_i - y_{N+i-k} &= a + bi + e_i - [a + b(N + i - k) + e_{N+i-k}] \\
&= -b[N-k] + e_i - e_{N+i-k}.
\end{aligned} \tag{10}
$$

Using (7), (9) and (10), we obtain

$$
\begin{aligned}
\mathcal{E}\left[ (\overline{y}_{BMSSEC} - \overline{Y})^2 \right] &= \mathcal{E}\left[ \left( \frac{bP}{n} + \overline{e}_i - \overline{\overline{e}} + \frac{P[-b(N-k) + e_i - e_{N+i-k}]}{n(N-k)} \right)^2 \right] \\
&= \mathcal{E}\left[ \left( \overline{e}_i - \overline{\overline{e}} + \frac{P[e_i - e_{N+i-k}]}{n(N-k)} \right)^2 \right].
\end{aligned} \tag{11}
$$

Applying Cochran's (1946) super-population model assumptions, results in

$$\mathcal{E}\left(\overline{e}_i^2\right) = \frac{1}{n^2}\left[\sum \mathcal{E}\left(e_i^2\right) + \sum \sum_{j \neq i} \mathcal{E}\left(e_i e_j\right)\right] = \frac{\sigma^2}{n}, \tag{12}$$

$$\mathcal{E}\left(\overline{\overline{e}}^2\right) = \frac{1}{N^2}\left[\sum_{j=1}^{N} \mathcal{E}\left(e_j^2\right) + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \mathcal{E}\left(e_i e_j\right)\right] = \frac{\sigma^2}{N}, \tag{13}$$

$$\mathcal{E}\left(\overline{e}_i \overline{\overline{e}}\right) = \frac{1}{nN} \sum \sum_{j=1}^{N} \mathcal{E}\left(e_i e_j\right) = \frac{n\sigma^2}{nN} = \frac{\sigma^2}{N}, \tag{14}$$

$$\mathcal{E}\left[\overline{e}_i\left(e_i - e_{i+(n-1)k}\right)\right] = \mathcal{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} e_{\{i+(j-1)k\}}\right)\left(e_i - e_{i+(n-1)k}\right)\right] = 0, \tag{15}$$

$$\mathcal{E}\left[\overline{\overline{e}}\left(e_i - e_{i+(n-1)k}\right)\right] = \mathcal{E}\left[\left(\frac{1}{N}\sum_{j=1}^{N} e_j\right)\left(e_i - e_{i+(n-1)k}\right)\right] = 0 \tag{16}$$

and

$$\mathcal{E}\left[\left(e_i - e_{i+(n-1)k}\right)^2\right] = \mathcal{E}\left[e_i^2 - 2e_i e_{i+(n-1)k} + e_{i+(n-1)k}^2\right] = 2\sigma^2. \tag{17}$$

Expanding (11) and then substituting (12) through to (17) into this expression, results in

$$\mathcal{E}\left[(\overline{y}_{BMSSEC} - \overline{Y})^2\right] = \sigma_e^2 + \frac{2\sigma^2 P^2}{n^2(N-k)^2}. \tag{18}$$

Finally, by substituting (18) into (8), we obtain

$$M_{BMSSEC} = \sigma_e^2 + \frac{2\sigma^2}{n^2 k(N-k)^2}\sum_{i=1}^{k} P^2 = \sigma_e^2 + \frac{2\sigma^2(k^2-1)}{3n^2(N-k)^2}.$$

Similarly, we can obtain $\mathcal{E}\left[(\overline{y}_{BMSSEC} - \overline{Y})^2\right]$ for Cases (C) to (E) and then substitute these expressions into (8).

***Proof of*** **Theorem 4**: An estimate of $\overline{Y}$ with random starts $ih$, for $h = 1, ..., m$ and $ih \in \{1, ..., k\}$, can be written as

$$\overline{y}_{MYEC} = \frac{1}{nm}\left(\lambda_1 \sum_{h=1}^{m} y_{ih} + \sum_{h=1}^{m} \sum_{j=1}^{(n-2)} y_{ih+jk} + \lambda_2 \sum_{h=1}^{m} y_{ih+(n-1)k}\right), \tag{19}$$

where $\lambda_1$ and $\lambda_2$ are the weights applied to the first and last sampling units of each sample, respectively. If we assume model (2.12), then $\overline{Y} = a + b(N+1)/2$. Thus, by substituting this model into (19) and then equating this expression to $\overline{Y}$, we obtain

$$\begin{aligned}\overline{y}_{MYEC} &= \frac{1}{nm}\left\{\lambda_1\sum_{h=1}^{m}(a+bih) + \sum_{h=1}^{m}\sum_{j=1}^{(n-2)}[a+b(ih+jk)] + \lambda_2\sum_{h=1}^{m}[a+b(ih+nk-k)]\right\} \\ &= nma + \frac{nmb(N+1)}{2}. \end{aligned} \tag{20}$$

On equating the coefficients of $a$ in (20), it follows that

$$\lambda_1 = 2 - \lambda_2. \tag{21}$$

Similarly, by equating the coefficients of $b$ in (20) we obtain

$$(\lambda_1 + n - 2 + \lambda_2)\sum_{h=1}^{m} ih + \frac{mk(n-1)(n-2)}{2} + \lambda_2 m(n-1)k = \frac{mn(N+1)}{2}.$$

Substituting (21) into this expression results in

$$2n\sum_{h=1}^{m} ih + mk(n-1)(n-2) + 2\lambda_2 m(n-1)k = mn(N+1),$$

which simplifies to

$$\lambda_2 = 1 - \frac{nV}{2m(n-1)k}. \tag{22}$$

Hence, by substituting (4) into (3) we obtain

$$\lambda_1 = 1 + \frac{nV}{2m(n-1)k}.$$

We then conclude the proof by substituting these weights into (19), i.e.

$$\begin{aligned}\overline{y}_{MYEC} = \frac{1}{nm}&\left\{\left[1+\frac{nV}{2m(n-1)k}\right]\sum_{h=1}^{m} y_{ih} + \sum_{h=1}^{m}\sum_{j=1}^{(n-2)} y_{ih+jk}\right. \\ &\left. + \left[1-\frac{nV}{2m(n-1)k}\right]\sum_{h=1}^{m} y_{ih+(n-1)k}\right\} \\ &= \overline{y}_{MLSS} + \frac{V}{2m^2(n-1)k}\left[\sum_{h=1}^{m}(y_{ih} - y_{ih+(n-1)k})\right],\end{aligned}$$

where

$$\overline{y}_{MLSS} = \frac{1}{nm}\sum_{h=1}^{m}\sum_{j=0}^{n-1} y_{ih+jk} = \frac{1}{nm}\left(\sum_{h=1}^{m} y_{ih} + \sum_{h=1}^{m}\sum_{j=1}^{n-2} y_{ih+jk} + \sum_{h=1}^{m} y_{ih+nk-k}\right). \tag{23}$$

***Proof of* Theorem 5**: By using (2.1) and (23) we obtain

$$\begin{aligned}
\overline{y}_{MLSS} - \overline{Y} &= \frac{1}{nm}\sum_{h=1}^{m}\sum_{j=0}^{(n-1)} y_{ih+jk} - a - \frac{b(N+1)}{2} - \overline{\overline{e}} \\
&= \frac{b}{m}\sum_{h=1}^{m} ih + \frac{bk(n-1)}{2} + \overline{e}_{MLSS} - \frac{b(N+1)}{2} - \overline{\overline{e}} \\
&= \frac{b}{m}\left[\sum_{h=1}^{m} ih - \frac{m(k+1)}{2}\right] + \overline{e}_{MLSS} - \overline{\overline{e}} = \frac{bV}{2m} + \overline{e}_{MLSS} - \overline{\overline{e}}, \qquad (24)
\end{aligned}$$

where $\overline{e}_{MLSS} = \sum_{h=1}^{m}\sum_{j=0}^{(n-1)} e_{ih+jk}/nm$. In addition,

$$\sum_{h=1}^{m}(y_{ih} - y_{ih+(n-1)k}) = -bm(n-1)k + \sum_{h=1}^{m}(e_{ih} - e_{ih+(n-1)k}). \qquad (25)$$

If we use (19), (24) and (25), then

$$\begin{aligned}
\mathcal{E}\left[\left(\overline{y}_{MYEC} - \overline{Y}\right)^2\right] &= \mathcal{E}\left\{\left[\overline{y}_{MLSS} - \overline{Y} + \frac{V\sum_{h=1}^{m}(y_{ih} - y_{ih+(n-1)k})}{2m^2(n-1)k}\right]^2\right\} \\
&= \mathcal{E}\left\{\left[\overline{e}_{MLSS} - \overline{\overline{e}} + \frac{V\sum_{h=1}^{m}(e_{ih} - e_{ih+(n-1)k})}{2m^2(n-1)k}\right]^2\right\}. \qquad (26)
\end{aligned}$$

Using the conditions of Cochran's (1946) super-population model, we obtain

$$\begin{aligned}
\mathcal{E}\left(\overline{e}^2_{MLSS}\right) &= \frac{1}{n^2m^2}\left[\sum_{h=1}^{m}\sum_{j=0}^{(n-1)}\mathcal{E}\left(e^2_{ih+jk}\right) + \sum_{h=1}^{m}\sum_{j=0}^{(n-1)}\sum_{p\neq j}\mathcal{E}\left(e_{ih+jk}e_{ih+pk}\right)\right] \\
&= \frac{\sigma^2}{nm}, \\
\mathcal{E}\left(\overline{\overline{e}}^2\right) &= \frac{1}{N^2}\left[\sum_{q=1}^{N}\mathcal{E}\left(e_q^2\right) + \sum_{z=1}^{N}\sum_{q\neq z}^{N}\mathcal{E}\left(e_q e_z\right)\right] = \frac{\sigma^2}{N}, \\
\mathcal{E}\left(\overline{e}_{MLSS}\overline{\overline{e}}\right) &= \frac{1}{nmN}\left[\sum_{h=1}^{m}\sum_{j=0}^{(n-1)}\sum_{q=1}^{N}\mathcal{E}\left(e_{ih+jk}e_q\right)\right] = \frac{\sigma^2}{N},
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}\left[\overline{e}_{MLSS}\sum_{h=1}^{m}\left(e_{ih} - e_{ih+(n-1)k}\right)\right] &= \frac{1}{nm}\sum_{h=1}^{m}\sum_{j=0}^{n-1}\mathcal{E}\left[e_{ih}e_{ih+jk} - e_{ih+jk}e_{ih+nk-k}\right] \\
&= \frac{m\sigma^2 - m\sigma^2}{nm} = 0, \\
\mathcal{E}\left[\overline{\overline{e}}\sum_{h=1}^{m}\left(e_{ih} - e_{ih+(n-1)k}\right)\right] &= \frac{1}{N}\sum_{q=1}^{N}\sum_{h=1}^{m}\mathcal{E}\left[e_q e_{ih} - e_q e_{ih+nk-k}\right] \\
&= \frac{m\sigma^2 - m\sigma^2}{N} = 0
\end{aligned}$$

and

$$\mathcal{E}\left[\sum_{h=1}^{m}\left(e_{ih}-e_{ih+(n-1)k}\right)^2\right]=\sum_{h=1}^{m}\mathcal{E}\left(e_{ih}^2-2e_{ih}e_{ih+nk-k}+e_{ih+nk-k}^2\right)=2m\sigma^2.$$

Expanding (26) and then substituting these expressions results in

$$\mathcal{E}\left[\left(\overline{y}_{MYEC}-\overline{Y}\right)^2\right]=\frac{\sigma^2}{nm}+\frac{\sigma^2}{N}-\frac{2\sigma^2}{N}+\frac{2m\sigma^2V^2}{4m^4(n-1)^2k^2}=\sigma_l^2+\frac{\sigma^2V^2}{2m^5(n-1)^2l^2}.$$

An expression for the expected MSE of $\overline{y}_{MYEC}$ is thus given as

$$\begin{aligned}M_{MYEC} &\triangleq \mathcal{E}\left\{\mathrm{E}\left[\left(\overline{y}_{MYEC}-\overline{Y}\right)^2\right]\right\}\\ &=\mathrm{E}\left\{\mathcal{E}\left[\left(\overline{y}_{MYEC}-\overline{Y}\right)^2\right]\right\}=\sigma_l^2+\frac{\sigma^2\mathrm{E}\left(V^2\right)}{2m^5(n-1)^2l^2}.\end{aligned}$$

We then conclude the proof since $\mathrm{E}\left(V^2\right)$ reduces to $m^2(l-1)(lm+1)/3$.

***Proof of*** **Theorem 6**: For MBSS and MMSS we respectively denote the $i$th ($i \in \{1,...,k\}$) sample totals by

$$BSST_i=\begin{cases}\sum_{j=0}^{(n-2)/2}(y_{i+2jk}+y_{2(j+1)k-i+1}), & \textit{if } n \textit{ is even}\\ \sum_{j=0}^{(n-3)/2}(y_{i+2jk}+y_{2(j+1)k-i+1})+y_{i+(n-1)k}, & \textit{if } n \textit{ is odd,}\end{cases}$$

and

$$MSST_i=\begin{cases}\sum_{j=0}^{(n-2)/2}(y_{i+jk}+y_{N-jk-i+1}), & \textit{if } n \textit{ is even}\\ \sum_{j=0}^{(n-3)/2}(y_{i+jk}+y_{N-jk-i+1})+y_{i+(n-1)k/2}, & \textit{if } n \textit{ is odd,}\end{cases}$$

Let us assume an indicator variable, given by

$$I_i=\begin{cases}1, & \text{if unit } y_{ih} \text{ is in the sample;}\\ 0, & \text{otherwise.}\end{cases}$$

If we assume that the $BSST_i$ and $MSST_i$ are fixed but unknown constants, then

$$\begin{aligned}\mathrm{E}(\overline{y}_{MBSS}) &= \mathrm{E}\left[\frac{1}{nm}\sum_{i=1}^{k}I_i(BSST_i)\right]\\ &=\frac{1}{nm}\sum_{i=1}^{k}\mathrm{E}(I_i)BSST_i\\ &=\frac{1}{nm}\sum_{i=1}^{k}\left(\frac{m}{k}\right)BSST_i=\frac{m}{nmk}\sum_{i=1}^{k}BSST_i=\frac{Y.}{nk}=\overline{Y}\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}(\overline{y}_{MMSS}) &= \mathrm{E}\left[\frac{1}{nm}\sum_{i=1}^{k} I_i(MSST_i)\right] \\
&= \frac{1}{nm}\sum_{i=1}^{k} \mathrm{E}(I_i)MSST_i \\
&= \frac{1}{nm}\sum_{i=1}^{k}\left(\frac{m}{k}\right)MSST_i = \frac{m}{nmk}\sum_{i=1}^{k} MSST_i = \frac{Y.}{nk} = \overline{Y},
\end{aligned}$$

since we are selecting $m$ samples from the $k$ possible samples, using SRS.

***Proof of* Theorem 7**: The single-start sample means ($BSST_i/n$ and $MSST_i/n$, for $i = 1, ..., k$) can now be viewed as population units. Remembering that SRS involves the random selection of $n$ sampling units from $N$ possible sampling units, the variance of the sample mean $\overline{y}_{SRS}$ is given by (1.5). Thus, by replacing $y_q$, $N$ and $n$ in (1.5) by $BSST_i/n$ (or $MSST_i/n$), $k$ and $m$ respectively, we obtain the variances of $\overline{y}_{MBSS}$ and $\overline{y}_{MMSS}$, which are respectively written as

$$\mathrm{V}(\overline{y}_{MBSS}) = \frac{S^2_{BSS}}{m}\left(\frac{k-m}{k}\right), \tag{27}$$

and

$$\mathrm{V}(\overline{y}_{MMSS}) = \frac{S^2_{MSS}}{m}\left(\frac{k-m}{k}\right), \tag{28}$$

where $S^2_{BSS} = \sum_{i=1}^{k}(BSST_i/n - \overline{Y})^2/(k-1)$ and $S^2_{MSS} = \sum_{i=1}^{k}(MSST_i/n - \overline{Y})^2/(k-1)$, such that the replacement of $y_q$ and $N$ in $\overline{Y} = \sum_{q=1}^{N} y_q/N$, by $BSST_i/n$ (or $MSST_i/n$) and $k$ respectively, results in $\sum_{i=1}^{k} BSST_i/nk = \sum_{i=1}^{k} MSST_i/nk = \overline{Y}$.

Now, if we compare $\mathrm{V}(\overline{y}_{MBSS})$ and $\mathrm{V}(\overline{y}_{MMSS})$ to $\widehat{\mathrm{V}}(\overline{y}_{MBSS})$ and $\widehat{\mathrm{V}}(\overline{y}_{MMSS})$, respectively, then it is clear that we need to show that $\sum_{h=1}^{m}(BSST_{ih} - \overline{BSST})^2/\left[n^2(m-1)\right]$ and $\sum_{h=1}^{m}(MSST_{ih} - \overline{MSST})^2/\left[n^2(m-1)\right]$ are unbiased estimates of $S^2_{BSS}$ and $S^2_{MSS}$, respectively. Using (27) and (28) as well as the property that $\mathrm{E}(I_i) = m/k$, we thus obtain

$$\begin{aligned}
\mathrm{E}\left[\frac{\sum_{h=1}^{m}\left(BSST_{ih} - \overline{BSST}\right)^2}{n^2(m-1)}\right] &= \frac{1}{m-1}\mathrm{E}\left[\sum_{h=1}^{m}\left(\frac{BSST_{ih}}{n} - \overline{y}_{MBSS}\right)^2\right] \\
&= \frac{1}{m-1}\mathrm{E}\left[\sum_{h=1}^{m}\left(\frac{BSST_{ih}}{n} - \overline{Y}\right)^2 - m\left(\overline{y}_{MBSS} - \overline{Y}\right)^2\right] \\
&= \frac{1}{m-1}\left\{\mathrm{E}\left[\sum_{i=1}^{k} I_i\left(\frac{BSST_i}{n} - \overline{Y}\right)^2\right] - m\mathrm{V}\left(\overline{y}_{MBSS}\right)\right\} \\
&= \frac{1}{m-1}\left[\frac{m(k-1)S_{BSS}^2}{k} - \frac{(k-m)S_{BSS}^2}{k}\right] = S_{BSS}^2.
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left[\frac{\sum_{h=1}^{m}\left(MSST_{ih} - \overline{MSST}\right)^2}{n^2(m-1)}\right] &= \frac{1}{m-1}\mathrm{E}\left[\sum_{h=1}^{m}\left(\frac{MSST_{ih}}{n} - \overline{y}_{MMSS}\right)^2\right] \\
&= \frac{1}{m-1}\mathrm{E}\left[\sum_{h=1}^{m}\left(\frac{MSST_{ih}}{n} - \overline{Y}\right)^2 - m\left(\overline{y}_{MMSS} - \overline{Y}\right)^2\right] \\
&= \frac{1}{m-1}\left\{\mathrm{E}\left[\sum_{i=1}^{k} I_i\left(\frac{MSST_i}{n} - \overline{Y}\right)^2\right] - m\mathrm{V}\left(\overline{y}_{MMSS}\right)\right\} \\
&= \frac{1}{m-1}\left[\frac{m(k-1)S_{MSS}^2}{k} - \frac{(k-m)S_{MSS}^2}{k}\right] = S_{MSS}^2.
\end{aligned}$$

***Proof of*** **Theorem 8**: If we consider the case of $n$ is even for MBSS and MMSS, such that $e_{MBSS} = \sum_{j=0}^{(n-2)/2}(e_{i+2jk} + e_{2(j+1)k-i+1})$ and $e_{MMSS} = \sum_{j=0}^{(n-2)/2}(e_{i+jk} + e_{N-jk-i+1})$ respectively denote the random errors associated with $i$th ($i \in \{1, ..., k\}$) balanced systematic sample and modified systematic sample, then substituting (2.1) into $BSST_i$ and $MSST_i$ results in

$$BSST_i = \sum_{j=0}^{(n-2)/2}[2a + b(4jk + 2k + 1)] + e_{MBSS} = an + bn\left(\frac{N+1}{2}\right) + e_{MBSS}$$

and

$$MSST_i = \sum_{j=0}^{(n-2)/2}[2a + b(N + 1)] + e_{MMSS} = an + bn\left(\frac{N+1}{2}\right) + e_{MMSS}.$$

Expressions for the expected MSEs of $\overline{y}_{MBSS}$ and $\overline{y}_{MMSS}$ are obtained by taking the expectation of (27) and (28), before substituting $BSST_i$, $MSST_i$ and $\overline{Y}$ into the relevant expressions, given by

$$\begin{aligned}
M_{MBSS} &= \frac{(k-m)}{mk(k-1)} \sum_{i=1}^{k} \mathcal{E}\left(\frac{BSST_i}{n} - \overline{Y}\right)^2 \\
&= \frac{(k-m)}{mk\,(k-1)} \sum_{i=1}^{k} \mathcal{E}\left[a + b\left(\frac{N+1}{2}\right) + \frac{e_{MBSS}}{n} - \left(a + \frac{b\,(N+1)}{2} + \overline{\overline{e}}\right)\right]^2 \\
&= \frac{(k-m)}{mk(k-1)} \sum_{i=1}^{k} \mathcal{E}\left(\frac{e_{MBSS}^2}{n^2} - \frac{2e_{MBSS}\overline{\overline{e}}}{n} + \overline{\overline{e}}^2\right) \qquad (29)
\end{aligned}$$

and

$$\begin{aligned}
M_{MMSS} &= \frac{(k-m)}{mk(k-1)} \sum_{i=1}^{k} \mathcal{E}\left(\frac{MSST_i}{n} - \overline{Y}\right)^2 \\
&= \frac{(k-m)}{mk\,(k-1)} \sum_{i=1}^{k} \mathcal{E}\left[a + b\left(\frac{N+1}{2}\right) + \frac{e_{MMSS}}{n} - \left(a + \frac{b\,(N+1)}{2} + \overline{\overline{e}}\right)\right]^2 \\
&= \frac{(k-m)}{mk(k-1)} \sum_{i=1}^{k} \mathcal{E}\left(\frac{e_{MMSS}^2}{n^2} - \frac{2e_{MMSS}\overline{\overline{e}}}{n} + \overline{\overline{e}}^2\right). \qquad (30)
\end{aligned}$$

Now, since there are $n$ terms in $e_{MBSS}$ and $e_{MMSS}$ as well as $N$ terms in $\overline{\overline{e}}$, it follows that

$$\begin{aligned}
&\mathcal{E}\left(e_{MBSS}^2\right) = n\sigma^2, \quad \mathcal{E}\left(e_{MBSS}\overline{\overline{e}}\right) = \frac{1}{N}\mathcal{E}[e_{MBSS}(e_1 + ... + e_N)] = \frac{n\sigma^2}{N}, \\
&\mathcal{E}\left(e_{MMSS}^2\right) = n\sigma^2, \quad \mathcal{E}\left(e_{MMSS}\overline{\overline{e}}\right) = \frac{1}{N}\mathcal{E}[e_{MMSS}(e_1 + ... + e_N)] = \frac{n\sigma^2}{N}, \\
&\mathcal{E}\left(\overline{\overline{e}}^2\right) = \frac{1}{N^2}\left[\sum_{q=1}^{N} \mathcal{E}\left(e_q^2\right) + \sum_{z=1}^{N}\sum_{q\neq z}^{N} \mathcal{E}\left(e_q e_z\right)\right] = \frac{\sigma^2}{N}.
\end{aligned}$$

Remembering that $k = lm$, we then substitute these relevant expressions into (29) and (30) to obtain

$$M_{MBSS} = \frac{(k-m)}{mk(k-1)} \sum_{i=1}^{k} \left(\frac{\sigma^2}{n} - \frac{2\sigma^2}{N} + \frac{\sigma^2}{N}\right) = \frac{(k-m)\sigma^2}{mN} = \frac{(l-1)\sigma^2}{N} = \sigma_l^2$$

and

$$M_{MMSS} = \frac{(k-m)}{mk(k-1)} \sum_{i=1}^{k} \left(\frac{\sigma^2}{n} - \frac{2\sigma^2}{N} + \frac{\sigma^2}{N}\right) = \frac{(k-m)\sigma^2}{mN} = \frac{(l-1)\sigma^2}{N} = \sigma_l^2.$$

Similarly, we can use the above method for the case of $n$ being odd and thus conclude the proof.

***Proof of*** **Theorem 10**: If we compare (3.15) to (3.16), then it is clear that we need to show that $\sum_{h=1}^{m}\left(Tj_{ih} - \overline{T}\right)^2/[n^2(m-1)]$ is an unbiased estimate of $S_T^2$, in order to prove the theorem. Using (3.15) and the property that $\mathrm{E}(I_i) = m/k$, we thus obtain

$$\begin{aligned}
\mathrm{E}\left[\frac{\sum_{h=1}^{m}\left(Tj_{ih}-\overline{T}\right)^2}{n^2(m-1)}\right] &= \frac{1}{m-1}\mathrm{E}\left[\sum_{h=1}^{m}\left(\frac{Tj_{ih}}{n}-\overline{y}_{MBMSS}\right)^2\right] \\
&= \frac{1}{m-1}\mathrm{E}\left[\sum_{h=1}^{m}\left(\frac{Tj_{ih}}{n}-\overline{Y}\right)^2 - m\left(\overline{y}_{MBMSS}-\overline{Y}\right)^2\right] \\
&= \frac{1}{m-1}\left\{\mathrm{E}\left[\sum_{i=1}^{k}I_i\left(\frac{Tj_i}{n}-\overline{Y}\right)^2\right] - m\mathrm{V}\left(\overline{y}_{MBMSS}\right)\right\} \\
&= \frac{1}{m-1}\left[\frac{m(k-1)S_T^2}{k} - \frac{(k-m)S_T^2}{k}\right] = S_T^2.
\end{aligned}$$

***Proof of* Theorem 11**: An estimate of $\overline{Y}$ with random starts $ih$, for $h = 1, ..., m$ and $ih \in \{1, ..., k\}$, can be written as

$$\overline{y}_{MBMSSEC} = \frac{1}{nm}\left(\lambda_1\sum_{h=1}^{m}y_{x1h} + \sum_{h=1}^{m}\sum_{i=2}^{n-1}y_{xih} + \lambda_2\sum_{h=1}^{m}y_{xnh}\right), \tag{31}$$

where $\lambda_1$ and $\lambda_2$ are the respective weights applied to the first and last sampling units of each selected sample and $x1h, ..., xnh$ are the indices belonging to the selected balanced modified systematic samples, which is arranged in ascending order. Substituting model (2.12) into (31) and then equating this expression to $\overline{Y}_B = a + b(N+1)/2$ results in

$$\begin{aligned}
\overline{y}_{MBMSSEC} &= \frac{1}{nm}\left[\lambda_1\sum_{h=1}^{m}(a+bx1h) + \sum_{h=1}^{m}\sum_{i=2}^{n-1}(a+bxih) + \lambda_2\sum_{h=1}^{m}(a+bxnh)\right] \\
&= a + \frac{b(N+1)}{2}.
\end{aligned} \tag{32}$$

By equating the coefficients of $a$ in (32), it follows that

$$\lambda_1 = 2 - \lambda_2. \tag{33}$$

Similarly, by equating the coefficients of $b$ in (32), we obtain

$$2\lambda_1\sum_{h=1}^{m}x1h + 2\sum_{h=1}^{m}\sum_{i=2}^{n-1}xih + 2\lambda_2\sum_{h=1}^{m}xnh = nm(N+1). \tag{34}$$

Substituting (33) into (34) results in

$$4\sum_{h=1}^{m}x1h - 2\lambda_2\sum_{h=1}^{m}x1h + 2\sum_{h=1}^{m}\sum_{i=2}^{n-1}xih + 2\lambda_2\sum_{h=1}^{m}xnh = nm(N+1),$$

which simplifies to

$$\lambda_2 = \frac{Q - 2\sum_{h=1}^{m} x1h}{\sum_{h=1}^{m}(xnh - x1h)}, \tag{35}$$

where $Q = nm(N+1)/2 - \sum_{h=1}^{m}\sum_{i=2}^{(n-2)} xih$. The weight applied to the first sampling units is thus obtained by substituting (35) into (33), such that

$$\lambda_2 = \frac{2\sum_{h=1}^{m} xnh - Q}{\sum_{h=1}^{m}(xnh - x1h)}. \tag{36}$$

On substituting (35) and (36) into (31) we obtain

$$\begin{aligned}\overline{y}_{MBMSSEC} &= \frac{1}{nm}\left[\frac{(2\sum_{h=1}^{m} xnh - Q)\sum_{h=1}^{m} y_{x1h}}{\sum_{h=1}^{m}(xnh - x1h)} + \sum_{h=1}^{m}\sum_{i=2}^{(n-1)} y_{xih}\right.\\ &\quad \left. + \frac{(Q - 2\sum_{h=1}^{m} x1h)\sum_{h=1}^{m} y_{xnh}}{\sum_{h=1}^{m}(xnh - x1h)}\right]\\ &= \overline{y}_{MBMSS} + \frac{[\sum_{h=1}^{m}(xnh - x1h) - Q]}{nm\sum_{h=1}^{m}(xnh - x1h)}\sum_{h=1}^{m}(y_{x1h} - y_{xnh}), \end{aligned} \tag{37}$$

where $nm\overline{y}_{MBMSS} = \sum_{h=1}^{m} y_{x1h} + \sum_{h=1}^{m}\sum_{i=2}^{(n-1)} y_{xih} + \sum_{h=1}^{m} y_{xnh}$. Now, if we consider Case (B), then $x1h = ih$, $xnh = N + ih - k$ and

$$\begin{aligned} Q &= \frac{nm(N+1)}{2} - \sum_{h=1}^{m}\sum_{i=2}^{n-1} xih\\ &= \frac{nm(N+1)}{2} - \sum_{h=1}^{m}\left[\sum_{j=1}^{(n-2)/4}(N + 2ih - k) + \sum_{j=0}^{(n-6)/4}(N + k - 2ih + 2)\right]\\ &= \frac{nm(N+1)}{2} - \sum_{h=1}^{m}\frac{2(N+1)(n-2)}{4} = m(N+1). \end{aligned}$$

Substituting these values into (37) results in

$$\begin{aligned}\overline{y}_{MBMSSEC} &= \overline{y}_{MBMSS} + \frac{[\sum_{h=1}^{m}(N - k + 2ih) - m(N+1)]}{nm\sum_{h=1}^{m}(N-k)}\sum_{h=1}^{m}(y_{x1h} - y_{xnh})\\ &= \overline{y}_{MBMSS} + \frac{V\sum_{h=1}^{m}(y_{ih} - y_{N+ih-k})}{nm^2(N-k)}. \end{aligned}$$

Similarly, we can find values of $x1h, xnh$ and $Q$ for Cases (C) to (E) and then substitute these values into (37), so as to conclude the proof.

***Proof of*** **Theorem 14**: If we consider Cases E and F, then an estimate of $\overline{Y}_1$ with random start $k1$, where $k1 \in \{1, ..., k\}$, can be written as

$$\overline{y}'_{k1} = \frac{1}{n-r}\Bigg[\phi_1 y_{k1} + y_{k1+(n-r-1)k/2} + \sum_{j=1}^{(n-r-3)/2}\left(y_{k1+jk} + y_{(n-r)k-jk-k1+1}\right) + \phi_2 y_{(n-r)k-k1+1}\Bigg], \tag{38}$$

where $\phi_1$ and $\phi_2$ are the weights applied to the first and last sampling units of $s_{k1}$, respectively. If we consider model (2.12), then

$$\overline{Y}_1 = \frac{1}{(n-r)k}\sum_{q=1}^{(n-r)k} y_q = a + \frac{b[(n-r)k+1]}{2}$$

and

$$\overline{Y}_2 = \frac{1}{r(k+1)}\sum_{q=(n-r)k+1}^{N} y_q = a + \frac{b[(n-r)k+N+1]}{2}.$$

By substituting model (2.12) into (38) and equating this result to $\overline{Y}_1$, we obtain

$$\phi_1(a+bk1) + \sum_{j=1}^{(n-r-3)/2}\{2a + b\left[(n-r)k+1\right]\} + a + b\left[k1 + \frac{(n-r-1)k}{2}\right] + \phi_2\left\{a + b\left[(n-r)k - k1 + 1\right]\right\} = (n-r)a + \frac{b(n-r)[(n-r)k+1]}{2}. \tag{39}$$

If we equate the coefficients of $a$ in (39), then

$$\phi_1 = 2 - \phi_2. \tag{40}$$

Similarly, by equating the coefficients of $b$ in (39), we obtain

$$2\phi_1 k1 - 2(n-r)k - 3 + 2k1 - k + 2\phi_2[(n-r)k - k1 + 1] = 0. \tag{41}$$

Substituting (40) into (41) results in

$$\phi_2 = 1 - \frac{(2k1-k-1)}{[2(n-r)k - 4k1 + 2]}. \tag{42}$$

The weight applied to the first sampling unit is thus obtained by substituting (42) into (40), i.e.

$$\phi_1 = 1 + \frac{(2k1-k-1)}{[2(n-r)k - 4k1 + 2]}. \tag{43}$$

Now, by substituting (42) and (43) into (38), we get

$$\begin{aligned}\overline{y}'_{k1} &= \frac{1}{n-r}\Bigg[y_{k1} + \sum_{j=1}^{(n-r-3)/2}\left(y_{k1+jk} + y_{(n-r)k-jk-k1+1}\right) + y_{k1+(n-r-1)k/2} \\ &\quad + \frac{(2k1-k-1)}{[2(n-r)k-4k1+2]}\left(y_{k1} - y_{(n-r)k-k1+1}\right)\Bigg] \\ &= \overline{y}_{k1} + \frac{(2k1-k-1)}{(n-r)[2(n-r)k-4k1+2]}\left(y_{k1} - y_{(n-r)k-k1+1}\right).\end{aligned}$$

Finally, we replace this expression for $\overline{y}_{k1}$ in (4.6), such that

$$\overline{y}_{RMSSEC} = \overline{y}_{RMSS} + \frac{k(2k1 - k - 1)}{N[2(n-r)k - 4k1 + 2]} \left( y_{k1} - y_{(n-r)k-k1+1} \right).$$

Note that estimator $\overline{y}_{k2}$ need not be adjusted, since $\overline{y}_{k2} = \overline{Y}_2$ for Cases E and F. Similarly, we can use the above method for Cases D and G to conclude the proof, where the only adjusted estimator will be denoted by $\overline{y}'_{k2}$.

# Bibliography

Bellhouse, D. R. (1984), 'On the choice of the sampling interval in circular systematic sampling', *Sankhyā: The Indian Journal of Statistics, Series B,* **46**(2), 247–248.

Bellhouse, D. R. (1988), *Systematic sampling*, Vol. 6, Elsevier, Amsterdam, pp. 125–145.

Bellhouse, D. R. & Rao, J. N. K. (1975), 'Systematic sampling in the presence of a trend', *Biometrika* **62**(3), 694–697.

Breidt, F. J. (1995), 'Markov chain designs for one-per-stratum sampling', *Survey Methodology* (21), 63–70.

Chang, H. J. & Huang, K. C. (2000), 'Remainder linear systematic sampling', *Sankhyā: The Indian Journal of Statistics, Series B,* **62**(2), 249–256.

Cochran, W. G. (1946), 'Relative accuracy of systematic and stratified random samples for a certain class of populations', *The Annals of Mathematical Statistics* **17**, 164–177.

Cochran, W. G. (1977), *Sampling Techniques*, 3rd edn, John Wiley and Sons, Inc., New York.

Deming, W. E. (1960), *Sample Design in Business Research*, Wiley, New York.

D'Orazio, M. (2003), 'Estimating the variance of the sample mean in two-dimensional systematic sampling', *Journal of Agricultural, Biological, and Environmental Statistics* **8**(3), 280–295.

Dunn, R. & Harrison, A. R. (1993), 'Two-dimensional systematic sampling of land use', *Journal of the Royal Statistical Society, Series C (Applied Statistics),* **42**(4), 585–601.

Finney, D. J. (1948), 'Random and systematic sampling in timber surveys', *Forestry* **22**, 64–99.

Fountain, R. L. & Pathak, P. K. (1989), 'Systematic and nonrandom sampling in the presence of linear trends', *Communications in Statistics - Theory and Methods* **18**(7), 2511–2526.

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

Gupta, A. K. & Kabe, D. G. (2011), *Theory of Sample Surveys*, World Scientific Publishing Company.

Hasel, A. A. (1938), 'Sampling error in timber surveys', *Journal of Agricultural Research* **57**(10), 713–736.

Horvitz, D. G. & Thompson, D. J. (1952), 'A generalization of sampling without replacement from a finite universe', *Journal of the American Statistical Association* **47**(260), 663–685.

Huang, K. C. (2004), 'Mixed random systematic sampling designs', *Metrika* **59**(1), 1–11.

Iachan, R. (1982), 'Systematic sampling: a critical review', *International Statistical Review* **50**(3), 293–303.

Jacobsen, J. S. (1998), Soil sampling, Technical report, Montana State University Extension Publications.

Kalton, G. (1983), *Introduction to Survey Sampling*, SAGE Publications, Beverly Hills.

Kao, F. F., Leu, C. H. & Ko, C. H. (2011*a*), 'Remainder markov systematic sampling', *Journal of Statistical Planning and Inference* **141**(11), 3595–3604.

Kao, F. F., Leu, C. H. & Ko, C. H. (2011*b*), 'Remainder systematic markov chain design', *International Journal of Information and Management Sciences* **22**(4), 327–341.

Khan, Z., Gupta, S. N. & Shabbir, J. (2014), 'A note on diagonal circular systematic sampling', *Journal of Statistical Theory and Practice* **8**(3), 439–443.

Khan, Z., Shabbir, J. & Gupta, S. N. (2013), 'A new sampling design for systematic sampling', *Communications in Statistics - Theory and Methods* **42**(18), 3359–3370.

Khan, Z., Shabbir, J. & Gupta, S. N. (2015), 'Generalized systematic sampling', *Communications in Statistics - Simulation and Computation* **44**(9), 2240–2250.

Kish, L. (1965), *Survey sampling*, John Wiley and Sons, Inc., New York.

Kouijn, H. S. (1973), 'On modified schemes of systematic sampling with several random starts', *Communications in Statistics* **1**(2), 133–148.

Lahiri, D. B. (1951), 'A method of sample selection providing unbiased ratio estimates', *Bulletin of the International Statistical Institute* **33**(2), 133–140.

Lehtonen, R. & Pahkinen, E. (2004), *Basic sampling techniques*, 2nd edn, John Wiley and Sons, Ltd, Chichester, pp. 9–58.

Leu, C. H. & Kao, F. F. (2006), 'Modified balanced circular systematic sampling', *Statistics & Probability Letters* **76**, 373–383.

Leu, C. H. & Tsui, K. W. (1996), 'New partially systematic sampling', *Statistica Sinica* **6**, 616–630.

Lohr, S. L. (2010), *Sampling: Design and Analysis*, 2nd edn, Brooks/Cole, Boston.

Madow, L. H. (1946), 'Systematic sampling and its relation to other sampling designs', *Journal of the American Statistical Association* **41**(234), 204–217.

Madow, W. G. (1953), 'On the theory of systematic sampling, iii. comparison of centered and random start systematic sampling', *The Annals of Mathematical Statistics* **24**(1), 101–106.

Madow, W. G. & Madow, L. H. (1944), 'On the theory of systematic sampling, i', *The Annals of Mathematical Statistics* **15**, 1–24.

Mahalanobis, P. C. (1946), 'Recent experiments in statistical sampling in the indian statistical institute', *Journal of the Royal Statistical Society* **109**(4), 325–370.

Mason, B. J. (1994), 'Preparation of soil sampling protocols: sampling techniques and strategies', *Houston Geological Society Bulletin* **36**(10), 22–23 and 33.

McArthur, R. D. (1987), 'An evaluation of sample designs for estimating a locally concentrated pollutant', *Communications in Statistics - Simulation and Computation* **16**, 735–759.

Mostafa, S. A. & Ahmad, I. A. (2016), 'Remainder linear systematic sampling with multiple random starts', *Journal of Statistical Theory and Practice* **10**(4), 824–851.

Mukerjee, R. & Sengupta, S. (1990), 'Optimal estimation of a finite population mean in the presence of linear trend', *Biometrika* **77**(3), 625–630.

Murthy, M. N. (1967), *Sampling: Theory and Methods*, Statistical Publishing Society, Calcutta.

Murthy, M. N. & Rao, T. J. (1988), *Systematic sampling with illustrative examples*, Vol. 6, Elsevier, Amsterdam, chapter 7, pp. 147–185.

Naidoo, L. R. (2013), Systematic Sampling from Finite Populations, Master's thesis, University of KwaZulu-Natal, South Africa.

Osborne, J. G. (1942), 'Sampling errors of systematic and random surveys of cover-type areas', *Journal of the American Statistical Association* **37**(218), 256–264.

Pawley, M. D. M. (2006), Systematic sampling in ecology, PhD thesis, The University of Auckland.

Rana, R. S. & Singh, R. (1989), 'Note on systematic sampling with supplementary observations', *Sankhyā: The Indian Journal of Statistics, Series B,* **51**, 205–211.

Ruiz Espejo, M. (1997), 'Uniqueness of the zinger strategy with estimable variance: Ranasingh estimator', *Sankhyā: The Indian Journal of Statistics, Series B,* **59**(1), 76–83.

Sampath, S. (2009), 'Finite population variance estimation under lss with multiple random starts', *Communications in Statistics - Theory and Methods* **38**(19), 3596–3607.

Sampath, S. & Ammani, S. (2010), 'Some systematic sampling strategies using multiple random starts', *Pakistan Journal of Statistics and Operation Research* **6**(2), 32–50.

Sampath, S. & 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.

Sampath, S. & Uthayakumaran, N. (1998), 'Markov systematic sampling', *Biometrical Journal* **40**(7), 883–895.

Sampath, S. & Varalakshmi, V. (2009), 'Diagonal circular systematic sampling', *Model Assisted Statistics and Applications* **3**(4), 345–352.

Sen, A. R. (1953), 'On the estimate of the variance in sampling with varying probabilities', *Indian Society of Agricultural Statistics* **5**, 119–127.

Sengupta, S. & Chattopadhyay, S. (1987), 'A note on circular systematic sampling', *Sankhyā: The Indian Journal of Statistics, Series B,* **49**(2), 186–187.

Sethi, V. K. (1965), 'On optimum pairing of units', *Sankhyā: The Indian Journal of Statistics, Series B,* **27**, 315–320.

Shiue, C. J. (1960), 'Systematic sampling with multiple random starts', *Forest Science* **6**, 42–50.

Singh, D., Jindal, K. K. & Garg, J. N. (1968), 'On modified systematic sampling', *Biometrika* **55**(3), 541–546.

Singh, D. & Singh, P. (1977), 'New systematic sampling', *Journal of Statistical Planning and Inference* **1**(2), 163–177.

Singh, P. & Garg, J. N. (1979), 'On balanced random sampling', *Sankhyā: The Indian Journal of Statistics, Series C,* **41**, 60–68.

Subramani, J. (2000), 'Diagonal systematic sampling scheme for finite populations', *Journal of the Indian Society of Agricultural Statistics* **53**(2), 187–195.

Subramani, J. (2009), 'Further results on diagonal systematic sampling scheme for finite populations', *Journal of the Indian Society of Agricultural Statistics* **63**(3), 277–282.

Subramani, J. (2010), 'Generalization of diagonal systematic sampling scheme for finite populations', *Model Assisted Statistics and Applications* **5**(2), 117–128.

Subramani, J. (2012), 'A modification on linear systematic sampling for odd sample size', *Bonfring International Journal of Data Mining* **2**(2), 32–36.

Subramani, J. (2013*a*), 'A further modification on linear systematic sampling for finite populations', *Journal of Statistical Theory and Practice* **7**(3), 471–479.

Subramani, J. (2013*b*), 'A modification on linear systematic sampling', *Model Assisted Statistics and Applications* **8**(3), 215–227.

Subramani, J. (2014), 'Star type systematic sampling for finite populations', *Communications in Statistics - Theory and Methods* **43**(1), 175–190.

Subramani, J. & Gupta, S. N. (2014), 'Generalized modified linear systematic sampling scheme for finite populations', *Hacettepe Journal of Mathematics and Statistics* **43**(3), 529–542.

Subramani, J., Gupta, S. N. & Prabavathy, G. (2014), 'Circular systematic sampling in the presence of linear trend', *American Journal of Mathematical and Management Sciences* **33**(1), 1–19.

Subramani, J. & Singh, S. (2014), 'Estimation of population mean in the presence of linear trend', *Communications in Statistics - Theory and Methods* **43**(15), 3095–3116.

Sudhakar, K. (1978), 'A note on circular systematic sampling design', *Sankhyā: The Indian Journal of Statistics, Series C,* **40**(1), 72–73.

Thompson, S. K. (1990), 'Adaptive cluster sampling', *Journal of the American Statistical Association* **85**, 1050–1059.

Thompson, S. K. & Seber, G. A. F. (1996), *Adaptive Sampling*, Wiley, New York.

Tornqvist, L. (1963), 'The theory of replicated systematic cluster sampling with random start', *Review of the International Statistical Institute* **31**(1), 11–23.

Tukey, J. W. (1950), 'Some sampling simplified', *Journal of the American Statistical Association* **45**(252), 501–519.

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

Wolter, K. M. (1984), 'An investigation of some estimators of variance for systematic sampling', *Journal of the American Statistical Association* **79**(388), 781–790.

Wolter, K. M. (2007), *Variance estimation for systematic sampling*, Springer, New York, pp. 298–353.

Wu, C. F. J. (1984), 'Estimation in systematic sampling with supplementary observations', *Sankhyā: The Indian Journal of Statistics, Series B,* **46**(3), 306–315.

Yates, F. (1948), 'Systematic sampling', *Philosophical Transactions of the Royal Society of London, Series A,* **241**(834), 345–377.

Yates, F. & Grundy, P. M. (1953), 'Selection without replacement from within strata with probability proportional to size', *Journal of the Royal Statistical Society, Series B (Methodological),* **15**(2), 253–261.

Zinger, A. (1963), 'Estimations de variances avec échantillonnage systématique', *Revue de Statistique Appliquée* **11**(2), 89–97.

Zinger, A. (1964), 'Systematic sampling in forestry', *Biometrics* **20**(3), 553–565.

Zinger, A. (1980), 'Variance estimation in partially systematic sampling', *Journal of the American Statistical Association* **75**(369), 206–211.