EXACT METHODS FOR COMPARISON OF ESTIMATION STRATEGIES IN SURVEY SAMPLING

by

Kaleem Uddin Khan, BSc.

A thesis submitted for the degree of Doctor of Philosophy in
Queen Mary, University of London

January 2010

Declaration

I, Kaleem Uddin Khan, declare that the work presented in the thesis are both my own, and have been generated by me as the result of my own original research.

Signed: ....

Acknowledgements

I would like to offer my sincerest gratitude to Dr Roger Sugden, my supervisor, for his invaluable support, encouragement, and stimulating discussions and suggestions throughout my research. It would have been next to impossible to successfully complete this thesis without his continuous help and guidance for which I am extremely grateful. I am also thankful to Prof Rosemary Bailey, my second supervisor, for her useful comments and suggestions about my work. I also like to thank her for running a Latex tutorial course which was very helpful for me when I wrote this thesis.

I wish to express my appreciation to Prof Steve Gilmour, who, as my undergraduate advisor and lecturer, helped me overcome my doubts in doing this thesis. Without his willing support and advice this research would not have been possible.

Many thanks are due to the staff at the ESD for providing a wide range of useful courses that helped improve my research skills. In particular I am indebted to Dr Cara Owens who gave me valuable advice for writing up a thesis. I also like to thank staff and colleagues at the School of Mathematical Sciences who provided the support I needed to produce and complete this thesis and the EPSRC for funding my study.

Finally, I thank my parents for their love and support throughout my life. I also wish to thank my brother and his wife for their support and understanding during my study.

Abstract

Many strategies in survey sampling depend on large sample approximation formulae for design-based inference on finite population parameters, which are not valid for small samples. We develop an approach using matrix algebra to tackle many problems for samples of any size.

Poststratification under a general unequal probability sampling design has received little attention and is an area that we will consider. We demonstrate that inference should be made conditional on the observed sampling allocation rather than unconditionally and examine different types of probability weights.

For certain strategies we give results that provide sufficient conditions for the superiority of one strategy over another. These methods are based on the exact mean square errors and are used to compare estimators under poststratification both conditionally and unconditionally.

We also present a result that gives an exact upper bound on the absolute bias ratio of a strategy which can be used at the design stage to assess the magnitude of the bias.

A general problem for unbiased variance estimators under unequal probability sampling is the possibility of obtaining a negative estimate. We show how the eigenvalues of the matrices given by a variance estimator for the ratio estimator under probability proportional to aggregate size sampling can be used to construct a class of nonnegative definite unbiased variance estimators. Our empirical studies show that estimators from this class are generally more efficient than the standard estimator, especially when the coefficient of variation of the size variable is large.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter we give a description of the conventional sample survey theory under the design-based approach giving standard notations and definitions in section 1.2. In section 1.3 we will discuss the concept of calibrated estimators and in section 1.4 we cover some aspects of poststratification. In section 1.5 it will be pointed out that nonnegative definite symmetric matrices can be useful for comparing the exact mean square errors of different strategies.

1.1 Basic ideas of design-based sampling theory

In sampling theory we study a finite set of individuals or units, called a population. Attached to these units are the values of several variables (defined for every unit in the population). The goal is to make inferences about the unknown values for parameters of the population that are based on the unknown values of the variables of interest or survey variables. A parameter is a function of a survey variable and is an unknown value of some measure of interest, for example the total, mean or variance of a survey variable. The units in a population can be listed where each unit has its own unique identifiable symbol or label. A collection of labelled units selected from the population is called a sample, a non-empty subset of the population. The general problem is to select a sample of units from the population, observe the values of the survey variables for those units and use these values to estimate the unknown values of parameters of interest.

At the selection or design stage the sampler chooses a method for sampling which involves probability selection to obtain a sample of units from the population. This procedure is called a sampling design or sampling scheme. The design is determined by assigning the probability of selection to every possible sample that can be selected. It can be based on prior information from previous studies, censuses or pilot surveys on variables thought to be related to the survey variables. The design can also be subject to cost and administrative constraints. Inferences are generally based on the design that has been executed.

At the estimation or inference stage the analyst has to choose estimators for the parameters of interest. An estimator is a function of the observed values of the survey variables from the sampled units. The choice of an estimator can depend on prior information as well as the design.

Once the variables of interest are observed for the sampled units, estimates for the parameters of interest are calculated according to the estimation formulae proposed. The usual measure of accuracy is given by a confidence interval (valid for large sample sizes) which is based on the sample estimate of the sampling variance. This latter quantity is also used as a measure of accuracy to compare strategies (either before the sample is drawn using prior knowledge or after the sample is drawn, to make a so called post-survey choice of an estimator). Another measure of accuracy is the mean square error of an estimator under the sample design.

A strategy is a combination of a design and an estimator. When analysts are free to choose a design they are in a position to select strategies that are, to their knowledge, strongly related to the population structure of the survey variables so that the design-based estimators will perform well.

1.2 Notations and definitions

Let the population of fixed size N be denoted by $\mathcal{U}$ and let the units of the population be identified by labelling them from 1 to N so that

$$\mathcal{U} = \{1, \ldots, N\}.$$

We will consider only one survey variable whose values will be denoted by y_i for each unit $i \in \mathcal{U}$. The $N \times 1$ column vector $\mathbf{y}$, where

$$\mathbf{y}^t = (y_1, \ldots, y_N),$$

will be called the survey vector.

Some population parameters of interest are:

- the population total

$$T_Y = \sum_{i \in \mathcal{U}} y_i = N\bar{Y}$$

- the population variance

$$S_Y^2 = \frac{\sum_{i \in \mathcal{U}} (y_i - \bar{Y})^2}{N - 1}$$

Associated with each unit in the population are the values of q, say, auxiliary variables which are a form of prior information about the population. The value of the k^{th} auxiliary variable for unit $i \in \mathcal{U}$ will be denoted by x_{ki} for $k = 1, \ldots, q$. The

$N \times q$ matrix $\mathbf{X}$ will be called the auxiliary matrix where the k^{th} column of $\mathbf{X}$ is equal to the k^{th} auxiliary vector $\mathbf{x}_k$ where

$$\mathbf{x}_k^t = (x_{k1}, x_{k2}, \ldots, x_{kN})$$

for $k = 1, \ldots, q$. The values of these variables can be incorporated into both the design and the estimators. We will assume that the values of the auxiliary variables are known for all units in the population (though this is not necessary for the implementation of many design-based strategies).

The sample will be denoted by s and the sample size will be denoted by a positive integer $n(s)$ which is less than N. Let p denote the design and let $p(s)$ denote the probability of selecting sample s under the design p. Then every sampling design has the property

$$\sum_{s \in \mathcal{S}} p(s) = 1$$

where $\mathcal{S}$, called a sampling support, is the set of all possible samples that could have been selected under the design p, i.e. $\mathcal{S}$ is the set of all s with $p(s) > 0$.

A design is of fixed size n if $n(s) = n$ for all $s \in \mathcal{S}$. Note that not all designs of fixed size n will give a sampling support $\mathcal{S}$ that contains every possible sample of fixed size n. Any sample of size $n(s) = n$ that this not contained in the support $\mathcal{S}$ will have probability $p(s) = 0$. We will only consider sampling designs of fixed size n.

Definition 1 *The single or first order inclusion probability of unit i, denoted by π_i, is the probability that unit i is included in the sample s,*

$$\pi_i = p(s : i \in s) = \sum_{s \ni i} p(s).$$

Similarly the r^{th} order inclusion probability, denoted by $\pi_{i_1 i_2 \ldots i_r}$, is the probability that r distinct units $\{i_1, i_2, \ldots, i_r\}$ from the population are included in the sample with

$r \leq n$ so that

$$\pi_{i_1 i_2 \dots i_r} = \sum_{s \ni i_1, i_2, \dots, i_r} p(s).$$

The randomness of an estimator is based on the design, so expectations are taken with respect to it. Let $\hat{\theta}_s$ be an estimator for some population parameter θ. The expected value and variance of $\hat{\theta}_s$, with respect to the design p, are given respectively by

$$\mathrm{E}(\hat{\theta}_s, p) = \sum_{s \in \mathcal{S}} p(s)\hat{\theta}_s$$

and

$$\begin{aligned} \mathrm{Var}(\hat{\theta}_s, p) &= \sum_{s \in \mathcal{S}} p(s)\hat{\theta}_s^2 - \left(\sum_{s \in \mathcal{S}} p(s)\hat{\theta}_s \right)^2 \\ &= \mathrm{E}(\hat{\theta}_s^2, p) - \mathrm{E}(\hat{\theta}_s, p)^2 \end{aligned}$$

Another measure of quality for an estimator is the bias:

Definition 2 *The bias of an estimator $\hat{\theta}_s$, under some design p, for some population parameter θ is defined as*

$$\mathrm{bias}(\hat{\theta}_s, p) = \mathrm{E}(\hat{\theta}_s, p) - \theta.$$

If the bias of $\hat{\theta}_s$ is zero for all possible populations, then $\hat{\theta}_s$ is unbiased for θ.

Unbiasedness or approximate unbiasedness are desirable properties of an estimator since this indicates that the average deviation of the estimated values from the true unknown value for the parameter of interest is zero. Therefore the distribution of an unbiased estimator is located around the unknown value of the parameter of interest. An estimator, $\hat{\theta}_s$, that varies little about the unknown value of θ is considered to be 'better' than one that varies a great deal. Hence, it is natural to desire an estimator that is unbiased for θ and of minimum variance. However unbiasedness or approximate unbiasedness does not tell us how widely dispersed the various values of the estimator are. In a discussion of Basu's (1971) paper, Hájek stated that

> the idea of unbiasedness is useful only to the extent that greatly biased estimates are poor no matter what other properties they have.

Cochran (1977, p.14) and Särndal *et al.* (1992, p.165) came to the conclusion that the effects of the bias on the coverage of confidence intervals can be ignored if the bias is relatively small compared to the variance of the estimator.

It is clear that an estimator is desirable if its sampling distribution is narrowly concentrated around the unknown value of the parameter of interest θ. Therefore a useful criterion for optimality is to choose an estimator, $\hat{\theta}_s$, which has the smallest mean square error

$$\mathrm{MSE}(\hat{\theta}_s, p) = \mathrm{E}[(\hat{\theta}_s - \theta)^2, p] = \mathrm{Var}(\hat{\theta}_s, p) + \Big(\mathrm{bias}(\hat{\theta}_s, p)\Big)^2 .$$

The mean square error measures the amount by which an estimator differs from the true value of θ. It is an average of the squared 'error' of the estimated values and incorporates both the variance and the bias of the estimator. A mean square error of zero would imply that the estimator gives exact estimates equal to θ regardless of which sample is selected under the design. Minimizing mean square error is a key criterion in choosing an estimator, and among unbiased estimators minimizing mean square error is equivalent to minimizing the variance.

In an interesting note, Padmawar (1998) showed that for the class of estimators $\alpha\hat{T}_s$, where $\alpha \in (0, 1]$ and $\hat{T}_s$ is unbiased for T_Y under some design p, the mean square error of $\alpha\hat{T}_s$ is at its minimum if we choose

$$\alpha = \frac{1}{1 + [\mathrm{cv}(\hat{T}_s, p)]^2}$$

where cv(.) denotes the coefficient of variation, i.e.

$$\text{cv}(\hat{T}_s, p) = \left(\frac{\text{Var}(\hat{T}_s, p)}{\text{E}(\hat{T}_s, p)^2} \right)^{\frac{1}{2}}.$$

Hence for any unbiased estimator for (nonzero) T_Y there will always be a biased estimator that has a smaller mean square error, provided $\text{Var}(\hat{T}_s) \neq 0$.

Another desirable property for an estimator is for it to be consistent. Formally an estimator, $\hat{\theta}_s$, is said to be consistent for θ if for any fixed $\delta > 0$,

$$\lim_{n \to \infty} \Pr(|\hat{\theta}_s - \theta| < \delta) = 1.$$

Consistency means that the probability of our estimate being within some small δ of θ can be made as close to one as we want by choosing a sufficiently large sample. We will adopt the sense of consistency in Rao (1985). The definition is as follows:

Definition 3 *An estimator, $\hat{\theta}_s$, is said to be consistent for θ if its mean square error approaches zero as the sample size increases.*

Then an estimator that is consistent will be more accurate, in terms of mean square error, as the sample size gets larger. Another type of consistency measure that may be of interest is Fisher consistency for finite populations which is a property of an estimator asserting that if the estimator were calculated using the entire population rather than a sample, the true value of the estimated parameter would be obtained.

1.3 Calibrated estimators

Godambe (1955) defined the following class of estimators for the population total T_Y that embraces many estimators used in practice including the ratio estimator.

Definition 4 *A general linear estimator $\hat{T}_s$ for the population total $T_Y = \sum_{i \in \mathcal{U}} y_i$ is of the form*

$$\hat{T}_s = \sum_{i \in s} b_{si} y_i$$

where the survey weights b_{si} for units $i \in s$ are independent of the survey variable.

Any estimator whose form is linear in the y_i's for each sample s falls into this class of estimators.

Example 1 *The ratio estimator, denoted by $\hat{T}_R$, can be written as*

$$\hat{T}_R = N\bar{X} \frac{\sum_{i \in s} y_i}{\sum_{i \in s} x_i} = \sum_{i \in s} b_{si} y_i$$

where the survey weight b_{si}, for unit $i \in s$, is equal to $N\bar{X} / \sum_{i \in s} x_i$ which depends on all units in the sample (but only through their sample sum), and here requires population level information namely the total of the x's.

For this wide class of estimators, Godambe (1955) proved his well-known non-optimality result that in the class of general linear estimators no minimum variance design-unbiased estimator exists (over all possible populations).

A special class of general linear estimators which can give zero mean square error, for some populations, are calibrated estimators. The following definition of calibrated estimators is from Sugden & Smith (2007).

Definition 5 *A general linear estimator $\hat{T}_s$ of T_Y is said to be calibrated for vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_q$ in $\mathcal{R}^N$ with respect to a set of samples $\mathcal{S}^*$ if*

$$\sum_{i \in s} b_{si} x_{ki} = \sum_{i \in \mathcal{U}} x_{ki}$$

for all $k = 1, \ldots, q$ and all $s \in \mathcal{S}^*$, where x_{ki} is the i^{th} element of the $N \times 1$ vector $\mathbf{x}_k$.

This property means that $\hat{T}_s$ is equal to T_Y for any sample in $\mathcal{S}^*$ whenever the survey vector $\mathbf{y}$ is equal to any linear combination of the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_q$.

Let $\mathbf{L}_s$ be an $N \times N$ diagonal matrix with its i^{th} diagonal element equal to 1 if unit i is selected in the sample s and 0 otherwise. A general class of estimators, is given by

$$\mathbf{1}_N^t \mathbf{A} \mathbf{L}_s \mathbf{y} + (\mathbf{1}_N^t \mathbf{X} - \mathbf{1}_N^t \mathbf{A} \mathbf{L}_s \mathbf{X}) \hat{\boldsymbol{\beta}}_s \tag{1.1}$$

where $\hat{\boldsymbol{\beta}}_s$ is given by the weighted least squares formula,

$$\hat{\boldsymbol{\beta}}_s = (\mathbf{X}^t \mathbf{W} \mathbf{L}_s \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \mathbf{L}_s \mathbf{y},$$

where $\mathbf{1}_N$ is the $N \times 1$ column vector of ones, $\mathbf{A}$ and $\mathbf{W}$ are positive definite diagonal matrices. Provided the $q \times q$ matrix $\mathbf{X}^t \mathbf{W} \mathbf{L}_s \mathbf{X}$ is nonsingular for all samples $s \in \mathcal{S}$ any estimator of the form (1.1) will be calibrated for any population vector that is a linear combination of the columns of $\mathbf{X}$.

Some special cases of (1.1) are:

1. The regression estimator

   $$\hat{T}_{REG} = \hat{\boldsymbol{\beta}}_s^t \mathbf{X}^t \mathbf{1}_N$$

   by letting $\mathbf{A} = \mathbf{0}$ and $\mathbf{W} = \text{diag}(\pi_i^{-1})$.

2. The general regression estimator, $\hat{T}_{GREG}$, by letting $\mathbf{A} = \text{diag}(\pi_i^{-1})$ and $\mathbf{W} = \text{diag}(q_i^{-1})$, where q_i is some positive arbitrary value for $i \in \mathcal{U}$.

3. When $\mathbf{A} = \mathbf{I}_N$ and $\mathbf{W} = \text{diag}(\sigma^2 v(\mathbf{x}_{(i)}^t))$, (1.1) reduces to the best linear model-unbiased predictor, $\hat{T}_{BLUP}$, under the tentative model

   $$\xi : \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

   with model expectations $\text{E}_\xi(\boldsymbol{\epsilon}) = \mathbf{0}_N$ and $\text{E}_\xi(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\boldsymbol{t}}) = \text{diag}(\sigma^2 v(\mathbf{x}_{(i)}^t))$, $\mathbf{x}_{(i)}^t$ is the i^{th} row of $\mathbf{X}$ for $i \in \mathcal{U}$ (Valliant *et al.* 2000 (2.2.1) on p.29 is this special case of uncorrelated errors).

The estimator $\hat{T}_{BLUP}$ in case 3 is considered when using the model-based approach to sampling where the values of the survey variables, $y_1, y_2, \ldots, y_N$, are realizations, of just one outcome, of random variables $Y_1, Y_2, \ldots, Y_N$ (whereas in the design-based approach the values of $y_1, y_2, \ldots, y_N$ are considered to be fixed but unknown values).

In the model-based approach to survey sampling the population model is given by the joint probability distribution or density which may depend on unknown parameters. The sampling design plays no role in the inference. This position is taken by Royall (1970) and Royall & Cumberland (1981a & b). Similar to the design-based theory, the purely model-based evaluation consists of finding the estimator to minimize the model mean square error, given a sample. However, we will only consider strategies under the design-based approach.

Hájek's (1981, p.157) definition of representative strategies with respect to $\mathbf{x}_k$, for $k = 1, \ldots, q$, is equivalent to estimators calibrated for $\mathbf{x}_k$ with respect to a sampling support given by the design.

Note that the calibration condition need not hold for any sample with probability zero and clearly no estimator is calibrated for the empty sample. We therefore give the following definition:

Definition 6 *We say that a strategy is calibrated for $\mathbf{x}_k$, for $k = 1, \ldots, q$, if its estimator is calibrated for the $\mathbf{x}_k$'s with respect to the sampling support given by the design.*

It is convenient to denote a strategy for T_Y by $stg(\hat{T}_s, p)$ where $\hat{T}_s$ is the estimator for T_Y under the design p. If $\hat{T}_s$ is a general linear estimator under the design p then $stg(\hat{T}_s, p)$ will be called a general linear strategy. We will only consider general linear strategies throughout this thesis.

Deville & Särndal (1992) defined a calibration estimator to be one with survey weights chosen to minimize some distance measure from the original sampling design weights π_i^{-1} under the constraints

$$\sum_{i \in s} b_{si} x_{ki} = \sum_{i \in \mathcal{U}} x_{ki},$$

for $k = 1, \ldots, q$. The equations are known as the calibration equations and any estimator of this type clearly falls in to the class of calibrated estimators given by definition 5. Deville & Särndal showed that $\hat{T}_{GREG}$ for a single auxiliary variable is a unique calibration estimator using the chi squared statistic

$$\mathrm{E}\left[\sum_{i \in s} \frac{(b_{si} - d_i)^2}{d_i}, p\right]$$

as a distance measure by putting $d_i = \pi_i^{-1}$ for all $i \in s$.

The idea behind calibrated strategies is a simple one. Suppose the survey variable $\mathbf{y}$ was a linear combination of the auxiliary variables $\mathbf{x}_k$, $k = 1, \ldots, q$. i.e. let $y_i = \sum_{k=1}^{q} c_k x_{ki}$ for all $i \in \mathcal{U}$ for some constant terms c_k, $k = 1, \ldots, q$. Then if we apply some strategy, $stg(\hat{T}_s, p)$, for the population total, T_Y, which is calibrated for the $\mathbf{x}_k$'s, we would always obtain an exact estimate of the population total of the y's regardless of which sample was selected. By definition 4 we have

$$\begin{aligned} \hat{T}_s = \sum_{i \in s} b_{si} y_i &= \sum_{i \in s} b_{si} \sum_{k=1}^{q} c_k x_{ki} = \sum_{k=1}^{q} c_k \sum_{i \in s} b_{si} x_{ki} \\ &= \sum_{k=1}^{q} c_k \sum_{i \in \mathcal{U}} x_{ki} = \sum_{i \in \mathcal{U}} \sum_{k=1}^{q} c_k x_{ki} = \sum_{i \in \mathcal{U}} y_i \end{aligned}$$

for all possible samples. Clearly also the mean square error of this strategy will be equal to zero. Of course in practice it is very unlikely that you would have a survey variable which is exactly proportional to a linear combination of auxiliary variables. But there are many cases where there can be a strong linear relationship between them and so applying a strategy which is calibrated would tend to have a small mean square error.

An appealing property of a calibration estimator is that it is more precise than the estimator with survey weights equal to $b_{si} = \pi_i^{-1}$ for each $i \in s$, i.e. the Horvitz-Thompson estimator (1952), provided the linear relationship between the survey and auxiliary variables is strong.

Sugden & Smith (2007) extended Godambe's nonexistence theorem to classes of linear calibrated strategies which implies that no unique minimum mean square error strategy exists under the design-based approach. So the choice of strategies under the criterion of minimum mean square error is also not clear.

1.4 Poststratification

Poststratification is a procedure that partitions the population into H, say, mutually exclusive and exhaustive subgroups, called domains, after the sample has been selected. This partition is done with respect to some categorical variable which is defined for each unit in the population, for example age, sex, educational level, etc. The observed sample is also partitioned according to the same categorical variable used to partition the population. Broadly defined by Smith (1991),

> poststratification could refer to any method of data analysis which involves forming units into homogeneous groups after observations of the sample.

Let the H population domains be denoted by $\mathcal{U}_1, \ldots, \mathcal{U}_H$ and let their sizes be $N_1, \ldots, N_H$ respectively, so that $\mathcal{U} = \bigcup_{h=1}^{H} \mathcal{U}_h$ and $N = \sum_{h=1}^{H} N_h$. Similarly the sample strata are denoted by $s_1, \ldots, s_H$ and their sizes are $n_1, \ldots, n_H$ respectively where $s_h = s \cap \mathcal{U}_h$ for $h = 1, \ldots, H$ and $\sum_{h=1}^{H} n_h = n$. The observed sampling configuration of the sampled strata sizes will be denoted by an $1 \times H$ row vector

$$\mathbf{n} = (n_1, \dots, n_H).$$

Note that the sample strata sizes are random at the design stage.

Poststratification requires information on the poststrata sizes and information for classifying the sampled units into those poststrata. The sample is cross-classified at the estimation stage, not the design stage. This information is used in choosing survey weights for units within poststrata which form an estimator that reduces sampling variance and bias through the influence of the homogeneity of the units within the poststrata. Here are some situations when poststratification can be employed:

- When the stratum membership for some stratifying variable is unknown for every unit in the population at the design stage. This information could be unavailable because it may be too difficult or expensive to use at the time of sampling.
- When the stratifying variable is known for each unit in the population, but not applied at the design stage. There may have been several variables to base the design on and the sampler chose some other one instead, or it could have just been overlooked, or there could be many other reasons why it was not used.

Kish (1965, p.91) describes poststratification as an 'adjustment' or 'correction' of the sample mean. If an inappropriate design was used at the design stage, we can still capture the effects of the population structure to recover the loss of efficiency by using a poststratified estimator at the estimation stage. Holt & Smith (1979) viewed poststratification as a robust technique that offers protection against unfavourable or extreme sampling configurations. They also stated that

> it is the structure of the population, rather than the sampling design, which an estimator should reflect.

Note that the use of calibrated estimators also reflects this.

1.4.1 Conditional and unconditional inferences

After the sample has been selected and the appropriate estimator has been chosen, inferences about the population parameters can be made with respect to either the unconditional design, that was used to obtain the sample, or (if possible) the design conditioned on only those samples with the same poststrata sample sizes as the observed sampling configuration $\mathbf{n}$. The observed sampling allocation $\mathbf{n}$ is considered to be an ancillary statistic if its distribution is known and independent of the parameter of interest, see Cox & Hinkley (1974, p.31). This is the case if the sampling design is completely known and the stratum membership is known for each unit in the population.

It has been argued by Durbin (1969) that inferences should not be drawn from irrelevant chance events. He says

> It seems self-evident that one should use the information available on sample size in the interpretation of the results. To average over variations in sample size which might have occurred but did not occur, when in fact the sample size is exactly known, seems quite wrong from the standpoint of the analysis of the data actually observed.

Holt & Smith (1979) strongly argued for conditional inference. They pointed out that the coverage of confidence intervals based on the conditional variance is more accurate with respect to those samples with the same sample size configuration as the observed one, than the coverage of confidence intervals based on the unconditional variance.

The coverage of confidence intervals given by the unconditional approach is correct when averaged over all possible samples that could have been selected, but this is also true for the conditional approach. As mentioned by Jagers *et al.* (1985),

> The main argument in favour of the unconditional approach is simplicity: the quality of a whole procedure is described by a single number.

That number being the unconditional variance or the unconditional mean square error. The general view is to choose a strategy based on the unconditional design at the design stage and make inferences with respect to the conditional design after the sample has been selected at the estimation stage.

The use of the conditional approach for small area estimation, or domains, has been studied by Rao (1985) and Särndal & Hidiroglou (1989) under simple random sampling. Consiglio *et al.* (2003) consider the conditional approach in small area estimation under two-stage sampling design with stratification of the primary sampling units. Rao (1985) and Hidiroglou & Srinath (1981) use the conditional approach to tackle the problem of 'outliers' in the case where there are a few units in a population with large or extreme values for some survey variable. For approaches to domain or small area estimation based on models see Rao & Ghosh (1994).

Conditional analysis has also been studied in areas outside poststratification. For example Robinson (1987) proposed an asymptotic conditional approach of inference for the ratio estimator under simple random sampling. Robinson's approach was applied to poststratified estimators by Casady & Valliant (1993). Montanari (2000) developed an asymptotic conditional framework, conditioning on the auxiliary sample means, for conditional analysis of several types of regression estimators.

1.4.2 Conditional and unconditional comparisons

Two estimators, under simple random sampling, that have been compared with one another in a poststratified setting, see Holt & Smith (1979), Sugden & Smith (2006), Smith (1991), Särndal *et al.* (1992), Jagers *et al.* (1985), Gelman & Carlin (2001) and others, are the stratified estimator

$$\hat{T}_{st} = \sum_{h=1}^{H} I_h N_h \bar{y}_h$$

where $\bar{y}_h$ is the sample poststrata mean of poststratum h, $h = 1, \ldots, H$, and I_h is equal to 1 if $n_h > 0$ or zero otherwise, and the expansion estimator

$$\hat{T}_0 = \frac{N}{n} \sum_{h=1}^{H} I_h n_h \bar{y}_h = N\bar{y}.$$

As noted by Bethlehem & Keller (1987), $\hat{T}_{st}$ is a special case of $\hat{T}_{GREG}$ where the auxiliary variables are stratum indicators.

The unconditional variance of $\hat{T}_{st}$ is usually given by an approximation formula, see Kish (1965, p.90), Cochran (1977, p.135), Hansen *et al.* (1953, p.232), Holt & Smith (1979), Särndal *et al.* (1992, p.266-267) and Thompson (2002, p.124). From these formulae it has been established that this variance is close to the variance of the estimator under stratified random sampling with proportional allocation, but never less than this. Sugden & Smith (2006) studied the exact unconditional mean square error of $\hat{T}_{st}$ and compared it with an unconditionally unbiased estimator under simple random sampling.

Unconditional comparisons of these estimators have been done by Särndal *et al.* (1992, p.268) using an approximation formula for the unconditional variance of $\hat{T}_{st}$

and the exact unconditional variance of $\hat{T}_0$. But for small population sizes Sugden & Smith (2006) gave examples showing that this result can be misleading. Conditional comparisons have been done via empirical studies by Holt & Smith (1979) and Jagers *et al.* (1985). The conclusion from both papers was that $\hat{T}_{st}$ performs better than $\hat{T}_0$ overall; however it is not uniformly better over every possible population.

There hasn't yet been a method for comparing these 'simple' strategies without the use of approximation formulae and empirical studies, both conditionally and unconditionally, except for Sugden & Smith (2006) under simple random sampling.

1.4.3 When $n_h = 0$ for some poststrata

One major problem we could face when poststratifying is that it may be possible to draw a sample which does not contain units from every poststratum since the achieved sampling configuration $\mathbf{n}$ is random. This can result in extreme estimation bias both conditionally and unconditionally. One way to overcome this difficulty is to merge or collapse similar poststrata together to ensure that the sample contains units from every poststratum. Fuller (1966) suggested a way of constructing an estimator obtained by collapsing two poststrata together when one of the two was empty. Little (1993) cover methods of collapsing using the model-based approach. Bethlehem & Keller (1987) gave an alternative approach to collapsing when there are several criteria for poststratification which involves removing higher order interactions between the poststratifying variables.

Doss *et al.* (1979), Rao (1985) and Tillé (1998) considered the problem of empty poststrata samples and studied the properties of a number of estimators under simple random sampling. One of these estimators was shown, by Sugden & Smith (2006), to represent a form of collapsing and falls into the class considered by Fuller (1966) when $H = 2$. In chapter 5 (section 5.5) we will generalize some of these estimators for any general unequal probability designs.

1.4.4 General designs

One of the main problems with an unequal probability design, where the π_i's are unequal, in a poststratification setting was pointed out by Smith (1991). Under an unequal probability sampling scheme the design conditional on $\mathbf{n}$ depends on knowing the poststratum membership for all units in the population, unlike the simple random sampling design which depends only on the poststratum population and sample sizes. Smith also gave a counterexample demonstrating that for an unequal probability design conditional on $\mathbf{n}$, the inclusion indicator variables of units from different poststrata are not conditionally independent (of one another). This would imply that the formula of the conditional variance would not be equal to the sum of poststratum variances which could make it difficult to estimate the variance. Rao (1985) concluded that under a complex sampling design it is difficult to investigate the conditional properties of estimators and the choice of statistic on which to condition might not be obvious.

Brewer (2002, p.33) made the following statement:

> When analysing poststrata using design-based inference, it is both possible and desirable to ignore the dependence of the sampling configuration within any poststratum on the outcomes of the sampling in the other strata. This is known as 'making conditional inferences conditional on the numbers of units selected within each poststratum'.

However conditional inferences under poststratification is still permitted even if there is dependence between the selection of units from different poststrata.

Särndal *et al.* (1992, p264-265) considered an estimator that can be used for poststratification under a general design but the unconditional properties for this estimator were given for the simple random sample design and not the general design. Also the conditional properties of the estimator were not studied.

There is very little work done on the design-based theory of poststratification under unequal probability sampling.

1.5 Mean square error matrix

It can be shown, see Hájek (1981, p.157), Gabler (1990, p.109), Cheng & Li (1983) and Sugden & Smith (2007), that the mean square error of any general linear strategy, $stg(\hat{T}_s, p)$, can be written as a quadratic form in $\mathbf{y}$. i.e. we can write

$$\text{MSE}(\hat{T}_s, p) = \mathbf{y}^t \mathbf{M}(\hat{T}_s, p)\mathbf{y} \tag{1.2}$$

where the $\mathbf{M}(\hat{T}_s, p)$ is an $N \times N$ real symmetric matrix which is independent of the y's. Bethlehem & Keller (1987) considered the variance-covariance matrix of an approximation formula for the variance of the regression estimator and they mentioned that this matrix can provide appropriate information for further analysis since it takes into account the sampling design. For the matrix given by the exact mean square error, Sugden & Smith (2007) stated that

> A remarkable consequence of our assumptions that the values of the auxiliary variables $\mathbf{X}$ are completely known for all units and that the sampling scheme is also completely known, is that the $N \times N$ matrices of the quadratic forms in "(1.2) here" depend only on $\mathbf{X}$ and are therefore known

for all possible $\mathbf{y}$. In particular the eigenvalues and associated eigenspaces, spanned by linearly independent normalized eigenvectors, are also known for each positive semidefinite matrix.

They also pointed out that if the strategy is calibrated for q vectors $\mathbf{x}_1, \ldots, \mathbf{x}_q$, then these vectors will also be eigenvectors of the matrix given by the mean square error of the strategy with corresponding eigenvalues equal to zero. i.e. $\mathbf{x}_1, \ldots, \mathbf{x}_q$ will belong in the nullspace of $\mathbf{M}(\hat{T}_s, p)$.

Cheng & Li (1983) used the eigenvalues of the mean square error matrices of certain ratio type strategies to obtain the minimax strategy between them. Gabler (1990, p.108-110) also obtained a modified minimax strategy for a certain class of strategies using the mean square error matrix. Hájek (1981, p.161) gave a result based on these matrices that can indicate how well an approximation formula for the mean square error of some strategy can perform. Berger (2005) also used this method to examine a variance approximation formula.

Gabler (1984) also used the eigenvalues of the mean square error matrices to give sufficient conditions for the superiority of sampling without replacement over sampling with replacement under unequal probability sampling.

Estimators for the mean square error can also be written as quadratic forms in the sampled values of the y's. Padmawar (1998) adopted this approach and by using the eigenvalues of this matrix he gave conditions for some mean square error estimators to be nonnegative definite and unbiased for a certain class of estimators under a cluster type sampling design. We will also consider the matrices given by variance estimators and attempt to construct nonnegative definite unbiased variance estimators.

Many approximation methods for analysing strategies are not valid for small sample sizes. It is therefore of interest to develop an exact method for comparison of strategies. In general, the mean square error matrix of any general linear strategy depends on the design and auxiliary information which is assumed to be known. It can therefore be considered to be a useful 'tool' to form a basis for further analysis of strategies. Our aim is to use the mean square error matrices to give some exact methods of analysing and comparing different strategies.

In poststratification under an unequal probability sampling design, provided the poststrata members are known for all units in the population, we can make conditional inferences conditional on $\mathbf{n}$. We intend to use the mean square error matrices of post-stratified strategies to compare poststratified estimators under unequal probability sampling, both conditionally and unconditionally.

Chapter 2

Conditional Designs under Poststratification

In this chapter we will consider designs conditional on the observed sample size configuration $\mathbf{n}$ under poststratification when using a general unequal probability design. We will give properties of such designs and the requirements that make it possible to make inferences conditional on $\mathbf{n}$.

We will also propose a subclass of general linear estimators which can be considered in a poststratification setting and give its properties.

A list of some desirable properties for a satisfactory unequal probability sampling scheme will be given and we will describe three well known unequal probability sampling schemes; two of which will be implemented in examples throughout this thesis.

In the last section we study some aspects of the conditional unequal probability design such as independence between selection of units from different poststrata and calibrated estimators under the conditional design.

2.1 General sampling designs without replacement

In this section we will give some standard results for any general sampling design with fixed sample size. These results will be used in subsequent sections to prove further results.

We will first define a random variable which is useful in the proof of theorems in this chapter.

Definition 7 *The sample indicator variable for unit i in the population, denoted by t_i, is a random variable which takes the value 1 if the unit labelled i is included in the sample or zero otherwise. i.e.*

$$t_i = \begin{cases} 1 & \textit{if} \quad i \in s \\ 0 & \quad \textit{else.} \end{cases}$$

Note that t_i is a function that depends on the sample s.

The statistical properties of the sample indicator variable depends on the sampling design under consideration and are given in the following lemma.

Lemma 1 *For any units $i, j \in \mathcal{U}$, with $i \neq j$, and design p we have*

a) $\mathrm{E}(t_i, p) = \pi_i$

b) $\mathrm{Var}(t_i, p) = \pi_i(1 - \pi_i)$

c) $\mathrm{Cov}(t_i, t_j, p) = \pi_{ij} - \pi_i\pi_j$

Proof The proof of this lemma follows from the fact that the sample indicator variable t_i has a Bernoulli distribution with probability of success equal to π_i. $\square$

The following lemma, from Yates & Grundy (1953), gives some properties for any general probability sampling design for fixed sample size n.

Lemma 2 *For any probability sampling design with fixed sample size n, we have*

a) $\sum\limits_{i\in\mathcal{U}}\pi_i = n$

b) $\sum\limits_{\substack{j\in\mathcal{U}\\ j\neq i}}\pi_{ij} = (n-1)\pi_i$

c) $\sum\limits_{i\in\mathcal{U}}\sum\limits_{\substack{j\in\mathcal{U}\\ j\neq i}}\pi_{ij} = n(n-1)$

Proof Note that a fixed sample size n can be written in terms of the sampling indicator variable $n = \sum_{i\in\mathcal{U}} t_i$.

a) Since the sample size n is fixed, we have

$$n = \mathrm{E}(n,p) = \sum_{i\in\mathcal{U}}\mathrm{E}(t_i,p) = \sum_{i\in\mathcal{U}}\pi_i.$$

b) For some $i \in \mathcal{U}$, consider the summation

$$\sum_{\substack{j\in\mathcal{U}\\ j\neq i}} t_j.$$

If $i \in s$, then this sum takes the value $n-1$, and if $i \notin s$, it takes the value n, thus

$$\begin{aligned}\sum_{\substack{j\in\mathcal{U}\\ j\neq i}}\pi_{ij} = \mathrm{E}\left[t_i\sum_{\substack{j\in\mathcal{U}\\ j\neq i}}t_j, p\right] &= 1\times(n-1)\times p(s:i\in s) + 0\times n\times p(s:i\notin s)\\ &= (n-1)\pi_i.\end{aligned}$$

The proof of part (c) follows directly from part (a) and (b). □

A feature of a general unequal probability sampling design is that the N units in the population can have different inclusion probabilities which can be taken into account to obtain the properties of estimators as we shall see in sections 2.4 and 2.6.

2.2 Properties of conditional designs

Tillé (2006, p.20) gave the following definition of a conditional sampling design with respect to a support.

Definition 8 *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two supports such that $\mathcal{S}_2 \subset \mathcal{S}_1$ and $p_1(.)$ a sampling design on $\mathcal{S}_1$. The conditional design of $\mathcal{S}_1$ with respect to $\mathcal{S}_2$, denote by $p_1(s \mid \mathcal{S}_2)$, is given by*

$$p_1(s \mid \mathcal{S}_2) = \frac{p_1(s)}{\sum_{s \in \mathcal{S}_2} p_1(s)} \quad \textit{for all} \quad s \in \mathcal{S}_2.$$

Conditional expectations are defined with respect to the conditional design. Typically, the support $\mathcal{S}_2$ from definition 8 would consist of all samples in the support $\mathcal{S}_1$ that respects some fixed values of an ancillary statistic.

Under a poststratification the support for which the conditional design is based on is the set of all samples which have the same sample size configuration or sampling allocation as the realized one $\mathbf{n}$. We will denote this set by $\mathcal{S}_{\mathbf{n}}$, i.e.

$$\mathcal{S}_{\mathbf{n}} = \{s \in \mathcal{S} : \mathbf{n}(s) = \mathbf{n}\}$$

where $\mathbf{n}(s)$ denotes the sample size configuration of the poststrata samples for sample s.

Definition 9 *Let $\mathcal{S}$ be the sampling support of some design p. Under a poststratification with H poststrata the conditional design on the achieved sampling configuration $\mathbf{n} = (n_1, \ldots, n_H)$, denoted by $p^*(s)$, is given by*

$$p^*(s) = \begin{cases} \frac{p(s)}{\sum_{s \in \mathcal{S}_\mathbf{n}} p(s)} & \textit{for} \quad s \in \mathcal{S}_\mathbf{n} \\ 0 & \textit{else.} \end{cases}$$

Throughout the rest of this thesis we will refer to $p^*(s)$ as the conditional design on $\mathbf{n}$ whose sample poststrata size n_h is fixed for all $h = 1, \ldots, H$ such that $\sum_{h=1}^{H} n_h = n$.

Example 2 *If the unconditional design is simple random sampling (SRS) without replacement where*

$$p(s) = \binom{N}{n}^{-1} \quad \textit{for all} \quad s \in \mathcal{S},$$

then the conditional design on $\mathbf{n}$ would be

$$\begin{aligned} p^*(s) &= \frac{p(s)}{\sum_{s \in \mathcal{S}_\mathbf{n}} p(s)} = \frac{\binom{N}{n}^{-1}}{\sum_{s \in \mathcal{S}_\mathbf{n}} \binom{N}{n}^{-1}} \\ &= \frac{1}{\prod_{h=1}^{H} \binom{N_h}{n_h}} \end{aligned}$$

for all $s \in \mathcal{S}_\mathbf{n}$. This design is a stratified random sampling design (StRS) for the sampling allocation $\mathbf{n}$.

Definition 10 *The r^{th} order conditional inclusion probabilities from the design conditional on $\mathbf{n}$ is defined as*

$$\pi^*_{i_1 i_2 \ldots i_r} = \sum_{s \ni i_1, i_2, \ldots, i_r} p^*(s) \quad \textit{for} \quad 1 \leq r \leq n$$

for r distinct unit(s) $i_1, i_2, \ldots, i_r \in \mathcal{U}$.

Note that for a general unequal probability design the conditional selection probabilities, in definition 9, depend on knowing the design probabilities for all samples with the same sample size configuration $\mathbf{n}$. In the case of simple random sampling, in example 2, we only needed to know the poststrata sizes and $\mathbf{n}$ to calculate $p^*(s)$ and so the conditional inclusion probabilities for all units in the population are easily calculated. But for an unequal probability design we need to know the poststratum membership for all units in the population in order to calculate $p^*(s)$ and hence the conditional inclusion probabilities.

Most estimators for the conditional variance of a strategy depend on the first or second order conditional inclusion probabilities for units selected in the sample. If we do not know the poststratum membership, under an unequal probability design, we won't be able to calculate the conditional design and hence we cannot calculate the conditional inclusion probabilities. This means that conditional inferences under an unequal probability sampling design may not be possible.

We will assume that the poststratum membership is known for all units in the population so that we can make our inferences conditional on $\mathbf{n}$.

Lemma 3 *Under the conditional design for any $i, j \in \mathcal{U}$, with $i \neq j$, and design p we have*

a) $\mathrm{E}(t_i, p \mid \mathcal{S}_\mathbf{n}) = \pi_i^*$

b) $\mathrm{Var}(t_i, p \mid \mathcal{S}_\mathbf{n}) = \pi_i^*(1 - \pi_i^*)$

c) $\mathrm{Cov}(t_i, t_j, p \mid \mathcal{S}_\mathbf{n}) = \pi_{ij}^* - \pi_i^* \pi_j^*$

Proof The proof of this lemma parallels the proof of lemma 1 by replacing π_i by π_i^* and π_{ij} by π_{ij}^* since here we take expectations with respect to the design conditional on $\mathbf{n}$. □

In the following lemma we give properties of a conditional design on $\mathbf{n}$ which are similar to those of the unconditional design given in lemma 2.

Lemma 4 *For any sampling design, p, conditional on $\mathbf{n} = (n_1, \ldots, n_H)$ where n_h's are fixed for all $h = 1, \ldots, H$, we have the following properties:*

a) $\sum_{i \in \mathcal{U}_h} \pi_i^* = n_h$

b) $\sum_{\substack{j \in \mathcal{U}_h \\ j \neq i}} \pi_{ij}^* = (n_h - 1)\pi_i^* \quad \textit{for} \quad i \in \mathcal{U}_h$

c) $\sum_{i \in \mathcal{U}_h} \pi_{ij}^* = n_h \pi_j^* \quad \textit{for} \quad j \in \mathcal{U}_g \quad \textit{with} \quad g \neq h$

d) $\sum_{i \in \mathcal{U}_h} \sum_{\substack{j \in \mathcal{U}_h \\ j \neq i}} \pi_{ij}^* = n_h(n_h - 1)$

e) $\sum_{i \in \mathcal{U}_h} \sum_{\substack{j \in \mathcal{U}_g \\ g \neq h}} \pi_{ij}^* = n_h n_g$

Proof For any design conditional on $\mathbf{n}$ the sample poststratum size n_h for $h = 1, \ldots, H$ can be written in terms of the sampling indicator variable $n_h = \sum_{i \in \mathcal{U}_h} t_i$.

a) If n_h is fixed, then under the conditional design we have

$$n_h = \mathrm{E}(n_h, p \mid \mathcal{S}_{\mathbf{n}}) = \sum_{i \in \mathcal{U}_h} \mathrm{E}(t_i, p \mid \mathcal{S}_{\mathbf{n}}) = \sum_{i \in \mathcal{U}_h} \pi_i^*$$

b) For some unit $i \in \mathcal{U}_h$ consider the summation

$$\sum_{\substack{j \in \mathcal{U}_h \\ j \neq i}} t_j.$$

If $i \in s_h$, where $s_h = s \cap \mathcal{U}_h$, then this takes the value $n_h - 1$ and if $i \notin s_h$ then it takes the value n_h. Thus under the conditional design for $i \in \mathcal{U}_h$ we have

$$\begin{aligned} \sum_{\substack{j \in \mathcal{U}_h \\ j \neq i}} \pi^*_{ij} &= \mathrm{E}\left(t_i \sum_{\substack{j \in \mathcal{U}_h \\ j \neq i}} t_j, p \mid \mathcal{S}_{\mathbf{n}} \right) \\ &= 1 \times (n_h - 1) \times p^*(s : i \in s_h) + 0 \times n_h \times p^*(s : i \notin s_h) \\ &= (n_h - 1)\pi^*_i. \end{aligned}$$

c) For $i \in \mathcal{U}_h$ and $j \in \mathcal{U}_g$ with $h \neq g$ we have

$$\sum_{j \in \mathcal{U}_g} \pi^*_{ij} = \mathrm{E}\left(\sum_{j \in \mathcal{U}_g} t_i t_j, p \mid \mathcal{S}_{\mathbf{n}} \right) = \mathrm{E}\left(t_i, p \mid \mathcal{S}_{\mathbf{n}}\right) n_g = \pi^*_i n_g.$$

The proof of part (d) and (e) follow from part (b) and (c) respectively by summing over the π^*_i for all $i \in \mathcal{U}_h$ using part (a). □

These results are valid for any $\mathbf{n}$ with fixed poststrata sample sizes. If $n_h = 0$ for some h then part (a) of lemma 4 implies $\pi^*_i = 0$ for all $i \in \mathcal{U}_h$. If $n_h = 1$ for some h then by part (b) all conditional joint inclusion probabilities for pairs of units in $\mathcal{U}_h$ are equal to zero.

We can also consider a design which is conditioned on all samples in $\mathcal{S}$ such that $n_h > 0$ for all $h = 1, \ldots, H$. It may be more appropriate to compare strategies under this design at the design stage, rather than the unconditional design, if the probability of obtaining a sample that doesn't contain units from all poststrata is small enough to neglect. This design (conditioned on s such that $n_h > 0$ for all poststrata) is under a weaker condition than the conditional design $p^*(s)$ and so we will call it the weak conditional design.

Definition 11 *Let* $\mathcal{S}_w = \{s \in \mathcal{S} : n_h(s) > 0, \quad \forall h = 1, \ldots, H\}$ *where* $n_h(s)$ *denotes the poststratum sample size of poststratum* h *for sample* s*. Then the weak conditional design, denoted by* $p'(s)$*, is given by*

$$p'(s) = \begin{cases} \frac{p(s)}{\sum_{s \in \mathcal{S}_w} p(s)} & \text{for} \quad s \in \mathcal{S}_w \\ 0 & \text{else.} \end{cases}$$

Note that if a strategy $stg(\hat{T}_s, p)$ is conditionally unbiased for T_Y (conditional on $\mathbf{n}$ with fixed positive n_h's) then it is also unbiased for T_Y under the weak conditional design which follows from properties of conditional expectations.

Definition 12 *The* r^{th} *order weak conditional inclusion probabilities is defined as*

$$\pi'_{i_1 i_2 \ldots i_r} = \sum_{s \ni i_1, i_2, \ldots, i_r} p'(s) \quad \text{for} \quad 1 \leq r \leq n$$

for r *distinct unit(s)* $i_1,\ i_2,\ \ldots,\ i_r \in \mathcal{U}$.

Lemma 5 *Under the weak conditional design for any* $i, j \in \mathcal{U}$*, with* $i \neq j$*, and design* p *we have*

a) $\mathrm{E}(t_i, p \mid \mathcal{S}_w) = \pi'_i$

b) $\mathrm{Var}(t_i, p \mid \mathcal{S}_w) = \pi'_i(1 - \pi'_i)$

c) $\mathrm{Cov}(t_i, t_j, p \mid \mathcal{S}_w) = \pi'_{ij} - \pi'_i \pi'_j$

Proof Again the proof of this parallels the proof of lemma 1 by taking expectations with respect to the weak conditional design. □

It is clear that the weak conditional inclusion probabilities satisfies the properties of lemma 2 but it is not true that the properties of lemma 4 will be satisfied, since under the weak conditional design the poststrata sample sizes are still random.

2.3 Separate general linear estimators

In this section we examine a special class of general linear estimators called separate general linear estimators.

Definition 13 *A separate general linear estimator* $\hat{T}_s$ *for the population total* T_Y *is of the form*

$$\hat{T}_s = \sum_{h=1}^{H} I_h \hat{T}_{s_h} \quad \textit{with} \quad \hat{T}_{s_h} = \sum_{i \in s_h} b_{s_h i} y_i \tag{2.1}$$

where I_h *is equal to one if* $n_h > 0$ *or zero otherwise and the survey weight* $b_{s_h i}$*, for unit* $i \in s_h$*, can depend on the sampled units but only through the auxiliary values attached to units in* s_h *and not in* s_g $(g \neq h)$ *for* $h, g = 1, \ldots, H$*.*

The survey weights, $b_{s_h i}$, can of course depend on population and domain measures.

Remark 1 $\hat{T}_{s_h}$ *defined in (2.1) is a general linear estimator of the domain total* $T_{Y_h} = \sum_{i \in \mathcal{U}_h} y_i$*, as* $b_{si} = b_{s_h i}$ *where* $s_h = s \cap \mathcal{U}_h$ *for* $i \in \mathcal{U}_h$ $(h = 1, \ldots, H)$*.*

Example 3 *The stratified estimator can be written as*

$$\hat{T}_{st} = \sum_{h=1}^{H} I_h N_h \bar{y}_h = \sum_{h=1}^{H} I_h \frac{N_h}{n_h} \sum_{i \in s_h} y_i = \sum_{h=1}^{H} I_h \sum_{i \in s_h} b_{s_h i} y_i$$

where the survey weight $b_{s_h i}$ *for unit* $i \in s_h$ *is equal to* N_h/n_h *which does not depend on auxiliary values associated with the sampled units in* s_g $(h \neq g)$*. Therefore it is a separate general linear estimator.*

Example 4 *The separate ratio estimator* $\hat{T}_{Rs}$*, which is of the form*

$$\hat{T}_{Rs} = \sum_{h=1}^{H} I_h N_h \bar{X}_h \frac{\sum_{i \in s_h} y_i}{\sum_{i \in s_h} x_i},$$

is a separate general linear estimator. Its survey weight,

$$b_{s_h i} = \frac{N_h \bar{X}_h}{\sum_{i \in s_h} x_i},$$

for unit $i \in s_h$ depends on the total of the x's in domain h and the auxiliary values of the sampled units in s_h, but not in s_g for $g \neq h$ (h, $g = 1, \ldots, H$).

Example 5 *The combined ratio estimator $\hat{T}_{Rc}$, which is of the form*

$$\hat{T}_{Rc} = N\bar{X} \frac{\sum_{h=1}^{H} I_h N_h \bar{y}_h}{\sum_{h=1}^{H} I_h N_h \bar{x}_h}$$

is not a separate general linear estimator since its survey weight,

$$b_{s_h i} = \frac{N\bar{X}}{n_h} \frac{N_h}{\sum_{h=1}^{H} I_h N_h \bar{x}_h},$$

for unit $i \in s_h$ does depend on the sampled units through the auxiliary values attached to units in s_g ($g \neq h$) for h, $g = 1, \ldots, H$.

It is clear that separate general linear estimators are general linear estimators. In the next section we cover some properties of general linear strategies.

2.3.1 Properties of general linear strategies

Sugden & Smith (2002) gave the following form of the variance and mean square error of a general linear strategy $stg(\hat{T}_s, p)$:

$$\text{Var}(\hat{T}_s, p) = \sum_{i \in \mathcal{U}} C_i y_i^2 + \sum_{i \in \mathcal{U}} \sum_{\substack{j \in \mathcal{U} \\ j \neq i}} D_{ij} y_i y_j \tag{2.2}$$

and

$$\text{MSE}(\hat{T}_s, p) = \sum_{i \in \mathcal{U}} [C_i + (B_i - 1)^2] y_i^2 + \sum_{i \in \mathcal{U}} \sum_{\substack{j \in \mathcal{U} \\ j \neq i}} [D_{ij} + (B_i - 1)(B_j - 1)] y_i y_j \tag{2.3}$$

where

$$B_i = \sum_{s \ni i} p(s) b_{si} \quad \text{for} \quad i \in \mathcal{U}, \tag{2.4}$$

$$C_i = \sum_{s \ni i} p(s) b_{si}^2 - B_i^2 \quad \text{for} \quad i \in \mathcal{U},$$

and

$$D_{ij} = \sum_{s \ni i,j} p(s) b_{si} b_{sj} - B_i B_j \quad \text{for} \quad i, j \in \mathcal{U} \quad (i \neq j).$$

By observing that the expected value of a general linear strategy can be written as

$$\mathrm{E}(\hat{T}_s, p) = \sum_{s \in \mathcal{S}} p(s) \left(\sum_{i \in s} b_{si} y_i \right) = \sum_{i \in \mathcal{U}} \left(\sum_{s \ni i} p(s) b_{si} \right) y_i = \sum_{i \in \mathcal{U}} B_i y_i$$

they evaluated the bias of a strategy as

$$\mathrm{bias}(\hat{T}_s, p) = \sum_{i \in \mathcal{U}} (B_i - 1) y_i \tag{2.5}$$

which led them to a result given in the following lemma.

Lemma 6 *A necessary and sufficient condition for a general linear strategy $stg(\hat{T}_s, p)$ to be unbiased for the population total $T_Y = \sum_{i \in \mathcal{U}} y_i$ over all possible $\mathbf{y} \in \mathcal{R}^N$ is that $B_i = 1$ for all $i \in \mathcal{U}$.*

Proof It is easily seen by substituting $B_i = 1$ for all $i \in \mathcal{U}$ into (2.5) that the bias of $stg(\hat{T}_s, p)$ is always zero for any population vector $\mathbf{y} \in \mathcal{R}^N$. The converse is also true since (2.5) is only zero for all $\mathbf{y}$ when the coefficient of each y_i is also zero. □

From lemma 6 we obtain the following theorem for unbiased strategies which are calibrated for constant population vectors.

Theorem 1 *Suppose a general linear strategy, $stg(\hat{T}_s, p)$, is calibrated for constant population vectors and is also such that B_i is a constant value for all $i \in \mathcal{U}$. Then $stg(\hat{T}_s, p)$ is also unbiased for T_Y for every $\mathbf{y} \in \mathcal{R}^N$.*

Proof If $B_i = B$, say, for all $i \in \mathcal{U}$ then

$$\text{bias}(\hat{T}_s, p) = \sum_{i \in \mathcal{U}} (B_i - 1) y_i = (B - 1) T_Y.$$

Since the strategy is calibrated for constant vectors its bias must be equal to zero when $y_i = c$, say, for all $i \in \mathcal{U}$. Then it follows that

$$\text{bias}(\hat{T}_s, p) = (B - 1) T_Y = (B - 1) N c = 0 \quad \Leftrightarrow \quad B = 1,$$

which implies that $B_i = 1$ for all $i \in \mathcal{U}$ and by lemma 6 the strategy $stg(\hat{T}_s, p)$ is unbiased for T_Y. □

Another measure of accuracy for a strategy is given by its absolute relative bias.

Definition 14 *Let $\hat{\theta}_s$ be an estimator for some population parameter θ under some design p. The absolute relative bias of a strategy, denoted by $\text{ARB}(\hat{\theta}_s, p)$, is defined as*

$$\text{ARB}(\hat{\theta}_s, p) = \left| \frac{\text{bias}(\hat{\theta}_s, p)}{\theta} \right|.$$

The absolute relative bias gives a measure that indicates the magnitude of the strategy's bias.

Lemma 7 *Let $\mathbf{y} \in \mathcal{R}^N$ be such that $y_i \geq 0$ for all $i \in \mathcal{U}$. Then an upper bound on the absolute relative bias for a general linear strategy for the population total T_Y is*

$$\text{ARB}(\hat{T}_s, p) \leq \max_{i \in \mathcal{U}} |B_i - 1|.$$

Proof By (2.5)

$$\text{bias}(\hat{T}_s, p) = \sum_{i \in \mathcal{U}} (B_i - 1) y_i.$$

From this we see that if $y_i \geq 0$ for all $i \in \mathcal{U}$ the absolute bias will be

$$\left|\sum_{i\in\mathcal{U}}(B_i - 1)y_i\right| \leq \max_{i\in\mathcal{U}}|B_i - 1|\sum_{i\in\mathcal{U}} y_i = \max_{i\in\mathcal{U}}|B_i - 1|T_Y.$$

Divide through by T_Y completes the proof. $\square$

This lemma is a generalization of Aires & Rosén (2005) who gave this result for the Horvitz-Thompson estimator under pareto πPS sampling, which is not an unbiased strategy for T_Y as this design is not an exact probability proportional to size sampling scheme, see p.61 in section 2.5.

Sugden & Smith (2002) proposed two classes of unbiased estimators for T_Y under the design p which are within the class of general linear estimators. The first is a 'shift' type estimator

$$\hat{T}_s(1) = \hat{T}_s - \sum_{i\in s}\frac{(B_i - 1)}{\pi_i},$$

and the other is a ratio correction for the bias

$$\hat{T}_s(2) = \sum_{i\in s}\frac{b_{si}y_i}{B_i}$$

where B_i is given by the survey weights of the estimator $\hat{T}_s$ and the design p.

The strategy $stg(\hat{T}_s(2), p)$ is calibrated for $(B_1 x_{k1}, \ldots, B_N x_{kN})^t$ provided $stg(\hat{T}_s, p)$ is calibrated for $\mathbf{x}_k$, $k = 1, \ldots, q$.

As pointed out by Rao (2002), a drawback of these estimators is that, in general, they are not calibrated for the auxiliary variables $\mathbf{x}_k$, $k = 1, \ldots, q$, and therefore might not be efficient relative to the linear relationship between $\mathbf{y}$ and the $\mathbf{x}_k$'s. But, as demonstrated by Sugden & Smith (2002), $stg(\hat{T}_s(2), p)$ can perform well if the points $(y_i, B_i x_{ki})$ for $i \in \mathcal{U}$, lie much closer to a regression line through the origin than the points (y_i, x_{ki}), $i \in \mathcal{U}$.

2.3.2 Induced design on strata

The variance and mean square error of a separate general linear strategy can be written in a more structured form than (2.2) and (2.3) which will be demonstrated. First we need to define the induced design on the strata as this design is part of the variance and mean square error structure.

Definition 15 *Consider a (post)stratification by some categorical variable of the population with H strata. Let the strata sampling sets $\mathcal{S}_h = \{s \cap \mathcal{U}_h$ over every $s \in \mathcal{S}\}$ and $\mathcal{S}_{hg} = \{s \cap (\mathcal{U}_h \cup \mathcal{U}_g)$ over every $s \in \mathcal{S}\}$ for $h,\, g = 1,\, \ldots,\, H$ $(h \neq g)$. The induced designs on $\mathcal{U}_h$ and $\mathcal{U}_h \cup \mathcal{U}_g$, denoted by $p_h(s_h)$ and $p_{hg}(s_h, s_g)$ respectively $(h \neq g)$, are given by*

$$p_h(s_h) = \sum_{\{s: s_h = s \cap \mathcal{U}_h\}} p(s) \quad \textit{for} \quad s_h \in \mathcal{S}_h$$

and

$$p_{hg}(s_h, s_g) = \sum_{\{s: s_h \cup s_g = s \cap (\mathcal{U}_h \cup \mathcal{U}_g)\}} p(s) \quad \textit{for} \quad s_h \cup s_g \in \mathcal{S}_{hg}$$

for $h,\, g = 1,\, \ldots,\, H$ with $h \neq g$.

The induced design on stratum h gives the probabilities of selecting samples with a specific sample stratum $s_h = s \cap \mathcal{U}_h$ for stratum h. These probabilities $p_h(s_h)$ are given for every possible s_h, including $s_h = \emptyset$, and are calculated by summing over the (unconditional) design probabilities for those samples s such that $s_h = s \cap \mathcal{U}_h$. Similarly the induced design on strata h and g $(h \neq g)$ gives the probabilities of selecting a sample with sample strata $s_h \cup s_g = s \cap (\mathcal{U}_h \cup \mathcal{U}_g)$.

Note that the probability of $s_h = \emptyset$ can be positive. If such a sample is selected then we could consider merging similar strata together so that our sample stratum s_h is nonempty.

Since the survey weight $b_{s_h i}$ for unit $i \in s_h$ of a separate general linear estimator is independent of units in s_g for $g = 1, \ldots, H$ $(g \neq h)$, it follows from (2.4) that

$$B_i = \sum_{s \ni i} p(s) b_{s_h i} = \sum_{s_h \ni i} \left[\sum_{\{s : s_h = s \cap \mathcal{U}_h\}} p(s) \right] b_{s_h i} = \sum_{s_h \ni i} p_h(s_h) b_{s_h i} = B_{hi},$$

say, for unit $i \in \mathcal{U}_h$ $(h = 1, \ldots, H)$. Because the summation in $B_{hi} = \sum_{s_h \ni i} p_h(s_h) b_{s_h i}$ is over fewer terms than that of $B_i = \sum_{s \ni i} p(s) b_{s_h i}$, provided the induced design probability $p_h(s_h)$ is known for all possible s_h, we can consider B_{hi} as a reduced form of B_i.

In the following example we illustrate the induced design on the strata under the simple random sampling design.

Example 6 *For a simple random sampling design on a population of size 5 and sample size 3 we have* $p(s) = \binom{5}{3}^{-1} = 0.1$ *for all possible samples. Now consider a stratification of the population with domain sizes* $\mathbf{N} = (3, 2)$ *where the first three units of the population fall into stratum 1 and the rest fall into stratum 2. The induced design on* $\mathcal{U}_h$*, for* $h = 1, 2$*, is given by*

$$p_h(s_h) = \sum_{\{s : s_h = s \cap \mathcal{U}_h\}} p(s) = \binom{N}{n}^{-1} \binom{N - N_h}{n - n_h}$$

for $s_h \in \mathcal{S}_h$ *and is calculated in table 2.1. The induced design on* $\mathcal{U}_1 \cup \mathcal{U}_2$ *is given by*

$$\begin{aligned} p_{12}(s_1, s_2) &= \sum_{\{s : s_1 \cup s_2 = s \cap (\mathcal{U}_1 \cup \mathcal{U}_2)\}} p(s) \\ &= \binom{N}{n}^{-1} \binom{N - N_1 - N_2}{n - n_1 - n_2} \\ &= \binom{N}{n}^{-1} = 0.1 \end{aligned}$$

for $s_1 \cup s_2 \in \mathcal{S}_{12}$.

Note from table 2.1 that the selection of units from different strata are not independent, i.e. $p_{12}(s_1, s_2) \neq p_1(s_1) p_2(s_2)$.

Table 2.1: Induced design on $\mathcal{U}_h$, for $h = 1, 2$, under simple random sampling with $\mathbf{N} = (3, 2)$ and $n = 3$ for example 6

Stratum h	Stratum sample s_h	$p_h(s_h)$
1	$\{1,2,3\}$	0.1
	$\{1,2\}$	0.2
	$\{1,3\}$	0.2
	$\{2,3\}$	0.2
	$\{1\}$	0.1
	$\{2\}$	0.1
	$\{3\}$	0.1
	$\emptyset$	0
2	$\{4,5\}$	0.3
	$\{4\}$	0.3
	$\{5\}$	0.3
	$\emptyset$	0.1

Note that when estimating the domain total $T_{Y_h} = \sum_{i \in \mathcal{U}_h} y_i$ of stratum h using an estimator of the form $\hat{T}_{s_h}$ from (2.1), we can make inference about T_{Y_h} with respect to the induced design on stratum h. This would be equivalent to making unconditional inference about the domain total with respect to the unconditional design, see Särndal & Hidiroglou (1989).

By following Sugden & Smith's (2006) approach we can write the following for a separate general linear strategy $stg(\hat{T}_s, p)$: Using (2.2), which says

$$\mathrm{Var}(\hat{T}_s, p) = \sum_{i \in \mathcal{U}} C_i y_i^2 + \sum_{i \in \mathcal{U}} \sum_{\substack{j \in \mathcal{U} \\ j \neq i}} D_{ij} y_i y_j,$$

we can write

$$C_i = \sum_{s_h \ni i} p_h(s_h) b_{s_h i}^2 - B_{hi}^2 = C_{hi},$$

say, for unit $i \in \mathcal{U}_h$,

$$D_{ij} = \sum_{s_h \ni i,j} p_h(s_h) b_{s_h i} b_{s_h j} - B_{hi} B_{hj} = D_{hij},$$

say, for units $i \neq j \in \mathcal{U}_h$ and

$$D_{ij} = \sum_{\substack{s_h \ni i \\ s_g \ni j}} p_{hg}(s_h, s_g) b_{s_h i} b_{s_g j} - B_{hi} B_{gj} = D_{hgij},$$

for units $i \in \mathcal{U}_h$ and $j \in \mathcal{U}_g$ $(h \neq g)$.

Thus it follows from (2.2) that the variance of any separate general linear strategy can be written as

$$\mathrm{Var}(\hat{T}_s, p) = \sum_{h=1}^{H} \left(\sum_{i \in \mathcal{U}_h} C_{hi} y_i^2 + \sum_{i \in \mathcal{U}_h} \sum_{\substack{j \in \mathcal{U}_h \\ j \neq i}} D_{hij} y_i y_j \right) + \sum_{h=1}^{H} \sum_{\substack{g=1 \\ g \neq h}}^{H} \left(\sum_{i \in \mathcal{U}_h} \sum_{j \in \mathcal{U}_g} D_{hgij} y_i y_j \right) \tag{2.6}$$

and the mean square error can be written as

$$\begin{aligned}\mathrm{MSE}(\hat{T}_s, p) &= \sum_{h=1}^{H}\left(\sum_{i\in\mathcal{U}_h}[C_{hi} + (B_{hi}-1)^2]y_i^2\right)\\ &\quad+\sum_{h=1}^{H}\left(\sum_{i\in\mathcal{U}_h}\sum_{\substack{j\in\mathcal{U}_h\\ j\neq i}}[D_{hij} + (B_{hi}-1)(B_{hj}-1)]y_iy_j\right)\\ &\quad+\sum_{h=1}^{H}\sum_{\substack{g=1\\ g\neq h}}^{H}\left(\sum_{i\in\mathcal{U}_h}\sum_{j\in\mathcal{U}_g}[D_{hgij} - (B_{hi}-1)(B_{gj}-1)]y_iy_j\right).\end{aligned}$$

For some complicated unequal probability sampling schemes it may be easier to calculate the induced design probabilities on the strata than the design probabilities $p(s)$. In which case calculating the exact mean square error of a separate general linear strategy would be simpler through the induced design on the strata.

Note that a feature of a separate general linear strategy is that its within stratum variance component, the first term in (2.6), is equal to the sum of the stratum variances for the domain total estimator $\hat{T}_{s_h}$, where $\hat{T}_{s_h}$ is defined as in (2.1). And also the between stratum variance component, the second term in (2.6), is equal to the covariance of the $\hat{T}_{s_h}$'s which can take positive or negative values.

Lemma 8 *If the selection of units from different (post)strata are independent, i.e. if*

$$p_{hg}(s_h, s_g) = p_h(s_h)p_g(s_g)$$

for all $h,\ g = 1,\ \ldots,\ H$ $(h \neq g)$ *such that* $s_h \cup s_g \in \mathcal{S}_{hg}$ *then the between stratum variance component is equal to zero for any separate general linear estimator.*

Proof From (2.6), the between stratum component is equal to

$$\sum_{h=1}^{H}\sum_{\substack{g=1\\g\neq h}}^{H}\left[\sum_{i\in\mathcal{U}_h}\sum_{j\in\mathcal{U}_g}\left(\sum_{\substack{s_h\ni i\\s_g\ni j}}p_{hg}(s_h,s_g)b_{s_hi}b_{s_gj}-B_{hi}B_{gj}\right)y_iy_j\right]=$$

$$\sum_{h=1}^{H}\sum_{\substack{g=1\\g\neq h}}^{H}\left[\sum_{i\in\mathcal{U}_h}\sum_{j\in\mathcal{U}_g}\left(\sum_{\substack{s_h\ni i\\s_g\ni j}}p_{hg}(s_h,s_g)b_{s_hi}b_{s_gj}-\sum_{s_h\ni i}\sum_{s_g\ni j}p_h(s_h)p_g(s_g)b_{s_hi}b_{s_gj}\right)y_iy_j\right]$$

and by independence this is equal to

$$\sum_{h=1}^{H}\sum_{\substack{g=1\\g\neq h}}^{H}\left[\sum_{i\in\mathcal{U}_h}\sum_{j\in\mathcal{U}_g}\left(\sum_{\substack{s_h\ni i\\s_g\ni j}}p_{hg}(s_h,s_g)b_{s_hi}b_{s_gj}-\sum_{\substack{s_h\ni i\\s_g\ni j}}p_{hg}(s_h,s_g)b_{s_hi}b_{s_gj}\right)y_iy_j\right]$$
$$=0.$$

That completes the proof. $\square$

If the between stratum variance component for any separate general linear strategy is zero then the variance of the strategy will be equal to the sum of the stratum variances.

It may be desirable for the variance of a separate general linear strategy to be equal to the sum of stratum variances as it may be easier to analyse the whole variance by analysing the individual stratum variances. As we have mentioned in section 1.4.4, p.32, it may be easier to estimate the whole variance by estimating the stratum variances.

2.4 The Horvitz-Thompson Estimator

For any general probability sampling design without replacement with $\pi_i > 0$ for all $i \in \mathcal{U}$, an unbiased estimator for the total $T_Y = \sum_{i\in\mathcal{U}} y_i$ is given by Horvitz &

Thompson (1952)

$$\hat{T}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i}.$$

Proof of its unbiasedness can be shown by using lemma 6. The survey weight for $\hat{T}_{HT}$ is $b_{si} = \pi_i^{-1}$ for all $i \in s$ and provided $\pi_i > 0$ for all $i \in \mathcal{U}$ we have

$$B_i = \sum_{s \ni i} p(s) b_{si} = \sum_{s \ni i} p(s) \frac{1}{\pi_i} = \frac{1}{\pi_i} \sum_{s \ni i} p(s) = \frac{1}{\pi_i} \pi_i = 1$$

for all $i \in \mathcal{U}$.

Note that for any fixed sample size design p with $\pi_i > 0$ for all $i \in \mathcal{U}$, $stg(\hat{T}_{HT}, p)$ will always be calibrated for vectors proportional to $\boldsymbol{\pi}^t = (\pi_1, \ldots, \pi_N)$.

The variance of the Horvitz-Thompson estimator can be written as, by using (2.2),

$$\text{Var}(\hat{T}_{HT}, p) = \sum_{i \in \mathcal{U}} C_i y_i^2 + \sum_{i \in \mathcal{U}} \sum_{\substack{j \in \mathcal{U} \\ j \neq i}} D_{ij} y_i y_j,$$

where

$$C_i = \sum_{s \ni i} p(s) b_{si}^2 - B_i^2 = \frac{1}{\pi_i^2} \sum_{s \ni i} p(s) - 1 = \frac{1}{\pi_i} - 1 = \frac{(1 - \pi_i)}{\pi_i}$$

for unit $i \in \mathcal{U}$ and

$$D_{ij} = \sum_{s \ni i,j} p(s) b_{si} b_{sj} - B_i B_j = \frac{1}{\pi_i \pi_j} \sum_{s \ni i,j} p(s) - 1 = \frac{\pi_{ij}}{\pi_i \pi_j} - 1 = \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j}$$

for units $i \neq j \in \mathcal{U}$. Hence we have

$$\text{Var}(\hat{T}_{HT}, p) = \sum_{i \in \mathcal{U}} \frac{y_i^2 (1 - \pi_i)}{\pi_i} + \sum_{i \in \mathcal{U}} \sum_{\substack{j \in \mathcal{U} \\ j \neq i}} \frac{y_i y_j (\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j}$$

which is the standard form for the variance of $stg(\hat{T}_{HT}, p)$.

The following unbiased estimator for this variance was given by Horvitz & Thompson (1952)

$$v_{HT} = \sum_{i \in s} \frac{y_i^2 (1 - \pi_i)}{\pi_i^2} + \sum_{i \in s} \sum_{\substack{j \in s \\ j \neq i}} \frac{y_i y_j (\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j \pi_{ij}}$$

provided $\pi_i > 0$ for $i \in \mathcal{U}$ and $\pi_{ij} > 0$ for all $i \neq j \in \mathcal{U}$.

Sen (1953), Yates & Grundy (1953) gave an alternative form for the variance of the Horvitz-Thompson estimator for fixed sample size n which is given by

$$\text{Var}(\hat{T}_{HT}, p) = \frac{1}{2} \sum_{i \in \mathcal{U}} \sum_{\substack{j \in \mathcal{U} \\ j \neq i}} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 . \tag{2.7}$$

From this form of the variance they gave another unbiased variance estimator which is given by

$$v_{SYG} = \frac{1}{2} \sum_{i \in s} \sum_{\substack{j \in s \\ j \neq i}} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2$$

provided $\pi_i > 0$ for all $i \in \mathcal{U}$ and $\pi_{ij} > 0$ for all $i \neq j \in \mathcal{U}$.

Both variance estimators, v_{HT} and v_{SYG}, can give negative values but a sufficient condition for v_{SYG} to be nonnegative is if $\pi_{ij} \leq \pi_i \pi_j$ for all units $i \neq j$ in the sample. The form of v_{SYG} was generalized by Vijayan (1975), see (6.1) in chapter 6.

For a design conditional on $\mathbf{n}$, we define the conditionally weighted Horvitz-Thompson estimator, denoted by $\hat{T}^*_{HT}$, using the conditional inclusion probabilities in place of the unconditional inclusion probabilities in $\hat{T}_{HT}$,

$$\hat{T}^*_{HT} = \sum_{i \in s} \frac{y_i}{\pi^*_i} = \sum_{h=1}^{H} I_h \sum_{i \in s_h} \frac{y_i}{\pi^*_i} = \sum_{h=1}^{H} I_h \hat{T}^*_{HT,h}$$

where $\hat{T}^*_{HT,h} = \sum_{i \in s_h} y_i / \pi^*_i$ is an estimator for the domain total of the y's in poststratum h for some $h = 1, \ldots, H$.

We will now consider the conditional properties of $\hat{T}^*_{HT}$ under some design p or $stg(\hat{T}^*_{HT}, p \mid \mathcal{S}_{\mathbf{n}})$.

Lemma 9 *The strategy $stg(\hat{T}^*_{HT}, p \mid \mathcal{S}_{\mathbf{n}})$ is (conditionally) unbiased for T_Y provided $\pi^*_i > 0$ for all $i \in \mathcal{U}_h$, $h = 1, \ldots, H$.*

Proof The survey weights for $\hat{T}^*_{HT}$ are equal to $b_{s_h i} = 1/\pi^*_i$ for unit $i \in s_h$ ($h = 1, \ldots, H$). Let $p^*_h(s_h)$ be the induced design on $\mathcal{U}_h$ given by the conditional design $p^*(s)$, then provided $\pi^*_i > 0$ for all $i \in \mathcal{U}$ by lemma 6 we have

$$B_{hi} = \sum_{s_h \ni i} p^*_h(s_h) b_{s_h i} = \sum_{s_h \ni i} p^*_h(s_h) \frac{1}{\pi^*_i} = \frac{1}{\pi^*_i} \sum_{s_h \ni i} p^*_h(s_h) = \frac{1}{\pi^*_i} \pi^*_i = 1$$

for all $i \in \mathcal{U}$ which implies that $stg(\hat{T}_{HT}, p \mid \mathcal{S}_{\mathbf{n}})$ will be unbiased for T_Y. □

Remark 2 *If $\pi^*_i > 0$ for all $i \in \mathcal{U}_h$ then $n_h > 0$, hence for $\hat{T}^*_{HT}$ to be conditionally unbiased for T_Y we must have $n_h > 0$ for all $h = 1, \ldots, H$, i.e. the sample must contain units from all poststrata.*

In fact $stg(\hat{T}^*_{HT,h}, p)$ is conditionally unbiased for T_{Y_h} for all $h = 1, \ldots, H$ provided $\pi^*_i > 0$ for $i \in \mathcal{U}_h$. Note that because $stg(\hat{T}^*_{HT}, p)$ is conditionally unbiased for T_Y, it is also unbiased for T_Y under the weak conditional design.

By using lemma 4 part (a) we see that $stg(\hat{T}^*_{HT}, p \mid \mathcal{S}_{\mathbf{n}})$ is calibrated for the within poststratum conditional inclusion probabilities $\boldsymbol{\pi}^*_h$ where $\boldsymbol{\pi}^*_h$ is an $N \times 1$ column vector whose i^{th} element is equal to

$$\boldsymbol{\pi}^*_{hi} = \begin{cases} \pi^*_i & \text{if} \quad i \in \mathcal{U}_h \\ 0 & \qquad \text{else} \end{cases}$$

for $h = 1, \ldots, H$. Hence, provided $\pi^*_i > 0$ for all $i \in \mathcal{U}$, the conditionally weighted Horvitz-Thompson estimator will always be calibrated for any linear combination of the $\boldsymbol{\pi}^*_h$'s, $h = 1, \ldots, H$.

The conditional variance of $stg(\hat{T}^*_{HT}, p)$ can be given in the form (2.6) where

$$\begin{aligned} C_{hi} = \sum_{s_h \ni i} p^*_h(s_h) b^2_{s_h i} - B^2_{hi} &= \frac{1}{(\pi^*_i)^2} \sum_{s_h \ni i} p^*_h(s_h) - 1 \\ &= \frac{(1 - \pi^*_i)}{\pi^*_i} \end{aligned}$$

for unit $i \in \mathcal{U}_h$,

$$D_{hij} = \sum_{s_h \ni i,j} p_h^*(s_h) b_{s_h i} b_{s_h j} - B_{hi} B_{hj} = \frac{1}{\pi_i^* \pi_j^*} \sum_{s_h \ni i,j} p_h^*(s_h) - 1 = \frac{(\pi_{ij}^* - \pi_i^* \pi_j^*)}{\pi_i^* \pi_j^*}$$

for units $i \neq j \in \mathcal{U}_h$ and

$$D_{hgij} = \sum_{\substack{s_h \ni i \\ s_g \ni j}} p_{hg}^*(s_h, s_g) b_{s_h i} b_{s_g j} - B_{hi} B_{gj} = \frac{1}{\pi_i^* \pi_j^*} \sum_{\substack{s_h \ni i \\ s_g \ni j}} p_{hg}^*(s_h, s_g) - 1 = \frac{(\pi_{ij}^* - \pi_i^* \pi_j^*)}{\pi_i^* \pi_j^*}$$

for units $i \in \mathcal{U}_h$ and $j \in \mathcal{U}_g$ $(h \neq g)$. This gives us

$$\begin{aligned} \mathrm{Var}(\hat{T}_{HT}^*, p \mid \mathcal{S}_{\mathbf{n}}) &= \sum_{h=1}^{H} \left[\sum_{i \in \mathcal{U}_h} \frac{y_i^2 (1 - \pi_i^*)}{\pi_i^*} + \sum_{i \in \mathcal{U}_h} \sum_{\substack{j \in \mathcal{U}_h \\ j \neq i}} \frac{y_i y_j (\pi_{ij}^* - \pi_i^* \pi_j^*)}{\pi_i^* \pi_j^*} \right] \\ &\quad + \sum_{h=1}^{H} \sum_{\substack{g=1 \\ g \neq h}}^{H} \left[\sum_{i \in \mathcal{U}_h} \sum_{j \in \mathcal{U}_g} \frac{y_i y_j (\pi_{ij}^* - \pi_i^* \pi_j^*)}{\pi_i^* \pi_j^*} \right] \end{aligned}$$

Note that the between stratum component is equal to zero when $\pi_{ij}^* = \pi_i^* \pi_j^*$ for all $i \in \mathcal{U}_h$ and $j \in \mathcal{U}_g$ $(h \neq g)$ $h, g = 1, \ldots, H$, i.e. when the event representing the selection of two units from different strata are independent, the between stratum variance component is equal to zero as expected due to lemma 8.

An unbiased estimator for $\mathrm{Var}(\hat{T}_{HT}^*, p \mid \mathcal{S}_{\mathbf{n}})$ is given by

$$\begin{aligned} v_{HT}^* &= \sum_{h=1}^{H} \left[\sum_{i \in s_h} \frac{y_i^2 (1 - \pi_i^*)}{\pi_i^{*2}} + \sum_{i \in s_h} \sum_{\substack{j \in s_h \\ j \neq i}} \frac{y_i y_j (\pi_{ij}^* - \pi_i^* \pi_j^*)}{\pi_i^* \pi_j^* \pi_{ij}^*} \right] \\ &\quad + \sum_{h=1}^{H} \sum_{\substack{g=1 \\ g \neq h}}^{H} \left[\sum_{i \in s_h} \sum_{j \in s_g} \frac{y_i y_j (\pi_{ij}^* - \pi_i^* \pi_j^*)}{\pi_i^* \pi_j^* \pi_{ij}^*} \right] \end{aligned}$$

provided $\pi_i^* > 0$ and $\pi_{ij}^* > 0$ for all $i \neq j \in \mathcal{U}$.

We can also write the conditional variance of $\hat{T}_{HT}^*$ in Sen, Yates and Grundy's form given in (2.7).

Theorem 2 *For any conditional design with $\pi_i^* > 0$ for all $i \in \mathcal{U}_h$ $h = 1, \ldots, H$ and sampling configuration $\mathbf{n}$, we have*

$$\begin{aligned}\mathrm{Var}(\hat{T}_{HT}^*, p \mid \mathcal{S}_{\mathbf{n}}) &= \sum_{h=1}^{H} \left[\frac{1}{2} \sum_{i \in \mathcal{U}_h} \sum_{\substack{j \in \mathcal{U}_h \\ j \neq i}} (\pi_i^* \pi_j^* - \pi_{ij}^*) \left(\frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*} \right)^2 \right] \\ &+ \sum_{h=1}^{H} \sum_{\substack{g=1 \\ g \neq h}}^{H} \left[\frac{1}{2} \sum_{i \in \mathcal{U}_h} \sum_{j \in \mathcal{U}_g} (\pi_i^* \pi_j^* - \pi_{ij}^*) \left(\frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*} \right)^2 \right].\end{aligned}$$

Proof The proof follows immediately by applying (2.7) to $stg(\hat{T}_{HT}^*, p^*)$ where $\hat{T}_{HT}^*$ is regarded as a general linear estimator and p^* is regarded as the sampling design on $\mathcal{S}$.

□

Hence another unbiased estimator for $\mathrm{Var}(\hat{T}_{HT}^*, p \mid \mathcal{S}_{\mathbf{n}})$ is given by

$$\begin{aligned}v_{SYG}^* &= \sum_{h=1}^{H} \left[\frac{1}{2} \sum_{i \in s_h} \sum_{\substack{j \in s_h \\ j \neq i}} \frac{(\pi_i^* \pi_j^* - \pi_{ij}^*)}{\pi_{ij}^*} \left(\frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*} \right)^2 \right] \\ &+ \sum_{h=1}^{H} \sum_{\substack{g=1 \\ g \neq h}}^{H} \left[\frac{1}{2} \sum_{i \in s_h} \sum_{j \in s_g} \frac{(\pi_i^* \pi_j^* - \pi_{ij}^*)}{\pi_{ij}^*} \left(\frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*} \right)^2 \right]\end{aligned}$$

provided $\pi_i^* > 0$ and $\pi_{ij}^* > 0$ for all $i \neq j \in \mathcal{U}$.

It is easy to see that when the unconditional design is simple random sampling $\hat{T}_{HT}$ reduces to the expansion estimator $N\bar{y}$ since $\pi_i = n/N$ for all units in the population.

Moreover its variance reduces to

$$\text{Var}(\hat{T}_{HT}, \text{SRS}) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) S_Y^2$$

since $\pi_{ij} = n(n-1)/N(N-1)$ for all distinct pairs. Also the conditionally weighted Horvitz-Thompson estimator reduces to the stratified estimator $\sum_{h=1}^{H} I_h N_h \bar{y}_h$ since $\pi_i^* = n_h/N_h$ for all units in $\mathcal{U}_h$ for $h = 1, \ldots, H$. Furthermore its conditional variance reduces to

$$\text{Var}(\hat{T}_{HT}^*, \text{SRS} \mid \mathcal{S}_\mathbf{n}) = \sum_{h=1}^{H} I_h N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_{Y_h}^2$$

where $S_{Y_h}^2$ is the variance of the y's in domain h for $h = 1, \ldots, H$. We will discuss more properties of these estimators in section 2.6.

2.5 Unequal probability sampling schemes

A number of unequal probability sampling schemes without replacement have been proposed. Hanif & Brewer (1980) studied 50 different sampling schemes and Tillé (2006) has looked at more recent sampling procedures. The following properties given by Sugden *et al.* (1996) and Tillé (1996a), are desirable for a satisfactory sampling scheme:

1. The sampling scheme should be exact in the sense that the units should be included with probabilities π_i, $i \in \mathcal{U}$, proportional to some size measure defined for each unit in the population. Any sampling scheme satisfying this condition is called a probability proportional to size sampling scheme and is denoted by $\pi\text{PS}(\mathbf{x})$ where $\mathbf{x}$ is the size variable.

2. The joint inclusion probabilities should satisfy $\pi_{ij} \leq \pi_i \pi_j$ for all $i, j \in \mathcal{U}$ $(i \neq j)$ so that v_{SYG} is always nonnegative.

3. The second order joint inclusion probability π_{ij} should be positive for all $i, j \in \mathcal{U}$ $(i \neq j)$ in order to permit the use of unbiased estimators for the variance or mean square error of strategies.

4. The sample size n should be fixed so that $\text{Var}(\hat{T}_{HT}, p)$ can be written as in (2.7).

5. The implementation of the scheme should be easy to understand and can be programmed on a computer.

6. The second order joint inclusion probability π_{ij} for all $i, j \in \mathcal{U}$ $(i \neq j)$ should be easy to compute without examining all the probabilities $p(s)$ of selecting the samples which contains unit i and j (however, this is not too difficult with a small population).

7. The scheme should be fast and the selection of the sample should be made without computing the $p(s)$ for all possible samples.

8. The scheme should be list sequential i.e. it should examine the units in accordance to some order of the population being implemented with a single pass of the population.

9. The design probabilities, $p(s)$, should not depend on the order of the units in the population.

10. The sampling scheme should give an estimator with a smaller variance than when sampling with replacement (n independent draws with unequal probability selection) with selection probability π_i/n for unit $i \in \mathcal{U}$ where π_i refers to the without replacement unequal probability sampling scheme.

Note that a sampling scheme that satisfies the first four conditions will ensure that the variance estimator v_{SYG} is unbiased and nonnegative.

Tillé (1996a) has shown that, in general, an unequal probability sampling scheme without replacement for fixed sample size that is list sequential does not satisfy conditions (3) and (9). Hence there is no sampling scheme that satisfies all of these conditions in general.

Let $\mathbf{x}$ be a size variable, or size vector, defined for each unit in the population. For an unequal probability sampling design based on $\mathbf{x}$, we could write $p(s \mid \mathbf{x})$ in place of $p(s)$ as there are many different size vectors to choose from and the selection probability for sample s can vary for different size vectors. Also for sampling designs that are dependent on the order of the population units, such as list sequential schemes, the probability of selecting a particular sample of units before reordering the population units is not necessarily the same as the probability of selecting that same sample of units after reordering the population units.

For the purpose of explaining the following definition and lemma we will use the notation $p(s \mid \mathbf{x})$ for the selection probability of sample s when using a sampling design based on a given size vector $\mathbf{x}$.

Note that if $\mathbf{x}$ is a constant vector and p is a probability proportional to size sampling scheme based on $\mathbf{x}$, then $p(s \mid \mathbf{x})$ is equal to the simple random sampling selection probabilities for all $s \in \mathcal{S}$.

Following Sugden (1993), let $\mathbf{x}_\tau$ be a vector whose i^{th} element is equal to $x_{\tau(i)}$ for every $i \in \mathcal{U}$ for some permutation τ of $\mathcal{U}$, i.e. τ is some bijective function such that $\tau : \mathcal{U} \rightarrow \mathcal{U}$. The original order of the population units will be listed, by their labels, as 1, 2, . . . , N. To change the order of the units in the population we first apply some permutation τ to their labels which replaces unit i with unit $\tau(i)$ in the i^{th} position of the population list for each $i \in \mathcal{U}$. Then we relabel unit $\tau(i)$, for each $i \in \mathcal{U}$, by their entry position in the population list. i.e. unit $\tau(i)$ in the original list is first moved to the i^{th} position of the population list and then relabelled as unit i for each

$i \in \mathcal{U}$. We will now illustrate this method of reordering the population units with a simple example.

For a population of size 5, let $s = \{2, 4\}$ so that the size variable attached to these units are $\{x_2, x_4\}$. Now consider the permutation τ_1 which is represented by

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 5 & 1 & 2 \end{pmatrix}.$$

By applying the permutation τ_1 to $\mathcal{U}$, we change the order of the units in the population list. So unit 4 will be relabelled as unit 1, unit 3 will be relabelled as unit 2, unit 5 will be relabelled as unit 3, unit 1 will be relabelled as unit 4 and unit 2 will be relabelled as unit 5. Also the size measures, and any other variables attached to each unit, are rearranged according to τ_1 so that

$$\mathbf{x}_{\tau_1}^t = (x_{\tau_1(1)}, x_{\tau_1(2)}, x_{\tau_1(3)}, x_{\tau_1(4)}, x_{\tau_1(5)}) = (x_4, x_3, x_5, x_1, x_2).$$

Now the sample $s = \{2, 4\}$, after reordering the population units according to τ_1, corresponds to the size measures $\{x_{\tau_1(2)}, x_{\tau_1(4)}\} = \{x_3, x_1\}$ which are clearly not the same as the size measures for those unit labels before reordering the population units. But the sample $\tau_1^{-1}(s) = \{5, 1\}$ corresponds to the size measures $\{x_{\tau_1(5)}, x_{\tau_1(1)}\} = \{x_2, x_4\}$ which are the same measures for $s = \{2, 4\}$ before reordering the population units according to τ_1. In fact the units corresponding to labels $\{2, 4\}$ before reordering the population units according to τ_1 and the units corresponding to the labels $\{5, 1\}$ after reordering the population units are the same units but with different labels. So if

$$p(s \mid \mathbf{x}) = p(\tau^{-1}(s) \mid \mathbf{x}_\tau)$$

for any permutation τ of $\mathcal{U}$ then the probability of selecting the units corresponding to $\{x_2, x_4\}$ will always be independent of the population order.

We now give a formal definition for sampling schemes that satisfy condition (9).

Definition 16 *Let* $\mathbf{x}$ *be any size vector in* $\mathcal{R}^N$ *and let* τ *be any permutation of* $\mathcal{U}$*. A sampling scheme is said to be invariant by permutation of the population order if*

$$p(s \mid \mathbf{x}) = p(\tau^{-1}(s) \mid \mathbf{x}_\tau)$$

for any $s \in \mathcal{S}$*.*

Hence a sampling scheme that is invariant by permutation of the population order is equivalent to saying that its selection probabilities $p(s)$ are independent of the population order. The reason why we have defined a sampling scheme that is invariant of the population order as in definition 16 is because it will be useful in the proof of the following lemma which gives an interesting property of designs that are independent of the population order.

Lemma 10 *Let* σ *be a permutation on* $\mathcal{U}$ *such that* $\mathbf{x} = \mathbf{x}_\sigma$*, i.e.* σ *is a permutation on* $\mathcal{U}$ *that preserves the values of each entry of the vector* $\mathbf{x}$*. Then for a sampling scheme that is invariant by permutation of the population order we have*

$$p(s \mid \mathbf{x}) = p(\sigma(s) \mid \mathbf{x})$$

for all $s \in \mathcal{S}$*.*

Proof Since the permutation σ is such that $x_i = x_{\sigma(i)}$ for all $i \in \mathcal{U}$, it follows that for any sampling scheme we must have

$$p(s \mid \mathbf{x}) = p(s \mid \mathbf{x}_\sigma)$$

for any $s \in \mathcal{S}$. For a sampling scheme that is invariant by permutation of the population order, we have

$$p(s \mid \mathbf{x}_\sigma) = p(\sigma(s) \mid \mathbf{x}_{\sigma\sigma^{-1}}) = p(\sigma(s) \mid \mathbf{x})$$

which implies

$$p(s \mid \mathbf{x}) = p(\sigma(s) \mid \mathbf{x})$$

for any $s \in \mathcal{S}$ and that completes the proof. $\square$

Note that the units in s corresponding to $p(s \mid \mathbf{x})$ are not necessarily the same units in s corresponding to $p(s \mid \mathbf{x}_\sigma)$ although there labels are the same. Also note that the units in s are not necessarily the same as the units in $\sigma(s)$ given a size vector $\mathbf{x}$.

For the rest of this thesis we will no longer use the notation $p(s \mid \mathbf{x})$ to denote the selection probability of sample s as it won't be necessary, and so we return to using $p(s)$ instead.

Lemma 10 shows that for a given population size vector $\mathbf{x}$, if two different samples $s_1 = \{u_1, \ldots, u_n\}$ and $s_2 = \{v_1, \ldots, v_n\}$ are such that the corresponding sets $\{x_{u_1}, \ldots, x_{u_n}\}$ and $\{x_{v_1}, \ldots, x_{v_n}\}$ are equal, then $p(s_1) = p(s_2)$ under a sampling scheme that is invariant by permutation of the population order. In particular, if we poststratify the population by the values of the population size measure $\mathbf{x}$, the conditional design on the observed sampling allocation $\mathbf{n}$ is just stratified random sampling.

Example 7 *Consider a population of size $N = 5$ and sample of size $n = 3$ and let the size variable for each unit be*

$$\mathbf{x}^t = (3, 4, 3, 5, 4).$$

The sample space is given in table 2.2.

Table 2.2: Sample space for $N = 5$ and $n = 3$ for example 7

$\mathcal{S}$	$\mathbf{x}_s$
$s_1 = \{1, 2, 3\}$	$(3, 4, 3)$
$s_2 = \{1, 2, 4\}$	$(3, 4, 5)$
$s_3 = \{1, 2, 5\}$	$(3, 4, 4)$
$s_4 = \{1, 3, 4\}$	$(3, 3, 5)$
$s_5 = \{1, 3, 5\}$	$(3, 3, 4)$
$s_6 = \{1, 4, 5\}$	$(3, 5, 4)$
$s_7 = \{2, 3, 4\}$	$(4, 3, 5)$
$s_8 = \{2, 3, 5\}$	$(4, 3, 4)$
$s_9 = \{2, 4, 5\}$	$(4, 5, 4)$
$s_{10} = \{3, 4, 5\}$	$(3, 5, 4)$

Then if the design is invariant to permutations of the population order, by lemma 10, we have

$$p(s_1) = p(s_5), \quad p(s_2) = p(s_6) = p(s_7) = p(s_{10}) \quad \textit{and} \quad p(s_3) = p(s_8).$$

Furthermore by poststratifying by the sizes of $\mathbf{x}$ *(*$H = 3$*), the design conditional on the sampling allocation* $\mathbf{n} = (1, 1, 1)$*, say, i.e. conditional on the support* $\mathcal{S}_\mathbf{n} = \{s_2, s_6, s_7, s_{10}\}$*, is equivalent to a stratified random sampling design on* $\mathbf{n}$ *since the probability of selecting any sample from* $\mathcal{S}_\mathbf{n}$ *will be equal to a constant.*

It is also clear that under the invariant property the joint order inclusion probabilities of any subset of units with the same sizes are always equal to a constant value.

We will consider the effects of designs which are invariant to permutations of the population order on estimators in section 2.6.

2.5.1 Systematic sampling

Systematic sampling was first proposed by Madow (1949) for fixed sample size and is one of the most commonly used unequal probability sampling schemes due to its simplicity. The method is as follows:

a) Assuming $0 < \pi_i < 1$ for all $i \in \mathcal{U}$, define

$$r_i = \sum_{m=1}^{i} \pi_m \quad \text{for all} \quad i \in \mathcal{U}$$

with $r_0 = 0$ and $r_N = n$.

b) Generate a value u from a uniform distribution $\mathrm{U}[0, 1]$.

c) For $j = 1, \ldots, n$ select the i_j^{th} units such that

$$r_{i_j - 1} \leq u + j - 1 < r_{i_j}.$$

This is the simplest way of implementing an unequal probability design, however the second order joint inclusion probabilities are often zero, see for example Tillé (2006, p.127-128), so that variance estimators may be badly biased.

Two unequal probability sampling schemes that satisfy most of the conditions (1) to (10) are Chao's scheme (1982), which is a list sequential scheme, and Tillé's elimination procedure (1996b), whose $p(s)$'s are independent of the population order.

In the following we give the description, implementation and calculations of design probabilities under Chao's and Tillé's scheme for unequal probability sampling without replacement.

2.5.2 Chao's scheme (1982)

This is a draw by draw list sequential procedure for fixed sample size n. The scheme is a generalization of Alan Waterman's reservoir sampling scheme (see Knuth, 1981, p.138-139) and reduces to it when the size measures are all equal. For this scheme the population size N does not need to be known in advance. However, second and higher-order inclusion probabilities depend on the ordering of the population through the size variable $\mathbf{x}$.

The procedure is as follows: starting at stage $k = n$ with the first n units of the population in the initial sample with probability 1, the following steps are performed for each stage $k = n + 1, \ldots, N - 1$.

a) Determine the first order inclusion probabilities at stage k, denoted by $\pi(k; i)$, where

$$\pi(k;i) = \frac{nx_i}{\sum_{j=1}^{k} x_j},$$

for $i = 1, \ldots, k$. If $\pi(k;i) \geq 1$, then we put $\pi(k;i) = 1$ and recalculate the other $\pi(k;j)$'s for $j = 1, \ldots, k$ $(j \neq i)$ with the reduced sample size until each $\pi(k;i)$ is in $[0,1]$.

b) Partition the units in the population at stage k into three sets:

$$\begin{aligned} A_k &= \{i \in \mathcal{U} : \pi(k;i) = 1 \quad \text{and} \quad \pi(k+1;i) = 1\} \\ B_k &= \{i \in \mathcal{U} : \pi(k;i) = 1 \quad \text{and} \quad \pi(k+1;i) < 1\} \\ C_k &= \{i \in \mathcal{U} : \pi(k;i) < 1 \quad \text{and} \quad \pi(k+1;i) < 1\}. \end{aligned}$$

c) Calculate replacement probabilities R_{ki} for each unit $i = 1, \ldots, k$ at each stage k as

$$R_{ki} = \begin{cases} 0 & \text{for} \quad i \in A_k \\ (1 - \pi(k+1;i))/W_k & \text{for} \quad i \in B_k \\ (1 - T_{B_k})/(n - L_k) & \text{for} \quad i \in C_k \end{cases} \tag{2.8}$$

where $W_k = \pi(k+1;k+1)$ is the probability of selecting a new unit $k+1$ at stage $k+1$, $T_{B_k} = \sum_{j \in B_k} R_{kj}$ and $L_k = \#(A_k) + \#(B_k)$ is the number of units in A_k and B_k.

d) Generate a random variable u from a uniform distribution U$[0,1]$. If $u < W_k$ then select unit $k+1$ and remove a unit i, say, from the sample with probability equal to R_{ki} replacing it by unit $k+1$. If $u > W_k$ then we retain all units in the sample.

The calculation of the r^{th} order inclusion probabilities at stage k are as follows: For all $r \leq n$ and all $1 \leq i_1 < \ldots < i_r \leq k$, $k > n$

$$\pi(k+1; i_1, \ldots, i_r) = \pi(k; i_1, \ldots, i_r) \left(1 - W_k \sum_{j=1}^{r} R_{ki_j}\right)$$

and for $1 \leq i_1 \leq \ldots < i_{r-1} \leq k$

$$\pi(k+1; i_1, \ldots, i_{r-1}, k+1) = \pi(k; i_1, \ldots, i_{r-1}) W_k \left(1 - \sum_{j=1}^{r-1} R_{ki_j}\right).$$

When $k = N - 1$,

$$\pi(k+1; i_1, \ldots, i_r) = \pi_{i_1 \ldots i_r}.$$

Sugden, *et al.* (1996) showed using simulation studies that from a statistical point of view, Chao's scheme (1982) for unequal probability sampling, although more complicated than systematic $\pi\mathrm{PS}(\mathbf{x})$ sampling, is more efficient. Chao's scheme is a list sequential procedure, i.e. it can be applied to a data file (or population) in only one reading by examining the units in accordance with their order on the data file. It satisfies all of the conditions from p.61-62 except (3) and (9). A necessary and sufficient conditions for strictly positive second-order inclusion probabilities is given by Bethlehem & Schuerhoff (1984) and Sengupta (1989). Chao's scheme has been generalized by the splitting method of Deville & Tillé (1998) which uses a faster algorithm.

Berger (1998) studied the structure of the Sen Yates & Grundy variance estimator, v_{SYG}, under Chao's scheme and showed that the calculation of this estimator can be simplified without having to compute all the second order joint inclusion probabilities. Berger also gave a simpler form of the replacement probabilities R_{ki}, defined in (2.8), for the case where there is no self-selecting units in the population at any stage $k \geq n+1$, i.e. all units are such that $0 < x_i < \frac{1}{n}\sum_{j=1}^{k} x_j$ for all $k \geq n+1$ and for all $i \leq k$. We give another simple form of R_{ki} that always holds even if there are self-selecting units in the population.

Theorem 3 *The replacement probabilities defined in (2.8) can be written as*

$$R_{ki} = \frac{1}{W_k}\left[1 - \frac{\pi(k+1; i)}{\pi(k; i)}\right] \tag{2.9}$$

for $i = 1, \ldots, k$ *and* $k \geq n+1$.

Proof We need to show that (2.9) gives the same values as (2.8) for all units at every stage.

For unit $i \in A_k$ we have $\pi(k+1;i) = \pi(k;i) = 1$ which implies that (2.9) is zero and agrees with (2.8).

For unit $i \in B_k$ we have $\pi(k+1;i) < 1$ and $\pi(k;i) = 1$ so that (2.9) reduces to (2.8) and hence they both give the same value.

For the case where unit $i \in C_k$ we need to consider two possibilities: $W_k < 1$ or $W_k = 1$. First we define the following quantities to help prove the theorem. Let $Q_k = n - L_k$, where L_k is defined in part (c) of Chao's procedure, and let $T_k = \sum_{j \in C_k} x_j$. Then it follows that at stage k

$$\pi(k;i) = \frac{Q_k x_i}{T_k}$$

for $k \geq n+1$ and $i \leq k$.

In the case where $W_k < 1$, and $i \in C_k$, we can write (2.8) as

$$\begin{aligned}
\frac{1 - T_{B_k}}{Q_k} &= \frac{1}{W_k Q_k}\left[W_k - \#(B_k) + \sum_{j \in B_k} \pi(k+1;j)\right] \\
&= \frac{1}{W_k Q_k}\left[\frac{Q_{k+1} x_{k+1}}{T_{k+1}} - \#(B_k) + \frac{Q_{k+1}\sum_{j \in B_k} x_j}{T_{k+1}}\right] \\
&= \frac{Q_{k+1}\left(\sum_{j \in B_k} x_j + x_{k+1}\right) - \#(B_k) T_{k+1}}{W_k Q_k T_{k+1}}, \qquad (2.10)
\end{aligned}$$

and (2.9) can be written as

$$\begin{aligned}
\frac{1}{W_k}\left[1 - \frac{\pi(k+1;i)}{\pi(k;i)}\right] &= \frac{1}{W_k}\left[1 - \frac{Q_{k+1} x_i}{T_{k+1}} \frac{T_k}{Q_k x_i}\right] \\
&= \frac{T_{k+1} Q_k - Q_{k+1} T_k}{W_k T_{k+1} Q_k}. \qquad (2.11)
\end{aligned}$$

Observe that for unit $k+1 \in C_{k+1}$, which is true when $W_k < 1$, we have

$$\sum_{j \in B_k} x_j = T_{k+1} - x_{k+1} - T_k.$$

This implies that the numerator of (2.10) is equal to

$$\begin{aligned}
& Q_{k+1}\left(\sum_{j \in B_k} x_j + x_{k+1}\right) - \#(B_k)T_{k+1} \\
= \; & Q_{k+1}(T_{k+1} - T_k) - \#(B_k)T_{k+1} \\
= \; & T_{k+1}\left(Q_{k+1} - \#(B_k)\right) - T_k Q_{k+1} \\
= \; & T_{k+1}[n - \#(A_{k+1}) - \#(B_{k+1}) - \#(B_k)] - T_k Q_{k+1}. \qquad (2.12)
\end{aligned}$$

It is easily verified for $k+1 \in C_{k+1}$ that

$$\#(B_{k+1}) = \#(A_k) - \#(A_{k+1}),$$

substitute this into (2.12) gives

$$T_{k+1}[n - \#(A_k) - \#(B_k)] - T_k Q_{k+1} = T_{k+1}Q_k - T_k Q_{k+1}$$

which is equal to the numerator of (2.11). Hence it follows that when $W_k < 1$ and $i \in C_k$, (2.9) is the same as (2.8).

If $W_k = 1$ and $i \in C_k$ then (2.8) can be written as

$$\begin{aligned}
\frac{1 - T_{B_k}}{Q_k} &= \frac{1}{Q_k}\left[1 - \#(B_k) + \sum_{j \in B_k} \pi(k+1; j)\right] \\
&= \frac{T_{k+1} - \#(B_k)T_{k+1} + Q_{k+1}\left(\sum_{j \in B_k} x_j\right)}{Q_k T_{k+1}} \\
&= \frac{Q_{k+1}\left(\sum_{j \in B_k} x_j\right) + T_{k+1}(1 - \#(B_k))}{Q_k T_{k+1}} \qquad (2.13)
\end{aligned}$$

and (2.9) can be written simply, from (2.11), as

$$\frac{T_{k+1}Q_k - Q_{k+1}T_k}{T_{k+1}Q_k}. \qquad (2.14)$$

Observe that if unit $k+1$ is in B_{k+1} or A_{k+1}, which will be true when $W_k = 1$, then

$$\sum_{j \in B_k} x_j = T_{k+1} - T_k.$$

This implies that the numerator of (2.13) can be written as

$$\begin{aligned} & Q_{k+1}\left(\sum_{j \in B_k} x_j\right) + T_{k+1}(1 - \#(B_k)) \\ = \; & Q_{k+1}(T_{k+1} - T_k) + T_{k+1}(1 - \#(B_k)) \\ = \; & T_{k+1}\left(Q_{k+1} - \#(B_k) + 1\right) - T_k Q_{k+1} \\ = \; & T_{k+1}[n - \#(A_{k+1}) - \#(B_{k+1}) - \#(B_k) + 1] - T_k Q_{k+1}. \end{aligned} \tag{2.15}$$

When unit $k+1$ is in A_{k+1} or B_{k+1} then it can be verified that

$$\#(B_{k+1}) = \#(A_k) - \#(A_{k+1}) + 1$$

and substituting this into (2.15) gives

$$T_{k+1}[n - \#(A_k) - \#(B_k)] - T_k Q_{k+1} = T_{k+1} Q_k - T_k Q_{k+1}$$

which is the same as the numerator of (2.14). Hence it follows that when $W_k = 1$ and $i \in C_k$, (2.9) is the same as (2.8). That completes the proof. $\square$

Remark 3 *Note that for* $1 \leq i \leq k$ *and* $k \geq n$,

$$\pi(k+1; i) = \pi(k; i)[(1 - W_k) + W_k(1 - R_{ki})]$$

which follows straight from the implementation of the scheme in part (d). This is equivalent to

$$\pi(k+1; i) = \pi(k; i)(1 - W_k R_{ki})$$

which implies that

$$R_{ki} = \frac{1}{W_k}\left[1 - \frac{\pi(k+1; i)}{\pi(k; i)}\right].$$

The form of R_{ki} given in theorem 3 was also given in Tillé (2006, p.120) but without the proof. It is clearly more simple to use theorem 3 to calculate the replacement probabilities when writing a computer algorithm for Chao's procedure than using the partition approach of (2.8).

2.5.3 Tillé's elimination procedure (1996)

One of the main feature of this sampling scheme is that its selection probabilities are independent of the order of the values in the size variable, unlike Chao's scheme. This sampling procedure for fixed sample size n starts with the whole population and then eliminates a unit from the population at each stage of the algorithm until n units are obtained.

The algorithm is as follows: starting at stage $k = N, \ldots, n$

a) Determine the first order inclusion probabilities at stage k as

$$\pi(i;k) = \frac{kx_i}{\sum_{j \in \mathcal{U}} x_j}, \quad \text{for} \quad i \in \mathcal{U}, \quad k = N, \ldots, n$$

If $\pi(k;i) \geq 1$, then put $\pi(k;i) = 1$ and recalculate the other $\pi(k;j)$'s for $j = 1, \ldots, k$ $(j \neq i)$ with the reduced sample size until $\pi(k;j) \leq 1$ for all $j \in \mathcal{U}$.

b) Calculate the elimination probabilities r_{ki} at stage k for each unit $i \in \mathcal{U}$ as

$$r_{ki} = \left(1 - \frac{\pi(i;k)}{\pi(i;k+1)}\right)$$

for $k = N - 1, \ldots, n$.

c) At each stage starting from $k = N - 1, \ldots, n$ a unit i, say, is eliminated from $\mathcal{U}$ with probability r_{ki}.

The calculation of the r^{th} order inclusion probabilities are as follows:

$$\pi_{i_1 i_2 \ldots i_r} = \prod_{k=n}^{N-1} \left(1 - \sum_{j=1}^{r} r_{ki_j}\right)$$

This sampling scheme satisfies all of the conditions from p.61-62 except for (8) and (3). However in his paper, Tillé (1996b) has suggested ways to avoid null joint inclusion probabilities. He also proposed a complementary draw by draw sampling procedure where at each of the first n stages, a unit of the population is to be included in the sample instead of being eliminated. We do not study this complementary scheme.

Tillé's elimination procedure is computationally easy to understand. The algorithm is faster than Chao's scheme and the probabilities $p(s)$ have an exact formula which is non-recursive.

2.6 More on πPS(x) designs

For an unconditional πPS($\mathbf{x}$) design, p, where $\pi_i \propto x_i$ over all $i \in \mathcal{U}$ the Horvitz-Thompson estimator

$$\hat{T}_{HT} = \sum_{i\in s} \frac{y_i}{\pi_i} = \frac{N\bar{X}}{n} \sum_{i\in s} \frac{y_i}{x_i} = \frac{N\bar{X}}{n} \sum_{h=1}^{H} I_h \sum_{i\in s_h} \frac{y_i}{x_i}$$

is calibrated for vectors proportional to $\mathbf{x}$. But if $y_i = c_h x_i$ for $i \in \mathcal{U}_h$ with $h = 1, \ldots, H$ (c_h not all the same) then

$$\hat{T}_{HT} = \frac{N\bar{X}}{n} \sum_{h=1}^{H} I_h c_h n_h \neq \sum_{h=1}^{H} c_h N_h \bar{X}_h = T_Y$$

in general unless $n_h = nN_h\bar{X}_h/N\bar{X}$, a kind of proportional allocation. Note that this allocation is equal to the expected value of the sample strata size $n_h(s)$ under a

πPS($\mathbf{x}$) design, p, since

$$\mathrm{E}[n_h(s), p] = \sum_{s \in \mathcal{S}} p(s) n_h(s) = \sum_{i \in \mathcal{U}_h} \left(\sum_{s \ni i} p(s) \right) = \sum_{i \in \mathcal{U}_h} \pi_i = \frac{n N_h \bar{X}_h}{N \bar{X}}.$$

So $stg(\hat{T}_{HT}, p)$ is only calibrated for $\mathbf{x}_h$ ($h = 1, \ldots, H$) for the 'proportionally' allocated samples where $\mathbf{x}_h$ is an $N \times 1$ column vector whose i^{th} element is equal to

$$\mathbf{x}_{hi} = \begin{cases} x_i & \text{if} \quad i \in \mathcal{U}_h \\ 0 & \quad\text{else.} \end{cases}$$

The conditionally weighted Horvitz-Thompson estimator, $\hat{T}^*_{HT} = \sum_{i \in s} y_i / \pi^*_i$, under the conditional design on $\mathbf{n}$ (with fixed $n_h > 0$ for all h) is calibrated for the $\mathbf{x}_h$'s in the special case of poststratification by the sizes of the x's, which is equivalent to $\mathbf{x}_h \propto \mathbf{1}_h$ for $h = 1, \ldots, H$, provided the unconditional design, p, is invariant to permutations of the population order. In this case $p^*(s)$ is just a stratified random sampling design which implies $\pi^*_i = n_h / N_h$ for all $i \in \mathcal{U}_h$, $h = 1, \ldots, H$, and $\hat{T}^*_{HT} = \sum_{h=1}^{H} I_h N_h \bar{y}_h$. We will now demonstrate this with the (order independent) Tillé procedure and give a counterexample using the order dependent Chao's scheme.

Consider a population of size 5 and a sample of size 3 with the size variable

$$\mathbf{x}^t = (6, 6, 4, 4, 4).$$

Table 2.3 gives the unconditional design probabilities given by Tillé's procedure and Chao's scheme on $\mathbf{x}$.

Table 2.3: Unconditional design given by Tillé's procedure and Chao's scheme with $N = 5$ and $n = 3$

$\mathcal{S}$	Tillé's $p(s)$	Chao's $p(s)$
$\{1, 2, 3\}$	1/6	1/5
$\{1, 2, 4\}$	1/6	1/5
$\{1, 2, 5\}$	1/6	2/15
$\{1, 3, 4\}$	1/12	1/20
$\{1, 3, 5\}$	1/12	1/12
$\{1, 4, 5\}$	1/12	1/12
$\{2, 3, 4\}$	1/12	1/20
$\{2, 3, 5\}$	1/12	1/12
$\{2, 4, 5\}$	1/12	1/12
$\{3, 4, 5\}$	0	1/30

Now suppose we poststratify by the values in $\mathbf{x}$ so that the poststrata sizes are $\mathbf{N} = (2, 3)$ and suppose the observed sampling configuration is $\mathbf{n} = (1, 2)$. The conditional design on $\mathbf{n}$ is given in table 2.4 and the first order conditional inclusion probabilities are given in table 2.5 for both Tillé's procedure and Chao's scheme.

Table 2.4: Conditional design given by Tillé's procedure and Chao's scheme with $\mathbf{N} = (2, 3)$ and $\mathbf{n} = (1, 2)$

$\mathcal{S}_{\mathbf{n}}$	Tillé's $p^*(s)$	Chao's $p^*(s)$
$\{1, 3, 4\}$	$1/6$	$3/26$
$\{1, 3, 5\}$	$1/6$	$5/26$
$\{1, 4, 5\}$	$1/6$	$5/26$
$\{2, 3, 4\}$	$1/6$	$3/26$
$\{2, 3, 5\}$	$1/6$	$5/26$
$\{2, 4, 5\}$	$1/6$	$5/26$

Table 2.5: π_i^*'s given by Tillé's procedure and Chao's scheme with $\mathbf{N} = (2, 3)$ and $\mathbf{n} = (1, 2)$

Unit i	Poststratum	Tillé's π_i^*	Chao's scheme π_i^*
1	1	1/2	1/2
2		1/2	1/2
3		2/3	8/13
4	2	2/3	8/13
5		2/3	10/13

We see from tables 2.4 and 2.5 that the conditional design given by Tillé's procedure is equivalent to a stratified random sample on $\mathbf{n}$ and it follows that the strategy $stg(\hat{T}^*_{HT}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ reduces to $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ since the conditional inclusion probabilities $\pi^*_i = n_h/N_h$ for $h = 1,\ 2$. For Chao's scheme the conditional selection probabilities from table 2.4 are not the same as a stratified random sample and the conditional inclusion probabilities are not constant within poststratum 2 so $stg(\hat{T}^*_{HT}, \text{Chao} \mid \mathcal{S}_{\mathbf{n}})$ does not reduce to $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$.

The following theorem gives some conditions for the selection of units in a domain to be independent of the selection of units from other domains.

Theorem 4 *Consider a poststratification of a population into H poststrata and a $\pi PS(\mathbf{x})$ design, which is invariant to permutations of the population order. Suppose $\mathbf{x}_h \propto \mathbf{1}_h$ for some poststratum $h = 1,\ \ldots,\ H$, then*

$$\pi^*_i = \frac{n_h}{N_h} \quad \text{and} \quad \pi^*_{ij} = \pi^*_i \pi^*_j$$

for all $i \in \mathcal{U}_h$ and some $j \in \mathcal{U}_g$ $(g \neq h)$ $g = 1,\ \ldots,\ H$.

Proof For a πPS($\mathbf{x}$) design which is invariant by permutation of the population order, if $\mathbf{x}_h \propto \mathbf{1}_h$ for some poststratum h then π^*_i must be constant for all $i \in \mathcal{U}_h$. By lemma 4 part (a) it is easily seen that for fixed n_h we have $\pi^*_i = n_h/N_h$ for all $i \in \mathcal{U}_h$. Also for some unit $j \in \mathcal{U}_g$ $(g \neq h)$ the sets of pairs $\{x_{i_1}, x_j\}$, $\{x_{i_2}, x_j\}$, $\ldots$, $\{x_{i_{N_h}}, x_j\}$ for $i_1, i_2, \ldots, i_{N_h} \in \mathcal{U}_h$ are all equal to one another since $\mathbf{x}_h \propto \mathbf{1}_h$ for some poststratum h. So by the invariance property the conditional joint inclusion probabilities π^*_{ij}, for some $j \in \mathcal{U}_g$ $(g \neq h)$, must be constant over all $i \in \mathcal{U}_h$. Hence, by lemma 4 part (c) we have

$$n_h \pi^*_j = \sum_{i \in \mathcal{U}_h} \pi^*_{ij} = N_h \pi^*_{ij}$$

which implies $\pi^*_{ij} = \pi^*_i \pi^*_j$ for some $j \in \mathcal{U}_g$ and all $i \in \mathcal{U}_h$. □

The following example illustrates that, although it may not be true in general (see Smith, 1991), the selection of units from different poststrata can still be independent for unequal probability sampling designs even if the inclusion probabilities are not constant within poststrata.

Example 8 *Consider a population of size 12 and a sample size 4 where Chao's scheme and Tillé's procedure are applied to the population size measure*

$$\mathbf{x}^t = (43, 58, 49, 42, 53, 46, 57, 14, 17, 21, 38, 30).$$

After calculating the design probabilities, i.e. the fourth order inclusion probabilities, we partitioned the population into two groups where the poststrata sizes are $\mathbf{N} = (7, 5)$ *with the first 7 units of* $\mathcal{U}$ *belonging to poststratum 1 and the rest belonging to poststratum 2. Conditioning on the sampling configuration* $\mathbf{n} = (2, 2)$ *we calculated the matrix of joint conditional inclusion probabilities whose* ij^{th} *entry is equal to* π^*_{ij}, $\pi^*_{ii} = \pi^*_i$, *for all* $i, j \in \mathcal{U}$.

Under Chao's scheme this matrix is equal to (all values rounded to three decimal places)

$$
\left(\begin{array}{ccccccc|ccccc}
.247 & .050 & .038 & .030 & .044 & .037 & .048 & .062 & .074 & .090 & .146 & .122 \\
.050 & .333 & .057 & .048 & .062 & .050 & .065 & .083 & .100 & .122 & .197 & .165 \\
.038 & .057 & .282 & .037 & .051 & .043 & .055 & .070 & .084 & .103 & .167 & .139 \\
.030 & .048 & .037 & .241 & .042 & .036 & .047 & .060 & .072 & .088 & .143 & .119 \\
.044 & .062 & .051 & .042 & .305 & .046 & .060 & .076 & .091 & .111 & .180 & .150 \\
.037 & .050 & .043 & .036 & .046 & .264 & .052 & .066 & .079 & .096 & .156 & .131 \\
.048 & .065 & .055 & .047 & .060 & .052 & .328 & .082 & .098 & .120 & .194 & .162 \\
\hline
.062 & .083 & .070 & .060 & .076 & .066 & .082 & .249 & .035 & .046 & .097 & .071 \\
.074 & .100 & .084 & .072 & .091 & .079 & .098 & .035 & .300 & .056 & .120 & .088 \\
.090 & .122 & .103 & .088 & .111 & .096 & .120 & .046 & .056 & .365 & .152 & .112 \\
.146 & .197 & .167 & .143 & .180 & .156 & .194 & .097 & .120 & .152 & .592 & .223 \\
.122 & .165 & .139 & .119 & .150 & .131 & .162 & .071 & .088 & .112 & .223 & .494
\end{array}\right)
$$

where the matrix partition indicates the within and between poststratum values. It can be verified that $\pi^*_{ij} = \pi^*_i \pi^*_j$ *for all* $i \in \mathcal{U}_1$ *and all* $j \in \mathcal{U}_2$ *so independence holds in this case.*

The matrix of joint conditional inclusion probability under Tillé's elimination procedure is equal to (all values rounded to three decimal places)

$$\left(\begin{array}{ccccccc|ccccc}
.226 & .049 & .035 & .024 & .041 & .030 & .047 & .053 & .064 & .079 & .143 & .113 \\
.049 & .358 & .061 & .047 & .069 & .055 & .077 & .083 & .101 & .125 & .227 & .179 \\
.035 & .061 & .282 & .033 & .052 & .040 & .059 & .066 & .080 & .098 & .178 & .140 \\
.024 & .047 & .033 & .217 & .039 & .029 & .045 & .051 & .061 & .076 & .137 & .108 \\
.041 & .069 & .052 & .039 & .315 & .047 & .067 & .074 & .089 & .110 & .200 & .158 \\
.030 & .055 & .040 & .029 & .047 & .254 & .053 & .059 & .072 & .089 & .161 & .127 \\
.047 & .077 & .059 & .045 & .067 & .053 & .349 & .081 & .099 & .122 & .221 & .175 \\
\hline
.053 & .083 & .066 & .051 & .074 & .059 & .081 & .233 & .019 & .035 & .108 & .072 \\
.064 & .101 & .080 & .061 & .089 & .072 & .099 & .019 & .283 & .046 & .131 & .087 \\
.079 & .125 & .098 & .076 & .110 & .089 & .122 & .035 & .046 & .350 & .162 & .108 \\
.143 & .227 & .178 & .137 & .200 & .161 & .221 & .108 & .131 & .162 & .633 & .233 \\
.113 & .179 & .140 & .108 & .158 & .127 & .175 & .072 & .087 & .108 & .233 & .500
\end{array}\right).$$

Again we observe that the selection of units is independent between poststrata.

It may also be of interest to see if the $\pi_i^* \propto x_i$ over all $i \in \mathcal{U}_h$, $h = 1, \ldots, H$. If this was true then $stg(\hat{T}^*_{HT}, p \mid \mathcal{S}_{\mathbf{n}})$ would be calibrated for any linear combination of the $\mathbf{x}_h$'s.

Example 9 *We will use Chao's scheme and Tillé's procedure as in example 8 on the same size variable with the same stratification and sampling allocation. Since there are no self-selecting units to begin with, if $\pi_i^* \propto x_i$ over all $i \in \mathcal{U}_h$ for $h = 1, \ldots, H$ then it follows from lemma 4 part (a) that*

$$\pi_i^* = p_i^* = \frac{n_h x_i}{N_h \bar{X}_h}$$

for all $i \in \mathcal{U}_h$.

Table 2.6 gives the first order conditional inclusion probabilities from both sampling *schemes as well as the values of* p_i^* *for all units in the population (values rounded to three decimal places).*

Table 2.6: π_i^* for Chao's scheme and Tillé's procedure for example 9

Unit	Poststratum	Chao π_i^*	Tillé π_i^*	p_i^*
1		0.247	0.226	0.247
2		0.333	0.358	0.333
3		0.282	0.282	0.282
4	1	0.241	0.217	0.241
5		0.305	0.315	0.305
6		0.264	0.254	0.264
7		0.328	0.349	0.328
8		0.249	0.233	0.233
9		0.300	0.283	0.283
10	2	0.365	0.350	0.350
11		0.592	0.633	0.633
12		0.494	0.500	0.500

We see that the conditional inclusion probabilities under Chao's scheme are proportional to the size variable over those units in poststratum 1 but not for poststratum 2, and the conditional inclusion probabilities under Tillé's procedure are proportional to the size variable over those units in poststratum 2 but not for poststratum 1.

The conditional inclusion probabilities are not quite proportional to the size variable over those units within poststratum 2 for Chao's scheme and poststratum 1 for Tillé's procedure. Plots of the conditional inclusion probabilities given by Chao's scheme against the p_i^*'s for those units in poststratum 2 are given in figure 2.1. Figure 2.2 gives a plot of the conditional inclusion probabilities given by Tillé's procedure against the p_i^*'s for those units in poststratum 1.

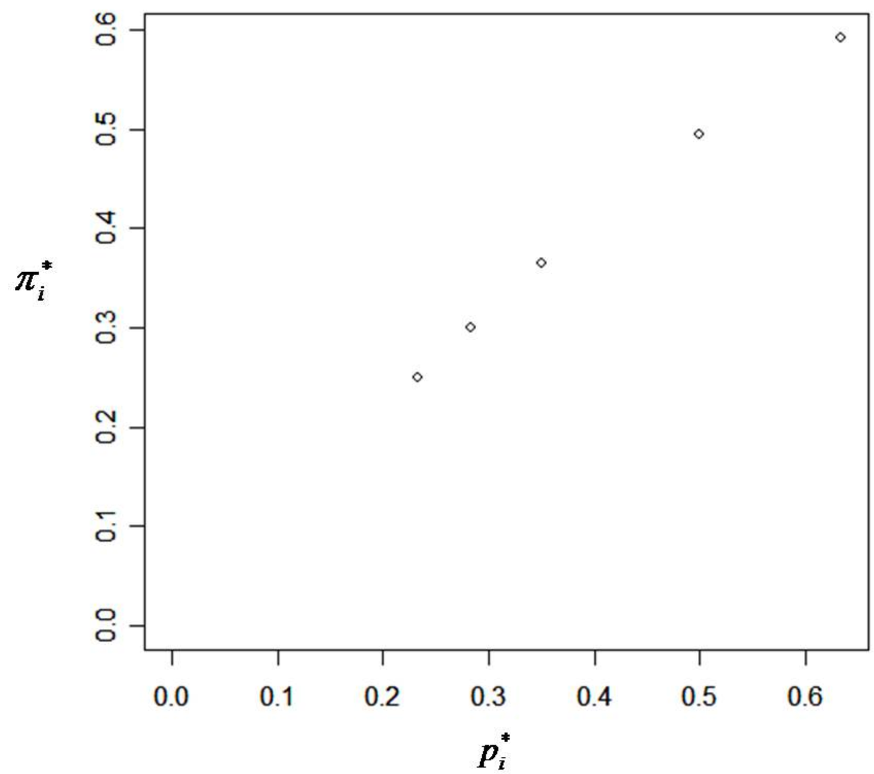

Figure 2.1: Chao's π_i^* against p_i^* over units in poststratum 2

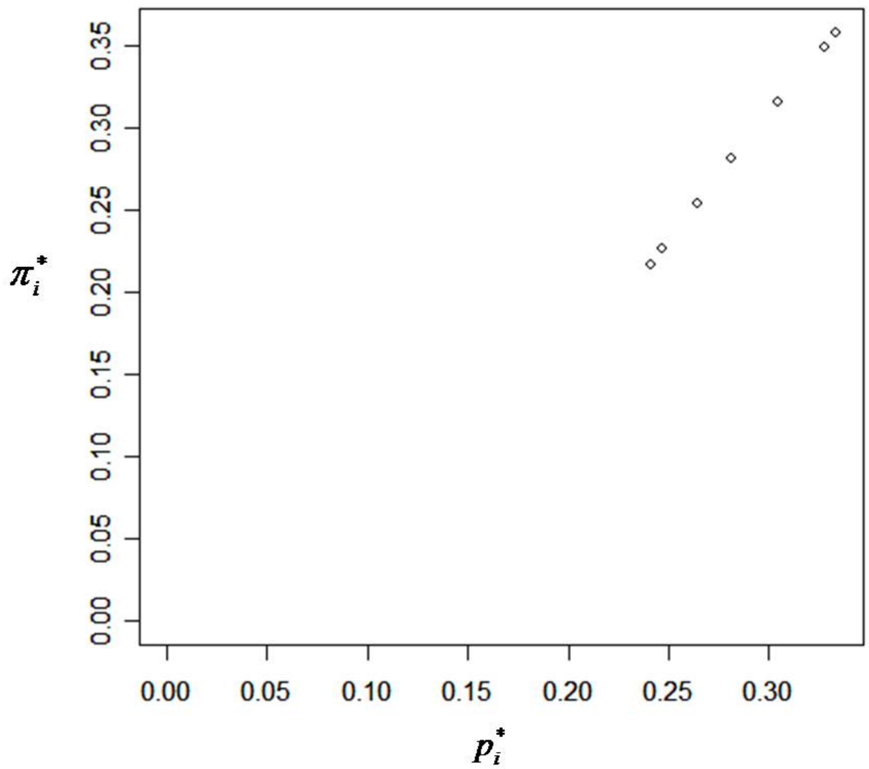

Figure 2.2: Tillé's π_i^* against p_i^* over units in poststratum 1

In figure 2.1 we see that the π_i^*'s given by Chao's scheme and p_i^* have a strong linear relationship closely through the origin over all units in $\mathcal{U}_2$. We also have a straight line relation in figure 2.2 for the π_i^*'s given by Tillé's procedure against p_i^* over all units in $\mathcal{U}_1$, but not through the origin. Hence it would be more appropriate to use Chao's scheme over Tillé's procedure for this size variable if we had observed that the variable of interest is more closely related to $\mathbf{x}$ *within poststrata.*

Example 9 illustrates that in general the conditionally weighted Horvitz-Thompson estimator under the conditional design might not be calibrated for the size variable $\mathbf{x}_h$ for all $h = 1, \ldots, H$, but it can be close to being calibrated.

If we are to consider an estimator that is calibrated for the $\mathbf{x}_h$'s we could use any of the following estimators as an alternative to $\hat{T}^*_{HT}$:

a) The general separate ratio estimator

$$\hat{T}_{GRs} = \sum_{h=1}^{H} I_h N_h \bar{X}_h \frac{\sum_{i \in s_h} w_i y_i}{\sum_{i \in s_h} w_i x_i}$$

where $w_i = 1/\pi_i$ or $1/\pi_i^*$ for (asymptotic) unbiasedness.

b) The separate mean of ratios estimator

$$\hat{T}_{Rsm} = \sum_{h=1}^{H} I_h N_h \bar{X}_h \left(\frac{1}{n_h} \sum_{i \in s_h} \frac{y_i}{x_i} \right).$$

For $\pi_i = n x_i / N\bar{X}$ this estimator can be written as

$$\hat{T}_{Rsm} = \sum_{h=1}^{H} I_h \frac{N_h \bar{X}_h n}{n_h N \bar{X}} \sum_{i \in s_h} \frac{y_i}{\pi_i}.$$

Note that for a 'proportionally' allocated sample $\frac{N_h \bar{X}_h n}{n_h N \bar{X}} = 1$ and this estimator reduces to $\hat{T}_{HT}$ which is not conditionally unbiased.

2.7 Conclusions

The conditional unequal probability design on the observed sample allocation is known provided the values of the size variable and the poststrata members are known for all units in the population. This allows us to make conditional inferences.

The conditional weighted Horvitz-Thompson estimator is conditionally unbiased provided the first order conditional inclusion probabilities are all positive. It is also calibrated for any linear combination of the $\boldsymbol{\pi}_h^*$'s whereas $\hat{T}_{HT}$ is calibrated only for $c\boldsymbol{\pi}$ for some constant c. So $\hat{T}_{HT}^*$ can be viewed as being less restricted than $\hat{T}_{HT}$ as it is exact for a wider range of vectors in $\mathcal{R}^N$.

The conditional first order inclusion probabilities π_i^*'s are not, in general, proportional to the size variable $\mathbf{x}$ and the within poststratum conditional inclusion probabilities $\boldsymbol{\pi}_h^*$ are not proportional to $\mathbf{x}_h$, for $h = 1, \ldots, H$. This implies that $\hat{T}_{HT}^*$ is not calibrated for the $\mathbf{x}_h^*$'s in general. However our example shows that the linear relationship between $\boldsymbol{\pi}_h^*$ and $\mathbf{x}_h$ can be strong so $\hat{T}_{HT}^*$ can be considered if the survey variable is strongly related to the $\mathbf{x}_h$'s.

It is true that in general the selection of units from different poststrata is not independent under the conditional design but it is possible for it to be independent as illustrated in our example.

Chapter 3

Mean Square Error Matrix

In this chapter we consider the matrix given by the mean square error of a general linear strategy. We cover some results for pairs of nonnegative definite symmetric matrices and interpret them in a survey sampling setting. We will use the matrix results to give bounds on the relative efficiency of calibrated strategies, compare variance approximation formulae and also give an exact upper bound for the absolute bias ratio of a strategy.

3.1 Introduction

By writing the expected value using (2.4) and the variance and mean square error as in (2.2) and (2.3), respectively, Sugden & Smith (2007) showed that these measures for any general linear strategy can be written respectively as:

$$\begin{aligned}
\mathrm{E}(\hat{T}_s, p) &= \sum_{i \in \mathcal{U}} B_i y_i = \mathbf{B}^t \mathbf{y} \\
\mathrm{Var}(\hat{T}_s, p) &= \mathbf{y}^t \mathbf{V}(\hat{T}_s, p) \mathbf{y} \\
\mathrm{MSE}(\hat{T}_s, p) &= \mathbf{y}^t \mathbf{M}(\hat{T}_s, p) \mathbf{y}
\end{aligned}$$

where the i^{th} element of the $N \times 1$ vector $\mathbf{B}$ is equal to $B_i = \sum_{s \ni i} p(s) b_{si}$, the ij^{th} element of the $N \times N$ variance-covariance matrix $\mathbf{V}(\hat{T}_s, p)$ is equal to

$$\mathbf{V}(\hat{T}_s, p)_{ij} = \begin{cases} \sum\limits_{s \ni i} p(s) b_{si}^2 - B_i^2 & \text{for} \quad i = j \in \mathcal{U} \\ \sum\limits_{s \ni i,j} p(s) b_{si} b_{sj} - B_i B_j & \text{for} \quad i \neq j \in \mathcal{U} \end{cases} \tag{3.1}$$

and the mean square error matrix $\mathbf{M}(\hat{T}_s, p)$ is equal to

$$\mathbf{M}(\hat{T}_s, p) = \mathbf{V}(\hat{T}_s, p) + (\mathbf{B} - \mathbf{1}_N)(\mathbf{B} - \mathbf{1}_N)^t.$$

The matrix $(\mathbf{B} - \mathbf{1}_N)(\mathbf{B} - \mathbf{1}_N)^t$ is called the bias square matrix of a strategy and the ij^{th} element of $\mathbf{M}(\hat{T}_s, p)$ is equal to

$$\mathbf{M}(\hat{T}_s, p)_{ij} = \begin{cases} \sum\limits_{s \ni i} p(s) b_{si}^2 - 2B_i + 1 & \text{for} \quad i = j \in \mathcal{U} \\ \sum\limits_{s \ni i,j} p(s) b_{si} b_{sj} - B_i - B_j + 1 & \text{for} \quad i \neq j \in \mathcal{U}. \end{cases} \tag{3.2}$$

Note that the mean square error matrix, $\mathbf{M}(\hat{T}_s, p)$, reduces to $\mathbf{V}(\hat{T}_s, p)$ if the strategy is unbiased for T_Y.

If $stg(\hat{T}_s, p)$ is a separate general linear strategy then the ij^{th} element of its mean square error matrix is given by

$$\mathbf{M}(\hat{T}_s, p)_{ij} = \begin{cases} \sum\limits_{s_h \ni i} p_h(s_h) b_{s_h i}^2 - 2B_{hi} + 1 & \text{for} \quad i = j \in \mathcal{U}_h \\ \sum\limits_{s_h \ni i,j} p_h(s_h) b_{s_h i} b_{s_h j} - B_{hi} - B_{hj} + 1 & \text{for} \quad i \neq j \in \mathcal{U}_h \\ \sum\limits_{\substack{s_h \ni i \\ s_g \ni j}} p_{hg}(s_h, s_g) b_{s_h i} b_{s_g j} - B_{hi} - B_{gj} + 1 & \text{for} \quad i \in \mathcal{U}_h \ \& \ j \in \mathcal{U}_g \end{cases} \tag{3.3}$$

where h, $g = 1, \ldots, H$ $(h \neq g)$.

Remark 4 *The expected value and mean square error of a general linear strategy are homogeneous linear functions of degree one and two, respectively, of the population vector* $\mathbf{y}$ *where a function* $f(\mathbf{y})$ *is said to be homogeneous linear of degree* k *if*

$$f(\alpha\mathbf{y}) = \alpha^k f(\mathbf{y})$$

for all nonzero $\alpha \in \mathcal{R}$ *and* $\mathbf{y} \in \mathcal{R}^N$.

In general, these matrices depend on the sampling design p and the survey weights of the estimator and can be calculated when the design and the population auxiliary vectors are known. For any general linear strategy that requires information on the (post)strata (e.g. a separate general linear estimator), we need to know the (post)stratum membership for all units in the population in order to calculate the variance-covariance and mean square error matrices. This is because the survey weights $b_{s_h i}$ will depend on knowledge of the (post)stratum members.

We will show in this chapter and the next how these matrices can be used to analyse and compare different strategies. But first we give a lemma from Sugden & Smith (2007).

Lemma 11 *A general linear strategy* $stg(\hat{T}_s, p)$ *for the population total* T_Y *is calibrated for vectors* $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_q$ *if and only if*

$$(\mathbf{B} - \mathbf{1}_N)^t\mathbf{X} = \mathbf{0}_q \quad \textit{and} \quad \mathbf{V}(\hat{T}_s, p)\mathbf{X} = \mathbf{0}.$$

This lemma is an interesting result because the columns of $\mathbf{X}$ are spanned by the normalized eigenvectors of $\mathbf{M}(\hat{T}_s, p)$ (and $\mathbf{V}(\hat{T}_s, p)$) with corresponding eigenvalues equal to zero. So the eigenspace of the zero eigenvalue of $\mathbf{M}(\hat{T}_s, p)$, i.e. the nullspace of $\mathbf{M}(\hat{T}_s, p)$ which will be denoted by $\mathcal{N}(\mathbf{M}(\hat{T}_s, p))$, contains all the vectors for which the strategy $stg(\hat{T}_s, p)$ is calibrated. This means that we can easily obtain the vector space that contains precisely those vectors for which a general linear strategy is calibrated. More importantly if the nullspaces of matrices (given by the mean square error of different strategies) are the same, then those strategies must be equally calibrated.

Definition 17 *Two or more strategies are said to be equally calibrated if the nullspaces of the mean square error matrix for each strategy are exactly the same.*

In the next section we cover some matrix results that can be applied to the mean square error matrices of equally calibrated strategies.

3.2 Pairs of nonnegative definite symmetric matrices

The mean square error matrix of a strategy is a real nonnegative definite symmetric matrix. It is therefore of interest to study nonnegative definite real symmetric matrices in general and interpret any results in a survey sampling setting. In this section we give some theorems for nonnegative definite symmetric matrices that can be applied to the mean square error matrices of general linear strategies.

The following theorem is a special case of a more general result by Rao & Mitra (1971, p.121) where they gave conditions for two hermitian matrices to be simultaneously diagonalised.

Theorem 5 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $N \times N$ *real symmetric matrices and let* $\mathbf{B}$ *be nonnegative definite with* $\mathcal{N}(\mathbf{B}) \subseteq \mathcal{N}(\mathbf{A})$ *and* $\text{rank}(\mathbf{B}) = r \leq N$*. Then there exists a* *real nonsingular $N \times N$ matrix $\mathbf{T}$ such that*

$$\mathbf{T}^t\mathbf{A}\mathbf{T} = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \textit{and} \quad \mathbf{T}^t\mathbf{B}\mathbf{T} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{\Lambda}$ is a diagonal matrix and $\mathbf{I}_r$ is the $r \times r$ identity matrix.

By assuming $\mathcal{N}(\mathbf{B}) \subseteq \mathcal{N}(\mathbf{A})$ and following Rao & Mitra's (1971, p.121) proof for the general case it can be verified that $\mathbf{T}$ is a real matrix.

The following definition will be useful in explaining a remark about the corollary at the end of this section.

Definition 18 *Let two $N \times N$ real symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, with $\mathbf{B}$ being nonnegative definite, satisfy the equation*

$$\mathbf{A}\mathbf{w} = \lambda\mathbf{B}\mathbf{w}$$

for some vector $\mathbf{w} \in \mathcal{R}^N$ such that $\mathbf{B}\mathbf{w} \neq \mathbf{0}_N$ and λ is a real number. Then λ is called a proper eigenvalue of $\mathbf{A}$ with respect to $\mathbf{B}$ with corresponding proper eigenvector $\mathbf{w}$.

From theorem 5, let $\lambda_1, \lambda_2, \ldots, \lambda_r$ be the diagonal elements of the $r \times r$ matrix $\mathbf{\Lambda}$ and let $\mathbf{t}_i$ be the i^{th} column of $\mathbf{T}$ for $i = 1, \ldots, r$. If we premultiply $\mathbf{T}^t\mathbf{A}\mathbf{T}$ and $\mathbf{T}^t\mathbf{B}\mathbf{T}$ by $(\mathbf{T}^{-1})^t$ we have

$$(\mathbf{T}^{-1})^t\mathbf{T}^t\mathbf{A}\mathbf{T} = \mathbf{A}\mathbf{T} = (\mathbf{T}^{-1})^t \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

and

$$(\mathbf{T}^{-1})^t\mathbf{T}^t\mathbf{B}\mathbf{T} = \mathbf{B}\mathbf{T} = (\mathbf{T}^{-1})^t \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

But

$$\mathbf{B}\mathbf{T}\begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = (\mathbf{T}^{-1})^t \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = (\mathbf{T}^{-1})^t \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

hence

$$\mathbf{AT} = \mathbf{BT}\begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

which implies

$$\mathbf{At}_i = \lambda_i \mathbf{Bt}_i$$

for $i = 1, \ldots, r$. i.e. λ_i is the proper eigenvalue of $\mathbf{A}$ with respect to $\mathbf{B}$ and its corresponding proper eigenvector is $\mathbf{t}_i$ for $i = 1, \ldots, r$. For more details about eigenvalues and eigenvectors of matrix $\mathbf{A}$ with respect to a nonnegative definite matrix $\mathbf{B}$ see Rao & Mitra's (1971, p.124-127).

Now we consider the following definition and a theorem of Rao & Mitra (1971, p.125).

Definition 19 *Let* $\mathbf{A}$ *be an* $N \times m$ *matrix. A generalized inverse matrix or g-inverse matrix of* $\mathbf{A}$*, denoted by* $\mathbf{A}^-$*, is such that*

$$\mathbf{A} = \mathbf{A}\mathbf{A}^-\mathbf{A}.$$

It can easily be shown that a g-inverse matrix will always exists for a real symmetric matrix $\mathbf{B}$ of rank r say. For example consider the Spectral Theorem (see Rao, C.R. 1967, p.36 or any standard linear algebra text book) which says there exists an orthogonal $N \times N$ matrix $\mathbf{P}$ such that

$$\mathbf{P}^t\mathbf{BP} = \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

where $\mathbf{D}$ is an $r \times r$ diagonal matrix of full rank. Then by choosing

$$(\mathbf{P}^t\mathbf{BP})^- = \begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

it is clear that

$$(\mathbf{P}^t\mathbf{BP})(\mathbf{P}^t\mathbf{BP})^-(\mathbf{P}^t\mathbf{BP}) = \mathbf{P}^t\mathbf{BP}$$

which implies

$$\mathbf{B}[\mathbf{P}(\mathbf{P}^t\mathbf{B}\mathbf{P})^-\mathbf{P}^t]\mathbf{B} = \mathbf{B}$$

and hence we can choose

$$\mathbf{B}^- = \mathbf{P}(\mathbf{P}^t\mathbf{B}\mathbf{P})^-\mathbf{P}^t.$$

For the existence of g-inverses for general matrices see Bapat (2001).

Theorem 6 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be nonnegative definite real* $N \times N$ *symmetric matrices such that* $\mathcal{N}(\mathbf{B}) \subseteq \mathcal{N}(\mathbf{A})$. *Then the proper eigenvalues of* $\mathbf{A}$ *with respect to* $\mathbf{B}$ *are the same as the eigenvalues of* $\mathbf{AB}^-$ *for any choice of the g-inverse* $\mathbf{B}^-$.

In their proof of this theorem, Rao & Mitra (1971, p.125) showed that if $\mathbf{w}$ is a proper eigenvector of $\mathbf{A}$ with respect to $\mathbf{B}$ with corresponding proper eigenvalue λ then $\mathbf{w} = \mathbf{B}^-\mathbf{v}$ where $\mathbf{v}$ is the eigenvector of $\mathbf{AB}^-$ with corresponding eigenvalue λ. They also showed that the eigenvector of $\mathbf{AB}^-$ with corresponding eigenvalue λ will be equal to $\mathbf{v} = \mathbf{Bw}$.

The following result will be useful for comparing different strategies and its proof can be found in Gabler (1990, p.109).

Corollary 1 *Let the matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *be as in theorem 5. Then we have the following bounds, which hold for all vectors* $\mathbf{y} \notin \mathcal{N}(\mathbf{A})$, *on the ratio of the quadratic forms*

$$\lambda_{min} \leq \frac{\mathbf{y}^t\mathbf{A}\mathbf{y}}{\mathbf{y}^t\mathbf{B}\mathbf{y}} \leq \lambda_{max} \qquad (3.4)$$

where λ_{min} *and* λ_{max} *are the respective (nonzero) minimum and maximum eigenvalues of* $\mathbf{AB}^-$.

Note that the upper bound in (3.4) will hold for all $\mathbf{y} \notin \mathcal{N}(\mathbf{B})$ since for any $\mathbf{y} \in \mathcal{N}(\mathbf{A})\backslash\mathcal{N}(\mathbf{B})$ the denominator of (3.4) is nonzero but the numerator is equal to zero. Hence, the minimum lower bound on (3.4) for every $\mathbf{y} \in \mathcal{N}(\mathbf{A}) \backslash \mathcal{N}(\mathbf{B})$ is zero but the maximum upper bound is still λ_{max}. When $\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{B})$ the bounds on the ratio in (3.4) will hold for any $\mathbf{y} \notin \mathcal{N}(\mathbf{B})$.

Remark 5 *The ratio in (3.4) achieves the value λ_{min} or λ_{max} when the population vector $\mathbf{y}$ is equal to their corresponding proper eigenvectors of $\mathbf{A}$ with respect to $\mathbf{B}$, respectively. This means that the bounds given by (3.4) are the best possible, based on the matrices $\mathbf{A}$ and $\mathbf{B}$, since they are attainable.*

Corollary 1 can clearly be applied to the mean square error matrices of strategies that are equally calibrated to obtain bounds on their relative mean square error. We will do this in the next section. Also note from corollary 1 that

$$\lambda_{max} \leq 1 \quad \Rightarrow \quad \mathbf{y}^t \mathbf{A} \mathbf{y} \leq \mathbf{y}^t \mathbf{B} \mathbf{y}$$

for all $\mathbf{y} \notin \mathcal{N}(\mathbf{B})$. So the quadratic form $\mathbf{y}^t \mathbf{A} \mathbf{y}$ will always be less than $\mathbf{y}^t \mathbf{B} \mathbf{y}$ for all possible vectors $\mathbf{y} \notin \mathcal{N}(\mathbf{B})$ if the maximum proper eigenvalue of $\mathbf{A}$ with respect to $\mathbf{B}$ is less than one. Note also that $\lambda_{min} = 1/\lambda_{max}$.

3.3 Applications of nonnegative definite symmetric matrices

In this section we apply the results for pairs of nonnegative definite symmetric matrices from section 3.2 to the mean square error matrices of general linear strategies and interpret results in a survey sampling setting.

It is clear that Corollary 1 can be applied to obtain upper and lower bounds on the relative mean square error of any pair of general linear strategies that are equally calibrated. We will illustrate this now in the following example.

Example 10 *For a population size 20 and a sample of size 4, consider the regression estimator $\hat{T}_{REG}$ (see ch.1, p.23) under two different designs. Let p_a be a sampling design that assigns zero probability to those samples such that the matrix $\mathbf{X}^t\mathbf{W}\mathbf{L}_s\mathbf{X}$ is singular and a constant positive probability to all other samples. i.e.*

$$p_a(s) = \begin{cases} 0 & \textit{if } \mathbf{X}^t\mathbf{W}\mathbf{L}_s\mathbf{X} \textit{ is singular} \\ \frac{1}{M} & \textit{otherwise} \end{cases}$$

where M is equal to the number of samples of fixed size n whose matrix $\mathbf{X}^t\mathbf{W}\mathbf{L}_s\mathbf{X}$ is nonsingular. This design will insure that the regression estimator will always be calibrated for any population vector that is a linear combination of the columns of $\mathbf{X}$. For estimators in the class (1.1) it is possible to obtain negative survey weights for some units in a sample. This may be undesirable and so in a similar way as before we define another sampling design, p_b, that avoids such samples and also attains the calibration property for $\hat{T}_{REG}$. i.e. let

$$p_b(s) = \begin{cases} 0 & \textit{if } \mathbf{X}^t\mathbf{W}\mathbf{L}_s\mathbf{X} \textit{ is singular or } b_{si} \leq 0 \textit{ for at least one unit in } s \\ \frac{1}{K} & \textit{otherwise} \end{cases}$$

where K is equal to the number of samples whose matrix $\mathbf{X}^t\mathbf{W}\mathbf{L}_s\mathbf{X}$ is nonsingular and whose survey weights of the estimator are positive for all units in s. We will compare the regression estimator under p_a and p_b. Let the auxiliary matrix be

$$\mathbf{X} = \begin{pmatrix} 1 & 44 \\ 1 & 44 \\ 1 & 45 \\ 1 & 43 \\ 1 & 40 \\ 1 & 52 \\ 1 & 43 \\ 1 & 47 \\ 1 & 54 \\ 1 & 43 \\ 1 & 51 \\ 1 & 50 \\ 1 & 61 \\ 1 & 47 \\ 1 & 62 \\ 1 & 34 \\ 1 & 51 \\ 1 & 48 \\ 1 & 51 \\ 1 & 57 \end{pmatrix}.$$

The nullspace of the mean square error matrices for both strategies are the same so the strategies are equally calibrated. Using Corollary 1 we find that their relative mean square error is bounded by

$$0.9250166 \leq \frac{\text{MSE}(\hat{T}_{REG}, p_a)}{\text{MSE}(\hat{T}_{REG}, p_b)} = \frac{\mathbf{y}^t\mathbf{M}(\hat{T}_{REG}, p_a)\mathbf{y}}{\mathbf{y}^t\mathbf{M}(\hat{T}_{REG}, p_b)\mathbf{y}} \leq 2.97573$$

and

$$0.336052 \leq \frac{\text{MSE}(\hat{T}_{REG}, p_b)}{\text{MSE}(\hat{T}_{REG}, p_a)} = \frac{\mathbf{y}^t\mathbf{M}(\hat{T}_{REG}, p_b)\mathbf{y}}{\mathbf{y}^t\mathbf{M}(\hat{T}_{REG}, p_a)\mathbf{y}} \leq 1.081062$$

for all $\mathbf{y} \in \mathcal{R}^N$ *that are not in the nullspaces of their mean square error matrix. This indicates that* $\text{MSE}(\hat{T}_{REG}, p_b)$ *can be at most* 8.1% *larger than* $\text{MSE}(\hat{T}_{REG}, p_a)$*, and* $\text{MSE}(\hat{T}_{REG}, p_a)$ *can be at most* 197.6% *larger than* $\text{MSE}(\hat{T}_{REG}, p_b)$*, hence it may be more appropriate to use* $stg(\hat{T}_{REG}, p_b)$ *since it can be much better or just as good as* $stg(\hat{T}_{REG}, p_a)$*.*

For equally calibrated strategies, $stg(\hat{T}_1, p_1)$ and $stg(\hat{T}_2, p_2)$, say, we define the maximum relative efficiency of $stg(\hat{T}_1, p_1)$ over $stg(\hat{T}_2, p_2)$ as

$$\text{MRE}(\hat{T}_1, p_1 \mid \hat{T}_2, p_2) = \max_{\text{MSE}(\hat{T}_2, p_2) \neq 0} \left\{ \frac{\text{MSE}(\hat{T}_1, p_1)}{\text{MSE}(\hat{T}_2, p_2)} \right\}.$$

Similarly let $\text{mre}(\hat{T}_1, p_1 \mid \hat{T}_2, p_2)$ denote the minimum nonzero value this ratio can take. Then for two equally calibrated strategies we have the following properties:

a) $\text{MRE}(\hat{T}_1, p_1 \mid \hat{T}_2, p_2) = 1/\text{mre}(\hat{T}_2, p_2 \mid \hat{T}_1, p_1)$

b) $\text{MRE}(\hat{T}_1, p_1 \mid \hat{T}_2, p_2) \leq 1 \Leftrightarrow \text{MSE}(\hat{T}_1, p_1) \leq \text{MSE}(\hat{T}_2, p_2)$

c) $\text{mre}(\hat{T}_1, p_1 \mid \hat{T}_2, p_2) \geq 1 \Leftrightarrow \text{MSE}(\hat{T}_1, p_1) \geq \text{MSE}(\hat{T}_2, p_2)$

Example 11 *The maximum relative efficiency has been computed to compare the following strategies that are equally calibrated:*

1. *The ratio estimator under simple random sampling:* $stg(\hat{T}_R, SRS)$

2. *The ratio estimator under probability proportional to aggregate size (PPAS) sampling due to Midzuno (1952),* $stg(\hat{T}_R, PPAS)$*, where* $p(s) = n\bar{x}/M_1 N\bar{X}$ *and* $M_1 = \binom{N-1}{n-1}$*. This strategy is unbiased for* T_Y *which can be verified by using lemma 6 as follows:*

$$B_i = \sum_{s \ni i} p(s) b_{si} = \sum_{s \ni i} \frac{n\bar{x}}{M_1 N\bar{X}} \frac{N\bar{X}}{n\bar{x}} = \frac{1}{M_1} \sum_{s \ni i} 1 = \frac{1}{M_1} M_1 = 1$$

for all $i \in \mathcal{U}$*.*

3. The Horvitz-Thompson estimator under Chao's scheme for unequal probability sampling: $stg(\hat{T}_{HT}, Chao)$

4. The Horvitz-Thompson estimator under Tillé's procedure for unequal probability sampling: $stg(\hat{T}_{HT}, Tillé)$

The auxiliary (or size) vector, $\mathbf{x}$ *used in this example is from Agarwal & Kumar (1998), data: A1-20, and is given as*

$$\mathbf{x}^t = (44, 44, 45, 43, 40, 52, 43, 47, 54, 43, 51, 50, 61, 47, 62, 34, 51, 48, 51, 57).$$

Table 3.1 contains the values of maximum relative efficiency for each pair of strategies.

Table 3.1: Maximum relative efficiency for strategies: 1, 2, 3, and 4 where the ij^{th} entry is equal to MRE(i^{th} strategy | j^{th} strategy) for example 11

MRE($i \mid j$)	$stg(\hat{T}_R, \text{SRS})$	$stg(\hat{T}_R, \text{PPAS})$	$stg(\hat{T}_{HT}, \text{Chao})$	$stg(\hat{T}_{HT}, \text{Tillé})$
$stg(\hat{T}_R, \text{SRS})$	1	1.1915	1.8289	1.7814
$stg(\hat{T}_R, \text{PPAS})$	1.0137	1	1.8498	1.8034
$stg(\hat{T}_{HT}, \text{Chao})$	2.7649	3.0620	1	1.2671
$stg(\hat{T}_{HT}, \text{Tillé})$	2.9198	3.2162	1.0929	1

Table 3.1 shows that $stg(\hat{T}_R, PPAS)$ may be preferred to $stg(\hat{T}_R, SRS)$ for this size variable as its mean square error can be at most only 1.37% larger than that of $stg(\hat{T}_R, SRS)$ but the mean square error of $stg(\hat{T}_R, SRS)$ can be 19.2% larger than that of $stg(\hat{T}_R, PPAS)$. The Horvitz-Thompson estimator under Tillé's procedure may be preferred over $stg(\hat{T}_{HT}, Chao)$ for similar reasons. The choice between strategies with estimator $\hat{T}_R$ and strategies with estimator $\hat{T}_{HT}$ is not clear here as their maximum relative efficiencies are very large.

Tillé (2006, p.143) compared the variances of the sampling sum $\sum_{i\in s} y_i$ under different πPS($\mathbf{x}$) sampling designs on a fixed size variable $\mathbf{x}$,

$$\sum_{i\in\mathcal{U}} \pi_i(1-\pi_i)y_i^2 + \sum_{i\in\mathcal{U}} \sum_{\substack{j\in\mathcal{U} \\ i\neq j}} (\pi_{ij} - \pi_i\pi_j)y_i y_j = \mathbf{y}^t\mathbf{\Pi}\mathbf{y}$$

where the ij^{th} element of the $N \times N$ matrix $\mathbf{\Pi}$ is equal to $\pi_{ij} - \pi_i\pi_j$ and $\pi_{ii} = \pi_i$ for i, $j = 1, \ldots, N$. For sampling designs with fixed sample size n, Tillé defined the largest possible deviation between two sampling designs, denoted by LPD$(1 \mid 2)$, as

$$\text{LPD}(1 \mid 2) = \max_{\mathbf{y}:\boldsymbol{\pi}^t\mathbf{y}\neq 0} \left\{ \frac{\mathbf{y}^t\mathbf{\Pi}_1\mathbf{y}}{\mathbf{y}^t\mathbf{\Pi}_2\mathbf{y}} \right\} - 1$$

where the $N \times N$ matrices $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are given by the sampling variance of two different designs, p_1 and p_2. Tillé used this technique to compare different designs in the same way as we have with the maximum relative efficiency for equally calibrated strategies in examples 10 and 11.

We can also apply Corollary 1 to compare a strategy with different sample sizes to see whether there is any gain in precision as the sample size increases, i.e. check for consistency (see ch.1 p.21), or to see how large a sample needs to be in order to use a large sample approximation formula for the variance. Berger (2005) used Corollary 1 to examine an approximation to the true variance of the Horvitz-Thompson estimator under Chao's scheme. Hájek (1981, p.157) called an approximation of MSE$(\hat{T}_s, p)$

tight if the nullspace of the matrix given by the approximation formula is the same as that of the mean square error matrix of $stg(\hat{T}_s, p)$. Hájek also gave the same upper bound on the ratio of two nonnegative definite quadratic forms as in Corollary 1 to compare a tight approximation formula for $\text{MSE}(\hat{T}_s, p)$ with the exact mean square error, and gave a class of tight approximation for $\text{MSE}(\hat{T}_s, p)$.

Many approximation formulae for the mean square error of a general linear strategy can be written as a quadratic form in $\mathbf{y}$. For the Hájek-Basu estimator, in a discussion of Basu (1971),

$$\hat{T}_{HB} = N\frac{\sum_{i\in s} y_i/\pi_i}{\sum_{i\in s} 1/\pi_i}$$

under some design p, the large sample approximation formula for its mean square error (derived using a Taylor series expansion, see Thompson 2002 p.74) is given by

$$\text{Var}(\hat{T}_{HB}, p) \approx \sum_{i\in\mathcal{U}} \frac{(y_i - \bar{Y})^2(1-\pi_i)}{\pi_i} + \sum_{i\in\mathcal{U}}\sum_{\substack{j\in\mathcal{U}\\ i\neq j}} \frac{(y_i - \bar{Y})(y_j - \bar{Y})(\pi_{ij} - \pi_i\pi_j)}{\pi_i\pi_j}.$$

This can be written as a quadratic form in $\mathbf{y}$ as follows,

$$\begin{aligned} \text{Var}(\hat{T}_{HB}, p) &\approx (\mathbf{y} - \bar{Y}\mathbf{1}_N)^t\mathbf{W}(\mathbf{y} - \bar{Y}\mathbf{1}_N) \\ &= \mathbf{y}^t(\mathbf{I}_N - \mathbf{J}_N/N)\mathbf{W}(\mathbf{I}_N - \mathbf{J}_N/N)\mathbf{y} \\ &= \mathbf{y}^t\mathbf{U}\mathbf{y}, \end{aligned} \quad (3.5)$$

where $\mathbf{J}_N$ is the $N \times N$ matrix of all ones and the ij^{th} element of the $N \times N$ matrix $\mathbf{W}$ is equal to

$$\mathbf{W}_{ij} = \begin{cases} \frac{1-\pi_i}{\pi_i} & \text{for} \quad i = j \\ \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} & \text{for} \quad i \neq j \end{cases} \quad (3.6)$$

for i, $j = 1, \ldots, N$ and $\mathbf{U} = (\mathbf{I}_N - \mathbf{J}_N/N)\mathbf{W}(\mathbf{I}_N - \mathbf{J}_N/N)$. Note that the $N \times N$ matrix $\mathbf{W}$, given by (3.6), is equal to the $N \times N$ matrix given by the variance of the Horvitz-Thompson estimator (see Tillé (2006), p.28). From (3.5) we see that the vector $\mathbf{y} - \bar{Y}\mathbf{1}_N$ will equal $\mathbf{0}_N$ whenever the population vector $\mathbf{y}$ is equal to a constant vector in which case $\mathbf{y}^t\mathbf{U}\mathbf{y} = 0$. This means that the constant vector belongs to the nullspace of $\mathbf{U}$. In general the nullspace of $\mathbf{U}$ will be the same as the nullspace of the mean square error matrix of $stg(\hat{T}_{HB}, p)$. Therefore we can compare the approximation formula with the exact mean square error of this strategy using Corollary 1.

Example 12 *Consider the Hájek-Basu estimator under Chao's scheme for probability proportional to size sampling using the size vector* $\mathbf{x}$ *from example 11 with a sample size equal to 4. The mean square error matrix given by* $\mathrm{MSE}(\hat{T}_{HB}, Chao)$ *and the matrix* $\mathbf{U}$ *given by the large sample variance approximation formula denoted by* $\mathrm{AVar}(\hat{T}_{HB}, Chao)$ *share the same nullspace. By corollary 1 we obtain*

$$0.97 \leq \frac{\mathrm{MSE}(\hat{T}_{HB}, Chao)}{\mathrm{AVar}(\hat{T}_{HB}, Chao)} \leq 1.07$$

and

$$0.93 \leq \frac{\mathrm{AVar}(\hat{T}_{HB}, Chao)}{\mathrm{MSE}(\hat{T}_{HB}, Chao)} \leq 1.03.$$

Thus the difference between the exact mean square error and approximation formula is small (less than 10%) which suggests that a sample size of 4 is sufficiently large so that the approximation formula can be used.

In the next section we will show how the matrix result of corollary 1 can be used to assess the bias of a strategy.

3.4 The absolute bias ratio of a strategy

The bias of a strategy is always of interest to the statistician since it could potentially have devastating effects on our inferences if it is large. A measure that can indicate the effects of the bias of a strategy on the coverage of a confidence interval for the total T_Y is the absolute bias ratio.

Definition 20 *The absolute bias ratio, denoted by* $\mathrm{ABR}(\hat{T}_s, p)$*, of a strategy* $stg(\hat{T}_s, p)$ *for* T_Y *is equal to*

$$\mathrm{ABR}(\hat{T}_s, p) = \frac{|\mathrm{bias}(\hat{T}_s, p)|}{\sqrt{\mathrm{Var}(\hat{T}_s, p)}}.$$

Cochran (1977, p.14) and Särndal *et al.* (1992, p.165) did simulation studies which suggested that the bias had little effect on the coverage of confidence intervals when the absolute bias ratio is less than 0.1. Many strategies that are calibrated might not be unbiased, and vice versa, and so it would be of interest to know the absolute bias ratio of such strategies.

For the ratio estimator $\hat{T}_R = N\bar{X}\hat{R}$, where $\hat{R} = \bar{y}/\bar{x}$ is an estimator for $R = \bar{Y}/\bar{X}$, under simple random sampling it has been shown by Hartely & Ross (1954) that the covariance of $\hat{R}$ and $\bar{x}$ is equal to

$$\begin{aligned}\mathrm{Cov}(\hat{R}, \bar{x}, \mathrm{SRS}) &= \mathrm{E}(\bar{y}, \mathrm{SRS}) - \mathrm{E}(\hat{R}, \mathrm{SRS})\mathrm{E}(\bar{x}, \mathrm{SRS}) \\ &= \bar{Y} - \mathrm{E}(\hat{R}, \mathrm{SRS})\bar{X} \\ &= -\frac{\mathrm{bias}(\hat{T}_R, \mathrm{SRS})}{N}. \end{aligned} \tag{3.7}$$

Provided $\bar{X} > 0$ and using equation (3.7) Hartely & Ross obtained their famous result for the absolute bias ratio of this strategy,

$$\mathrm{ABR}(\hat{T}_R, \mathrm{SRS}) \leq \mathrm{cv}(\bar{x}, \mathrm{SRS}). \tag{3.8}$$

Equality in (3.8) holds when the correlation between $\hat{R}$ and $\bar{x}$ over all $s \in \mathcal{S}$ is equal to 1.

The inequality (3.8) allows us to assess the limit and magnitude of the bias ratio of $stg(\hat{T}_R, \mathrm{SRS})$ at the design stage. Särndal *el al.* (1992, p.177) considered the general ratio estimator

$$\hat{T}_{GR} = N\bar{X}\frac{\hat{T}_{HT,\mathbf{y}}}{\hat{T}_{HT,\mathbf{x}}} = N\bar{X}\hat{R}_\pi$$

where $\hat{R}_\pi = \hat{T}_{HT,\mathbf{y}}/\hat{T}_{HT,\mathbf{x}}$, $\hat{T}_{HT,\mathbf{y}} = \sum_{i\in s} y_i/\pi_i$ and $\hat{T}_{HT,\mathbf{x}} = \sum_{i\in s} x_i/\pi_i$. They presented a similar result to (3.8) for the case of an arbitrary design p, which was given as

$$\mathrm{ABR}(\hat{T}_{GR}, p) \leq \mathrm{cv}(\hat{T}_{HT,\mathbf{x}}, p). \tag{3.9}$$

The inequality (3.8) has also been generalized by Meng (1993) to that of the separate ratio estimator, denoted by $\hat{T}_{Rs}$, under stratified random sampling where

$$\hat{T}_{Rs} = \sum_{h=1}^{H} I_h N_h \bar{X}_h \frac{\bar{y}_h}{\bar{x}_h} = \sum_{h=1}^{H} I_h \hat{T}_{R,h}, \quad \text{with} \quad \hat{T}_{R,h} = N_h \bar{X}_h \frac{\bar{y}_h}{\bar{x}_h}$$

and $\bar{X}_h > 0$ for all $h = 1, \ldots, H$. The upper bound on the absolute bias ratio of $stg(\hat{T}_{Rs}, \mathrm{StRS})$ was given as

$$\mathrm{ABR}(\hat{T}_{Rs}, \mathrm{StRS}) \leq \left(\sum_{h=1}^{H} [\mathrm{cv}(\bar{x}_h, \mathrm{StRS})]^2 \right)^{\frac{1}{2}}.$$

We can generalize this result for the separate general ratio estimator

$$\hat{T}_{GRs} = \sum_{h=1}^{H} I_h N_h \bar{X}_h \frac{\hat{T}_{HT,\mathbf{y}_h}}{\hat{T}_{HT,\mathbf{x}_h}} = \sum_{h=1}^{H} I_h \hat{T}_{GR,h} \quad \text{with} \quad \hat{T}_{GR,h} = N_h \bar{X}_h \frac{\hat{T}_{HT,\mathbf{y}_h}}{\hat{T}_{HT,\mathbf{x}_h}},$$

for $h = 1, \ldots, H$, under any arbitrary stratified design (with independent selection of units from different strata).

Theorem 7 *Provided $\bar{X}_h > 0$ for all $h = 1, \ldots, H$, an upper bound for the absolute bias ratio of the separate general ratio estimator under any arbitrary stratified design p_{st}, say, with $n_h > 0$ for all $h = 1, \ldots, H$ is given by*

$$\mathrm{ABR}(\hat{T}_{GRs}, p_{st}) \leq \left(\sum_{h=1}^{H} [\mathrm{cv}(\hat{T}_{HT,\mathbf{x}_h}, p_{st})]^2 \right)^{\frac{1}{2}}.$$

Proof Observe that

$$\mathrm{bias}(\hat{T}_{GRs}, p_{st}) = \sum_{h=1}^{H} \mathrm{E}(\hat{T}_{GR,h}, p_{st}) - \sum_{h=1}^{H} N_h \bar{X}_h \frac{\bar{Y}_h}{\bar{X}_h} = \sum_{h=1}^{H} \mathrm{bias}(\hat{T}_{GR,h}, p_{st}).$$

Now

$$\begin{aligned}\text{Cov}(\hat{T}_{GR,h}, \hat{T}_{HT,\mathbf{x}_h}, p_{st}) &= N_h\bar{X}_h\text{E}(\hat{T}_{HT,\mathbf{y}_h}, p_{st}) - \text{E}(\hat{T}_{HT,\mathbf{x}_h}, p_{st})\text{E}(\hat{T}_{GR,h}, p_{st}) \\ &= N_h^2\bar{X}_h\bar{Y}_h - N_h\bar{X}_h\text{E}(\hat{T}_{GR,h}, p_{st})\end{aligned}$$

so it can easily be seen that

$$\text{bias}(\hat{T}_{GR,h}, p_{st}) = -\frac{\text{Cov}(\hat{T}_{GR,h}, \hat{T}_{HT,\mathbf{x}_h}, p_{st})}{N_h\bar{X}_h}$$

and hence

$$\text{bias}(\hat{T}_{GRs}, p_{st}) = -\sum_{h=1}^{H}\frac{\text{Cov}(\hat{T}_{GR,h}, \hat{T}_{HT,\mathbf{x}_h}, p_{st})}{N_h\bar{X}_h}.$$

Since

$$|\text{Cov}(\hat{T}_{GR,h}, \hat{T}_{HT,\mathbf{x}_h}, p_{st})| \leq \sqrt{\text{Var}(\hat{T}_{GR,h}, p_{st})}\sqrt{\text{Var}(\hat{T}_{HT,\mathbf{x}_h}, p_{st})}$$

and $\bar{X}_h > 0$ for each $h = 1, \ldots, H$ it follows that

$$|\text{bias}(\hat{T}_{GRs}, p_{st})| \leq \sum_{h=1}^{H}\sqrt{\text{Var}(\hat{T}_{GR,h}, p_{st})}\text{cv}(\hat{T}_{HT,\mathbf{x}_h}, p_{st})$$

and by the Cauchy-Schwarz inequality

$$\sum_{h=1}^{H}\sqrt{\text{Var}(\hat{T}_{GR,h}, p_{st})}\text{cv}(\hat{T}_{HT,\mathbf{x}_h}, p_{st}) \leq \left(\sum_{h=1}^{H}\text{Var}(\hat{T}_{GR,h}, p_{st})\right)^{\frac{1}{2}}\left(\sum_{h=1}^{H}[\text{cv}(\hat{T}_{HT,\mathbf{x}_h}, p_{st})]^2\right)^{\frac{1}{2}}.$$

Because the selection of units from different strata are independent

$$\text{Var}(\hat{T}_{GRs}, p_{st}) = \sum_{h=1}^{H}\text{Var}(\hat{T}_{GR,h}, p_{st}),$$

and so it follows that

$$\text{ABR}(\hat{T}_{GRs}, p_{st}) = \frac{|\text{bias}(\hat{T}_{GRs}, p_{st})|}{\sqrt{\text{Var}(\hat{T}_{GRs}, p_{st})}} \leq \left(\sum_{h=1}^{H}[\text{cv}(\hat{T}_{HT,\mathbf{x}_h}, p_{st})]^2\right)^{\frac{1}{2}}.$$

□

It can easily be seen that the upper bound on $\mathrm{ABR}(\hat{T}_{GRs}, p_{st})$ from theorem 7 reduces to the upper bound for $\mathrm{ABR}(\hat{T}_{GR}, p)$ given by (3.9) when $H = 1$.

When poststratifying under unequal probability sampling, theorem 7 can still be applied to estimators of the same form as $\hat{T}_{GRs}$ under the conditional design on $\mathbf{n}$ provided that the selection of units from different poststrata are independent. But in general it is not true that they are independent.

Another application of corollary 1 is to calculate an exact upper bound, whose value can be achieved for some $\mathbf{y} \in \mathcal{R}^N$, for the absolute bias ratio of a general linear strategy.

Theorem 8 *Let the $N \times N$ matrices $\mathbf{D} = (\mathbf{B} - \mathbf{1}_N)(\mathbf{B} - \mathbf{1}_N)^t$ and $\mathbf{V}(\hat{T}_s, p)$ be , respectively, the bias square matrix and variance-covariance matrix of some general linear strategy $stg(\hat{T}_s, p)$. Provided $\mathcal{N}(\mathbf{V}(\hat{T}_s, p)) \subseteq \mathcal{N}(\mathbf{D})$, an exact upper bound of the absolute bias ratio of $stg(\hat{T}_s, p)$ for every $\mathbf{y} \notin \mathcal{N}(\mathbf{V}(\hat{T}_s, p))$ is given by*

$$\mathrm{ABR}(\hat{T}_s, p) \leq \sqrt{\lambda}.$$

where λ is the largest eigenvalue of the matrix $\mathbf{D}\mathbf{V}(\hat{T}_s, p)^-$.

Proof Observe that

$$\mathrm{ABR}(\hat{T}_s, p) = \left| \sqrt{\frac{[\mathrm{bias}(\hat{T}_s, p)]^2}{\mathrm{Var}(\hat{T}_s, p)}} \right| = \left| \sqrt{\frac{\mathbf{y}^t\mathbf{D}\mathbf{y}}{\mathbf{y}^t\mathbf{V}(\hat{T}_s, p)\mathbf{y}}} \right| .$$

If $\mathcal{N}(\mathbf{V}(\hat{T}_s, p)) \subseteq \mathcal{N}(\mathbf{D})$, then by corollary 1

$$\frac{[\mathrm{bias}(\hat{T}_s, p)]^2}{\mathrm{Var}(\hat{T}_s, p)} = \frac{\mathbf{y}^t\mathbf{D}\mathbf{y}}{\mathbf{y}^t\mathbf{V}(\hat{T}_s, p)\mathbf{y}} \leq \lambda$$

over every $\mathbf{y} \notin \mathcal{N}(\mathbf{V}(\hat{T}_s, p))$ where λ is the largest eigenvalue of $\mathbf{D}\mathbf{V}(\hat{T}_s, p)^-$ and so it follows that

$$\mathrm{ABR}(\hat{T}_s, p) \leq \sqrt{\lambda}.$$

That completes the proof. □

Remark 6 *The upper bound in theorem 8 is attained when the population vector* $\mathbf{y}$ *is proportional to the proper eigenvector of* $\mathbf{D}$ *with respect to* $\mathbf{V}(\hat{T}_s, p)$ *corresponding to the proper eigenvalue* λ.

Note that if the upper bound of the absolute bias ratio is very large this does not imply that the actual absolute bias ratio is large. The usefulness of this upper bound is when the upper bound is small.

Many strategies used in practice satisfy the conditions of theorem 8.

If the estimator is biased, then the matrix $\mathbf{D} = (\mathbf{B} - \mathbf{1}_N)(\mathbf{B} - \mathbf{1}_N)^t$ from theorem 8 is of rank 1 and therefore has one nonzero eigenvalue equal to $\sum_{i \in \mathcal{U}}(B_i - 1)^2$ with corresponding normalized eigenvector $(\mathbf{B} - \mathbf{1}_N)/\sqrt{\sum_{i \in \mathcal{U}}(B_i - 1)^2}$. If the condition of theorem 8, i.e. $\mathcal{N}(\mathbf{V}(\hat{T}_s, p)) \subseteq \mathcal{N}(\mathbf{D})$, was violated then there exist a vector $\mathbf{a} \in \mathcal{N}(\mathbf{V}(\hat{T}_s, p)) \backslash \mathcal{N}(\mathbf{D})$ such that

$$\mathbf{a}^t \mathbf{V}(\hat{T}_s, p)\mathbf{a} = 0 \quad \text{and} \quad \mathbf{a}^t \mathbf{D} \mathbf{a} > 0.$$

Hence, any strategy that does not satisfy the condition of theorem 8 can have zero variance but nonzero bias. This 'instability' property of such strategies is undesirable as the bias can be arbitrarily large. Furthermore, we cannot obtain an upper bound on the absolute bias ratio of these strategies which is independent of the y's.

In the following examples we compare Särndal, Swensson and Wretman's (SSW) upper bound on $\text{ABR}(\hat{T}_{GR}, p)$ given in (3.9) with the exact method of theorem 8. We also compare Meng's method with the exact method for calculating the upper bound on $\text{ABR}(\hat{T}_{Rs}, \text{StRS})$.

Example 13 *We will compare the upper bound on the absolute bias ratio given by SSW in (3.9) with the exact upper bound given by theorem 8 for different strategies.*

Table 3.2 gives the strategies and values for the upper bounds for various size variables whose description are given in table 3.3.

Table 3.2: Upper bounds on the absolute bias ratio for example 13

x	strategy	n	SSW upper bound	Exact upper bound
1	$stg(\hat{T}_R, \text{SRS})$	4	0.1600	0.1560
		5	0.1339	0.1316
2	$stg(\hat{T}_R, \text{SRS})$	4	0.0638	0.0636
		5	0.0553	0.0551
3	$stg(\hat{T}_{HB}, \text{Chao})$	2	0.0979	0.0975
		3	0.0784	0.0782
		4	0.0666	0.0665
3	$stg(\hat{T}_{HB}, \text{Tillé})$	2	0.0980	0.0974
		3	0.0786	0.0781
		4	0.0668	0.0663
4	$stg(\hat{T}_{HB}, \text{Chao})$	2	0.2928	0.2856
		3	0.2353	0.2291
		4	0.2002	0.1953
4	$stg(\hat{T}_{HB}, \text{Tillé})$	2	0.2912	0.2841
		3	0.2326	0.2264
		4	0.1969	0.1922
		5	0.1719	0.1684
		6	0.1531	0.1503
		7	0.1380	0.1359
5	$stg(\hat{T}_{HB}, \text{Tillé})$	2	0.9185	0.8973
		3	0.7419	0.7191
		4	0.6355	0.6159
		5	0.5620	0.5465
		6	0.5072	0.4952

Table 3.3: Description of the size vectors, $\mathbf{x}$, for examples 13, 14 & 15. The population for size vectors 3 & 4 was partitioned into two groups where $\mathcal{U}_1 = \{1, \ldots, 17\}$ and $\mathcal{U}_2 = \{18, \ldots, 33\}$, the population for size variable 5 was partitioned into two groups where $\mathcal{U}_1 = \{1, \ldots, 25\}$ and $\mathcal{U}_2 = \{26, \ldots, 49\}$, the population for size vector 6 was stratified by sex and the population for size vectors 7, 8 & 9 was stratified by geographical region

$\mathbf{x}$	Source	cv($\mathbf{x}$)	N
1	Example 8 p.83	0.3919	12
2	Agarwal & Kumar (1998): A1-20	0.1427	20
3	Cochran (1977, p.34): Family income,	0.1458	33
4	Cochran (1977, p.34): Family size,	0.4095	33
5	Cochran (1977, p.152): City size,	1.0122	49
6	Brewer (2002, p.298): Estimated weights of Basu's elephants	0.1186	50
7	Särndal *et al.* (1992, p.652): MU284 pop., CS82	0.5428	284
8	Särndal *et al.* (1992, p.652): MU284 pop., SS82	0.3268	284
9	Särndal *et al.* (1992, p.652): MU284 pop., S82	0.2325	284

Table 3.2 shows that the upper bound given by the inequality in (3.9) is not attainable. There are very little differences however between the values of the two bounds on the absolute bias ratio.

Example 14 *We will now compare Meng's method for calculating the upper bound on the absolute bias ratio for the separate ratio estimator under stratified random sampling. Table 3.4 gives the values for the upper bounds when using this strategy with various auxiliary vectors of different sizes described in table 3.3.*

Table 3.4: Upper bounds on the absolute bias ratio for $stg(\hat{T}_{Rs}, \text{StRS})$ in example 14

x	**n**	Meng's upper bound	Exact upper bound
6	(2,2)	0.1144	0.1140
	(3,3)	0.0914	0.0911
	(4,4)	0.0773	0.07712
	(5,5)	0.0675	0.0673
7	(2,2,2,2,2,2,2,2)	0.8280	0.8280
	(3,3,3,3,3,3,3,3)	0.7000	0.6624
	(4,4,4,4,4,4,4,4)	0.5649	0.5647
	(5,5,5,5,5,5,5,5)	0.5217	0.4972
8	(2,2,2,2,2,2,2,2)	0.5852	0.5852
	(3,3,3,3,3,3,3,3)	0.4873	0.4702
	(4,4,4,4,4,4,4,4)	0.4147	0.4012
	(5,5,5,5,5,5,5,5)	0.3642	0.3532
9	(2,2,2,2,2,2,2,2)	0.4405	0.4335
	(3,3,3,3,3,3,3,3)	0.3533	0.3477
	(4,4,4,4,4,4,4,4)	0.3003	0.2958
	(5,5,5,5,5,5,5,5)	0.2634	0.2598

Table 3.4 shows that there is very little difference between the exact upper bound and Meng's upper bound for the absolute bias ratio.

Example 13 & 14 demonstrate that it is appropriate to use SSW and Meng's upper bound on the absolute bias ratio over the exact method since their values are easier to calculate and are very close to the exact value. However the upper bound on the absolute bias ratio given by SSW and Meng or theorem 7 only applies to generalized separate ratio estimator under a design that gives independent selection of units in different strata whereas the exact method given by theorem 8 can be applied to any general linear strategy that satisfy its condition. In the next example we consider a poststratification and calculate the exact upper bound of the absolute conditional bias ratio for the separate means of ratio estimator conditional on the poststrata sample sizes.

Example 15 *Consider the separate means of ratio estimator,*

$$\hat{T}_{Rsm} = \sum_{h=1}^{H} I_h \frac{N_h \bar{X}_h}{n_h} \sum_{i \in s_h} \frac{y_i}{x_i},$$

under a poststratification. This estimator is conditionally biased on $\mathbf{n}$*.*

Table 3.5 gives the exact upper bound on the absolute conditional bias ratio, conditional on $\mathbf{n}$*, using various size vectors under the Tillé procedure for unequal probability sampling.*

Table 3.5: Exact upper bound on the absolute bias ratio for $stg(\hat{T}_{Rsm}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ in example 15

x	**n**	Exact upper bound
3	(2,2)	0.0010
	(3,3)	0.0125
	(4,4)	0.0153
4	(2,2)	0.0277
	(3,3)	0.0350
	(4,4)	0.0425
5	(2,2)	0.0881
	(3,3)	0.1537

Table 3.5 shows that in most cases the bias of stg($\hat{T}_{Rsm}$, Tillé | $\mathcal{S}_{\mathbf{n}}$) may be neglected however, there seems to be a trend where the exact upper bound on the absolute bias ratio of stg($\hat{T}_{Rsm}$, Tillé | $\mathcal{S}_{\mathbf{n}}$) increases with the n_h's.

Consider the following estimator under poststratification with H poststrata,

$$\hat{T}_{pst} = \sum_{h=1}^{H} I_{.h} N_{.h} \frac{\sum_{k=1}^{L} I_{k.} N_{k.} \bar{y}_{kh}}{\sum_{k=1}^{L} I_{k.} N_{k.} \frac{n_{kh}}{n_{k.}}}$$

where the unconditional design is a stratified random sample design with L strata of sizes $\mathbf{N}_{k.} = (N_{1.}, N_{2.}, \ldots, N_{L.})$ such that $\sum_{k=1}^{L} N_{k.} = N$, and strata sample sizes $\mathbf{n}_{k.} = (n_{1.}, n_{2.}, \ldots, n_{L.})$ such that $\sum_{k=1}^{L} n_{k.} = n$ with $I_{k.}$ equal to one if $n_{k.} > 0$ or zero otherwise. Similarly the poststrata sizes are $\mathbf{N}_{.h} = (N_{.1}, N_{.2}, \ldots, N_{.H})$ with $\sum_{h=1}^{H} N_{.h} = N$ and the poststrata sample sizes are $\mathbf{n}_{.h} = (n_{.1}, n_{.2}, \ldots, n_{.H})$ such that $\sum_{h=1}^{H} n_{.h} = n$ with $I_{.h}$ equal to one for $n_{.h} > 0$ or zero otherwise. N_{kh} denotes the number of units in stratum k and poststratum h and is such that $\sum_{h=1}^{H} N_{kh} = N_{k.}$ and $\sum_{k=1}^{L} N_{kh} = N_{.h}$. Similarly n_{kh} is the number of units that fall into the sample stratum k and sample poststratum h with $\sum_{h=1}^{H} n_{kh} = n_{k.}$ and $\sum_{k=1}^{L} n_{kh} = n_{.h}$, and $\bar{y}_{kh}$ is the sample stratum-poststratum mean for stratum k and poststratum h.

Note that $\hat{T}_{pst}$ is calibrated for the constant vector provided $n_{.h} > 0$ for all $h = 1, \ldots, H$.

Rao (1985) illustrated the difficulty of investigating the conditional properties of $stg(\hat{T}_{pst}, \text{StRS})$, when conditioning on the observed values of $\mathbf{n}_{.h}$ and concluded that the strategy is conditionally biased. Since the estimator $\hat{T}_{pst}$ falls into the class of general linear estimators, provided the conditions of theorem 8 are satisfied, we can apply theorem 8 to give the exact upper bound on the absolute conditional bias of $stg(\hat{T}_{pst}, \text{StRS})$ conditional on $\mathbf{n}_{.h}$. We do this now in the following example.

Example 16 *Consider the strategy stg($\hat{T}_{pst}$, StRS) with $H = L = 2$, $N_{kh} = 10$ for* *all h, $g = 1, 2$, $\mathbf{n}_{k.} = (5,3)$ and $\mathbf{n}_{.h} = (4,4)$. The variance-covariance matrix and the bias square matrix of $stg(\hat{T}_{pst}, StRS)$ under the conditional design on $\mathbf{n}_{.h}$ satisfied the condition of theorem 8, and the exact upper bound on the absolute conditional bias ratio of this strategy is given as 0.73333096208. This upper bound is large which indicates that $stg(\hat{T}_{pst}, StRS)$ can be badly conditionally biased, but not necessarily.*

3.5 Conclusions

In this chapter we demonstrated how nonnegative definite symmetric matrices can be used to compare the exact mean square errors of equally calibrated strategies by calculating their maximum relative efficiencies. Our examples show that this method may be more useful when comparing an estimator under different designs. We also used this method to measure the accuracy of a variance approximation formula.

We gave an upper bound for the absolute bias ratio which is exact in the sense that the value of the upper bound can be attained for some vectors in $\mathcal{R}^N$. We compared our upper bound with some standard methods for ratio type estimators. Our studies showed that the standard results were not exact but very close to the exact upper bound. For this reason it is more appropriate to use the standard method since the calculation of the exact method is more complicated, especially for large samples, as the computational burden can be huge. However our method can be applied to a wide range of strategies not just the ratio type.

For a separate general linear strategy under poststratification and unequal probability sampling, it is possible to calculate the exact upper bound on the absolute bias ratio but standard methods will fail if the selection of units from different poststrata are not independent.

Chapter 4

Comparison of Strategies

In this chapter we will consider the exact mean square error of a general linear strategy in the form of a linear combination of eigenvalues given by the mean square error matrix. From this form of the mean square error we give exact upper and lower bounds on it which will be used to prove further results that give sufficient conditions that indicate the superior strategy between two different strategies.

Our results can also be applied to compare approximation formulae discussed in the last section.

4.1 Eigenvalues of certain matrices

In general the eigenvalues of a matrix given by the mean square error of a general linear estimator under an unequal probability design cannot be written in a simple way. However there are special cases where we can obtain an expression for most or all of the eigenvalues of certain matrices. The following lemma is a well known result given as an exercise in Rao, C.R. (1967, p.53).

Lemma 12 *The eigenvalues of the $N \times N$ matrix which is of the form*

$$(a-b)\mathbf{I}_N + b\mathbf{J}_N \tag{4.1}$$

are equal to $a-b$ of multiplicity $N-1$ and $a+(N-1)b$ of multiplicity one.

Example 17 *Consider the expansion estimator $\hat{T}_0 = N\bar{y}$ under simple random sampling whose survey weight is $b_{si} = N/n$ for unit $i \in s$. Then for unit $i \in \mathcal{U}$ we have*

$$B_i = \sum_{s \ni i} p(s) b_{si} = \sum_{s \ni i} \binom{N}{n}^{-1} \frac{N}{n} = \binom{N}{n}^{-1} \binom{N-1}{n-1} \frac{N}{n} = 1,$$

$$\sum_{s \ni i} p(s) b_{si}^2 = \frac{N}{n},$$

and for $i \neq j$ we have

$$\sum_{s \ni i,j} p(s) b_{si} b_{sj} = \sum_{s \ni i,j} \binom{N}{n}^{-1} \frac{N^2}{n^2} = \binom{N}{n}^{-1} \binom{N-2}{n-2} \frac{N^2}{n^2} = \frac{(n-1)N}{(N-1)n}.$$

Hence, by (3.2), the ij^{th} element of the mean square error matrix $\mathbf{M}(\hat{T}_0, SRS)$ is equal to

$$\mathbf{M}(\hat{T}_0, SRS)_{ij} = \begin{cases} \frac{N-n}{n} & \text{for} \quad i = j \in \mathcal{U} \\ -\frac{(N-n)}{n(N-1)} & \text{for} \quad i \neq j \in \mathcal{U} \end{cases}$$

which is of the same form as (4.1) with

$$a = \frac{N-n}{n} \quad \text{and} \quad b = -\frac{(N-n)}{n(N-1)}.$$

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ be the eigenvalues of $\mathbf{M}(\hat{T}_0, SRS)$ with corresponding normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N$ respectively. Then by lemma 12 we have

$$\lambda_l = \begin{cases} \frac{N-n}{n} - \left[\frac{-(N-n)}{n(N-1)}\right] = \frac{N(N-n)}{n(N-1)} & \text{for} \quad l = 1, ..., N-1 \\ \frac{N-n}{n} + \left[\frac{-(N-n)(N-1)}{n(N-1)}\right] = 0 & \text{for} \quad l = N. \end{cases} \tag{4.2}$$

Since the mean square error matrix of a strategy, $\mathbf{M}(\hat{T}_s, p)$, is a real symmetric matrix it can be written as a linear combination of idempotent matrices

$$\mathbf{M}(\hat{T}_s, p) = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^t + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^t + \ldots + \lambda_N \mathbf{e}_N \mathbf{e}_N^t$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ are eigenvalues of $\mathbf{M}(\hat{T}_s, p)$ with corresponding normalized eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\ldots$, $\mathbf{e}_N$, respectively. Hence the mean square error of a general linear strategy can be written as a linear combination of eigenvalues

$$\text{MSE}(\hat{T}_s, p) = \mathbf{y}^t \mathbf{M}(\hat{T}_s, p) \mathbf{y} = \lambda_1 c_1^2 + \lambda_2 c_2^2 + \ldots + \lambda_N c_N^2 \tag{4.3}$$

where $c_l = \mathbf{y}^t \mathbf{e}_l$ for $l = 1, \ldots, N$. The Spectral Theorem shows that the normalized eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\ldots$, $\mathbf{e}_N$, which forms an orthonormal set, satisfies

$$\mathbf{e}_1 \mathbf{e}_1^t + \ldots + \mathbf{e}_N \mathbf{e}_N^t = \mathbf{I}_N$$

which implies

$$\sum_{l=1}^{N} c_l^2 = \mathbf{y}^t (\mathbf{e}_1 + \ldots + \mathbf{e}_N)(\mathbf{e}_1 + \ldots + \mathbf{e}_N)^t \mathbf{y} = \sum_{i \in \mathcal{U}} y_i^2 = (N-1) S_Y^2 + N \bar{Y}^2.$$

Then it follows from (4.3) that

$$\lambda_N [(N-1) S_Y^2 + N \bar{Y}^2] \leq \sum_{l=1}^{N} \lambda_l c_l^2 \leq \lambda_1 [(N-1) S_Y^2 + N \bar{Y}^2]. \tag{4.4}$$

From example 17 and (4.3) we can write the mean square error of $stg(\hat{T}_0, \text{SRS})$ as

$$\text{MSE}(\hat{T}_0, \text{SRS}) = \mathbf{y}^t \mathbf{M}(\hat{T}_0, \text{SRS}) \mathbf{y} = \sum_{l=1}^{N} \lambda_l c_l^2 = \frac{N(N-n)}{n(N-1)} \sum_{l=1}^{N-1} c_l^2 \tag{4.5}$$

where $c_l = \mathbf{y}^t \mathbf{e}_l$, $l = 1, \ldots, N$. However, we know from any standard text book in survey sampling theory that the variance of the expansion estimator under simple random sampling is equal to

$$\text{MSE}(\hat{T}_0, \text{SRS}) = N^2 \frac{(1-f)}{n} S_Y^2 = \frac{N(N-n)}{n} S_Y^2 \tag{4.6}$$

where $f = n/N$ is the sampling fraction. Comparing (4.5) and (4.6) we see that the sum of squares of the first $N-1$ coefficients of the eigenvalues is equal to the population corrected sum of squares, i.e.

$$\sum_{l=1}^{N-1} c_l^2 = (N-1)S_Y^2 = \sum_{i\in\mathcal{U}}(y_i - \bar{Y})^2$$

In fact this is true for any general linear strategy which is calibrated only for constant population vectors, and an extension of this is given by the following lemma.

Lemma 13 *Let the general linear strategy $stg(\hat{T}_s, p)$ be calibrated only for vectors that are proportional to some $\mathbf{x} \in \mathcal{R}^N \backslash \{\mathbf{0}\}$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ be the eigenvalues of the mean square error matrix $\mathbf{M}(\hat{T}_s, p)$ with corresponding normalized eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$, ..., $\mathbf{e}_N$, respectively, so that*

$$\mathrm{MSE}(\hat{T}_s, p) = \mathbf{y}^t\mathbf{M}(\hat{T}_s, p)\mathbf{y} = \lambda_1 c_1^2 + \ldots + \lambda_N c_N^2$$

where $c_l = \mathbf{y}^t\mathbf{e}_l$ for $l = 1$, ..., N. Then

$$\sum_{l=1}^{N-1} c_l^2 = \sum_{i\in\mathcal{U}} y_i^2 - B\sum_{i\in\mathcal{U}} y_i x_i$$

where

$$B = \frac{\sum_{i\in\mathcal{U}} y_i x_i}{\sum_{i\in\mathcal{U}} x_i^2}$$

is the population ordinary least squares regression (through the origin) coefficient. Furthermore when $\mathbf{x}$ is a constant vector then

$$\sum_{l=1}^{N-1} c_l^2 = (N-1)S_Y^2 = \sum_{i\in\mathcal{U}}(y_i - \bar{Y})^2.$$

Proof Observe that

$$c_1^2 + c_2^2 + \ldots + c_N^2 = \sum_{i\in\mathcal{U}} y_i^2$$

Since $stg(\hat{T}_s, p)$ is calibrated only for $\mathbf{x}$, by lemma 11 λ_N must be equal to zero and $\mathbf{e}_N$ must be proportional to $\mathbf{x}$. But $\mathbf{e}_N$ is a normalized eigenvector, this implies $\mathbf{e}_N = \mathbf{x}/\sqrt{\mathbf{x}^t\mathbf{x}}$. Hence

$$c_N = \mathbf{y}^t\mathbf{e}_N = \frac{\sum_{i\in\mathcal{U}} y_i x_i}{\sqrt{\sum_{i\in\mathcal{U}} x_i^2}}$$

so that

$$c_N^2 = \frac{(\sum_{i\in\mathcal{U}} y_i x_i)^2}{\sum_{i\in\mathcal{U}} x_i^2} = B\sum_{i\in\mathcal{U}} y_i x_i.$$

Thus we have

$$\begin{aligned} c_1^2 + c_2^2 + \ldots + c_N^2 - c_N^2 &= \sum_{i\in\mathcal{U}} y_i^2 - c_N^2 \\ \Leftrightarrow \quad c_1^2 + c_2^2 + \ldots + c_{N-1}^2 &= \sum_{i\in\mathcal{U}} y_i^2 - B\sum_{i\in\mathcal{U}} y_i x_i. \end{aligned}$$

When $\mathbf{x}$ is a constant vector $B\sum_{i\in\mathcal{U}} y_i x_i$ reduces to $N\bar{Y}^2$ and

$$\sum_{i\in\mathcal{U}} y_i^2 - B\sum_{i\in\mathcal{U}} y_i x_i = \sum_{i\in\mathcal{U}} y_i^2 - N\bar{Y}^2 = (N-1)S_Y^2$$

and that completes the proof. □

Remark 7 *In fact if $stg(\hat{T}_s, p)$ is calibrated only for vectors that are spanned by the column space of the $N \times q$ auxiliary matrix $\mathbf{X}$, then*

$$\sum_{l=1}^{N-q} c_l^2 = \mathbf{y}^t(\mathbf{I}_N - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)\mathbf{y}.$$

From lemma 13 we observe that if $stg(\hat{T}_s, p)$ is calibrated only for constant vectors, then for any $\mathbf{y} \in \mathcal{R}^N$ that is not constant we have

$$\lambda_{N-1}(N-1)S_Y^2 \leq \text{MSE}(\hat{T}_s, p) \leq \lambda_1(N-1)S_Y^2. \tag{4.7}$$

The bounds on $\text{MSE}(\hat{T}_s, p)$ could also be given by Corollary 1, on p.99, since $(N-1)S_Y^2$ can be written as the quadratic form $\mathbf{y}^t(\mathbf{I}_N - \mathbf{J}_N/N)\mathbf{y}$ and the matrix $\mathbf{I}_N - \mathbf{J}_N/N$ is orthogonal only to constant vectors, thus has the same nullspace as the mean square error matrix of any strategy that is calibrated for only constant vectors. This would give

$$\lambda_{N-1} \leq \frac{\text{MSE}(\hat{T}_s, p)}{(N-1)S_Y^2} \leq \lambda_1$$

then multiply through by $(N-1)S_Y^2$ gives us the exact bounds on $\text{MSE}(\hat{T}_s, p)$ since they are attainable when $\mathbf{y}$ is proportional to the proper eigenvectors of $\mathbf{M}(\hat{T}_s, p)$ with respect to $\mathbf{I}_N - \mathbf{J}_N/N$ corresponding to the proper eigenvalues λ_{N-1} and λ_1.

The following well known result, see Padmawar (1998), gives the formulae for the eigenvalues of a particular type of matrix.

Theorem 9 *Let $N_h = N/H = N_\varsigma$ for all $h = 1, \ldots, H$. Then the eigenvalues of an $N \times N$ symmetric matrix $\mathbf{M}$ which is of the form*

$$\mathbf{M} = \begin{pmatrix} \mathbf{D} & d\mathbf{J}_{N_\varsigma} & \cdots & d\mathbf{J}_{N_\varsigma} \\ d\mathbf{J}_{N_\varsigma} & \mathbf{D} & \cdots & d\mathbf{J}_{N_\varsigma} \\ \vdots & \vdots & \ddots & \vdots \\ d\mathbf{J}_{N_\varsigma} & d\mathbf{J}_{N_\varsigma} & \cdots & \mathbf{D} \end{pmatrix} \tag{4.8}$$

where $\mathbf{D} = (a-b)\mathbf{I}_{N_\varsigma} + b\mathbf{J}_{N_\varsigma}$ for $a, b, d \in \mathcal{R}$, are equal to $a - b$ of multiplicity $N - H$ with corresponding eigenspace $\left\langle \mathbf{v}_{11}, \ldots, \mathbf{v}_{1(N_\varsigma-1)}, \ldots, \mathbf{v}_{H1}, \ldots, \mathbf{v}_{H(N_\varsigma-1)} \right\rangle$ where

$$\mathbf{v}_{hl_i} = \begin{cases} v_{hl_i} & \textit{for} \quad i \in \mathcal{U}_h \\ 0 & \textit{else} \end{cases} \tag{4.9}$$

such that $\sum_{i \in \mathcal{U}_h} v_{hl_i} = 0$ and $\mathbf{v}_{hl}^t \mathbf{v}_{gk} = 0$ for h, $g = 1, \ldots, H$ and l, $k = 1, \ldots, N_\varsigma - 1$ ($l \neq k$ when $h = g$); $(a-b) + N_\varsigma(b-d)$ of multiplicity $H-1$ with corresponding eigenspace $\langle \mathbf{u}_1, \ldots, \mathbf{u}_H \rangle$ where

$$\mathbf{u}_{hi} = \begin{cases} 1 & \textit{for} \quad i \in \mathcal{U}_h \\ 0 & \textit{else} \end{cases}$$

for $h = 1, \ldots, H$*; and* $(a-b) + N_\varsigma(b-d) + Nd$ *of multiplicity* 1 *with eigenspace equal to constant vectors.*

Remark 8 *It can easily be seen that for unequal* N_h*'s the vectors* $\mathbf{v}_{h1}, \ldots, \mathbf{v}_{h(N_h-1)}$*, which are defined in a similar way as in (4.9), form an eigenspace of the matrix*

$$\mathbf{M} = \begin{pmatrix} \mathbf{D}_1 & b_{12}\mathbf{J}_{N_1N_2} & \cdots & b_{1H}\mathbf{J}_{N_1N_H} \\ b_{21}\mathbf{J}_{N_2N_1} & \mathbf{D}_2 & \cdots & b_{2H}\mathbf{J}_{N_2N_H} \\ \vdots & \vdots & \ddots & \vdots \\ b_{H1}\mathbf{J}_{N_HN_1} & b_{H2}\mathbf{J}_{N_HN_2} & \cdots & \mathbf{D}_H \end{pmatrix} \tag{4.10}$$

where $\mathbf{D}_h = (a_h - b_h)\mathbf{I}_{N_h} + b_h\mathbf{J}_{N_h}$ *for* $a_h, b_h, b_{hg} \in \mathcal{R}$ $(b_{hg} = b_{gh})$*, with corresponding eigenvalue equal to* $a_h - b_h$ *for* $h = 1, \ldots, H$*.*

Note that in general the matrix given in (4.10) does not have a constant vector as an eigenvector even if $b_{hg} = 0$ for all $h \neq g = 1, \ldots, H$.

Padmawar (1998) applied theorem 9 to the mean square error matrix of an estimator for the population total under a cluster sampling design to find conditions for some estimators for the mean square error to be nonnegative definite and unbiased.

Lemma 14 *Let* $stg(\hat{T}_s, p)$ *be calibrated only for vectors which are constant within strata and let the mean square error matrix* $\mathbf{M}(\hat{T}_s, p)$ *have eigenvalues* $\lambda_1 \geq \ldots \geq \lambda_N$ *with corresponding normalized eigenvectors* $\mathbf{e}_1, \ldots, \mathbf{e}_N$*, respectively, so that*

$$\text{MSE}(\hat{T}_s, p) = \mathbf{y}^t\mathbf{M}(\hat{T}_s, p)\mathbf{y} = \lambda_1 c_1^2 + \ldots + \lambda_N c_N^2$$

where $c_l = \mathbf{y}^t\mathbf{e}_l$ *for* $l = 1, \ldots, N$*. Then*

$$\sum_{l=1}^{N-H} c_l^2 = \sum_{h=1}^{H}(N_h - 1)S_{Y_h}^2.$$

Proof Since $stg(\hat{T}_s, p)$ is calibrated only for vectors that are constant within strata it follows from lemma 11 that $\mathcal{N}(\mathbf{M}(\hat{T}_s, p)) = \langle \mathbf{u}_1, \ldots, \mathbf{u}_H \rangle$ where $\mathbf{u}_h$ is defined as in theorem 9. Hence we can choose $\mathbf{e}_{N-H+l} = \mathbf{u}_l / \sqrt{\mathbf{u}_l^t \mathbf{u}_l}$ for $l = 1, \ldots, H$ since $\lambda_{N-H+1} = \ldots = \lambda_N = 0$. Now

$$\sum_{l=1}^{H} c_{N-H+l}^2 = \sum_{l=1}^{H} (\mathbf{y}^t \mathbf{e}_{N-H+l})^2 = \sum_{h=1}^{H} N_h \bar{Y}_h^2$$

and so

$$\begin{aligned} \sum_{l=1}^{N} c_l^2 - \sum_{l=1}^{H} c_{N-H+l}^2 &= \sum_{i \in \mathcal{U}} y_i^2 - \sum_{h=1}^{H} N_h \bar{Y}_h^2 \\ &= \sum_{h=1}^{H} \left(\sum_{i \in \mathcal{U}_h} y_i^2 - N_h \bar{Y}_h^2 \right) \\ &= \sum_{h=1}^{H} (N_h - 1) S_{Y_h}^2, \end{aligned}$$

that completes the proof. □

Lemma 15 *Suppose the mean square error matrix, $\mathbf{M}(\hat{T}_s, p)$, for some general linear strategy $stg(\hat{T}_s, p)$ was of the same form as (4.10). Let $\mathbf{e}_{hl}$ for $l = 1, \ldots, N_h - 1$ be the normalized eigenvectors of $\mathbf{M}(\hat{T}_s, p)$ with corresponding eigenvalue $a_h - b_h$ such that the i^{th} element of $\mathbf{e}_{hl}$ is equal to zero if $i \notin \mathcal{U}_h$, but nonzero otherwise. Then*

$$\sum_{l=1}^{N_h - 1} c_{hl}^2 = (N_h - 1) S_{Y_h}^2$$

where $c_{hl} = \mathbf{y}^t \mathbf{e}_{hl}$ for $l = 1, \ldots, N_h - 1$ and $h = 1, \ldots, H$.

Proof Let the $N \times 1$ vector $\mathbf{y}_h$ be such that its i^{th} element is

$$\mathbf{y}_{hi} = \begin{cases} y_i & \text{if} \quad i \in \mathcal{U}_h \\ 0 & \text{else.} \end{cases}$$

Since the $\mathbf{e}_{hl}$'s are all orthogonal to $\mathbf{u}_h$ as well as each other and they have nonzero entry for all $i \in \mathcal{U}_h$ but zero otherwise. We can write $\mathbf{y}_h$ as a linear combination of $\mathbf{e}_{h1}, \ldots, \mathbf{e}_{h(N_h-1)}$ and $\mathbf{u}_h$, i.e.

$$\mathbf{y}_h = c_{h1}\mathbf{e}_{h1} + \ldots + c_{h(N_h-1)}\mathbf{e}_{h(N_h-1)} + c_{hN_h}\mathbf{e}_{hN_h}$$

where $\mathbf{e}_{hN_h}$ is the normalized vector of $\mathbf{u}_h$, i.e. $\mathbf{e}_{hN_h} = \frac{1}{\sqrt{N_h}}\mathbf{u}_h$, $c_{hl} = \mathbf{y}_h^t\mathbf{e}_{hl} = \mathbf{y}^t\mathbf{e}_{hl}$ for $l = 1, \ldots, N_h - 1$ and $c_{hN_h} = \mathbf{y}_h^t\mathbf{e}_{hN_h} = \frac{N_h}{\sqrt{N_h}}\bar{Y}_h$. Hence

$$\begin{aligned} \mathbf{y}_h^t\mathbf{y}_h = c_{h1}^2 + \ldots + c_{h(N_h-1)}^2 + N_h\bar{Y}_h^2 &= \sum_{i\in\mathcal{U}_h} y_i^2 \\ \Leftrightarrow \quad c_{h1}^2 + \ldots + c_{h(N_h-1)}^2 &= \sum_{i\in\mathcal{U}_h} y_i^2 - N_h\bar{Y}^2 \\ &= (N_h - 1)S_{Y_h}^2 \end{aligned}$$

for $h = 1, \ldots, H$ and that completes the proof. □

Example 18 *Consider the stratified estimator* $\hat{T}_{st} = \sum_{h=1}^{H} I_h N_h \bar{y}_h$ *under stratified random sampling. For fixed positive* n_h*'s the induced design on the strata under stratified random sampling are given by*

$$p_h(s_h) = \sum_{\{s:s_h=s\cap\mathcal{U}_h\}} \left[\prod_{l=1}^{H}\binom{N_l}{n_l}\right]^{-1} = \left[\prod_{l=1}^{H}\binom{N_l}{n_l}\right]^{-1} \prod_{\substack{g=1\\g\neq h}}^{H}\binom{N_g}{n_g} = \binom{N_h}{n_h}^{-1}$$

for $s_h \in \mathcal{S}_h$, $h = 1, \ldots, H$, *and similarly*

$$p_{hg}(s_h, s_g) = \sum_{\{s:s_h\cup s_g=s\cap(\mathcal{U}_h\cup\mathcal{U}_g)\}} \left[\prod_{l=1}^{H}\binom{N_l}{n_l}\right]^{-1} = \binom{N_h}{n_h}^{-1}\binom{N_g}{n_g}^{-1}$$

for $s_h \cup s_g \in \mathcal{S}_{hg}$, *with* $h, g = 1, \ldots, H$ $(h \neq g)$. *Then*

$$B_{hi} = \sum_{s_h\ni i} p_h(s_h) b_{s_h i} = \sum_{s_h\ni i}\binom{N_h}{n_h}^{-1}\frac{N_h}{n_h} = \frac{n_h}{N_h}\frac{N_h}{n_h} = 1,$$

which indicates the unbiasedness of this strategy, and by (3.3) the mean square error matrix $\mathbf{M}(\hat{T}_{st}, StRS)$ *has* ij^{th} *element equal to*

$$\mathbf{M}(\hat{T}_{st}, StRS)_{ij} = \begin{cases} \sum\limits_{s_h \ni i} \binom{N_h}{n_h}^{-1} \frac{N_h^2}{n_h^2} - 1 & = \frac{N_h - n_h}{n_h} \quad \text{for} \quad i = j \in \mathcal{U}_h \\ \sum\limits_{s_h \ni i,j} \binom{N_h}{n_h}^{-1} \frac{N_h^2}{n_h^2} - 1 & = -\frac{(N_h - n_h)}{n_h(N_h - 1)} \quad \text{for} \quad i \neq j \in \mathcal{U}_h \\ \sum\limits_{\substack{s_h \ni i \\ s_g \ni j}} \binom{N_h}{n_h}^{-1} \binom{N_g}{n_g}^{-1} \frac{N_h N_g}{n_h n_g} - 1 & = 0 \quad \text{for} \quad i \in \mathcal{U}_h \ \& \ j \in \mathcal{U}_g \end{cases}$$

for $h,\ g = 1, \ldots, H$ $(h \neq g)$*. It is easily seen that* $\mathbf{M}(\hat{T}_{st}, StRS)$ *is of the same form as the matrix given by (4.10) with* $a_h = \frac{N_h - n_h}{n_h}$, $b_h = -\frac{(N_h - n_h)}{n_h(N_h - 1)}$ *for* $h = 1, \ldots, H$, *and* $b_{hg} = 0$ *for all* $g \neq h$*. Since* $\mathbf{M}(\hat{T}_{st}, StRS)$ *is a block diagonal matrix with the* h^{th} *block equal to the* $N_h \times N_h$ *matrix* $(a_h - b_h)\mathbf{I}_{N_h} - b_h \mathbf{J}_{N_h}$, *it follows that its eigenvalues are equal to the eigenvalues of* $(a_h - b_h)\mathbf{I}_{N_h} - b_h \mathbf{J}_{N_h}$ *for* $h = 1, \ldots, H$*. Then by lemma 12 the eigenvalues of* $\mathbf{M}(\hat{T}_{st}, StRS)$ *are*

$$\lambda_h = a_h - b_h = \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \tag{4.11}$$

of multiplicity $N_h - 1$ *for* $h = 1, \ldots, H$ *and* $a_h + (N_h - 1)b_h = 0$ *of multiplicity* H *for* $h = 1, \ldots, H$.

Since $stg(\hat{T}_{st}, \text{StRS})$ is calibrated only for constant strata vectors, by using lemma 15 we can write

$$\begin{aligned} \text{MSE}(\hat{T}_{st}, \text{StRS}) = \sum_{l=1}^{N} \lambda_l c_l^2 &= \sum_{h=1}^{H} \lambda_h \sum_{j=1}^{N_h - 1} c_{hj}^2 = \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \sum_{j=1}^{N_h - 1} c_{hj}^2 \\ &= \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h} S_{Y_h}^2 \end{aligned}$$

which is the standard formula for $\text{MSE}(\hat{T}_{st}, \text{StRS})$.

4.2 Sufficient conditions for superior strategies

In this section we compare pairs of calibrated strategies and give theorems that provide sufficient conditions for superiority of one strategy over the other. Firstly note that we can write the following as quadratic forms:

$$(N-1)S_Y^2 = \mathbf{y}^t\mathbf{R}\mathbf{y} \tag{4.12}$$

where $\mathbf{R} = \mathbf{I}_N - \mathbf{J}_N/N$ and

$$\sum_{h=1}^{H}(N_h-1)S_{Y_h}^2 = \mathbf{y}^t\mathbf{P}\mathbf{y} \tag{4.13}$$

where $\mathbf{P}$ is a block diagonal matrix of H blocks with the h^{th} block equal to the $N_h \times N_h$ matrix

$$\mathbf{I}_{N_h} - \mathbf{J}_{N_h}/N_h$$

for $h = 1, \ldots, H$. Also note that $\mathbf{R}$ is orthogonal only to constant vectors and $\mathbf{P}$ is orthogonal only to vectors that are constant within strata.

The following theorem gives some sufficient conditions for one quadratic form to be smaller than another and a special case of this result is given in a corollary that follows.

Theorem 10 *Let $\mathbf{M}_1$ and $\mathbf{M}_2$ be two real symmetric matrices and let $\gamma_1(\kappa)$ and $\gamma_2(\kappa)$ be the respective smallest and largest eigenvalue of the matrix*

$$\mathbf{Q}_1 = \mathbf{M}_2 - \mathbf{M}_1 + \kappa\mathbf{R}$$

for some $\kappa \in \mathcal{R}$ where the matrix $\mathbf{R}$ is given as in (4.12). Then we have the following:

a) *if $\mathbf{Q}_1$ is of full rank and*

$$\kappa(N-1)S_V^2 \leq \gamma_1(\kappa)[(N-1)S_V^2 + N\bar{V}^2]$$

for some $\mathbf{v} \in \mathcal{R}^N$ then $\mathbf{v}^t\mathbf{M}_1\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_2\mathbf{v}$ and

b) *if*

$$\kappa(N-1)S_V^2 \geq \gamma_2(\kappa)[(N-1)S_V^2 + N\bar{V}^2]$$

for some $\mathbf{v} \in \mathcal{R}^N$ *then* $\mathbf{v}^t\mathbf{M}_2\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v}$.

Proof For part (a), under the assumption that $\mathbf{Q}_1$ is of full rank with $\gamma_1(\kappa)$ being its smallest eigenvalue it follows from (4.4) that

$$\gamma_1(\kappa)[(N-1)S_V^2 + N\bar{V}^2] \leq \mathbf{v}^t\mathbf{Q}_1\mathbf{v} = \mathbf{v}^t[\mathbf{M}_2 - \mathbf{M}_1 + \kappa\mathbf{R}]\mathbf{v}.$$

Then

$$\kappa(N-1)S_V^2 = \kappa\mathbf{v}^t\mathbf{R}\mathbf{v} \leq \gamma_1(\kappa)[(N-1)S_V^2 + N\bar{V}^2]$$

will imply

$$\kappa\mathbf{v}^t\mathbf{R}\mathbf{v} \leq \mathbf{v}^t[\mathbf{M}_2 - \mathbf{M}_1 + \kappa\mathbf{R}]\mathbf{v}$$

and hence $\mathbf{v}^t\mathbf{M}_1\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_2\mathbf{v}$.

Similarly for part (b), since $\gamma_2(\kappa)$ is the largest eigenvalue of $\mathbf{Q}_1$ it follows from (4.4) that

$$\gamma_2(\kappa)[(N-1)S_V^2 + N\bar{V}^2] \geq \mathbf{v}^t\mathbf{Q}_1\mathbf{v} = \mathbf{v}^t[\mathbf{M}_2 - \mathbf{M}_1 + \kappa\mathbf{R}]\mathbf{v}.$$

Hence

$$\kappa(N-1)S_V^2 = \kappa\mathbf{v}^t\mathbf{R}\mathbf{v} \geq \mathbf{v}^t[\mathbf{M}_2 - \mathbf{M}_1 + \kappa\mathbf{R}]\mathbf{v}$$

which implies $\mathbf{v}^t\mathbf{M}_2\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v}$ and that completes the proof. $\square$

Note that the condition of part (a) in theorem 10 will always hold whenever $\kappa < \gamma_1(\kappa)$, and if $\kappa > \gamma_1(\kappa) > 0$ then for $\bar{V} \neq 0$ we have

$$\mathrm{cv}(\mathbf{v}) = \frac{S_V}{\bar{V}} \leq \sqrt{\frac{\gamma_1(\kappa)N}{(\kappa - \gamma_1(\kappa))(N-1)}} \quad \Leftrightarrow \quad \kappa(N-1)S_V^2 \leq \gamma_1(\kappa)[(N-1)S_V^2 + N\bar{V}^2].$$

Similarly from part (b) of theorem 10, if $\kappa \geq \gamma_2(\kappa) > 0$ then

$$\text{cv}(\mathbf{v}) \geq \sqrt{\frac{\gamma_2(\kappa)N}{(\kappa - \gamma_2(\kappa))(N-1)}} \quad \Leftrightarrow \quad \kappa(N-1)S_V^2 \geq \gamma_2(\kappa)[(N-1)S_V^2 + N\bar{V}^2].$$

From this we have the following result that can be applied to strategies that are not calibrated for the same vectors.

Corollary 2 *Let* $\mathbf{v} \in \mathcal{R}^N$ *such that* $\bar{V} \neq 0$ *and let* $\mathbf{M}_1$ *and* $\mathbf{M}_2$ *be two real symmetric matrices.*

a) *Suppose* $\alpha > \gamma_1(\alpha) > 0$ *for* $\alpha \in \mathcal{R}$ *where* $\gamma_1(\alpha)$ *is the smallest eigenvalue of the matrix*

$$\mathbf{Q}_1 = \mathbf{M}_2 - \mathbf{M}_1 + \alpha\mathbf{R}.$$

Provided $\mathbf{Q}_1$ *is of full rank we have*

$$\text{cv}(\mathbf{v}) = \frac{S_V}{\bar{V}} \leq \sqrt{\frac{\gamma_1(\alpha)N}{(\alpha - \gamma_1(\alpha))(N-1)}} \quad \Rightarrow \quad \mathbf{v}^t\mathbf{M}_1\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_2\mathbf{v}.$$

b) *Suppose* $\beta > \gamma_2(\beta) > 0$ *for some* $\beta \in \mathcal{R}$ *where* $\gamma_2(\beta)$ *is the largest eigenvalue of the matrix*

$$\mathbf{Q}_1 = \mathbf{M}_2 - \mathbf{M}_1 + \beta\mathbf{R}.$$

Then we have

$$\text{cv}(\mathbf{v}) = \frac{S_V}{\bar{V}} \geq \sqrt{\frac{\gamma_2(\beta)N}{(\beta - \gamma_2(\beta))(N-1)}} \quad \Rightarrow \quad \mathbf{v}^t\mathbf{M}_2\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v}.$$

It is easy to see that corollary 2 may be applied to any pair of strategies where one is calibrated only for constant vectors and the other is not calibrated for constant vectors. For example if $\mathbf{M}_1$ was the mean square error matrix for a strategy that is calibrated for constant vectors and $\mathbf{M}_2$ was the mean square error matrix for a strategy that was not calibrated for constant vectors, then choosing α and β to be equal to the maximum eigenvalue of $\mathbf{M}_1$ would imply that $\mathbf{Q}_1$ is positive definite and of full rank. Provided $\alpha > \gamma_1(\alpha)$ and $\beta > \gamma_2(\beta)$ we can use corollary 2 to compare the strategies.

It is reasonable to believe that when $\mathrm{cv}(\mathbf{y})$ is small, a strategy that is calibrated for constant vectors should be more efficient than one that isn't calibrated for constant vectors. Part (a) of corollary 2 indicates how small $\mathrm{cv}(\mathbf{y})$ would have to be in order for a strategy that is calibrated for constant vectors to be better than one that isn't. However if $\mathrm{cv}(\mathbf{y})$ is large it is not necessarily true that a strategy that is calibrated for constant vectors will always have a larger mean square error than one that isn't calibrated for constant vectors. So part (b) of corollary 2 might not always apply for these types of strategies as we will see in an example that follows.

Note that there can be a wide range of values for α and β that satisfy the conditions of corollary 2. The appropriate choice for $\alpha = \alpha_0$ would be one that maximizes the ratio

$$r_1(\alpha) = \frac{\gamma_1(\alpha)}{\alpha - \gamma_1(\alpha)}$$

as then the sufficient condition of corollary 2 part (a) will cover the most vectors in $\mathcal{R}^N$. Similarly for part (b) of corollary 2, the appropriate value for $\beta = \beta_0$ would be the one that minimizes the ratio

$$r_2(\beta) = \frac{\gamma_2(\beta)}{\beta - \gamma_2(\beta)}.$$

Unfortunately it isn't obvious how to calculate α_0 and β_0. Instead we will use an algorithm to approximate these values in our examples and give plots of α against $r_1(\alpha)$ and β against $r_2(\beta)$.

Example 19 *For a population size 20 and sample size 4, consider the expansion estimator under simple random sampling,* $stg(\hat{T}_0, SRS)$*, and the ratio estimator under* *simple random sampling,* $stg(\hat{T}_R, SRS)$*, with auxiliary vector from example 11,*

$$\mathbf{x}^t = (44, 44, 45, 43, 40, 52, 43, 47, 54, 43, 51, 50, 61, 47, 62, 34, 51, 48, 51, 57)$$

whose coefficient of variation is 0.1427. Since $stg(\hat{T}_0, SRS)$ *is calibrated for constant vectors and* $stg(\hat{T}_R, SRS)$ *is calibrated for* $\mathbf{x}$*, then* $\mathbf{M}(\hat{T}_0, SRS)$ *is orthogonal to constant vectors and* $\mathbf{M}(\hat{T}_R, SRS)$ *is orthogonal to* $\mathbf{x}$*. We applied corollary 2 part (a) to these strategies.*

Figure 4.1 shows a plot of the values of $r_1(\alpha)$ *against* α *and we approximated the values* $\alpha_0 = 8.511656$ *which gives* $\gamma_1(\alpha_0) = 0.04140370176$ *and* $r_1(\alpha_0) = 0.004888130$*. Hence by corollary 2 part (a) a sufficient condition for* $\mathrm{MSE}(\hat{T}_0, SRS)$ *to be less than* $\mathrm{MSE}(\hat{T}_R, SRS)$ *is if*

$$\mathrm{cv}(\mathbf{y}) \leq \sqrt{r_1(\alpha_0)\frac{N}{N-1}} = 0.07173145$$

provided $\bar{Y} \neq 0$*.*

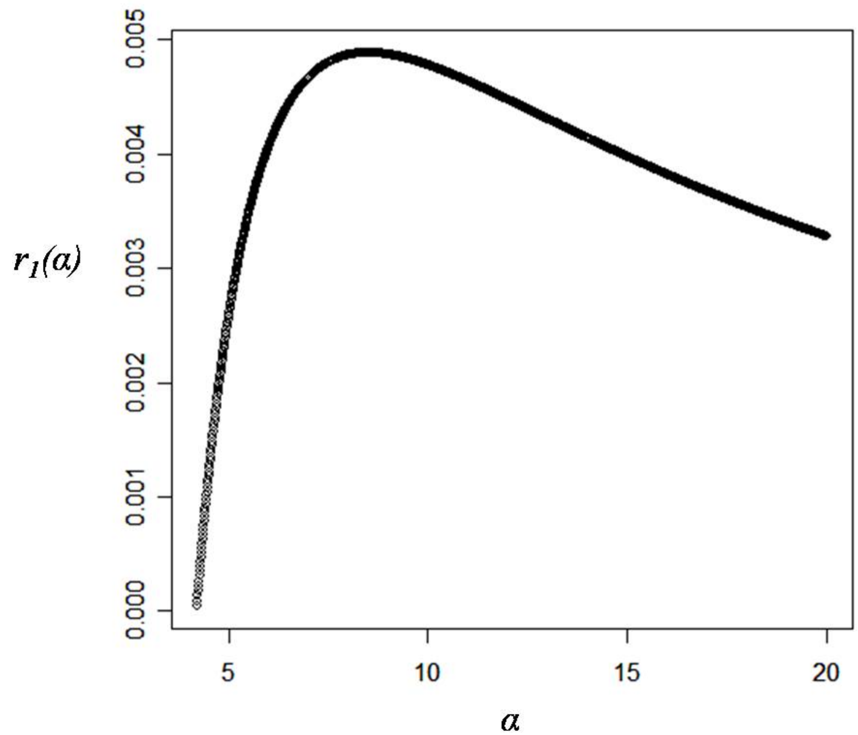

Figure 4.1: Plot of $r_1(\alpha)$ against α for example 19

For example 19 we could not apply corollary 2 part (b) to give a sufficient condition for $\text{MSE}(\hat{T}_R, \text{SRS}) \leq \text{MSE}(\hat{T}_0, \text{SRS})$ because there doesn't exist a β such that $\beta > \gamma_2(\beta) > 0$ for this example. We will see (in ch.5, p.183) an example where part (b) of corollary 2 can be applied to compare strategies.

Note that if we consider the nonsingular transformation $\mathbf{z} = \text{diag}(\mathbf{x})^{-1}\mathbf{y}$ then for any symmetric matrix $\mathbf{M}$ that is orthogonal to $\mathbf{x}$ we have

$$\mathbf{y}^t\mathbf{M}\mathbf{y} = \mathbf{z}^t[\text{diag}(\mathbf{x})\mathbf{M}\text{diag}(\mathbf{x})]\mathbf{z} = \mathbf{z}^t\mathbf{A}\mathbf{z}$$

where the matrix $\mathbf{A} = \text{diag}(\mathbf{x})\mathbf{M}\text{diag}(\mathbf{x})$ is orthogonal to constant vectors. Hence from example 19 we can give a sufficient condition for $stg(\hat{T}_R, \text{SRS})$ to be better than $stg(\hat{T}_0, \text{SRS})$, based on part (a) of corollary 2, by changing variables from $\mathbf{y}$ to $\mathbf{z}$ which we now illustrate in the next example.

Example 20 *Following example 19 but now applying the nonsingular transformation* $\mathbf{z} = \text{diag}(\mathbf{x})^{-1}\mathbf{y}$ *so we can write*

$$\text{MSE}(\hat{T}_0, SRS) = \mathbf{z}^t[\text{diag}(\mathbf{x})\mathbf{M}(\hat{T}_0, SRS)\text{diag}(\mathbf{x})]\mathbf{z} = \mathbf{z}^t\mathbf{A}(\hat{T}_0, SRS)\mathbf{z}$$

and

$$\text{MSE}(\hat{T}_R, SRS) = \mathbf{z}^t[\text{diag}(\mathbf{x})\mathbf{M}(\hat{T}_R, SRS)\text{diag}(\mathbf{x})]\mathbf{z} = \mathbf{z}^t\mathbf{A}(\hat{T}_R, SRS)\mathbf{z}$$

where

$$\mathbf{A}(\hat{T}_0, SRS) = \text{diag}(\mathbf{x})\mathbf{M}(\hat{T}_0, SRS)\text{diag}(\mathbf{x})$$

and

$$\mathbf{A}(\hat{T}_R, SRS) = \text{diag}(\mathbf{x})\mathbf{M}(\hat{T}_R, SRS)\text{diag}(\mathbf{x}).$$

Now $\mathbf{A}(\hat{T}_0, SRS)$ *is orthogonal to* $\mathbf{x}^{-1} = (x_1^{-1}, x_2^{-1}, \ldots, x_N^{-1})^t$ *and* $\mathbf{A}(\hat{T}_R, SRS)$ *is orthogonal to constant vectors so corollary 2 can be applied to give a sufficient condition for the superiority of* $stg(\hat{T}_R, SRS)$.

Figure 4.2 gives the plot of $r_1(\alpha)$ *against* α *and we approximated the values* $\alpha_0 = 20804.46$ *which gives* $\gamma_1(\alpha_0) = 93.94339104$ *and* $r_1(\alpha) = 0.004536023549$*. Hence, by corollary 2 part (a) a sufficient condition for* $\mathrm{MSE}(\hat{T}_R, SRS)$ *to be less than* $\mathrm{MSE}(\hat{T}_0, SRS)$ *is if*

$$\mathrm{cv}(\mathbf{z}) \leq 0.06909965$$

provided $\bar{Z} \neq 0$*.*

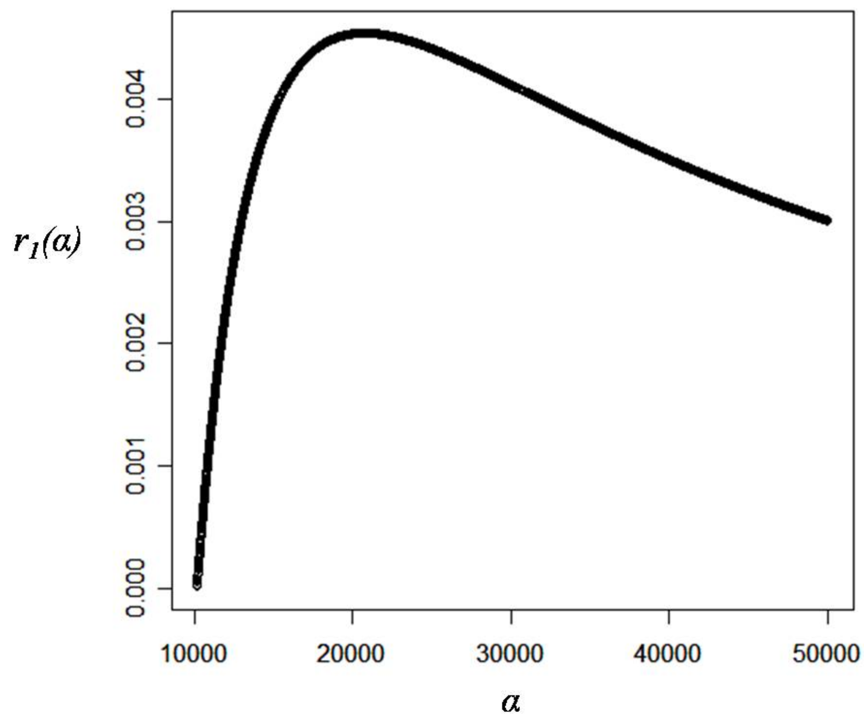

Figure 4.2: Plot of $r_1(\alpha)$ against α for example 20

Another result for comparing $stg(\hat{T}_0, \text{SRS})$ and $stg(\hat{T}_R, \text{SRS})$ was given by Cochran (1977, p.157), which is based on a large sample approximation formula for the variance of $stg(\hat{T}_R, \text{SRS})$ and says that if the correlation coefficient, $\rho(\mathbf{x}, \mathbf{y})$, between $\mathbf{x}$ and $\mathbf{y}$ is such that

$$\rho(\mathbf{x}, \mathbf{y}) > \frac{\text{cv}(\mathbf{x})}{2\text{cv}(\mathbf{y})}$$

then $stg(\hat{T}_R, \text{SRS})$ has a smaller mean square error than $stg(\hat{T}_0, \text{SRS})$. Deng & Chhikara (1990) used a different approximation formula for $\text{Var}(\hat{T}_R, \text{SRS})$ and gave a similar result for $stg(\hat{T}_R, \text{SRS})$ to be more efficient than $stg(\hat{T}_0, \text{SRS})$.

The advantages of using corollary 2, over Cochran's method, to compare strategies is that we only need to estimate one unknown parameter, $\text{cv}(\mathbf{z})$, which is equal to zero when $\mathbf{y} \propto \mathbf{x}$. It is an exact method since it is based on the exact mean square error of strategies valid for any sample size and it can be applied to any pair of strategies that are not calibrated for the same vector. However a disadvantage of corollary 2 is that calculating the mean square error matrices for large population and sample sizes can be a massive computational burden for complicated strategies.

Cochran's method is only valid for large sample size and we need to estimate two unknown parameters: $\rho(\mathbf{y}, \mathbf{x})$ and $\text{cv}(\mathbf{y})$. Note also that $\rho(\mathbf{x}, \mathbf{y}) = 1$ does not necessarily imply $\mathbf{y} \propto \mathbf{x}$ since the straight line relationship between $\mathbf{y}$ and $\mathbf{x}$ might not go through the origin. Cochran's method is only valid for comparing $stg(\hat{T}_R, \text{SRS})$ with $stg(\hat{T}_0, \text{SRS})$, however it is computationally simple.

We will give a result that compares a separate general linear strategy with another strategy that are both calibrated for some $\mathbf{x} \in \mathcal{R}^N$, but not equally calibrated. But first we consider the ANOVA decomposition of the total sum of squares of the y's, see Cochran (1977, p.100):

$$(N-1)S_Y^2 = \sum_{h=1}^{H}(N_h - 1)S_{Y_h}^2 + \sum_{h=1}^{H} N_h(\bar{Y}_h - \bar{Y})^2. \tag{4.14}$$

From the identity given in (4.14) we obtain the measure

$$R_Y = \frac{\sum_{h=1}^{H}(N_h - 1)S_{Y_h}^2}{(N-1)S_Y^2} \tag{4.15}$$

which could indicate a potentially 'good' stratification when its value is small. Note that R_Y is equal to $1 - R^2$ in the usual linear model notation. Holt & Smith (1979) also considered the measure (4.15) in their empirical study of the conditional mean square errors of $stg(\hat{T}_{st}, \mathrm{SRS})$ and $stg(\hat{T}_0, \mathrm{SRS})$, we discuss this more on p.201. Note that R_Y takes values between 0 and 1.

The following theorem can be used to give a sufficient condition, which is based on R_Y from (4.15), for the superiority of a separate general linear strategy.

Theorem 11 *Suppose* $\mathbf{M}_1$ *and* $\mathbf{M}_2$ *are real* $N \times N$ *symmetric matrices such that*

$$\mathbf{Q}_2 = \mathbf{M}_1 - \mathbf{M}_2 + \eta\mathbf{P}$$

is nonnegative definite, where the matrix $\mathbf{P}$ *is given by (4.13), and let* $\gamma_3(\eta)$ *denote the smallest nonzero eigenvalue of* $\mathbf{Q}_2$ *for some real number* $\eta > 0$. *Provided* $\mathbf{Q}_2$ *is orthogonal only to constant vectors we have*

$$R_V = \frac{\sum_{h=1}^{H}(N_h - 1)S_{V_h}^2}{(N-1)S_V^2} \leq \frac{\gamma_3(\eta)}{\eta} \quad \Rightarrow \quad \mathbf{v}^t\mathbf{M}_2\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v}.$$

Proof Observe that for $\eta > 0$ and $\gamma_3(\eta) > 0$ we have

$$\frac{\sum_{h=1}^{H}(N_h - 1)S_{V_h}^2}{(N-1)S_V^2} \leq \frac{\gamma_3(\eta)}{\eta} \quad \Leftrightarrow \quad \eta\sum_{h=1}^{H}(N_h - 1)S_{V_h}^2 \leq \gamma_3(\eta)(N-1)S_V^2.$$

Under the assumption that $\mathbf{Q}_2$ is only orthogonal to constant vectors with $\gamma_3(\eta)$ being its smallest nonzero eigenvalue it follows from (4.7) that

$$\gamma_3(\eta)(N-1)S_V^2 \leq \mathbf{v}^t\mathbf{Q}_2\mathbf{v} = \mathbf{v}^t[\mathbf{M}_1 - \mathbf{M}_2 + \eta\mathbf{P}]\mathbf{v}.$$

Hence

$$\eta \sum_{h=1}^{H}(N_h - 1)S_{V_h}^2 = \eta \mathbf{v}^t\mathbf{P}\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v} - \mathbf{v}^t\mathbf{M}_2\mathbf{v} + \eta \mathbf{v}^t\mathbf{P}\mathbf{v}$$

which implies

$$\mathbf{v}^t\mathbf{M}_2\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v}$$

and that completes the proof. □

Note that if the intersection of the nullspaces of the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ from theorem 11 is equal to constant vectors, i.e. if $\mathcal{N}(\mathbf{M}_1) \cap \mathcal{N}(\mathbf{M}_2) = \langle \mathbf{1}_N \rangle$, then the matrix $\mathbf{Q}_2$ will also be orthogonal to constant vectors. Furthermore if $\mathcal{N}(\mathbf{M}_2) = \langle \mathbf{u}_1, \ldots, \mathbf{u}_H \rangle$, where $\mathbf{u}_h$ for $h = 1, \ldots, H$ is defined as in theorem 9, and if η is greater than the maximum eigenvalue of $\mathbf{M}_2$ then the conditions imposed on $\mathbf{Q}_2$ in theorem 11 will be satisfied. To see this observe that as $\mathbf{M}_2$ is orthogonal only to vectors that are constant within strata, it follows from lemma 14 that

$$\mathbf{y}^t\mathbf{M}_2\mathbf{y} \leq \lambda_{max}\mathbf{y}^t\mathbf{P}\mathbf{y} = \lambda_{max}\sum_{h=1}^{H}(N_h - 1)S_{Y_h}^2$$

where λ_{max} is the maximum eigenvalue of $\mathbf{M}_2$. Choosing η to be greater than λ_{max} gives us $\mathbf{y}^t[-\mathbf{M}_2 + \eta\mathbf{P}]\mathbf{y} \geq 0$ for all $\mathbf{y} \in \mathcal{R}^N$, and so it follows that the value of $\mathbf{y}^t\mathbf{Q}_2\mathbf{y}$ will always be nonnegative which implies $\mathbf{Q}_2$ will be nonnegative definite. Since the intersection of the nullspaces of $\mathbf{M}_1$ and $\mathbf{M}_2$ is equal to constant vectors, the quadratic form

$$\mathbf{y}^t\mathbf{Q}_2\mathbf{y} = \mathbf{y}^t\mathbf{M}_1\mathbf{y} - \mathbf{y}^t\mathbf{M}_2\mathbf{y} + \eta\mathbf{y}^t\mathbf{P}\mathbf{y}$$

can only be equal to zero when $\mathbf{y}$ is a constant vector which implies that $\mathbf{Q}_2$ is only orthogonal to constant vectors.

Then it is clear that we can apply theorem 11 to give a sufficient condition for a separate general linear strategy that is calibrated only for those vectors that are constant within strata with strategies that are calibrated for constant vectors. We will only consider theorem 11 to compare these types of strategies.

Note that, as with the previous corollary, there can be a wide range of values for η that satisfy the condition of theorem 11. Furthermore it is not necessary, but sufficient, for η to be greater than the maximum eigenvalue of $\mathbf{M}_2$ in order to satisfy the conditions imposed on $\mathbf{Q}_2$, where $\mathbf{M}_1$ and $\mathbf{M}_2$ are described on the previous page. The appropriate choice for $\eta = \eta_0$ should be such that the ratio

$$r_3(\eta) = \frac{\gamma_3(\eta)}{\eta}$$

is at its maximum so that the sufficient condition of theorem 11 holds for the maximum possible number of vectors in $\mathcal{R}^N$. We will denote $\gamma_3(\eta_0)$ by γ_0.

Example 21 *For a population size* $\mathbf{N} = (8, 12)$ *sample size 4, consider the stratified estimator* $\hat{T}_{st} = \sum_{h=1}^{H} I_h N_h \bar{y}_h$ *under a stratified random sampling design on* $\mathbf{n} = (2, 2)$ *and the regression estimator* $\hat{T}_{REG}$ *under the equal probability design given by* p_b *in example 10 on p.101. Let the auxiliary matrix* $\mathbf{X}$ *be the same as in example 10. Since* $stg(\hat{T}_{st}, StRS)$ *is calibrated only for vectors that are constant within strata and* $stg(\hat{T}_{REG}, p_b)$ *is calibrated for constant vectors we can apply theorem 11 to give a sufficient condition for* $stg(\hat{T}_{st}, StRS)$ *to be better than* $stg(\hat{T}_{REG}, p_b)$*.*

Figure 4.3 gives a plot of $r_3(\eta)$ *against* η *and we approximated the values* $\eta_0 = 11.888966$ *and* $\gamma_0 = 3.936793$ *which gives* $r_3(\alpha_0) = 0.3311302$*. Then by theorem 11 a sufficient condition for* $\mathrm{MSE}(\hat{T}_{st}, StRS)$ *to be less than* $\mathrm{MSE}(\hat{T}_{REG}, p_b)$ *is if*

$$R_Y \leq \frac{\gamma_0}{\eta_0} = 0.3311302$$

provided $S_Y \neq 0$*.*

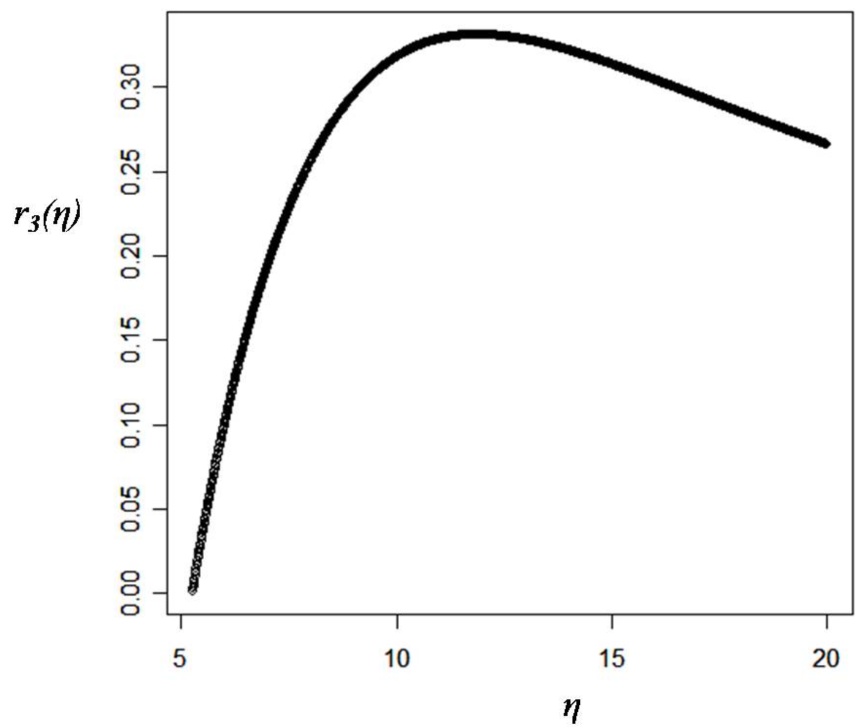

Figure 4.3: Plot of $r_3(\eta)$ against η for example 21

A special case of theorem 11 is when comparing the expansion estimator under simple random sampling with the stratified estimator under stratified random sampling. In this case the value of η_0 that maximizes $r_3(\eta_0)$ can be easily calculated and is given in theorem 12. But first we give a lemma that will be useful in the proof of theorem 12.

Lemma 16 *Let μ_{max} be the maximum eigenvalue of $\mathbf{M}(\hat{T}_{st}, StRS)$, with $n_h > 0$ for all $h = 1, \ldots, H$, and let λ be the nonzero eigenvalue of $\mathbf{M}(\hat{T}_0, SRS)$. Then $\lambda < \mu_{max}$ is always true.*

Proof By (4.2) the nonzero eigenvalue of $\mathbf{M}(\hat{T}_0, \text{SRS})$ is equal to

$$\lambda = \frac{N(N-n)}{n(N-1)}$$

and by (4.11) the eigenvalues, μ_h for $h = 1, \ldots, H$, of $\mathbf{M}(\hat{T}_{st}, \text{StRS})$ are equal to

$$\mu_h = \frac{N_h(N_h - n_h)}{n_h(N_h - 1)}$$

with $\mu_{max} = \max_{h=1,\ldots,H}\{\mu_h\}$.

Observe that

$$\begin{aligned} \lambda < \mu_h \quad &\Leftrightarrow \quad \frac{N(N-n)}{n(N-1)} < \frac{N_h(N_h-n_h)}{n_h(N_h-1)} \\ &\Leftrightarrow \quad \frac{f_h(1-f)}{f(1-f_h)} < \frac{N_h(N-1)}{N(N_h-1)} \end{aligned} \tag{4.16}$$

where $f = n/N$ is the sampling fraction for the population and $f_h = n_h/N_h$ is the sampling fraction for stratum h for some $h = 1, \ldots, H$. Note that the expression on the right hand side of (4.16) is greater than one since $N/(N-1)$ is a decreasing function in N.

Now there are three cases of the sampling fraction for stratum h that we need to consider. The first case is when $f_h = f$. Then the left hand side of (4.16) will equal to one and so it follows from (4.16) that $\lambda < \mu_h$ which implies $\lambda < \mu_{max}$.

The second case is when $f_h < f$. Since the expression on the left hand side of (4.16) is an increasing function in f_h it follows that

$$\frac{f_h(1-f)}{f(1-f_h)} < \frac{f(1-f)}{f(1-f)} = 1$$

and so the inequality (4.16) will hold implying $\lambda < \mu_{max}$.

For the third case where $f_h > f$ notice that this implies that there must be at least one stratum $g \neq h$ such that $f_g < f$. To see this suppose $f_h > f$ were true for all $h = 1, \ldots, H$. Then this will imply that $Nn_h > N_h n$ for all $h = 1, \ldots, H$ which is impossible because we must have

$$\sum_{h=1}^{H} Nn_h = \sum_{h=1}^{H} N_h n = Nn.$$

Hence for any $f_h > f$ there must be at least one f_g ($g \neq h$) such that $f_g < f$ and so if follows from (4.16) that $\lambda < \mu_g \leq \mu_{max}$ and that completes the proof. □

Theorem 12 *The values of η_0 and γ_0 for theorem 11 that maximizes $r_3(\eta) = \gamma_3(\eta)/\eta$ so that the sufficient condition for stg($\hat{T}_{st}$, StRS) to be better than stg($\hat{T}_0$, SRS) holds for the maximum possible number of vectors in $\mathcal{R}^N$ are given by*

$$\eta = \eta_0 = \max_{h=1,\ldots,H} \left\{ \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \right\} \quad \textit{and} \quad \gamma_3(\eta_0) = \gamma_0 = \frac{N(N-n)}{n(N-1)}.$$

Proof Observe that the mean square error matrix of $stg(\hat{T}_{st}, \text{StRS})$, when $n_h > 0$ for all $h = 1, \ldots, H$, is equal to

$$\mathbf{M}(\hat{T}_{st}, \text{StRS}) = \sum_{h=1}^{H} \mu_h \mathbf{R}_h$$

where $\mathbf{R}_h = \text{diag}(\mathbf{u}_h) - \mathbf{u}_h \mathbf{u}_h^t / N_h$ with $\mathbf{u}_h$ defined as in theorem 9 and from (4.11)

$$\mu_h = \frac{N_h(N_h - n_h)}{n_h(N_h - 1)}$$

for $h = 1, \ldots, H$. The mean square error matrix of $stg(\hat{T}_0, \text{SRS})$ is equal to

$$\mathbf{M}(\hat{T}_0, \text{SRS}) = \lambda \mathbf{R}$$

where, from (4.2),

$$\lambda = \frac{N(N-n)}{n(N-1)}.$$

Then from theorem 11 we have

$$\begin{aligned}
\mathbf{Q}_2 &= \lambda \mathbf{R} - \sum_{h=1}^{H} \mu_h \mathbf{R}_h + \eta \mathbf{P} \\
&= \lambda \mathbf{R} - \sum_{h=1}^{H} (\mu_h - \eta) \mathbf{R}_h
\end{aligned}$$

and since

$$\mathbf{R} = \mathbf{I}_N - \mathbf{J}_N/N = \sum_{h=1}^{H} [\text{diag}(\mathbf{u}_h) - \mathbf{u}_h \mathbf{u}_h^t / N_h] + \left[\sum_{h=1}^{H} \mathbf{u}_h \mathbf{u}_h^t / N_h - \mathbf{J}_N/N \right]$$

we have

$$\mathbf{Q}_2 = \sum_{h=1}^{H} (\lambda - \mu_h + \eta) [\text{diag}(\mathbf{u}_h) - \mathbf{u}_h \mathbf{u}_h^t / N_h] + \lambda \left[\sum_{h=1}^{H} \mathbf{u}_h \mathbf{u}_h^t / N_h - \mathbf{J}_N/N \right]$$

which implies

$$\mathbf{v}^t \mathbf{Q}_2 \mathbf{v} = \sum_{h=1}^{H} (\lambda - \mu_h + \eta)(N_h - 1) S_{V_h}^2 + \lambda \sum_{h=1}^{H} N_h (\bar{V}_h - \bar{V})^2.$$

The matrix $\mathbf{Q}_2$ is orthogonal to constant vectors only and is of the same form as (4.10) with

$$\mathbf{Q}_{2_{ij}} = \begin{cases} \lambda(1 - \frac{1}{N}) - (\mu_h - \eta)(1 - \frac{1}{N_h}) & \text{for} \quad i = j \in \mathcal{U}_h \\ -\frac{\lambda}{N} + \frac{(\mu_h - \eta)}{N_h} & \text{for} \quad i \neq j \in \mathcal{U}_h \\ -\frac{\lambda}{N} & \text{for} \quad i \in \mathcal{U}_h \text{ for } j \in \mathcal{U}_g \end{cases}$$

for $h, g = 1, \ldots, H$ $(h \neq g)$. Therefore its eigenvalues are equal to $\lambda - \mu_h + \eta$ of multiplicity $N_h - 1$ for $h = 1, \ldots, H$ and since the matrix $\sum_{h=1}^{H} \mathbf{u}_h \mathbf{u}_h^t / N_h - \mathbf{J}_N/N$, which is idempotent and of rank $H-1$, is orthogonal to $\sum_{h=1}^{H}[\text{diag}(\mathbf{u}_h) - \mathbf{u}_h\mathbf{u}_h^t/N_h]$ it follows that λ is an eigenvalue of $\mathbf{Q}_2$ of multiplicity $H-1$. Then the minimum nonzero eigenvalue of $\mathbf{Q}_2$ must be

$$\gamma_3(\eta) = \min\{\lambda - \mu_{max} + \eta, \lambda\}$$

where $\mu_{max} = \max_{h=1,\ldots,H}\{\mu_h\}$.

If we choose η such that $\eta \geq \mu_{max}$ then $\gamma_3(\eta) = \lambda$ and the ratio

$$r_3(\eta) = \frac{\gamma_3(\eta)}{\eta} = \frac{\lambda}{\eta}$$

is at its maximum when $\eta = \mu_{max}$. If we choose η such that $\eta \leq \mu_{max}$ then $\gamma_3(\eta) = \lambda - \mu_{max} + \eta$ and since, by lemma 16, μ_{max} is always greater than λ the ratio

$$r_3(\eta) = \frac{\gamma_3(\eta)}{\eta} = \frac{\lambda - \mu_{max} + \eta}{\eta} = \frac{\lambda - \mu_{max}}{\eta} + 1$$

is at its maximum when $\eta = \mu_{max}$ in which case $\gamma_3(\eta) = \lambda$.

Hence by choosing $\eta = \eta_0 = \mu_{max}$ and using theorem 11, a sufficient condition for $\text{MSE}(\hat{T}_{st}, \text{StRS}) \leq \text{MSE}(\hat{T}_0, \text{SRS})$ for the maximum possible number of vectors in $\mathcal{R}^N$ is if $R_Y \leq r_3(\eta_0) = \lambda/\mu_{max}$. □

Cochran (1977, p.99-101) also compared $stg(\hat{T}_{st}, \text{StRS})$ with $stg(\hat{T}_0, \text{SRS})$ and gave a large sample approximation result that says that the variance of the stratified estimator under proportional allocation where $n_h = N_h n/N$ for all $h = 1, \ldots, H$ is always less than that of the expansion estimator under simple random sampling.

The following theorem gives another sufficient condition for the superiority of a separate general linear strategy.

Theorem 13 *Let $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{Q}_2$ be as in theorem 11. Then for some $\mathbf{v} \in \mathcal{R}^N$ if*

$$\text{cv}(\mathbf{v}_h) = \frac{S_{V_h}}{\bar{V}} \leq \sqrt{\frac{(\gamma_0 - \tau)N_h}{(\eta_0 - \gamma_0 + \tau)(N_h - 1)}} \tag{4.17}$$

for all $h = 1, \ldots, H$ *and* $\mathrm{cv}(\mathbf{v})$ *is such that*

$$\mathrm{cv}(\mathbf{v}) \geq \sqrt{\frac{(\gamma_0 - \tau)N}{\tau(N-1)}}$$

then

$$\mathbf{v}^t\mathbf{M}_2\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v}$$

where τ *is any value between* $(0, \gamma_0)$*.*

Proof Observe that for all $h = 1, \ldots, H$ we have

$$\mathrm{cv}(\mathbf{v}_h)^2 \leq \frac{(\gamma_0 - \tau)N_h}{(\eta_0 - \gamma_0 + \tau)(N_h - 1)} \quad \Leftrightarrow \quad (\eta_0 - \gamma_0 + \tau)(N_h - 1)S_{V_h}^2 \leq (\gamma_0 - \tau)N_h\bar{V}_h^2$$

which implies

$$(\eta_0 - \gamma_0 + \tau)\sum_{h=1}^{H}(N_h - 1)S_{V_h}^2 \leq (\gamma_0 - \tau)\sum_{h=1}^{H}N_h\bar{V}_h^2, \tag{4.18}$$

and that

$$\mathrm{cv}(\mathbf{v})^2 \geq \frac{(\gamma_0 - \tau)N}{\tau(N-1)} \quad \Leftrightarrow \quad \tau(N-1)S_V^2 \geq (\gamma_0 - \tau)N\bar{V}^2. \tag{4.19}$$

Combining (4.18) and (4.19) gives

$$\eta_0\sum_{h=1}^{H}(N_h - 1)S_{V_h}^2 + (\gamma_0 - \tau)N\bar{V}^2 \leq (\gamma_0 - \tau)[(N-1)S_V^2 + N\bar{V}^2] + \tau(N-1)S_V^2$$

and hence

$$\eta_0\sum_{h=1}^{H}(N_h - 1)S_{V_h}^2 \leq \gamma_0(N-1)S_V^2 \quad \Rightarrow \quad \mathbf{v}^t\mathbf{M}_2\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v}$$

and that completes the proof. □

Suppose that $\mathrm{cv}(\mathbf{v}) \geq d$ where d is known. Then from theorem 13, if we let

$$d = \sqrt{\frac{(\gamma_0 - \tau)N}{\tau(N-1)}} \quad \Leftrightarrow \quad \tau_0 = \tau = \frac{\gamma_0 N}{d^2(N-1) + N}$$

and substitute τ_0 into (4.17) to get

$$\sqrt{\frac{(\gamma_0-\tau_0)N_h}{(\eta_0-\gamma_0+\tau_0)(N_h-1)}} = r\sqrt{\frac{N_h}{N_h-1}},$$

where

$$r = \sqrt{\frac{(\gamma_0-\tau_0)}{(\eta_0-\gamma_0+\tau_0)}},$$

we get the corresponding upper bounds on $\text{cv}(\mathbf{v}_h)$ for each $h = 1, \ldots, H$ so that $\mathbf{v}^t\mathbf{M}_2\mathbf{v} \leq \mathbf{v}^t\mathbf{M}_1\mathbf{v}$.

Example 22 *For a population of size 284 and sample size 43, consider the expansion estimator under simple random sampling, stg($\hat{T}_0, SRS$), and the stratified estimator under stratified random sampling, stg($\hat{T}_{st}, StRS$), with strata sizes equal to* $\mathbf{N} = (25, 48, 32, 38, 56, 41, 15, 29)$ *and sample strata sizes* $\mathbf{n} = (4, 4, 5, 5, 8, 9, 2, 6)$. *From (4.2) the nonzero eigenvalue of* $\mathbf{M}(\hat{T}_0, SRS)$ *is equal to*

$$\lambda = \frac{N(N-n)}{n(N-1)} = \frac{284(284-43)}{43(284-1)} = 5.624455584$$

and from (4.11) the nonzero eigenvalues of $\mathbf{M}(\hat{T}_{st}, StRS)$ *are given by*

$$\mu_h = \frac{N_h(N_h-n_h)}{n_h(N_h-1)}$$

for $h = 1, \ldots, H$. The maximum eigenvalue, μ_{max}, in this case is for $h = 2$ and is equal to

$$\mu_{max} = \frac{48(48-4)}{4(48-1)} = 11.23404.$$

Hence it follows from theorem 12 that a sufficient condition for $\text{MSE}(\hat{T}_{st}, StRS)$ *to be less than* $\text{MSE}(\hat{T}_0, SRS)$ *is if*

$$R_Y \leq \frac{\lambda}{\mu_{max}} = 0.5006618$$

and by theorem 13 if

$$\text{cv}(\mathbf{y}) \geq d \quad \textit{and} \quad \text{cv}(\mathbf{y}_h) \leq r\sqrt{\frac{N_h}{N_h-1}}$$

for all $h = 1, \ldots, H$ then $\text{MSE}(\hat{T}_{st}, StRS)$ *will be less than* $\text{MSE}(\hat{T}_0, SRS)$ *where the values of d and r are given in table 4.1.*

Table 4.1: Values of d and r for example 22

d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
r	0.005	0.020	0.043	0.074	0.111	0.152	0.197	0.242	0.288	0.333

Example 23 *Consider the Horvitz-Thompson estimator under Tillé's procedure for unequal probability sampling with* $\pi_i \propto x_i$ *over all* $i \in \mathcal{U}$ *where the size vector* $\mathbf{x}$ *is the MU284-S82 data from Särndal et al. (1992, p.652), and the Horvitz-Thompson estimator under a stratified Tillé scheme on* $\mathbf{x}$ *where the population is partitioned by geographical region. The strata sizes are* $\mathbf{N} = (25, 48, 32, 38, 56, 41, 15, 29)$ *with sample size allocation* $\mathbf{n} = (4, 4, 5, 5, 8, 9, 2, 6)$. *After applying a nonsingular transformation* $\mathbf{z} = \mathrm{diag}(\mathbf{x})^{-1}\mathbf{y}$ *on the mean square error matrices of these strategies as we did in example 20, we approximated the values* $\eta_0 = 34280.38$ *and* $\gamma_0 = 10794.93$. *Then by theorem 11 a sufficient condition for* $\mathrm{MSE}(\hat{T}_{HT}, \mathit{St.Tillé})$ *to be less than* $\mathrm{MSE}(\hat{T}_{HT}, \mathit{Tillé})$ *is if*

$$R_Z = \frac{\sum_{h=1}^{H}(N_h - 1)S^2_{Z_h}}{(N-1)S^2_Z} \leq \frac{\gamma_0}{\eta_0} = 0.3149012$$

and by theorem 13 if

$$\mathrm{cv}(\mathbf{z}) \geq d \quad \textit{and} \quad \mathrm{cv}(\mathbf{z}_h) \leq r\sqrt{\frac{N_h}{N_h - 1}}$$

for all $h = 1, \ldots, H$, *then* $\mathrm{MSE}(\hat{T}_{HT}, \mathit{St.Tillé})$ *will be less than* $\mathrm{MSE}(\hat{T}_{HT}, \mathit{Tillé})$ *where the values of* d *and* r *are given in table 4.2.*

Table 4.2: Values of d and r for example 23

d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
r	0.003	0.012	0.027	0.045	0.067	0.091	0.115	0.140	0.164	0.186

Example 24 *Consider the strategies stg*$(\hat{T}_R, SRS)$ *and stg*$(\hat{T}_{Rs}, StRS)$ *using the same auxiliary vector from example 19. Let the population be partitioned into two groups such that* $\mathcal{U}_1 = \{1, 2, \ldots, 10\}$ *and let* $\mathcal{U}_2 = \{11, 12, \ldots, 20\}$ *and let the sample size allocation be* $\mathbf{n} = (2, 2)$. *Since stg*$(\hat{T}_{Rs}, StRS)$ *is calibrated for* $\mathbf{x}_h$, $h = 1,\ 2,$ *and stg*$(\hat{T}_R, SRS)$ *is calibrated for* $\mathbf{x}$ *we can apply theorem 11, after changing variables from* $\mathbf{y}$ *to* $\mathbf{z} = \mathrm{diag}(\mathbf{x})^{-1}\mathbf{y}$, *to give a sufficient condition for the superiority of stg*$(\hat{T}_{Rs}, StRS)$. *We approximated the values* $\eta_0 = 7551.056$ *and* $\gamma_0 = 9976.285$ *so a sufficient condition for* $\mathrm{MSE}(\hat{T}_{Rs}, StRS)$ *to be less than* $\mathrm{MSE}(\hat{T}_R, SRS)$ *is if*

$$R_Z = \frac{\sum_{h=1}^{H} (N_h - 1) S^2_{Z_h}}{(N-1) S^2_Z} \le 0.7569007$$

and by theorem 13 if

$$\mathrm{cv}(\mathbf{z}) \ge d \quad \textit{and} \quad \mathrm{cv}(\mathbf{z}_h) \le r\sqrt{\frac{N_h}{N_h - 1}}$$

for all $h = 1,\ 2,$ *then* $\mathrm{MSE}(\hat{T}_{Rs}, StRS)$ *will be less than* $\mathrm{MSE}(\hat{T}_R, SRS)$ *where the values of* d *and* r *are given in table 4.3.*

Table 4.3: Values of d and r for example 24

d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
r	0.007	0.028	0.063	0.111	0.170	0.239	0.317	0.401	0.491	0.584

Note that if the sufficient conditions of theorem 13 are true than so is the sufficient condition of theorem 11 since

$$\mathrm{cv}(\mathbf{v}) \geq d \quad \text{and} \quad \mathrm{cv}(\mathbf{v}_h) \leq r\sqrt{\frac{N_h}{N_h - 1}}$$

for all $h = 1, \ldots, H$ implies that

$$R_V \leq \frac{\gamma_0}{\eta_0}.$$

However the opposite is not true, so we consider theorem 11 to be a more general condition than that of theorem 13 since it covers more vectors in $\mathcal{R}^N$. But it may be within the statisticians knowledge of the population, possibly due to previous studies, to be able to give a reasonable value of d such that $\mathrm{cv}(\mathbf{y}) > d$ and hence theorem 13 may be more applicable (whereas it may be difficult to say how large R_Y could be for the survey vector).

The combined ratio estimator $\hat{T}_{Rc}$ is of the form

$$\hat{T}_{Rc} = N\bar{X}\frac{\sum_{h=1}^{H} I_h N_h \bar{y}_h}{\sum_{h=1}^{H} I_h N_h \bar{x}_h}$$

and is clearly calibrated for $\mathbf{x}$ with respect to a support of any design. In the next example we compare this estimator under stratified random sampling with the separate ratio estimator under stratified random sampling.

Example 25 *For a population of size 20 and sample of size 4 we compute the mean square error matrices for stg(*$\hat{T}_{Rc}$*, StRS) and stg(*$\hat{T}_{Rs}$*, StRS) using the auxiliary vector from example 19 with* $\mathbf{N} = (10, 10)$ *and* $\mathbf{n} = (2, 2)$*. Using the nonsingular transformation* $\mathbf{z} = \mathrm{diag}(\mathbf{x})^{-1}\mathbf{y}$*, as in example 20, it follows that* $\mathbf{A}(\hat{T}_{Rs}, StRS)$ *is orthogonal only to vectors that are constant within strata and the intersection of the nullspaces of* $\mathbf{A}(\hat{T}_{Rs}, StRS)$ *and* $\mathbf{A}(\hat{T}_{Rc}, StRS)$ *is equal to constant vectors. So we can apply theorem 11 to these strategies. The following values were approximated:* $\eta_0 = 19888.35$ *and*

$\gamma_0 = 79.87337$. *Then by theorem 11 a sufficient condition for* $\mathrm{MSE}(\hat{T}_{Rs}, StRS)$ *to be less than* $\mathrm{MSE}(\hat{T}_{Rc}, StRS)$ *is if* $R_Z \leq 0.004016088$ *and by theorem 13 if*

$$\mathrm{cv}(\mathbf{z}) \geq d \quad \textit{and} \quad \mathrm{cv}(\mathbf{z}_h) \leq r\sqrt{\frac{N_h}{N_h - 1}}$$

for all $h = 1, \ldots, H$, *then* $\mathrm{MSE}(\hat{T}_{Rs}, StRS)$ *will be less than* $\mathrm{MSE}(\hat{T}_{Rc}, StRS)$ *where the values of* d *and* r *are given in table 4.4.*

Table 4.4: Values of d and r for example 25

d	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
r	$10^{-3} >$	$10^{-3} >$	$10^{-3} >$	$10^{-3} >$	0.001	0.001	0.002	0.002	0.002

Note that the values of the bounds in example 25, although valid for the sufficient conditions of the superiority of $stg(\hat{T}_{Rs}, \text{StRS})$, are extremely small. These results may seem strange since $stg(\hat{T}_{Rs}, \text{StRS})$ is calibrated for any linear combination of the $\mathbf{x}_h$'s (whereas the only linear combination of the $\mathbf{x}_h$'s for which $stg(\hat{T}_{Rc}, \text{StRS})$ is calibrated is $\mathbf{x}$) we would expect $\text{MSE}(\hat{T}_{Rs}, \text{StRS})$ to be less than $\text{MSE}(\hat{T}_{Rc}, \text{StRS})$ when R_Z or the $\text{cv}(\mathbf{z}_h)$'s are small. But we wouldn't expect the upper bounds on R_Z and $\text{cv}(\mathbf{z}_h)$ to be as small as they are in example 25.

The reason why the values of the bounds in example 25 were so small can be explained by examining the nullspace of the matrix

$$\mathbf{A}(\hat{T}_{Rc}, \text{StRS}) = \text{diag}(\mathbf{x})\mathbf{M}(\hat{T}_{Rc}, \text{StRS})\text{diag}(\mathbf{x}).$$

We found that the nullspace of the matrix $\mathbf{A}(\hat{T}_{Rc}, \text{StRS})$ in example 25 is of dimension 2. This means that the nullspace of $\mathbf{A}(\hat{T}_{Rc}, \text{StRS})$ is not equal to constant vectors but rather the constant vectors are part of the nullspace of $\mathbf{A}(\hat{T}_{Rc}, \text{StRS})$. Therefore there are vectors that belong in $\mathcal{N}(\mathbf{A}(\hat{T}_{Rc}, \text{StRS}))$ that are not necessarily equal to a constant vector. This implies that for some $\mathbf{z} \in \mathcal{N}(\mathbf{A}(\hat{T}_{Rc}, \text{StRS}))$ that is not equal to a constant vector, we have

$$\text{MSE}(\hat{T}_{Rc}, \text{StRS}) = \mathbf{z}^t\mathbf{A}(\hat{T}_{Rc}, \text{StRS})\mathbf{z} = \mathbf{y}^t\mathbf{M}(\hat{T}_{Rc}, \text{StRS})\mathbf{y} = 0$$

where $y_i = x_i z_i$ for all $i \in \mathcal{U}$. Since $\mathbf{y} \in \mathcal{N}(\mathbf{M}(\hat{T}_{Rc}, \text{StRS}))$ it follows from lemma 11 that the combined ratio estimator under stratified random sampling is calibrated for

$$\mathbf{y}^t = (x_1 z_1, \ldots, x_N z_N).$$

But $\mathbf{z}$ is not equal to a constant vector here and so $\mathbf{y}$ is not proportional to $\mathbf{x}$. This means that $stg(\hat{T}_{Rc}, \text{StRS})$ is calibrated for other vectors which are not necessarily proportional to $\mathbf{x}$.

Now by taking a vector $\mathbf{z}$ that is not equal to a constant vector and belongs in

$\mathcal{N}(\mathbf{A}(\hat{T}_{Rc}, \text{StRS}))$ where

$$\mathbf{z} = \begin{pmatrix} 99.68093 \\ 99.68093 \\ 99.68463 \\ 99.67705 \\ 99.66426 \\ 99.70657 \\ 99.67705 \\ 99.69157 \\ 99.71180 \\ 99.67705 \\ 99.99142 \\ 99.99429 \\ 99.96784 \\ 100.00366 \\ 99.96590 \\ 100.06332 \\ 99.99142 \\ 100.00040 \\ 99.99142 \\ 99.97628 \end{pmatrix}$$

we found that its value for R_Z is equal to 0.01769623 and also $\text{cv}(\mathbf{z}) = 0.001604046$, $\text{cv}(\mathbf{z}_1) = 0.0001449614$ and $\text{cv}(\mathbf{z}_2) = 0.0002737544$. The values of these measures are very small and yet we have

$$\text{MSE}(\hat{T}_{Rc}, \text{StRS}) = \mathbf{z}^t \mathbf{A}(\hat{T}_{Rc}, \text{StRS})\mathbf{z} = 0$$

whereas

$$\text{MSE}(\hat{T}_{Rs}, \text{StRS}) = \mathbf{z}^t \mathbf{A}(\hat{T}_{Rs}, \text{StRS})\mathbf{z} = 78.90973.$$

Hence this can explain why the values of the bounds in example 25 are so small. Although $stg(\hat{T}_{Rs}, \text{StRS})$ is calibrated for any linear combination of $\mathbf{x}_h$, $h = 1, \ldots, H$, and should perform well when R_Z or the $\text{cv}(\mathbf{z}_h)$'s are small, there still can be vectors in $\mathcal{R}^N$ with small values of R_Z or $\text{cv}(\mathbf{z}_h)$, $h = 1, \ldots, H$, for which $stg(\hat{T}_{Rc}, \text{StRS})$ can perform better.

4.3 Approximations & simulation of the MSE matrix

For large population and sample sizes calculating the mean square error matrix of some strategies can be difficult and time consuming. However corollary 2 and theorems 11 and 13 can be applied to matrices that are given by large sample approximation formulae for the variance of a general linear strategy.

Consider the ratio estimator under simple random sampling. It can be shown (see Cochran 1977 p.153) that a large sample approximation formula for the variance of this strategy is given by

$$\text{Var}(\hat{T}_R, \text{SRS}) \approx N^2 \frac{(1-f)}{n} \left[\frac{\sum_{i \in \mathcal{U}} (y_i - Rx_i)^2}{N-1} \right]. \tag{4.20}$$

We will denote this variance approximation by $\text{AVar}(\hat{T}_R, \text{SRS})$. It can be verified that the expression (4.20) can be written as the following quadratic form:

$$\text{AVar}(\hat{T}_R, \text{SRS}) = \mathbf{y}^t [b(\mathbf{I}_N - a\mathbf{J}_{\mathbf{x}})(\mathbf{I}_N - a\mathbf{J}_{\mathbf{x}}^t)]\mathbf{y} \tag{4.21}$$

where $a = 1/T_X$, $b = N(N-n)/n(N-1)$ and the $N \times N$ matrix $\mathbf{J}_{\mathbf{x}}^t$ is equal to $\mathbf{x}\mathbf{1}_N^t$. Note that the matrix given in (4.21) is orthogonal to $\mathbf{x}$.

The formula for a large sample variance approximation of the separate ratio estimator under stratified random sampling is given by

$$\text{AVar}(\hat{T}_{Rs}, \text{StRS}) = \sum_{h=1}^{H} N_h^2 \frac{(1-f_h)}{n_h} \left[\frac{\sum_{i \in \mathcal{U}_h} (y_i - R_h x_i)^2}{N_h - 1} \right]$$

where $R_h = \bar{Y}_h / \bar{X}_h$ and $f_h = n_h / N_h$ for $h = 1, \ldots, H$. This follows directly from (4.20) since sampling is independent in each stratum with simple random sampling applied to each stratum.

The matrix given by the quadratic form of $\text{AVar}(\hat{T}_{Rs}, \text{StRS})$ is just a block diagonal $N \times N$ matrix with the h^{th} block of size $N_h \times N_h$ equivalent to (4.21) with population size N_h, sample size n_h and $a_h = 1/T_{X_h}$ for $h = 1, \ldots, H$. The matrix given by $\text{AVar}(\hat{T}_{Rs}, \text{StRS})$ will be orthogonal to the strata vectors $\mathbf{x}_h$ for $h = 1, \ldots, H$. Hence we can apply theorems 11 and 13 to obtain sufficient conditions for the superiority of $stg(\hat{T}_{Rs}, \text{StRS})$ over $stg(\hat{T}_R, \text{SRS})$. We will illustrate this now in the following example.

Example 26 *Consider the matrices given by the variance approximation formula for* $stg(\hat{T}_R, SRS)$ *and* $stg(\hat{T}_{Rs}, StRS)$*. Using the same population data as in example 24 with* $\mathbf{N} = (10, 10)$ *and* $\mathbf{n} = (2, 2)$ *we approximated the values* $\eta_0 = 9921.015$ *and* $\gamma_0 = 7465.619$ *so by theorem 11 a sufficient condition for* $\text{AVar}(\hat{T}_{Rs}, StRS)$ *to be less than* $\text{AVar}(\hat{T}_R, SRS)$ *is if* $R_Z \leq 0.7525056$ *or, by theorem 13, if*

$$\text{cv}(\mathbf{z}) \geq d \quad \textit{and} \quad \text{cv}(\mathbf{z}_h) \leq r \sqrt{\frac{N_h}{N_h - 1}}$$

for all $h = 1, \ldots, H$*, then* $\text{AVar}(\hat{T}_{Rs}, StRS)$ *will be less than* $\text{AVar}(\hat{T}_R, SRS)$ *where the values of* d *and* r *are given in table 4.5.*

Table 4.5: Values of d and r for example 26

d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
r	0.007	0.028	0.063	0.110	0.169	0.237	0.314	0.398	0.486	0.579

From example 26 we see that the sufficient condition for the superiority of the separate ratio estimator, based on the variance approximations, are practically the same as the ones based on the exact mean square errors from example 24.

For large populations we can simulate the mean square error matrices of a general linear strategy. We illustrate this in the following example.

Example 27 *Using the same population data as in the previous example, we simulated 100,000 samples to calculate the mean square error matrices of* $stg(\hat{T}_R, SRS)$ *and* $stg(\hat{T}_{Rs}, StRS)$*. We approximated the values* $\eta_0 = 9988.101$ *and* $\gamma_0 = 7575.263$ *so by theorem 11 a sufficient condition for* $\text{MSE}(\hat{T}_{Rs}, StRS)$ *to be less than* $\text{MSE}(\hat{T}_R, StRS)$ *given by the simulated mean square error matrices is if* $R_Z \leq 0.7584287$ *or, by theorem 13 if*

$$\text{cv}(\mathbf{z}) \geq d \quad \textit{and} \quad \text{cv}(\mathbf{z}_h) \leq r\sqrt{\frac{N_h}{N_h - 1}}$$

for all $h = 1, \ldots, H$ *then* $\text{MSE}(\hat{T}_{Rs}, StRS)$ *will be less than* $\text{MSE}(\hat{T}_R, StRS)$ *where the values of* d *and* r *are given in table 4.6.*

Table 4.6: Values of d and r for example 27

d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
r	0.007	0.029	0.064	0.111	0.170	0.240	0.317	0.402	0.492	0.586

From example 27 we see that the sufficient condition for the superiority of the separate ratio estimator, based on the simulated mean square error matrices, are practically the same as the ones based on the exact mean square errors from example 24.

4.4 Conclusions

The method for comparing strategies using corollary 2 and theorems 11 and 13 is an alternative to simulation studies where population vectors which are thought to be similar to $\mathbf{y}$ are simulated in order to compare different strategies. Corollary 2 and theorems 11 and 13 are based on the exact mean square error of general linear strategies and are theoretically valid for any population and sample sizes. They can also be applied to variance approximations and simulated mean square error matrices. For complicated strategies, corollary 2 and theorems 11 and 13 can be more practically applied to moderate population and small sample sizes where approximation methods are not valid.

However simulation studies focus more on the assumed population structure of the y's and can reveal things that may be concealed by the exact methods of corollary 2 and theorems 11 and 13. For example, for a population that is stratified by some categorical variable we could simulate various population vectors with different values of R_Y. By calculating the mean square errors of a stratified strategy and an unstratified strategy for each simulated vector, the simulation can provide evidence to see how small R_Y would have to be in order for the stratified strategy to always be better. This can be achieved by means of theorem 11 which gives a sufficient condition for the superiority of a stratified strategy based on R_Y. But theorem 11 does not tell us the magnitude of the difference between the mean square errors of the strategies for different values of R_Y, whereas simulation studies can offer evidence that show how much these differences can be.

Simulation studies of $\mathbf{y}$ could indicate the general performance of strategies for vectors with a particular structure. For example our simulation study could show that the stratified strategy is better than the unstratified strategy for 95% of the simulated vectors with the same value for R_Y, whereas the exact methods of corollary 2 and theorems 11 and 13 are only useful if their conditions are satisfied. But there is no harm in calculating or simulating the mean square error matrices and using corollary 2 and theorems 11 or 13 at the same time if we do a simulation study.

Chapter 5

Poststratification

In this chapter we will analyse estimators for the total T_Y under poststratification. In the first three sections we compare estimators unconditionally, weak conditionally and conditionally on $\mathbf{n}$ under simple random sampling using methods described in chapter 4. We extend the use of these methods for comparison to domains of study conditionally on n_h in section 5.4. In section 5.5 we cover poststratification under unequal probability sampling designs and compare estimators unconditionally, weak conditionally and conditionally on $\mathbf{n}$. In the last section we empirically investigate the differences between estimators with different inclusion probability weights.

5.1 Unconditional comparison (SRS)

The following five strategies $stg(\hat{T}_0, \text{SRS})$, $stg(\hat{T}_{st}, \text{SRS})$, $stg(\hat{T}_U, \text{SRS})$, $stg(\hat{T}_{AD}, \text{SRS})$ and $stg(\hat{T}_D, \text{SRS})$ were studied by Rao (1985) who advocates $stg(\hat{T}_D, \text{SRS})$ which is invariant under translation and consistent. Tillé (1998) also studied these strategies and advocates $stg(\hat{T}_{AD}, \text{SRS})$ since it is calibrated for vectors that are constant within strata with respect to all samples with $n_h > 0$ for all $h = 1, \ldots, H$.

In this section we consider estimators for T_Y under unconditional simple random sampling. We begin with a theorem that can be used to compare strategies unconditionally for the special case where the poststrata sizes are equal.

Theorem 14 *Suppose the poststrata sizes N_h are constant i.e. $N_h = N_\varsigma = N/H$ for all $h = 1, \ldots, H$, and let $stg(\hat{T}_1, p_1)$ and $stg(\hat{T}_2, p_2)$ be two general linear strategies for the total T_Y such that their mean square error matrix can be written in the same form as (4.8) so that:*

$$\mathrm{MSE}(\hat{T}_1, p_1) = \lambda_1 (N_\varsigma - 1) \sum_{h=1}^{H} S^2_{Y_h} + \lambda_2 N_\varsigma \sum_{h=1}^{H} (\bar{Y}_h - \bar{Y})^2 + \lambda_3 N \bar{Y}^2$$

and

$$\mathrm{MSE}(\hat{T}_2, p_2) = \mu_1 (N_\varsigma - 1) \sum_{h=1}^{H} S^2_{Y_h} + \mu_2 N_\varsigma \sum_{h=1}^{H} (\bar{Y}_h - \bar{Y})^2 + \mu_3 N \bar{Y}^2$$

where λ_l and μ_l, for $l = 1$, 2 and 3, are the eigenvalues of $\mathbf{M}(\hat{T}_1, p_1)$ and $\mathbf{M}(\hat{T}_2, p_2)$, respectively and $\bar{Y} \neq 0$. When

a) $\mu_1 > \lambda_1$, $\mu_2 > \lambda_2$ and $\mu_3 < \lambda_3$, a sufficient condition for

$$\mathrm{MSE}(\hat{T}_2, p_2) \leq \mathrm{MSE}(\hat{T}_1, p_1)$$

is that

$$\mathrm{cv}(\mathbf{y}) \leq \sqrt{\frac{(\lambda_3 - \mu_3)N}{\underset{l=1,2}{max}\{\mu_l - \lambda_l\}(N-1)}}$$

and a sufficient condition for

$$\mathrm{MSE}(\hat{T}_1, p_1) \leq \mathrm{MSE}(\hat{T}_2, p_2)$$

is that

$$\mathrm{cv}(\mathbf{y}) \geq \sqrt{\frac{(\lambda_3 - \mu_3)N}{\underset{l=1,2}{min}\{\mu_l - \lambda_l\}(N-1)}}.$$

b) $\mu_1 > \lambda_1$, $\mu_2 < \lambda_2$ and $\mu_3 > \lambda_3$, then a sufficient condition for

$$\mathrm{MSE}(\hat{T}_1, p_1) \leq \mathrm{MSE}(\hat{T}_2, p_2)$$

is that

$$R_Y \geq \frac{(\lambda_2 - \mu_2)}{[(\lambda_2 - \mu_2) + (\mu_1 - \lambda_1)]}$$

where R_Y is given in (4.15). Furthermore this is also a necessary and sufficient condition when $\lambda_3 = \mu_3$ in which case a necessary and sufficient condition for

$$\mathrm{MSE}(\hat{T}_2, p_2) \leq \mathrm{MSE}(\hat{T}_1, p_1)$$

is that

$$R_Y \leq \frac{(\lambda_2 - \mu_2)}{[(\lambda_2 - \mu_2) + (\mu_1 - \lambda_1)]}.$$

Proof For part (a) observe that

$$\begin{aligned}
&\mathrm{MSE}(\hat{T}_2, p_2) - \mathrm{MSE}(\hat{T}_1, p_1) = \\
&(\mu_1 - \lambda_1)(N_\varsigma - 1)\sum_{h=1}^{H} S_{Y_h}^2 + (\mu_2 - \lambda_2)N_\varsigma \sum_{h=1}^{H}(\bar{Y}_h - \bar{Y})^2 + (\mu_3 - \lambda_3)N\bar{Y}^2.
\end{aligned}$$

Since the ANOVA identity in (4.14) gives

$$\begin{aligned}
(N-1)S_Y^2 &= \sum_{h=1}^{H}(N_h - 1)S_{Y_h}^2 + \sum_{h=1}^{H} N_h(\bar{Y}_h - \bar{Y})^2 \\
&= (N_\varsigma - 1)\sum_{h=1}^{H} S_{Y_h}^2 + N_\varsigma \sum_{h=1}^{H}(\bar{Y}_h - \bar{Y})^2
\end{aligned}$$

we have

$$\begin{aligned}
&\mathrm{cv}(\mathbf{y})^2 = \frac{S_Y^2}{\bar{Y}^2} \leq \frac{(\lambda_3 - \mu_3)N}{\max\limits_{l=1,2}\{\mu_l - \lambda_l\}(N-1)} \\
\Leftrightarrow \quad & \max_{l=1,2}\{\mu_l - \lambda_l\}(N-1)S_Y^2 + (\mu_3 - \lambda_3)N\bar{Y}^2 \leq 0 \\
\Leftrightarrow \quad & \max_{l=1,2}\{\mu_l - \lambda_l\}\left[(N_\varsigma - 1)\sum_{h=1}^{H} S_{Y_h}^2 + N_\varsigma \sum_{h=1}^{H}(\bar{Y}_h - \bar{Y})^2\right] + (\mu_3 - \lambda_3)N\bar{Y}^2 \leq 0 \\
\Rightarrow \quad & \mathrm{MSE}(\hat{T}_2, p_2) - \mathrm{MSE}(\hat{T}_1, p_1) \leq 0.
\end{aligned}$$

Thus the condition on the overall coefficient of variation is sufficient for the superiority of $stg(\hat{T}_2, p_2)$ over $stg(\hat{T}_1, p_1)$.

The proof of

$$\text{cv}(\mathbf{y}) \geq \sqrt{\frac{(\lambda_3 - \mu_3)N}{\underset{l=1,2}{min}\{\mu_l - \lambda_l\}(N-1)}} \quad \Rightarrow \quad \text{MSE}(\hat{T}_1, p_1) \leq \text{MSE}(\hat{T}_2, p_2)$$

follows in a similar fashion.

For part (b) observe that if $\text{MSE}(\hat{T}_2, p_2) \geq \text{MSE}(\hat{T}_1, p_1)$ then this not only implies but is equivalent to

$$\begin{aligned}
& (\mu_1 - \lambda_1)(N_\varsigma - 1)\sum_{h=1}^{H} S_{Y_h}^2 \geq (\lambda_2 - \mu_2)N_\varsigma \sum_{h=1}^{H}(\bar{Y}_h - \bar{Y})^2 + (\lambda_3 - \mu_3)N\bar{Y}^2 \\
\Leftrightarrow \quad & (\mu_1 - \lambda_1)R_Y \geq (\lambda_2 - \mu_2)(1 - R_Y) + \frac{(\lambda_3 - \mu_3)N}{(N-1)\text{cv}(\mathbf{y})^2} \\
\Leftrightarrow \quad & R_Y \geq [(\lambda_2 - \mu_2) + (\mu_1 - \lambda_1)]^{-1}\left((\lambda_2 - \mu_2) + \frac{(\lambda_3 - \mu_3)N}{(N-1)\text{cv}(\mathbf{y})^2}\right). \qquad (5.1)
\end{aligned}$$

When $\mu_3 > \lambda_3$ the second term in the right hand side of (5.1) is always less than zero and so a sufficient lower bound on R_Y is given by

$$R_Y \geq \frac{(\lambda_2 - \mu_2)}{[(\lambda_2 - \mu_2) + (\mu_1 - \lambda_1)]}.$$

When $\lambda_3 = \mu_3$ the second term on the right hand side of (5.1) will always be equal to zero and then the condition above will be necessary and sufficient for $\text{MSE}(\hat{T}_1, p_1) \leq \text{MSE}(\hat{T}_2, p_2)$. Similarly

$$R_Y \leq \frac{(\lambda_2 - \mu_2)}{[(\lambda_2 - \mu_2) + (\mu_1 - \lambda_1)]}$$

will be a necessary and sufficient condition for $\text{MSE}(\hat{T}_2, p_2) \leq \text{MSE}(\hat{T}_1, p_1)$ when $\lambda_3 = \mu_3$. □

Note that there are eight different possible relationships between the λ's and the μ's that can be used to derive conditions for superiority of one strategy over another for pairs of strategies whose mean square error matrices are of the same form as (4.8). In theorem 14 we only considered two of the eight possible cases, case (a) and case (b). This is because these are the only two cases that are applicable when comparing the five strategies mentioned at the beginning of this section.

Theorem 14 is based on pairs of general linear strategies whose mean square error matrices are of the same form as (4.8) and part (a) is just a special case of corollary 2. Since there are explicit formulae for the eigenvalues of these matrices (see theorem 9 on p.128), we can calculate them without having to evaluate the matrices.

We now consider comparison of the expansion estimator and the stratified estimator under the unconditional simple random sampling design.

For the general case of unequal poststrata sizes it has been shown by Sugden & Smith (2006) that the mean square error of $\hat{T}_{st}$ under unconditional simple random sampling is given by

$$\begin{aligned}\text{MSE}(\hat{T}_{st}, \text{SRS}) = \sum_{h=1}^{H} V_h^* p_h \quad &+ \quad \sum_{h=1}^{H} T_{Y_h}^2 (1 - p_h) \\ &+ \quad \sum_{h=1}^{H} \sum_{\substack{g=1 \\ g \neq h}}^{H} T_{Y_h} T_{Y_g} (p_{hg} + 1 - p_h - p_g) \qquad (5.2)\end{aligned}$$

with

$$V_h^* = \sum_{n_h=1}^{n} N_h^2 S_{Y_h}^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \frac{\Pr(n_h)}{p_h}$$

and

$$\Pr(n_h) = \frac{\binom{N-N_h}{n-n_h} \binom{N_h}{n_h}}{\binom{N}{n}}$$

for $h = 1, \ldots, H$, and where

$$p_h = \sum_{n_h=1}^{n} \Pr(n_h)$$

is the probability of selecting a sample such that $n_h > 0$,

$$\begin{aligned} p_{hg} &= \Pr(n_h > 0 \cap n_g > 0) = 1 - \Pr(n_h = 0 \cup n_g = 0) \\ &= 1 - [\Pr(n_h = 0) + \Pr(n_g = 0) - \Pr(n_h = 0 \cap n_g = 0)] \\ &= \Pr(n_h > 0) + \Pr(n_g > 0) + \Pr(n_h = 0 \cap n_g = 0) - 1 \\ &= p_h + p_g + p^0_{hg} - 1 \end{aligned}$$

so that

$$p^0_{hg} = \Pr(n_h = 0 \cap n_g = 0) = \frac{\binom{N-N_h-N_g}{n}}{\binom{N}{n}} = p_{hg} + 1 - p_h - p_g.$$

First we derive the mean square error matrix for $stg(\hat{T}_{st}, \text{SRS})$. For unit $i \in \mathcal{U}_h$, $h = 1, \ldots, H$, we have

$$\begin{aligned} B_{hi} = \sum_{s_h \ni i} p_h(s_h) b_{s_h i} &= \binom{N}{n}^{-1} \sum_{s_h \ni i} \binom{N - N_h}{n - n_h} \frac{N_h}{n_h} \\ &= \binom{N}{n}^{-1} \sum_{n_h=1}^{n} \binom{N - N_h}{n - n_h} \binom{N_h - 1}{n_h - 1} \frac{N_h}{n_h} \\ &= \binom{N}{n}^{-1} \sum_{n_h=1}^{n} \binom{N - N_h}{n - n_h} \binom{N_h}{n_h} \\ &= \sum_{n_h=1}^{n} \Pr(n_h) = p_h \end{aligned}$$

and similarly

$$\sum_{s_h \ni i} p_h(s_h) b^2_{s_h i} = \binom{N}{n}^{-1} \sum_{s_h \ni i} \binom{N - N_h}{n - n_h} \frac{N_h^2}{n_h^2} = \sum_{n_h=1}^{n} \Pr(n_h) \frac{N_h}{n_h}.$$

For $i \neq j \in \mathcal{U}_h$ we have

$$\begin{aligned}
\sum_{s_h \ni i,j} p_h(s_h) b_{s_h i} b_{s_h j} &= \binom{N}{n}^{-1} \sum_{s_h \ni i,j} \binom{N - N_h}{n - n_h} \frac{N_h^2}{n_h^2} \\
&= \binom{N}{n}^{-1} \sum_{n_h=2}^{n} \binom{N - N_h}{n - n_h} \binom{N_h - 2}{n_h - 2} \frac{N_h^2}{n_h^2} \\
&= \binom{N}{n}^{-1} \sum_{n_h=2}^{n} \binom{N - N_h}{n - n_h} \binom{N_h}{n_h} \frac{N_h(n_h - 1)}{n_h(N_h - 1)} \\
&= \sum_{n_h=1}^{n} \Pr(n_h) \frac{N_h(n_h - 1)}{n_h(N_h - 1)}
\end{aligned}$$

and for $i \in \mathcal{U}_h$ and $j \in \mathcal{U}_g$ for $g \neq h$ we have

$$\begin{aligned}
\sum_{\substack{s_h \ni i \\ s_g \ni j}} p_{hg}(s_h, s_g) b_{s_h i} b_{s_g j} &= \binom{N}{n}^{-1} \sum_{\substack{s_h \ni i \\ s_g \ni j}} \binom{N - N_h - N_g}{n - n_h - n_g} \frac{N_h N_g}{n_h n_g} \\
&= \binom{N}{n}^{-1} \sum_{\substack{n_h > 0 \\ n_g > 0}} \binom{N - N_h - N_g}{n - n_h - n_g} \binom{N_h - 1}{n_h - 1} \binom{N_g - 1}{n_g - 1} \frac{N_h N_g}{n_h n_g} \\
&= \binom{N}{n}^{-1} \sum_{\substack{n_h > 0 \\ n_g > 0}} \binom{N - N_h - N_g}{n - n_h - n_g} \binom{N_h}{n_h} \binom{N_g}{n_g} \\
&= p_{hg}.
\end{aligned}$$

Hence by (3.3) the mean square error matrix $\mathbf{M}(\hat{T}_{st}, \mathrm{SRS})$ has ij^{th} element

$$\mathbf{M}(\hat{T}_{st}, \mathrm{SRS})_{ij} = \begin{cases} \sum_{n_h=1}^{n} \frac{N_h}{n_h} \Pr(n_h) - 2p_h + 1 & \text{for} \quad i = j \in \mathcal{U}_h \\ \sum_{n_h=1}^{n} \frac{N_h(n_h-1)}{n_h(N_h-1)} \Pr(n_h) - 2p_h + 1 & \text{for} \quad i \neq j \in \mathcal{U}_h \\ p_{hg} + 1 - p_h - p_g = p_{hg}^0 & \text{for} \quad i \in \mathcal{U}_h \ \&\ j \in \mathcal{U}_g \end{cases}$$

with h, $g = 1, \ldots, H$ $(h \neq g)$. Since the matrix $\mathbf{M}(\hat{T}_{st}, \mathrm{SRS})$ is of the same form as (4.10) it follows that it has eigenvalues equal to

$$\sum_{n_h=1}^{n} \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \Pr(n_h)$$

of multiplicity $N_h - 1$ for $h = 1, \ldots, H$.

We now consider an important special case where $N_h = N/H = N_\varsigma$ for all $h = 1, \ldots, H$. Then the matrix $\mathbf{M}(\hat{T}_{st}, \text{SRS})$ reduces to

$$\mathbf{M}(\hat{T}_{st}, \text{SRS})_{ij} = \begin{cases} a = N_\varsigma \sum_{n_h=1}^{n} \frac{\Pr(n_h)}{n_h} - 2p_\varsigma + 1 & \text{for} \quad i = j \in \mathcal{U}_h \\ b = \frac{N_\varsigma}{(N_\varsigma - 1)} \sum_{n_h=1}^{n} \frac{(n_h - 1)}{n_h} \Pr(n_h) - 2p_\varsigma + 1 & \text{for} \quad i \neq j \in \mathcal{U}_h \\ c = p_{\varsigma\varsigma} + 1 - 2p_\varsigma = p^0 & \text{for} \quad i \in \mathcal{U}_h \ \& \ j \in \mathcal{U}_g \end{cases}$$

where

$$p_\varsigma = 1 - \binom{N}{n}^{-1} \binom{N - N_\varsigma}{n}$$

for all $h = 1, \ldots, H$,

$$p^0 = \binom{N}{n}^{-1} \binom{N - 2N_\varsigma}{n}$$

so that

$$p_{\varsigma\varsigma} = p^0 + 2p_\varsigma - 1$$

for all $h \neq g = 1, \ldots, H$. Then by theorem 9 the eigenvalues of $\mathbf{M}(\hat{T}_{st}, \text{SRS})$ can be written as

$$\lambda_1 = a - b = \frac{N_\varsigma}{(N_\varsigma - 1)} \sum_{n_h=1}^{n} \frac{(N_\varsigma - n_h)}{n_h} \Pr(n_h) \tag{5.3}$$

of multiplicity $(N_\varsigma - 1)H = N - H$,

$$\lambda_2 = (a - b) + N_\varsigma(b - c) = N_\varsigma(p_\varsigma - p_{\varsigma\varsigma}) \tag{5.4}$$

of multiplicity $H - 1$ and

$$\lambda_3 = (a - b) + N_\varsigma(b - c) + Nc = N_\varsigma[(p_\varsigma - p_{\varsigma\varsigma}) + Hp^0] \tag{5.5}$$

of multiplicity one. Hence we can write

$$\text{MSE}(\hat{T}_{st}, \text{SRS}) = \lambda_1(N_\varsigma - 1) \sum_{h=1}^{H} S_{Y_h}^2 + \lambda_2 N_\varsigma \sum_{h=1}^{H} (\bar{Y}_h - \bar{Y})^2 + \lambda_3 N \bar{Y}^2.$$

The mean square error matrix of $stg(\hat{T}_0, \text{SRS})$ is given by (see example 17 p.124)

$$\mathbf{M}(\hat{T}_0, \text{SRS}) = \frac{N(N - n)}{n(N - 1)} (\mathbf{I}_N - \mathbf{J}_N / N)$$

and its only nonzero eigenvalue is equal to

$$\mu = \frac{N(N-n)}{n(N-1)}$$

of multiplicity $N-1$. Then we can write

$$\mathrm{MSE}(\hat{T}_0, \mathrm{SRS}) = \mu(N_\varsigma - 1)\sum_{h=1}^{H} S_{Y_h}^2 + \mu N_\varsigma \sum_{h=1}^{H} (\bar{Y}_h - \bar{Y})^2 + 0 \times N\bar{Y}^2.$$

We now compare these strategies using theorem 14 in the following example:

Example 28 *For* $\mathbf{N} = (10, 10, 10)$ *and* $n = 6$ *the eigenvalues of* $\mathbf{M}(\hat{T}_{st}, SRS)$*, given by (5.3), (5.4) and (5.5), are respectively*

$$\mu_1 = 4.962018, \quad \mu_2 = 0.6492358, \quad \mu_3 = 0.6598459$$

so that

$$\mathrm{MSE}(\hat{T}_{st}, SRS) = 9\mu_1 \sum_{h=1}^{H} S_{Y_h}^2 + 10\mu_2 \sum_{h=1}^{H} (\bar{Y}_h - \bar{Y})^2 + 30\mu_3 \bar{Y}^2.$$

The eigenvalues of $\mathbf{M}(\hat{T}_0, SRS)$ *are*

$$\lambda_1 = 4.137931, \quad \lambda_2 = 4.137931, \quad \lambda_3 = 0$$

so that

$$\mathrm{MSE}(\hat{T}_0, SRS) = 9\lambda_1 \sum_{h=1}^{H} S_{Y_h}^2 + 10\lambda_2 \sum_{h=1}^{H} (\bar{Y}_h - \bar{Y})^2 + 30\lambda_3 \bar{Y}^2.$$

Since $\mu_1 > \lambda_1$*,* $\mu_2 < \lambda_2$ *and* $\mu_3 > \lambda_3$ *we can apply theorem 14 part (b) which gives a sufficient condition for* $\mathrm{MSE}(\hat{T}_0, SRS) \leq \mathrm{MSE}(\hat{T}_{st}, SRS)$*:*

$$R_Y \geq \frac{(\lambda_2 - \mu_2)}{[(\lambda_2 - \mu_2) + (\mu_1 - \lambda_1)]} = 0.8089198662.$$

We can also apply corollary 2 part (a) to give a sufficient condition for the superiority of $stg(\hat{T}_0, SRS)$ *over* $stg(\hat{T}_{st}, SRS)$*. We approximated the values* $\alpha_0 = 4.148541$ *and* $\gamma_1(\alpha_0) = 0.6598458841$ *which gives*

$$r_1(\alpha_0) = \frac{\gamma_1(\alpha_0)}{(\alpha_0 - \gamma(\alpha_0))} = 0.1891383059.$$

Hence by corollary 2 part (a) we have

$$\begin{aligned} \text{cv}(\mathbf{y}) &\leq \sqrt{r_1(\alpha_0)\frac{N}{N-1}} = 0.4430092432 \\ \Rightarrow \quad \text{MSE}(\hat{T}_0, SRS) &\leq \text{MSE}(\hat{T}_{st}, SRS). \end{aligned}$$

In a side note if the number of poststrata was $H = 2$ then $p_{hg}^0 = 0$ and the matrix $\mathbf{M}(\hat{T}_{st}, \text{SRS})$ is a block diagonal with eigenvalues (by lemma 12) equal to

$$\sum_{n_h=1}^{n} \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \Pr(n_h)$$

of multiplicity $N_h - 1$ and

$$\begin{aligned} \left(\sum_{n_h=1}^{n} \frac{N_h}{n_h}\Pr(n_h) - 2p_h + 1\right) &+ (N_h - 1)\left(\sum_{n_h=1}^{n} \frac{N_h(n_h - 1)}{n_h(N_h - 1)}\Pr(n_h) - 2p_h + 1\right) \\ &= N_h(1 - p_h) \end{aligned}$$

of multiplicity one for $h = 1, 2$.

Now back to the general case. Särndal *et al.* (1992, p.268) gave the following approximation for large n

$$\frac{\text{MSE}(\hat{T}_{st}, \text{SRS})}{\text{MSE}(\hat{T}_0, \text{SRS})} \approx R_Y$$

and R_Y is always less than or equal to one (implying that $\hat{T}_{st}$ is always better than $\hat{T}_0$ for simple random sampling). However Sugden & Smith (2006) demonstrated, via examples, that this result can be very misleading for small N.

Our method for comparing $stg(\hat{T}_{st}, \text{SRS})$ with $stg(\hat{T}_0, \text{SRS})$, using either corollary 2 part (a) or theorem 14 part (b), gives sufficient conditions which are based on the exact mean square errors of these strategies for $stg(\hat{T}_0, \text{SRS})$ to be the superior strategy.

As an alternative to the unconditionally biased stratified estimator $\hat{T}_{st}$, an adjusted unconditionally unbiased stratified estimator, which we will denote by $\hat{T}_U$, was given by Doss *et al.* (1979) as

$$\hat{T}_U = \sum_{h=1}^{H} I_h N_h \frac{\bar{y}_h}{p_h}.$$

Sugden & Smith (2006) gave an expression for its exact mean square error as

$$\text{MSE}(\hat{T}_U, \text{SRS}) = \sum_{h=1}^{H} \frac{V_h^*}{p_h} + \sum_{h=1}^{H} T_h^2 \frac{(1-p_h)}{p_h} + \sum_{h=1}^{H} \sum_{\substack{g=1 \\ g \neq h}}^{H} T_h T_g \frac{(p_{hg} - p_h p_g)}{p_h p_g} \tag{5.6}$$

and noted that (since the coefficient of $T_h T_g$ in the last term of (5.6) is negative and positive in (5.2)) no one of these strategies is uniformly better than the other unconditionally. We will show how we can use corollary 2 and theorem 14 to give us sufficient conditions for one of these strategies, $stg(\hat{T}_{st}, \text{SRS})$ and $stg(\hat{T}_U, \text{SRS})$, to be better than the other.

We can derive the mean square error matrix $\mathbf{M}(\hat{T}_U, \text{SRS})$ in a similar way to that of $\mathbf{M}(\hat{T}_{st}, \text{SRS})$. The survey weight $b_{s_h i}$ of $\hat{T}_U$ for unit i in s_h is equal to $N_h/n_h p_h$ so that

$$B_{hi} = \sum_{s_h \ni i} p_h(s_h) b_{s_h i} = \binom{N}{n}^{-1} \sum_{s_h \ni i} \binom{N - N_h}{n - n_h} \frac{N_h}{n_h p_h} = \frac{p_h}{p_h} = 1$$

which shows the unconditional unbiasedness of $stg(\hat{T}_U, \text{SRS})$. Also

$$\sum_{s_h \ni i} p_h(s_h) b_{s_h i}^2 = \binom{N}{n}^{-1} \sum_{s_h \ni i} \binom{N - N_h}{n - n_h} \left(\frac{N_h}{n_h p_h}\right)^2 = \frac{1}{p_h^2} \sum_{n_h=1}^{n} \frac{N_h}{n_h} \Pr(n_h).$$

For units $i \neq j \in \mathcal{U}_h$ we have

$$\begin{aligned} \sum_{s_h \ni i,j} p_h(s_h) b_{s_h i} b_{s_h j} &= \binom{N}{n}^{-1} \sum_{s_h \ni i,j} \binom{N - N_h}{n - n_h} \left(\frac{N_h}{n_h p_h}\right)^2 \\ &= \frac{1}{p_h^2} \sum_{n_h=1}^{n} \frac{N_h(n_h - 1)}{n_h(N_h - 1)} \Pr(n_h) \end{aligned}$$

and for units $i \in \mathcal{U}_h$ and $j \in \mathcal{U}_g$, $g \neq h$ we have

$$\begin{aligned} \sum_{\substack{s_h \ni i \\ s_g \ni j}} p_{hg}(s_h, s_g) b_{s_h i} b_{s_g j} &= \binom{N}{n}^{-1} \sum_{\substack{s_h \ni i \\ s_g \ni j}} \binom{N - N_h - N_g}{n - n_h - n_g} \frac{N_h N_g}{n_h n_g p_h p_g} \\ &= \frac{p_{hg}}{p_h p_g}. \end{aligned}$$

Hence by (3.3) the ij^{th} element of $\mathbf{M}(\hat{T}_U, \text{SRS})$ is equal to

$$\mathbf{M}(\hat{T}_U, \text{SRS})_{ij} = \begin{cases} \frac{1}{p_h^2}\sum_{n_h=1}^{n} \frac{N_h}{n_h}\Pr(n_h) - 1 & \text{for} \quad i = j \in \mathcal{U}_h \\ \frac{1}{p_h^2}\sum_{n_h=1}^{n} \frac{N_h(n_h-1)}{n_h(N_h-1)}\Pr(n_h) - 1 & \text{for} \quad i \neq j \in \mathcal{U}_h \\ (p_{hg} - p_h p_g)/(p_h p_g) & \text{for} \quad i \in \mathcal{U}_h \ \&\ j \in \mathcal{U}_g \end{cases}$$

for h, $g = 1, \ldots, H$ $(h \neq g)$.

When the poststrata sizes are all equal this matrix reduces to

$$\mathbf{M}(\hat{T}_U, \text{SRS})_{ij} = \begin{cases} a = \frac{N_\varsigma}{p_\varsigma^2}\sum_{n_h=1}^{n} \frac{\Pr(n_h)}{n_h} - 1 & \text{for} \quad i = j \in \mathcal{U}_h \\ b = \frac{N_\varsigma}{(N_\varsigma-1)p_\varsigma^2}\sum_{n_h=1}^{n} \frac{\Pr(n_h)(n_h-1)}{n_h} - 1 & \text{for} \quad i \neq j \in \mathcal{U}_h \\ c = (p_{\varsigma\varsigma} - p_\varsigma^2)/p_\varsigma^2 & \text{for} \quad i \in \mathcal{U}_h \ \&\ j \in \mathcal{U}_g \end{cases}$$

and the eigenvalues of $\mathbf{M}(\hat{T}_U, \text{SRS})$ can be written as

$$\mu_1 = a - b = \frac{N_\varsigma}{(N_\varsigma - 1)p_\varsigma^2}\sum_{n_h=1}^{n} \frac{(N_\varsigma - n_h)}{n_h}\Pr(n_h) \tag{5.7}$$

of multiplicity $N - H$,

$$\mu_2 = (a - b) + N_\varsigma(b - c) = \frac{N_\varsigma}{p_\varsigma^2}(p_\varsigma - p_{\varsigma\varsigma}) \tag{5.8}$$

of multiplicity $H - 1$ and

$$\mu_3 = (a - b) + N_\varsigma(b - c) + Nc = \frac{N_\varsigma}{p_\varsigma^2}[(p_\varsigma - p_{\varsigma\varsigma}) + H(p_{\varsigma\varsigma} - p_\varsigma^2)] \tag{5.9}$$

of multiplicity one.

The mean square error of $stg(\hat{T}_U, \text{SRS})$ is

$$\text{MSE}(\hat{T}_U, \text{SRS}) = \mu_1(N_\varsigma - 1)\sum_{h=1}^{H} S_{Y_h}^2 + \mu_2 N_\varsigma \sum_{h=1}^{H}(\bar{Y}_h - \bar{Y})^2 + \mu_3 N\bar{Y}^2$$

so comparing with $\text{MSE}(\hat{T}_{st}, \text{SRS})$ it is easily seen from (5.3) and (5.7) that $\lambda_1 \leq \mu_1$ and from (5.4) and (5.8) we see that $\lambda_2 \leq \mu_2$ since $0 < p_\varsigma \leq 1$ and equality holds when $p_\varsigma = 1$ which can happen if the sample size is large enough. Hence provided

$\mu_3 \leq \lambda_3$ by theorem 14 part (a) a sufficient condition for $stg(\hat{T}_U, \text{SRS})$ to be better than $stg(\hat{T}_{st}, \text{SRS})$ is if

$$\text{cv}(\mathbf{y}) \leq \sqrt{\frac{(\lambda_3 - \mu_3)N}{\max\limits_{l=1,2}\{\mu_l - \lambda_l\}(N-1)}}$$

and a sufficient condition for $stg(\hat{T}_{st}, \text{SRS})$ to be better than $stg(\hat{T}_U, \text{SRS})$ is if

$$\text{cv}(\mathbf{y}) \geq \sqrt{\frac{(\lambda_3 - \mu_3)N}{\min\limits_{l=1,2}\{\mu_l - \lambda_l\}(N-1)}}.$$

Note that if $\lambda_3 < \mu_3$ then it follows that $\text{MSE}(\hat{T}_U, \text{SRS})$ will always be less than $\text{MSE}(\hat{T}_{st}, \text{SRS})$ since $\lambda_1 \leq \mu_1$ and $\lambda_2 \leq \mu_3$ will always be true.

Example 29 *Let* $\mathbf{N} = (6, 6, 6)$ *and* $n = 4$. *Then the eigenvalues of the matrix* $\mathbf{M}(\hat{T}_{st}, SRS)$*, given by (5.3), (5.4) and (5.5), are respectively*

$$\lambda_1 = 3.4618 \quad \lambda_2 = 0.9412 \quad \lambda_3 = 1.0294$$

and the eigenvalues of $\mathbf{M}(\hat{T}_U, SRS)$*, given by (5.7), (5.8) and (5.9), are respectively*

$$\mu_1 = 4.1298 \quad \mu_2 = 1.3395 \quad \mu_3 = 0.7947.$$

Then applying theorem 14 part (a),

$$\text{cv}(\mathbf{y}) \leq \sqrt{\frac{(\lambda_3 - \mu_3)N}{max\limits_{l=1,2}\{\mu_l - \lambda_l\}(N-1)}} = 0.6099$$

is a sufficient condition for

$$\text{MSE}(\hat{T}_U, SRS) \leq \text{MSE}(\hat{T}_{st}, SRS)$$

and

$$\text{cv}(\mathbf{y}) \geq \sqrt{\frac{(\lambda_3 - \mu_3)N}{min\limits_{l=1,2}\{\mu_l - \lambda_l\}(N-1)}} = 0.7899$$

implies

$$\text{MSE}(\hat{T}_{st}, SRS) \leq \text{MSE}(\hat{T}_U, SRS).$$

In general both of these mean square error matrices are of full rank. Then by using corollary 1, on p.99, we calculate their maximum relative efficiency as

$$\text{MRE}(\hat{T}_{st}, SRS \mid \hat{T}_U, SRS) = \max_{\mathbf{y} \in \mathcal{R}^N} \left\{ \frac{\text{MSE}(\hat{T}_{st}, SRS)}{\text{MSE}(\hat{T}_U, SRS)} \right\} = 1.2953365$$

and

$$\text{MRE}(\hat{T}_U, SRS \mid \hat{T}_{st}, SRS) = \max_{\mathbf{y} \in \mathcal{R}^N} \left\{ \frac{\text{MSE}(\hat{T}_U, SRS)}{\text{MSE}(\hat{T}_{st}, SRS)} \right\} = 1.423105889.$$

Hence the mean square error of $stg(\hat{T}_{st}, SRS)$ *can be at most 29.5% larger than that of* $stg(\hat{T}_U, SRS)$ *and the mean square error of* $stg(\hat{T}_U, SRS)$ *can be at most 42.3% larger than that of* $stg(\hat{T}_{st}, SRS)$*. Since these values are large it isn't clear which strategy to choose based on them.*

When the poststrata sizes are unequal we can apply corollary 2 to compare $stg(\hat{T}_U, \text{SRS})$ and $stg(\hat{T}_{st}, \text{SRS})$. We do this now in the following example.

Example 30 *Let* $\mathbf{N} = (16, 20, 24)$ *and* $n = 8$*. We apply corollary 2 to compare* $stg(\hat{T}_{st}, SRS)$ *with* $stg(\hat{T}_U, SRS)$*.*

Figure 5.1 shows a plot of $r_1(\alpha)$ *against* α *and figure 5.2 shows a plot of* $r_2(\beta)$ *against* β*. We approximated the values* $\alpha_0 = 0.5898242$*,* $\gamma_1(\alpha_0) = 0.01137159257$ *which gives* $r_1(\alpha_0) = 0.01965864174$ *and also* $\beta_0 = 0.03060958$*,* $\gamma_2(\beta_0) = 0.0156870383$ *which gives* $r_2(\beta_0) = 1.051230992$*. Hence by corollary 2 part (a) a sufficient condition for*

$$\text{MSE}(\hat{T}_U, SRS) \leq \text{MSE}(\hat{T}_{st}, SRS)$$

is if

$$\mathrm{cv}(\mathbf{y}) \leq \sqrt{r_1(\alpha_0)\frac{N}{N-1}} = 0.1413925$$

and by part (b), a sufficient condition for

$$\mathrm{MSE}(\hat{T}_{st}, \mathit{SRS}) \leq \mathrm{MSE}(\hat{T}_U, \mathit{SRS})$$

is if

$$\mathrm{cv}(\mathbf{y}) \geq \sqrt{r_2(\beta_0)\frac{N}{N-1}} = 1.033948.$$

The maximum relative efficiencies of these strategies are

$$\mathrm{MRE}(\hat{T}_U, \mathit{SRS} \mid \hat{T}_{st}, \mathit{SRS}) = 1.1103448$$

and

$$\mathrm{MRE}(\hat{T}_{st}, \mathit{SRS} \mid \hat{T}_U, \mathit{SRS}) = 1.0303971$$

which indicates that $\mathrm{MSE}(\hat{T}_{st}, \mathit{SRS})$ *can be at most 3.0% larger than* $\mathrm{MSE}(\hat{T}_U, \mathit{SRS})$ *and* $\mathrm{MSE}(\hat{T}_U, \mathit{SRS})$ *can be at most 11.0% larger than* $\mathrm{MSE}(\hat{T}_{st}, \mathit{SRS})$*. Since the potential gains in efficiency of stg*$(\hat{T}_U, \mathit{SRS})$ *is little, compared to the gains in efficiency of stg*$(\hat{T}_{st}, \mathit{SRS})$*, we may prefer to use stg*$(\hat{T}_{st}, \mathit{SRS})$ *instead of stg*$(\hat{T}_U, \mathit{SRS})$*.*

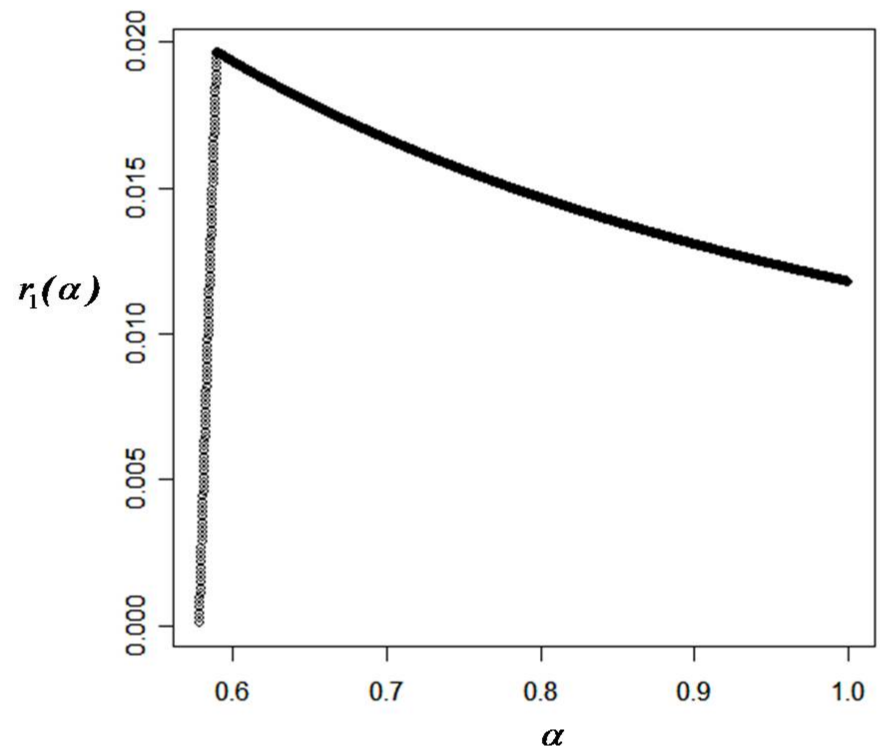

Figure 5.1: Plot of $r_1(\alpha)$ against α for example 30

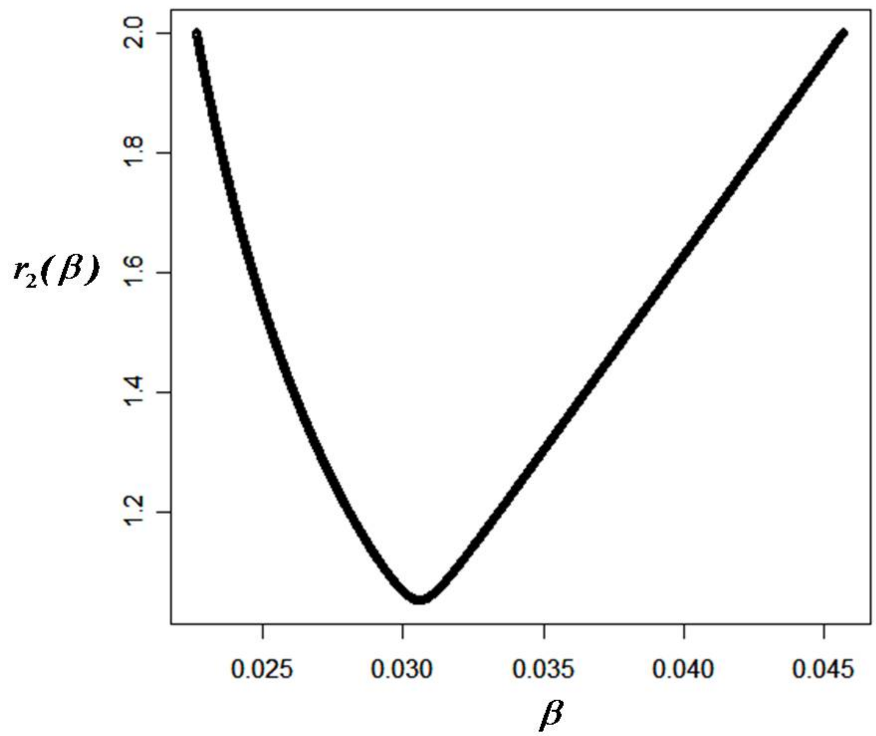

Figure 5.2: Plot of $r_2(\beta)$ against β for example 30

Another estimator proposed by Doss *et al.* (1979), denoted by $\hat{T}_D$, which is an adjustment of $\hat{T}_{st}$ and calibrated for the constant population vectors is given by

$$\hat{T}_D = N \frac{\sum_{h=1}^{H} I_h N_h \bar{y}_h / p_h}{\sum_{h=1}^{H} I_h N_h / p_h}.$$

This estimator can be written as a general linear estimator, but it is not a separate general linear estimator. Its survey weight for unit $i \in \mathcal{U}$,

$$b_{si} = N \frac{N_h / n_h p_h}{\sum_{h=1}^{H} I_h N_h / p_h},$$

depends on units in the sample from different poststrata through the denominator of b_{si}. Now

$$\begin{aligned} B_i &= \sum_{s \ni i} p(s) b_{si} \\ &= \sum_{s \ni i} \binom{N}{n}^{-1} N \frac{N_h / p_h n_h}{\sum_{h=1}^{H} I_h N_h p_h} \\ &= \sum_{n_h=1}^{n} \binom{N}{n}^{-1} \binom{N - N_h}{n - n_h} \binom{N_h - 1}{n_h - 1} \frac{N_h}{n_h p_h} \left(\sum_{s_{n_h} \ni i} \frac{N}{\sum_{h=1}^{H} I_h N_h / p_h} \right) \\ &= \sum_{n_h=1}^{n} \frac{\Pr(n_h)}{p_h} \left(\sum_{s_{n_h} \ni i} \frac{N}{\sum_{h=1}^{H} I_h N_h / p_h} \right) \end{aligned}$$

where s_{n_h} is the set of all samples s with fixed poststratum sample size n_h, for unit i in stratum h ($h = 1, \ldots, H$). The value of $B_i = B_h$ will be constant over all units in $\mathcal{U}_h$, for each $h = 1, \ldots, H$, since b_{si} doesn't depend on the individual units in the poststratum sample s_h, but only its size n_h. The strategy $stg(\hat{T}_D, \text{SRS})$ is unconditionally biased unless the poststrata sizes are of equal size, see Doss *et al.* (1979) for the proof of this. This will imply that $B_i = 1$ for all $i \in \mathcal{U}$.

For unit $i \in \mathcal{U}_h$ we have

$$\begin{aligned}
\sum_{s \ni i} p(s) b_{si}^2 &= \sum_{s \ni i} \binom{N}{n}^{-1} \left[\frac{N}{(\sum_{h=1}^H I_h N_h/p_h)} \frac{N_h}{n_h p_h} \right]^2 \\
&= \sum_{n_h=1}^{n} \binom{N}{n}^{-1} \binom{N-N_h}{n-n_h} \binom{N_h-1}{n_h-1} \frac{N_h^2}{n_h^2 p_h^2} \left[\sum_{s_{n_h} \ni i} \left(\frac{N}{\sum_{h=1}^H I_h N_h/p_h} \right)^2 \right] \\
&= N_h \sum_{n_h=1}^{n} \frac{\Pr(n_h)}{n_h p_h^2} D_{n_h i}
\end{aligned}$$

where

$$D_{n_h i} = \sum_{s_{n_h} \ni i} \left(\frac{N}{\sum_{h=1}^H I_h N_h/p_h} \right)^2$$

for $i \in \mathcal{U}_h$. Similarly for $i \neq j \in \mathcal{U}_h$, $h = 1, \ldots, H$, we have

$$\begin{aligned}
\sum_{s \ni i,j} p(s) b_{si} b_{sj} &= \sum_{s \ni i,j} \binom{N}{n}^{-1} \left[\frac{N}{(\sum_{h=1}^H I_h N_h/p_h)} \frac{N_h}{n_h p_h} \right]^2 \\
&= \sum_{n_h=2}^{n} \binom{N}{n}^{-1} \binom{N-N_h}{n-n_h} \binom{N_h-2}{n_h-2} \frac{N_h^2}{n_h^2 p_h^2} \left[\sum_{s_{n_h} \ni i,j} \frac{N^2}{(\sum_{h=1}^H I_h N_h/p_h)^2} \right] \\
&= N_h \sum_{n_h=1}^{n} \frac{\Pr(n_h)(n_h-1)}{n_h(N_h-1)p_h^2} D_{n_h ij}
\end{aligned}$$

where

$$D_{n_h ij} = \sum_{s_{n_h} \ni i,j} \left(\frac{N}{\sum_{h=1}^H I_h N_h/p_h} \right)^2 .$$

For $i \in \mathcal{U}_h$ and $j \in \mathcal{U}_g$ with $h \neq g = 1, \ldots, H$ we have

$$\begin{aligned}
\sum_{s \ni i,j} p(s) b_{si} b_{sj} &= \sum_{s \ni i,j} \binom{N}{n}^{-1} \left[\frac{N}{\sum_{h=1}^H I_h N_h/p_h} \right]^2 \frac{N_h}{n_h p_h} \frac{N_g}{n_g p_g} \\
&= \sum_{\substack{n_h > 0 \\ n_g > 0}} \binom{N}{n}^{-1} \binom{N-N_h-N_g}{n-n_h-n_g} \binom{N_h-1}{n_h-1} \binom{N_g-1}{n_g-1} \frac{N_h N_g D_{n_{hg}}}{n_h n_g p_h p_g} \\
&= \sum_{\substack{n_h > 0 \\ n_g > 0}} \Pr(n_h, n_g) \frac{D_{n_{hg}}}{p_h p_g}
\end{aligned}$$

where $\Pr(n_h, n_g) = \binom{N}{n}^{-1}\binom{N-N_h-N_g}{n-n_h-n_g}\binom{N_h}{n_h}\binom{N_g}{n_g}$ for $h \neq g$,

$$D_{n_{hg}} = \sum_{s_{n_{hg}} \ni i,j} \left(\frac{N}{\sum_{h=1}^{H} I_h N_h / p_h} \right)^2$$

and $s_{n_{hg}}$ is the set of all samples with n_h and n_g fixed for $h \neq g$. Hence the ij^{th} element of the mean square error matrix $\mathbf{M}(\hat{T}_D, \text{SRS})$ is equal to

$$\mathbf{M}(\hat{T}_D, \text{SRS})_{ij} = \begin{cases} N_h \sum_{n_h=1}^{n} \frac{\Pr(n_h)}{n_h p_h^2} D_{n_h i} - 2B_h + 1 & \text{for} \quad i = j \in \mathcal{U}_h \\ N_h \sum_{n_h=1}^{n} \frac{\Pr(n_h)(n_h-1)}{n_h(N_h-1)p_h^2} D_{n_h ij} - 2B_h + 1 & \text{for} \quad i \neq j \in \mathcal{U}_h \\ \sum_{n_h>0, n_g>0} \frac{\Pr(n_h,n_g)}{p_h p_g} D_{n_{hg}} - B_h - B_g + 1 & \text{for} \quad i \in \mathcal{U}_h \;\&\; j \in \mathcal{U}_g \end{cases}$$

for h, $g = 1. \ldots, H$. The matrix $\mathbf{M}(\hat{T}_D, \text{SRS})$ is of the same form as in (4.10) and in the case of equal sized poststrata its eigenvalues, denoted by α_l for $l = 1, 2, 3$, are equal to

$$\alpha_1 = \frac{N_\varsigma}{p_\varsigma^2} \left(\sum_{n_h=1}^{n} \frac{\Pr(n_h)}{n_h} D_{n_h i} - \sum_{n_h=1}^{n} \frac{\Pr(n_h)(n_h-1)}{n_h(N_\varsigma - 1)} D_{n_h ij} \right)$$

of multiplicity $N - H$,

$$\alpha_2 = \alpha_1 + \frac{N_\varsigma^2}{p_\varsigma^2} \left(\sum_{n_h=1}^{n} \frac{\Pr(n_h)(n_h-1)}{n_h(N_\varsigma - 1)} D_{n_h ij} - \frac{1}{N_\varsigma} \sum_{\substack{n_h>0 \\ n_g>0}} \Pr(n_h, n_g) D_{hg} \right)$$

of multiplicity $H - 1$ and

$$\alpha_3 \;\; = \;\; \alpha_2 + N \left(\sum_{\substack{n_h>0 \\ n_g>0}} \frac{\Pr(n_h, n_g)}{p_\varsigma^2} D_{hg} - 1 \right)$$

of multiplicity one. Hence when $N_h = N_\varsigma$ for all $h = 1, \ldots, H$ the mean square error of $stg(\hat{T}_D, \text{SRS})$ can be written as

$$\text{MSE}(\hat{T}_D, \text{SRS}) = \alpha_1 (N_\varsigma - 1) \sum_{h=1}^{H} S_{Y_h}^2 + \alpha_2 N_\varsigma \sum_{h=1}^{H} (\bar{Y}_h - \bar{Y})^2 + \alpha_3 N \bar{Y}^2.$$

In the following examples we compare $stg(\hat{T}_D, \text{SRS})$ with strategies $stg(\hat{T}_{st}, \text{SRS})$ and $stg(\hat{T}_0, \text{SRS})$.

Example 31 *Let* $\mathbf{N} = (6,6,6)$ *and* $n = 4$. *Then the eigenvalues of the matrices* $\mathbf{M}(\hat{T}_{st}, SRS)$ *and* $\mathbf{M}(\hat{T}_D, SRS)$ *are*

$$\lambda_1 = 3.4618 \quad \lambda_2 = 0.9412 \quad \lambda_3 = 1.0294$$

for $\mathbf{M}(\hat{T}_{st}, SRS)$ *and*

$$\alpha_1 = 4.6324 \quad \alpha_2 = 2.3162 \quad \alpha_3 = 0$$

for $\mathbf{M}(\hat{T}_D, SRS)$. *Since* $\alpha_1 > \lambda_1$, $\alpha_2 > \lambda_2$ *and* $\lambda_3 > \alpha_3$ *we can apply theorem 14 part (a) which says if the coefficient of variation of* $\mathbf{y}$ *is*

$$\mathrm{cv}(\mathbf{y}) \leq \sqrt{\frac{(\lambda_3 - \alpha_3)N}{\underset{l=1,2}{max}\{\alpha_l - \lambda_l\}(N-1)}} = 0.8903$$

then $\mathrm{MSE}(\hat{T}_D, SRS) \leq \mathrm{MSE}(\hat{T}_{st}, SRS)$ *and if*

$$\mathrm{cv}(\mathbf{y}) \geq \sqrt{\frac{(\lambda_3 - \alpha_3)N}{\underset{l=1,2}{min}\{\alpha_l - \lambda_l\}(N-1)}} = 0.9311$$

then $\mathrm{MSE}(\hat{T}_{st}, SRS) \leq \mathrm{MSE}(\hat{T}_D, SRS)$.

The maximum relative efficiency of $stg(\hat{T}_D, SRS)$ *over* $stg(\hat{T}_{st}, SRS)$ *is equal to*

$$\mathrm{MRE}(\hat{T}_D, SRS \mid \hat{T}_{st}, SRS) = 2.46938$$

which means that the mean square error of $stg(\hat{T}_D, SRS)$ *can be at most 146.9% larger than that of* $stg(\hat{T}_{st}, SRS)$.

Example 32 *Consider the strategies* $stg(\hat{T}_D, SRS)$ *and* $stg(\hat{T}_0, SRS)$ *with poststrata sizes* $\mathbf{N} = (10, 10)$ *and* $n = 4$. *Then the eigenvalues of* $\mathbf{M}(\hat{T}_D, SRS)$ *are equal to*

$$\alpha_1 = 5.265451, \quad \alpha_2 = 1.733746, \quad \alpha_3 = 0$$

and the eigenvalue of $\mathbf{M}(\hat{T}_0, SRS)$ *is equal to*

$$\mu_1 = 4.21052 \quad \mu_2 = 4.21052, \quad \mu_3 = 0.$$

Since $\alpha_1 > \mu_1$, $\alpha_2 < \mu_2$ *and* $\alpha_3 = \mu_3$, *by theorem 14 part (b), a necessary and sufficient condition for* $\mathrm{MSE}(\hat{T}_D, SRS) \leq \mathrm{MSE}(\hat{T}_0, SRS)$ *is*

$$R_Y \leq \frac{\mu_2 - \alpha_2}{[(\mu_2 - \alpha_2) + (\alpha_1 - \mu_1)]} = 0.701298$$

and a necessary and sufficient condition for $\mathrm{MSE}(\hat{T}_D, SRS) \geq \mathrm{MSE}(\hat{T}_0, SRS)$ *is*

$$R_Y \geq 0.701298.$$

The maximum relative efficiency for these strategies are

$$\mathrm{MRE}(\hat{T}_D, SRS \mid \hat{T}_0, SRS) = 1.25$$

and

$$\mathrm{MRE}(\hat{T}_0, SRS \mid \hat{T}_D, SRS) = 1.60.$$

Another estimator that is similar to $\hat{T}_D$ is given by

$$\hat{T}_{AD} = N \frac{\sum_{h=1}^{H} I_h N_h \bar{y}_h}{\sum_{h=1}^{H} I_h N_h}$$

which is also calibrated for the constant vector. This estimator is equal to $\hat{T}_D$ when $N_h = N_\varsigma$ for all $h = 1, \ldots, H$ and its mean square error matrix will also be the same as that of $\hat{T}_D$. Sugden & Smith (2006) showed that $\hat{T}_{AD}$ can be written as

$$\hat{T}_{AD} = N \frac{\sum_{h=1}^{H} I_h N_h \bar{y}_h}{\sum_{h=1}^{H} I_h N_h} = \sum_{\substack{h=1 \\ n_h > 0}}^{H} N_h \bar{y}_h + \sum_{\substack{h=1 \\ n_h = 0}}^{H} N_h \left(\sum_{\substack{h=1 \\ n_h > 0}}^{H} w_h \bar{y}_h \right),$$

where $w_h = N_h / N(\sum_{h=1}^{H} I_h N_h)$, which represents a form of collapsing the unsampled poststrata. Rao (1985) said that $stg(\hat{T}_{AD}, \mathrm{SRS})$ is inconsistent unconditionally, however this is not necessarily true and we illustrate this by giving a counterexample:

For $\mathbf{N} = (6, 5, 4)$ apply corollary 1 to give the maximum relative efficiency of $stg(\hat{T}_{AD}, \mathrm{SRS})$ using different overall sample sizes and let

$$\lambda(n) = \max \left\{ \frac{\mathrm{MSE}(\hat{T}_{AD}, SRS), \text{ with sample size } n}{\mathrm{MSE}(\hat{T}_{AD}, SRS), \text{ with sample size } n-1} \right\}.$$

The values of $\lambda(n)$ are calculated in Table 5.1 for $n = 2, \ldots, 15$.

Table 5.1: Values of $\lambda(n)$ for $\mathbf{N} = (6, 5, 4)$ using $stg(\hat{T}_{AD}, \text{SRS})$

n	$\lambda(n)$
2	0.56
3	0.72
4	0.80
5	0.84
6	0.86
7	0.85
8	0.83
9	0.80
10	0.76
11	0.72
12	0.68
13	0.61
14	0.46
15	0

Since $\lambda(n) < 1$ for all $n = 2, \ldots, 15$ this implies that $stg(\hat{T}_{AD}, \text{SRS})$ is consistent when $\mathbf{N} = (6, 5, 4)$.

5.2 Weak conditional comparison (SRS)

In situations where the sample size is large enough so that the probability of selecting a sample with at least one empty sample poststratum is negligible, we may consider making our inferences with respect to the weak conditional design (conditioned on the set $\mathcal{S}_w$ which contains all samples with $n_h > 0$ for all $h = 1, \ldots, H$) instead of the unconditional design. If the probability of selecting a sample with $n_h = 0$ for any poststratum is zero then the unconditional design is the same as the weak conditional design. Holt & Smith (1979) used the weak conditional variance of the stratified estimator and the unconditional variance of the expansion estimator when they compared $\hat{T}_{st}$ with $\hat{T}_0$ under simple random sampling at the design stage.

Under the weak conditional simple random sampling design $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_w)$ is calibrated for vectors that are constant within poststrata and $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_w)$ is only calibrated for constant vectors. Hence we can apply theorem 11 to obtain a sufficient condition for $\text{MSE}(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_w)$ to be less than $\text{MSE}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_w)$. We illustrate this in the following example.

Example 33 *Let* $\mathbf{N} = (10, 10)$ *and* $n = 4$. *To compare* $stg(\hat{T}_{st}, SRS \mid \mathcal{S}_w)$ *with* $stg(\hat{T}_0, SRS \mid \mathcal{S}_w)$*, we approximated the values from p.145 as* $\eta_0 = 3.86698$ *and* $\gamma_0 = 2.711908047$*. Hence by theorem 11 a sufficient condition for*

$$\text{MSE}(\hat{T}_{st}, SRS \mid \mathcal{S}_w) \leq \text{MSE}(\hat{T}_0, SRS \mid \mathcal{S}_w)$$

is if

$$R_Y < \frac{\gamma_0}{\eta_0} = \frac{2.711908047}{3.86698} = 0.7012987$$

which means that the poststratification accounts for more than two thirds (70%) of the total variation.

5.3 Conditional comparison (SRS)

In this section we compare the stratified estimator and the expansion estimator under the conditional simple random sampling design. We will only consider fixed sample size configurations $\mathbf{n}$ with $n_h > 0$ for all $h = 1, \ldots, H$. Holt & Smith (1979) pointed out that no one of these strategies is uniformly better than the other and they carried out an empirical investigation comparing $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ with $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ which suggested that when $\text{MSE}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is smaller than $\text{MSE}(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$, it isn't by much, but when $\text{MSE}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is larger than $\text{MSE}(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ it can be by a huge amount. Jagers *et al.* (1985) also had the same conclusion.

Sugden & Smith (2006) stated that a sufficient condition for the mean square error of $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ to be less than that of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is if for those poststrata $h = 1, \ldots, H$ with $n_h > 0$ it is either true that $\hat{N}_h > N_h$ or else, when $\hat{N}_h < N_h$ it is true that

$$\text{cv}(\mathbf{y}_h)^2 < \frac{N_h - \hat{N}_h}{(N_h + \hat{N}_h)(n_h^{-1} - N_h^{-1})} \tag{5.10}$$

where $\hat{N}_h = Nn_h/n$. However this sufficient condition for the superiority of the strategy $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is not necessarily true. This can be demonstrated via a counter example. Consider the population survey vector

$$\mathbf{y}^t = (1, 2, 1, 1, 2, 1, 2, 1, 2, 1, 1, 2)$$

which is partitioned into two groups with $\mathbf{N} = (7, 5)$ where the first seven units of the population belong to poststratum one and the rest belonging to poststratum two. Let the observed sampling configuration be $\mathbf{n} = (2, 2)$. Then we have $\hat{N}_1 = \hat{N}_2 = 6$

so that $N_2 < \hat{N}_2$, and since $\hat{N}_1 < N_1$ we calculated $\text{cv}(\mathbf{y}_1)^2 = 0.14$ which is less than

$$\frac{N_1 - \hat{N}_1}{(1/n_1 - 1/N_1)(N_1 + \hat{N}_1)} = 0.2154$$

so that condition (5.10) holds. But $\text{MSE}(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}}) = 7.2496$ which is greater than $\text{MSE}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}}) = 6.9332$. This shows that Sugden & Smith's condition for the superiority of $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is not sufficient.

We will give a theorem which can be used to give a sufficient condition for $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ to be better than $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ for the case where the number of poststrata is equal to two. But first we examine the mean square error matrices of these strategies. The matrix $\mathbf{M}(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is clearly identical to $\mathbf{M}(\hat{T}_{st}, \text{StRS})$, see example 18 on p.131. For $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ we have $b_{s_h i} = N/n$ and $p^*(s) = 1/\Pi_{h=1}^{H}\binom{N_h}{n_h}$ so that the induced design on the strata are the same as that of a stratified random sampling design with $p_h^*(s_h) = \binom{N_h}{n_h}^{-1}$ and $p_{hg}^*(s_h, s_g) = \binom{N_h}{n_h}^{-1}\binom{N_g}{n_g}^{-1}$. Then

$$\begin{aligned} B_{hi} &= \sum_{s_h \ni i} p_h^*(s) b_{s_h i} = \sum_{s_h \ni i} \binom{N_h}{n_h}^{-1} \frac{N}{n} \\ &= \frac{N}{n} \frac{\binom{N_h - 1}{n_h - 1}}{\binom{N_h}{n_h}} = \frac{N n_h}{n N_h} = \hat{N}_h / N_h \end{aligned}$$

for every unit $i \in \mathcal{U}_h$, $h = 1, \ldots, H$. So the bias of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is equal to

$$\text{bias}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}}) = \sum_{h=1}^{H} (\hat{N}_h / N_h - 1) T_h.$$

Similarly for $i \in \mathcal{U}_h$ we have

$$\sum_{s_h \ni i} p_h^*(s_h) b_{s_h i}^2 = \frac{N^2}{n^2} \frac{n_h}{N_h},$$

for $i \neq j \in \mathcal{U}_h$ we have

$$\begin{aligned} \sum_{s_h \ni i,j} p_h^*(s_h) b_{s_h i} b_{s_h j} &= \frac{N^2}{n^2} \sum_{s_h \ni i,j} \binom{N_h}{n_h}^{-1} \\ &= \frac{N^2}{n^2} \frac{\binom{N_h - 2}{n_h - 2}}{\binom{N_h}{n_h}} = \frac{N^2}{n^2} \frac{n_h(n_h - 1)}{N_h(N_h - 1)} \end{aligned}$$

for $h = 1, \ldots, H$ and for $i \in \mathcal{U}_h$ and $j \in \mathcal{U}_g$ $(h \neq g)$ we have

$$\begin{aligned}\sum_{\substack{s_h \ni i \\ s_g \ni j}} p^*_{hg}(s_h, s_g) b_{s_h i} b_{s_g j} &= \frac{N^2}{n^2} \sum_{\substack{s_h \ni i \\ s_g \ni j}} \binom{N_h}{n_h}^{-1} \binom{N_g}{n_g}^{-1} \\ &= \frac{N^2}{n^2} \binom{N_h - 1}{n_h - 1} \binom{N_g - 1}{n_g - 1} \binom{N_h}{n_h}^{-1} \binom{N_g}{n_g}^{-1} \\ &= \frac{N^2}{n^2} \frac{n_h}{N_h} \frac{n_g}{N_g} = \frac{\hat{N}_h}{N_h} \frac{\hat{N}_g}{N_g}.\end{aligned}$$

Since $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is a separate general linear strategy and uses the conditional simple random sampling design (equivalent to stratified random sampling on $\mathbf{n}$), its between poststrata variance component is equal to zero. Hence the ij^{th} element of the mean square error matrix for $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is equal to

$$\mathbf{M}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})_{ij} = \begin{cases} \frac{N^2 n_h}{n^2 N_h^2}(N_h - n_h) + (\hat{N}_h/N_h - 1)^2 & \text{for} \quad i = j \in \mathcal{U}_h \\ -\frac{N^2 n_h}{n^2 N_h^2}\frac{(N_h - n_h)}{(N_h - 1)} + (\hat{N}_h/N_h - 1)^2 & \text{for} \quad i \neq j \in \mathcal{U}_h \\ (\hat{N}_h/N_h - 1)(\hat{N}_g/N_g - 1) & \text{for} \quad i \in \mathcal{U}_h \text{ and } j \in \mathcal{U}_g. \end{cases}$$

This matrix is of the same form as (4.10) and observe the that vector $(\mathbf{B} - \mathbf{1}_N)$ is orthogonal to variance-covariance matrix $\mathbf{V}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ where, by (3.1),

$$\mathbf{V}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})_{ij} = \begin{cases} \frac{N^2 n_h}{n^2 N_h^2}(N_h - n_h) & \text{for} \quad i = j \in \mathcal{U}_h \\ -\frac{N^2 n_h}{n^2 N_h^2}\frac{(N_h - n_h)}{(N_h - 1)} & \text{for} \quad i \neq j \in \mathcal{U}_h \\ 0 & \text{for} \quad i \in \mathcal{U}_h \text{ and } j \in \mathcal{U}_g, \end{cases}$$

for $h, g = 1, \ldots, H$ $(h \neq g)$. Since the bias square matrix $(\mathbf{B} - \mathbf{1}_N)(\mathbf{B} - \mathbf{1}_N)^t$ of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$, which is of rank one, is orthogonal to $\mathbf{V}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ it follows that the eigenvalues of $\mathbf{M}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ are equal to

$$\lambda_h = \frac{\hat{N}_h^2}{N_h n_h} \frac{(N_h - n_h)}{(N_h - 1)}$$

of multiplicity $N_h - 1$ for $h = 1, \ldots, H$,

$$\lambda = \sum_{h=1}^{H} N_h \left(\frac{\hat{N}_h}{N_h} - 1\right)^2 \tag{5.11}$$

of multiplicity one and zero of multiplicity $H-1$. Then the mean square error of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ can be written as

$$\text{MSE}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}}) = \sum_{h=1}^{H} \lambda_h (N_h - 1) S_{Y_h}^2 + \lambda \mathbf{y}^t \mathbf{e}\mathbf{e}^t \mathbf{y} \tag{5.12}$$

where $\mathbf{e}$ is the $(N \times 1)$ normalized eigenvector corresponding to λ whose i^{th} element is equal to $(\hat{N}_h/N_h - 1)/\sqrt{\lambda}$ for $i \in \mathcal{U}_h$, $h = 1, \ldots, H$. The second term of (5.12) is equal to the bias square of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$.

The expansion estimator under the conditional simple random sampling design is unbiased for T_Y either when T_{Y_h} is constant over all $h = 1, \ldots, H$ and $N_1 = \ldots = N_H$ or if n_h is equal to a proportional allocation, i.e. if $n_h = N_h n/N$ for all $h = 1, \ldots, H$, in which case $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ will be the same as $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$.

Note that because the bias square matrix of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is orthogonal to the variance-covariance matrix $\mathbf{V}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$, we cannot apply theorem 8 to $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ to obtain an exact upper bound on its absolute bias ratio.

Now the following theorem can be used to give a sufficient condition for the strategy $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ to be better than $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ in the case where $H = 2$.

Theorem 15 *Let $stg(\hat{T}_1, p_1)$ and $stg(\hat{T}_2, p_2)$ be two different general linear strategies whose mean square error can be written as*

$$\text{MSE}(\hat{T}_1, p_1) = \sum_{h=1}^{H} \lambda_h (N_h - 1) S_{Y_h}^2 + \lambda \sum_{h=1}^{H} N_h (\bar{Y}_h - \bar{Y})^2$$

and

$$\text{MSE}(\hat{T}_2, p_2) = \sum_{h=1}^{H} \mu_h (N_h - 1) S_{Y_h}^2.$$

Provided $\omega = \max_{h=1,\ldots,H} \{\mu_h - (\lambda_h - \lambda)\}$ is greater than zero, a sufficient condition for $stg(\hat{T}_2, p_2)$ to be better than $stg(\hat{T}_1, p_1)$ is if

$$R_Y = \frac{\sum_{h=1}^{H} (N_h - 1) S_{Y_h}^2}{(N-1) S_Y^2} \leq \frac{\lambda}{\omega}.$$

Proof Observe that

$$R_Y \leq \frac{\lambda}{\omega} \quad \Leftrightarrow \quad \omega \sum_{h=1}^{H} (N_h - 1) S_{Y_h}^2 \leq \lambda (N-1) S_Y^2.$$

But

$$\sum_{h=1}^{H} [\mu_h - (\lambda_h - \lambda)](N_h - 1) S_{Y_h}^2 \leq \omega \sum_{h=1}^{H} (N_h - 1) S_{Y_h}^2$$

which implies that

$$\begin{aligned}
& \sum_{h=1}^{H} \mu_h (N_h - 1) S_{Y_h}^2 - \sum_{h=1}^{H} (\lambda_h - \lambda)(N_h - 1) S_{Y_h}^2 \leq \lambda (N-1) S_Y^2 \\
\Leftrightarrow \quad & \sum_{h=1}^{H} \mu_h (N_h - 1) S_{Y_h}^2 \leq \sum_{h=1}^{H} \lambda_h (N_h - 1) S_{Y_h}^2 + \lambda (N-1) S_Y^2 - \sum_{h=1}^{H} \lambda (N_h - 1) S_{Y_h}^2 \\
\Leftrightarrow \quad & \sum_{h=1}^{H} \mu_h (N_h - 1) S_{Y_h}^2 \leq \sum_{h=1}^{H} \lambda_h (N_h - 1) S_{Y_h}^2 + \lambda \sum_{h=1}^{H} N_h (\bar{Y}_h - \bar{Y})^2 \\
\Leftrightarrow \quad & \mathrm{MSE}(\hat{T}_2, p_2) \leq \mathrm{MSE}(\hat{T}_1, p_1)
\end{aligned}$$

and that completes the proof. □

In order to use theorem 15 to compare $stg(\hat{T}_{st}, \mathrm{SRS} \mid \mathcal{S}_{\mathbf{n}})$ with $stg(\hat{T}_0, \mathrm{SRS} \mid \mathcal{S}_{\mathbf{n}})$ we first show that when $H = 2$ the bias square of $stg(\hat{T}_0, \mathrm{SRS} \mid \mathcal{S}_{\mathbf{n}})$ can be written as

$$\lambda \sum_{h=1}^{2} N_h (\bar{Y}_h - \bar{Y})^2.$$

First observe that for $H = 2$ we have

$$\begin{aligned}
\mathrm{bias}(\hat{T}_0, \mathrm{SRS} \mid \mathcal{S}_{\mathbf{n}}) &= \sum_{h=1}^{2} (\hat{N}_h - N_h) \bar{Y}_h \\
&= (\hat{N}_1 - N_1)\bar{Y}_1 + (\hat{N}_2 - N_2)\bar{Y}_2 \\
&= (\hat{N}_1 - N_1)\bar{Y}_1 + (N - \hat{N}_1 - N + N_1)\bar{Y}_2 \\
&= (\hat{N}_1 - N_1)(\bar{Y}_1 - \bar{Y}_2)
\end{aligned}$$

which implies that the bias square is equal to $(\hat{N}_1 - N_1)^2 (\bar{Y}_1 - \bar{Y}_2)^2$.

Now for $H = 2$ the eigenvalue λ, given in (5.11), can be written as

$$\begin{aligned}\sum_{h=1}^{2} N_h(\hat{N}_h/N_h - 1)^2 &= N_1^{-1}(\hat{N}_1 - N_1)^2 + N_2^{-1}(\hat{N}_2 - N_2)^2 \\ &= (N_1^{-1} + N_2^{-1})(\hat{N}_1 - N_1)^2 \\ &= \frac{N}{N_1 N_2}(\hat{N}_1 - N_1)^2. \end{aligned} \tag{5.13}$$

And since

$$\bar{Y}_1 - \bar{Y} = \bar{Y}_1 - (W_1\bar{Y}_1 + W_2\bar{Y}_2) = W_2(\bar{Y}_1 - \bar{Y}_2)$$

where $W_h = N_h/N$ and similarly

$$\bar{Y}_2 - \bar{Y} = \bar{Y}_2 - (W_1\bar{Y}_1 + W_2\bar{Y}_2) = -W_1(\bar{Y}_1 - \bar{Y}_2)$$

if follows that

$$\begin{aligned}\sum_{h=1}^{2} N_h(\bar{Y}_h - \bar{Y})^2 &= (N_1W_2^2 + N_2W_1^2)(\bar{Y}_1 - \bar{Y}_2)^2 \\ &= \frac{N_1N_2(N_2 + N_1)}{N^2}(\bar{Y}_1 - \bar{Y}_2)^2 \\ &= \frac{N_1N_2}{N}(\bar{Y}_1 - \bar{Y}_2)^2. \end{aligned} \tag{5.14}$$

Hence when we multiply (5.13) with (5.14) we get

$$\lambda \sum_{h=1}^{2} N_h(\bar{Y}_h - \bar{Y})^2 = (\hat{N}_1 - N_1)^2(\bar{Y}_1 - \bar{Y}_2)^2$$

which is equal to the bias square of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$. This implies that the mean square error of the expansion estimator under the conditional simple random sampling design can be written in the form

$$\sum_{h=1}^{H} \lambda_h(N_h - 1)S_h^2 + \lambda \sum_{h=1}^{H} N_h(\bar{Y}_h - \bar{Y})^2$$

when $H = 2$ and so we can apply theorem 15 to compare it with the stratified estimator.

When $H > 2$ it is not true that the bias square of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ can be written in the form

$$\lambda \sum_{h=1}^{H} N_h(\bar{Y}_h - \bar{Y})^2$$

for example let $H = 3$ with $N_1 = N_2 = N_3 = 10$ and let $\bar{Y}_1 = 6$, $\bar{Y}_2 = 15$ and $\bar{Y}_3 = 9$ with $\bar{Y} = 10$. For the sample size configuration $\mathbf{n} = (2, 6, 3)$ we have $\hat{N}_1 = 5.45$, $\hat{N}_2 = 16.36$ and $\hat{N}_3 = 8.18$. Then the bias square of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ is equal to

$$\begin{aligned} \left[\text{bias}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})\right]^2 &= [(5.45 - 10) \times 6 + (16.36 - 10) \times 15 + (8.18 - 10) \times 9]^2 \\ &= 51.72^2 = 2674.9584. \end{aligned}$$

But $\lambda \sum_{h=1}^{3} N_h(\bar{Y}_h - \bar{Y})^2$ is equal to

$$\begin{aligned} & 10^{-1}[(5.45 - 10)^2 + (16.36 - 10)^2 + (8.18 - 10)^2] \\ & \times 10[(6 - 10)^2 + (15 - 10)^2 + (9 - 10)^2] \\ = \; & 2703.6912 \end{aligned}$$

which is not the same as the bias square of $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$. This means that we cannot apply theorem 15 to this strategy.

Example 34 *For $stg(\hat{T}_0, SRS \mid \mathcal{S}_{\mathbf{n}})$ and $stg(\hat{T}_{st}, SRS \mid \mathcal{S}_{\mathbf{n}})$ with $H = 2$, $\mathbf{N} = (7, 13)$ and $\mathbf{n} = (2, 3)$ we have*

$$\text{MSE}(\hat{T}_0, SRS \mid \mathcal{S}_{\mathbf{n}}) = \sum_{h=1}^{2} \lambda_h (N_h - 1) S_{Y_h}^2 + \lambda \sum_{h=1}^{2} N_h(\bar{Y}_h - \bar{Y})^2$$

and

$$\text{MSE}(\hat{T}_{st}, SRS \mid \mathcal{S}_{\mathbf{n}}) = \sum_{h=1}^{2} \mu_h (N_h - 1) S_{Y_h}^2$$

where

$$\lambda_1 = 3.076923, \quad \lambda_2 = 3.809524, \quad \lambda = 2.197802$$

and

$$\mu_1 = 3.611111, \quad \mu_2 = 2.916667.$$

Then we calculate the value of ω *which is equal to*

$$\begin{aligned}\omega &= \max\{3.611111 - (3.076923 - 2.197802), 2.916667 - (3.809524 - 2.197802)\} \\ &= \max\{2.73199, 1.3049446\} = 2.73199\end{aligned}$$

and so a sufficient condition for the superiority of $stg(\hat{T}_{st}, SRS \mid \mathcal{S}_{\mathbf{n}})$ *is if*

$$R_Y \leq \frac{\lambda}{\omega} = \frac{2.197802}{2.73199} = 0.8044692402.$$

Since the nullspace of the mean square error matrix $\mathbf{M}(\hat{T}_0, SRS \mid \mathcal{S}_{\mathbf{n}})$ *is contained in the nullspace of* $\mathbf{M}(\hat{T}_{st}, SRS \mid \mathcal{S}_{\mathbf{n}})$ *we can apply corollary 1 on p.99 to give the maximum relative efficiency of* $stg(\hat{T}_{st}, SRS \mid \mathcal{S}_{\mathbf{n}})$ *over* $stg(\hat{T}_0, SRS \mid \mathcal{S}_{\mathbf{n}})$ *which is equal to*

$$\text{MRE}[(\hat{T}_{st}, SRS \mid \mathcal{S}_{\mathbf{n}}) \mid (\hat{T}_0, SRS \mid \mathcal{S}_{\mathbf{n}})] = 1.173611.$$

Hence the mean square error of the stratified estimator can be at most 17% larger than that of the expansion estimator under the conditional simple random sampling design.

In their empirical study of the stratified and expansion estimator Holt & Smith (1979) looked at 13 different populations of various sizes, calculating the value of R_Y for each population, and simulated the distribution of the ratio

$$K = \frac{\text{MSE}(\hat{\bar{Y}}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})}{\text{MSE}(\hat{\bar{Y}}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})}$$

for each population using every configuration of $\mathbf{n}$ ignoring those samples with $n_h = 0$ for some $h = 1, \ldots, H$. They pointed out the strong relationship between the value of R_Y and the 90, 95 and 99 percentile of the distribution of K for the various cases considered in their paper. For populations with small values of R_Y, those percentile values were exceptionally large compared to the percentile values for those populations with large values of R_Y.

One of these populations which we found to be of interest was case 9 which was the only population out of the 13 that was split into two poststrata. The value of R_Y for this population was equal to 0.5 and out of all the 13 cases studied case 9 was the only population whose value for the 1 percentile, of the distribution of K, that was equal to 1 for all the various sample sizes considered in their studies. This would suggest that most of the times $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ would be better than $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ for case 9. In this particular case ($H = 2$) we could apply theorem 15 to obtain a sufficient condition for the superiority of $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ for every possible configuration of $\mathbf{n}$ such that $n_h \neq 0$ for all $h = 1, \ldots, H$, similarly to example 34.

5.4 Domains of study

In this section we compare the stratified estimator, $\hat{T}_{st,h}$, with the expansion estimator, $\hat{T}_{0,h}$, for the domain total T_{Y_h} where

$$\hat{T}_{st,h} = N_h \bar{y}_h$$

and

$$\hat{T}_{0,h} = \frac{N}{n} n_h \bar{y}_h$$

for some $h = 1, \ldots, H$ under the conditional simple random sampling design. We will assume that the poststratum size is known for all poststrata. The conditional mean square error matrix, on n_h, for $stg(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ and $stg(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ will be equal to the h^{th} diagonal block of $\mathbf{M}(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ and $\mathbf{M}(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ respectively where

$$\mathcal{S}_{n_h} = \{s \in \mathcal{S} : n_h(s) = n_h\}$$

for some poststratum h. Hence, for some $h = 1, \ldots, H$, the ij^{th} element of the mean square error matrix $\mathbf{M}(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ will be equal to

$$\mathbf{M}(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})_{ij} = \begin{cases} \frac{N_h - n_h}{n_h} & \text{for} \quad i = j \in \mathcal{U}_h \\ -\frac{(N_h - n_h)}{n_h(N_h - 1)} & \text{for} \quad i \neq j \in \mathcal{U}_h \end{cases}$$

and the ij^{th} element of $\mathbf{M}(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ is equal to

$$\mathbf{M}(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})_{ij} = \begin{cases} \frac{\hat{N}_h^2(N_h - n_h)}{N_h^2 n_h} + (\hat{N}_h/N_h - 1)^2 & \text{for} \quad i = j \in \mathcal{U}_h \\ -\frac{\hat{N}_h^2(N_h - n_h)}{N_h^2 n_h(N_h - 1)} + (\hat{N}_h/N_h - 1)^2 & \text{for} \quad i \neq j \in \mathcal{U}_h. \end{cases}$$

These $N_h \times N_h$ matrices are of the same form as (4.1) on p.124 and it follows from lemma 12 that the eigenvalues of $\mathbf{M}(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ are equal to

$$\lambda_1 = \frac{N_h(N_h - n_h)}{n_h(N_h - 1)}$$

of multiplicity $N_h - 1$, and zero of multiplicity one, so that

$$\text{MSE}(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h}) = \lambda_1(N_h - 1)S_{Y_h}^2.$$

Similarly, by lemma 12, the eigenvalues of $\mathbf{M}(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ are equal to

$$\mu_1 = \frac{\hat{N}_h^2(N_h - n_h)}{N_h n_h(N_h - 1)}$$

of multiplicity $N_h - 1$ and

$$\mu_2 = N_h(\hat{N}_h/N_h - 1)^2$$

of multiplicity one so that

$$\text{MSE}(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h}) = \mu_1(N_h - 1)S_{Y_h}^2 + \mu_2 N_h \bar{Y}_h^2.$$

If $\lambda_1 < \mu_1$, which will be true when $N_h^2 < \hat{N}_h^2$, then $\text{MSE}(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ will be less than $\text{MSE}(\hat{T}_{0,h}, \text{SRS} \mid n_h)$. This was also pointed out by Sugden & Smith (2006). But if $\lambda_1 > \mu_1$, in which case $N_h > \hat{N}_h$, then it is easily seen that a necessary and sufficient condition for $stg(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ to be better than $stg(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$, provided $\bar{Y}_h \neq 0$, is if

$$\text{cv}(\mathbf{y}_h)^2 = \frac{S^2_{Y_h}}{\bar{Y}_h^2} < \frac{\mu_2 N_h}{(\lambda_1 - \mu_1)(N_h - 1)}. \tag{5.15}$$

Furthermore the right hand side of (5.15) is equal to

$$\begin{aligned} \frac{N_h^2(\hat{N}_h/N_h - 1)^2}{(N_h - \hat{N}_h^2/N_h)(N_h - n_h)/n_h} &= \frac{(\hat{N}_h - N_h)^2}{(N_h^2 - \hat{N}_h^2)(n_h^{-1} - N_h^{-1})} \\ &= \frac{N_h - \hat{N}_h}{(N_h + \hat{N}_h)(n_h^{-1} - N_h^{-1})} \end{aligned} \tag{5.16}$$

which was the upper bound on $\text{cv}(\mathbf{y}_h)^2$ given by Sugden & Smith (2006) as a sufficient condition for the mean square error of $stg(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ to be less than that of $stg(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$. Since $stg(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ is calibrated for the constant $(N_h \times 1)$ vector and $\text{MSE}(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ is of full rank, we can apply corollary 2 part (a) to these strategies to obtain the same condition as (5.16) for the superiority of $stg(\hat{T}_{st,h}, \text{SRS} \mid \mathcal{S}_{n_h})$.

Särndal & Hidiroglou (1989) also studied the expansion estimator for the domain total T_{Y_h} under the conditional simple random sampling design. They pointed out that the variance of $stg(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$

$$\begin{aligned} \text{Var}(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h}) &= \hat{N}_h^2(n_h^{-1} - N_h^{-1})S^2_{Y_h} \\ &= \left(\frac{N}{n}\right)^2 \left[\frac{n_h}{N_h}(N_h - n_h)\right] S^2_{Y_h}, \end{aligned} \tag{5.17}$$

is an increasing function over the practical range for n_h with $0 \leq f_h \leq 0.5$, where $f_h = n_h/N_h$ is the sampling fraction for (post)stratum h. To see this we take the second derivative of (5.17) with respect to n_h which gives a quadratic function with negative leading coefficient

$$-\frac{2}{N_h}\left(\frac{N}{n}\right)^2 S^2_{Y_h}$$

and implies that $\text{Var}(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ is a concave (down) function in n_h. Hence differentiating (5.17) with respect to n_h and solving gives

$$f_h = \frac{n_h}{N_h} = \frac{1}{2}$$

which shows that the variance increases as n_h increases from 0 to $N_h/2$ at which point (5.17) is at its maximum. For this reason, and the fact that $stg(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$ is conditionally biased, Hidiroglou & Särndal deemed this strategy to be unsuitable for the conditional approach. Also note that, for the same reason as $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$, we can not apply theorem 8 to obtain an exact upper bound on the absolute bias ratio of $stg(\hat{T}_{0,h}, \text{SRS} \mid \mathcal{S}_{n_h})$. So we cannot say how large the absolute bias ratio can be.

A more general form of $\hat{T}_{0,h}$ is the Hájek-Basu estimator for the domain total T_{Y_h},

$$\hat{T}_{HB,h} = N \frac{\sum_{i \in s_h} y_i/\pi_i}{\sum_{g=1}^{H} I_g \sum_{i \in s_g} 1/\pi_i},$$

which reduces to $\hat{T}_{0,h}$ when π_i is constant for all $i \in \mathcal{U}$. And a more general form of $\hat{T}_{st,h}$ is the separate Hájek-Basu estimator for T_{Y_h}, also see Särndal *et al.* (1992, p.185),

$$\hat{T}_{HBs,h} = N_h \frac{\sum_{i \in s_h} y_i/\pi_i}{\sum_{i \in s_h} 1/\pi_i}$$

which reduces to $\hat{T}_{st,h}$ when π_i is constant for all $i \in \mathcal{U}_h$.

Suppose for some order independent design p is such that $\pi_i = a_h$, say, for all $i \in \mathcal{U}_h$ for some $h = 1, \ldots, H$. The Hájek-Basu estimator for the domain total of poststratum h is then equal to

$$\hat{T}_{HB,h} = N \frac{\sum_{i \in s_h} y_i/\pi_i}{\sum_{g=1}^{H} I_g \sum_{i \in s_g} 1/\pi_i} = N \frac{\frac{1}{a_h} \sum_{i \in s_h} y_i}{\sum_{g=1}^{H} I_g \frac{n_g}{a_g}} = N \frac{\frac{n_h}{a_h} \bar{y}_h}{\sum_{g=1}^{H} I_g \frac{n_g}{a_g}}$$

and the separate Hájek-Basu estimator for the domain total of poststratum h is equal to

$$\hat{T}_{HBs,h} = N_h \frac{\sum_{i \in s_h} y_i/\pi_i}{\sum_{i \in s_h} 1/\pi_i} = N_h \frac{\frac{1}{a_h} \sum_{i \in s_h} y_i}{n_h/a_h} = N_h \bar{y}_h = \hat{T}_{st,h}.$$

The conditional variance of $\hat{T}_{HB,h}$ is

$$\begin{aligned}\text{Var}(\hat{T}_{HB,h}, p \mid \mathcal{S}_{n_h}) &= \left(\frac{N}{\sum_{g=1}^{H} I_g \frac{n_g}{a_g}}\right)^2 \text{Var}\left(\frac{n_h}{a_h}\bar{y}_h, p \mid \mathcal{S}_{n_h}\right) \\ &= \left(\frac{N}{\sum_{g=1}^{H} I_g \frac{n_g}{a_g}}\right)^2 \left(\frac{n_h}{a_h}\right)^2 \frac{(N_h - n_h)}{N_h n_h} S_{Y_h}^2\end{aligned}$$

and its conditional bias is equal to

$$\text{bias}(\hat{T}_{HB,h}, p \mid \mathcal{S}_{n_h}) = \left(\frac{N}{(\sum_{g=1}^{H} I_g \frac{n_g}{a_g})} \frac{n_h}{a_h} - N_h\right) \bar{Y}_h.$$

Hence, its mean square error is equal to

$$\text{MSE}(\hat{T}_{HB,h}, p \mid \mathcal{S}_{n_h}) = \frac{\tilde{N}_h^2 (N_h - n_h)}{N_h n_h} S_{Y_h}^2 + (\tilde{N}_h - N_h)^2 \bar{Y}_h^2$$

where $\tilde{N}_h = N n_h / [(\sum_{g=1}^{H} I_g n_g / a_g) a_h]$. Now, $\hat{T}_{HBs,h}$ will be unbiased under the conditional design and its conditional mean square error is given by

$$\text{MSE}(\hat{T}_{HBs,h}, p \mid \mathcal{S}_{n_h}) = \frac{N_h (N_h - n_h)}{n_h} S_{Y_h}^2.$$

Conditional on $\mathcal{S}_{n_h}$ we see that for $S_{Y_h} \neq 0$ we have

$$\begin{aligned}\frac{\text{MSE}(\hat{T}_{HB,h}, p \mid \mathcal{S}_{n_h})}{\text{MSE}(\hat{T}_{HBs,h}, p \mid \mathcal{S}_{n_h})} &= \frac{\frac{\tilde{N}_h^2 (N_h - n_h)}{N_h n_h} S_{Y_h}^2 + (\tilde{N}_h / N_h - 1)^2 T_h^2}{\frac{N_h (N_h - n_h)}{n_h} S_{Y_h}^2} \\ &= \frac{\left(\frac{\tilde{N}_h}{N_h}\right)^2 (n_h^{-1} - N_h^{-1}) \text{cv}(\mathbf{y}_h)^2 + (\tilde{N}_h / N_h - 1)^2}{(n_h^{-1} - N_h^{-1}) \text{cv}(\mathbf{y}_h)^2}.\end{aligned}$$

It is easy to see that this ratio is always greater than one when $N_h < \tilde{N}_h$, and when $N_h > \tilde{N}_h$ we can apply the following theorem.

Theorem 16 *Under a design p that is order independent and applied to a size variable that is constant within poststratum h for some $h = 1, \ldots, H$. Provided $\tilde{N}_h < N_h$ and $\bar{Y}_h \neq 0$ a sufficient condition for the mean square error of $stg(\hat{T}_{HBs,h}, p \mid \mathcal{S}_{n_h})$ to be less than that of $stg(\hat{T}_{HB,h}, p \mid \mathcal{S}_{n_h})$ is*

$$\text{cv}(\mathbf{y}_h)^2 \leq \frac{N_h - \tilde{N}_h}{(N_h + \tilde{N}_h)(n_h^{-1} - N_h^{-1})}.$$

Proof If this condition holds then

$$\begin{aligned} & (N_h^2 - \tilde{N}_h^2)(n_h^{-1} - N_h^{-1})\mathrm{cv}(\mathbf{y}_h)^2 \le (N_h - \tilde{N}_h)^2 \\ \Leftrightarrow \quad & N_h^2(n_h^{-1} - N_h^{-1})\mathrm{cv}(\mathbf{y}_h)^2 \le \tilde{N}_h^2(n_h^{-1} - N_h^{-1})\mathrm{cv}(\mathbf{y}_h)^2 + (N_h - \tilde{N}_h)^2 \\ \Leftrightarrow \quad & (n_h^{-1} - N_h^{-1})S_{Y_h}^2 \le \left(\frac{\tilde{N}_h}{N_h}\right)^2 (n_h^{-1} - N_h^{-1})S_{Y_h}^2 + (\tilde{N}_h/N_h - 1)^2\bar{Y}_h^2. \end{aligned}$$

Hence

$$\mathrm{MSE}(\hat{T}_{HBs,h}, p \mid \mathcal{S}_{n_h}) \le \mathrm{MSE}(\hat{T}_{HB,h}, p \mid \mathcal{S}_{n_h})$$

and that completes the proof. □

This result reduces to Sugden and Smith's result, given by (5.16), when the design is simple random sampling with $\pi_i = n/N$ for all $i \in \mathcal{U}$.

5.5 General designs

In this section we consider poststratification under a general unequal probability design. We will apply corollary 2 and theorem 11, from chapter 4, to give sufficient conditions for the superiority of one strategy over another and analyse strategies unconditionally, weak conditionally and conditionally.

The Horvitz-Thompson estimator will perform well whenever the linear relationship between the y's and the π's is strong. If under some poststratification the relationship between the y's and the π's is stronger within the poststrata then an estimator that is calibrated for the $\boldsymbol{\pi}_h$'s will be more efficient. However if after observing the sampled units we found that the assumption of the y's being strongly related to the π's is not true and rather there was a strong relationship according to some poststratification variable, i.e. the $\mathbf{y}_h$'s are strongly related to the $\mathbf{1}_h$'s for some poststratification, then there are a number of Hájek-Basu type estimators we could consider. In the following subsection we give details of the Hájek-Basu type estimators that we will consider and compare.

5.5.1 Description of estimators

The first estimator we will consider is the Hájek-Basu estimator

$$\hat{T}_{HB} = N\frac{\sum_{i\in s} y_i/\pi_i}{\sum_{i\in s} 1/\pi_i}.$$

This estimator is calibrated for constant vectors with respect to the support given by any without replacement sample design. When the design probabilities are from a simple random sample, in which case $\pi_i = n/N$ for all $i \in \mathcal{U}$, this estimator reduces to the expansion estimator. Thompson (2002, p.56) and Särndal *et al.* (1992, p.183) say that the Hájek-Basu estimator should be considered instead of the Horvitz-Thompson estimator if the y's are not linearly related to the π's because then it should have a smaller variance. This will be true provided the variance of the y's is small.

Suppose we realized that the population can be partitioned by some categorical variable after observing the sampled units, then we can consider the separate Hájek-Basu estimator for the total T_Y,

$$\hat{T}_{HBs} = \sum_{h=1}^{H} I_h N_h \frac{\sum_{i\in s_h} y_i/\pi_i}{\sum_{i\in s_h} 1/\pi_i}.$$

This estimator is calibrated for vectors that are constant within strata with respect to $\mathcal{S}_w$. This estimator will reduce to the stratified estimator when the design probabilities are from a simple random sample.

Another estimator we will consider is the Doss type separate Hájek-Basu estimator, denoted by $\hat{T}_{HBsD}$ where

$$\hat{T}_{HBsD} = \left(\frac{N}{\sum_{h=1}^{H} I_h N_h}\right) \sum_{h=1}^{H} I_h N_h \frac{\sum_{i\in s_h} y_i/\pi_i}{\sum_{i\in s_h} 1/\pi_i},$$

which is always calibrated for the constant population vectors. Note that this estimator is a form of a collapsed estimator since it can be written as

$$\hat{T}_{HBsD} = \sum_{\substack{h=1 \\ n_h>0}}^{H} \hat{T}_{HBs,h} + \sum_{\substack{h=1 \\ n_h=0}}^{H} N_h(\hat{T}_{HBsD}/N)$$

and it reduces to $\hat{T}_{AD}$, from p.191, when the design probabilities are from a simple random sample.

There is no simple expression for the mean square error for these estimators under an unequal probability design but since they are general linear estimators we can write their exact mean square error as quadratic forms in $\mathbf{y}$.

5.5.2 Conditional and unconditional analysis

It has been strongly argued by Holt & Smith (1979) that inferences should be made with respect to the conditional design, by conditioning on $\mathbf{n}$, rather than the unconditional design which is over all possible samples of fixed size n. Although the confidence interval given by the unconditional variance gives us the correct coverage probability over every possible sample of fixed size n, it does not give the correct coverage over those samples with fixed allocation $\mathbf{n}$. They illustrated this point via an example with a population of two poststrata using the stratified estimator under simple random sampling by calculating the confidence intervals using the conditional and unconditional variances, and comparing the coverage of each interval with respect to those samples with the same sampling allocation. They used an extreme sampling allocation in their example with $\mathbf{n} = (1, 19)$ and said that for the 95% confidence interval based on the unconditional variance, i.e.

$$\hat{\hat{Y}}_{st} \pm 1.96\sqrt{\text{Var}(\hat{\hat{Y}}_{st}, \text{SRS})}, \tag{5.18}$$

about one third of all samples with this sampling configuration did not produce a 95% confidence interval that contained the true value of the mean $\bar{Y}$. This is equivalent to saying that the conditional coverage probability based on the 95% confidence interval (5.18), which uses the unconditional variance, is close to $(1 - 1/3) = 66.67\%$ which is a great undercoverage. i.e.

$$\mathrm{CCP}_1 = \sum_{s \in \mathcal{T}_1} p^*(s) = \sum_{s \in \mathcal{T}_1} \left(\prod_{h=1}^{H} \binom{N_h}{n_h} \right)^{-1} \approx (1 - \frac{1}{3}) = 0.6667$$

where CCP_1 denotes the conditional coverage probability which is calculated by summing over the conditional selection probabilities of all samples in the set

$$\mathcal{T}_1 = \left\{ s \in \mathcal{S}_{\mathbf{n}} : \bar{Y} \in \left(\hat{\bar{Y}}_{st} - 1.96\sqrt{\mathrm{Var}(\hat{\bar{Y}}_{st}, \mathrm{SRS})}, \hat{\bar{Y}}_{st} + 1.96\sqrt{\mathrm{Var}(\hat{\bar{Y}}_{st}, \mathrm{SRS})} \right) \right\}.$$

But the conditional coverage probability based on the 95% confidence interval which uses the conditional variance,

$$\hat{\bar{Y}}_{st} \pm 1.96\sqrt{\mathrm{Var}(\hat{\bar{Y}}_{st}, \mathrm{SRS} \mid \mathcal{S}_{\mathbf{n}})},$$

is more accurate to the correct 0.95 coverage provided the distribution of $\hat{\bar{Y}}_{st}$ over all samples in $\mathcal{S}_{\mathbf{n}}$ is close to a normal distribution. i.e.

$$\mathrm{CCP}_2 = \sum_{s \in \mathcal{T}_2} p^*(s) = \sum_{s \in \mathcal{T}_2} \left(\prod_{h=1}^{H} \binom{N_h}{n_h} \right)^{-1} \approx 0.95$$

where CCP_2 denotes the conditional coverage probability calculated by summing over the conditional selection probabilities of all samples in the set $\mathcal{T}_2$ which is equal to

$$\left\{ s \in \mathcal{S}_{\mathbf{n}} : \bar{Y} \in \left(\hat{\bar{Y}}_{st} - 1.96\sqrt{\mathrm{Var}(\hat{\bar{Y}}_{st}, \mathrm{SRS} \mid \mathcal{S}_{\mathbf{n}})}, \hat{\bar{Y}}_{st} + 1.96\sqrt{\mathrm{Var}(\hat{\bar{Y}}_{st}, \mathrm{SRS} \mid \mathcal{S}_{\mathbf{n}})} \right) \right\}.$$

The conditional coverage probability, CCP_1, based on (5.18) is not theoretically correct but it can serve as a means to give us some degree of comparison between the conditional and unconditional approach.

Let us redefine the sets $\mathcal{T}_1$ and $\mathcal{T}_2$ for any general strategy not just the stratified estimator under simple random sampling. i.e. for some strategy $stg(\hat{T}_s, p)$ let

$$\mathcal{T}_1 = \left\{ s \in \mathcal{S}_{\mathbf{n}} : T_Y \in \left(\hat{T}_s - 1.96\sqrt{\mathrm{Var}(\hat{T}_s, p)}, \hat{T}_s + 1.96\sqrt{\mathrm{Var}(\hat{T}_s, p)} \right) \right\}$$

and

$$\mathcal{T}_2 = \left\{ s \in \mathcal{S}_{\mathbf{n}} : T_Y \in \left(\hat{T}_s - 1.96\sqrt{\operatorname{Var}(\hat{T}_s, p \mid \mathcal{S}_{\mathbf{n}})}, \hat{T}_s + 1.96\sqrt{\operatorname{Var}(\hat{T}_s, p \mid \mathcal{S}_{\mathbf{n}})} \right) \right\}.$$

As an analogy of Holt & Smith, in the following example we will make conditional and unconditional comparisons of the separate Hájek-Basu estimator for T_Y under Tillé's procedure for unequal probability sampling. Provided that the distribution of

$$Z = \frac{\hat{T}_{HBs} - T_Y}{\operatorname{Var}(\hat{T}_{HBs}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})}$$

is close to that of $\mathrm{N}(0,1)$ we would expect the value of CCP_2 to be close to 95% for each allocation of $\mathbf{n}$. But for the confidence intervals based on the unconditional variance we would expect the value of CCP_1 to be less accurate, at least for those samples with extreme allocations.

Example 35 *Consider a population of size 26 which is poststratified into two groups where units* 1, 2, ..., 13 *belong to poststratum* 1 *and the rest belong to poststratum* 2. *For each unit in* $\mathcal{U}$ *we independently and randomly generated the value of the auxiliary or size variable* x_i *from* $\mathrm{N}(100, 10)$. *For each unit in* $\mathcal{U}_1$ *and each unit in* $\mathcal{U}_2$ *we independently and randomly generated the value of the survey variable* y_i *from* $\mathrm{N}(50, 5)$ *and* $\mathrm{N}(80, 8)$ *respectively.*

For a sample of size 10 *we calculated the unconditional variance of the separate Hájek-Basu estimator under the Tillé procedure, which is equal to*

$$\operatorname{Var}(\hat{T}_{HBs}, \textit{Tillé}) = 2644.135$$

and the absolute bias ratio of this strategy is equal to

$$\operatorname{ABR}(\hat{T}_{HBs}, \textit{Tillé}) = 0.00359815.$$

Table 5.2 gives the conditional variances and absolute conditional bias ratios for every possible sample size configuration and table 5.3 gives the 95% *conditional coverage probabilities,* CCP_1 *and* CCP_2*, based on the normal distribution.*

Table 5.2: Conditional variances and absolute conditional bias ratios for example 35

$\mathbf{n}$	$\text{Var}(\hat{T}_{HBs}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$	$\text{ABR}(\hat{T}_{HBs}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$
$(10, 0)$	264.908	39.92039
$(9, 1)$	5494.035	0.07645
$(8, 2)$	2883.461	0.07037
$(7, 3)$	2183.567	0.05103
$(6, 4)$	2033.067	0.02497
$(5, 5)$	2201.460	0.00260
$(4, 6)$	2698.799	0.02772
$(3, 7)$	3732.918	0.04812
$(2, 8)$	6059.509	0.06312
$(1, 9)$	13963.180	0.07507
$(0, 10)$	115.775	97.73143

Table 5.3: Conditional coverage probabilities CCP_1 and CCP_2 for example 35

$\mathbf{n}$	CCP_1	CCP_2
$(10, 0)$	0.0000	0.0000
$(9, 1)$	0.8271	0.8956
$(8, 2)$	0.9527	0.9613
$(7, 3)$	0.9724	0.9545
$(6, 4)$	0.9769	0.9517
$(5, 5)$	0.9704	0.9514
$(4, 6)$	0.9491	0.9516
$(3, 7)$	0.8976	0.9506
$(2, 8)$	0.7723	0.9595
$(1, 9)$	0.4917	0.9997
$(0, 10)$	0.0000	0.0000

We see that the conditional coverage probabilities CCP_2 are in general closer to the correct 95% coverage than the values given by CCP_1, especially for the extreme allocations $(9,1)$, $(2,8)$ and $(1,9)$. The absolute conditional bias ratios are all less than 10% apart for those samples whose sampling allocation is $n_h = 0$ for some poststrata.

Figure 5.3 gives a histogram of

$$Z = \frac{\hat{T}_{HBs} - T_Y}{\sqrt{\text{Var}(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_{\mathbf{n}})}}$$

for $\mathbf{n} = (5,5)$ and figure 5.4 gives a quantile-quantile normal plot of Z. We see that the distribution of Z is close to a bell shape and the q-q plot is slightly 's' shaped. The Kolmogorov-Smirnov test statistic for normality of Z is equal to $D = 0.0122$ and the p-value is less than 0.001. So under the Kolmogorov-Smirnov test for normality, there is overwhelming evidence against the hypothesis that Z is normally distributed with mean zero and standard deviation 1. But theoretically the central limit theorem does not apply here because in general the terms $b_{si}y_i$ of the distinct units in the sample are not independent (results on the asymptotic normality of the Horvitz-Thompson estimator that can be extended to other estimators under unequal probability sampling are given in Rosén, 1972a & b, but these results do not apply here as our population size is too small). However the purpose of this example is to show that the conditional coverage is more accurate by using the conditional variance for confidence intervals which has been demonstrated here.

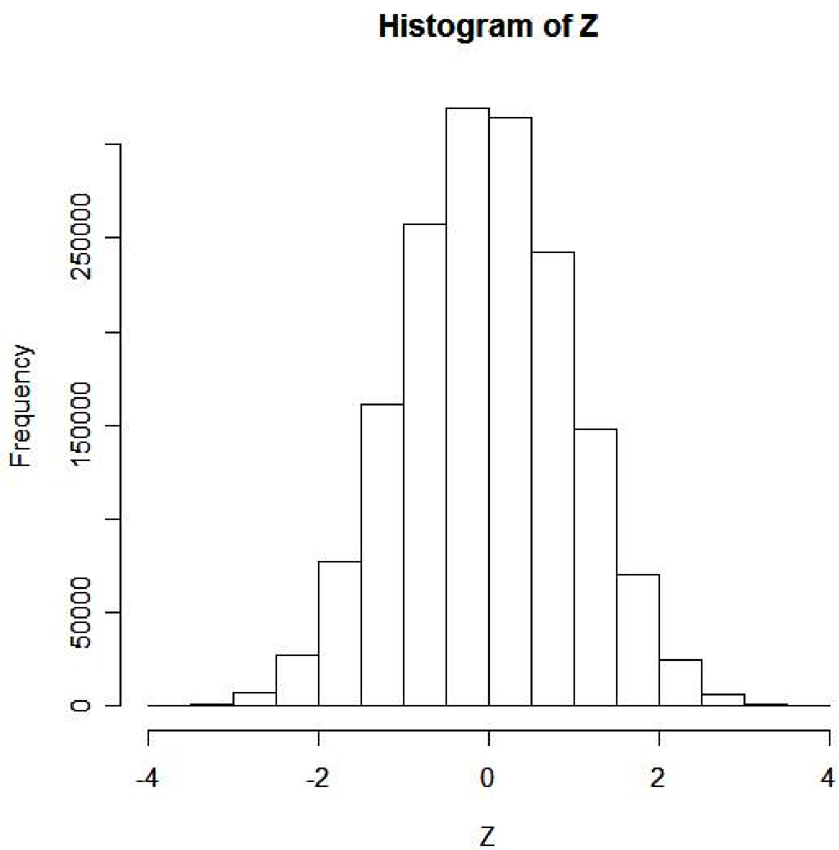

Figure 5.3: Histogram of Z for example 35

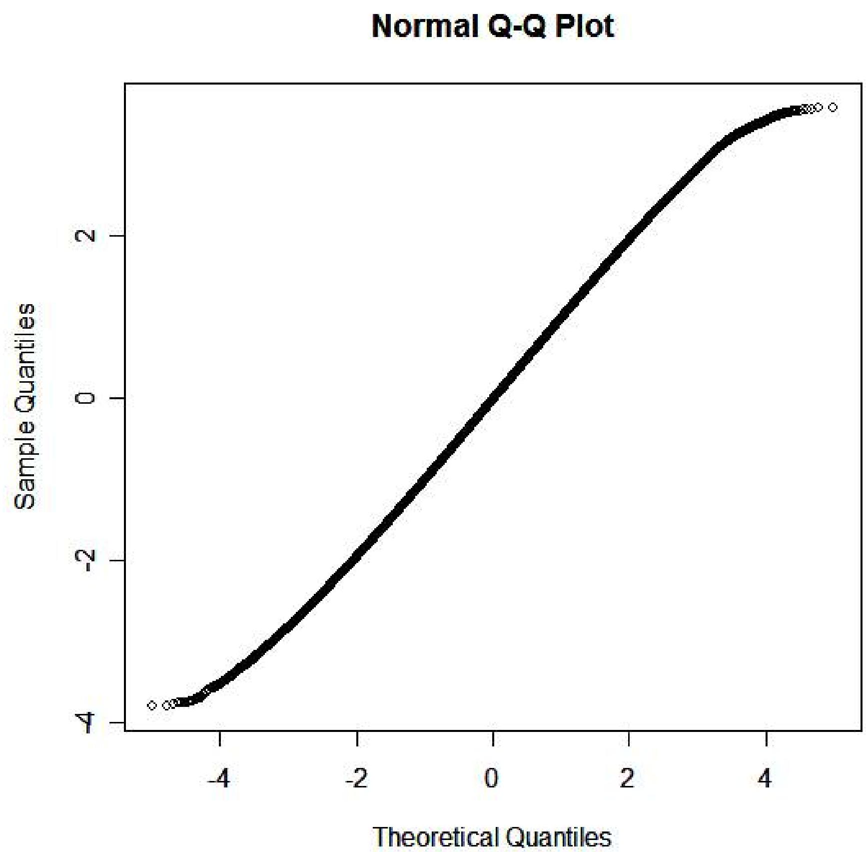

Figure 5.4: q-q normal plot of Z for example 35

5.5.3 Unconditional comparison

Since the estimators $\hat{T}_{HB}$ and $\hat{T}_{HBsD}$ are calibrated for constant vectors with respect to a sampling support given by any without replacement design and $\hat{T}_{HBs}$ is not calibrated for any vectors in $\mathcal{R}^N$ unconditionally, we can apply corollary 2 to compare $\hat{T}_{HB}$ and $\hat{T}_{HBsD}$ with $\hat{T}_{HBs}$ under the unconditional design.

Example 36 *Consider a population of size 26 which is partitioned into two groups and the auxiliary vector as in example 35. For a sample of size 6 we will use corollary 2 to compare* $stg(\hat{T}_{HB}, Tillé)$ *with* $stg(\hat{T}_{HBs}, Tillé)$*. The approximated value of* α_0 *that maximizes* $r_1(\alpha)$*, from p.136, is equal to 3.641557 with* $\gamma_1(\alpha_0) = 0.0945352498$*. Hence by corollary 2 part (a) a sufficient condition for*

$$\mathrm{MSE}(\hat{T}_{HB}, Tillé) \leq \mathrm{MSE}(\hat{T}_{HBs}, Tillé)$$

is if

$$\mathrm{cv}(\mathbf{y}) \leq \sqrt{\frac{r_1(\alpha_0)N}{(N-1)}} = \sqrt{\frac{\gamma_1(\alpha_0)N}{(\alpha_0 - \gamma_1(\alpha_0))(N-1)}} = 0.1664875.$$

Part (b) of corollary 2 could not be applied here to give a sufficient condition for the superiority of $stg(\hat{T}_{HBs}, Tillé)$ *over* $stg(\hat{T}_{HB}, Tillé)$ *as there does not exist a* β *such that* $\beta > \gamma_2(\beta) > 0$*.*

Since $\mathbf{M}(\hat{T}_{HBs}, Tillé)$ *is of full rank we can calculate the maximum relative efficiency of* $stg(\hat{T}_{HB}, Tillé)$ *over* $stg(\hat{T}_{HBs}, Tillé)$*. This is equal to*

$$\mathrm{MRE}(\hat{T}_{HB}, Tillé \mid \hat{T}_{HBs}, Tillé) = 37.12002$$

which is an extremely large value.

The variance-covariance and bias square matrices of these strategies do satisfy the condition of theorem 8, i.e. the nullspace of the variance-covariance matrix is *contained in the nullspace of the bias square matrix. Then by theorem 8 we have the following exact upper bounds on the absolute bias ratio of these strategies,*

$$\text{ABR}(\hat{T}_{HB}, \textit{Tillé}) \leq 0.04580392996$$

and

$$\text{ABR}(\hat{T}_{HBs}, \textit{Tillé}) \leq 0.1703467053.$$

This suggests that the Hájek-Basu estimator under the Tillé scheme is more robust against biases for this population.

Example 37 *Using the same population as in example 36 but now we compare* $\hat{T}_{HBsD}$ *with* $\hat{T}_{HBs}$ *unconditionally. Here we approximated the values* $\alpha_0 = 0.3967376$ *and* $\gamma_1(\alpha_0) = 0.09089722127$*. So by corollary 2 part (a) a sufficient condition for*

$$\text{MSE}(\hat{T}_{HBsD}, \textit{Tillé}) \leq \text{MSE}(\hat{T}_{HBs}, \textit{Tillé})$$

is if

$$\text{cv}(\mathbf{y}) \leq \sqrt{\frac{\gamma_1(\alpha_0)N}{(\alpha_0 - \gamma_1(\alpha_0))(N-1)}} = 0.5559613.$$

We also approximated the values $\beta_0 = 0.1152843$ *and* $\gamma_2(\beta_0) = 0.09839015076$*. So by corollary 2 part (b) a sufficient condition for*

$$\text{MSE}(\hat{T}_{HBs}, \textit{Tillé}) \leq \text{MSE}(\hat{T}_{HBsD}, \textit{Tillé})$$

is if

$$\text{cv}(\mathbf{y}) \geq \sqrt{\frac{\gamma_2(\beta_0)N}{(\beta_0 - \gamma_2(\beta_0))(N-1)}} = 2.461072.$$

The maximum relative efficiency of stg($\hat{T}_{HBsD}$, *Tillé*) *over stg*($\hat{T}_{HBs}$, *Tillé*) *is equal to*

$$\text{MRE}(\hat{T}_{HBsD}, \textit{Tillé} \mid \hat{T}_{HBs}, \textit{Tillé}) = 4.039217$$

which is still large. The upper bounds on the absolute bias ratio of these strategies are

$$\text{ABR}(\hat{T}_{HBsD}, \textit{Tillé}) \leq 0.1160474041$$

and

$$\text{ABR}(\hat{T}_{HBs}, \textit{Tillé}) \leq 0.1703467053.$$

Example 38 *Since stg($\hat{T}_{HB}$, Tillé) and stg($\hat{T}_{HBsD}$, Tillé) are equally calibrated we can use corollary 1, on p.99, to calculate the maximum relative efficiencies. We have*

$$\text{MRE}(\hat{T}_{HB}, \textit{Tillé} \mid \hat{T}_{HBsD}, \textit{Tillé}) = 9.198286$$

and

$$\text{MRE}(\hat{T}_{HBsD}, \textit{Tillé} \mid \hat{T}_{HB}, \textit{Tillé}) = 1.53038$$

which tells us that the mean square error of stg($\hat{T}_{HB}$, Tillé) can be at most 819% larger than that of stg($\hat{T}_{HBsD}$, Tillé), but the mean square error of stg($\hat{T}_{HBsD}$, Tillé) can be at most 53% larger than that of stg($\hat{T}_{HB}$, Tillé).

5.5.4 Weak conditional comparison

If the sample size was large enough so that the probability of selecting a sample such that $n_h = 0$ for any poststrata was negligible, then at the design stage we can consider comparing $\hat{T}_{HB}$ with $\hat{T}_{HBs}$ under the weak conditional design. In this case $\hat{T}_{HBs}$ will be calibrated for constant within strata vectors and so we can apply theorem 11 to give a sufficient condition for $\hat{T}_{HBs}$ to be better than $\hat{T}_{HB}$.

Example 39 *For the same population data from previous examples we will compare stg($\hat{T}_{HB}$, Tillé $\mid \mathcal{S}_w$) with stg($\hat{T}_{HBs}$, Tillé $\mid \mathcal{S}_w$) weak conditionally. The values from* *p.145 are approximately equal to* $\eta_0 = 4.395499$ *and* $\gamma_0 = 2.72740317$ *so by theorem 11 a sufficient condition for*

$$\text{MSE}(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_w) \leq \text{MSE}(\hat{T}_{HB}, \textit{Tillé} \mid \mathcal{S}_w)$$

is if

$$R_Y \leq \frac{\gamma_0}{\eta_0} = \frac{2.72740314}{4.395499} = 0.6204991$$

which means that the poststratification accounts for 62% of the total variation. The nullspace of the matrix $\mathbf{M}(\hat{T}_{HB}, \textit{Tillé} \mid \mathcal{S}_w)$ *is contained in the nullspace of the matrix* $\mathbf{M}(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_w)$ *so we can calculate the maximum relative efficiency as*

$$\text{MRE}[(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_w) \mid (\hat{T}_{HB}, \textit{Tillé} \mid \mathcal{S}_w)] = 1.534843.$$

Hence we know that the mean square error of stg$(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_w)$ *can be almost 53.5% larger than that of stg*$(\hat{T}_{HB}, \textit{Tillé} \mid \mathcal{S}_w)$*.*

The exact upper bound on the absolute bias ratio, given by theorem 8, of these strategies are equal to

$$\text{ABR}(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_w) \leq 0.04710626285$$

and

$$\text{ABR}(\hat{T}_{HB}, \textit{Tillé} \mid \mathcal{S}_w) \leq 0.1176392792$$

which suggest that stg$(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_w)$ *is more robust against biases for this population.*

5.5.5 Conditional comparison

Although comparisons of strategies should be made unconditionally (or weak conditionally for large n) at the design stage it may be of interest to make conditional comparisons, at the analysis stage. We saw in section 5.3 that a conditional comparison can be made between $stg(\hat{T}_{st}, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ and $stg(\hat{T}_0, \text{SRS} \mid \mathcal{S}_{\mathbf{n}})$ if the number of poststrata was equal to two. But under a general unequal probability design we can compare $\hat{T}_{HB}$ with $\hat{T}_{HBs}$ for any number of poststrata. This is because in general the conditions for theorem 11 will be satisfied, i.e. the matrix $\mathbf{Q}_2$ from theorem 11 will be nonnegative definite and orthogonal to constant vectors only, and hence theorem 11 can be applied to give a sufficient condition for $stg(\hat{T}_{HBs}, p \mid \mathcal{S}_{\mathbf{n}})$ to be better than $stg(\hat{T}_{HB}, p \mid \mathcal{S}_{\mathbf{n}})$.

Example 40 *Consider a population of size 32 that has been poststratified into three groups of sizes* $\mathbf{N} = (10, 12, 10)$. *We independently generated the values of the size variable from* $\text{N}(100, 10)$. *For the sample size configuration* $\mathbf{n} = (3, 3, 3)$ *we applied theorem 11 to give a sufficient condition for* $stg(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ *to be better than* $stg(\hat{T}_{HB}, \textit{Tillé} \mid \mathcal{S}_{\mathbf{n}})$. *We approximated the values* $\eta_0 = 5.551926$ *and* $\gamma_0 = 0.005163941866$ *so by theorem 11 a sufficient condition for the superiority of* $stg(\hat{T}_{HBs}, \textit{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ *is if*

$$R_Y \leq \frac{0.005163941866}{5.551926} = 0.0009301172.$$

Although the sufficient condition for the superiority of $stg(\hat{T}_{HBs}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ in example 40 holds, the value for the upper bound on R_Y is extremely small. This also happened when we compared $stg(\hat{T}_{Rs}, \text{StRS})$ with $stg(\hat{T}_{Rc}, \text{StRS})$ in example 25 on p.158 and has happen here for the same reason. The nullspace of the matrix $\mathbf{M}(\hat{T}_{HB}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ is of dimension 2 and contains vectors that are close to being constant within strata. This implies that $stg(\hat{T}_{HB}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ is not just calibrated for constant vectors but it's also calibrated for vectors that are close to being constant within strata. Hence it is possible for $stg(\hat{T}_{HB}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ to be better than $stg(\hat{T}_{HBs}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})$ for some vectors with small values of R_Y.

5.6 Choice of inclusion probability weights

When considering estimators that are functions of the inclusion probabilities under a poststratification we have a genuine choice between which inclusion probabilities to use: π_i, π_i' or π_i^*; and it isn't always clear which inclusion probability weights to use. Also it is not obvious how to compare estimators with different probability weights, but one possible way we could compare these estimators is by using corollary 1 to compute their maximum relative efficiencies and theorem 8 to calculate the exact upper bound on their absolute bias ratios.

In this section we will empirically compare an estimator with different inclusion probability weights under the unconditional, weak conditional and conditional design using an unequal probability sampling scheme.

The estimators we will consider are:

1. The Hájek-Basu estimator
$$\hat{T}_{HB} = N\frac{\sum_{i\in s} y_i/w_i}{\sum_{i\in s} 1/w_i}$$
with $w_i = \pi_i$, π_i' and π_i^*.

2. The separate Hájek-Basu estimator
$$\hat{T}_{HBs} = \sum_{h=1}^{H} I_h N_h \frac{\sum_{i\in s_h} y_i/w_i}{\sum_{i\in s_h} 1/w_i}$$
with $w_i = \pi_i$, π_i' and π_i^*.

3. The Doss type separate Hájek-Basu estimator
$$\hat{T}_{HBsD} = \frac{N}{\left(\sum_{h=1}^{H} I_h N_h\right)} \sum_{h=1}^{H} I_h N_h \frac{\sum_{i\in s_h} y_i/w_i}{\sum_{i\in s_h} 1/w_i}$$
with $w_i = \pi_i$, π_i' and π_i^*.

Note that the numerator of $\hat{T}_{HB}$, which is the Horvitz-Thompson estimator, when $w_i = \pi_i$, π_i' and π_i^* will respectively be unconditionally unbiased, weak conditionally unbiased and conditionally unbiased for T_Y. But this tells us nothing about the variation of $\sum_{i\in s} y_i/w_i$. The same thing can be said about the numerators of the terms in $\hat{T}_{HBs}$ and $\hat{T}_{HBsD}$, i.e. $\sum_{i\in s_h} y_i/w_i$, for $h = 1, \ldots, H$.

For convenience we will denote any estimator, $\hat{T}_s$, that uses probability weights π_i, π_i' and π_i^* by $\hat{T}_s$, $\hat{T}_s'$ and $\hat{T}_s^*$ respectively.

Provided the probability of $n_h = 0$ is small for every poststratum we would expect π_i to be close to π_i' and little differences between $\hat{T}_s$ and $\hat{T}_s'$. But π_i and π_i^* could differ greatly and so it is of interest to see how $\hat{T}_s$ and $\hat{T}_s^*$ will perform.

In our empirical studies we will consider a population of size 26 and three auxiliary or size vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ for which the sampling design will be based on. The values of $\mathbf{x}_l$ for $l = 1$, 2 and 3 were independently and randomly generated from $\mathrm{N}(100, [10 + 20(l-1)])$. The coefficient of variation of these size variables are

$$\mathrm{cv}(\mathbf{x}_1) = 0.127269, \quad \mathrm{cv}(\mathbf{x}_2) = 0.2730108 \quad \text{and} \quad \mathrm{cv}(\mathbf{x}_3) = 0.4544409.$$

Like in example 35 we poststratified the population into two groups where units 1, 2, ..., 13 fall into poststratum 1 and the rest fall into poststratum 2 and we used Tillé's procedure for unequal probability sampling for a sample of size 6.

Using Corollary 1, on p.99, we calculated the maximum relative efficiency of each estimator under different probability weights and used theorem 8 to give an exact upper bound on the absolute bias ratio for each strategy.

5.6.1 Unconditional comparison

We begin with the Hájek-Basu estimator under the unconditional design. Table 5.4 gives the maximum relative efficiencies of this estimator under the different probability weights by applying Tillé's procedure to each size variable. Table 5.5 gives the exact upper bound on the absolute bias ratio for each strategy.

Table 5.4: The maximum relative efficiencies for $\hat{T}_{HB}$, $\hat{T}'_{HB}$ and $\hat{T}^{*}_{HB}$ under the unconditional design using Tillé's procedure on each size variable: $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. The ij^{th} entry of the block of size 3×3 in the last three columns for each size vector is equal to MRE(i^{th} estimator, Tillé | j^{th} estimator, Tillé)

Size vector		$\hat{T}_{HB}$	$\hat{T}'_{HB}$	$\hat{T}^{*}_{HB}$
$\mathbf{x}_1$	$\hat{T}_{HB}$	1.0000	1.0038	8.9614
	$\hat{T}'_{HB}$	1.0043	1.0000	8.9606
	$\hat{T}^{*}_{HB}$	1.3599	1.3650	1.0000
$\mathbf{x}_2$	$\hat{T}_{HB}$	1.0000	1.0114	8.9707
	$\hat{T}'_{HB}$	1.0099	1.0000	8.9676
	$\hat{T}^{*}_{HB}$	1.5256	1.5427	1.0000
$\mathbf{x}_3$	$\hat{T}_{HB}$	1.0000	1.0040	10.2151
	$\hat{T}'_{HB}$	1.0047	1.0000	10.2097
	$\hat{T}^{*}_{HB}$	1.5223	1.5281	1.0000

Table 5.5: The exact upper bounds on the absolute unconditional bias ratio for $\hat{T}_{HB}$, $\hat{T}'_{HB}$ and $\hat{T}^*_{HB}$ under Tillé's procedure using the size vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$

Size vector	Estimator $\hat{T}_s$	Upper bound on ABR($\hat{T}_s$, Tillé)
$\mathbf{x}_1$	$\hat{T}_{HB}$	0.0458
	$\hat{T}'_{HB}$	0.0450
	$\hat{T}^*_{HB}$	0.0580
$\mathbf{x}_2$	$\hat{T}_{HB}$	0.1058
	$\hat{T}'_{HB}$	0.1056
	$\hat{T}^*_{HB}$	0.1336
$\mathbf{x}_3$	$\hat{T}_{HB}$	0.5535
	$\hat{T}'_{HB}$	0.5524
	$\hat{T}^*_{HB}$	0.5798

We see from table 5.4 that there is very little difference in the mean square error of $\hat{T}_{HB}$ and $\hat{T}^{'}_{HB}$. But the differences can be huge between $\hat{T}^{*}_{HB}$ and $\hat{T}_{HB}$ or $\hat{T}^{'}_{HB}$. In particular the mean square error of $\hat{T}_{HB}$ and $\hat{T}^{'}_{HB}$ can be much larger (around 796% and 921% larger) than the mean square error of $\hat{T}^{*}_{HB}$ compared to how large the mean square error of $\hat{T}^{*}_{HB}$ can be over that of $\hat{T}_{HB}$ and $\hat{T}^{'}_{HB}$. The differences in the mean square errors of these strategies seem to increase as the coefficient of variation of the size variable increases.

Table 5.5 shows little differences in the exact upper bound on the absolute bias ratio of $\hat{T}_{HB}$ and $\hat{T}^{'}_{HB}$ with the same size vector. But the upper bound on the absolute bias ratio of $\hat{T}^{*}_{HB}$ is always larger than that of $\hat{T}_{HB}$ and $\hat{T}^{'}_{HB}$.

Now we consider the estimators $\hat{T}_{HBs}$, $\hat{T}^{'}_{HBs}$ and $\hat{T}^{*}_{HBs}$ under the unconditional design using Tillé's procedure for unequal probability sampling. Table 5.6 gives the maximum relative efficiencies of these strategies and table 5.7 gives the exact upper bound on the absolute bias ratio.

Table 5.6: The maximum relative efficiencies for $\hat{T}_{HBs}$, $\hat{T}'_{HBs}$ and $\hat{T}^*_{HBs}$ under the unconditional design using Tillé's procedure on each size variable: $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. The ij^{th} entry of the block of size 3×3 in the last three columns for each size vector is equal to MRE(i^{th} estimator, Tillé | j^{th} estimator, Tillé)

Size vector		$\hat{T}_{HBs}$	$\hat{T}'_{HBs}$	$\hat{T}^*_{HBs}$
$\mathbf{x}_1$	$\hat{T}_{HBs}$	1.0000	1.0009	1.0114
	$\hat{T}'_{HBs}$	1.0014	1.0000	1.0105
	$\hat{T}^*_{HBs}$	1.0114	1.0103	1.0000
$\mathbf{x}_2$	$\hat{T}_{HBs}$	1.0000	1.0016	1.0314
	$\hat{T}'_{HBs}$	1.0024	1.0000	1.0301
	$\hat{T}^*_{HBs}$	1.0175	1.0164	1.0000
$\mathbf{x}_3$	$\hat{T}_{HBs}$	1.0000	1.0018	1.0353
	$\hat{T}'_{HBs}$	1.0029	1.0000	1.0337
	$\hat{T}^*_{HBs}$	1.0223	1.0215	1.0000

Table 5.7: The exact upper bounds on the absolute unconditional bias ratio for $\hat{T}_{HBs}$, $\hat{T}'_{HBs}$ and $\hat{T}^{*}_{HBs}$ under Tillé's procedure using the size vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$

Size vector	Estimator $\hat{T}_s$	Upper bound on ABR($\hat{T}_s$, Tillé)
$\mathbf{x}_1$	$\hat{T}_{HBs}$	0.1703
	$\hat{T}'_{HBs}$	0.1698
	$\hat{T}^{*}_{HBs}$	0.1688
$\mathbf{x}_2$	$\hat{T}_{HBs}$	0.2899
	$\hat{T}'_{HBs}$	0.2899
	$\hat{T}^{*}_{HBs}$	0.2878
$\mathbf{x}_3$	$\hat{T}_{HBs}$	1.1720
	$\hat{T}'_{HBs}$	1.1710
	$\hat{T}^{*}_{HBs}$	1.1763

We see from table 5.6 that there is small difference in the efficiency of $\hat{T}_{HBs}$, $\hat{T}'_{HBs}$ and $\hat{T}^*_{HBs}$. But the differences between $\hat{T}^*_{HBs}$ with $\hat{T}_{HBs}$ and the differences between $\hat{T}^*_{HBs}$ with $\hat{T}'_{HBs}$ are larger than the differences between $\hat{T}_{HBs}$ and $\hat{T}'_{HBs}$.

The upper bounds on the absolute bias ratios in table 5.7 are practically the same for strategies with the same size variable.

Now we compare the estimators $\hat{T}_{HBsD}$, $\hat{T}'_{HBsD}$ and $\hat{T}^*_{HBsD}$ under the unconditional design using Tillé's procedure for unequal probability sampling. Table 5.8 gives the maximum relative efficiencies of these strategies and table 5.9 gives the exact upper bound on the absolute bias ratio.

Table 5.8: The maximum relative efficiencies for $\hat{T}_{HBsD}$, $\hat{T}'_{HBsD}$ and $\hat{T}^{*}_{HBsD}$ under the unconditional design using Tillé's procedure on each size variable: $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. The ij^{th} entry of the block of size 3×3 in the last three columns for each size vector is equal to $\text{MRE}(i^{\text{th}} \text{ estimator}, \text{Tillé} \mid j^{\text{th}} \text{ estimator}, \text{Tillé})$

Size vector		$\hat{T}_{HBsD}$	$\hat{T}'_{HBsD}$	$\hat{T}^{*}_{HBsD}$
$\mathbf{x}_1$	$\hat{T}_{HBsD}$	1.0000	1.0009	1.0122
	$\hat{T}'_{HBsD}$	1.0014	1.0000	1.0118
	$\hat{T}^{*}_{HBsD}$	1.0118	1.0112	1.0000
$\mathbf{x}_2$	$\hat{T}_{HBsD}$	1.0000	1.0016	1.0317
	$\hat{T}'_{HBsD}$	1.0024	1.0000	1.0306
	$\hat{T}^{*}_{HBsD}$	1.0195	1.0188	1.0000
$\mathbf{x}_3$	$\hat{T}_{HBsD}$	1.0000	1.0018	1.0369
	$\hat{T}'_{HBsD}$	1.0029	1.0000	1.0363
	$\hat{T}^{*}_{HBsD}$	1.0215	1.0207	1.0000

Table 5.9: The exact upper bounds on the absolute unconditional bias ratio for $\hat{T}_{HBsD}$, $\hat{T}'_{HBsD}$ and $\hat{T}^*_{HBsD}$ under Tillé's procedure using the size vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$

Size vector	Estimator $\hat{T}_s$	Upper bound on ABR($\hat{T}_s$, Tillé)
$\mathbf{x}_1$	$\hat{T}_{HBsD}$	0.1160
	$\hat{T}'_{HBsD}$	0.1151
	$\hat{T}^*_{HBsD}$	0.0308
$\mathbf{x}_2$	$\hat{T}_{HBsD}$	0.2620
	$\hat{T}'_{HBsD}$	0.2605
	$\hat{T}^*_{HBsD}$	0.2605
$\mathbf{x}_3$	$\hat{T}_{HBsD}$	1.1369
	$\hat{T}'_{HBsD}$	1.1358
	$\hat{T}^*_{HBsD}$	1.1449

We see from table 5.8 that there is very little difference in efficiency between $\hat{T}_{HBsD}$, $\hat{T}'_{HBsD}$ and $\hat{T}^*_{HBsD}$.

From table 5.9 the exact upper bounds on the absolute bias ratios are similar for all strategies with the same size variable apart from $\hat{T}^*_{HBsD}$ with $\mathbf{x}_1$ whose values for the upper bound is much smaller than the others.

5.6.2 Weak conditional comparison

Under the weak conditional design it is assumed that $n_h > 0$ for all $h = 1, \ldots, H$ in which case $\hat{T}_{HBsD}$ will reduce to $\hat{T}_{HBs}$. So in this section we will compare the Hájek-Basu estimator and the separate Hájek-Basu estimator with different probability weights under the weak conditional design using Tillé's procedure for unequal probability sampling.

Table 5.10 gives the maximum relative efficiencies of the Hájek-Basu strategies and table 5.11 gives the exact upper bound on the absolute bias ratio.

Table 5.10: The maximum relative efficiencies for $\hat{T}_{HB}$, $\hat{T}'_{HB}$ and $\hat{T}^*_{HB}$ under the weak conditional design using Tillé's procedure on each size variable: $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. The ij^{th} entry of the block of size 3×3 in the last three columns for each size vector is equal to $\text{MRE}[(i^{\text{th}}\text{ estimator}, \text{Tillé} \mid \mathcal{S}_w) \mid (j^{\text{th}}\text{ estimator}, \text{Tillé} \mid \mathcal{S}_w)]$

Size vector		$\hat{T}_{HB}$	$\hat{T}'_{HB}$	$\hat{T}^*_{HB}$
$\mathbf{x}_1$	$\hat{T}_{HB}$	1.0000	1.0038	3474.479
	$\hat{T}'_{HB}$	1.0043	1.0000	3474.180
	$\hat{T}^*_{HB}$	1.3619	1.3670	1.0000
$\mathbf{x}_2$	$\hat{T}_{HB}$	1.0000	1.0115	690.7440
	$\hat{T}'_{HB}$	1.0100	1.0000	690.5025
	$\hat{T}^*_{HB}$	1.5260	1.5433	1.0000
$\mathbf{x}_3$	$\hat{T}_{HB}$	1.0000	1.0040	435.7694
	$\hat{T}'_{HB}$	1.0047	1.0000	435.5415
	$\hat{T}^*_{HB}$	1.5285	1.5285	1.0000

Table 5.11: The exact upper bounds on the absolute weak conditional bias ratio for $\hat{T}_{HB}$, $\hat{T}'_{HB}$ and $\hat{T}^*_{HB}$ under Tillé's procedure using the size vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$

Size vector	Estimator $\hat{T}_s$	Upper bound on ABR($\hat{T}_s$, Tillé $\mid \mathcal{S}_w$)
$\mathbf{x}_1$	$\hat{T}_{HB}$	0.0471
	$\hat{T}'_{HB}$	0.0460
	$\hat{T}^*_{HB}$	0.0572
$\mathbf{x}_2$	$\hat{T}_{HB}$	0.1071
	$\hat{T}'_{HB}$	0.1058
	$\hat{T}^*_{HB}$	0.1287
$\mathbf{x}_3$	$\hat{T}_{HB}$	0.5567
	$\hat{T}'_{HB}$	0.5555
	$\hat{T}^*_{HB}$	0.5823

We see from table 5.10 that there is very little difference in efficiency between $\hat{T}_{HB}$ and $\hat{T}'_{HB}$ but there can be massive gains in efficiency for $\hat{T}^*_{HB}$ over $\hat{T}_{HB}$ and $\hat{T}'_{HB}$ (up to 347348% when using $\mathbf{x}_1$).

Table 5.11 shows that the exact upper bounds on the absolute bias ratios are similar for $\hat{T}_{HB}$ and $\hat{T}'_{HB}$. But the value of the upper bound for $\hat{T}^*_{HB}$ is always larger than the others.

We now compare the separate Hájek-Basu estimator with different probability weights under the weak conditional design using Tillé's sampling scheme. Table 5.12 gives the maximum relative efficiencies of these strategies and table 5.13 gives the exact upper bound on the absolute bias ratio.

Table 5.12: The maximum relative efficiencies for $\hat{T}_{HBs}$, $\hat{T}'_{HBs}$ and $\hat{T}^*_{HBs}$ under the weak conditional design using Tillé's procedure on each size variable: $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. The ij^{th} entry of the block of size 3×3 in the last three columns for each size vector is equal to $\text{MRE}[(i^{\text{th}} \text{ estimator}, \text{Tillé} \mid \mathcal{S}_w) \mid (j^{\text{th}} \text{ estimator}, \text{Tillé} \mid \mathcal{S}_w)]$

Size vector		$\hat{T}_{HBs}$	$\hat{T}'_{HBs}$	$\hat{T}^*_{HBs}$
$\mathbf{x}_1$	$\hat{T}_{HBs}$	1.0000	1.0009	1.0116
	$\hat{T}'_{HBs}$	1.0014	1.0000	1.0106
	$\hat{T}^*_{HBs}$	1.0112	1.0099	1.0000
$\mathbf{x}_2$	$\hat{T}_{HBs}$	1.0000	1.0016	1.0312
	$\hat{T}'_{HBs}$	1.0023	1.0000	1.0299
	$\hat{T}^*_{HBs}$	1.0166	1.0143	1.0000
$\mathbf{x}_3$	$\hat{T}_{HBs}$	1.0000	1.0018	1.0351
	$\hat{T}'_{HBs}$	1.0029	1.0000	1.0335
	$\hat{T}^*_{HBs}$	1.0193	1.0193	1.0000

Table 5.13: The exact upper bounds on the absolute weak conditional bias ratio for $\hat{T}_{HBs}$, $\hat{T}'_{HBs}$ and $\hat{T}^*_{HBs}$ under Tillé's procedure using the size vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$

Size vector	Estimator $\hat{T}_s$	Upper bound on ABR($\hat{T}_s$, Tillé $\mid \mathcal{S}_w$)
$\mathbf{x}_1$	$\hat{T}_{HBs}$	0.1176
	$\hat{T}'_{HBs}$	0.1168
	$\hat{T}^*_{HBs}$	0.1155
$\mathbf{x}_2$	$\hat{T}_{HBs}$	0.2632
	$\hat{T}'_{HBs}$	0.2618
	$\hat{T}^*_{HBs}$	0.2595
$\mathbf{x}_3$	$\hat{T}_{HBs}$	1.1683
	$\hat{T}'_{HBs}$	1.1672
	$\hat{T}^*_{HBs}$	1.1718

Table 5.12 shows little differences in efficiency between all strategies with the same size variable and table 5.13 shows that the exact upper bounds on the absolute bias ratios are similar for strategies with the same size variable.

5.6.3 Conditional comparison

In this section we compare the Hájek-Basu estimator and the separate Hájek-Basu estimator with different probability weights under the conditional design on $\mathbf{n} = (3, 3)$ using Tillé's sampling scheme.

We begin with the Hájek-Basu estimator. Table 5.14 gives the maximum relative efficiency of $\hat{T}_{HB}$, $\hat{T}'_{HB}$ and $\hat{T}^*_{HB}$ under the conditional design and table 5.15 gives the upper bound on the absolute bias ratio.

Table 5.14: The maximum relative efficiencies for $\hat{T}_{HB}$, $\hat{T}'_{HB}$ and $\hat{T}^*_{HB}$ under the conditional design using Tillé's procedure on each size variable: $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. The ij^{th} entry of the block of size 3×3 in the last three columns for each size vector is equal to $\text{MRE}[(i^{\text{th}}\ \text{estimator}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}}) \mid (j^{\text{th}}\ \text{estimator}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})]$

Size vector		$\hat{T}_{HB}$	$\hat{T}'_{HB}$	$\hat{T}^*_{HB}$
$\mathbf{x}_1$	$\hat{T}_{HB}$	1.0000	1.2301	–
	$\hat{T}'_{HB}$	1.0046	1.0000	–
	$\hat{T}^*_{HB}$	1.0483	1.0435	1.0000
$\mathbf{x}_2$	$\hat{T}_{HB}$	1.0000	1.2246	–
	$\hat{T}'_{HB}$	1.01123	1.0000	–
	$\hat{T}^*_{HB}$	1.1244	1.1119	1.0000
$\mathbf{x}_3$	$\hat{T}_{HB}$	1.0000	1.4379	–
	$\hat{T}'_{HB}$	1.2210	1.0000	–
	$\hat{T}^*_{HB}$	10.1353	9.9908	1.0000

Table 5.15: The exact upper bounds on the absolute conditional bias ratio for $\hat{T}_{HB}$, $\hat{T}'_{HB}$ and $\hat{T}^*_{HB}$ under Tillé's procedure using the size vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$

Size vector	Estimator $\hat{T}_s$	Upper bound on ABR($\hat{T}_s$, Tillé $\mid \mathcal{S}_{\mathbf{n}}$)
$\mathbf{x}_1$	$\hat{T}_{HB}$	–
	$\hat{T}'_{HB}$	–
	$\hat{T}^*_{HB}$	0.0487
$\mathbf{x}_2$	$\hat{T}_{HB}$	–
	$\hat{T}'_{HB}$	–
	$\hat{T}^*_{HB}$	0.1081
$\mathbf{x}_3$	$\hat{T}_{HB}$	–
	$\hat{T}'_{HB}$	–
	$\hat{T}^*_{HB}$	0.5785

The conditional mean square error matrices of $\hat{T}^{*}_{HB}$ for each size variable are orthogonal to some nonconstant vectors whereas the conditional mean square error matrices for $\hat{T}_{HB}$ and $\hat{T}^{'}_{HB}$ are orthogonal to constant vectors only. This means we cannot use Corollary 1, on p.99, to give the conditional maximum relative efficiency of $\hat{T}_{HB}$ or $\hat{T}^{'}_{HB}$ over $\hat{T}^{*}_{HB}$.

We see from table 5.14 that for the size vector $\mathbf{x}_1$ and $\mathbf{x}_2$ the estimator $\hat{T}^{'}_{HB}$ may be preferred over $\hat{T}_{HB}$ since its conditional mean square error can respectively be at most only 0.5% and 1.1% larger than that of $\hat{T}_{HB}$ whereas the conditional mean square error of $\hat{T}_{HB}$ can be up to 23% and 22.5% larger than that of $\hat{T}^{'}_{HB}$.

The nullspaces of the conditional variance-covariance matrices for $\hat{T}_{HB}$ and $\hat{T}^{'}_{HB}$ were not contained in the nullspace of their bias square matrices so we cannot apply theorem 8 to give an exact upper bound on their absolute bias ratio.

Now we consider the estimators $\hat{T}_{HBs}$, $\hat{T}^{'}_{HBs}$ and $\hat{T}^{*}_{HBs}$ under the conditional design on $\mathbf{n} = (3, 3)$ using Tillé's procedure for unequal probability sampling. Table 5.16 gives the maximum relative efficiency of these strategies and table 5.17 gives the exact upper bound on the absolute bias ratio.

Table 5.16: The maximum relative efficiencies for $\hat{T}_{HBs}$, $\hat{T}'_{HBs}$ and $\hat{T}^*_{HBs}$ under the conditional design using Tillé's procedure on each size variable: $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. The ij^{th} entry of the block of size 3×3 in the last three columns for each size vector is equal to $\text{MRE}[(i^{\text{th}} \text{ estimator}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}}) \mid (j^{\text{th}} \text{ estimator}, \text{Tillé} \mid \mathcal{S}_{\mathbf{n}})]$

Size vector		$\hat{T}_{HBs}$	$\hat{T}'_{HBs}$	$\hat{T}^*_{HBs}$
$\mathbf{x}_1$	$\hat{T}_{HBs}$	1.0000	1.0014	1.0147
	$\hat{T}'_{HBs}$	1.0017	1.0000	1.0133
	$\hat{T}^*_{HBs}$	1.0177	1.0160	1.0000
$\mathbf{x}_2$	$\hat{T}_{HBs}$	1.0000	1.0030	1.0273
	$\hat{T}'_{HBs}$	1.0027	1.0000	1.0243
	$\hat{T}^*_{HBs}$	1.0298	1.0270	1.0000
$\mathbf{x}_3$	$\hat{T}_{HBs}$	1.0000	1.0034	1.0393
	$\hat{T}'_{HBs}$	1.0032	1.0000	1.0358
	$\hat{T}^*_{HBs}$	1.0389	1.0355	1.0000

Table 5.17: The exact upper bounds on the absolute conditional bias ratio for $\hat{T}_{HBs}$, $\hat{T}'_{HBs}$ and $\hat{T}^*_{HBs}$ under Tillé's procedure using the size vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$

Size vector	Estimator $\hat{T}_s$	Upper bound on ABR($\hat{T}_s$, Tillé $\mid \mathcal{S}_{\mathbf{n}}$)
$\mathbf{x}_1$	$\hat{T}_{HBs}$	0.1092
	$\hat{T}'_{HBs}$	0.1082
	$\hat{T}^*_{HBs}$	0.0989
$\mathbf{x}_2$	$\hat{T}_{HBs}$	0.2384
	$\hat{T}'_{HBs}$	0.2367
	$\hat{T}^*_{HBs}$	0.2216
$\mathbf{x}_3$	$\hat{T}_{HBs}$	1.1890
	$\hat{T}'_{HBs}$	1.1881
	$\hat{T}^*_{HBs}$	1.1782

Table 5.16 shows that there is very little difference in efficiency between $\hat{T}_{HBs}$, $\hat{T}'_{HBs}$ and $\hat{T}^*_{HBs}$. Table 5.17 shows that the exact upper bound on the absolute bias ratio is similar for all strategies with the same size vector but $\hat{T}^*_{HB}$ always has the smallest value.

Our studies show that there were no major differences between using the different probability weights for the separate Hájek-Basu estimator with the same size variable either conditionally on $\mathbf{n}$, weak conditionally or unconditionally. Because calculating the π'_i's and π^*_i's can be difficult it is more appropriate to use the π_i's instead for this estimator.

The Doss type separate Hájek-Basu estimator also performed relatively the same in terms of efficiency for each probability weight under the unconditional design. However $\hat{T}^*_{HBsD}$ can be more robust against bias compared to $\hat{T}_{HBsD}$ and $\hat{T}'_{HBsD}$ for example when using the size variable $\mathbf{x}_1$ which has a small coefficient of variation.

For the Hájek-Basu estimator there was hardly any differences between $\hat{T}_{HB}$ and $\hat{T}'_{HB}$ unconditionally and weak conditionally. But the gains in efficiency can be huge for $\hat{T}^*_{HB}$ over $\hat{T}_{HB}$ and $\hat{T}'_{HB}$ both unconditionally and weak conditionally.

Under the conditional design on $\mathbf{n}$ the estimator $\hat{T}'_{HB}$ seems to be more favourable than $\hat{T}_{HB}$ since $\hat{T}'_{HB}$ can potentially have larger gains in efficiency over $\hat{T}_{HB}$. Conditionally $\hat{T}^*_{HB}$ isn't calibrated just for constant vectors so in this sense it is in a different class to $\hat{T}_{HB}$ and $\hat{T}'_{HB}$ under the conditional design.

In general the differences in efficiency of the strategies increase with the coefficient of variation of the size variable. In almost every case the exact upper bounds on the absolute bias ratio for the Hájek-Basu estimators are much smaller than those of the separate Hájek-Basu estimators with the same size variable.

5.7 Conclusions

For the special case where the poststrata sizes are all equal, theorem 14 can be used to make unconditional comparisons by giving sufficient conditions for superiority of one strategy over another for some estimators under simple random sampling. Also part (a) of theorem 14 is a special case of corollary 2.

For conditional comparisons under simple random sampling we give a result that provides a sufficient condition for the stratified estimator to be better than the expansion estimator when there are a total of two poststrata. Also, for designs which are independent of the population order, we give a result that provides a sufficient condition for the separate Hájek-Basu estimator to be better than the Hájek-Basu estimator for the domain total under the conditional design where the poststratification is by the values of the size variable.

In section 5.5 it is demonstrated that when poststratifying under an unequal probability design the conditional approach gives a more accurate coverage probability, with respect to those samples with the same sampling allocation as the observed one, than the unconditional approach. We also make unconditional, weak conditional and conditional comparisons of the Hájek-Basu type estimators under Tillé's sampling procedure by means of corollary 2 and theorem 11.

In general it is not obvious how to choose between the different inclusion probability weights for an estimator. But applying corollary 1 to compare the maximum relative efficiencies of estimators with different probability weights can give us an indication of which weights to use as demonstrated in our numerical examples.

In this chapter we only considered Tillé's procedure for unequal probability sampling. This is because of its simplicity as the design probabilities from this scheme are easy to calculate compared to other sampling schemes. However the main focus in this chapter are the techniques used to make comparisons of strategies under a post-stratification. These techniques (theorems 8 and 11 and corollaries 1 and 2) can be applied to estimators under any general unequal probability design. Tillé's procedure was used here to demonstrate these techniques.

Chapter 6

Nonnegative Definite Variance Estimation for $stg(\hat{T}_R, \text{PPAS})$

The topic of this chapter is somewhat different from the rest of the thesis. In this chapter we introduce a class of nonnegative definite unbiased estimators for the variance of the ratio estimator under probability proportional to aggregate size sampling. The way in which this class of estimators is derived is by analysing the matrix given by a variance estimator and the details of this will be described. We will then do an empirical study to examine the performance of estimators from this class.

6.1 Introduction

One of the problems with unbiased variance estimators when using an unequal probability design is the possibility of obtaining an estimate for the variance that is a negative value. This is clearly a problem since calculating confidence intervals is not possible with a negative variance estimate. Also if the probability of obtaining a negative estimate for the variance is large then this would reduce the coverage probability of the strategy.

Many strategies have a large sample approximation formula for their variance which is approximately unbiased (e.g. see (4.20) for $stg(\hat{T}_R, \text{SRS})$) and can be estimated nonnegatively but these formulae are not valid for small samples. It is therefore of interest to use an estimator for the variance which is a nonnegative definite unbiased estimator (NNDUE).

Definition 21 *An estimator for the variance or mean square error of a strategy is said to be a nonnegative definite unbiased estimator if it is uniformly nonnegative and unbiased for all* $\mathbf{y} \in \mathcal{R}^N$.

Vijayan (1975) considered a class of unbiased estimators for the variance of the Horvitz-Thompson estimator, $\text{Var}(\hat{T}_{HT}, p)$, with fixed sample size n which is given by

$$-\frac{1}{2}\sum_{i\in s}\sum_{\substack{j\in s\\ j\neq i}} a_{ij}(s)\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \tag{6.1}$$

where the $a_{ij}(s)$'s do not depend on the y's. For unbiasedness every pair of units $i, j \in \mathcal{U}$ $(i \neq j)$ must satisfy

$$\sum_{s\ni i,j} p(s)a_{ij}(s) = \pi_{ij} - \pi_i\pi_j.$$

The Sen, Yates and Grundy estimator v_{SYG} (see ch.2 p.57) for $\text{Var}(\hat{T}_{HT}, p)$ is a special case of (6.1) where

$$a_{ij}(s) = \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}}$$

for all $i, j \in \mathcal{U}$.

Note that any estimator given by (6.1) will be zero for all samples in the support whenever the y's are proportional to the π's. Vijayan (1975) showed that it is necessary that a nonnegative definite unbiased estimator for $\mathrm{Var}(\hat{T}_{HT}, p)$ is of the form (6.1). He also proved that when $n = 2$, a nonnegative definite unbiased estimator exists if and only if $\pi_{ij} \leq \pi_i \pi_j$ for all $i, j \in \mathcal{U}$, $i \neq j$ and $a_{ij}(s)$ is as in v_{SYG}.

Rao & Vijayan (1977) generalized (6.1) for any general linear strategy calibrated for some fixed $\mathbf{x} \in \mathcal{R}^N$, such that $x_i > 0$ for all $i \in \mathcal{U}$. They observed that the mean square error of $stg(\hat{T}_s, p)$ can be written as

$$\mathrm{MSE}(\hat{T}_s, p) = \mathbf{y}^t \mathbf{M}(\hat{T}_s, p)\mathbf{y} = \mathbf{z}^t \mathbf{A}(\hat{T}_s, p)\mathbf{z} = -\frac{1}{2} \sum_{i \in \mathcal{U}} \sum_{\substack{j \in \mathcal{U} \\ j \neq i}} a_{ij}(z_i - z_j)^2$$

where $z_i = y_i / x_i$ for every $i \in \mathcal{U}$, and

$$a_{ij} = \left(\sum_{s \ni i,j} p(s) b_{si} b_{sj} - B_i - B_j + 1 \right) x_i x_j$$

is the ij^{th} element of the $N \times N$ matrix $\mathbf{A}(\hat{T}_s, p)$. They also gave the form of a nonnegative definite unbiased estimator for $\mathrm{MSE}(\hat{T}_s, p)$ as

$$-\frac{1}{2} \sum_{i \in s} \sum_{\substack{j \in s \\ j \neq i}} a_{ij}(s)(z_i - z_j)^2 \tag{6.2}$$

where $a_{ij}(s)$ does not depend on the y's and satisfies the unbiasedness condition

$$\sum_{s \ni i,j} p(s) a_{ij}(s) = a_{ij} \tag{6.3}$$

for all $i, j \in \mathcal{U}$ $(i \neq j)$. Note that estimators of the class (6.2) will be zero for all $s \in \mathcal{S}$ whenever $\mathbf{z}$ is equal to a constant vector. This means that for any general linear strategy which is calibrated for vectors that are proportional to $\mathbf{x}$, not only will its mean square error be equal to zero whenever $\mathbf{y}$ is proportional to $\mathbf{x}$ but any estimator in the class (6.2) will also be zero for all $s \in \mathcal{S}$.

Rao (1979) gave some formulae for $a_{ij}(s)$ that achieve the unbiasedness condition (6.3) but none of them achieve uniform nonnegativity.

Remark 9 *The estimators of the class (6.2) can be written as the quadratic form*

$$\mathbf{z}_s^t \mathbf{A}(s) \mathbf{z}_s$$

where $\mathbf{z}_s$ *is an* $n \times 1$ *vector with* $z_i = y_i/x_i$ *for* $i \in s$*, and the* ij^{th} *element of the* $n \times n$ *matrix* $\mathbf{A}(s)$ *is equal to*

$$\mathbf{A}(s)_{ij} = \begin{cases} -\sum\limits_{\substack{k \in s \\ k \neq i}} a_{ik}(s) & \textit{for} \quad i = j \in s \\ \\ a_{ij}(s) & \textit{for} \quad i \neq j \in s. \end{cases} \tag{6.4}$$

It is easily seen that a sufficient condition for estimators of the class (6.2) to be nonnegative definite is that the coefficients $a_{ij}(s)$ are negative for all $i, j \in \mathcal{U}$. However, a necessary and sufficient condition for these estimators to be nonnegative definite is the nonnegative definiteness of the matrix $\mathbf{A}(s)$ given in (6.4) for all samples in the support given by the design. Padmawar (1998) realized this and stated that one of the desired properties of a mean square error estimator is that the smallest eigenvalue of the matrix given by the mean square error estimator should be nonnegative, as this will avoid negative estimates.

In the next section we will construct a class of unbiased estimators for the variance of the ratio estimator under probability proportional to aggregate size sampling (see example 11 on p.103) that is nonnegative definite. The way in which the class of estimators is constructed is by satisfying the nonnegative definiteness condition on the matrices given by the variance estimator for each sample in $\mathcal{S}$.

6.2 NNDUEs for Var($\hat{T}_R$, PPAS)

In this section we are concerned with the ratio estimator under probability proportional to aggregate size sampling where the selection probability for each sample in $\mathcal{S}$

is equal to $p(s) = n\bar{x}/M_1 N\bar{X}$ with $M_1 = \binom{N-1}{n-1}$ and $x_i > 0$ for all $i \in \mathcal{U}$. As shown in example 11 this strategy is unbiased for the population total T_Y. The reason why we have chosen this strategy will become clear when we construct a class of nonnegative definite unbiased estimators for its variance.

Rao & Vijayan (1977) also focused their attention on this strategy in their paper and proposed two unbiased estimators for its variance which are of the same form as (6.2). In an empirical study they showed that one of their estimators performed better than the other in terms of higher efficiency and a lower probability of getting a negative estimate. This estimator is given by

$$v_{RV} = -\frac{1}{2}\sum_{i\in s}\sum_{\substack{j\in s\\ j\neq i}}\frac{a_{ij}}{\pi_{ij}}(z_i - z_j)^2$$

where

$$\begin{aligned} a_{ij} &= \left(\sum_{s\ni i,j} p(s)b_{si}b_{sj} - 1\right)x_i x_j = \left(\sum_{s\ni i,j}\frac{n\bar{x}}{M_1 N\bar{X}}\left(\frac{N\bar{X}}{n\bar{x}}\right)^2 - 1\right)x_i x_j \\ &= \left(\sum_{s\ni i,j}\frac{N\bar{X}}{M_1 n\bar{x}} - 1\right)x_i x_j = \left(\sum_{s\ni i,j}\frac{1}{M_1^2 p(s)} - 1\right)x_i x_j \qquad (6.5) \end{aligned}$$

with $M_1 = \binom{N-1}{n-1}$.

It is clear by the form of v_{RV} that when a_{ij} is greater than zero then a negative contribution will be made towards the estimate v_{RV} by units $i, j \in s$, $i \neq j$. Since it is assumed that $x_i > 0$ for all $i \in \mathcal{U}$ observe from (6.5) that

$$\begin{aligned} a_{ij} > 0 \quad &\Leftrightarrow \quad \sum_{s\ni i,j}\frac{1}{M_1^2 p(s)} - 1 > 0 \\ &\Leftrightarrow \quad \sum_{s\ni i,j}\frac{1}{p(s)} > M_1^2. \qquad (6.6) \end{aligned}$$

This suggests that when some or most of the samples involved in the summation of (6.6) have small selection probabilities, then a_{ij} can be greater than zero and hence a negative contribution is made towards v_{RV}. It is this negative contribution we intend to reduce but at the same time preserve the unbiasedness.

Rao & Vijayan (1977) also observed, from their empirical studies, that negative estimates given by v_{RV} correspond to those samples with small selection probabilities. Similarly, when some or most of the samples involved in the summation have large probabilities of selection then a_{ij} can be less than zero, and hence a positive contribution is made towards v_{RV}. It is this relation that v_{RV} tends to be negative for small $p(s)$ and nonnegative for large $p(s)$ that we wish to exploit in an effort to construct a nonnegative definite unbiased estimator for $\text{Var}(\hat{T}_R, \text{PPAS})$.

Define $a_{ij}(s)$ in (6.2) by

$$\frac{a_{ij}}{\pi_{ij}} - \frac{\alpha}{N}\left(\frac{1}{M_2 p(s)} - \frac{1}{\pi_{ij}}\right)$$

where $M_2 = \binom{N-2}{n-2}$, α is a constant term over all $s \in \mathcal{S}$ and let $\mathcal{S}_n$ be the sampling support that contains all $\binom{N}{n}$ samples of fixed size n (this is the case for probability proportional to aggregate size sampling). Then provided $p(s) > 0$ for all $s \in \mathcal{S}_n$ we have

$$\sum_{s \ni i,j} p(s)\left[\frac{a_{ij}}{\pi_{ij}} - \frac{\alpha}{N}\left(\frac{1}{M_2 p(s)} - \frac{1}{\pi_{ij}}\right)\right] = a_{ij} - 0$$

so that the unbiasedness condition (6.3) is satisfied. We then have the following class of unbiased estimators for $\text{Var}(\hat{T}_R, \text{PPAS})$ which is contained in the class of (6.2)

$$v_\alpha = -\frac{1}{2}\sum_{i \in s}\sum_{\substack{j \in s \\ j \neq i}}\left[\frac{a_{ij}}{\pi_{ij}} - \frac{\alpha}{N}\left(\frac{1}{M_2 p(s)} - \frac{1}{\pi_{ij}}\right)\right](z_i - z_j)^2,$$

where $\alpha = 0$ is the special case of v_{RV}. Note that the value of α does not have to be greater than zero in order for v_α to be unbiased.

Observe that when α decreases to zero the negative contribution made towards v_α by units $i, j \in s$ $(i \neq j)$ for those samples s which contains units i and j where

(6.6) holds increases if $\pi_{ij} > M_2 p(s)$. The inequality $\pi_{ij} > M_2 p(s)$, for $i, j \in s$, will in general be true for samples with small $p(s)$.

The idea of v_α is basically to take the 'standard' estimator v_{RV} and add a multiple of an unbiased estimator of zero. We know that v_{RV} can give negative estimates for samples with small probabilities of selection but v_α can avoid negative estimates for those samples with a suitably large α. However if α is too large v_α could give negative estimates for other samples which do not have a small probability of selection. Then the question now is, does there exist an α such that v_α is nonnegative definite? The following steps can help us to obtain a range of possible values for α, provided such an α exists.

First we will need the following lemma.

Lemma 17 *Two nonnegative definite unbiased estimators for* $(N-1)S_Z^2$ *are:*

$$m_1 = \frac{1}{2} \sum_{i \in s} \sum_{\substack{j \in s \\ j \neq i}} \frac{(z_i - z_j)^2}{N M_2 p(s)}$$

provided $p(s) > 0$ *for all* $s \in \mathcal{S}_n$*, and*

$$m_2 = \frac{1}{2} \sum_{i \in s} \sum_{\substack{j \in s \\ j \neq i}} \frac{(z_i - z_j)^2}{N \pi_{ij}}$$

provided $\pi_{ij} > 0$ *for all pairs of units* $i, j \in \mathcal{U}$*, where* $M_2 = \binom{N-2}{n-2}$.

Proof Observe that

$$\begin{aligned}
(N-1)S_Z^2 &= \sum_{i\in\mathcal{U}} z_i^2 - \frac{1}{N}\left(\sum_{i\in\mathcal{U}} z_i\right)^2 \\
&= \frac{1}{N}\left(N\sum_{i\in\mathcal{U}} z_i^2 - \sum_{i\in\mathcal{U}} z_i^2 - \sum_{i\in\mathcal{U}}\sum_{\substack{j\in\mathcal{U}\\ j\neq i}} z_i z_j\right) \\
&= \frac{1}{N}\left[\frac{1}{2}\sum_{i\in\mathcal{U}} z_i^2(N-1) + \frac{1}{2}\sum_{j\in\mathcal{U}} z_j^2(N-1) - \sum_{i\in\mathcal{U}}\sum_{\substack{j\in\mathcal{U}\\ j\neq i}} z_i z_j\right] \\
&= \frac{1}{2N}\sum_{i\in\mathcal{U}}\sum_{\substack{j\in\mathcal{U}\\ j\neq i}}(z_i - z_j)^2. \qquad (6.7)
\end{aligned}$$

The expected value of m_1 is equal to

$$\sum_{s\in\mathcal{S}_n} p(s)\left(\frac{1}{2}\sum_{i\in s}\sum_{\substack{j\in s\\ j\neq i}}\frac{(z_i - z_j)^2}{NM_2 p(s)}\right).$$

Provided $p(s) > 0$ for all $s \in \mathcal{S}_n$ we can write this as

$$\begin{aligned}
\frac{1}{2NM_2}\sum_{s\in\mathcal{S}_n}\left(\sum_{i\in s}\sum_{\substack{j\in s\\ j\neq i}}(z_i - z_j)^2\right) &= \frac{M_2}{2NM_2}\sum_{i\in\mathcal{U}}\sum_{\substack{j\in\mathcal{U}\\ j\neq i}}(z_i - z_j)^2 \\
&= \frac{1}{2N}\sum_{i\in\mathcal{U}}\sum_{\substack{j\in\mathcal{U}\\ j\neq i}}(z_i - z_j)^2
\end{aligned}$$

which is equal to $(N-1)S_Z^2$ in the form of (6.7).

The expected value of m_2 is equal to

$$\sum_{s\in\mathcal{S}_n} p(s)\left(\frac{1}{2}\sum_{i\in s}\sum_{\substack{j\in s\\ j\neq i}}\frac{(z_i - z_j)^2}{N\pi_{ij}}\right).$$

Provided $\pi_{ij} > 0$ for all $i, j \in \mathcal{U}$ $(i \neq j)$ we can write this as

$$\frac{1}{2}\sum_{i\in\mathcal{U}}\sum_{\substack{j\in\mathcal{U}\\ j\neq i}}\frac{\mathrm{E}(t_i t_j, p)}{\pi_{ij}}(z_i - z_j)^2 = \frac{1}{2N}\sum_{i\in\mathcal{U}}\sum_{\substack{j\in\mathcal{U}\\ j\neq i}}(z_i - z_j)^2$$

which is equal to $(N-1)S_Z^2$ and that completes the proof. □

Remark 10 *Under simple random sampling $M_2 p(s) = \pi_{ij}$ for all $i, j \in \mathcal{U}$ $(i \neq j)$ which implies m_1 and m_2 are the same and v_α is equal to v_{RV}. Also under any general design with $n = 2$, $M_2 p(s)$ will be equal to π_{ij} for all $i, j \in \mathcal{U}$, $i \neq j$.*

Now we will write the quadratic form of v_α. Note that m_1 and m_2, from lemma 17, can be written as the following quadratic forms in $\mathbf{z}_s$

$$m_1 = \mathbf{z}_s^t \mathbf{P}_1(s)\mathbf{z}_s$$

where the ij^{th} element of the $n \times n$ matrix $\mathbf{P}_1(s)$ is equal to

$$\mathbf{P}_1(s)_{ij} = \begin{cases} \frac{(n-1)}{NM_2p(s)} & \text{for} \quad i = j \in s \\ -\frac{1}{NM_2p(s)} & \text{for} \quad i \neq j \in s, \end{cases}$$

and

$$m_2 = \mathbf{z}_s^t \mathbf{P}_2(s)\mathbf{z}_s$$

where the ij^{th} element of the $n \times n$ matrix $\mathbf{P}_2(s)$ is equal to

$$\mathbf{P}_2(s)_{ij} = \begin{cases} \frac{1}{N}\sum\limits_{\substack{k\in s\\ k\neq i}}\frac{1}{\pi_{ik}} & \text{for} \quad i = j \in s \\ -\frac{1}{N\pi_{ij}} & \text{for} \quad i \neq j \in s. \end{cases}$$

Hence we can write the estimator v_α as the following quadratic form

$$v_\alpha = \mathbf{z}_s^t(\mathbf{A}(s) + \alpha[\mathbf{P}_1(s) - \mathbf{P}_2(s)])\mathbf{z}_s = \mathbf{z}_s^t\mathbf{D}(s)\mathbf{z}_s$$

where $\mathbf{A}(s)$ is the $n \times n$ symmetric matrix given by v_{RV} with ij^{th} elements

$$\mathbf{A}(s)_{ij} = \begin{cases} -\sum\limits_{\substack{k \in s \\ k \neq i}} \frac{a_{ik}}{\pi_{ik}} & \text{for} \quad i = j \in s \\ \frac{a_{ij}}{\pi_{ij}} & \text{for} \quad i \neq j \in s \end{cases}$$

and $\mathbf{D}(s) = \mathbf{A}(s) + \alpha(\mathbf{P}_1(s) - \mathbf{P}_2(s))$ for $s \in \mathcal{S}_n$.

If, for a given α which is constant over all $s \in \mathcal{S}_n$, the eigenvalues for $\mathbf{D}(s)$ are all nonnegative for every sample in $\mathcal{S}_n$ then the estimator v_α will also be a nonnegative definite unbiased estimator.

Now we will find some conditions for $\mathbf{D}(s)$ to be nonnegative definite for all $s \in \mathcal{S}_n$. Observe that $\mathbf{P}_1(s)$ is of the same form as (4.1) on p.124. Since it is orthogonal to constant vectors by lemma 12 it has one nonzero eigenvalue, denoted by $\gamma(s)$, which is equal to

$$\gamma(s) = \frac{n}{NM_2 p(s)} \quad \text{for} \quad s \in \mathcal{S}_n.$$

The nullspace of $\mathbf{P}_1(s)$ is spanned by only the constant vector in $\mathcal{R}^n$. Hence it is easily seen by using lemma 13 that

$$\mathbf{z}_s^t \mathbf{P}_1(s) \mathbf{z}_s = \gamma(s)(n-1) S_{z_s}^2 \quad \text{for} \quad s \in \mathcal{S}_n \tag{6.8}$$

where S_{z_s} is the sample standard deviation of the sampled units of $\mathbf{z}$. Also provided $\mathbf{P}_2(s)$ is orthogonal to constant vectors only we see that, from (4.7),

$$\mu_{min}(s)(n-1)S_{z_s}^2 \leq \mathbf{z}_s^t \mathbf{P}_2(s) \mathbf{z}_s \leq \mu_{max}(s)(n-1)S_{z_s}^2 \quad \text{for} \quad s \in \mathcal{S}_n \tag{6.9}$$

where $\mu_{min}(s)$ and $\mu_{max}(s)$ are the respective minimum and maximum nonzero eigenvalues of $\mathbf{P}_2(s)$. Since $v_{RV} = 0$ when $\mathbf{z}_s$ is constant this implies that $\mathbf{A}(s)$ is orthogonal to constant vectors for all $s \in \mathcal{S}_n$. Provided the dimension of the nullspace of $\mathbf{A}(s)$ is equal to one we have,

$$v_{RV} = \mathbf{z}_s^t \mathbf{A}(s) \mathbf{z}_s \geq \lambda_{min}(s)(n-1)S_{z_s}^2 \quad \text{for} \quad s \in \mathcal{S}_n \tag{6.10}$$

where $\lambda_{min}(s)$ is the minimum nonzero eigenvalue of $\mathbf{A}(s)$.

Now, let $t(s) = \gamma(s) - \mu_{max}(s)$ and $r(s) = \gamma(s) - \mu_{min}(s)$ for $s \in \mathcal{S}_n$. The range of values for α, such that v_α is nonnegative definite, may be over positive and negative values. If $\alpha \geq 0$ then it is easily seen by (6.8)-(6.10) that

$$v_\alpha = \mathbf{z}_s^t \mathbf{D}(s) \mathbf{z}_s \geq [\lambda_{min}(s) + \alpha t(s)](n-1) S_{z_s}^2$$

for every $s \in \mathcal{S}_n$. Similarly if $\alpha \leq 0$ then

$$v_\alpha = \mathbf{z}_s^t \mathbf{D}(s) \mathbf{z}_s \geq [\lambda_{min}(s) + \alpha r(s)](n-1) S_{z_s}^2.$$

Hence a sufficient condition for $\mathbf{D}(s)$ to be nonnegative definite for some $s \in \mathcal{S}_n$ is if either

$$\alpha > 0 \quad \text{and} \quad \lambda_{min}(s) + \alpha t(s) \geq 0 \tag{6.11}$$

or if

$$\alpha < 0 \quad \text{and} \quad \lambda_{min}(s) + \alpha r(s) \geq 0. \tag{6.12}$$

We need to consider both of these cases. If there exists a constant value α over all $s \in \mathcal{S}_n$ such that each sample in $\mathcal{S}_n$ satisfies one of the conditions (6.11) or (6.12) then this implies that $\mathbf{D}(s)$ will be nonnegative definite for all samples in $\mathcal{S}_n$ and hence, v_α will be a nonnegative definite unbiased estimator for $\text{Var}(\hat{T}_R, \text{PPAS})$.

We will now describe how we find a range of values for α such that v_α is nonnegative definite. Let the interval (α_L, α_U), for $\alpha_L, \alpha_U \in \mathcal{R}$, be a range of values for α such that the estimator v_α is nonnegative definite. The following definition will be useful in explaining how we obtain such an interval.

Definition 22 *The interval $(\alpha_L(s), \alpha_U(s))$, for some $s \in \mathcal{S}_n$ with $\alpha_L(s), \alpha_U(s) \in \mathcal{R}$, is called an interval for nonnegative definiteness of $\mathbf{D}(s)$ if for every $\alpha \in (\alpha_L(s), \alpha_U(s))$ the matrix $\mathbf{D}(s)$ is nonnegative definite.*

Suppose an interval for nonnegative definiteness of $\mathbf{D}(s)$ is known for each sample in $\mathcal{S}_n$. If the intersection of all the $\binom{N}{n}$ intervals is nonempty, i.e. if there are some specific values contained in every $\binom{N}{n}$ interval, then the range of the values in this intersection will be given by the interval $(\max_{s\in\mathcal{S}_n}\{\alpha_L(s)\}, \min_{s\in\mathcal{S}_n}\{\alpha_U(s)\})$. It is clear that for any value of α from this interval the matrix $\mathbf{D}(s)$ will be nonnegative definite for all samples in $\mathcal{S}_n$ and v_α will also be nonnegative definite. So provided the intersection of intervals for nonnegative definiteness of $\mathbf{D}(s)$ for all samples is nonempty we can obtain the interval (α_L, α_U) by letting $\alpha_L = \max_{s\in\mathcal{S}_n}\{\alpha_L(s)\}$ and $\alpha_U = \min_{s\in\mathcal{S}_n}\{\alpha_U(s)\}$.

By using (6.11) and (6.12), we can obtain an interval for nonnegative definiteness of $\mathbf{D}(s)$ for each sample in $\mathcal{S}_n$. It will be desirable to obtain intervals with the largest range in order to obtain a large range for the interval (α_L, α_U). Each sample in $\mathcal{S}_n$ will satisfy one of the following cases for which we give the interval $(\alpha_L(s), \alpha_U(s))$ where possible:

1. If for some sample $s \in \mathcal{S}_n$ we have

$$\lambda_{min}(s) > 0, \quad t(s) < 0 \quad \text{and} \quad r(s) > 0$$

   then by using (6.11) we can choose

$$\alpha_U(s) = \frac{-\lambda_{min}(s)}{t(s)}$$

   and by (6.12) we can choose

$$\alpha_L(s) = \frac{-\lambda_{min}(s)}{r(s)}$$

   so that the interval $(\alpha_L(s), \alpha_U(s))$ will be an interval for nonnegative definiteness of $\mathbf{D}(s)$.

2. If for some sample $s \in \mathcal{S}_n$ we have

$$\lambda_{min}(s) > 0, \quad t(s) < 0 \quad \text{and} \quad r(s) < 0$$

then by using (6.11) we can choose

$$\alpha_U(s) = \frac{-\lambda_{min}(s)}{t(s)}$$

and from (6.12) we can choose $\alpha_L(s) = -\infty$. Then the interval $(\alpha_L(s), \alpha_U(s))$ will be an interval for nonnegative definiteness of $\mathbf{D}(s)$.

3. If for some sample $s \in \mathcal{S}_n$ we have

$$\lambda_{min}(s) > 0, \quad t(s) > 0 \quad \text{and} \quad r(s) > 0$$

then from (6.11) we can choose $\alpha_U(s) = \infty$ and from (6.12) we can choose

$$\alpha_L(s) = \frac{-\lambda_{min}(s)}{r(s)}.$$

Then the interval $(\alpha_L(s), \alpha_U(s))$ will be an interval for nonnegative definiteness of $\mathbf{D}(s)$.

4. If for some sample $s \in \mathcal{S}_n$ we have

$$\lambda_{min}(s) < 0, \quad t(s) > 0 \quad \text{and} \quad r(s) > 0$$

then by using (6.11) we can choose

$$\alpha_L(s) = \frac{-\lambda_{min}(s)}{t(s)}$$

and $\alpha_U(s) = \infty$ to give an interval $(\alpha_L(s), \alpha_U(s))$ which is an interval for non-negative definiteness of $\mathbf{D}(s)$. Note that for this particular case the inequalities in (6.12) do not hold and so it doesn't apply here.

5. If for some sample $s \in \mathcal{S}_n$ we have

$$\lambda_{min}(s) < 0, \quad t(s) < 0 \quad \text{and} \quad r(s) < 0$$

then the inequalities in (6.11) do not hold but from (6.12) we can choose $\alpha_L(s) = -\infty$ and

$$\alpha_U(s) = \frac{-\lambda_{min}(s)}{r(s)}$$

to give an interval $(\alpha_L(s), \alpha_U(s))$ which is an interval for nonnegative definiteness of $\mathbf{D}(s)$.

6. If for some sample $s \in \mathcal{S}_n$ we have

$$\lambda_{min}(s) < 0, \quad t(s) < 0 \quad \text{and} \quad r(s) > 0$$

then (6.11) and (6.12) cannot be applied here to obtain an interval for nonnegative definiteness of $\mathbf{D}(s)$.

If any sample in $\mathcal{S}_n$ satisfies the conditions of case 6 then the method of obtaining a range of values for α such that v_α is nonnegative definite fails. Also if there is a sample for which case 4 holds and another sample for which case 5 holds then $\alpha_L < \alpha_U$ cannot be true and the intervals for nonnegative definiteness of $\mathbf{D}(s)$ for these samples do not overlap one another. This means that we cannot apply the method of obtaining a range of values for α such that v_α is nonnegative definite here either. But provided none of these situations occur we can obtain the interval (α_L, α_U) for which v_α is a nonnegative definite unbiased estimator for $\mathrm{Var}(\hat{T}_R, \mathrm{PPAS})$ for any choice of $\alpha \in (\alpha_L, \alpha_U)$.

Example 41 *Consider a population of size $N = 5$ and a sample of size $n = 3$ and the size variable*

$$\mathbf{x}^t = (3, 4, 5, 7, 10).$$

For the ratio estimator under probability proportional to the aggregate size of $\mathbf{x}$*, we calculated the values of $\lambda_{min}(s)$, $r(s)$, $t(s)$, $\alpha_L(s)$, $\alpha_U(s)$ and $p(s)$ in table 6.1.*

Table 6.1: Values for $\lambda_{min}(s)$, $r(s)$, $t(s)$, $\alpha_L(s)$, $\alpha_U(s)$ and $p(s)$ for example 41

s	$\lambda_{min}(s)$	$r(s)$	$t(s)$	$\alpha_L(s)$	$\alpha_U(s)$	$p(s)$
$\{1,2,3\}$	-2.1644	0.6366	0.5173	**4.1843**	∞	0.0690
$\{1,2,4\}$	5.2888	0.3994	0.1683	-13.2411	∞	0.0805
$\{1,2,5\}$	19.6267	0.1832	-0.1942	-107.1326	**101.0876**	0.0977
$\{1,3,4\}$	9.6952	0.2811	0.0787	-34.4950	∞	0.0862
$\{1,3,5\}$	25.2887	0.1037	-0.2288	-243.8640	110.5199	0.1034
$\{1,4,5\}$	38.1322	-0.0186	-0.2920	$-\infty$	130.5902	0.1149
$\{2,3,4\}$	19.6776	0.1700	0.0212	-115.7597	∞	0.0920
$\{2,3,5\}$	37.9657	0.0322	-0.2506	-1179.0590	151.5046	0.1092
$\{2,4,5\}$	57.3913	-0.0754	-0.3009	$-\infty$	190.7422	0.1207
$\{3,4,5\}$	79.3845	-0.1256	-0.3084	$-\infty$	257.3745	0.1264

Note that none of the samples in $\mathcal{S}_n$ satisfy condition 6 on p.261. There is only one sample, with the smallest probability of selection, that gives a matrix $\mathbf{A}(s)$ *with a negative eigenvalue* $\lambda_{min}(s)$*. The value of* α_L *is equal to 4.1843 and the values of* α_U *is equal to 101.0876. Since* $\alpha_L < \alpha_U$ *we have the range of values* $(4.1843, 101.0876)$ *for* α *such that* v_α *is a nonnegative definite unbiased estimator for* $\mathrm{Var}(\hat{T}_R, \mathrm{PPAS})$.

From (6.6) we showed that the value of v_{RV} is more likely to be negative when $p(s)$ is small for some $s \in \mathcal{S}_n$. This could also suggest that $\lambda_{min}(s)$ is likely to be negative when $p(s)$ is small. Also the terms

$$\frac{1}{M_2 p(s)} - \frac{1}{\pi_{ij}}$$

for all $i, j \in s$ $(i \neq j)$ are likely to be positive when $p(s)$ is small implying that the term

$$\frac{n-1}{M_2 p(s)} - \sum_{\substack{k \in s \\ k \neq i}} \frac{1}{\pi_{ij}}$$

will be positive for each $i \in s$. Hence the symmetric matrix $\mathbf{P}_1(s) - \mathbf{P}_2(s)$, whose ij^{th} elements are equal to

$$(\mathbf{P}_1(s) - \mathbf{P}_2(s))_{ij} = \begin{cases} \frac{1}{N}\left[\frac{(n-1)}{M_2 p(s)} - \sum\limits_{\substack{k \in s \\ k \neq i}} \frac{1}{\pi_{ik}}\right] & \text{for} \quad i = j \in s \\ -\frac{1}{N}\left[\frac{1}{M_2 p(s)} - \frac{1}{\pi_{ij}}\right] & \text{for} \quad i \neq j \in s, \end{cases}$$

is likely to be nonnegative definite when $p(s)$ is small. The nonnegative definiteness of this matrix can easily be seen by writing $\mathbf{z}_s^t(\mathbf{P}_1(s) - \mathbf{P}_2(s))\mathbf{z}_s$ as in (6.2) and observing that the off-diagonal terms of $\mathbf{P}_1(s) - \mathbf{P}_2(s)$ are all negative which is a sufficient condition for this quadratic form to be nonnegative definite. If $\mathbf{P}_1(s) - \mathbf{P}_2(s)$ is nonnegative definite then the value of $r(s)$ cannot be negative.

Hence the value of $\lambda_{min}(s)$ is more likely to be negative when $p(s)$ is small in which case the matrix $\mathbf{P}_1(s) - \mathbf{P}_2(s)$ is more likely to be nonnegative definite so that

$r(s)$ is also nonnegative. This implies that the conditions on $\lambda_{min}(s)$ and $r(s)$ in case 5 might not always apply to $\hat{T}_R$ under a probability proportional to aggregate size design as demonstrated in example 41.

It is important to note that the unbiasedness of v_α depends on $p(s)$ being greater than zero for all samples in $\mathcal{S}_n$ and for the method described in this section to obtain the range (α_L, α_U) for nonnegative definiteness, we require the matrices $\mathbf{A}(s)$ and $\mathbf{P}_2(s)$ to be of rank $n-1$ for all $s \in \mathcal{S}_n$.

6.3 Efficiency of v_α

The choice of α that minimizes $\text{Var}(v_\alpha)$ would be desirable as the corresponding estimator will be the most efficient in its class over all values of α. Observe that

$$\begin{aligned}
\text{Var}(v_\alpha) &= \sum_{s\in\mathcal{S}_n} p(s)[(v_\alpha - \text{Var}(\hat{T}_R, \text{PPAS}))^2] \\
&= \sum_{s\in\mathcal{S}_n} p(s)[v_\alpha^2 - 2v_\alpha \text{Var}(\hat{T}_R, \text{PPAS}) + \text{Var}(\hat{T}_R, \text{PPAS})^2] \\
&= \sum_{s\in\mathcal{S}_n} p(s)[(\alpha(m_1 - m_2) + v_{RV})^2 - 2(\alpha(m_1 - m_2) + v_{RV})\text{Var}(\hat{T}_R, \text{PPAS}) \\
&\quad +\text{Var}(\hat{T}_R, \text{PPAS})^2]
\end{aligned}$$

where m_1 and m_2 are defined as in lemma 17. By differentiating with respect to α we have

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\text{Var}(v_\alpha) &= \sum_{s\in\mathcal{S}_n} p(s)[2\alpha(m_1 - m_2)^2 + 2(m_1 - m_2)v_{RV} \\
&\quad -2\alpha(m_1 - m_2)\text{Var}(\hat{T}_R, \text{PPAS})] \\
&= 2\alpha\text{E}[(m_1 - m_2)^2] + 2\text{E}[(m_1 - m_2)v_{RV}],
\end{aligned}$$

equating this to zero and solving for α we get

$$\alpha = \alpha_E = \frac{-\text{E}[(m_1 - m_2)v_{RV}]}{\text{E}[(m_1 - m_2)^2]}.$$

Taking second derivatives with respect to α gives

$$\frac{\partial^2}{\partial\alpha^2}\text{Var}(v_\alpha) = 2\text{E}[(m_1 - m_2)^2] \geq 0.$$

Hence $\text{Var}(v_\alpha)$ is a convex function in α which implies that $\text{Var}(v_\alpha)$ is at its minimum when $\alpha = \alpha_E$, i.e. v_{α_E} is the most efficient estimator in the class of v_α over all values of α.

Although in practice α_E is unknown since it depends on knowing the values of all the y's in the population, an estimate may be available. Moreover even a crude estimate of α_E is likely to perform better (in terms of efficiency) than when $\alpha = 0$ as $\text{Var}(v_\alpha)$ is convex in α. However, the value of α_E does not necessarily lie within the interval (α_L, α_U) which is an undesirable feature. We will use the median, lower and upper quartile, and the end points of the interval (α_L, α_U) for values of α in an empirical study in the next section.

6.4 Empirical studies

In this section we empirically investigate the performances of v_{RV} and v_α for different values of α by randomly generating population auxiliary and survey vectors. We will use the lower end, lower quartile, median, upper quartile, and upper end of the interval (α_L, α_U), if it exists, as values for α denoting them by α_L, α_{LQ}, α_M, α_{UQ} and α_U respectively, and we will also use v_{α_E} in our study.

For each population the following performance measures will be calculated:

1. The probability of obtaining a negative estimate, $\Pr(v_\alpha < 0)$;
2. The probability of obtaining an estimate which is below the lower bound on $\text{Var}(\hat{T}_R, \text{PPAS})$ given by theorem 17, on p.266, which is based on the sampled

units, $\Pr(v_\alpha < LB_s)$;

3. The coverage probability of each estimator based on a 95% confidence interval using the t-distribution with $n-1$ degrees of freedom,

$$(\hat{T}_R - t_{.975;n-1}\sqrt{v_\alpha}, \hat{T}_R + t_{.975;n-1}\sqrt{v_\alpha});$$

4. The relative efficiency of v_α over v_{RV}, $\text{RE}_\alpha = \text{MSE}(v_\alpha)/\text{MSE}(v_{RV})$.

The formal derivation of the asymptotic normality of the ratio estimator under probability proportional to aggregate size was given by Scott & Wu (1981). However we will be using a small sample size and so the normal assumption may not be valid. We will look at the coverage given by the t-distribution with $n-1$ degrees of freedom to see how accurate the coverage can be. Although there is no theoretical justification for using the t-distribution with $n-1$ degrees of freedom here, in general it may be used for small sample sizes as a pragmatic rule since the t-distribution tends to give more accurate coverage than the normal, see Royall & Cumberland (1985) and Thompson (2002, p.69).

The following result, from Gabler (1990, p.113), can be used to obtain a lower bound for a nonnegative definite quadratic form in $\mathbf{y}$, based on the observed sample.

Theorem 17 *If* $\mathbf{y}^t = (\mathbf{y}_1^t, \mathbf{y}_2^t)$ *and the* $N \times N$ *nonnegative definite symmetric matrix* $\mathbf{A}$ *is correspondingly partitioned as*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

then provided $\mathbf{A}_{22}$ *is nonsingular we have*

$$\mathbf{y}_1^t(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})\mathbf{y}_1 \leq \mathbf{y}^t\mathbf{A}\mathbf{y}. \tag{6.13}$$

By letting $\mathbf{y}_1$ be the sampled values of $\mathbf{y}$, we can use this result to see whether an estimator for the quadratic form $\mathbf{y}^t\mathbf{A}\mathbf{y}$ gives an estimated value which is less than the lower bound given by (6.13).

Lemma 18 *Let* $\mathbf{A}$ *be defined as in theorem 17. Suppose* $\mathbf{A}$ *is only orthogonal to vectors that are proportional to* $\mathbf{x}$ *with* $x_i > 0$ *for all* $i \in \mathcal{U}$*. Then the matrix* $\mathbf{A}_{22}$ *will always be nonsingular.*

Proof Let the $N \times 1$ vector $\mathbf{v}^t = (\mathbf{v}_1^t, \mathbf{v}_2^t)$ be such that the entries of $\mathbf{v}_1$ are all equal to zero. Then it follows that $\mathbf{v}^t\mathbf{A}\mathbf{v} = \mathbf{v}_2^t\mathbf{A}_{22}\mathbf{v}_2$ and hence

$$\mathbf{v}^t\mathbf{A}\mathbf{v} = 0 \quad \Leftrightarrow \quad \mathbf{v}_2^t\mathbf{A}_{22}\mathbf{v}_2 = 0.$$

But $\mathbf{v}^t\mathbf{A}\mathbf{v} = 0$ only holds if $\mathbf{v} \propto \mathbf{x}$ which cannot be true since $x_i > 0$ for all $i \in \mathcal{U}$. Hence $\mathbf{A}_{22}$ must be of full rank if $\mathbf{A}$ is orthogonal only to vectors that are proportional to $\mathbf{x}$ and that completes the proof. □

Lemma 18 shows that theorem 17 can be applied to any nonnegative definite symmetric matrix $\mathbf{A}$ whose nullspace is spanned by a single vector. In general the nullspace of the mean square error matrix for the ratio estimator under probability proportional to aggregate size sampling will be spanned by the size vector, and so we can apply theorem 17 to this matrix.

For each population we will use theorem 17 to calculate the lower bound (6.13) for $\text{Var}(\hat{T}_R, \text{PPAS})$ for each sample in $\mathcal{S}_n$ and then check to see if the corresponding estimates, using the same sample, is less than or greater than this lower bound. Note that this lower bound is known after the sample is selected and not before. Therefore it is reasonable to require that any sample estimate of $\text{Var}(\hat{T}_R, \text{PPAS})$ should not be smaller than the bound given in (6.13). It is clearly undesirable to use an estimator that gives values that are less than the lower bound given in (6.13) as this could indicate a poor estimate. Therefore the estimator with the least probability of obtaining a sample who's estimate for $\mathrm{Var}(\hat{T}_R, \mathrm{PPAS})$ is less than the lower bound in (6.13) would be more preferable than others.

We generated six auxiliary vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_6$ each of size 25 where the values of $\mathbf{x}_l$ were independently and randomly generated from $\mathrm{N}(100, 10l)$ for $l = 1, \ldots, 6$. The value for each entry of $\mathbf{x}_l$, for $l = 1, \ldots, 6$, were all greater than zero which implies that for any of these auxiliary vectors $p(s) > 0$ for all $s \in \mathcal{S}_n$ and hence v_α is unbiased for $\mathrm{Var}(\hat{T}_R, \mathrm{PPAS})$.

Table 6.2 gives the coefficient of variation of the auxiliary variables and the interval values of (α_L, α_U) for the nonnegative definiteness of our estimators.

Table 6.2: Coefficient of variation of $\mathbf{x}_l$, $l = 1, \ldots, 6$, and intervals (α_L, α_U) for nonnegative definiteness of v_α

Pop.	N	n	cv($\mathbf{x}$)	(α_L, α_U)
$\mathbf{x}_1$	25	4	0.10	(-208776,560232)
$\mathbf{x}_2$	25	4	0.23	(-7105,207031)
$\mathbf{x}_3$	25	4	0.35	(2940,100313)
$\mathbf{x}_4$	25	4	0.44	(5076,112510)
$\mathbf{x}_5$	25	4	0.50	–
$\mathbf{x}_6$	25	4	0.56	(15526,19279)

We see, from table 6.2, that as $\mathrm{cv}(\mathbf{x})$ increases the estimator v_{RV} (v_α when $\alpha = 0$) falls out of our class of nonnegative definite unbiased estimators. It is clear that estimators in the class v_α, with $\alpha \in (\alpha_L, \alpha_U)$, are less restricted than v_{RV} as they can be nonnegative definite for values of $\mathrm{cv}(\mathbf{x})$ up to 0.44 whereas v_{RV} can lose its nonnegative definiteness when $\mathrm{cv}(\mathbf{x})$ is only 0.35.

Because the size vector $\mathbf{x}_5$ doesn't give an interval (α_L, α_U) for nonnegative definiteness of v_α we will only focus our attention on the size vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$ and $\mathbf{x}_6$ in our study.

For each of the auxiliary vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$ and $\mathbf{x}_6$ we generated five survey variables as follows: $\mathbf{y}_{lj} = \mathbf{x}_l \times \mathbf{w}_{lj}/100$ (i.e. $\mathbf{y}_{lj}$ is an entry by entry product of $\mathbf{x}_l$ and $\mathbf{w}_{lj}/100$) for $l = 1$, 2, 3, 4 and 6 where the values of $\mathbf{w}_{lj}$ are independently and randomly generated from $\mathrm{N}(100, q_j)$ with $q_1 = 5$, $q_2 = 10$, $q_3 = 15$, $q_4 = 20$ and $q_5 = 30$.

In table 6.3 we give the values of the correlation coefficient between the x's and y's, the coefficient of variation of $\mathbf{z}$ where $z_i = y_i/x_i$ for all $i \in \mathcal{U}$, and the values of α_E.

Table 6.3: Values for $\rho(\mathbf{x}, \mathbf{y})$, $\text{cv}(\mathbf{z})$ and α_E

Pop.	$\rho(\mathbf{x}, \mathbf{y})$	$\text{cv}(\mathbf{z})$	α_E
$(\mathbf{x}_1, \mathbf{y}_{11})$	0.9229	0.0436	432612.6
$(\mathbf{x}_1, \mathbf{y}_{12})$	0.7348	0.0855	343075
$(\mathbf{x}_1, \mathbf{y}_{13})$	0.6561	0.1243	-61024.96
$(\mathbf{x}_1, \mathbf{y}_{14})$	0.4114	0.1815	219277.4
$(\mathbf{x}_1, \mathbf{y}_{15})$	0.1056	0.3464	189875.7
$(\mathbf{x}_2, \mathbf{y}_{21})$	0.9711	0.0495	300988.2
$(\mathbf{x}_2, \mathbf{y}_{22})$	0.9287	0.1051	126355.2
$(\mathbf{x}_2, \mathbf{y}_{23})$	0.8724	0.1228	296204.7
$(\mathbf{x}_2, \mathbf{y}_{24})$	0.6632	0.2304	366668.4
$(\mathbf{x}_2, \mathbf{y}_{25})$	0.6495	0.2870	290210.7
$(\mathbf{x}_3, \mathbf{y}_{31})$	0.9954	0.0427	48082.64
$(\mathbf{x}_3, \mathbf{y}_{32})$	0.9583	0.0980	201313.5
$(\mathbf{x}_3, \mathbf{y}_{33})$	0.9133	0.1757	147871.7
$(\mathbf{x}_3, \mathbf{y}_{34})$	0.8800	0.1741	124145.5
$(\mathbf{x}_3, \mathbf{y}_{35})$	0.7986	0.3313	50293.89
$(\mathbf{x}_4, \mathbf{y}_{41})$	0.9965	0.0362	175299.8
$(\mathbf{x}_4, \mathbf{y}_{42})$	0.9591	0.1175	179348.6
$(\mathbf{x}_4, \mathbf{y}_{43})$	0.9760	0.1204	138513.2
$(\mathbf{x}_4, \mathbf{y}_{44})$	0.9184	0.2170	98844.7
$(\mathbf{x}_4, \mathbf{y}_{45})$	0.7601	0.3744	133528.7
$(\mathbf{x}_6, \mathbf{y}_{61})$	0.9977	0.0418	77079.38
$(\mathbf{x}_6, \mathbf{y}_{62})$	0.9886	0.0777	171198
$(\mathbf{x}_6, \mathbf{y}_{63})$	0.9558	0.1525	140708.9
$(\mathbf{x}_6, \mathbf{y}_{64})$	0.9302	0.1750	172116.7
$(\mathbf{x}_6, \mathbf{y}_{65})$	0.9298	0.2590	65649.6

From table 6.3 we see that not all of the populations give values of α_E that are contained in the interval (α_L, α_U). For the size variable $\mathbf{x}_1$ which has $\text{cv}(\mathbf{x}_1) = 0.1$ the value of α_E for each survey vector $y_{11}, \ldots, y_{15}$ were contained inside the interval (α_L, α_U). For the size vector $\mathbf{x}_2$, with $\text{cv}(\mathbf{x}_2) = 0.23$, only one vector y_{22} gave a value for α_E inside the range (α_L, α_U). For $\mathbf{x}_3$, with $\text{cv}(\mathbf{x}_3) = 0.35$, there were two vectors $\mathbf{y}_{31}$ and $\mathbf{y}_{35}$ that gave values for α_E inside (α_L, α_U). For $\mathbf{x}_4$ with $\text{cv}(\mathbf{x}_4) = 0.44$ there was only one vector $\mathbf{y}_{44}$ such that $\alpha_E \in (\alpha_L, \alpha_U)$ and for $\mathbf{x}_6$, which has $\text{cv}(\mathbf{x}_6) = 0.56$, none of the survey vectors $\mathbf{y}_{61}, \ldots, \mathbf{y}_{65}$ gave values of α_E inside the interval (α_L, α_U).

The numerical examples suggests that the value of α_E for populations with small $\text{cv}(\mathbf{x})$ is more likely to lie inside the interval (α_L, α_U) than that for populations with large $\text{cv}(\mathbf{x})$.

For each auxiliary variable, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$ and $\mathbf{x}_6$, table 6.4 gives the probability of obtaining a negative estimate when applying v_{RV} and v_{α_E} to the survey variables.

Table 6.4: Probabilities of obtaining negative estimate from v_{RV} and v_{α_E}

Pop.	$\Pr(v_{RV} < 0)$	$\Pr(v_{\alpha_E} < 0)$
$(\mathbf{x}_1, \mathbf{y}_{11})$	0.00000	0.00000
$(\mathbf{x}_1, \mathbf{y}_{12})$	0.00000	0.00000
$(\mathbf{x}_1, \mathbf{y}_{13})$	0.00000	0.00000
$(\mathbf{x}_1, \mathbf{y}_{14})$	0.00000	0.00000
$(\mathbf{x}_1, \mathbf{y}_{15})$	0.00000	0.00000
$(\mathbf{x}_2, \mathbf{y}_{21})$	0.00000	0.00000
$(\mathbf{x}_2, \mathbf{y}_{22})$	0.00000	0.00000
$(\mathbf{x}_2, \mathbf{y}_{23})$	0.00000	0.00024
$(\mathbf{x}_2, \mathbf{y}_{24})$	0.00000	0.00016
$(\mathbf{x}_2, \mathbf{y}_{25})$	0.00000	0.00033
$(\mathbf{x}_3, \mathbf{y}_{31})$	0.00057	0.00000
$(\mathbf{x}_3, \mathbf{y}_{32})$	0.00062	0.00024
$(\mathbf{x}_3, \mathbf{y}_{33})$	0.00039	0.00088
$(\mathbf{x}_3, \mathbf{y}_{34})$	0.00070	0.00000
$(\mathbf{x}_3, \mathbf{y}_{35})$	0.00061	0.00000
$(\mathbf{x}_4, \mathbf{y}_{41})$	0.00004	0.00137
$(\mathbf{x}_4, \mathbf{y}_{42})$	0.00087	0.00451
$(\mathbf{x}_4, \mathbf{y}_{43})$	0.00014	0.00009
$(\mathbf{x}_4, \mathbf{y}_{44})$	0.00065	0.00000
$(\mathbf{x}_4, \mathbf{y}_{45})$	0.00046	0.00000
$(\mathbf{x}_6, \mathbf{y}_{61})$	0.02622	0.02028
$(\mathbf{x}_6, \mathbf{y}_{62})$	0.03324	0.05291
$(\mathbf{x}_6, \mathbf{y}_{63})$	0.01849	0.02875
$(\mathbf{x}_6, \mathbf{y}_{64})$	0.02496	0.03140
$(\mathbf{x}_6, \mathbf{y}_{65})$	0.02398	0.00873

Table 6.4 shows that although v_{α_E} is more efficient than v_{RV}, its probability of obtaining a negative estimate can be larger for some populations. For the size vector $\mathbf{x}_2$ the estimator v_{RV} performs just as well or better than v_{α_E} for all survey vectors. But for the size vector $\mathbf{x}_3$ the estimator v_{RV} does not perform as well as v_{α_E} in terms of obtaining nonnegative estimates for all populations apart from the survey vector $\mathbf{y}_{33}$.

The general conclusion from table 6.4 is that v_{α_E} and v_{RV} can perform as bad as each other in term of obtaining negative estimates. But the differences in the probabilities of obtaining negative estimates can be much larger when using a size vector with a large coefficient of variation compared to that of a size vector with small coefficient of variation.

Table 6.5 gives the probability $\Pr(v_\alpha < LB_s)$ of obtaining estimates that are less than the lower bound on $\text{Var}(\hat{T}_R, \text{PPAS})$ which is calculated, by theorem 17, using the sample values of the y's. The probability $\Pr(v_\alpha < LB_s)$ for each estimator, v_α with $\alpha = RV, \alpha_L, \alpha_{LQ}, \alpha_M, \alpha_{UQ}, \alpha_U$ and α_E, is calculated by

$$\Pr(v_\alpha < LB_s) = \sum_{s \in \mathcal{L}_\alpha} p(s)$$

where $\mathcal{L}_\alpha$ is the set of all samples that gives a value of v_α which is less than the lower bound on $\text{Var}(\hat{T}_R, \text{PPAS})$ given by (6.13).

Table 6.5: Values for $\Pr(v_\alpha < LB_s)$ for $\alpha = RV, \alpha_L, \alpha_{LQ}, \alpha_M, \alpha_{UQ}, \alpha_U$, and α_E

Pop.	v_{RV}	v_{α_L}	$v_{\alpha_{LQ}}$	v_{α_M}	$v_{\alpha_{UQ}}$	v_{α_U}	v_{α_E}
$(\mathbf{x}_1, \mathbf{y}_{11})$	0.00000	0.00007	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_1, \mathbf{y}_{12})$	0.00000	0.00007	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_1, \mathbf{y}_{13})$	0.00000	0.00007	0.00000	0.00000	0.00000	0.00017	0.00000
$(\mathbf{x}_1, \mathbf{y}_{14})$	0.00000	0.00007	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_1, \mathbf{y}_{15})$	0.00000	0.00007	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_2, \mathbf{y}_{21})$	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00016
$(\mathbf{x}_2, \mathbf{y}_{22})$	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_2, \mathbf{y}_{23})$	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00072
$(\mathbf{x}_2, \mathbf{y}_{24})$	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00016
$(\mathbf{x}_2, \mathbf{y}_{25})$	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00087
$(\mathbf{x}_3, \mathbf{y}_{31})$	0.00274	0.00098	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_3, \mathbf{y}_{32})$	0.00320	0.00100	0.00006	0.00006	0.00006	0.00006	0.00130
$(\mathbf{x}_3, \mathbf{y}_{33})$	0.00254	0.00075	0.00000	0.00000	0.00000	0.00008	0.00259
$(\mathbf{x}_3, \mathbf{y}_{34})$	0.00225	0.00116	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_3, \mathbf{y}_{35})$	0.00307	0.00130	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_4, \mathbf{y}_{41})$	0.00312	0.00012	0.00007	0.00007	0.00007	0.00007	0.00274
$(\mathbf{x}_4, \mathbf{y}_{42})$	0.00542	0.00130	0.00000	0.00000	0.00000	0.00018	0.00753
$(\mathbf{x}_4, \mathbf{y}_{43})$	0.00290	0.00018	0.00000	0.00000	0.00000	0.00000	0.00025
$(\mathbf{x}_4, \mathbf{y}_{44})$	0.00269	0.00080	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_4, \mathbf{y}_{45})$	0.00323	0.00050	0.00000	0.00000	0.00000	0.00000	0.00000
$(\mathbf{x}_6, \mathbf{y}_{61})$	0.03906	0.00033	0.00022	0.00011	0.00018	0.00018	0.02586
$(\mathbf{x}_6, \mathbf{y}_{62})$	0.04398	0.00154	0.00122	0.00671	0.00456	0.00471	0.06080
$(\mathbf{x}_6, \mathbf{y}_{63})$	0.02774	0.00000	0.00000	0.00000	0.00000	0.00006	0.03314
$(\mathbf{x}_6, \mathbf{y}_{64})$	0.03636	0.00654	0.00023	0.00018	0.00006	0.00006	0.03867
$(\mathbf{x}_6, \mathbf{y}_{65})$	0.03701	0.00034	0.00017	0.00006	0.00006	0.00006	0.01170

We see from table 6.5 that in general the value of $\Pr(v_\alpha < LB_s)$ seems to increase with the coefficient of variation of the auxiliary variables. The estimators v_{RV}, v_{α_L} and v_{α_E} performed poorly compared to v_{α_M} in all cases and the only differences between $v_{\alpha_{LQ}}$, v_{α_M} and $v_{\alpha_{UQ}}$ is for the size vector $\mathbf{x}_6$. But these differences are small so in general the estimators $v_{\alpha_{LQ}}$, v_{α_M} and $v_{\alpha_{UQ}}$ seem to perform the best.

Table 6.6 gives the coverage probabilities for each estimator which is based on the 95% confidence intervals using the t-distribution with $n-1$ degrees of freedom.

Table 6.6: Coverage probabilities for estimators v_{RV}, v_{α_L}, $v_{\alpha_{LQ}}$, v_{α_M}, $v_{\alpha_{UQ}}$, v_{α_U} and v_{α_E}

Pop.	v_{RV}	v_{α_L}	$v_{\alpha_{LQ}}$	v_{α_M}	$v_{\alpha_{UQ}}$	v_{α_U}	v_{α_E}
$(\mathbf{x}_1, \mathbf{y}_{11})$	0.9495	0.9445	0.9490	0.9501	0.9485	0.9442	0.9470
$(\mathbf{x}_1, \mathbf{y}_{12})$	0.9345	0.9327	0.9343	0.9353	0.9348	0.9329	0.9351
$(\mathbf{x}_1, \mathbf{y}_{13})$	0.9418	0.9393	0.9417	0.9434	0.9449	0.9454	0.9413
$(\mathbf{x}_1, \mathbf{y}_{14})$	0.9715	0.9691	0.9717	0.9716	0.9709	0.9696	0.9718
$(\mathbf{x}_1, \mathbf{y}_{15})$	0.9654	0.9623	0.9653	0.9665	0.9663	0.9657	0.9667
$(\mathbf{x}_2, \mathbf{y}_{21})$	0.9623	0.9615	0.9646	0.9675	0.9696	0.9703	0.9688
$(\mathbf{x}_2, \mathbf{y}_{22})$	0.9213	0.9207	0.9236	0.9243	0.9235	0.9234	0.9241
$(\mathbf{x}_2, \mathbf{y}_{23})$	0.9634	0.9631	0.9662	0.9675	0.9682	0.9676	0.9658
$(\mathbf{x}_2, \mathbf{y}_{24})$	0.9515	0.9512	0.9547	0.9560	0.9572	0.9583	0.9579
$(\mathbf{x}_2, \mathbf{y}_{25})$	0.9471	0.9466	0.9488	0.9483	0.9446	0.9416	0.9346
$(\mathbf{x}_3, \mathbf{y}_{31})$	0.9593	0.9614	0.9646	0.9658	0.9674	0.9666	0.9655
$(\mathbf{x}_3, \mathbf{y}_{32})$	0.8832	0.8844	0.8967	0.9059	0.9108	0.9151	0.9193
$(\mathbf{x}_3, \mathbf{y}_{33})$	0.9551	0.9566	0.9638	0.9678	0.9692	0.9698	0.9677
$(\mathbf{x}_3, \mathbf{y}_{34})$	0.9280	0.9294	0.9370	0.9402	0.9428	0.9441	0.9441
$(\mathbf{x}_3, \mathbf{y}_{35})$	0.9347	0.9367	0.9441	0.9458	0.9446	0.9477	0.9459
$(\mathbf{x}_4, \mathbf{y}_{41})$	0.8986	0.9001	0.9027	0.9050	0.9064	0.9091	0.9078
$(\mathbf{x}_4, \mathbf{y}_{42})$	0.8814	0.8839	0.8913	0.8982	0.9001	0.8989	0.8887
$(\mathbf{x}_4, \mathbf{y}_{43})$	0.9013	0.9022	0.9056	0.9102	0.9134	0.9153	0.9150
$(\mathbf{x}_4, \mathbf{y}_{44})$	0.9518	0.9544	0.9619	0.9675	0.9702	0.9718	0.9712
$(\mathbf{x}_4, \mathbf{y}_{45})$	0.8882	0.8916	0.8995	0.9022	0.9006	0.8984	0.8960
$(\mathbf{x}_6, \mathbf{y}_{61})$	0.8413	0.8665	0.8674	0.8676	0.8682	0.8693	0.8728
$(\mathbf{x}_6, \mathbf{y}_{62})$	0.8496	0.8744	0.8755	0.8766	0.8773	0.8780	0.8594
$(\mathbf{x}_6, \mathbf{y}_{63})$	0.8283	0.8471	0.8479	0.8484	0.8493	0.8496	0.8523
$(\mathbf{x}_6, \mathbf{y}_{64})$	0.8818	0.9102	0.9112	0.9123	0.9131	0.9135	0.9010
$(\mathbf{x}_6, \mathbf{y}_{65})$	0.8738	0.8941	0.8948	0.8961	0.8966	0.8969	0.9033

From table 6.6 we see that using the t-distribution does give accurate coverage probability close to 95% for all estimators for population with small cv($\mathbf{x}$). However this isn't true for those populations with large values of cv($\mathbf{x}$) and in those cases v_{RV} seems to perform the worse of all. This may be because the probability of obtaining a negative estimate with v_{RV} using $\mathbf{x}_6$ is large. However the probability of obtaining a negative estimate with v_{α_E} is also large and its coverage probabilities are more accurate than that of v_{RV} in all cases when using $\mathbf{x}_6$.

Table 6.7 gives the relative efficiency of v_α over v_{RV} which, as v_α and v_{RV} are unbiased for $\mathrm{Var}(\hat{T}_R, \mathrm{PPAS})$, is calculated as

$$\mathrm{RE}_\alpha = \mathrm{Var}(v_\alpha)/\mathrm{Var}(v_{RV})$$

for each value of $\alpha = \alpha_L$, α_{LQ}, α_M, α_{UQ}, α_U and α_E.

Table 6.7: Relative efficiency of v_α, for $\alpha = \alpha_L$, α_{LQ}, α_M, α_{UQ}, α_U and α_E, over v_{RV}

Pop.	RE_{α_L}	$\mathrm{RE}_{\alpha_{LQ}}$	RE_{α_M}	$\mathrm{RE}_{\alpha_{UQ}}$	RE_{α_U}	RE_{α_E}
$(\mathbf{x}_1, \mathbf{y}_{11})$	1.1680	1.0109	0.9092	0.8629	0.8719	0.8597
$(\mathbf{x}_1, \mathbf{y}_{12})$	1.0849	1.0053	0.9593	0.9468	0.9680	0.9465
$(\mathbf{x}_1, \mathbf{y}_{13})$	1.0217	0.9979	1.0626	1.2157	1.4573	0.9955
$(\mathbf{x}_1, \mathbf{y}_{14})$	1.0933	1.0052	0.9681	0.9821	1.0471	0.9668
$(\mathbf{x}_1, \mathbf{y}_{15})$	1.1062	1.0056	0.9690	0.9963	1.0874	0.9688
$(\mathbf{x}_2, \mathbf{y}_{21})$	1.0065	0.9613	0.9247	0.8967	0.8773	0.8641
$(\mathbf{x}_2, \mathbf{y}_{22})$	1.0099	0.9486	0.9180	0.9182	0.9492	0.9143
$(\mathbf{x}_2, \mathbf{y}_{23})$	1.0043	0.9745	0.9505	0.9322	0.9197	0.9117
$(\mathbf{x}_2, \mathbf{y}_{24})$	1.0068	0.9591	0.9187	0.8858	0.8601	0.8274
$(\mathbf{x}_2, \mathbf{y}_{25})$	1.0063	0.9623	0.9270	0.9004	0.8826	0.8720
$(\mathbf{x}_3, \mathbf{y}_{31})$	0.9857	0.9019	0.8800	0.9199	1.0217	0.8793
$(\mathbf{x}_3, \mathbf{y}_{32})$	0.9943	0.9506	0.9125	0.8802	0.8536	0.8044
$(\mathbf{x}_3, \mathbf{y}_{33})$	0.9948	0.9555	0.9234	0.8985	0.8808	0.8970
$(\mathbf{x}_3, \mathbf{y}_{34})$	0.9924	0.9363	0.8928	0.8618	0.8433	0.8373
$(\mathbf{x}_3, \mathbf{y}_{35})$	0.9856	0.9000	0.8736	0.9065	0.9986	0.8735
$(\mathbf{x}_4, \mathbf{y}_{41})$	0.9889	0.9360	0.8922	0.8574	0.8316	0.8068
$(\mathbf{x}_4, \mathbf{y}_{42})$	0.9909	0.9471	0.9106	0.8814	0.8595	0.8369
$(\mathbf{x}_4, \mathbf{y}_{43})$	0.9821	0.8988	0.8340	0.7880	0.7606	0.7518
$(\mathbf{x}_4, \mathbf{y}_{44})$	0.9824	0.9050	0.8534	0.8277	0.8280	0.8246
$(\mathbf{x}_4, \mathbf{y}_{45})$	0.9803	0.8886	0.8184	0.7696	0.7421	0.7356
$(\mathbf{x}_6, \mathbf{y}_{61})$	0.9236	0.9196	0.9156	0.9116	0.9077	0.7892
$(\mathbf{x}_6, \mathbf{y}_{62})$	0.9733	0.9717	0.9702	0.9687	0.9672	0.8456
$(\mathbf{x}_6, \mathbf{y}_{63})$	0.9460	0.9429	0.9399	0.9369	0.9339	0.7411
$(\mathbf{x}_6, \mathbf{y}_{64})$	0.9648	0.9628	0.9608	0.9588	0.9568	0.7959
$(\mathbf{x}_6, \mathbf{y}_{65})$	0.9111	0.9065	0.9020	0.8976	0.8932	0.7869

The data from table 6.7 show that the estimator v_{α_L} is not as efficient as v_{RV} when the coefficient of variation of the size variable is small. However when the coefficient of variation of the size variable is large, practically all the estimators used in this study are more efficient than v_{RV}. Since v_{α_E} is unknown in practice we could consider v_{α_U} for populations with large values (greater than 0.3) of $\text{cv}(\mathbf{x})$. For populations with smaller values of $\text{cv}(\mathbf{x})$ we could consider using v_{α_M}.

6.5 Conclusions

Our method for constructing a class of nonnegative definite unbiased variance estimators of $\text{Var}(\hat{T}_R, \text{PPAS})$ has proven to be useful as it demonstrates that it may be possible to obtain such estimators when the standard unbiased estimator fails to be nonnegative definite. For large population and sample sizes the construction of this class of estimators is computationally involved. But for small sample sizes, where variance approximations may be badly biased, our method is more practically applied.

We found that the variance of our estimator is a convex function in one of its parameters implying that more efficient estimators are available. However it isn't obvious how to estimate this unknown parameter that maximizes efficiency, but our empirical studies show that estimators in our class of nonnegative definite unbiased estimators perform well in terms of efficiency compared to the standard estimator.

The method of constructing our class of nonnegative definite unbiased variance estimators for $\text{Var}(\hat{T}_R, \text{PPAS})$ is independent of the y's. This means we can obtain this class of estimators at the design stage.

Conclusions

In this thesis we have shown that we can make inferences conditional on the observed sample size configuration when poststratifying under unequal probability sampling provided the poststratum membership is known for all units in the population. As in the case of simple random sampling, under an unequal probability design the conditional coverage probabilities given by confidence intervals based on the conditional variances are more accurate over those samples in $\mathcal{S}_{\mathbf{n}}$ compared to that given by the confidence intervals based on the unconditional variance. This confirms the view that inferences should be made conditionally on the observed sample size configuration.

Using matrix algebra we presented a result that gives an exact upper bound on the absolute bias ratio for a wide range of strategies which can be useful in assessing the bias. We also gave exact bounds on the relative efficiency of two different strategies that indicates the maximum gains in efficiency one strategy can have over the other.

In chapter 4 we gave theorems that can be applied to the mean square error matrices of two different strategies to give sufficient conditions for the superiority of one strategy over another. These methods can also be used to compare variance approximation formulae as well as pairs of strategies under poststratification conditionally, weak conditionally and unconditionally.

One of our results from chapter 4 gives a sufficient condition for a separate general linear strategy that is calibrated for auxiliary stratum vectors to be more efficient than a strategy that is calibrated for the overall population auxiliary vector.

In chapter 5 we applied the results of chapter 4 to compare strategies based on poststratification under the unconditional, weak conditional and conditional designs. Both equal and unequal probability sampling designs were considered.

For the special case of equal poststrata sizes, we derived a result that can be used to make unconditional comparisons of some estimators under simple random sampling. We also gave a result that can be used to make conditional comparisons between the stratified and expansion estimators under simple random sampling when there are a total of two poststrata.

For an unequal probability design we made comparisons of estimators with different probability weights under poststratification. This was done by computing their maximum relative efficiencies which gave us some indication of the appropriate probability weights to use.

In chapter 6 we showed that there can be cases where it is possible to construct a class of nonnegative definite unbiased estimators for the variance of the ratio estimator under probability proportional to aggregate size sampling. Some of the estimators in this class perform well in terms of efficiency compared to a standard estimator. For estimators considered in our empirical study, in section 6.4, the coverage probabilities given by the 95% confidence intervals based on the t-distribution with $n-1$ degrees of freedom are adequate for those size variables whose coefficient of variation is small.

We hope that the method of constructing a class of nonnegative definite unbiased variance estimators in chapter 6 can be extended to other strategies under unequal probability sampling. Also it will be interesting to see whether it is more likely that conditions 4 and 5, on p.260, hold simultaneously when the population or sample size becomes larger.

Bibliography

[1] Agarwal, S.K. and Kumar, P. (1998). On the relative efficiency of estimators of population total in unequal probability sampling when study variable has weak relationship with size variable. *Computational Statistics & Data Analysis*, **28**, 271-281.

[2] Aires, N. and Rosén, B. (2005). On inclusion probabilities and relative estimator bias for Pareto πps sampling. *Journal of Statistical Planning and Inference*, **128**, 543-567.

[3] Bapat, R.B, (2001). Existence of generalized inverses: Ten proofs and some remarks. *Resonance*, **6**, 19-28.

[4] Berger, Y. (1998). Variance estimation using list sequential scheme for unequal probability sampling. *Journal of Official Statistics*, **14**, 315-323.

[5] Berger, Y. (2005). Variance estimation with Chao's sampling scheme. *Journal of Statistical Planning and Inference*, **127**, 253-277.

[6] Bethlehem, J.G. and Keller, W.J. (1987). Linear weighting of sample survey data. *Journal of Official Statistics*, **3**, 141-153.

[7] Bethlehem, J.G. and Schuerhoff, M.H. (1984). Second-order inclusion probabilities in sequential sampling without replacement with unequal probabilities. *Biometrika*, **71**, 642-644.

[8] Brewer, K. (2002). *Combined Survey Sampling Inference, Weighing Basu's Elephants*. London: Arnold.

[9] Casady, R. and Valliant, R. (1993). Conditional properties of post-stratified estimators under normal theory. *Survey Methodology*, **19**, 183-192.

[10] Chao, M.T. (1982). A general purpose unequal probability sampling plan. *Biometrika*, **69**, 653-656.

[11] Cheng, C.S. and Li, K.C. (1983). A minimax approach to sample surveys. *The Annuals of Statistics*, **11**, 552-563.

[12] Cochran, W.G. (1977). *Sampling Techniques*, 3rd edition. New York: Wiley.

[13] Consiglio, L.D., Falorsi, P.D., Falorsi, S. and Russo, A. (2003). Conditional and unconditional analysis of some small area estimators in complex sampling. *Survey Methodology*, **29**, 53-61.

[14] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

[15] Deng, L.Y. and Chhikara, R.S. (1990). On the ratio and regression estimation in finite population sampling. *The American Statistician*, **44**, 282-284.

[16] Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87**, 376-382.

[17] Deville, J.C. and Tillé, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika*, **85**, 89-101.

[18] Doss, D.C., Hartley, H.O. and Somayajulu, G.R. (1979). An exact small sample theory for post-stratification. *Journal of Statistical Planning and Inference*, **3**, 235-247.

[19] Durbin, J. (1969). Inferential aspects of the randomness of sample size in survey sampling. In *New Developments in Survey Sampling*, ed. N.L. Johnson and H. Smith. New York: Wiley-Interscience.

[20] Fuller, W.A. (1966). Estimation employing post strata. *Journal of the American Statistical Association*, **61**, 1172-1183.

[21] Gabler, S. (1984). On unequal probability sampling: sufficient conditions for the superiority of sampling without replacement. *Biometrika*, **71**, 171-175.

[22] Gabler, S. (1990). *Minimax Solutions in Sampling from Finite Populations.* Berlin: Springer-Verlag.

[23] Gelman, A. and Carlin, J.B. (2001). Poststratification and Weighting Adjustments. p. 289-302 in Survey Nonresponse, edited by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little. New York: Wiley.

[24] Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55-76.

[25] Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society, Series B (Methodological)*, **17**, 269-278.

[26] Hájek, J. (1981). *Sampling from a Finite Population.* New York: Marcel Dekker.

[27] Hájek, J. (1971). Comments on a paper by D. Basu. In Godambe, V.P. and Sprott, D.A. editors. *An essay on the logical foundations of statistical inference.* Toronto: Holt, Rinehart and Winston, p.236.

[28] Hanif, M. and Brewer, K.R. (1980). Sampling with unequal probabilities without replacement: a review. *International Statistical Review*, **48**, 317-335.

[29] Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sampling Survey Methods and Theory, Vol I*. New York: Wiley.

[30] Hartley, H.O. and Ross, A. (1954). Unbiased ratio estimators. *Nature*, **174**, 270-271.

[31] Hidiroglou, M.A. and Srinath, K.P. (1981). Some estimators of a population total from simple random samples containing large units. *Journal of the American Statistical Association*, **76**, 690-695.

[32] Holt, D. and Smith, T.M.F. (1979). Post stratification. *Journal of the Royal Statistical Society: Series A (General)*, **142**, 33-46.

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

[34] Jagers, P., Odén, A. and Trulsson, L. (1985). Post-stratification and ratio estimation: usages of auxiliary information in survey sampling and opinion polls. *International Statistical Review*, **53**, 221-238.

[35] Kish, L. (1965). *Survey Sampling*. New York: Wiley.

[36] Knuth, D.E. (1981). *The Art of Computer Programming, Vol II*, 2nd edition. Reading, Massachusetts: Addison-Wesley.

[37] Little, R.J.A. (1993). Post-stratification: a modeler's perspective. *Journal of the American Statistical Association*, **88**, 1001-1012.

[38] Madow, W.G. (1949). On the theory of systematic sampling, II. *Annals of Mathematical Statistics*, **20**, 333-354.

[39] Meng, X.L. (1993). On the absolute bias ratio of ratio estimators. *Statistics & Probability Letters*, **18**, 345-348.

[40] Midzuno, H. (1952). On the sampling system with probability proportionate to sum of sizes. *Annals of the Institute of Statistical Mathematics*, **3**, 99-107.

[41] Montanari, G.E. (2000), Conditioning on auxiliary variable means in finite population inference. *Australian and New Zealand Journal of Statistics*, **42**, 407-421.

[42] Padmawar, V.R. (1998). On estimating nonnegative definite quadratic forms. *Metrika*, **48**, 231-244.

[43] Rao, C.R. (1967). *Linear Statistical Inference and its Applications*. New York: Wiley.

[44] Rao, C.R. and Mitra, S.K. (1971). *Generalized Inverse of Matrices and its Applications*. New York: Wiley.

[45] Rao, J.N.K. (1979). On deriving mean square errors and their non-negative unbiased estimators in finite population sampling. *Journal of the Indian Statistical Association*, **17**, 125-136.

[46] Rao, J.N.K. (1985). Conditional inferences in survey sampling. *Survey Methodology*, **11**, 15-31.

[47] Rao, J.N.K. (2002). Discussion of "Exact linear unbiased estimation in survey sampling". *Journal of Statistical Planning and Inference*, **102**, 39-40.

[48] Rao, J.N.K. and Vijayan, K. (1977). On estimating the variance in sampling with probability proportional to aggregate size. *Journal of the American Statistical Association*, **72**, 579-584.

[49] Robinson, J. (1987). Conditioning ratio estimates under simple random sampling. *Journal of the American Statistical Association*, **82**, 826-831.

[50] Rosén, B. (1972a). Asymptotic theory for successive sampling with varying probabilities without replacement, I. *The Annals of Mathematical Statistics*, **43**, 373-397.

[51] Rosén, B. (1972b). Asymptotic theory for successive sampling with varying probabilities without replacement, II. *The Annals of Mathematical Statistics*, **43**, 748-776.

[52] Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, **57**, 377-387.

[53] Royall, R.M. and Cumberland W.G. (1981a). An empirical study of the ratio estimator and estimator of its variance. *Journal of the American Statistical Association*, **76**, 66-77.

[54] Royall, R.M. and Cumberland W.G. (1981b). The finite-population linear regression estimator and estimators of its variance - an empirical study. *Journal of the American Statistical Association*, **76**, 924-930.

[55] Royall, R.M. and Cumberland W.G. (1985). Conditional coverage properties of finite population confidence intervals. *Journal of the American Statistical Association*, **80**, 355-359.

[56] Särndal, C.E. and Hidiroglou, M.A. (1989). Small domain estimation: conditional analysis. *Journal of the American Statistical Association*, **84**, 266-275.

[57] Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Spring-Verlag.

[58] Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, **5**, 119-127.

[59] Sengupta, S. (1989). On Chao's unequal probability sampling plan. *Biometrika*, **76**, 192-196.

[60] Smith, T.M.F. (1991). Post-stratification. *The Statistician*, **40**, 315-323.

[61] Sugden, R.A. (1993). Partial exchangeability and survey sampling inference. *Biometrika*, **80**, 451-455.

[62] Sugden, R.A. and Smith, T.M.F. (2002). Exact linear unbiased estimation in survey sampling. *Journal of Statistical Planning and Inference*, **102**, 25-38.

[63] Sugden, R.A. and Smith, T.M.F. (2006). Domains of study and poststratification. *Journal of Statistical Planning and Inference*, **136**, 3307-3317.

[64] Sugden, R.A. and Smith, T.M.F. (2007). Design-based properties of linear calibrated estimators of a finite population total. *International Statistical Review*, **75**, 218-223.

[65] Sugden, R.A., Smith, T.M.F. and Brown, R.P. (1996). Chao's list sequential scheme for unequal probability sampling. *Journal of Applied Statistics*, **23**, 413-421.

[66] Tillé, Y. (1996a). Some remarks on unequal probability sampling designs without replacement. *Annales D'économie et de statistique*, **44**, 177-189.

[67] Tillé, Y. (1996b). An elimination procedure for unequal probability sampling without replacement. *Biometrika*, **83**, 238-241.

[68] Tillé, Y. (1998). Estimation in surveys using conditional inclusion probabilities: simple random sampling. *International Statistical Review*, **66**, 303-322.

[69] Tillé, Y. (2006). *Sampling Algorithms*. New York: Springer.

[70] Thompson, S. K. (2002). *Sampling*, 2nd edition. New York: Wiley.

[71] Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference, a Prediction Approach*. New York: Wiley.

[72] Vijayan, K. (1977). On estimating the variance in unequal probability sampling. *Journal of the American Statistical Association*, **70**, 713-716.

[73] Yates, F. and Grundy, R.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society. Series B (Methodological)*, **15**, 253-261.

Glossary of Estimators

The following list of estimators, and where they are defined in the text, are used throughout the thesis.