Estimating the Variance of the Horvitz-Thompson Estimator

Tamie Henderson

A thesis submitted in partial fulfillment of the requirements
for the degree requirements of Bachelor of Commerce with
Honours in Statistics.

School of Finance and Applied Statistics
The Australian National University

October 2006

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge and belief, contains no material published or written by another person, except where due reference is made in the thesis:

.....

Tamie Henderson

27^{th} October 2006

Abstract

Unequal probability sampling was introduced by Hansen and Hurwitz (1943) as a means of reducing the mean squared errors of survey estimators. For simplicity they used sampling with replacement only. Horvitz and Thompson (1952) extended this methodology to sampling without replacement, however the knowledge of the joint inclusion probabilities of all pairs of sample units was required for the variance estimation process. The calculation of these probabilities is typically problematic.

Sen (1953) and Yates and Grundy (1953) independently suggested the use of a more efficient variance estimator to be used when the sample size was fixed, but this estimator again involved the calculation of the joint inclusion probabilities. This requirement has proved to be a substantial disincentive to its use.

More recently, efforts have been made to find useful approximations to this fixed-size sample variance, which would avoid the need to evaluate the joint inclusion probabilities. These approximate variance estimators have been shown to perform well under high entropy sampling designs, however, there is now an ongoing dispute in the literature regarding the preferred approximate estimator. This thesis examines in detail nine of these approximate estimators, and their empirical performances under two high entropy sampling designs, namely Conditional Poisson Sampling and Randomised Systematic Sampling.

These nine approximate estimators were separated into two families based on their variance formulae. It was hypothesised, due to the derivation of these variance estimators, that one family would perform better under Randomised Systematic Sampling and the other under Conditional Poisson Sampling. The two families of approximate variance estimators showed within group

similarities, and they usually performed better under their respective sampling designs.

Recently algorithms have been derived to efficiently determine the exact joint inclusion probabilities under Conditional Poisson Sampling. As a result, this study compared the Sen-Yates-Grundy variance estimator to the other approximate estimators to determine whether the knowledge of these probabilities could improve the estimation process. This estimator was found to avoid serious inaccuracies more consistently than the nine approximate estimators, but perhaps not to the extent that would justify its routine use, as it also produced estimates of variance with consistently higher mean squared errors than the approximate variance estimators.

The results of the more recent published papers, Matei and Tillé (2005), have been shown to be largely misleading. This study also shows that the relationship between the variance and the entropy of the sampling design is more complex than was originally supposed by Brewer and Donadio (2003). Finally, the search for a best all-round variance estimator has been somewhat inconclusive, but it has been possible to indicate which estimators are likely to perform well in certain definable circumstances.

Acknowledgments

I would like to take this opportunity to thank my wonderful supervisor Dr. Ken Brewer. It has been such a joy and privilege to learn from his vast survey sampling experience and knowledge. Throughout this year Ken's intuition always amazed me, and it has greatly helped in the development of this thesis. I would like to thank him for his constant encouragement, his meticulous review of this thesis, for refining my English, and also for the long discussions we had on the side.

A huge thank you to John Anakotta for his love, exceeding patience, and providing me with many enjoyable, and much needed, breaks from study. He has constantly encouraged me and supported me at times when I needed it most.

To Emily Brown and John Anakotta for their helpful feedback after reading over this thesis. To the staff at the ABS, in particular to John Preston, for their valuable thoughts on some interesting concepts that arose throughout this year. To the wonderful friends I have made with my fellow honours students. It has been such joy to share in this experience together.

I would also like to thank my amazing family, and in particular my parents for their love, prayers and support throughout my life.

Finally, I would like to thank God for giving me the strength to complete this year, teaching me to always trust in Him, and for providing me with the amazing support of the people I have already thanked.

Contents

List of Tables

1 Introduction

1.1 Outline of the Problem

1.1.1 Background

Thousands of surveys are conducted each year across many fields of studies such as marketing, agriculture, businesses, households and health. Information is vital in making decisions within these fields, and survey sampling provides an effective method of obtaining this information by analysing only a sample of units from a population. Since only a sample of units is being analysed, the information gathered is not exact, but it is important that it should be as exact as possible. This thesis focuses on the process of variance estimation which provides one measure of this exactness.

Departure from exactness is usually considered under two headings, bias and imprecision. The bias of a sample estimator is the difference between its expectation over all possible samples, and the actual value of the parameter that is being estimated. In this thesis the parameter being estimated is the sum over all population units of a particular characteristic, such as the incomes of taxpayers or the sales of retail stores. The estimator of total being used (the Horvitz-Thompson Estimator) is unbiased over all possible samples, so the bias of this estimator is not an issue. The imprecision of a sample estimator is measured by its variance, which is the expectation over all possible samples of the squared difference between the sample estimate and the population value being estimated.

The variance of an estimator is estimated from the sample by a variance estimator. To estimate the departure from exactness of a variance estimator itself, there is often the possibility that this estimator of variance may be biased. In that

case it is necessary to consider both the bias and the variance of that variance estimator. The mean squared error is an overall measure of the variance estimator's inaccuracy, and is the sum of its variance and its squared bias. The theory of variance estimation has developed greatly over recent decades to simplify the variance estimation process and improve its accuracy. The topic of this thesis, therefore, can appropriately be described as the estimation of the variance of the unbiased (Horvitz-Thompson) estimator of total.

The algorithm chosen to select a sample from the population can greatly improve the accuracy of an estimator. Unequal probability sampling was introduced by Hansen and Hurwitz (1943) as this procedure can provide more precise estimates than is possible when sample units are included with equal probabilities. Hansen and Hurwitz derived an unbiased estimator for sampling algorithms that used unequal probability sampling with replacement. Horvitz and Thompson (1952) extended this research by deriving an unbiased estimator for sampling algorithms that used unequal probability sampling without replacement. This estimator is commonly referred to as the Horvitz-Thompson estimator. The same authors also derived the variance of this estimator, and an estimator of its variance. This variance estimator was applicable for without replacement unequal probability sampling algorithms, but it was severely inefficient for algorithms which provided samples of a fixed size. Sen (1953) and Yates and Grundy (1953) independently derived a more efficient variance estimator for the Horvitz-Thompson estimator for fixed-size sampling algorithms.

To calculate the Horvitz-Thompson and the Sen-Yates-Grundy variance estimators, knowledge of the joint inclusion probabilities is required for all possible pairs of the units sampled. That is, the probability that any pair of units is included in the sample. These probabilities are usually problematic to calculate for complex sampling algorithms, such as unequal probability sampling. It is also particularly

difficult to devise sample algorithms that are simultaneously easy to implement, produce efficient estimates of variance, and for which the joint inclusion probabilities are easy to evaluate.

Two approaches can be implemented to overcome the difficulty in calculating the joint inclusion probabilities. The first approach is to use the approximate joint inclusion probabilities derived by Hartley and Rao (1962), Asok and Sukhatme (1976) or Hájek (1964) directly in the Sen-Yates-Grundy variance estimator. The second approach is to use one of the numerous approximate variance estimators that are independent of the joint inclusion probabilities. These approximate variance estimators have been shown to provide sufficiently accurate estimates of the true variance under high entropy fixed-size sampling algorithms (Brewer and Donadio (2003) and Donadio (2002)), where entropy is a measure of the "randomness" of a sampling algorithm.

Chen, Dempster, and Liu (1994), Aires (1999) and Deville (2000) developed algorithms to determine the exact joint inclusion probabilities for the complex fixed-size sampling algorithm, Conditional Poisson Sampling. It is therefore of interest now, to determine whether the Sen-Yates-Grundy variance estimator is more efficient than the approximate variance estimators. Extensive research has not been conducted in this developing area, especially with regard to comparing the performance of these approximate variance estimators.

Matei and Tillé (2005) provided an extensive study comparing twenty different variance estimators under Conditional Poisson Sampling. They compared many approximate variance estimators and also the Sen-Yates-Grundy variance estimator with the exact joint inclusion probabilities. Their results indicated that the approximate variance estimators have, for the most part, similar properties to the

Sen-Yates-Grundy variance estimator. However, some of their results regarding the behaviour of certain approximate variance estimators were inconsistent with other empirical results produced by Brewer and Donadio (2003).

1.1.2 Aims and intentions

The first aim of this study is to resolve the discrepancy between the results produced by Brewer and Donadio (2003) and by Matei and Tillé (2005). The second, and main objective, is to determine whether there is one approximate variance estimator that consistently produces accurate estimates, by comparing the behaviour of nine different approximate variance estimators. The final objective of this study is analyse whether knowing the exact joint inclusion probabilities under Conditional Poisson Sampling can significantly improve the variance estimation process by using the Sen-Yates-Grundy variance estimator.

In addition to comparing the nine approximate variance estimators individually they will also be compared as two groups. The nine estimators are divided into two groups based on similarities between their variance formulae: the Brewer Family and the Hájek-Deville Family. The Brewer Family estimators are related to an approximation of the joint inclusion probabilities realised under Randomised Systematic Sampling, while the Hájek-Deville Family estimators are based on approximations to the joint inclusion probabilities realised under Conditional Poisson Sampling. It is hypothesised that these two families will perform better under their corresponding sampling algorithm. To date this study is the only research conducted to determine whether variance estimators are more accurate under certain sampling algorithms.

This study uses simulations to generate variance estimates as the properties of

these estimators cannot simply be determined algebraically. The simulations are first conducted for the same populations and sampling algorithm as those used by Brewer and Donadio (2003) and by Matei and Tillé (2005). The properties of the variance estimators under these populations are directly compared with the properties published in the two papers to explain the discrepancy between their results. A further five real populations are studied to compare the behaviour of the variance estimators both individually, and within their respective families.

During the process of this simulation study a further two objectives were established. The first of these objectives is to mathematically analyse the relationship between some of the approximate estimators. The second is to analyse the relationship between the entropy and variance of a sampling design. It is assumed that, except in some unusual and easily recognisable circumstances, an increase in the entropy should also increase the variance. Neither of these concepts have been discussed in the literature before.

1.1.3 Thesis outline

The remainder of chapter 1 details the notation and common statistical terms used within this thesis. Chapter 2 describes two unequal probability sampling algorithms, namely Randomised Systematic Sampling and Conditional Poisson Sampling. The Brewer Family and Hájek-Deville Family of approximate variance estimators are also defined. In addition, chapter 2 describes the major discrepancy between the results of Brewer and Donadio (2003) and of Matei and Tillé (2005).

Chapter 3 describes the methodology and results of the major simulation study. The discrepancies between the results of Brewer and Donadio and of Matei and Tillé is resolved before comparing the variance estimators. Chapter 4 discusses the two

additional discoveries made during the simulation study, and chapter 5 provides a summary of the major findings and explains the contribution of these results to the study of variance estimation.

1.2 Notation

This section is used as a reference for the notation and general statistical terms used throughout this thesis.

Consider a *finite population*, U, containing N distinct and measurable units, where the i^{th} unit is represented by the label i, such that $U = \{1, 2, \ldots, i, \ldots, N\}$. It is assumed that the *population size*, N, is known. Let y denote a variable of the population, where y_i represents the value of y for the i^{th} unit, and $Y = \{y_1, \ldots, y_N\}$. For example, if U is the population of taxpayers in Australia, and y represents the income, then y_i is the income of the i^{th} taxpayer. It is assumed that y_i is unknown for $i \in U$ prior to sampling.

A finite *population parameter* is an unknown characteristic of the population. For example, the total of y for all units in the population, denoted by $Y_\bullet$,

$$Y_\bullet = \sum_{i \in U} y_i,$$

or the average of y across the population, denoted $\bar{Y}_\bullet$,

$$\bar{Y}_\bullet = \frac{Y_\bullet}{N} = \frac{1}{N} \sum_{i \in U} y_i,$$

where $\sum_{i \in U}$ is the summation over all units within the population and y is known as the *study variable*. A population parameter must be estimated as only the sampled units of y are known. Throughout this thesis, only the population total is considered for estimation because the analysis for the average is virtually the same since N is

known.

A *sample*, s, is a subset of units of the population U, in which y_i is known for all $i \in s$. The set of possible samples, $\mathcal{S}$, has 2^N distinct elements. The *sample size*, $n(s)$, is the number of units included in the sample, s. The objective of survey sampling is to provide precise and accurate estimates of the population parameters based only on the units sampled. This is achieved in two stages - the design stage and the estimation stage. The *design stage* describes how the sample is selected, and the *estimation stage* describes how the parameters are estimated from that sample selected. First consider the design stage. The function $p(\cdot)$ is known as the *sampling design*, where $p(s)$ is the probability that the sample, s, is selected from the population. The properties of the sampling design are,

$$(i) \quad p(s) \geq 0 \tag{1.1}$$

$$(ii) \quad \sum_{s \subset \mathcal{S}} p(s) = 1. \tag{1.2}$$

The *sampling algorithm* is the process in which the samples are selected to produce this sampling design. There are many different sampling algorithms such as simple random sampling without replacement (*srswor*) and Poisson sampling. A sampling algorithm with replacement can include the same unit more than once in a sample. In contrast, in a sampling algorithm without replacement units cab only appear once in a sample. It is possible for different sampling algorithms to result in the same sampling design. The sampling design can be represented as a mathematical formula for some sampling algorithms.

It should be noted that some statistical literature uses the term "sampling design" to represent the sampling algorithm, whilst other literature uses this expression to represent both the sampling algorithm and the method of estimation

combined. To avoid confusion, throughout this thesis, the term "sampling design" is represented only by the function $p(\cdot)$, and the "sampling algorithm" describes how the sampling process is implemented.

Entropy is a measure of spread of the sampling design $p(\cdot)$, and is computed by

$$e = -\sum_{s \in \mathcal{S}} p(s) ln(p(s)). \tag{1.3}$$

A sampling algorithm with a high entropy sampling design is an algorithm where there is a high amount of uncertainty or "randomness" in the samples which will be selected.

The *support*, Ω, of a sampling algorithm is the set of possible samples for which properties (1.1) and (1.2) are satisfied for $s \in \Omega$. A fixed-size sampling algorithm only selects samples of a given fixed size, say n. Therefore, the support is the set of samples of size n, $\Omega = \mathcal{S}_n$. For without replacement fixed-size sampling algorithms there are $\frac{N!}{n!(N-n)!}$ samples in the support.

A sample could be selected simply by randomly selecting one of the possible samples with given probabilities $p(s)$. However, for large populations the number of possible samples makes this approach infeasible. As a result, inclusion probabilities are assigned to each unit and are used to select a sample. An indicator variable, δ_i, takes the value of one if the i^{th} units is included in the sample and zero otherwise.

The *first order inclusion probability*, π_i, is the probability that the i^{th} unit is included in the sample, that is

$$\pi_i = P(i \in s) = \sum_{s \ni i} p(s),$$

where $s \ni i$ is the sum over all samples including unit i. The first order inclusion probabilities are often simply referred to as the inclusion probabilities. If all the

inclusion probabilities are known and greater than zero, implying each population unit has some probability of being selected, then the sample is known as a probability sample. The *second order inclusion probability*, or the *joint inclusion probability*, π_{ij}, is the probability that the i^{th} and j^{th} units are both included in the sample,

$$\pi_{ij} = P(i \in s, j \in s) = \sum_{s \ni i,j} p(s)$$

It is clear that $\pi_{ij} = \pi_{ji}$ as each unit is selected independently.

Example 1.1 combines these notations together by considering *srswor* where the support is $\mathcal{S}_n$.

Example 1.1. *For* srswor *the sampling design is,*

$$p(s) = \begin{cases} \binom{N}{n}, & \text{if} \quad s \in \mathcal{S}_n \\ 0, & \text{otherwise,} \end{cases}$$

and the first and second order inclusion probabilities are

$$\begin{aligned} \pi_i &= \frac{n}{N} \quad \forall i \in U \\ \pi_{ij} &= \frac{n(n-1)}{N(N-1)} \quad \text{for all } j \neq i. \end{aligned}$$

The quantity $n(s)$ is a random variable under some sampling algorithms, like Poisson Sampling. Only fixed-size sampling algorithms are considered in this thesis so the sample size is simply denoted by n. Under fixed-size sampling algorithms the following properties hold

$$\sum_{i \in U} \pi_i = n \tag{1.4}$$

$$\sum_{j(\neq i)=1}^{N} \pi_{ij} = (n-1)\pi_i \tag{1.5}$$

$$\sum_{i=1}^{N} \sum_{j>i} \pi_{ij} = \frac{n(n-1)}{2}. \tag{1.6}$$

The *estimation stage* is the second stage in survey sampling where an appropriate estimator is chosen. An *estimator* is a formula which estimates a population parameter based on the sampled units. An estimator of a particular parameter, θ, is denoted by adding a circumflex, $\hat{\theta}$, and an approximation of a particular parameter is denoted by adding a tilde. Therefore, $\tilde{\theta}$ is an approximation to the parameter θ, and $\hat{\tilde{\theta}}$, is an estimator of the approximation to the population parameter θ. Hence $\hat{Y}_\bullet$ is an estimator of the population total $Y_\bullet$.

The statistical properties of an estimator can be described by the sampling design as

$$E(\hat{\theta}) = \sum_{s \subset \mathcal{S}} p(s)\hat{\theta} \tag{1.7}$$

$$Var(\hat{\theta}) = \sum_{s \subset \mathcal{S}} p(s)(\hat{\theta} - \theta)^2. \tag{1.8}$$

where $E(\hat{\theta})$ and $V(\hat{\theta})$ is the expectation and variance of the estimates of all possible samples. The precision of an estimator is commonly determined by its variance. Two desirable properties of an estimator are the bias and the mean squared error (MSE). The bias, $B(\cdot)$ is a measure of how far the expected value of the estimator is from the true parameter

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta. \tag{1.9}$$

The MSE is a measure of the stability of an estimator and involves both the bias and variance of the estimator

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + [B(\hat{\theta})]^2. \tag{1.10}$$

The MSE and the bias are used throughout this study to compare the properties of different variance estimators.

1.3 Variance of the Horvitz-Thompson Estimator

Horvitz and Thompson (1952) showed that the only linear unbiased estimator for any without replacement sampling algorithm, where the inclusion probabilities are well defined as π_i for $i = 1, \ldots, N$, was

$$\hat{Y}_{\bullet HT} = \sum_{i \in s} \frac{y_i}{\pi_i}. \tag{1.11}$$

This is commonly referred to as the Horvitz-Thompson estimator (HTE). These authors also showed the variance of this estimator to be

$$V_{HTE}(\hat{Y}_{\bullet HT}) = \frac{1}{2} \sum_{i \in U} \sum_{\substack{j \in U \\ j \neq i}} (\pi_i \pi_j - \pi_{ij}) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j}, \tag{1.12}$$

with the corresponding variance estimator

$$\hat{V}_{HTE}(\hat{Y}_{\bullet HT}) = \frac{1}{2} \sum_{i \in U} \sum_{\substack{j \in U \\ j \neq i}} \frac{\delta_i \delta_j (\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \tag{1.13}$$

which is unbiased provided $\pi_{ij} > 0$ for all $i, j \in U$. If any of the joint inclusion probabilities are equal to zero, then this variance estimator will be negatively biased.

Sen (1953) and Yates and Grundy (1953), (SYG), independently demonstrated that the above estimator was inefficient for fixed-size sampling algorithms and has the undesirable property of producing negative values. They then both independently derived the following variance of the HTE for fixed sampling designs

$$V_{SYG}(\hat{Y}_{\bullet HT}) = \frac{1}{2} \sum_{i \in U} \sum_{\substack{j \in U \\ j \neq i}} (\pi_i \pi_j - \pi_{ij})(y_i \pi_i^{-1} - y_j \pi_j^{-1})^2, \tag{1.14}$$

with the corresponding variance estimator of

$$\hat{V}_{SYG}(\hat{Y}_{\bullet HT}) = \frac{1}{2} \sum_{i \in U} \sum_{\substack{j \in U \\ j \neq i}} \delta_i \delta_j (\pi_i \pi_j - \pi_{ij}) \pi_{ij}^{-1} (y_i \pi_i^{-1} - y_j \pi_j^{-1})^2. \tag{1.15}$$

This variance estimator is also unbiased provided $\pi_{ij} > 0$ for all $i, j \in U$, and is non-negative for any sampling algorithm where $\pi_{ij} < \pi_i \pi_j$ for all $j \neq i$.

2 Variance Estimation

2.1 Unequal Probability Sampling

The precision of the an estimator is greatly dependent upon the sampling algorithm. Until the 1940s, samples were generated by assigning each unit an equal probability of selection. This provided simple, however not necessarily efficient, estimators. Hansen and Hurwitz (1943) first suggested using unequal probability sampling, showing that this improved the estimation of the population total. Horvitz and Thompson (1952) extended this to unequal probability sampling without replacement by deriving the unbiased HTE (1.11) of the total of the population for these sampling algorithms.

It is easy to show that the variance of the HTE, equation (1.14), is zero when the inclusion probabilities are proportional to the study variable, that is $\pi_i \propto y_i$. A zero variance implies that there is no error at all in the HTE. To implement this design, however, knowledge of all the values of Y is required, which is not possible (or if it were there would be no need to draw a sample to estimate their total). In many sampling situations there is an auxiliary variable, $X = \{x_1, \ldots, x_N\}$, which is known for each unit in the population and is believed to be approximately proportional to the study variable, Y. Designing the inclusion probabilities proportional to this auxiliary variable ensures that they are also approximately proportional to the study variable; hence reducing the variance. Sampling algorithms which use these inclusion probabilities are known as unequal probability sampling algorithms because each unit has its own individual probability of being selected. The inclusion probabilities under fixed-size unequal probability sampling algorithms are

$$\pi_i = \frac{nx_i}{\sum_{j \in U} x_j} \tag{2.1}$$

which ensures that $\sum_{i \in U} \pi_i = n$. In this situation X is referred to as a measure of

size variable. These probabilities are referred to as the desired inclusion probabilities as they usually reduce the variance compared with any other set of inclusion probabilities, as X is approximately proportional to Y.

Inclusion probabilities can not exceed unity, so if there is a unit which has a large x_i value then it may be the case that $nx_i > \sum_{j \in U} x_j$, implying the impossibility of $\pi_i > 1$. In such a case, it is necessary to set this unit's inclusion probability to unity, and recalculate the remaining inclusion probabilities again using (2.1), with that unit excluded and with one fewer units to be included by this procedure; that is, reduce n by one. If necessary, this process is repeated until all units have an inclusion probability such that $0 < \pi_i \leq 1$. Units with an inclusion probability of unity are, by definition, included in the sample with certainty, and may be referred to as completely enumerated units. That is, these units will be included in every possible sample from the given population. The variance of an estimator is defined as the variability across all possible samples and since these units are included in every possible sample, they do not contribute to that variance.

Brewer and Hanif (1983) proposed many unequal probability sampling algorithms which produce the desired inclusion probabilities exactly, including Randomised Systematic Sampling. Conditional Poisson Sampling is another unequal probability sampling algorithm. Both these algorithms are used within this study and are explained in more detail later in this section. For some sampling algorithms it is not possible to produce the first and second order inclusion probabilities exactly. Knowing the first order inclusion probabilities exactly ensures an unbiased estimator of the total. If the joint inclusion probabilities are also known exactly and $\pi_{ij} > 0$ for all pairs of units in the population, then there is also an unbiased estimator of the variance.

One disadvantage of unequal probability sampling designs is that the joint inclusion probabilities are problematic to calculate (Särndal (1996) and Brewer (1999)). Hence it is difficult to calculate the variance of the HTE as both (1.13) and (1.15) require knowledge of these probabilities. Recently, however, algorithms have been developed to determine the exact joint inclusion probabilities under certain sampling algorithms such as for Conditional Poisson Sampling and Pareto πps sampling (Aires, 1999).

2.1.1 Randomised Systematic Sampling

Systematic Sampling is a sampling algorithm where the population is ordered before the units are systematically selected. As Systematic Sampling involves two stages, the ordering and the sampling, it represents a group of sampling algorithms depending on how the two stages are implemented. In equal probability systematic sampling the N units are listed and a skip interval k is chosen, such that N/k is as nearly as possible the desired sample size. A random start, r, is then selected between 1 and k, and the sample units are the r^{th} and every k^{th} thereafter.

There are two main options to order the population. The first option is to order the units in a meaningful order, such as in the order of size of some auxiliary variable. This ensures that the sample selected is a good representation of the population in terms of size, as both large and small units are included in the sample. However, this approach makes it impossible to estimate the variance unbiasedly as many of the joint inclusion probabilities are zero. The second main option is to list the population units in random order. This makes the selected sample virtually equivalent to one chosen with *srswor*. The variance under this approach will typically be higher than when the units are meaningfully ordered, but it is possible to estimate it almost unbiasedly.

Similar strategies can be applied in the context of systematic sampling with unequal probability sampling. The only difference is that the skip interval is defined in terms of the variable used to determine the inclusion probabilities. The first main option is to list the entire population in some meaningful order to ensure a highly representative sample will be selected. This is known as Ordered Systematic Sampling (OSYS), and is the most commonly known algorithm in this group. The variance of the HTE under this sampling algorithm will be smaller than under *srswor* if the population is well ordered, however once again it is difficult to unbiasedly estimate this variance as many of the joint inclusion probabilities can be zero.

The second main option, which overcomes this problem of joint inclusion probabilities being zero, is Randomised Systematic Sampling (RANSYS). This sampling algorithm, introduced by Goodman and Kish (1950), is implemented by randomly ordering the units in the population before systematically selecting the units. Algorithm 2.1 describes how to implement RANSYS and Example 2.1 provides an example of this sampling algorithm for a small population.

Algorithm 2.1. *Randomised Systematic Sampling*

(i) Assign each unit a probability of inclusion by (2.1).

(ii) Randomly order the population units and let $k = 1, 2, \ldots, N$ *denote the* k^{th} *unit in the randomly ordered population.*

(iii) Determine $W_k = \sum_{j=1}^{k} \pi_j$ *for each unit, where* $W_0 = 0$ *and* $W_N = n$.

(iv) Select a uniform random number u *from the interval* $(0, 1]$ *as a starting point.*

(v) Select each unit k *which satisfies*

$$W_{k-1} \leq u + i < W_k \quad for \quad i = 0, 1, \ldots, n-1.$$

Example 2.1. *Suppose a sample of 3 units is to be selected from the population* $U = \{1, 2, 3, 4, 5\}$ *using RANSYS. Column (1) in Table 2.1 shows the random order of these units with their corresponding inclusion probabilities in column (3). For the random starting position* $u = 0.58$ *the sampled obtained was* $s = 1, 2, 4$.

Units	k	π_i	W_k	Selections
4	1	0.73	0.73	$u = 0.58$
5	2	0.81	1.54	
1	3	0.24	1.78	$u = 1.58$
3	4	0.65	2.43	
2	5	0.57	3.00	$u = 2.58$

Table 2.1: Sample selected using RANSYS

Systematically selecting units from a large randomly ordered population automatically guarantees a high entropy sampling algorithm. If a population consists of 10 names, then there is only a 1 in $10! = 3628800$ chance that the units will be ordered alphabetically (Brewer, 2002, p. 147). There is a large amount of uncertainty in which sample will be selected as there are a many possible random permutations of the population, therefore, this sampling algorithm has a high entropy. It is not possible to define a simple equation for the sampling design, $p(s)$, for RANSYS. In addition, the joint inclusion probabilities for RANSYS can only be calculated exactly for small populations. It is too difficult to determine all possible permutations and all possible samples from each permutation for a large population. Despite this disadvantage, RANSYS is commonly used in practice as it is very simple and fast to implement.

2.1.2 Conditional Poisson Sampling

Poisson Sampling (PO) is an unequal probability sampling design. It is implemented by assigning each unit in the population an inclusion probability, and conducting N independent Bernoulli trials using these probabilities to determine which units are included in the sample. For clarity, the inclusion probabilities for Poisson Sampling will be denoted by p_i. Since each Bernoulli trial is independent the sample size is random, and the joint inclusion probabilities are $p_{ij} = p_i p_j$. The sampling design, $p_{PO}(s)$, for this algorithm is

$$p_{PO}(s) = \prod_{k \in s} p_k \prod_{j \notin s} (1 - p_j) \tag{2.2}$$

where $s \in \mathcal{S}$, any possible subset of U.

Conditional Poisson Sampling (CPS) is Poisson Sampling conditioned on the sample being of a given size, say, n. That is, only samples such that $s \in \mathcal{S}_n$, where $\mathcal{S}_n$ is the set of samples of size n, are accepted and all other samples are rejected. Hájek (1964) introduced this sampling algorithm under the name of Rejective Sampling. The sampling design, $p_{CPS}(s)$, for this algorithm is

$$\begin{aligned} p_{CPS}(s) &= p_{PO}(s|s \in \mathcal{S}_n) \\ &= \frac{\prod_{j \in s} p_j \prod_{k \notin s} (1 - p_k)}{\sum_{s \in \mathcal{S}_n} \prod_{j \in s} p_j \prod_{k \notin s} (1 - p_k)}. \end{aligned} \tag{2.3}$$

It is important to note that there are two first order inclusion probabilities in CPS. There are the inclusion probabilities, p_i, used to select the Poisson Samples which result in the CPS inclusion probabilities after samples have been rejected. The inclusion probabilities for the Poisson Sampling algorithm will be referred to as the working probabilities of CPS. The CPS inclusion probabilities, $\breve{\pi}_i$, are calculated

from the working probabilities, $\boldsymbol{p} = \{p_1, \ldots, p_N\}$, by

$$\begin{aligned} \breve{\pi}_i(\boldsymbol{p}) &= P(i \in s | s \in \mathcal{S}_n) \\ &= \frac{\sum_{s \in \mathcal{S}_n^i} \prod_{j \in s} p_j \prod k \notin s (1 - p_k)}{\sum_{s \in \mathcal{S}_n} \prod_{j \in s} p_j \prod_{k \notin s} (1 - p_k)} \end{aligned} \tag{2.4}$$

where $\mathcal{S}_n^i$ is the set of samples of size n which include unit i. Equation (2.4) requires the enumeration of all possible samples, therefore, is not feasible to calculate the above formula for large populations. Hájek (1964) proposed approximations to the first and second order inclusion probabilities for CPS based on the working probabilities. It is rarely true that $p_j = \breve{\pi}_j$, however Hájek showed that as $N \to \infty$ uniformly on j

$$\breve{\pi}_j / p_j \to 1 \quad j = 1, \ldots, N \tag{2.5}$$

provided that $\sum_{j \in U} p_j(1 - p_j) \to \infty$.

One major disadvantage with CPS at the time it was first studied by Hájek was that the exact inclusion probabilities could only be approximated. In addition, if the exact inclusion probabilities were known, for instance if they were defined by (2.1), then it was not possible to determine the working probabilities which would guarantee these exact probabilities were obtained. To determine the working probabilities, $\boldsymbol{p} = (p_1, p_2, \ldots, p_N)$, to produce the exact inclusion probabilities, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$,

$$\boldsymbol{\pi} = \breve{\pi}_i(\boldsymbol{p}) \quad \text{for} \quad i = 1, \ldots, N \tag{2.6}$$

must be solved, where $\breve{\pi}_i(\boldsymbol{p})$ is defined by equation (2.4). Dupačová (1979) showed that (2.6) has a unique solution when $\sum_{i=1}^{N} p_i = n$. These unique working probabilities are denoted by the vector, $\tilde{\boldsymbol{p}}$. Hájek proposed a method to adjust the p_i's if the exact inclusion probabilities were known, however this only ensured the exact inclusion probabilities were approximately obtained.

Recently, algorithms have been developed to determine the exact first and second order inclusion probabilities under CPS. These recursive algorithms do not require all samples to be enumerated, and allow the exact inclusion probabilities to be calculated from the working probabilities and vice versa. Two algorithms are considered in this thesis, one developed by Chen *et al.* (1994) which was later improved by Deville (2000), and another developed by Aires (1999). A greater emphasis will be placed on the first algorithm as it is faster to implement, although both algorithms can be implemented within an acceptable time even for moderately large populations.

For the following sections, $\breve{\boldsymbol{\pi}}$ denotes the CPS inclusion probabilities calculated from the given working probabilities, $\boldsymbol{p}$. The desired inclusion probabilities are denoted by $\boldsymbol{\pi}$, and the working probabilities needed to produce these desired probabilities are denoted by $\tilde{\boldsymbol{p}}$.

Chen and Deville's algorithm

The algorithm developed by Chen *et al.* (1994) and later improved by Deville (2000), was developed as a result of noting the relationship between CPS and the exponential family of distributions. If the working probabilities, $\boldsymbol{p}$, are given, the first order inclusion probabilities, $\breve{\boldsymbol{\pi}} = (\breve{\pi}_1, \ldots, \breve{\pi}_N)$, can be determined for any permissable sample size n by calculating

$$\breve{\pi}_i = \psi_i(\boldsymbol{p}, n) = n \frac{\frac{p_i}{1-p_i}[1 - \psi_i(\boldsymbol{p}, n-1)]}{\sum_{k \in U} \frac{p_k}{1-p_k}[1 - \psi_k(\boldsymbol{p}, n-1)]} \tag{2.7}$$

recursively, where $\psi_k(\boldsymbol{p}, 0) = 0$ for all $k \in U$. Table 2.2 shows the first order inclusion probabilities calculated by equation (2.7), for when the working probabilities are known for a sampling situation of $N = 5$ and $n = 3$. Table 2.2 also shows that $p_i \neq \breve{\pi}_i$ as expected for a small population.

$\boldsymbol{p}$	$\breve{\boldsymbol{\pi}}$
0.24	0.1750
0.57	0.5536
0.65	0.6654
0.73	0.7616
0.81	0.8444
3.00	3.0000

Table 2.2: Inclusion probabilities, $\breve{\boldsymbol{\pi}}$ given the working probabilities $\boldsymbol{p}$

The joint inclusion probabilities are then determined recursively by,

$$\begin{aligned}\pi_{ij} &= \psi_{ij}(\boldsymbol{p}, n) \\ &= \frac{n(n-1)\frac{p_i}{1-p_i}\frac{p_j}{1-p_j}[1-\psi_i(\boldsymbol{p},n-2)-\psi_j(\boldsymbol{p},n-2)+\psi_{ij}(\boldsymbol{p},n-2)]}{\sum_{k\in U}\sum_{\substack{l\in U\\ l\neq k}}\frac{p_k}{1-p_k}\frac{p_l}{1-p_l}[1-\psi_k(\boldsymbol{p},n-2)-\psi_l(\boldsymbol{p},n-2)+\psi_{kl}(\boldsymbol{p},n-2)]} \quad (2.8)\end{aligned}$$

where $\psi_{ij}(\boldsymbol{p},0) = \psi_{ij}(\boldsymbol{p},1) = 0$ and $\psi_{ij}(\boldsymbol{p},2) = 2\frac{\frac{p_i}{1-p_i}\frac{p_j}{1-p_j}}{\sum_{k\in U}\sum_{l\in U l\neq k}\frac{p_k}{1-p_k}\frac{p_l}{1-p_l}}$.

It is usually the case, however, that the inclusion probabilities are known and the working probabilities need to be determined. In this situation the required working probabilities are determined by the Newton method. Let $\tilde{\boldsymbol{p}}^{(0)} = \boldsymbol{\pi}$, then iterate using

$$\tilde{\boldsymbol{p}}^{(k+1)} = \tilde{\boldsymbol{p}}^{(k)} + (\boldsymbol{\pi} - \psi_i(\tilde{\boldsymbol{p}}^{(k)}, n)) \quad (2.9)$$

for $k = 1, 2, \ldots$ until convergence; that is, until $\sum_{i\in U}|\tilde{p}_i^{(k)} - \tilde{p}_i^{(k+1)}|$ is less than a predetermined precision. The complete derivation of the above equations is comprehensively explained by Tillé (2006, pp. 79-86).

Table 2.3 shows the working probabilities obtained from (2.9) given the exact inclusion probabilities, $\boldsymbol{\pi}$. To indicate the accuracy of this iterative method

the inclusion probabilities were then determined by (2.7) from these working probabilities and are also shown in Table 2.3. This table indicates that the recalculated inclusion probabilities, $\breve{\boldsymbol{\pi}}$, agree with the original inclusion probabilities to the fourth decimal place.

$\boldsymbol{\pi}$	$\tilde{\boldsymbol{p}}$	$\breve{\boldsymbol{\pi}}$
0.24	0.3034	0.2400
0.57	0.5783	0.5700
0.65	0.6378	0.6500
0.73	0.7034	0.7300
0.81	0.7772	0.8100
3.00	3.0001	3.0000

Table 2.3: Working probabilities, $\tilde{\boldsymbol{p}}$, given the exact inclusion probabilities, $\boldsymbol{\pi}$, and recalculated inclusion probabilities $\breve{\boldsymbol{\pi}}$

Finally, Table 2.4 shows the joint inclusion probabilities determined using the desired inclusion probabilities in Table 2.3. A simple check verifies that these joint inclusion probabilities satisfy the fixed-size sampling property (1.6), as $\sum_{i=1}^{N}\sum_{j>i}\pi_{ij} = 3.0002$ and $n(n-1)/2 = 3$.

	1	2	3	4	5
1		0.0864	0.1052	0.1298	0.1587
2			0.2884	0.3508	0.4145
3				0.4195	0.4869
4					0.5600

Table 2.4: Joint inclusion probabilities given the inclusion probabilities in Table 2.3

Aires' algorithm

Aires (1999) derived an alternative recursive algorithm to calculate the exact first and second order inclusion probabilities. To determine the inclusion probabilities from the working probabilities under Aires' algorithm, first consider

$$\breve{\pi}_i = \phi_i(\boldsymbol{p}) = \frac{p_i S_{n-1}^{N-1}(p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_N)}{S_n^N(p_i, \ldots, p_N)} \tag{2.10}$$

$$S_n^N(p_1, \ldots, p_N) = \sum_{s \in \mathcal{S}_n(N)} \prod_{i \in s} p_i \prod_{j \notin s} (1 - p_j) \tag{2.11}$$

where $S_n(N)$ is the subset of samples of size $n \leq N$ from $\{1, \ldots, N\}$. S_n^N is defined for $N = 0, 1, 2, \ldots$ and $n = 0, \ldots, N$, and can be calculated recursively by

$$S_n^N(p_1, \ldots, p_N) = p_N S_{n-1}^{N-1}(p_i, \ldots, p_{N-1}) + (1 - p_N) S_n^{N-1}(p_1, \ldots, p_{N-1}) \tag{2.12}$$

using $S_0^N = (1 - p_1)(1 - p_2) \cdots (1 - p_N)$ and $S_N^N = p_1 p_2 \cdots p_N$.

The joint inclusion probabilities can be computed using a similar approach to the algorithm for the inclusion probabilities above, however, Aires developed a faster algorithm jointly with Prof. O. Nerman (Aires, 1999). This algorithm proceeded as follows: let $\gamma_i = p_i/(1 - p_i)$ then

$$\pi_{ij} = \frac{\gamma_i \breve{\pi}_j - \gamma_j \breve{\pi}_i}{\gamma_i - \gamma_j} \tag{2.13}$$

for all $i \neq j$ and $\gamma_i \neq \gamma_j$. For the case when $\gamma_i = \gamma_j$ and $j \neq i$ the fixed sampling design property $\sum_{\substack{j \in U \\ j \neq i}} \pi_{ij} = (n - 1)\pi_i$, equation (1.5), is required. Let $\pi_{ij} = \pi_{ij_0}$ for all units $j \neq i$ in which $\gamma_i = \gamma_j$ for a fixed unit i. Assume that for this fixed value there are k_i units satisfying this case, therefore

$$(n - 1)\pi_i = \sum_{\substack{j \in U \\ j \neq i}} \pi_{ij} = \sum_{\substack{j \in U \\ \gamma_j \neq \gamma_i}} \pi_{ij} + k_i \pi_{ij_0}$$

which implies

$$\pi_{ij_0} = \frac{(n - 1)\pi_i - \sum_{\substack{j \in U \\ \gamma_j \neq \gamma_i}} \pi_{ij}}{k_i} \tag{2.14}$$

There is no need to specify the value for $\breve{\pi}_{ii}$ as this is not required for the SYG variance estimator.

To determine the working probabilities based on the exact inclusion probabilities, Aires proposed two methods. The first method is to solve a non-linear system of $N + 1$ equations; the N equations from (2.6) and the additional equation that $\sum_{i \in U} p_i = n$. This method is exact, however, quite slow for large populations. The second method proposed is similar to the approach used by Chen and Deville. Let $\tilde{\boldsymbol{p}}^{(0)} = \boldsymbol{\pi}$ and iterate

$$\tilde{\boldsymbol{p}}^{(k+1)} = \tilde{\boldsymbol{p}}^{(k)} + (\boldsymbol{\pi} - \phi(\tilde{\boldsymbol{p}}^{(k)})) \tag{2.15}$$

until $max_{i \in U} |\frac{\phi_i(\tilde{\boldsymbol{p}}^{(k)})}{\pi_i} - 1|$ is less than or equal to a predetermined precision. After each iteration $\phi_i(\tilde{\boldsymbol{p}}^{(k)})$ is normalised by setting

$$\phi_i(\tilde{\boldsymbol{p}}^{(k)}) = n \frac{\phi_i(\tilde{\boldsymbol{p}}^{(k)})}{\sum_{j \in U} \phi_j(\tilde{\boldsymbol{p}}^{(k)})} \tag{2.16}$$

which ensures that the probabilities $\tilde{\boldsymbol{p}}$ are well defined. The values obtained under Aires algorithm are the same as those produced in Table 2.3 and Table 2.4 to the four decimal places shown.

Implementing Conditional Poisson Sampling

Algorithm 2.2 below, describes the implementation of CPS when the desired inclusion probabilities are defined by (2.1). Example 2.2 provides a small example of the use of this algorithm.

Algorithm 2.2. *Conditional Poisson Sampling*

(i) Assign each unit a probability of inclusion by (2.1).

(ii) Determine the working probabilities, $\tilde{\boldsymbol{p}}$ given these desired inclusion probabilities, from (2.9) or (2.15).

(iii) For each unit k, generate a random Bernoulli trial with probability $\tilde{p}_k$ and accept the unit if the trial is successful.

(iv) If the number of units in the sample is n accept the sample, otherwise reject this sample and repeat (iii).

Example 2.2. *Consider the same sampling situation as in Example 2.1 where $N = 5$ and $n = 3$. Using Chen and Deville's algorithm, the sample selected in Trial 1, $s = (1, 4)$, was rejected as only 2 units were included. The sample selected in Trial 2 was $s = (2, 4, 5)$, and was accepted as the desired sample was of 3 units.*

Units	π_i	$\tilde{\boldsymbol{p}}$	Trial 1	Trial 2
1	0.24	0.3034	1	0
2	0.57	0.5783	0	1
3	0.65	0.6378	0	0
4	0.73	0.7034	1	1
5	0.81	0.7772	0	1

Table 2.5: Sample selected using CPS

Originally, CPS was not commonly used because the first and second order inclusion probabilities could only be approximated. Clearly, this is no longer an issue. One important property of CPS is that it attains the maximum possible entropy among sampling algorithms with the same inclusion probabilities and the same support (Hájek, 1981, p. 29). This is an advantageous property because

approximate variance estimators perform well under high entropy sampling designs. Chen *et al.* (1994) showed that under maximal entropy $\pi_{ij} \leq \pi_i \pi_j$ for all $i \neq j$, which ensures that the SYG variance estimator (1.15) is always positive.

2.2 Variance Estimation

As the calculations of the joint inclusion probabilities are usually not straightforward, many approximations to (1.14) have been developed for fixed-size sampling algorithms. This was initially done by approximating the joint inclusion probabilities in terms of the first order inclusion probabilities only as will be discussed in section (2.2.1). These approximations were used in the SYG variance estimator, however, this did not remove the cumbersome double summation of that estimator. As a result, many approximate variance estimators have been derived that exclude both the joint inclusion probabilities and the double summation.

Within this research nine different approximate variance estimators shall be considered. These estimators are divided into two groups, the Brewer Family and the Hájek-Deville Family, as outlined in sections (2.2.2) and (2.2.3) respectively. These families are defined specifically for this thesis and are not used explicitly within the literature. The estimators are grouped together due to the within-group similarities in the formulae of the variance estimators. This allows the behaviour of these two types of estimators to be meaningfully compared under both sampling algorithms.

The behaviour of variance estimators can be compared by their relative bias (RB) and their mean squared error (MSE). The RB is considered as it is desirable that an estimator be as nearly unbiased as possible. The MSE is considered as it provides a measure of how close the variance estimates are to the true variance. A variance estimator with an RB close to zero, and a small MSE is preferable. The

calculation of these properties is discussed in more detail in the next chapter (see section 3.3).

2.2.1 Approximations to the joint inclusion probabilities

Hartley and Rao (1962) derived an approximation to the joint inclusion probabilities with precision of order $O(N^{-4})$. This approximation is given by

$$\begin{aligned}\pi_{ij} &= \frac{n-1}{n}\pi_i\pi_j + \frac{n-1}{n^2}(\pi_i^2\pi_j + \pi_i\pi_j^2) - \frac{n-1}{n^3}\pi_i\pi_j\sum_{k\in U}\pi_k^2 \\ &\quad + \frac{2(n-1)}{n^3}(\pi_i^3\pi_j + \pi_i\pi_j^3 + \pi_i^3\pi_j^2) - \frac{3(n-1)}{n^4}(\pi_i^2\pi_j + \pi_i\pi_j^2)\sum_{k\in U}\pi_k^2 \\ &\quad + \frac{3(n-1)}{n^5}\pi_i\pi_j(\sum_{k\in U}\pi_k^2)^2 - \frac{2(n-1)}{n^4}\pi_i\pi_j\sum_{k\in U}\pi_k^3. \qquad (2.17)\end{aligned}$$

Hartley and Rao derived this asymptotic approximation under RANSYS by keeping n fixed while letting N increase, therefore it is only applicable if the population is large compared with the sample size. Chen *et al.* (1994) stated that this approximation does not satisfy $\pi_{ij} \leq \pi_i\pi_j$ except when $n = 2$, indicating that it may produce negative estimates of the variance when used in the SYG variance estimator.

Asok and Sukhatme (1976) examined a completely different sampling algorithm devised by Samford (1967), and to order $O(n^3N^{-3})$ produced the identical approximation to (2.17) given by

$$\pi_{ij} \doteq \frac{1}{2}\pi_i\pi_j(c_i + c_j) \qquad (2.18)$$

$$\text{where} \quad c_i = n^{-1}(n-1)(1 - n^{-2}\sum_{k\in U}\pi_k^2 + 2n^{-1}\pi_i). \qquad (2.19)$$

It can be shown that (2.18) is the same as the first three terms of (2.17). This approximation is referred to as Hartley and Rao's third order approximation throughout this thesis. The only relationship between the derivation of the above

two approximations is that both sampling algorithms have a high entropy. This suggests that high entropy sampling algorithms can produce similar joint inclusion probabilities and hence similar variance estimates.

Hájek (1964) derived another approximation to the joint inclusion probabilities under CPS given by

$$\pi_{ij} \dot{=} \pi_i \pi_j \big[1 - (1-\pi_i)(1-\pi_j)\big\{\sum_{i \in U} \pi_k(1-\pi_k)\big\}^{-1}\big]. \tag{2.20}$$

His derivation was based on letting both $n \to \infty$ and $(N-n) \to \infty$, hence the population does not need to be large compared with the sample size.

The above approximations can be substituted into the SYG variance estimator (1.15) to produce approximate variance estimators. Berger (2004) showed that the approximate variance estimator produced by substituting (2.20) into (1.15) produces estimates close to the exact SYG variance estimator when the joint inclusion probabilities are known.

2.2.2 The Brewer Family

The Brewer Family is defined as variance estimators which include the term $(y_i\pi_i^{-1} - n^{-1}\hat{Y}_{\bullet HT})^2$, where $\hat{Y}_{\bullet HT}$ is the HTE of equation (1.11). Brewer (2002) modified Hartley and Rao's third order approximation of the joint inclusion probabilities in (2.18) to derive the following approximation to (1.14),

$$\tilde{V}(\hat{Y}_{\bullet HT}) = \sum_{i \in U} (1 - c_i\pi_i)(y_i\pi_i^{-1} - n^{-1}\hat{Y}_{\bullet HT})^2 \tag{2.21}$$

where c_i is defined as one of the following,

$$BR1 \qquad c_i = \frac{n-1}{n - n^{-1}\sum_{k\in U}\pi_k^2} \tag{2.22}$$

$$BR2 \qquad c_i = \frac{n-1}{n-\pi_i} \tag{2.23}$$

$$BR3 \qquad c_i = \frac{n-1}{n - 2\pi_i + n^{-1}\sum_{k\in U}\pi_k^2} \tag{2.24}$$

$$BR4 \qquad c_i = \frac{n-1}{n - (2n-1)(n-1)^{-1}\pi_i + (n-1)^{-1}\sum_{k\in U}\pi_k^2}. \tag{2.25}$$

The values of c_i were determined by using the properties of fixed-size sampling algorithms, and the ratios of sums of the $\pi'_{ij}s$ to the corresponding sums of $\pi_i\pi'_j s$. Under *srswor* (2.22), (2.23) and (2.24) produce the correct SYG variance estimator when the exact joint inclusion probabilities are used (see Example 1.1). Brewer (2002, p. 153, 158) recommends that to provide the greatest possible accuracy (2.25) should be used, but mentions that (2.24) is nearly as accurate. Complete derivation of these approximate variances is given in Brewer (2002, Chap. 9).

Brewer's suggested estimator for the approximation (2.21) is

$$\hat{\hat{V}}(\hat{Y}_{\bullet HT}) = \sum_{i\in s}(c_i^{-1} - \pi_i)(y_i\pi_i^{-1} - n^{-1}\hat{Y}_{\bullet HT})^2 \tag{2.26}$$

where the c_i's are defined as above. Throughout this thesis $\hat{\hat{V}}_{BR1}$, $\hat{\hat{V}}_{BR2}$, $\hat{\hat{V}}_{BR3}$ and $\hat{\hat{V}}_{BR4}$ denoted the approximate variance estimator in (2.26) with c_i defined by (2.22), (2.23), (2.24) and (2.25) respectively.

In some sampling situations, the first order inclusion probabilities are unknown for every unit in the population. Under this situation only $\hat{\hat{V}}_{BR2}$ can be used, as the corresponding values of c_i for the other estimators cannot be determined. Berger (2004) states that $\hat{\hat{V}}_{BR2}$ does not take into consideration the correction for the degrees of freedom, which implies it would not be a good estimator for small populations. Other empirical studies, however, show that this estimator is still comparable to

other variance estimators for $n = 10$ and even $n = 2$ (Brewer and Donadio (2003) and Donadio (2002)).

Another variance estimator that has been included in this family, for the purpose of this study, is one which was derived by Deville (1999). The reason for its inclusion in this family is due to the term $(y_i\pi_i^{-1} - n^{-1}\hat{Y}_{\bullet HT})^2$. This estimator is

$$\hat{\tilde{V}}_{BR-Dev}(\hat{Y}_{\bullet HT}) = \frac{1}{1-\sum_{k\in s} a_k^2}\sum_{i\in s}(1-\pi_i)(y_i\pi_i^{-1} - n^{-1}\hat{Y}_{\bullet HT})^2 \quad (2.27)$$

$$\text{where} \qquad a_k = \frac{1-\pi_k}{\sum_{j\in s}(1-\pi_j)} \quad (2.28)$$

To avoid confusion this estimator is denoted by $\hat{\tilde{V}}_{BR-Dev}$ to indicate that it is in the Brewer Family, albeit derived by Deville.

2.2.3 The Hájek-Deville Family

The second group of variance estimators is the Hájek-Deville Family. These estimators are based on the Hàjek variance approximation determined by substituting (2.20) into (1.14), which after some simple manipulation gives

$$\tilde{V}_{Haj}(\hat{Y}_{\bullet HT}) = \sum_{i\in U}\pi_i(1-\pi_i)(y_i\pi_i^{-1} - A_u)^2 \quad (2.29)$$

$$\text{where} \qquad A_u = \sum_{k\in U}\Big(\frac{1-\pi_k}{\sum_{j\in U}(1-\pi_j)}\Big)y_k\pi_k^{-1}.$$

The Hájek-Deville Family variance estimators are defined as those which include the term $(y_i\pi_i^{-1} - A_s)^2$ where A_s is the same as A_u, only it is the summation over the sample instead of the population; that is,

$$A_s = \sum_{k\in s} a_k y_k \pi_k^{-1} \quad (2.30)$$

where a_k is defined by (2.28).

Two variance estimators have been derived based on the approximate variance (2.29). The first was derived by Hájek (1964), and shall be denoted by $\hat{\tilde{V}}_{Dev1}$

because $\tilde{V}_{Haj}$ represents the approximate variance of (2.29) and Deville (1999) has also worked within this area. The second variance estimator was proposed by Deville (1999) and is denoted by $\hat{\tilde{V}}_{Dev2}$. These estimators are,

$$\hat{\tilde{V}}_{Dev1} = \frac{n}{n-1}\sum_{i\in s}(1-\pi_i)(y_i\pi_i^{-1}-A_s)^2 \tag{2.31}$$

$$\hat{\tilde{V}}_{Dev2} = \frac{1}{1-\sum_{k\in s}a_k^2}\sum_{i\in s}(1-\pi_i)(y_i\pi_i^{-1}-A_s)^2. \tag{2.32}$$

A further two estimators included in this family due to the similarity in the variance estimators, even though they do not include the term $(y_i\pi_i^{-1}-A_s)^2$. One of these was proposed by Berger (1998) namely,

$$\hat{\tilde{V}}_{Ber} = \sum_{i\in s}\frac{b_i}{\pi_i^2}(y_i-\hat{y}_i^*)^2 \tag{2.33}$$

$$\text{where}\quad \hat{y}_i^* = \pi_i\frac{\sum_{k\in s}b_k y_k/\pi_k}{\sum_{j\in U}b_j} \tag{2.34}$$

$$b_i = (1-\pi_i)\frac{n}{n-1}\frac{\sum_{k\in s}(1-\pi_k)}{\sum_{j\in U}\pi_j(1-\pi_j)}.$$

This is similar to the previous two estimators as both $\hat{\tilde{V}}_{Dev1}$ and $\hat{\tilde{V}}_{Dev2}$ are also of the form (2.33) but with $b_i=(1-\pi_i)\frac{n}{n-1}$ and $b_i=(1-\pi_i)\Big[1-\sum_{j\in s}\big(\frac{1-\pi_j}{\sum_{k\in s}(1-\pi_k)}\big)^2\Big]^{-1}$ respectively.

The other estimator in this family was proposed by Rosén (1991), namely

$$\hat{\tilde{V}}_{Ros} = \frac{n}{n-1}\sum_{i\in s}(1-\pi_i)(y_i\pi_i^{-1}-A)^2 \tag{2.35}$$

$$\text{where}\quad A = \frac{\sum_{k\in s}y_k\frac{1-\pi_k}{\pi_k}log(1-\pi_k)}{\sum_{j\in s}\frac{1-\pi_j}{\pi_j}log(1-\pi_j)}$$

which is the same as (2.31) with A replacing A_s.

2.3 Previous Empirical Studies

The behaviour of the above approximate variance estimators have been studied by Donadio (2002), Brewer and Donadio (2003), Matei and Tillé (2005) and Berger

(2004). These studies showed that the knowledge of joint inclusion probabilities is not necessary to produce accurate variance estimates.

Berger (2004) used a high entropy sampling design, Chao (1982) Sampling, where the joint inclusion probabilities were known exactly. Of the variance estimators considered in this thesis, Berger only studied $\hat{\tilde{V}}_{Dev1}$ and $\hat{\tilde{V}}_{BR2}$. He stated that the bias of $\hat{\tilde{V}}_{Dev1}$ should be lower than the bias of $\hat{\tilde{V}}_{BR2}$, and this was confirmed by his empirical results. There was, however, no noticeable difference among the MSEs for these variance estimators.

Donadio (2002) analysed the properties of (2.31), (2.32) and the first three Brewer estimators (2.26); $\hat{\tilde{V}}_{BR4}$ had not been derived at the time of his research. Monte Carlo simulations were used to analyse the variance estimators under both RANSYS and Pareto πps Sampling. The analysis was conducted over 10,000 simulations for three artificial populations. Donadio concluded that there was no superior choice between the approximate variance estimators. The variance estimators $\hat{\tilde{V}}_{BR1}$, $\hat{\tilde{V}}_{Dev1}$ and $\hat{\tilde{V}}_{Dev2}$ generally had smaller MSEs, and $\hat{\tilde{V}}_{BR2}$ was never the optimal choice. Overall, $\hat{\tilde{V}}_{BR1}$, $\hat{\tilde{V}}_{BR3}$, $\hat{\tilde{V}}_{Dev1}$ and $\hat{\tilde{V}}_{Dev2}$ have a good balance between their RB and MSE across all sample sizes of $n = 10$, 20 and 40. Donadio also analysed the situation of sampling 2 units, and under this approach $\hat{\tilde{V}}_{BR3}$ and $\hat{\tilde{V}}_{Dev1}$ performed the best.

There is a major discrepancy in relation to the performance of the Brewer Family estimators between the results of Brewer and Donadio (2003) and of Matei and Tillé (2005). These authors considered nearly identical populations yet produced conflicting conclusions. Although these studies used two different sampling algorithms, there should not have been such a large effect on the relative performance of the variance estimators, as both algorithms were of high entropy. The methods

used by each of these papers is explained before specifying the nature of this major discrepancy.

2.3.1 Brewer and Donadio

Brewer and Donadio (2003) specifically analysed the properties of the Brewer Family variance estimators and showed empirically that these estimators produced acceptable estimates. They studied $\hat{\tilde{V}}_{BR1}$, $\hat{\tilde{V}}_{BR2}$, $\hat{\tilde{V}}_{BR3}$, $\hat{\tilde{V}}_{BR4}$, $\hat{\tilde{V}}_{Dev1}$ and $\hat{\tilde{V}}_{Dev2}$ which they denoted as $\hat{\tilde{V}}_{(16.9)}$, $\hat{\tilde{V}}_{(16.10)}$, $\hat{\tilde{V}}_{(16.11)}$, $\hat{\tilde{V}}_{(16.18)}$, $\hat{\tilde{V}}_{Haj}$ and $\hat{\tilde{V}}_{Dev}$ respectively. They also considered the SYG variance estimator with the complete Hartley and Rao approximation of the joint inclusion probabilities. They used the RB and coefficient of variation to analyse the properties of the variance estimators. The formula they used was for the RB was

$$RB_{BD} = E(\hat{V})/V_s(\hat{Y}_{\bullet HT}) - 1 \tag{2.36}$$

where $V_s(\hat{Y}_{\bullet HT})$ was the empirical variance of the simulated estimates of the total.

They analysed two real populations, including the MU281 population, which as described further in the next chapter, is a data set containing information about 281 municipalities in Sweden. The revenue from the 1985 municipal taxation ($RMT85$) was used as the study variable, and the population in 1975 ($P75$) as a measure of size variable, where the correlation between $RMT85$ and $P75$ was 0.99.

RANSYS and Tillé (1996a) Sampling, (the latter is also a high entropy fixed-size sampling design), were used to simulate 50,000 independent random samples. For each sample, the variance estimates were computed and the RBs and coefficient of variations were used to compare their behaviour. As the variance estimators had similar coefficient of variations, this property was not a useful measure of comparison. However, it is of interest that $\hat{\tilde{V}}_{BR1}$ consistently had the smallest

coefficient of variation for the MU281 population, despite the differences being trivial. For the MU281 population, $\hat{\hat{V}}_{BR2}$, $\hat{\hat{V}}_{BR3}$ and $\hat{\hat{V}}_{BR4}$ outperformed both $\hat{\hat{V}}_{Dev1}$, and $\hat{\hat{V}}_{Dev2}$ under RANSYS in relation to their RBs. For Tillé Sampling, however, the reverse was true. Overall there were no substantial differences among the RBs of these variance estimators.

One further point discussed briefly in their paper was the entropy of the two sampling algorithms considered. As the RBs under Tillé sampling were always positive and tending to increase with the sample size, Brewer and Donadio believed that this design had a lower entropy than RANSYS. Monte Carlo variances were calculated under both sampling algorithms, and the variance was always smaller under Tillé sampling, which Brewer and Donadio assumed indicated that this design had a lower entropy. These concepts of variance and entropy are discussed further in Chapter 4.

2.3.2 Matei and Tillé

Matei and Tillé (2005) extended the above research further by considering CPS, and the benefit of using the exact joint order inclusion probabilities. They compared twenty different variance estimators, including those mentioned in section 2.2. The notation they used is the same as that of this thesis, except that $\hat{\hat{V}}_{BR-Dev}$ was defined by $\hat{\hat{V}}_{Dev3}$, and $\hat{\hat{V}}_{Ros}$ was defined by $\hat{\hat{V}}_{R}$. As the joint inclusion probabilities were explicitly calculated, the variance of the HTE was computed using (1.14). In addition, both the Horvitz-Thompson variance estimator (1.13) and the SYG variance estimator (1.15) were calculated to see whether they were more precise than the approximate variance estimators.

Two artificial populations and the MU281 population were analysed. For the MU281 population $RMT85$ was used as their study variable and their measure of

size variable was the population measured in 1985 ($P85$). Analyses were conducted over 10,000 independent samples for each population. The relative bias and mean squared error were used as comparative measures; however, the relative bias was defined by these authors as being

$$RB_{MT} = \frac{E(\hat{V}) - V_{MT}}{\sqrt{V_s(\hat{V})}} \tag{2.37}$$

where V_{MT} was the true variance determined from (1.14) using the exact joint inclusion probabilities, and V_s was the variance of the simulated variance estimates.

The RBs indicated that the performance of $\hat{V}_{SYG}$ and $\hat{V}_{HTE}$ was similar to the Hájek-Deville variance estimators. This confirmed that the knowledge of the joint inclusion probabilities was not necessary to produce more precise estimates than those provided by other approximate estimators. However, for the MU281 population, the Brewer Family estimators had extremely high RBs (of the order of 12%) for samples of size $n = 40$, which did not agree with Brewer and Donadio's results. For the two artificial populations, however, the estimators in both families were comparably even for $n = 40$. As an aside, Matei and Tillé concluded that the correlation between the study variable and the auxiliary variable did not have an affect on the overall performance of the variance estimators.

2.3.3 The major discrepancy between the papers

The major discrepancy between these two papers was in the behaviour of the Brewer Family variance estimators for $n = 40$ in the MU281 population. Although the papers used different measure of size variables, the correlation between $P85$ and $P75$ was 0.995, therefore the results of the two studies should be similar. As stated earlier, under high entropy the sampling algorithms used should have produced reasonably similar results. Therefore the overall results should have been similar.

The results published in the two papers are shown side-by-side in Table 2.6. As some of the variance estimators in this thesis were not considered by Brewer and Donadio these RBs have been denoted by the symbol "-". Table 2.6 indicates that for the first two sample sizes the relative biases are similar. It is shown later that the true causes of these differences are due partly to the relatively small number of simulations conducted, and partly to the inconsistency between what Matei and Tillé described as being their sampling procedure, and the procedures they actually used.

	Matei and Tillé			**Brewer and Donadio**		
	RB (%)			RB (%)		
	n=10	n=20	n=40	n=10	n=20	n=40
$\hat{V}_{SYG}$	-0.67	-0.01	-0.64	-0.27	-0.43	0.77
$\hat{\hat{V}}_{Br1}$	-0.87	0.18	12.75	-0.34	-0.51	-0.67
$\hat{\hat{V}}_{Br2}$	-0.75	0.37	12.28	-0.40	-0.58	0.58
$\hat{\hat{V}}_{Br3}$	-0.63	0.56	11.81	-0.27	-0.43	0.76
$\hat{\hat{V}}_{Br4}$	-0.61	0.57	11.80	-0.27	-0.43	0.75
$\hat{\hat{V}}_{BR-Dev}$	-0.70	0.54	12.95	.	.	.
$\hat{\hat{V}}_{Dev1}$	-0.86	-0.34	-0.87	-0.40	-0.75	-0.59
$\hat{\hat{V}}_{Dev2}$	-0.81	-0.17	-0.19	-0.37	-0.68	-0.39
$\hat{\hat{V}}_{Ros}$	-0.84	-0.18	1.48	.	.	.
$\hat{\hat{V}}_{Ber}$	-0.70	-0.12	-0.76	.	.	.

Table 2.6: Comparison of the relative biases (%) produced in the corresponding papers

Table 2.6 shows that there is, however, a major discrepancy for the sample size of $n = 40$. For Brewer and Donadio the absolute size of their RBs for the estimators

are comparable, but Matei and Tillé's results indicate that the Brewer Family do not perform well at all. Therefore, it seems there may be an error in one or both of these papers. It is important to resolve this discrepancy because these approximate estimators of the HTE otherwise appear to be convenient and reliable replacements for the unbiased, and comparably accurate but distinctly more cumbersome, $\hat{V}_{SYG}$.

3 Simulation Study

3.1 Introduction

A simulation study was conducted to compare the behaviour of the variance estimators. The three aims of this section are

- To resolve the discrepancy between the results of Brewer and Donadio (2003) and of Matei and Tillé (2005).
- To compare the behaviour of the approximate variance estimators, and to determine whether there is one which is superior.
- To determine whether the exact joint inclusion probabilities, known exactly under CPS, can significantly improve the SYG variance estimator compared with the approximate estimators.

The first two sections of this chapter describe the data and methodology used within this study. The third section investigates the discrepancies between the results produced by Matei and Tillé. Finally, the behaviour of the variance estimators are compared under both RANSYS and CPS.

3.2 Data

It is important to study a variety of populations when comparing variance estimators, as the population used in a sampling situation will affect how well the variance estimators perform. This variety is achieved by constructing seven different real populations from two data sets; the MU284 Population and the CO124 Population. These populations, available from Särndal *et al.* (2003, pp. 652, 662), are described below.

The MU284 Population represents information regarding 284 municipalities in Sweden. A municipality typically consists of a single town and its surrounding area. This data set was provided by Statistics Sweden and is commonly used in survey sampling literature. The variables that are of interest in this study are as follows

- *P85* - 1985 population (in thousands);
- *P75* - 1975 population (in thousands);
- *RMT85* - Revenues from the 1985 municipal taxation (in millions of kronor);
- *SS82* - Number of Social-Democratic seats in municipal council; and
- *CS82* - Number of Conservative seats in municipal council.

The MU281 Population is the same as the MU284 Population, excluding the three largest municipalities, Stockholm, Göteborg and Malmö.

The CO124 Population, provided by the University of Lund, Sweden consists of variables provided for 124 countries. The variables of interest for this study are

- IMP - 1983 imports (in millions of U.S. dollars);
- EXP - 1983 exports (in millions of U.S. dollars);
- MEX - 1981 military expenditure (in millions of U.S. dollars); and
- P80 - 1980 population (in millions).

A summary of the seven populations constructed from these data sets is provided in Table 3.1. Column 1 shows the label used to represent the population throughout this study, and column 2 contains the data set from which the population was constructed by using the study variable, Y and the measure of size variable, X. For each population the true population total, $Y_\bullet$, which is to be estimated, is known.

The final column shows the correlation between the study variable and measure of size variable.

Label	Data Set	Y	X	$Y_\bullet$	$X_\bullet$	ρ_{XY}
(a)	MU281	RMT85	P85	53151	7033	0.9920
(b)	MU281	RMT85	P75	53151	6818	0.9870
(c)	MU284	P85	P75	7033	6818	0.9984
(d)	CO124	EXP	IMP	1770336	1810957	0.9760
(e)	CO124	EXP	MEX	1770336	514616	0.6970
(f)	MU284	SS82	CS82	6301	2583	0.2053
(g)	CO124	EXP	P80	1770336	4308	0.2296

Table 3.1: Populations used for the simulation study

These populations were analysed in this study. Populations (a) and (b) were considered as they were used in the two key papers under discussion. The results simulated using these populations were directly compared with the results already produced by Matei and Tillé and by Brewer and Donadio to explain any discrepancies between them. Populations (f) and (g) were chosen for their low correlations between Y and X. The magnitude of the correlation affects the precision of the Horvitz-Thompson estimator; a higher correlation will produce a more precise estimate and hence a lower variance. However, it is believed that the correlation should not effect the comparisons of the variance estimators as the same inclusion probabilities and study variable are used for each estimator. The remaining populations were included simply to ensure that a variety of situations could be studied.

3.3 Methodology

In this study, Monte Carlo simulations were used to compare the variance estimators. The simulations were run using the R statistical language, and the relevant code is provided in Appendix A. Three different sample sizes were considered; $n = 10$, $n = 20$ and $n = 40$. For each population, 50,000 independent samples of size n were generated under both RANSYS and CPS to ensure precise results were obtained in a feasible time. RANSYS was implemented by Algorithm 2.1 (see section 2.1.1), as this algorithm was used by Brewer and Donadio. CPS was implemented by Algorithm 2.2 using Chen and Deville's algorithm to calculate the working probabilities and the joint inclusion probabilities (see section 2.1.2), as this was the procedure used by Matei and Tillé.

The R *sampling* package written by Tillé and Matei (2006), available from http://cran.r-project.org, was used to implement this algorithm. This package implements the complete CPS design by determining the working and joint inclusion probabilities given the desired inclusion probabilities, and also selecting a sample. The precision chosen for the Newton method, (2.9), was 10^{-6}. To guarantee greater control over the sampling procedure, only two functions from the *sampling* package were used. The first function, *UPMEpiktildefrompik*, was used to calculate the working probabilities given the desired inclusion probabilities and a predetermined precision for the Newton method. The second function, *UPmaxentropypi2*, determined the joint inclusion probabilities based on the desired inclusion probabilities and the working probabilities. After the working probabilities were obtained, sets of N Bernoulli trials were generated to select the samples. Sets which produced the incorrect sample size were rejected. As samples were rejected it was faster to implement RANSYS than CPS.

A problem was encountered when trying to implement Chen and Deville's Algorithm without using the *sampling* package. It was found that for the calculation of $\tilde{\boldsymbol{p}}$ in the iterative algorithm (2.9), one could not simply start with $\psi(\boldsymbol{p}, 0) = 0$, then calculate $\psi(\boldsymbol{p}, 1)$ using equation (2.7) and continue similarly until arriving at $\psi(\boldsymbol{p}, n)$. Newton's iterative method occasionally resulted in $\tilde{p}_i > 1$ for some units i, which implied that on the next iteration some of the working probabilities would be negative while others would be appreciably greater than 1. Consequently, the iterative process never converged. This problem was first encountered for population (a) and a sample size of $n = 40$. It is important to normalise ψ after each iteration as in Aires' algorithm. Since this problem is overcome in the *sampling* package it was used in this study.

For each sample, eleven estimates were calculated; an estimate of the total of Y, nine variance estimates using the approximate estimators in the previous chapter and the SYG variance estimate. Calculating the SYG variance estimate under CPS was straightforward, as the exact joint inclusion probabilities were known. Hartley and Rao's full approximation of the joint inclusion probabilities, (2.17), was used to calculate the SYG variance estimate for RANSYS. The implementation time to determine this variance estimator under RANSYS was considerably longer than under CPS. The variance estimators were computed for the same individual samples to improve comparisons.

A slightly different methodology is required for sampling situations where units are assigned an inclusion probability of one. It has been shown that if $nx_i > \sum_{i \in U} x_i$, then this unit needs to be included in the sample with certainty. This is the common approach in survey sampling, but how to deal with these units in relation to the variance estimators does not seem to have been directly discussed in the literature. This issue is important as it affects the performance of the variance estimators. If

$\pi_i = 1$ then this unit is included in every possible sample selected for the population, U. If a sample size of n is required and n_E units have an inclusion probability of one, then only n_* units need to be selected from U_*, where U_* is the population of units excluding those included with certainty. This ensures that the desired sample size of $n = n_* + n_E$ is obtained.

A unit included with certainty does not add to the variance, hence the variance of the HTE should be calculated only for the population units that were not included with certainty. That is, the variance of $\hat{Y}_{\bullet HT}$ defined in (1.11) is the same as the variance of $\hat{Y}^*_{\bullet HT}$, which is the HTE for U_*. If U_E is the set of units included with certainty, then

$$\begin{aligned}
\hat{V}(\hat{Y}_{\bullet HT}) &= \hat{V}(\sum_{i\in U} \delta_i \frac{y_i}{\pi_i}) \\
&= \hat{V}(\sum_{i\in U_*} \delta_i \frac{y_i}{\pi_i} + \sum_{i\in U_E} y_i) \\
&= \hat{V}(\sum_{i\in U_*} \delta_i \frac{y_i}{\pi_i}) \\
&= \hat{V}(\hat{Y}^*_{\bullet HT})
\end{aligned}$$

where the first equality holds because $\delta_i = 1$ for units included with certainty, and the second inequality holds because $\sum_{i\in U_E} y_i$ is constant with respect to all possible samples. Hence, the variance of the HTE of U must be calculated as the variance of the HTE of U_*. As a result of this, each variance estimator formula was modified in this situation by replacing n with n_*, $\sum_U$ with $\sum_{U_*}$ and $\hat{Y}_{\bullet HT}$ with $\hat{Y}^*_{\bullet HT}$. For example, $\hat{\tilde{V}}_{BR1}$ is calculated by

$$\begin{aligned}
\hat{\tilde{V}}_{BR1} &= \sum_{U_*} \delta_i (c^{*-1}_i - \pi_i)(y_i \pi_i^{-1} - \hat{Y}^*_{\bullet HT} n^{-1})^2 \\
c^*_i &= \frac{n_* - 1}{n_* - \sum_{k\in U_*} \pi_k}.
\end{aligned}$$

This is an important modification because most of the variance estimators have

not been designed to ensure that units included with certainty do not have an additive effect on the variance estimate. This may not seem clear at first because most of the variance estimators include the term $(1-\pi_i)$, which implies that if $\pi_i = 1$ the component of the sum for this unit would be zero. These units still, however, have an additive effect through other terms, for instance $\hat{Y}_{\bullet HT}$. Hence, the variance estimators must be modified to ensure any additive effect is removed.

As in Chapter 2, the two properties of the variance estimators that are computed to compare their behaviour are

(i) the relative bias (RB) and

(ii) the mean squared error (MSE).

The RB is considered as it is desirable for an estimator to be as nearly unbiased as possible. The MSE is considered as it provides a measure of how close the variance estimates are to the true variance. These properties are computed for each variance estimator $\hat{V}$ by

$$RB(\hat{V}) = \frac{E_s(\hat{V})}{V_T} - 1 = \frac{B(\hat{V})}{V_T} \tag{3.1}$$

$$MSE(\hat{V}) = E_s(\hat{V} - V_T)^2 = V_s(\hat{V}) + [B(\hat{V})]^2 \tag{3.2}$$

where E_s and V_s are the expectation and variance over the 50,000 samples, $B(\hat{V})$ is the bias of the variance estimator (see equation (1.9)) and V_T is the true variance being estimated. The expression "absolute RB" is defined here as the size of the relative bias irrespective of its sign. An unbiased estimator is desirable, however in many situations it is not possible to derive an exactly unbiased estimator. In addition, an unbiased estimator with a small MSE is usually preferred to an unbiased estimator with a large MSE. Therefore an exactly unbiased estimator is not always necessary. An estimator with a high bias should be avoided. A good balance is required between the MSE and bias of an estimator. As the MSE affects every single sample estimate, whereas the RB affect only the average over larger numbers

of samples, the former property is generally treated more seriously. Consequently the choice of the best estimator is subjective, depending on whether one prefers a nearly unbiased estimator or a more stable estimator.

Under RANSYS the value of V_T was approximated by the variance of 500,000 HTEs of the total, $Y_\bullet$, generated by RANSYS for each population. As the precision of the variance is inversely proportional to the number of estimates, this should be a precise approximation of the true variance. Under CPS, V_T is computed exactly by the SYG variance formula (1.14) using the exact joint inclusion probabilities calculated. Although V_T is only an approximation of the variance under RANSYS, for the purposes of this study it is treated as though it were the true variance.

The relative bias formula (3.1), was used as opposed to Matei and Tillé's formula, (2.37), because it provides a measure of the ratio of the bias of the variance estimator to the true variance being estimated, as opposed to the square root of the variance of the simulated estimates of the total. Formula (3.1) is very similar to Brewer and Donadio's relative bias formula, (2.36), however the true variance is determined over 500,000 estimates rather than only the 50,000 estimates from the simulation study.

3.4 The Matei and Tillé Study

The discrepancy in the results produced by Matei and Tillé is explained by analysing population (a), the same population they used in their study. As they used CPS, the variance of the HTE could be determined exactly. Table 3.2 shows their calculated variances for each sample size, compared with several other approximate values. This table also shows the exact variance calculated under the CPS algorithm described in this thesis (see section 2.1.2). The magnitudes of the true variance estimates and of the MSEs produced in this study are different from those given in Matei and Tillé's

paper due to the units of measurement used for each variable. In this study $P85$ is measured in thousands of people and $RMT85$ is measured in millions of kronors, whereas Matei and Tillé used a single unit measure for both. For example, when this study used $P85 = 10$, Matei and Tillé used $P85 = 10,000$.

In Table 3.2, V_{MT} is the variance produced in Matei and Tillé's paper, $\tilde{V}_{RANSYS}$ is the variance of 500,000 simulated estimates under RANSYS, and $\tilde{V}_{SYG-HR}$, $\tilde{V}_{SYG-HR2}$ and $\tilde{V}_{SYG-Haj}$ are the SYG variance estimator (1.14) with the joint inclusion probabilities approximated by (2.17), (2.18) with (2.19), and (2.20) respectively. The last two columns are of particular interest as they represent the exact variance calculated under CPS within this study, V_{CPS}, and the difference between this variance and V_{MT}.

n	V_{MT}	$\tilde{V}_{RANSYS}$	$\tilde{V}_{SYG-HR}$	$\tilde{V}_{SYG-HR2}$	$\tilde{V}_{SYG-Haj}$	V_{CPS}	Diff (%)
10	3817000	3962793.0	3963544.6	3962768.2	3934120.0	3958045	3.56
20	1782000	1851227.1	1848322.0	1851467.9	1832723.0	1843522	3.45
40	1007000	795022.2	792482.3	795817.8	775647.0	779899.5	29.12

Table 3.2: Comparison of the variances calculated by Matei and Tillé with those calculated within this thesis

Table 3.2 shows there are substantial discrepancies between the variances produced by Matei and Tillé and those produced in this study. There is almost a 30% difference[1] between V_{MT} and V_{CPS} for $n = 40$. In addition, especially when $n = 40$ V_{MT} is noticeably different from the other approximate variances calculated. As the approximated variances are similar to V_{CPS}, this provides strong evidence that this variance is accurate, as opposed to V_{MT}.

[1] $Diff = 100\frac{|V_{CPS} - V_{MT}|}{V_{CPS}}$

Table 3.2 also shows that $\tilde{V}_{SYG-HR}$ and $\tilde{V}_{SYG-HR2}$ are similar to V_{RANSYS}, and that $\tilde{V}_{Haj}$ is similar to V_{CPS}. This is expected because the first two approximations were designed under RANSYS and the latter approximation under CPS. This suggests that if RANSYS is being utilised one should use Hartley and Rao's approximation, whereas Hájek's approximation should be used for CPS. These similarities also indicate that the approximations of the joint inclusion probabilities used in the SYG variance estimator are acceptable. It is interesting to see, however, that Hartley and Rao's full approximation, $\tilde{V}_{SYG-HR}$, is not as accurate as the simpler third-order approximation, $\tilde{V}_{SYG-HR2}$, in estimating $\tilde{V}_{RANSYS}$. This result is not analysed further in this study, but should be considered for further research.

Upon examination of the algorithm for CPS design, it was discovered that the incorrect variances computed by Matei and Tillé for sample sizes of 10 and 20 could be reproduced. The much greater discrepancy for a sample size of 40 could not, however, be reproduced. The algorithm proposed by Matei and Tillé in their paper to implement CPS is the same as the algorithm described in this study, however, this is not the process that they used to calculate their variances. The process that appears to have been used by Matei and Tillé to calculate their variances is outlined as follows:

(i) Define the working probabilities, $\boldsymbol{p} = (p_1, \ldots, p_N)$, to be

$$p_i = \frac{nX_i}{\sum_{j \in U} X_j}. \tag{3.3}$$

(ii) Given $\boldsymbol{p}$, calculate the inclusion probabilities $\breve{\boldsymbol{\pi}} = \psi(\boldsymbol{p}, n)$ by equation (2.7).

(iii) Calculate the joint inclusion probabilities from $\boldsymbol{p}$ and $\breve{\boldsymbol{\pi}}$ by using (2.8).

(iv) Compute V_{SYG} using the above joint inclusion probabilities and $\breve{\boldsymbol{\pi}}$ as the first order inclusion probabilities.

In other words, it seems that what is usually defined as the desired inclusion

probabilities, $\boldsymbol{\pi}$, were actually used for the working probabilities. Technically there is nothing wrong with this method, as under CPS the working probabilities can be known rather than the desired inclusion probabilities. It is, however, generally the case in unequal probability sampling that the desired inclusion probabilities are defined by $\boldsymbol{\pi}$, to yield a small variance.

Table 3.3 shows the variances, V_{alg}, calculated under the above algorithm. The joint inclusion probabilities to determine V_{alg} were calculated using Aires algorithm, because the *sampling* package is only designed to use the desired inclusion probabilities and not the working probabilities. (The code to calculate these joint inclusion probabilities is given in Appendix A).

n	V_{MT}	V_{alg}
10	3817000	3817235.0
20	1782000	1782865.0
40	1007000	758932.3

Table 3.3: SYG variances for the algorithm used by Matei and Tillé

It is clear from Table 3.3 that the variance calculated by Matei and Tillé for $n = 40$ was not produced with the above algorithm, indicating there must be an additional error. The source of this additional error in the variance for $n = 40$ has not been traced. This error was not due to the change in methodology required when units are included with certainty as this did not occur for this population when $n = 40$.

As an aside, the similarity between V_{MT} and V_{CPS} is an indication of the relationship between the desired inclusion probabilities and the working probabilities. As explained earlier, the inclusion probabilities and the working probabilities are

similar when N is large (see equation (2.5)). As $N = 281$ is moderately large, the inclusion probabilities used by Matei and Tillé should be similar to their working probabilities, that is the desired inclusion probabilities used in V_{CPS}. This is why V_{MT} for the two smaller sample sizes, where the additional error did not occur, are similar to V_{CPS}.

The error in V_{MT} for $n = 40$ cannot, however, explain why the Brewer Family estimators have appreciably higher RBs compared with the Hájek-Deville Family estimators. The magnitude of the RB used by Matei and Tillé, (2.37) is determined by the numerator $E(\hat{V}) - V_{MT}$. The denominator will not have a noticeable effect when comparing the magnitude of the RBs due to the similarity between the MSEs of the variance estimators, which are approximations to the variances. As V_{MT} was larger than expected, $E(\hat{V}) - V_{MT}$ should be smaller indicating that the RBs, if calculated in this fashion, would have been even larger than those recorded by Matei and Tillé. Hence, this does not explain why the Brewer Family estimators perform poorly.

3.4.1 Matei and Tillé's sampling procedure

It is also important to determine the sampling procedure used for Matei and Tillé's simulations to determine whether this has had any effect on the results they produced. Although the algorithm described above used by Matei and Tillé to determine their variances is known, there is no conclusive evidence that this algorithm was used within the sampling procedure, as no sampling was needed to determine V_{MT}.

A small simulation was conducted to determine whether this algorithm was used by Matei and Tillé for their simulation study. The working probabilities defined in (3.3) were used to generate 10,000 independent samples using CPS for the two

sample sizes of $n = 10$ and $n = 20$. (The sample size of $n = 40$ was not included in these simulations because of the previously stated unknown error in Matei and Tillé's results). The number of simulations was chosen to be the same as that used by Matei and Tillé. The RBs and MSEs were determined for each variance estimator and compared with Matei and Tillé's original results (see Table 2.6 for their original results). For comparability, Matei and Tillé's relative bias formula, (2.37), was used for these simulations instead of (3.1).

As the results for one trial of 10,000 simulations were different from the results of Matei and Tillé, the process was repeated another two times to ensure that the differences were not simply due to simulation error. Table 3.4 show the RBs for the three trials. Trial 2 produces the closest results to those produced by Matei and Tillé. The order of the variance estimators across the three trials are similar for each sample size when they are ordered by size of the RBs. The problem, however, is that the preferred variance estimator defined as having the lowest absolute RB varies. $\hat{\tilde{V}}_{Dev2}$, $\hat{\tilde{V}}_{BR4}$ and $\hat{\tilde{V}}_{Ros}$ perform best in at least one of the trials. The results suggest, however, that $\hat{\tilde{V}}_{BR4}$ would be the preferred estimator in a larger simulation study, as this estimator usually has the lowest absolute RB.

Table 3.5 shows the MSEs of the three trials, where the MSE of $\hat{V}_{SYG}$ for a sample size of 10 is $4.8050 \times (10^{12})$, that is the column label 10 (10^{12}) represents the results of size 10^{12} for a sample size of 10. The order of the variance estimators for each sample size is also similar when they are ordered by the size of their MSEs. As the MSE is a positive measure, if the order of the variance estimators is the same for each trial then the preferred estimator will not change. The results suggest that $\hat{\tilde{V}}_{BR1}$ is the most stable for $n = 10$ and $\hat{\tilde{V}}_{Dev1}$ for $n = 20$ consistently across the three trials.

To measure the amount of variation of the RBs and MSEs for each variance

	Trial 1		Trial 2		Trial 3	
	10	20	10	20	10	20
$\hat{V}_{SYG}$	0.1665	-1.9929	-0.9568	-1.2473	0.1497	-1.2905
$\hat{\hat{V}}_{BR1}$	-0.0697	-1.8538	-1.1437	-1.0417	-0.0246	-1.1352
$\hat{\hat{V}}_{BR2}$	0.0469	-1.6581	-1.0246	-0.8501	0.0981	-0.9427
$\hat{\hat{V}}_{BR3}$	0.1632	-1.4628	-0.9057	-0.6588	0.2207	-0.7505
$\hat{\hat{V}}_{BR4}$	0.1762	**-1.4525**	**-0.8925**	**-0.6488**	0.2343	**-0.7404**
$\hat{\hat{V}}_{BR-Dev}$	0.0943	-1.4865	-0.9754	-0.6805	0.1471	-0.7719
$\hat{\hat{V}}_{Dev1}$	-0.0641	-2.3722	-1.1413	-1.5625	-0.0164	-1.6577
$\hat{\hat{V}}_{Dev2}$	**-0.0166**	-2.2005	-1.0922	-1.3928	0.0326	-1.4868
$\hat{\hat{V}}_{Ros}$	-0.0376	-2.2120	-1.1134	-1.4027	**0.0110**	-1.4972
$\hat{\hat{V}}_{Ber}$	0.1353	-2.1006	-0.9878	-1.3524	0.1182	-1.3972

Table 3.4: RB (%) for three trials of 10,000 simulations

estimator, the variance across the three trials for $n = 10$ and the variance across the three trials for $n = 20$ was evaluated for both the RB and the MSE. The variances for the MSE were calculated over the values shown in Table 3.5 excluding the constant size, for example for $n = 10$ and $\hat{\hat{V}}_{BR1}$ the variance of 6.2349, 5.9358 and 5.9325 was calculated, not the variance of 6.2349×10^{12}, 5.9358×10^{12} and 5.9325×10^{12}. This was done to indicate that the variances were small relative to the size of the MSEs. Table 3.6 shows that the variances of the RBs for $n = 10$ is large compared to the actual RBs. This could easily cause the RB to vary from negative to positive values across different sets of 10,000 simulations, thus affecting the preferred estimator.

Table 3.6 also indicates that as the sample size increases from 10 to 20, the amount of variability measured over the three simulations decreases. A larger sample size should provide more precise variance estimators, and hence less variability. As

	Trial 1		Trial 2		Trial 3	
	10 (10^{12})	20 (10^{11})	10 (10^{12})	20 (10^{11})	10 (10^{12})	20 (10^{11})
$\hat{V}_{SYG}$	6.2929	5.9148	5.9939	6.0780	5.9750	5.9843
$\hat{\hat{V}}_{BR1}$	**6.2349**	5.8038	**5.9358**	5.9907	**5.9325**	5.8857
$\hat{\hat{V}}_{BR2}$	6.2441	5.8139	5.9450	6.0019	5.9410	5.8954
$\hat{\hat{V}}_{BR3}$	6.2533	5.8241	5.9543	6.0132	5.9495	5.9052
$\hat{\hat{V}}_{BR4}$	6.2544	5.8246	5.9553	6.0137	5.9505	5.9058
$\hat{\hat{V}}_{BR-Dev}$	6.2476	5.8213	5.9482	6.0100	5.9443	5.9033
$\hat{\hat{V}}_{Dev1}$	6.2405	**5.8022**	5.9414	**5.9876**	5.9375	**5.8818**
$\hat{\hat{V}}_{Dev2}$	6.2440	5.8095	5.9446	5.9956	5.9409	5.8895
$\hat{\hat{V}}_{Ros}$	6.2414	5.8051	5.9422	5.9911	5.9384	5.8852
$\hat{\hat{V}}_{Ber}$	6.2881	5.9051	5.9893	6.0682	5.9706	5.9747

Table 3.5: MSE for three trials of 10,000 simulations

the variances across the three trials only have two degrees of freedom these results are only suggestive of the amount of variability. It is interesting to observe that the amount of variation is similar among the variance estimators for each sample size, indicating a similar profile of the variance estimators. The profile of the variance estimators was constructed by subtracting the smallest RB (or MSE) for each sample size from the RB (or MSE) of each variance estimator. Table 3.7 shows that the RB profiles for a given sample size are similar, indicating that the order of the estimators, in terms of the size of their RBs, are similar for each trial, and that the distance between the values of the RBs are also similar. The range between the variance estimator with maximum RB and the one with the minimum RB is similar for a given sample size. Hence, the estimator with the lowest absolute RB will depend on how the range is located relative to zero. The profiles for the MSEs show the same pattern and are shown in Table B.1 (Appendix B).

	RB		MSE	
	10	20	10	20
$\hat{V}_{SYG}$	0.4144	0.1752	0.0318	0.0067
$\hat{\hat{V}}_{BR1}$	0.4013	0.1974	0.0302	0.0088
$\hat{\hat{V}}_{BR2}$	0.4018	0.1955	0.0302	0.0089
$\hat{\hat{V}}_{BR3}$	0.4024	0.1937	0.0303	0.0090
$\hat{\hat{V}}_{BR4}$	0.4025	0.1936	0.0303	0.0090
$\hat{\hat{V}}_{BR-Dev}$	0.4012	0.1948	0.0303	0.0090
$\hat{\hat{V}}_{Dev1}$	0.4047	0.1958	0.0302	0.0086
$\hat{\hat{V}}_{Dev2}$	0.4040	0.1951	0.0303	0.0087
$\hat{\hat{V}}_{Ros}$	0.4040	0.1958	0.0302	0.0087
$\hat{\hat{V}}_{Ber}$	0.4142	0.1761	0.0317	0.0067

Table 3.6: Variances of the RBs and the MSEs across the 3 trials of 10,000 simulations

In conclusion, this study indicates that 10,000 simulations is not large enough to produce consistent estimates of the RBs of variance estimators for a given population. It is largely for this reason that the sampling method used by Matei and Tillé cannot be determined. The results also indicates that similar studies with only 10,000 simulations should be treated with caution. This does not, however, imply the RBs and the MSEs are incorrect, but that these results are constrained to describing the behaviour of the variance estimators for that particular set of 10,000 samples only and not other samples. That is, the performance of the variance estimators for a set of 10,000 simulations cannot be extrapolated to describe how well the competing variance estimators might perform over the whole population.

	n=10			n=20		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
$\hat{V}_{SYG}$	0.2362	0.1869	0.1743	0.3672	0.3793	0.3152
$\hat{\hat{V}}_{BR1}$	0.0000	0.0000	0.0000	0.5225	0.5184	0.5208
$\hat{\hat{V}}_{BR2}$	0.1166	0.1191	0.1227	0.7150	0.7141	0.7124
$\hat{\hat{V}}_{BR3}$	0.2329	0.2380	0.2453	0.9072	0.9094	0.9037
$\hat{\hat{V}}_{BR4}$	0.2458	0.2512	0.2589	0.9173	0.9197	0.9137
$\hat{\hat{V}}_{BR-Dev}$	0.1640	0.1683	0.1717	0.8858	0.8857	0.8820
$\hat{\hat{V}}_{Dev1}$	0.0056	0.0023	0.0082	0.0000	0.0000	0.0000
$\hat{\hat{V}}_{Dev2}$	0.0531	0.0515	0.0572	0.1709	0.1717	0.1697
$\hat{\hat{V}}_{Ros}$	0.0321	0.0303	0.0356	0.1605	0.1602	0.1598
$\hat{\hat{V}}_{Ber}$	0.2050	0.1558	0.1428	0.2606	0.2715	0.2101

Table 3.7: Profile of the RBs for three trials of 10,000 simulations

3.5 Simulation Results

In this section the Brewer Family and Hájek-Deville Family of variance estimators are compared. The SYG variance estimator (1.15) is also considered to determine whether the knowledge of the exact joint inclusion probabilities significantly improves this estimator compared with the other approximate estimators. In this section the RB and MSE of each of the variance estimators being evaluated is presented for each population and each sample size. In each table of results, for each sampling situation, the smallest absolute RB and smallest MSE for the nine approximate variance estimators is highlighted in bold to emphasise the estimator that appears in this instance to be the best approximate variance estimator. If the SYG variance estimator has a smaller absolute RB or MSE then this value is also highlighted. This ensures that the best approximate variance estimator can be easily

identified in each situation, and that the SYG is also identified if it is better than all the approximate estimators.

The SYG estimator should be unbiased for CPS as all the joint inclusion probabilities are greater than zero and known exactly. Hence its true RBs are all zero. The RBs of the SYG estimator in these results, however, are not zero due to the size of the simulation study. If all possible samples had been simulated then the corresponding RB would be zero.

3.5.1 Summary results for all populations

The true variances to be estimated for each population are shown in this section. The true variances were determined as described earlier in the methodology section, hence the "true variance" under RANSYS is only an approximation, whereas it is exact for CPS. Table 3.8 shows the true variances for populations (a), (b) and (c) for the two sampling algorithms.

	RANSYS			CPS		
Population	(a)	(b)	(c)	(a)	(b)	(c)
	(10^6)	(10^6)	(10^4)	(10^6)	(10^6)	(10^4)
n=10	3.9628	5.6569	4.4956	3.9580	5.6539	4.4731
n=20	1.8512	2.6425	1.9644	1.8435	2.6385	1.9653
n=40	0.7950	1.1415	0.8192	0.7799	1.1248	0.8184

Table 3.8: RANSYS and CPS true variances - populations (a) to (c)

Tables 3.9 and 3.10 show the true variances for the remaining populations for RANSYS and CPS respectively. Populations (d), (e) and (g) have the same study variable but differing auxiliary variables. The correlation between the study variable

and the auxiliary variable is the highest for population (d) and the lowest for population (g). Hence the variance for population (d) is smaller than for populations (e) and (g). The results from each table indicates that the true variance of an estimator is clearly dependent upon the population structure, and especially on the correlation between X and Y.

Population	(d)	(e)	(f)	(g)
	(10^{10})	(10^{11})	(10^{6})	(10^{11})
n=10	2.0423	1.8048	1.5765	8.1520
n=20	0.7229	0.4406	0.7632	2.8744
n=40	0.1603	0.1028	0.3600	0.7944

Table 3.9: RANSYS true variances - populations (d) to (g)

Population	(d)	(e)	(f)	(g)
	(10^{10})	(10^{11})	(10^{6})	(10^{11})
n=10	2.0470	1.8133	1.6431	8.1282
n=20	0.7206	0.4428	0.7981	2.8705
n=40	0.1607	0.1002	0.3748	0.7853

Table 3.10: CPS true variances - populations (d) to (g)

An interesting result is the relationship between the true variances under RANSYS and CPS. It was originally believed that since CPS maximises the entropy it should produce higher variances than RANSYS. This, however, is not consistently observed in this section; for the twenty-one different variances, CPS only has a larger variance eight times. This result will be discussed in detail in chapter 4.

3.5.2 Effects of simulation size

Before comparing the variance estimators it is important to consider how the number of simulations conducted affects the results. It was initially assumed that 50,000 simulations would be sufficient to provide precise estimates of the RBs and the MSEs, but that was prior to finding that 10,000 simulations was not sufficient. This section examines whether 50,000 simulations is large enough. From the previous analysis with 10,000 simulations is was clear that the major concern is with the RB of the variance estimators, hence the relevant results for the MSE are provided in Appendix C.

Three independent simulations of 50,000 samples were generated for one population from each data set; populations (a) and (d) were chosen. The simulations were conducted under RANSYS only as this method is faster to implement. It was assumed that the results would be similar for CPS as that is also a high entropy sampling algorithm. The SYG estimator was excluded in order to reduce computation time. The RBs over the three trials for both populations (see Tables C.3 and C.7 in Appendix C) indicate that there are still some inconsistencies among the RBs, as different variance estimators are preferred under the three trials. Table 3.11 shows the variability of the RBs across the three trials for each variance estimator under both populations. These variances are smaller compared to the results for 10,000 simulations (see Table 3.6) indicating that 50,000 simulations is better than 10,00 simulations, as expected.

Once again the variance is basically the same for each variance estimator, and it decreases as the sample size increases. This similarity indicates that the variance estimators vary the same amount between samples. As the variances here only have two degrees of freedom these results are only suggestive. However, more

	Population (a)			Population (d)		
	10	20	40	10	20	40
$\hat{V}_{BR1}$	0.0784	0.0599	0.0167	0.0230	0.0540	0.0055
$\hat{V}_{BR2}$	0.0784	0.0600	0.0168	0.0252	0.0562	0.0058
$\hat{V}_{BR3}$	0.0785	0.0601	0.0170	0.0274	0.0585	0.0060
$\hat{V}_{BR4}$	0.0785	0.0601	0.0170	0.0277	0.0587	0.0060
$\hat{V}_{BR-Dev}$	0.0785	0.0600	0.0169	0.0259	0.0577	0.0060
$\hat{V}_{Dev1}$	0.0784	0.0603	0.0162	0.0290	0.0565	0.0051
$\hat{V}_{Dev2}$	0.0785	0.0603	0.0162	0.0300	0.0580	0.0053
$\hat{V}_{Ros}$	0.0784	0.0602	0.0163	0.0283	0.0565	0.0052
$\hat{V}_{Ber}$	0.0765	0.0696	0.0173	0.0620	0.0548	0.0048

Table 3.11: Variances of the RBs across 3 trials of 50,000 simulations

conclusive results can be seen by examining the profile of the variance estimators as accomplished for 10,000 simulations. Tables 3.12 shows the profile for $n = 10$ in populations (a) and (d). Once again the profiles are similar for the given sample size. The profiles were also similar for other sample sizes and for the MSEs (see Tables C.1, C.2, C.5 and C.6 in Appendix C)

In conclusion, there are inconsistencies in the RBs over sets of 50,000 simulations, however overall is it considerably better than 10,000 simulations. There is no concern in relation to comparing the MSE for 50,000 simulations. The only way to remove the variation of the RBs across the simulations would be to simulate every possible sample which clearly is not feasible. The RBs of variance estimators for a set of 50,000 simulations correctly describe the behaviour of the estimators but for the selected samples only. Hence, the results are still meaningful, but they will vary for other simulation studies under the same population. Therefore, the behaviour

	Population (a)			Population (d)		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
$\hat{V}_{BR1}$	0.0220	0.0207	0.0217	0	0	0
$\hat{V}_{BR2}$	0.0809	0.0808	0.0807	2.8614	2.8522	2.8659
$\hat{V}_{BR3}$	0.1397	0.1408	0.1397	5.7229	5.7043	5.7318
$\hat{V}_{BR4}$	0.1463	0.1475	0.1463	6.0805	6.0608	6.0900
$\hat{V}_{BR-Dev}$	0.1115	0.1116	0.1112	5.2929	5.2825	5.3017
$\hat{V}_{Dev1}$	0	0	0	0.9790	0.9633	1.0032
$\hat{V}_{Dev2}$	0.0306	0.0308	0.0304	3.3579	3.3407	3.3869
$\hat{V}_{Ros}$	0.0193	0.0193	0.0192	1.2877	1.2740	1.3095
$\hat{V}_{Ber}$	0.0708	0.0607	0.0751	4.6109	4.4331	4.6001

Table 3.12: Profile of the RBs for three trials of 50,000 simulations - populations (a) and (d)

of variance estimators under 50,000 simulations cannot be extrapolated to provide conclusive evidence of the behavior of the estimators in the whole population. For the main simulation study 50,000 simulations were used as this was feasible given the time frame. The results of this section are considered when comparing the RBs, in the sense that the results are not conclusive for the behaviour over the entire population only the given samples simulated.

3.5.3 Population (a)

Population (a) was the population used to reproduce Matei and Tillé's results. Tables 3.13 and 3.14 show the RBs and MSEs respectively for the two sampling algorithms considered.

	RANSYS			CPS		
	10	20	40	10	20	40
$\hat{V}_{SYG}$	0.2435	0.1933	0.4043	0.2470	0.0677	0.1225
$\hat{\hat{V}}_{BR1}$	0.0460	-0.0764	**-0.0040**	0.1811	0.2298	1.6608
$\hat{\hat{V}}_{BR2}$	0.1049	**-0.0094**	0.0772	0.2403	0.2972	1.7432
$\hat{\hat{V}}_{BR3}$	0.1637	0.0577	0.1584	0.2995	0.3646	1.8256
$\hat{\hat{V}}_{BR4}$	0.1703	0.0612	0.1605	0.3060	0.3681	1.8277
$\hat{\hat{V}}_{BR-Dev}$	0.1355	0.0626	0.2814	0.2711	0.3695	1.9508
$\hat{\hat{V}}_{Dev1}$	**0.0240**	-0.3484	-1.7212	**0.1589**	-0.0443	-0.0858
$\hat{\hat{V}}_{Dev2}$	0.0546	-0.2767	-1.5208	0.1897	0.0277	0.1180
$\hat{\hat{V}}_{Ros}$	0.0433	-0.2725	-1.4128	0.1784	0.0322	0.2277
$\hat{\hat{V}}_{Ber}$	0.0948	-0.2775	-1.6416	0.2297	**0.0266**	**-0.0078**

Table 3.13: Relative biases (%) - population (a)

In relation to the RBs of the approximate variance estimators, it is difficult to determine a preferred estimator across both sampling algorithms. The Brewer Family estimators, and in particular $\hat{\hat{V}}_{BR1}$ and $\hat{\hat{V}}_{BR2}$ have the lowest absolute RBs for $n = 40$ and $n = 20$ respectively, under RANSYS. This is interesting in regards to Brewer's belief that $\hat{\hat{V}}_{BR3}$ and $\hat{\hat{V}}_{BR4}$ should be more accurate than $\hat{\hat{V}}_{BR1}$ and $\hat{\hat{V}}_{BR2}$ (Brewer, 2002, p. 153, 158). This result, however, may be due to the variability in 50,000 simulations. In the previous section this population was analysed over three sets of 50,000 simulations under RANSYS. In Table C.3 the results for Trial 2 and Trial 3 indicate that $\hat{\hat{V}}_{BR3}$ and $\hat{\hat{V}}_{BR4}$ are more accurate for this population (Note Trial 1 in Table C.3 is the same as Table 3.13).

For CPS, the Hájek-Deville Family performs better than the Brewer Family, with $\hat{\hat{V}}_{Dev1}$ and $\hat{\hat{V}}_{Ber}$ having the lowest absolute RB, although both have a slightly

negative RB. For both sampling designs it is clear that the Brewer Family estimators have similar properties to each other, and the Hájek-Deville Family estimators have similar properties. In addition, the difference between the behaviour of these groups tends to increase as the sample size increases. That is, when $n = 10$ the estimators have similar RBs, however when $n = 40$ there is a clear difference between the RBs of the two families.

In regards to $\hat{V}_{SYG}$, this estimator never has the lowest absolute RB, however, it does perform consistently well under both sampling designs and across all sample sizes. It also has the desirable property that the RB is always positive, as it is better to overestimate the variance than underestimate it. To underestimate the variance is to claim a greater accuracy than actually obtained. It appears that the knowledge of the joint inclusion probabilities does not significantly improve the variance estimation process, but that it still provides a good estimator. Overall the MSEs of the estimators for both sampling algorithm are similar, however, $\hat{V}_{SYG}$ is always achieves the highest MSE followed by $\hat{\hat{V}}_{Ber}$. Excluding $\hat{\hat{V}}_{Ber}$, the MSEs of the Hájek-Deville Family estimators are usually lower than the Brewer Family estimators.

These results do not need to be directly compared to those produced by Matei and Tillé, as section 3.4.1 indicates that their results cannot be replicated, due to the small simulation size and the different algorithm used. The Brewer Family estimators have higher absolute RBs than the Hájek-Deville Family under CPS when $n = 40$, however the differences are not extreme as were those produced by Matei and Tillé. This provides more support that Matei and Tillé's results are misleading.

	RANSYS			CPS		
	10 (10^{12})	20 (10^{11})	40 (10^{10})	10 (10^{12})	20 (10^{11})	40 (10^{10})
$\hat{V}_{SYG}$	6.4598	6.3257	4.9952	6.3276	6.2229	4.9171
$\hat{\hat{V}}_{BR1}$	**6.3950**	6.2152	4.7785	**6.2753**	6.1440	4.8166
$\hat{\hat{V}}_{BR2}$	6.4036	6.2254	4.7913	6.2839	6.1542	4.8309
$\hat{\hat{V}}_{BR3}$	6.4122	6.2358	4.8042	6.2927	6.1646	4.8454
$\hat{\hat{V}}_{BR4}$	6.4132	6.2364	4.8046	6.2936	6.1652	4.8457
$\hat{\hat{V}}_{BR-Dev}$	6.4071	6.2339	4.8104	6.2874	6.1628	4.8544
$\hat{\hat{V}}_{Dev1}$	6.3991	**6.2076**	**4.7370**	6.2795	**6.1359**	**4.7357**
$\hat{\hat{V}}_{Dev2}$	6.4026	6.2159	4.7512	6.2830	6.1442	4.7542
$\hat{\hat{V}}_{Ros}$	6.4002	6.2119	4.7475	6.2806	6.1403	4.7533
$\hat{\hat{V}}_{Ber}$	6.4422	6.2837	4.9095	6.3229	6.2132	4.8989

Table 3.14: Mean squared errors - population (a)

3.5.4 Population (b)

The next population considered is population (b) which was used by Brewer and Donadio. Tables 3.15 and 3.16 show the RBs and MSEs for this population respectively. For the RBs, it is once again the case that the Brewer Family tends to perform better under RANSYS and the Hájek-Deville Family estimators under CPS.

Regarding the absolute RBs, $\hat{\hat{V}}_{BR3}$ and $\hat{\hat{V}}_{BR4}$ tend to perform better across all sample sizes under RANSYS, which is expected according to Brewer and Donadio. There are, however, a few differences in the RBs for CPS. Under this sampling algorithm $\hat{\hat{V}}_{BR1}$ has the lowest absolute RB for $n = 10$ yet performs poorly for $n = 40$, and $\hat{\hat{V}}_{Ros}$ is performing well across all sample sizes.

For $\hat{V}_{SYG}$ the absolute RB decreases as the sample size increases for both sampling algorithms. This may be indicating that this estimator performs better for large populations. This is expected for RANSYS because the approximation of the joint inclusion probabilities should be more precise as the sample size increases, provided that the the population size is large compared with the sample size. In comparison to the other approximate estimators, $\hat{V}_{SYG}$ was always comparable with the estimator with the lowest absolute RB for both sampling algorithms.

	RANSYS			CPS		
	10	20	40	10	20	40
$\hat{V}_{SYG}$	**-0.3201**	0.2268	**-0.0753**	0.1977	0.1306	0.0231
$\hat{\hat{V}}_{BR1}$	-0.5230	-0.0546	-0.5035	**0.0990**	0.1680	0.9506
$\hat{\hat{V}}_{BR2}$	-0.4587	**0.0211**	-0.4140	0.1644	0.2430	1.0403
$\hat{\hat{V}}_{BR3}$	-0.3943	0.0968	-0.3246	0.2297	0.3180	1.1300
$\hat{\hat{V}}_{BR4}$	**-0.3872**	0.1008	-0.3223	0.2369	0.3219	1.1323
$\hat{\hat{V}}_{BR-Dev}$	-0.4283	0.0924	**-0.2128**	0.1948	0.3145	1.2447
$\hat{\hat{V}}_{Dev1}$	-0.5210	-0.2239	-1.6503	0.1022	**-0.0031**	-0.2214
$\hat{\hat{V}}_{Dev2}$	-0.4907	-0.1527	-1.4517	0.1326	0.0681	-0.0196
$\hat{\hat{V}}_{Ros}$	-0.5061	-0.1688	-1.4332	0.1170	0.0522	**-0.0002**
$\hat{\hat{V}}_{Ber}$	-0.4493	-0.1496	-1.5547	0.1810	0.0889	-0.1142

Table 3.15: Relative biases (%) - population (b)

The MSE of $\hat{V}_{SYG}$ is once again the largest followed by $\hat{\hat{V}}_{Ber}$, and $\hat{\hat{V}}_{BR1}$ has the smallest MSE for all but one of the cases. If $\hat{V}_{SYG}$ and $\hat{\hat{V}}_{Ber}$ are excluded the differences of the MSEs among the remaining estimators are trivial. The MSEs of the Hájek-Deville Family estimators tend to be smaller than most of those of the Brewer Family estimators.

	RANSYS			CPS		
	10 (10^{12})	20 (10^{11})	40 (10^{10})	10 (10^{12})	20 (10^{11})	40 (10^{10})
$\hat{V}_{SYG}$	9.6353	9.6618	8.3672	9.7012	9.8061	8.2684
$\hat{\hat{V}}_{BR1}$	**9.5027**	**9.4012**	**7.8744**	**9.5756**	**9.5645**	7.8625
$\hat{\hat{V}}_{BR2}$	9.5271	9.4322	7.9123	9.6004	9.5965	7.9034
$\hat{\hat{V}}_{BR3}$	9.5519	9.4635	7.9507	9.6254	9.6290	7.9447
$\hat{\hat{V}}_{BR4}$	9.5546	9.4652	7.9517	9.6282	9.6307	7.9458
$\hat{\hat{V}}_{BR-Dev}$	9.5322	9.4444	7.9400	9.6056	9.6092	7.9389
$\hat{\hat{V}}_{Dev1}$	9.5228	9.4185	7.9140	9.5957	9.5815	**7.8566**
$\hat{\hat{V}}_{Dev2}$	9.5278	9.4304	7.9351	9.6009	9.5939	7.8853
$\hat{\hat{V}}_{Ros}$	9.5239	9.4220	7.9160	9.5969	9.5853	7.8671
$\hat{\hat{V}}_{Ber}$	9.6136	9.6150	8.3297	9.6935	9.7884	8.2331

Table 3.16: Mean squared errors - population (b)

The above results are similar to those produced by Brewer and Donadio (see Table 2.6), however not identical. For example, $\hat{\hat{V}}_{BR2}$ has the lowest absolute RB for $n = 20$ in this study, whereas $\hat{\hat{V}}_{BR3}$ and $\hat{\hat{V}}_{BR4}$ have the lowest absolute RB for Brewer and Donadio's results. These differences should not be due to the different RB formulas used because (3.1) and (2.36) are similar. For clarity, however, the RB results are also produced using Brewer and Donadio's relative bias formula (2.36). These results are shown in Appendix D for RANSYS only as this was the algorithm used by Brewer and Donadio. These results are very similar to Table 3.15 as expected. The differences between the results of this study and those of Brewer and Donadio are most likely due to the different sets of 50,000 simulations used in the two studies.

3.5.5 Further populations

Populations (c) to (e) encountered the problem that for some units $nx_i > \sum_{j \in U} x_j$, hence these units were included with certainty. Table 3.17 shows the number of units included with certainty, n_E, for each population. For instance, to ensure that a sample of 40 units are obtained for population (e) only 24 units are selected from the population excluding the 16 units included with certainty. (The tables of results for these populations are at the end the of this section due to their size.)

$\mathbf{n_E}$	Pop (c)	Pop (d)	Pop (e)
10	0	1	3
20	2	7	7
40	3	18	16

Table 3.17: Number of units included with certainty
- populations (c) to (e)

Tables 3.18 and 3.19 show the RBs under RANSYS and CPS respectively. For populations (c) and (d), $\hat{\tilde{V}}_{BR3}$ and $\hat{\tilde{V}}_{BR4}$, perform well across all the sample sizes for both sampling algorithms. These two variance estimators also perform well in population (e) under CPS. It is difficult to choose a preferred estimator for population (e) which has the lowest absolute RB for all sample sizes under RANSYS. In relation to comparing the performance of the two families, the Brewer Family estimators clearly perform better than the Hájek-Deville Family estimators for CPS in both populations (d) and (e). The performance of the two families is similar for population (c), however, the Brewer Family estimators are slightly better. This casts some doubt on the hypothesis that the Hájek-Deville Family variance estimators perform better under this sampling algorithm. As populations (d) and (e) are from the CO124 population, this may indicate that the performance of the estimators is more dependent upon the population than upon the sampling algorithm.

In terms of $\hat{V}_{SYG}$, this estimator performed quite well under populations (c) and (d) for both sampling designs, especially for CPS, with regard to the RB. In population (e), $\hat{V}_{SYG}$ again performed consistently well under CPS except for $n = 40$.

Tables 3.20 and 3.21 show the MSE for populations (c) to (e) under RANSYS and CPS respectively. $\hat{V}_{SYG}$ and $\hat{\tilde{V}}_{Ber}$ again have the two largest MSE values under both sampling algorithms for populations (c) and (d). $\hat{\tilde{V}}_{BR1}$ has the smallest MSE in each situation under both sampling algorithms except for population (e) when $n = 40$; when $\hat{\tilde{V}}_{Dev1}$ had the smallest MSE here. Once again the Hájek-Deville estimators generally have a lower MSE than the Brewer Family estimators for each population and the two sampling algorithms.

	Pop (c)			Pop (d)			Pop (e)		
	10	20	40	10	20	40	10	20	40
$\hat{V}_{SYG}$	0.4443	0.2135	**0.0400**	1.0066	-0.0456	1.3210	4.6714	0.8129	2.6384
$\hat{\hat{V}}_{BR1}$	-1.1662	-0.3372	-0.7671	-5.8835	-4.2768	-1.7398	-3.1708	-2.1234	-1.7549
$\hat{\hat{V}}_{BR2}$	-0.5188	-0.1487	-0.4951	-3.0221	-2.5596	-0.8813	-0.0729	-1.2521	**0.0241**
$\hat{\hat{V}}_{BR3}$	**0.1287**	0.0398	-0.2231	**-0.1606**	-0.8424	-0.0229	3.0250	-0.3808	1.8032
$\hat{\hat{V}}_{BR4}$	0.2006	0.0508	**-0.2156**	0.1970	**-0.6993**	**0.0144**	3.5414	-0.3082	1.8879
$\hat{\hat{V}}_{BR-Dev}$	0.3160	-0.0231	-0.3197	-0.5906	-0.9946	0.4458	3.6625	**-0.1311**	2.2951
$\hat{\hat{V}}_{Dev1}$	-1.3174	-0.2884	-0.5889	-4.9045	-3.4478	-1.5573	-3.5951	-2.4112	-5.7585
$\hat{\hat{V}}_{Dev2}$	-0.4909	-0.1629	-0.4136	-2.5256	-1.8982	-0.2395	**-0.0179**	-1.3038	-3.6201
$\hat{\hat{V}}_{Ros}$	-1.2188	-0.2658	-0.5704	-4.5958	-3.3321	-1.4296	-3.0151	-2.2748	-5.1383
$\hat{\hat{V}}_{Ber}$	-0.7841	**-0.0007**	-0.3316	-1.2726	-1.9215	-0.6988	0.1183	-0.7795	-3.4208

Table 3.18: Relative biases (%) for RANSYS - populations (c) to (e)

	Pop (c)			Pop (d)			Pop (e)		
	10	20	40	10	20	40	10	20	40
$\hat{V}_{SYG}$	**0.0394**	-0.0440	**-0.3415**	**-0.3876**	**-0.1691**	0.2408	-0.8078	0.8560	-3.7727
$\hat{\hat{V}}_{BR1}$	-1.0331	-0.4298	-0.9201	-6.5976	-3.9525	-1.8578	-6.1704	-1.5065	-3.3830
$\hat{\hat{V}}_{BR2}$	-0.3900	-0.2421	-0.6492	-3.7642	-2.2297	-0.9973	-3.2095	-0.6052	-1.6644
$\hat{\hat{V}}_{BR3}$	**0.2530**	-0.0543	-0.3783	-0.9307	-0.5070	-0.1368	-0.2486	0.2961	**0.0541**
$\hat{\hat{V}}_{BR4}$	0.3245	**-0.0432**	**-0.3708**	**-0.5765**	**-0.3634**	**-0.0994**	**0.2449**	0.3712	0.1360
$\hat{\hat{V}}_{BR-Dev}$	0.4472	-0.1156	-0.4737	-1.2917	-0.6453	0.3377	0.5262	0.5340	0.5897
$\hat{\hat{V}}_{Dev1}$	-1.1870	-0.3819	-0.7416	-5.6604	-3.1240	-1.6861	-6.6938	-1.7665	-7.4879
$\hat{\hat{V}}_{Dev2}$	-0.3580	-0.2557	-0.5663	-3.2410	-1.5550	-0.3606	-3.1144	-0.6410	-5.3683
$\hat{\hat{V}}_{Ros}$	-1.0893	-0.3594	-0.7234	-5.3509	-3.0076	-1.5554	-6.1246	-1.6287	-6.8684
$\hat{\hat{V}}_{Ber}$	-0.6722	-0.1232	-0.5027	-2.8673	-1.8285	-0.9816	-4.0773	**-0.1347**	-5.8963

Table 3.19: Relative biases (%) for CPS - populations (c) to (e)

	Pop (c)			Pop (d)			Pop (e)		
	10 (10^9)	20 (10^8)	40 (10^7)	10 (10^{20})	20 (10^{19})	40 (10^{17})	10 (10^{23})	20 (10^{22})	40 (10^{20})
$\hat{V}_{SYG}$	1.4367	1.3581	1.1950	5.1495	2.6187	4.5366	7.8611	1.8862	8.5025
$\hat{\hat{V}}_{BR1}$	**1.3489**	**1.3055**	**1.1337**	**4.1502**	**2.2413**	**3.7613**	**6.3358**	**1.6789**	7.5879
$\hat{\hat{V}}_{BR2}$	1.3772	1.3183	1.1447	4.4596	2.3405	3.8769	6.9522	1.7671	7.9760
$\hat{\hat{V}}_{BR3}$	1.4060	1.3312	1.1560	4.7879	2.4452	3.9984	7.5979	1.8576	8.3745
$\hat{\hat{V}}_{BR4}$	1.4092	1.3320	1.1563	4.8303	2.4542	4.0039	7.7084	1.8652	8.3937
$\hat{\hat{V}}_{BR-Dev}$	1.3997	1.3215	1.1485	4.6711	2.4091	3.9764	7.4749	1.8076	8.3462
$\hat{\hat{V}}_{Dev1}$	1.3616	1.3148	1.1426	4.3980	2.3440	3.8424	6.5196	1.7291	**7.3271**
$\hat{\hat{V}}_{Dev2}$	1.3834	1.3179	1.1463	4.6016	2.4111	3.9363	7.0035	1.7686	7.6638
$\hat{\hat{V}}_{Ros}$	1.3636	1.3154	1.1430	4.4108	2.3452	3.8464	6.5934	1.7378	7.4240
$\hat{\hat{V}}_{Ber}$	1.4105	1.3531	1.1944	5.2025	2.6352	4.6245	7.5428	1.8759	8.2876

Table 3.20: Mean squared errors for RANSYS - populations (c) to (e)

	Pop (c)			Pop (d)			Pop (e)		
	10 (10^9)	20 (10^8)	40 (10^7)	10 (10^{20})	20 (10^{19})	40 (10^{17})	10 (10^{23})	20 (10^{22})	40 (10^{20})
$\hat{V}_{SYG}$	1.4312	1.3609	1.1650	5.1165	2.6352	4.4762	5.9578	2.2460	6.5343
$\hat{\tilde{V}}_{BR1}$	**1.3460**	**1.3139**	**1.1049**	**4.0348**	**2.2078**	**3.6823**	**5.1354**	**1.9353**	6.0119
$\hat{\tilde{V}}_{BR2}$	1.3741	1.3267	1.1156	4.3326	2.3058	3.7939	5.6297	2.0369	6.3178
$\hat{\tilde{V}}_{BR3}$	1.4027	1.3397	1.1264	4.6488	2.4093	3.9114	6.1474	2.1412	6.6320
$\hat{\tilde{V}}_{BR4}$	1.4059	1.3405	1.1267	4.6896	2.4182	3.9167	6.2360	2.1500	6.6472
$\hat{\tilde{V}}_{BR-Dev}$	1.3965	1.3300	1.1192	4.5484	2.3765	3.8928	6.1359	2.0843	6.6198
$\hat{\tilde{V}}_{Dev1}$	1.3585	1.3231	1.1135	4.2682	2.3078	3.7597	5.2143	1.9941	**5.7764**
$\hat{\tilde{V}}_{Dev2}$	1.3803	1.3263	1.1172	4.4759	2.3769	3.8528	5.6775	2.0404	6.0494
$\hat{\tilde{V}}_{Ros}$	1.3604	1.3238	1.1140	4.2816	2.3092	3.7636	5.2859	2.0037	5.8576
$\hat{\tilde{V}}_{Ber}$	1.4078	1.3572	1.1602	4.8450	2.5421	4.3405	5.5527	2.1958	6.2509

Table 3.21: Mean squared errors for CPS - populations (c) to (e)

3.5.6 Low correlation populations

The remaining two populations (populations (f) and (g)) have a low correlation between the study variable and the auxiliary variable. Table 3.22 represent the number of units included with certainty for both populations.

$\mathbf{n_E}$	Pop (f)	Pop (g)
10	0	2
20	0	5
40	0	15

Table 3.22: Number of units included with certainty - populations (f) and (g)

Tables 3.23 and 3.24 show the RBs for RANSYS and CPS respectively. For population (f), $\hat{\tilde{V}}_{BR1}$ and $\hat{\tilde{V}}_{BR2}$ perform well under RANSYS and $\hat{\tilde{V}}_{Ber}$ under CPS. For population (g) $\hat{\tilde{V}}_{BR4}$ has the lowest absolute RB for all sample sizes under RANSYS. For CPS, however, it is difficult to choose a preferred estimator. Finally, $\hat{V}_{SYG}$ has a low absolute RB in all situations.

Overall, the low correlation has not effected the behaviour of the estimators. The Brewer Family estimators still perform well under RANSYS and the Hájek-Deville Family under CPS. It is surprising that $\hat{\tilde{V}}_{BR1}$ and $\hat{\tilde{V}}_{BR2}$ have the lowest absolute RB for population (f), however, this was also observed for population (a).

Tables 3.25 and 3.26 show the MSE for RANSYS and CPS respectively. These tables indicate that $\hat{\tilde{V}}_{BR1}$ and $\hat{\tilde{V}}_{Dev1}$ perform well for the MSE under both sampling algorithms, whereas $\hat{V}_{SYG}$ and $\hat{\tilde{V}}_{BR4}$ usually have the two highest MSEs. $\hat{\tilde{V}}_{Ber}$ generally has one of the higher MSEs, again indicating that the correlation has not greatly affected the performance of the variance estimators.

	Pop (f)			Pop (g)		
	10	20	40	10	20	40
$\hat{V}_{SYG}$	0.8379	0.9729	0.3091	**-0.9760**	0.5491	**-0.0255**
$\hat{\hat{V}}_{BR1}$	**0.4492**	**0.5490**	-0.1736	-5.9633	-2.9154	-2.8624
$\hat{\hat{V}}_{BR2}$	0.6625	0.7362	**0.0281**	-3.7725	-1.5781	-1.5604
$\hat{\hat{V}}_{BR3}$	0.7958	0.9234	0.2298	-1.5818	-0.2408	-0.2585
$\hat{\hat{V}}_{BR4}$	0.8150	0.9332	0.2350	**-1.2688**	**-0.1453**	**-0.2042**
$\hat{\hat{V}}_{BR-Dev}$	0.6287	0.7491	0.0590	-2.0941	-0.3526	-0.3428
$\hat{\hat{V}}_{Dev1}$	0.5951	0.6317	-0.4277	-4.8053	-2.5293	-3.8430
$\hat{\hat{V}}_{Dev2}$	0.6012	0.6446	-0.3969	-3.1479	-1.3163	-2.6542
$\hat{\hat{V}}_{Ros}$	0.6018	0.6569	-0.3232	-4.6277	-2.3641	-3.5618
$\hat{\hat{V}}_{Ber}$	0.7934	0.8359	-0.2023	-2.5591	-0.9576	-2.4380

Table 3.23: Relative biases (%) for RANSYS - populations (f) and (g)

	Pop (f)			Pop (g)		
	10	20	40	10	20	40
$\hat{V}_{SYG}$	-1.5133	0.04758	0.0706	**-0.2944**	1.5303	0.6705
$\hat{\hat{V}}_{BR1}$	-1.8385	-0.2344	0.0805	-5.4157	-1.4437	-0.6529
$\hat{\hat{V}}_{BR2}$	-1.6731	-0.0498	0.2827	-3.2130	**-0.0752**	0.6809
$\hat{\hat{V}}_{BR3}$	-1.5078	0.1348	0.4848	-1.0103	1.2933	2.0147
$\hat{\hat{V}}_{BR4}$	**-1.4894**	0.1445	0.4900	**-0.6956**	1.3910	2.0703
$\hat{\hat{V}}_{BR-Dev}$	-1.6671	-0.0369	0.3137	-1.5112	1.1808	1.9356
$\hat{\hat{V}}_{Dev1}$	-1.6999	-0.1541	-0.1744	-4.2443	-1.0544	-1.6745
$\hat{\hat{V}}_{Dev2}$	-1.6939	-0.1412	-0.1434	-2.5636	0.1888	**-0.4497**
$\hat{\hat{V}}_{Ros}$	-1.6933	-0.1289	-0.0696	-4.0685	-0.8848	-1.3840
$\hat{\hat{V}}_{Ber}$	-1.5189	**0.0350**	**0.0392**	-2.0044	0.3973	-0.5767

Table 3.24: Relative biases (%) for CPS - populations (f) and (g)

	Pop (f)			Pop (g)		
	10 (10^{12})	20 (10^{11})	40 (10^{10})	10 (10^{24})	20 (10^{23})	40 (10^{22})
$\hat{V}_{SYG}$	4.8050	5.7014	6.6846	4.5238	2.8635	1.3060
$\hat{\tilde{V}}_{BR1}$	**4.7329**	**5.5998**	6.5287	**3.9463**	**2.6005**	1.2066
$\hat{\tilde{V}}_{BR2}$	4.7634	5.6385	6.5759	4.1917	2.7174	1.2488
$\hat{\tilde{V}}_{BR3}$	4.7941	5.6774	6.6234	4.4454	2.8372	1.2919
$\hat{\tilde{V}}_{BR4}$	4.7975	5.6795	6.6247	4.4823	2.8459	1.2937
$\hat{\tilde{V}}_{BR-Dev}$	4.7641	5.6401	6.5803	4.3370	2.7863	1.2800
$\hat{\tilde{V}}_{Dev1}$	4.7607	5.6277	**6.5274**	4.0753	2.6411	**1.1838**
$\hat{\tilde{V}}_{Dev2}$	4.7614	5.6293	6.5317	4.2155	2.7078	1.2131
$\hat{\tilde{V}}_{Ros}$	4.7614	5.6303	6.5389	4.0963	2.6548	1.1936
$\hat{\tilde{V}}_{Ber}$	4.8025	5.6917	6.6429	4.4307	2.8174	1.2654

Table 3.25: Mean squared errors for RANSYS - populations (f) and (g)

	Pop (f)			Pop (g)		
	10 (10^{12})	20 (10^{11})	40 (10^{10})	10 (10^{24})	20 (10^{23})	40 (10^{22})
$\hat{V}_{SYG}$	4.4584	5.6237	6.5570	4.5931	3.0492	1.2878
$\hat{\hat{V}}_{BR1}$	**4.3961**	**5.5342**	6.4458	**3.9179**	**2.7446**	1.2172
$\hat{\hat{V}}_{BR2}$	4.4241	5.5721	6.4926	4.1616	2.8684	1.2601
$\hat{\hat{V}}_{BR3}$	4.4522	5.6101	6.5397	4.4134	2.9953	1.3040
$\hat{\hat{V}}_{BR4}$	4.4553	5.6122	6.5409	4.4501	3.0044	1.3059
$\hat{\hat{V}}_{BR-Dev}$	4.4247	5.5736	6.4969	4.3080	2.9417	1.2924
$\hat{\hat{V}}_{Dev1}$	4.4216	5.5613	**6.4447**	4.0474	2.7874	**1.1937**
$\hat{\hat{V}}_{Dev2}$	4.4222	5.5629	6.4490	4.1888	2.8583	1.2239
$\hat{\hat{V}}_{Ros}$	4.4222	5.5639	6.4561	4.0677	2.8019	1.2036
$\hat{\hat{V}}_{Ber}$	4.4578	5.6219	6.5519	4.4319	2.9731	1.2556

Table 3.26: Mean squared errors for CPS - populations (f) and (g)

3.5.7 Conclusions

This section provides a summary of the main findings for the comparisons between the variance estimators. As discussed in the methodology, the choice of the best estimator is subjective depending on whether one prefers a low MSE or a low absolute RB. As a result it is difficult to find any one of the nine approximate variance estimators that is uniformly superior to the rest. Table 3.27 shows the preferred estimator for RANSYS and CPS, and the preferred estimator within both families across both sampling algorithms. The RB and MSE are considered separately for each situation as it is difficult to determine one estimator which consistently has both a low absolute RB and low MSE. For some situations two estimators were chosen.

	RB	MSE
RANSYS	$\hat{\hat{V}}_{BR4}$	$\hat{\hat{V}}_{BR1}$
CPS	$\hat{\hat{V}}_{BR4}$, $\hat{\hat{V}}_{Ber}$	$\hat{\hat{V}}_{BR1}$
Brewer Family	$\hat{\hat{V}}_{BR4}$, $\hat{\hat{V}}_{BR3}$	$\hat{\hat{V}}_{BR1}$
Hájek-Deville Family	$\hat{\hat{V}}_{Dev2}$, $\hat{\hat{V}}_{Ber}$	$\hat{\hat{V}}_{Dev1}$

Table 3.27: The preferred approximate estimators

Among the Brewer Family estimators, $\hat{\hat{V}}_{BR1}$ obtained the lowest absolute RB for some populations, however, it performed poorly in populations (d), (e) and (g). On the other hand, $\hat{\hat{V}}_{BR3}$ and $\hat{\hat{V}}_{BR4}$ rarely had high absolute RBs as compared with the other Brewer Family estimator, and $\hat{\hat{V}}_{BR4}$ frequently had a low absolute RB. In the Hájek-Deville Family $\hat{\hat{V}}_{Dev1}$ achieved the lowest absolute RB in some situations, however, it performed poorly for populations (d) and (e). $\hat{\hat{V}}_{Dev2}$ and $\hat{\hat{V}}_{Ber}$ performed the best across all populations in the Hájek-Deville Family with regard to the RB, however the latter approximate variance estimator consistently had a high MSE.

$\hat{\tilde{V}}_{Dev1}$ always achieved the lowest MSE among the Hájek-Deville Family.

The MSE affects every single sample estimate, therefore it is important to consider this property in more detail. The approximate variance estimators had much the same MSEs for any given situation, therefore they should all be about equally liable to produce a poor estimate from time to time. This is consistent with the empirical results as each estimator has a high RB on at least one occasion. As $\hat{\tilde{V}}_{BR1}$ has a low MSE for both RANSYS and CPS it should be slightly less likely to produce a poor estimate from time to time.

The properties of the variance estimators within each family are generally very similar. Under RANSYS the Brewer Family estimators usually performed better across all sample sizes than the Hájek-Deville Family estimators. This was expected, as the Brewer Family estimators were designed using Hartley and Rao's approximation of joint inclusion probabilities realised under this sampling algorithm. The Hájek-Deville estimators performed just as well or better under CPS than the Brewer Family, except for populations (d) and (e) where this family performed rather poorly. Excluding $\hat{\tilde{V}}_{Ber}$, the Hájek-Deville Family estimators tend to be more stable, shown by their MSE, than the Brewer Family estimators (excluding $\hat{\tilde{V}}_{BR1}$). Therefore the Hájek-Deville Family estimators (excluding $\hat{\tilde{V}}_{Ber}$) should be less likely to produce a poor estimate than the Brewer Family estimators (excluding $\hat{\tilde{V}}_{BR1}$).

Finally, comparing $\hat{V}_{SYG}$ with the approximate variance estimators, this estimator nearly always had the largest MSE. The RB of this estimator is generally not noticeably greater in magnitude than the lowest absolute RB, and occasionally is itself the lowest. For RANSYS using Hartley and Rao's full approximations of the joint inclusion probabilities has not significantly improved this estimator to justify the extra computational effort required. The main interest in $\hat{V}_{SYG}$, however, is in

its performance under CPS, as the joint inclusion probabilities are known exactly. Under this sampling algorithm $\hat{V}_{SYG}$, unlike any of the other approximate variance estimators, had an absolute RB greater than 1% in only three occasions, the largest in magnitude being -3.7727%. All the approximate variance estimators had an absolute RB greater than 1% on numerous occasions. Thus, although the knowledge of the joint inclusion probabilities does not imply that $\hat{V}_{SYG}$ is the best estimator for minimising the absolute RB, it does appear to guarantee a consistently low absolute RB which is itself desirable. However, it is difficult to justify using this estimator due to its high MSE.

4 Further Results

4.1 Introduction

During the simulation study two interesting discoveries were made which do not seem to have appeared in the literature before. The first discovery is concerned with the relationship between some of the approximate variance estimators under consideration. The second discovery is the relationship between the entropy of a sampling design and the true variance of the HTE under this sampling design. Entropy is a measure of the "randomness" of a sampling design, so it was thought that if a sampling design had a greater entropy then it would also have a higher variance. This, however, has not been found to be true empirically. This chapter will discuss both these discoveries in detail, and their relevance to the estimation of variance.

4.2 Relationships among the Variance Estimators

One focus of this thesis the was of use Monte Carlo simulations to compare the variance estimators in the two families, and their performances under the two sampling algorithms. In the process of this analysis it was discovered that for every single sample generated under both sampling designs, the following properties held:

$$\begin{aligned}
\hat{\hat{V}}_{BR-Dev} &> \hat{\hat{V}}_{BR2} > \hat{\hat{V}}_{Dev1} \\
\hat{\hat{V}}_{BR-Dev} &> \hat{\hat{V}}_{Dev2} \\
\hat{\hat{V}}_{BR-Dev} &> \hat{\hat{V}}_{Ros} > \hat{\hat{V}}_{Dev1}.
\end{aligned}$$

Although the results in the previous chapter show that the Brewer Family estimators have very similar properties, it was not true that $\hat{\hat{V}}_{BR1}$, $\hat{\hat{V}}_{BR3}$ and $\hat{\hat{V}}_{BR4}$ were always greater than $\hat{\hat{V}}_{Dev1}$, like $\hat{\hat{V}}_{BR2}$ and $\hat{\hat{V}}_{BR-Dev}$ were. However, all the Brewer

Family variance estimators were greater than $\hat{\tilde{V}}_{Dev1}$ for a majority of the samples. Berger (2004) stated that the bias of $\hat{\tilde{V}}_{Dev1}$ should be less than $\hat{\tilde{V}}_{BR2}$, however, his reasoning was only based on approximations.

Mathematical analysis of these inequalities led to the following theorem below. For this theorem the second condition regarding equality holds because the variance estimators are designed to equal zero when x_i is exactly proportional to y_i. Empirically it was also found that $\hat{\tilde{V}}_{BR-Dev} > \hat{\tilde{V}}_{Ros}$ and $\hat{\tilde{V}}_{Ros} > \hat{\tilde{V}}_{Dev1}$, however these inequalities are not included in the theorem as mathematical proofs could not be derived because of the complexity of the formula for $\hat{\tilde{V}}_{Ros}$.

Theorem 4.1. *For any given sampling algorithm and any given set of first order inclusion probabilities,*

$$\text{(T1)}\ \ \hat{\tilde{V}}_{BR2} \geq \hat{\tilde{V}}_{Dev1}$$

$$\text{(T2)}\ \ \hat{\tilde{V}}_{BR-Dev} \geq \hat{\tilde{V}}_{BR2}$$

$$\text{(T3)}\ \ \hat{\tilde{V}}_{BR-Dev} \geq \hat{\tilde{V}}_{Dev2}$$

where equality holds if and only if: (i) the inclusion probabilities are all the same or (ii) if $\pi_i = nx_i / \sum_{i \in U} x_i$ and x_i is exactly proportional to y_i.

Proof of Theorem 1. The summations within this proof are summations over the sampled units unless indicated otherwise, therefore $\sum_{j \in s}$ is simplified to $\sum$. A proof of each inequality is given separately.

(T1) Consider the two variance estimators,

$$\begin{aligned}
\hat{\tilde{V}}_{BR2} &= \frac{n}{n-1} \sum (1-\pi_i)(y_i \pi_i^{-1} - n^{-1} \sum y_j \pi_j^{-1})^2 \\
\hat{\tilde{V}}_{Dev1} &= \frac{n}{n-1} \sum (1-\pi_i)\Big(y_i \pi_i^{-1} - \frac{\sum (1-\pi_j) y_j \pi_j^{-1}}{\sum (1-\pi_j)}\Big)^2
\end{aligned}$$

where $\hat{\tilde{V}}_{BR2}$ is simplified after the substitution of (2.23) into (2.26). These two estimators are both of the form

$$V(E) = \frac{n}{n-1}\sum(1-\pi_i)(y_i\pi_i^{-1} - E)^2 \tag{4.1}$$

for some value E which is constant for all values of i. Differentiating this equation with respect to E gives,

$$\frac{dV}{dE} = -2\frac{n}{n-1}\sum(1-\pi_i)(y_i\pi_i^{-1} - E).$$

Equating this derivative to zero implies that (4.1) is a minimised when

$$E = \frac{\sum(1-\pi_i)y_i\pi_i^{-1}}{\sum(1-\pi_j)},$$

as the second derivative is always positive. Therefore $\hat{\tilde{V}}_{Dev1}$ is the minimum for all equations of the form (4.1), and hence it is less than $\hat{\tilde{V}}_{BR2}$.

Equality holds if and only if

$$n^{-1}\sum y_i\pi_i^{-1} = \frac{\sum(1-\pi_i)y_i\pi_i^{-1}}{\sum(1-\pi_j)},$$

which occurs when π_i is the same for all i, or when $\pi_i = \frac{nx_i}{\sum_u x_j}$ and x_i is exactly proportional to y_i.

(T2) Consider,

$$\begin{aligned}\hat{\tilde{V}}_{BR2} &= \frac{n}{n-1}\sum(1-\pi_i)(y_i\pi_i^{-1} - n^{-1}\sum y_j\pi_j^{-1})^2\\ \hat{\tilde{V}}_{BR-Dev} &= \frac{1}{1-\sum a_i^2}\sum(1-\pi_i)(y_i\pi_i^{-1} - n^{-1}\sum y_j\pi_j^{-1})^2\end{aligned}$$

where $a_k = \frac{1-\pi_k}{\sum(1-\pi_j)}$. These two estimators are exactly the same except for the constant factor outside the summations. Therefore $\hat{\tilde{V}}_{BR-Dev} \geq \hat{\tilde{V}}_{BR2}$ if and only if $\frac{1}{1-\sum a_i^2} \geq \frac{n}{n-1}$, hence if $n\sum a_i^2 \geq 1$.

First, if the inclusion probabilities are all equal then, $a_i = \frac{1}{n}$ for all i and hence

$n\sum a_i^2 = n \cdot n(\frac{1}{n})^2 = 1$ and $\hat{\tilde{V}}_{BR-Dev} = \hat{\tilde{V}}_{BR2}$. Secondly,

$$\begin{aligned} n\sum a_i^2 &= n\sum\Big(\frac{1-\pi_i}{\sum(1-\pi_k)}\Big)^2 \\ &= n\frac{\sum(1-\pi_i)^2}{(\sum(1-\pi_k))^2} \\ &= n\frac{\sum(1+\pi_i^2-2\pi_i)}{(n-\sum\pi_k)^2} \\ &= \frac{n^2+n\sum\pi_i^2-2n\sum\pi_i}{n^2+(\sum\pi_k)^2-2n\sum\pi_k} \end{aligned}$$

As the variance of π_i is clearly greater than or equal to zero, this implies that $(\sum\pi_i)^2 \leq n\sum\pi_i^2$, and this in turn implies that,

$$n^2+(\sum\pi_k)^2-2n\sum\pi_k \leq n^2+n\sum\pi_i^2-2n\sum\pi_i$$

and hence $n\sum a_i^2 \geq 1$. Thus $\hat{\tilde{V}}_{BR-Dev} \geq \hat{\tilde{V}}_{BR2}$.

(T3) Finally consider

$$\begin{aligned} \hat{\tilde{V}}_{BR-Dev} &= \frac{1}{1-\sum a_i^2}\sum(1-\pi_i)(y_i\pi_i^{-1}-n^{-1}\sum y_j\pi_j^{-1})^2 \\ \hat{\tilde{V}}_{Dev2} &= \frac{1}{1-\sum a_i^2}\sum(1-\pi_i)\Big(y_i\pi_i^{-1}-\frac{\sum(1-\pi_j)y_j\pi_j^{-1}}{\sum(1-\pi_j)}\Big)^2. \end{aligned}$$

It is clear from (T1) that

$$\sum(1-\pi_i)(y_i\pi_i^{-1}-n^{-1}\sum y_k\pi_k^{-1})^2 \geq \sum(1-\pi_i)\Big(y_i\pi_i^{-1}-\frac{\sum(1-\pi_j)y_j\pi_j^{-1}}{\sum(1-\pi_j)}\Big)^2,$$

and from (T2) that $\frac{1}{1-\sum a_i^2} \geq \frac{n}{n-1}$, hence $\hat{\tilde{V}}_{BR-Dev} \geq \hat{\tilde{V}}_{Dev2}$.

□

One of the two conditions required for equality is impossible in unequal probability sampling; if the inclusion probabilities are the same then the sampling design would not be unequal probabilities sampling. In addition, it would be highly coincidental for x_i to be exactly proportional to y_i.

To understand the implications of the above inequalities, consider two arbitrary variance estimators $\hat{V}_A$ and $\hat{V}_B$, where $\hat{V}_A \geq \hat{V}_B$ for all samples, then

$$\begin{aligned} \hat{V}_A &\geq \hat{V}_B \\ \Rightarrow \quad E(\hat{V}_A) &\geq E(\hat{V}_B) \\ \Rightarrow \quad E(\hat{V}_A)/V_T - 1 &\geq E(\hat{V}_B)/V_T - 1 \\ \Rightarrow \quad RB(\hat{V}_A) &\geq RB(\hat{V}_B) \end{aligned}$$

Note this is not the absolute RB, so $RB(\hat{V}_A)$ may be positive while $RB(\hat{V}_B)$ is negative. Corollary 1 below is a consequence of Theorem 1.

Corollary 4.1. *For any given sampling algorithm the following is true for all simulations where the variances are estimated for the same samples,*

(C1) $RB(\hat{\hat{V}}_{BR2}) \geq RB(\hat{\hat{V}}_{Dev1})$ *from (T1),*

(C2) $RB(\hat{\hat{V}}_{BR-Dev}) \geq RB(\hat{\hat{V}}_{BR2})$ *from (T2),*

(C3) $RB(\hat{\hat{V}}_{BR-Dev}) \geq RB(\hat{\hat{V}}_{Dev1})$ *from (T1) and (T2), and*

(C4) $RB(\hat{\hat{V}}_{BR-Dev}) \geq RB(\hat{\hat{V}}_{Dev2})$ *from (T3).*

These relationships between the RBs of the variance estimators described in Corollary 1 are observed in the results in this study, and also in the results produced by Brewer and Donadio (2003) and by Matei and Tillé (2005). For example, consider Table 4.1 which is a subset of the relative biases from population (a) under RANSYS, reproduced from Table 3.13 for convenience. It is clear from this table that the above inequalities hold for all sample sizes. The estimator $\hat{\hat{V}}_{Ros}$ has also been included in Table 4.1, as it has already been noted that empirically it appears always to be intermediate in value between $\hat{\hat{V}}_{BR-Dev}$ and $\hat{\hat{V}}_{Dev1}$.

These results have implications when selecting the preferred variance estimator. First, it is desired that the HTE has a small variance. This does not imply that $\hat{\hat{V}}_{Dev1}$

	$n = 10$	$n = 20$	$n = 40$
$\hat{\tilde{V}}_{BR2}$	0.1049	-0.0094	0.0772
$\hat{\tilde{V}}_{BR-Dev}$	0.1355	0.0626	0.2814
$\hat{\tilde{V}}_{Dev1}$	0.0240	-0.3484	-1.7212
$\hat{\tilde{V}}_{Dev2}$	0.0546	-0.2767	-1.5208
$\hat{\tilde{V}}_{Ros}$	0.0433	-0.2725	-1.4128

Table 4.1: Relative biases (%) - subset of population (a)

is better than $\hat{\tilde{V}}_{BR2}$ from Theorem 1; one is concerned with how well the variance estimator estimates the variance, not in estimating it to be as small as possible. A good variance estimator would itself have a small variance, therefore the variance of the HTE should be estimated with as small an MSE as possible. It is important to reiterate that the choice of the best estimator is subjective, some prefer a low MSE whilst others prefer a low bias. If all the variance estimators have much the same MSE, then one would usually prefer the one with the smallest RB. It is clear from the results in the previous chapter that the MSE for all the variance estimators are very similar, therefore the RB can be used to determine the preferred estimator.

A small absolute RB is desired for a good variance estimator. This does not, however, imply that $\hat{\tilde{V}}_{BR2}$ is a better estimator than $\hat{\tilde{V}}_{Dev1}$ by Corollary 1. The absolute RB is important, thus each possible case needs to be considered on its merit. If $\hat{\tilde{V}}_{BR2}$ has a positive RB then $\hat{\tilde{V}}_{Dev1}$ will either have a smaller positive RB or a negative RB (see $n = 10$ and $n = 40$ respectively in Table 4.1). In the first case $\hat{\tilde{V}}_{Dev1}$ will be the better estimator, but in the latter case it will depend on the magnitude of the RBs ($\hat{\tilde{V}}_{BR2}$ is preferred in Table 4.1). If, however, $\hat{\tilde{V}}_{BR2}$ is negatively biased, then $\hat{\tilde{V}}_{Dev1}$ will have an even greater negative bias, in which case $\hat{\tilde{V}}_{BR2}$ is preferred (see $n = 20$ in Table 4.1).

A variance estimator with a positive bias is generally preferred as it is better to overestimate the variance than to underestimate it. If the variance is underestimated a greater precision is claimed than is actually obtained. If this is the case then $\hat{\hat{V}}_{Br-Dev}$ would be the choice estimator out of those represented in Corollary 1. As the Brewer Family estimators have been shown to have similar properties, this result also suggests that these estimators may be conservative choices compared with the Hájek-Deville Family estimators.

4.2.1 Effects of sample size

It was observed for populations (a) and (b) only, that as the sample size increased from 10 to 40 the differences between the RBs of the paired variance estimators in Corollary 1 increased for the two sampling algorithms. This implies that as n increases $\hat{\hat{V}}_{BR2}$ and $\hat{\hat{V}}_{Dev1}$, for example, become more dissimilar. Intuitively this does not make sense because the precision of a variance estimator should increase with the sample size as more information about the population is known. Thus, the differences should decrease.

To analyse this relationship empirically, a smaller simulation of R=10,000 independent samples of a large range of sample sizes was considered under RANSYS. Only RANSYS was considered because of its quick implementation. For this analysis, using 10,000 simulations is sufficient as the differences between the RBs of the estimators have been shown to be consistent for this simulation size (see Table 3.7). For each sample the differences between the variance estimates were first calculated, and then the mean was determined for each sample size. For instance,

$$\frac{1}{R}\sum_{i=1}^{R}(\hat{\hat{V}}_{BR2_i} - \hat{\hat{V}}_{Dev1_i}) \tag{4.2}$$

n	$\hat{\hat{V}}_{BR2} - \hat{\hat{V}}_{Dev1}$	$\hat{\hat{V}}_{BR-Dev} - \hat{\hat{V}}_{Dev1}$	$\hat{\hat{V}}_{BR-Dev} - \hat{\hat{V}}_{Dev2}$
10	3237.2	4466.5	3238.5
20	6329.4	7662.2	6334.5
30	9823.0	11281.3	9836.3
40	14254.0	15877.1	14284.1
50	15181.4	16631.4	15220.3
60	11255.7	12438.1	11287.1
70	11578.8	12411.4	11608.7
80	10267.1	10963.1	10296.2
90	7806.0	8283.9	7826.1
100	6734.8	7071.0	6751.1
120	2460.2	2660.6	2466.7
140	1461.2	1580.6	1464.8
160	1138.8	1235.5	1142.5

Table 4.2: Mean differences between pairs of estimators as the sample size increases - population (a)

where $\hat{\hat{V}}_{BR2_i}$ and $\hat{\hat{V}}_{Dev1_i}$ are the variance estimates for the i^{th} sample in the R simulated samples. The difference between $\hat{\hat{V}}_{BR-Dev}$ and $\hat{\hat{V}}_{BR2}$ was not considered because the main concern was the difference between the Brewer Family and the Hájek-Deville Family of estimators.

Three pairs of differences were considered in Tables 4.2 and 4.3 for populations (a) and (b) respectively. As the sample size increases, the difference between the paired variance estimators considered eventually decreases as expected. It is not clear, however, why the differences initially increase.

n	$\hat{\hat{V}}_{BR2} - \hat{\hat{V}}_{Dev1}$	$\hat{\hat{V}}_{BR-Dev} - \hat{\hat{V}}_{Dev1}$	$\hat{\hat{V}}_{BR-Dev} - \hat{\hat{V}}_{Dev2}$
10	3513.6	5240.9	3514.9
20	6509.4	8400.4	6514.6
30	9950.4	12018.2	9963.8
40	14124.4	16417.1	14153.8
50	13181.3	15373.0	13216.6
60	10745.4	12493.3	10775.7
70	12808.4	13995.0	12840.4
80	12874.0	13811.9	12907.2
90	9663.4	10301.4	9685.9
100	11523.6	12125.6	11556.5
120	2382.0	2676.4	2388.0
140	1184.3	1382.7	1187.5
160	1039.6	1189.5	1042.8

Table 4.3: Mean differences between pairs of estimators as the sample size increases - population (b)

4.3 Entropy and Variance

4.3.1 Entropy

Entropy is a measure of spread of the sampling design, $p(\cdot)$, and is computed by

$$e = -\sum_{s \in \mathcal{S}} p(s) ln(p(s)). \tag{4.3}$$

This definition of entropy was introduced by Shannon (1948), and is also known as Information Entropy and Shannon's Entropy. He introduced this concept of entropy for use in the communication of messages, and in particular the ability to reconstruct a message which has been distorted by random noise. A message can be represented by a Markov chain of letters, or outcomes. Shannon aimed to measure the amount of uncertainty removed when the next outcome in a sequence becomes known, or alternatively the average amount of information which is contained in the next outcome. That is, if there is a set of possible outcomes $X = \{x_1, \ldots, x_n\}$ with the corresponding probability density function defined by $\{p_1, \ldots, p_n\}$, then the amount of uncertainty, $H(X) \equiv H(p_1, \ldots, p_n)$ should have following properties (Shannon, 1948, p. 388):

(i) $H(X)$ is continuous,

(ii) $H(X)$ is maximised when the p_i are equal, so that

$$H(p_1, \ldots, p_n) \leq H(\frac{1}{n}, \ldots, \frac{1}{n}), \text{ and}$$

(iii) if an outcome is broken down into two possible outcomes then the entropy does not change.

Shannon (1948, p. 389) showed that the only function which satisfies these properties is

$$H(X) = -K \sum_{i=1}^{n} p_i log(p_i) \tag{4.4}$$

where K is a positive constant. The choice $K = 1$ is commonly made, and if the natural logarithm is used, equation (4.3) results. One final property of the entropy

is that if an outcome has $p_i = 0$, then the entropy remains unchanged if this outcome is removed, as it is known that this outcome will never occur.

Another illustration which assists in understanding the meaning of entropy is as follows[2]. Consider the situation of a box containing coloured balls. If all the balls are of different colours, then the uncertainty about the colour of the next ball selected is a maximum. If, however, it is known that a large proportion of the balls are red, then the amount of uncertainty of the next ball selected is reduced. As there is a greater amount of uncertainty when all the balls are of different colours, being informed of the colours obtained at consecutive draws will provide more information about the population than it would if it was already known that a majority of the balls were red. Thus the first situation has a higher entropy.

To provide relevance to a survey sampling situation, the balls represent each possible sample in the support, $\mathcal{S}$, where the sampling design prescribes the probability of selecting each sample. A sampling design maximises the entropy when there is the greatest amount of uncertainty in the sample which will be selected. The situation where the balls are all of different colours is equivalent to the situation when all samples have the same probability of being selected, that is if $p(s)$ is the same for all samples.

High entropy sampling designs are beneficial as the approximate variance estimators described in this study are designed to perform well under these algorithms. In Hajék's posthumous book (1981) he proved that CPS maximises the entropy among all sampling algorithms having the same inclusion probabilities and support. The second property of entropy implies that the entropy of a sampling design is maximised under *srswor*, when $p(s)$ is the same for all samples. This does

[2]Idea sourced from Wikipedia, http://en.wikipedia.org

not contradict Hájek's proof, since if the inclusion probabilities are all equal CPS produces the same sampling design as *srswor*, and hence the same entropy.

It was initially assumed that, except for in some unusual and easily recognisable circumstances, there would be a one-to-one relationship between the variance of an estimator and the entropy. That is, it was conjectured that in situations where the entropy was already close to a maximum, then increasing the entropy should increase the variance. This belief was also portrayed by Brewer and Donadio (2003), where they believed Tillé Sampling had a lower entropy than RANSYS because this algorithm produced lower variances.

If a population is ordered in a meaningful way then this will almost certainly decrease the variance of a sample drawn from it. For example, consider Ordered Systematic Sampling (OSYS) when the units are ordered by the size of the auxiliary variable. The variance under OSYS is typically smaller than the variance under RANSYS, especially when there is a trend in the study variable (Bellhouse and Rao, 1975). Therefore, OSYS is a low entropy sampling design and typically provides accurate estimates that have a low variance. However, if the ordered population has a periodic trend then this algorithm may also produce a high variance. For instance, given a list of soldiers in which every tenth is a sergeant and the rest are privates, a one-in-ten systematic sampling algorithm will increase the variance of most survey study variables. Thus there is conclusive evidence against a direct one-to-one relationship.

The true variances calculated under RANSYS and CPS shown in section (3.5.1) indicate that although CPS has the higher entropy it typically produces a lower variance. Among the twenty-one different sampling situations considered in this thesis (seven populations and three sample sizes), CPS had a larger variance than

RANSYS only eight times. In addition, the variance under RANSYS is 2.5% larger for population (e) when $n = 40$. Although the "true variance" for RANSYS is an approximation, it is difficult to believe that this would cause the variance to be greater so frequently and sometimes to be larger by more than 2%. There are four possible explanations for this: (A) The simulated variances for RANSYS are misleading, (B) CPS is not always the method of highest entropy, (C) RANSYS has nearly maximal entropy and hence its variances should be similar to those of CPS, or (D) there is not a simple relationship between the entropy and variance.

Table 3.2 provides clear evidence to refute (A). The simulated "true variances" under RANSYS are close to the SYG estimates of variance obtained using Hartley and Rao's approximations (2.17) to the joint inclusion probabilities. With respect to (B), the required conditions of Hájek's proof that CPS maximises the entropy are met in this analysis. That is, the two sampling algorithms considered have the same inclusion probabilities defined by equation (2.1), and the same support, $\mathcal{S}_n$. Two approaches are considered to analyse situations (C) and (D) further. The first is to examine both the entropy and the variance for small populations, and the second is to consider "pairwise entropy".

4.3.2 Previous studies

Before examining these approaches, other results produced by Aires (2000) and Tillé (1996b) are discussed. Although neither of these papers explicitly discuss the relationship between the concepts of entropy and variance, their results assist in understanding these concepts.

Aires (1999) compared two high entropy sampling algorithms, namely Pareto πps Sampling and CPS. The exact first and second order inclusion probabilities

were calculated for both of these sampling algorithms. The working probabilities for CPS were calculated using Aires' algorithm by solving the system of $N + 1$ non-linear equations, rather than using the iterative algorithm in equation (2.15). Aires compared the true variances for the two sampling algorithms under eleven different sampling situations ranging from $N = 5$ and $n = 2$ up to $N = 271$ and $n = 65$. The true SYG variance calculated for Pareto πps Sampling was larger than the corresponding true variances calculated under CPS nine of the eleven situations. The largest difference between these variances, however, was only 0.1245%.

On balance, Aires recommends Pareto πps Sampling over CPS, because although the results were very similar under both algorithms Pareto πps Sampling has the advantage of being easy to use with Permanent Random Number (PRN) Sampling (Ohlsson, 1995). In survey sampling PRN Sampling makes it particularly easy for some units to be rotated out of (and others into) the sample for repeating a survey. CPS does permit sample rotation (Chen *et al.*, 1994) however not as easily as in Pareto πps Sampling.

Tillé (1996b) analysed the exact joint inclusion probabilities under RANSYS and CPS, as well as approximations to these probabilities. As it is difficult to determine the exact inclusion probabilities under RANSYS, only a small sampling situation with $N = 7$ and $n = 3$ was considered. The most interesting result from this paper is that Tillé found the exact joint inclusion probabilities under CPS and RANSYS to be close, which would imply that their variances should also be close. In addition, Tillé also stated that he believed RANSYS to have an entropy close to that of CPS.

4.3.3 Small populations

In order to calculate the entropy of a sampling design, the values of p(s) for each sample in the support are required. This is straightforward for CPS - see equation (2.3). To determine the sampling design under RANSYS, however, simulations are required. In theory, this is accomplished by generating a large number of samples under RANSYS and determining the proportion of each possible sample. However, this is not possible for large populations because there are $\binom{N}{n}$ possible samples in any given without replacement sampling algorithm, and a larger number of simulations would be required to obtain precise estimates of $p(s)$. As a result, three small sample situations were considered for this analysis, two with $N = 5$ and $n = 3$, and one with $N = 6$ and $n = 3$. For these situations, 150,000 samples were simulated to approximate the sampling design for RANSYS.

OSYS is also included in this study as it is known to have both a low entropy and a low variance for well ordered populations, which should be valuable when studying the relationships between these concepts. In this study the units are ordered by the sizes of the known auxiliary variables for this sampling algorithm. As with RANSYS, 150,000 samples are simulated using OSYS to approximate the sampling design. It is common under OSYS to find $p(s) = 0$ for some samples within the support, $\mathcal{S}_n$, depending on the ordering of the population units. If $p(s) = 0$ then this sample, s, can still be included in the support as the conditions (1.1) and (1.2) are still satisfied. The reason why this sampling algorithm generally has a low variance is that atypical samples can easily be given zero probability of selection.

The first population considered here consists of 5 units from the MU281 population, and is shown in Table 4.4. The second column indicates which units were chosen from the MU281 population. For each unit, the inclusion probability,

π_i determined by (2.1), and the working probability, $\tilde{p}_i$ are shown. The working probabilities for CPS are determined from the desired inclusion probabilities using Chen and Deville's algorithm. These probabilities were required to calculate the sampling design for CPS. The population is ordered by the size of the X variable to assist in later discussions. The ten possible samples and the corresponding values of the sampling design, $p(s)$, for CPS, RANSYS and OSYS, together with estimates of the total for each sample, are given in Table 4.5.

Label	MU281 Unit	Y	X	π_i	$\tilde{p}$
1	1	288	33	0.3474	0.3977
2	5	536	56	0.5895	0.5911
3	10	467	60	0.6316	0.6239
4	8	517	66	0.6947	0.6754
5	7	623	70	0.7368	0.7119
Total		2431	285	3.0000	3.0000

Table 4.4: First small population

The entropy, expectation and variance of the estimator across all possible samples are calculated by (4.3), (1.7) and (1.8) respectively. As the HTE is unbiased, the expectation of this estimator is used to indicate the accuracy of the approximate sampling designs simulated. If the weighted mean of the sample estimates is appreciably different from the true population total, this indicates that the approximate values of $p(s)$ are imprecise. If the sampling design imprecise then the entropy and the variance will also be imprecise. As the sampling design is known exactly under CPS, the mean for this algorithm is always unbiased. Table 4.6 displays these results for this population.

Sample	Estimate	Values of $p(s)$		
		CPS	RANSYS	OSYS
(1,2,3)	2477.7933	0.0384	0.0363	0
(1,2,4)	2482.5433	0.0482	0.0508	0
(1,2,5)	2583.8766	0.0573	0.0636	0
(1,3,4)	2312.6742	0.0553	0.0505	0.2605
(1,3,5)	2414.0076	0.0657	0.0651	0.0839
(1,4,5)	2418.7576	0.0824	0.0787	0
(2,3,4)	2392.8690	0.1212	0.1253	0
(2,3,5)	2494.2024	0.1439	0.1396	0.2226
(2,4,5)	2498.9524	0.1805	0.1732	0.3703
(3,4,5)	2329.0833	0.2071	0.2170	0.0627

Table 4.5: Sampling design - the first small population having a known total of 2431.00

	Entropy	Mean	Variance
CPS	2.1498	2431.00	6091.8939
RANSYS	2.1460	2430.74	6112.9460
OSYS	1.4343	2431.60	6921.2710

Table 4.6: Entropy, mean and variance - first small population

The weighted mean of the estimates for RANSYS and OSYS in Table 4.6 are both close to the true total of 2431.00 indicating that the sampling designs are well approximated, and hence that the entropy and the variances are reliable measures. The calculated values of the entropy are expected, with CPS having the maximum entropy. The variances, however, are increasing as the entropy decreases. The entropy under RANSYS is close to the maximum entropy. The reason that OSYS does not minimise the variance is because the population considered is not well ordered. OSYS will cause a reduction in the variance compared with RANSYS if the ratio of y_i to x_i tends to increase or decrease with the size of x_i. For this population, as x_i increases the ratios were $y_i/x_i = \{8.73, 9.57, 7.78, 7.83, 8.9\}$ which does not tend to increase or decrease with x_i. This is an indication that the variance is dependent upon both the sampling design and the structure of the population.

The second small population considered was constructed to ensure that OSYS would produce a lower variance than RANSYS. The auxiliary variables were chosen to increase in multiples of one hundred, however, the last unit was assigned the value 490, instead of 500, to ensure that the inclusion probability was less than unity. The corresponding values of Y were chosen so the ratio of y_i to x_i would increase with the size of X. As x_i increases these ratios were $y_i/x_i = \{0.75, 1, 1.33, 1.5, 1.84\}$. As the y_i values were not specifically chosen to be proportional to x_i this may increase the variances. Table 4.7 shows this ordered population, with the corresponding

probabilities.

Label	Y	X	π_i	$\tilde{p}$
1	75	100	0.2013	0.2624
2	200	200	0.4027	0.4384
3	400	300	0.6040	0.5723
4	600	400	0.8054	0.7483
5	900	490	0.9866	0.9787
Total	1630	1490	3.000	3.0001

Table 4.7: Second small population

Table 4.8 displays the entropy, mean and variance for the second small population. The sampling design and estimates for each sample are provided in Table E.1 (Appendix E). The variance under OSYS is the smallest as expected due to the ordered population. Despite maximising the entropy CPS, once again, has a smaller variance than RANSYS.

	Entropy	Mean	Variance
CPS	1.5336	2175.00	22006.6742
RANSYS	1.4892	2175.56	27437.2108
OSYS	1.1401	2174.74	20452.73

Table 4.8: Entropy, mean and variance - second small population

The final small population is shown in Table 4.9, where N=6 and n=3. The population was constructed from the inclusion probabilities regularly used by Tillé (2006) in his examples. This population was chosen because the second order

inclusion probabilities were known exactly for the three sampling algorithms (see Tillé, 2006, pp. 86,127,128), which may explain the behaviour of the variances. The values of X were constructed based on the inclusion probabilities and a chosen value $X_\bullet = 1500$. The corresponding values of Y were chosen to be approximately proportional to X, where the difference between X and Y was greater for larger units, as this is a good representation of reality. This method was chosen to ensure a suitable population was constructed for the desired inclusion probabilities to reduce the variance. As a results, however, this population was order well for OSYS to have a smaller variance than RANSYS. There are 20 possible samples in this sampling situation. The values of the sampling design and estimates for this population are provided in Table E.2 (Appendix E). The results in Table 4.10 display the same behaviour with regard to the order of the size of the entropies and variances of three algorithms as that of the first population considered. That is, as the entropy increases as the variance decreases.

Label	Y	X	π_i	$\tilde{p}$
1	37	35	0.07	0.1021
2	89	85	0.17	0.2238
3	200	205	0.41	0.4417
4	315	305	0.61	0.5796
5	425	415	0.83	0.7794
6	440	455	0.91	0.8734
Total	1506	1500	3.0000	3.0000

Table 4.9: Third small population

The joint inclusion probabilities may be used to explain why the variances do not increase with the entropy. The joint inclusion probabilities for the three sampling

	Entropy	Mean	Variance
CPS	1.8897	1506.00	280.3127
RANSYS	1.8281	1506.07	294.1470
OSYS	1.3610	1505.92	487.1576

Table 4.10: Entropy, mean and variance - third small population

algorithms under this population are shown below. The diagonal of these matrices are simply the first order inclusion probabilities, as the π_{ii} are not required for the variance estimates. The values below the diagonal are excluded as the matrices are symmetric.

$$\pi_{ij}(CPS) = \begin{pmatrix} 0.07 & 0.0049 & 0.0130 & 0.0215 & 0.0447 & 0.0559 \\ & 0.17 & 0.0324 & 0.0537 & 0.1113 & 0.1377 \\ & & 0.41 & 0.1407 & 0.2888 & 0.3452 \\ & & & 0.61 & 0.4691 & 0.5351 \\ & & & & 0.83 & 0.7461 \\ & & & & & 0.91 \end{pmatrix}$$

$$\pi_{ij}(RANSYS) = \begin{pmatrix} 0.07 & 0.0140 & 0.0257 & 0.0257 & 0.0373 & 0.0373 \\ & 0.17 & 0.0623 & 0.0623 & 0.0740 & 0.1273 \\ & & 0.41 & 0.0873 & 0.2957 & 0.3490 \\ & & & 0.61 & 0.4957 & 0.5490 \\ & & & & 0.83 & 0.7573 \\ & & & & & 0.91 \end{pmatrix}$$

$$\pi_{ij}(OSYS) = \begin{pmatrix} 0.07 & 0 & 0 & 0.07 & 0.07 & 0 \\ & 0.17 & 0 & 0.17 & 0.02 & 0.15 \\ & & 0.41 & 0.02 & 0.39 & 0.41 \\ & & & 0.61 & 0.44 & 0.52 \\ & & & & 0.83 & 0.74 \\ & & & & & 0.91 \end{pmatrix}$$

The population units are increasing in size as the unit label increases, therefore $\pi_{i,i+1}$ is the probability of selecting two adjacent, and hence similar, units. Comparing the above inclusion probabilities, under CPS the probability that two small units (π_{12} or π_{23}) or alternatively two large units (π_{45} or π_{56}) are selected together is always smaller than for RANSYS[3]. Thus CPS is more likely to chose two dissimilar units than two similar units within a sample, hence obtaining a better representation of the population and reducing the variance. For OSYS to have a smaller variance the population must be well ordered, in which case the inclusion probabilities for adjacent units will be smaller than they are for the other two sampling algorithms. Although the third small population was not well ordered, its joint inclusion probabilities did tend to be smaller for adjacent units compared to RANSYS and CPS. There was not, however, an appreciable difference among the inclusion probabilities of the two pairs of larger units which may explain why OSYS still has a higher variance. The same relationships can also be observed for the other two populations where the joint inclusion probabilities are shown in Appendix E.

The property of CPS observed above is similar to the dumbbell effect. Dumbbell Sampling (Foreman, 1991) is a form of cluster sampling where the between cluster

[3]This idea of comparing the joint inclusion probabilities was suggested by John Preston, ABS.

variation is minimised. This provides more efficient estimators and hence a lower variance. Dumbbell Sampling works by placing the largest and smallest units in the first cluster, then the second largest and second smallest units in the second cluster and so on. For example consider the population $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$, then the clusters will be defined as $(1, 8), (2, 7), (3, 6)$ and $(4, 5)$. This ensures that a good spread of the population is sampled. Clearly under this situation the probability of selecting two dissimilar units is high. Therefore CPS appears to be behaving similar to Dumbbell Sampling as it is likely to select two dissimilar units. One difference between CPS and Dumbbell Sampling, however, is that the latter algorithm has a low entropy, as some $p(s) = 0$, where as CPS has a high entropy.

4.3.4 Pairwise entropy

It is also of interest to examine the main populations in this study. Clearly this is not possible for samples of size $n = 10$, 20 or 40. However, if samples of size $n = 2$ are considered then the joint inclusion probabilities are the same as the probabilities of selecting those samples, that is $\pi_{ij} = p(s)$ where $s = \{i, j\}$. This allows the entropy for a large population to be evaluated relatively easy, as the joint inclusion probabilities for RANSYS can be approximated by Hartley and Rao's formula, (2.17), which is designed for situations where the population size is large compared with the sample size. The entropy measured for sampling situations with $n = 2$ is known as "pairwise entropy" in this thesis.

First, however, consider a smaller sampling situation of $N = 8$ and $n = 2$ (see Appendix E for the population, estimates and sampling design). As the population size was not large compared with the sample size, 150,000 simulations were used instead of Hartley and Rao's formula to estimate the joint inclusion probabilities for RANSYS. Table 4.11 shows that CPS still has the maximum entropy and RANSYS

the largest variance.

	Entropy	Mean	Variance
CPS	2.6909	1827.00	7248.7830
RANSYS	2.6703	1826.84	8392.1110
OSYS	2.0002	1827.00	4532.3810

Table 4.11: Entropy, mean and variance - population N=8 and n=2

Finally, one population from the MU281 data set and another from the CO124 data set were considered to ensure that two quite different population structures were analysed. Populations (a) and (d) were considered. Only CPS and RANSYS were analysed as the joint inclusion probabilities cannot be simulated for OSYS when the population is large.

	Population (a)			Population (d)		
	Entropy	Mean	Variance	Entropy	Mean	Variance
CPS	9.8864	53151.00	2.0852 (10^7)	6.4198	1770336	1.3127 (10^{11})
RANSYS	9.8857	53145.87	2.1111 (10^7)	6.4045	1762077	1.4718 (10^{11})

Table 4.12: Pairwise entropy - populations (a) and (d)

Table 4.12 shows that the entropy under RANSYS is only slightly lower than that for CPS, but that the variance is appreciably larger for RANSYS. There is only a 0.01% bias in the expectation for population (a) and 0.47% for population (d). This indicates that Hartley and Rao's full approximation under these situations are acceptable, and that the variances should be well approximated. These biases are quite small, however, as they are both negative there may also be slight reductions in the variances shown. It is difficult to believe that this decrease will cause RANSYS

to have a lower variance than CPS.

4.3.5 Conclusions

In conclusion, the possible situations described earlier to explain why CPS has a higher entropy yet lower variance are revisited: (A) The simulated variances for RANSYS are misleading, (B) CPS is not always the method of highest entropy, (C) RANSYS has nearly maximal entropy and hence its variances should be similar to those of CPS, or (D) there is not a simple relationship between the entropy and variance. (A) has already been refuted, and the empirical results agree with Hájek's proof that CPS maximises the entropy, refuting (B). The results indicated that RANSYS almost achieves the maximum entropy (C), however this does not explain why the variances under RANSYS is higher than CPS, only that the variances may be similar. The results show that there is not a simple relationship between the entropy and the variance. Therefore (D) is the appropriate response to the explanation as to why the variances under CPS are typically lower than RANSYS. The results in relation to OSYS clearly indicate that the variance is dependent upon both the sampling algorithm and the structure of the population.

Additional investigations on other sampling algorithms to arrive at more conclusive results regarding the relationship between entropy and variance is required. In particular, it would be appropriate to also consider Tillé sampling which Brewer and Donadio believed had a lower entropy than RANSYS due to the fact that it had a lower variance.

Entropy measures the amount of uncertainty in a sequence of events, or the amount of "randomness" in this sequence. Thus the entropy of a sampling design measures the amount of uncertainty in the sample that is to be selected, or the

amount of "randomness" in the selection of more than one sample. The entropy does not, however, explicitly describe the structure of the sample itself. That is, a high entropy sampling designs does not imply the units within a selected sample are highly random. Therefore it was a misconception to assume that a highly random sampling design, implies a highly random sample, and hence a high variance. The amount of "randomness" within a sample may be related to its probability of being selected, and hence the entropy of a sampling design, however this relationship is unclear.

To further analyse these concepts the entropy of each individual sample could be considered by using Shannon's formula,

$$-\sum_{i \in s} \pi_i log(\pi_i). \tag{4.5}$$

The set of possible samples selectable under CPS and RANSYS is the same, therefore both the entropy of each sample and its corresponding probability of begin selected, $p(s)$, would need to be considered to conclude whether the variance should be lower under certain algorithms. The results observed in this section indicate that CPS is more likely to chose samples with dissimilar units, therefore produce a lower variance than RANSYS. As this concept was only developed at the conclusion of this study it was not feasible to analyse it further.

4.4 Combining these Results

In the first discovery it was found that two variance estimators from the Brewer Family, $\hat{\tilde{V}}_{BR-Dev}$ and $\hat{\tilde{V}}_{BR2}$, always had a greater variance estimate than some estimators in the Hájek-Deville Family (see Theorem 1). As the estimators within each family have similar properties, this suggests that generally the RBs of the Brewer Family estimators will be higher than those of the Hájek-Deville Family. Assuming these variance estimators are designed to perform well under their

corresponding sampling algorithm, this indicates that the variances under RANSYS should be greater than under CPS. The second discovery also indicates that this is generally true. Thus these two discoveries combine to support the initial hypothesis that the Brewer Family estimators should perform better under RANSYS and the Hájek-Deville Family estimators under CPS.

5 Conclusion

Variance estimation is needed in survey sampling to provide a measure of the precision of an estimator. As a result it is important to precisely estimate the variance itself. When the sampling algorithms are complex, the joint inclusion probabilities that are required to obtain an unbiased estimate of this variance are problematic to compute. To overcome this problem approximate variance estimators have been developed which are independent of these probabilities.

The first major result of this study was resolving the discrepancy in the results produced by Brewer and Donadio (2003) and by Matei and Tillé (2005). It can be concluded that Matei and Tillé misused the CPS algorithm, which resulted in their variances of the HTE for $n = 10$ and $n = 20$. There was also an additional error in their calculation of the variance for $n = 40$, which is so far unknown. As a result, it was not possible to conclude how they obtained high RBs for the Brewer Family estimators for $n = 40$, yet clearly their results are erroneous.

The principal aim of this thesis was to compare the behaviour of the nine different approximate variances estimators in the Brewer Family and the Hájek-Deville Family. It was initially conjectured that these families would perform better under the sampling algorithms they were designed under. A further aim was to determine whether the knowledge of the exact joint inclusion probabilities under CPS, could provide improve the variance estimation process. This final chapter provides a summary of the results for these aims, and the implications that they have for survey sampling. The next two sections discuss the behaviour of the variance estimators, and the relationship between the variance and entropy of a sampling design. The final section refers to areas of further research to extend the results of this study.

5.1 Comparison of Variance Estimators

It was difficult to find one of the nine approximate variance estimators that was uniformly superior to the rest. In regard to the RB, $\hat{\tilde{V}}_{BR4}$ performed well under RANSYS and CPS, and $\hat{\tilde{V}}_{Ber}$ also performed well under CPS. $\hat{\tilde{V}}_{BR1}$ nearly always achieved the lowest MSE for both RANSYS and CPS, however the difference compared with the other estimators was usually only trivial. Under CPS, $\hat{\tilde{V}}_{Ber}$ nearly always maximised the MSE among the approximate variance estimators. Among the Hájek-Deville Family $\hat{\tilde{V}}_{Dev1}$ always minimised the MSE. Therefore, overall $\hat{\tilde{V}}_{BR1}$ and $\hat{\tilde{V}}_{Dev1}$ were the better estimators with regard to the MSE, and $\hat{\tilde{V}}_{BR4}$ and $\hat{\tilde{V}}_{Dev2}$ with regard to the RB.

Theorem 1 in chapter 4 indicates that the $\hat{\tilde{V}}_{BR-Dev}$ is a conservative choice for the best estimator, as this estimator guarantees the lowest negative bias compared to another three estimators. This suggests that the Brewer Family is the conservative choice based on obtaining an estimator with a smaller negative bias as there were within group similarities. The empirical results of the major simulation showed that the Hájek-Deville Family usually has a lower MSE than the Brewer Family, indicating that the Hájek-Deville Family are less likely to produce poor estimates than the Brewer Family (excluding $\hat{\tilde{V}}_{BR1}$). Therefore both families have desirable properties.

This research provides evidence both for and against the hypothesis that the Brewer Family estimators perform better under RANSYS, and Hájek-Deville Family under CPS. The Brewer Family estimators always perform better under RANSYS, however, they also perform considerably better than the Hájek-Family for two populations under CPS. The hypothesis is also support by combining the results in Chapter 4.

Finally, the knowledge of the exact joint inclusion probabilities under CPS did not significantly improve the SYG estimator in regard to either the RB or the MSE. In fact, this estimator generally had the largest MSE. This estimator, however, has the desirable property that it never performed poorly in relation to the RB compared with the approximate variance estimators. The little extra effort required to compute the joint inclusion probabilities under CPS can ensure that a consistent variance estimator is obtained, but perhaps not to the extent that would justify its routine use as it often maximises the MSE.

These results have implications on the process of variance estimation in survey sampling, and in particular the choice of variance estimators. First the results indicate which estimators are likely to perform well in regard with the RB or MSE. Second they indicate that certain approximate variance estimators tend to perform better under some given sampling algorithms. This allows appropriate variance estimators to be chosen depending on the sampling algorithm in use.

5.2 Entropy and Variance

This studied disproved the original belief that a sampling design with a higher entropy will typically produce a higher variance. Although CPS maximises the entropy it usually had a smaller variance than RANSYS. The reason for this is that CPS tends to select samples containing dissimilar units more than RANSYS. These results may have implications when choosing a sampling algorithm. CPS maximises the entropy, however it was found RANSYS almost achieves this maximum entropy as well. Thus the approximate variance estimators should perform well under both designs, and they generally do. CPS, however, has the desirably property that it usually produces lower variances than RANSYS, that is, it provides more precise

estimates.

RANSYS is commonly used due to its simplicity and the ease with which it is possible to rotate samples. Although the implementation of CPS is not as simple or as fast as RANSYS, it is still clearly feasible for moderately large populations. In addition, it is also possible to rotate samples under CPS. The choice of sampling algorithm may also depend upon the preferred approximate estimator. For instance, if the conservative estimator of $\hat{\tilde{V}}_{BR-Dev}$ is chosen then RANSYS should be used. Overall there is no clear choice between RANSYS and CPS; both have their advantages.

5.3 Further Research

This thesis concludes by discussing further areas of research. There were a few interesting discoveries in Chapter 3 which were not examined further, but should definitely be considered. The first is to compare the accuracy of Hartley and Rao's full approximation of the joint inclusion probabilities with the simpler third-order approximation, the latter approximation may in fact be more accurate. The second is to determine the number of simulations which are needed to ensure that the RBs are consistent over different sets of simulations for the same population. This would enable a firm conclusion to be arrived at about the behaviour of the variance estimators over all possible samples in a population. As simulations are the main approach to comparing the behaviour of variance estimators, it is important that a sufficient number of simulations are used to ensure the comparisons are reliable.

In regards to the relationship between the entropy and the variance of sampling designs, other designs should also be considered. In particular the relationship

between the entropy and the variance should be analysed for individual samples, not just the sampling design as a whole. This may assist in finding sampling algorithms which guarantee a high entropy as well as a low variance.

References

Aires, N. (1999). Algorithms to Find Exact Inclusion Probabilities for Conditional Poisson Sampling and Pareto πps Sampling Designs, *Methodology and Computing in Applied Probability*, **4**, 457–469.

Aires, N. (2000). Comparisons between Conditional Poisson Sampling and Pareto πps Sampling Designs, *Journal of Statistical Planning and Inference*, **82**, 1–15.

Asok, C. and Sukhatme, B. (1976). On Sampford's Procedure of Unequal Probability Sampling Without Replacement, *Journal of the American Statistical Association*, **71**, 912–918.

Bellhouse, D. and Rao, J. (1975). Systematic Sampling in the Presence of a Trend, *Biometrika*, **62**, 690–697.

Berger (2004). A Simple Variance Estimator for Unequal Probability Sampling Without Replacement, *Journal of Applied Statistics*, **31**, 305–315.

Brewer, K. (2002). *Combined Survey Sampling Inference; Weighing Basu's Elephants*, London: Arnold.

Brewer, K. and Donadio, M. E. (2003). The High Entropy Variance of the Horvitz-Thompson Estimator, *Survey Methodology*, **29**, 189–196.

Brewer, K. and Hanif, M. (1983). *Sampling with Unequal Probabilities*, New York: Springer-Verlag.

Chao, M. (1982). A General Purpose Unequal Probability Sampling Plan, *Biometrika*, **69**, 653–656.

Chen, X.-H., Dempster, A. P., and Liu, J. S. (1994). Weighted Finite Population Sampling to Maximize Entropy, *Biometrika*, **81**, 457–469.

Deville, J.-C. (1999). Estimation de la Variance pour les Enquêtes en Deux Phases, note Interne Manusctrite. France: INSEE.

Deville, J.-C. (2000). Note sur l'Algorithme de Chen, Technical report, Dempster et Liu, France CREST-ENSAI.

Donadio, M. E. (2002). *Variance Estimation in πps Sampling*, Master's thesis, The University of Melbourne.

Dupačová, J. (1979). *A note on Rejective Sampling, Contributions to Statistics, Jaroslav Hájek Memorial Volume*, Prague: Reidal, Holland and Academia.

Foreman, E. (1991). *Survey Sampling Principles*, New York: Marcel Dekker, Inc.

Goodman, R. and Kish, L. (1950). Controlled Selection - a Technique in Probability Sampling, *Journal of the American Statistical Association*, **45**, 350–372.

Hájek, J. (1964). Asymptotic Theory of Rejection Sampling with Varying Probabilities from a Finite Population, *Annals of Mathematical Statistics*, **35**, 1491 – 1523.

Hájek, J. (1981). *Sampling from a Finite Population*, New York: Marcel Dekker, Inc.

Hansen, M. and Hurwitz, W. (1943). On the Theory of Sampling from Finite Populations, *Annals of Mathematical Statistics*, **14**, 333–362.

Hartley, H. and Rao, J. (1962). Sampling with Unequal Probabilities and without Replacement, *Annals of Mathematical Statistics*, **33**, 350–374.

Horvitz, D. and Thompson, D. (1952). A Generalisation of Sampling Without Replacement from a Finite Universe, *Journal of the American Statistical Association*, **47**, 663–685.

Matei, A. and Tillé, Y. (2005). Evaluation of Variance Approximations and Estimators in Maximum Entropy Sampling with Unequal Probability and Fixed Sample Size, *Journal of Official Statistics*, **21**, 543–570.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers, in *Business Survey Methods*, New York, Wiley.

Rosén, B. (1991). Variance Estimation for Systematic pps-sampling, Technical Report 15, Statistics Sweden.

Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model Assisted Survey Sampling*, New York: Springer.

Sen, A. (1953). On the Estimate of the Variance in Sampling with Varying Probabilities, *Journal of the Indian Society of Agricultural Statistics*, **5**, 119–127.

Shannon, C. E. (1948). A Mathematical Theory of Communication, *The Bell System*, **27**, 379–423,623–656.

Tillé, Y. (1996a). An Elimination Procedure for Unequal Probability Sampling Without Replacement, *Biometrika*, **83**, 238–241.

Tillé, Y. (1996b). Some Remarks on Unequal Probability Sampling Designs Without Replacement, *Annales D'Économie et de Statistique*, **44**, 177–189.

Tillé, Y. (2006). *Sampling Algorithms*, New York: Springer.

Tillé, Y. and Matei, A. (2006). *Sampling: Survey Sampling*, Department of Statistics and Mathematics of the WU Wien, Retrieved September 2, 2006, http://cran.r-project.org.

Wikipedia (2006). *Information Entropy*, Wikipedia, Retrieved September 8, 2006, http://en.wikipedia.org.

YATES, F. and GRUNDY, P. (1953). Selection Without Replacement from Within Strata with Probabilities Proportion to Size, *Journal of the Royal Statistical Society*, **15**, 235–261, series B.

A Simulation Code

A.1 Introduction

This appendix includes the code use to simulate 50,000 independent samples under RANSYS and CPS for population (a). The code can be easily modified for other populations.

The code to calculate the joint inclusion probabilities under Aires' algorithm is also provided.

A.2 RANSYS Code

```
#Simulates 50,000 samples by RANSYS
#and determines the variance estimates for each sample.

#Read in the appropriate data set.
pop <- read.table(file= "H:Data\\mu284.txt",sep ="\t", header = T)

#Obtain the Y and X variable from the population.
#This example represents Population (a)
pop <- pop[-c(16,114,137),c(4,2)]

#The simulations is conducted for all three sample sizes considered.
for(n in c(10,20,40)){
cat(n,"\n",date(),"\n") #progress output
#Note Dev3 is the BR-Dev variance estimator in this code
#Files to store all the relevant results
filenameSYG <- paste("H:\\Pop a\\Results.SYG",".",n,".save",sep="")
filenamec1 <- paste("H:\\Pop a\\Results.c1",".",n,".save",sep="")
filenamec2 <- paste("H:\\Pop a\\Results.c2",".",n,".save",sep="")
filenamec3 <- paste("H:\\Pop a\\Results.c3",".",n,".save",sep="")
filenamec4 <- paste("H:\\Pop a\\Results.c4",".",n,".save",sep="")
filenameDev <- paste("H:\\Pop a\\Results.Dev",".",n,".save",sep="")
filenameDev2 <- paste("H:\\Pop a\\Results.Dev2",".",n,".save",sep="")
filenameDev3 <- paste("H:\\Pop a\\Results.Dev3",".",n,".save",sep="")
filenameRos <- paste("H:\\Pop a\\Results.Ros",".",n,".save",sep="")
filenameBer <- paste("H:\\Pop a\\Results.Ber",".",n,".save",sep="")

#Variables to store relevant information
pop_samp <- NULL; resSYG <- NULL; resc1 <- NULL; resc2 <- NULL
resc3 <- NULL; resc4 <- NULL; resDev <- NULL; resDev2 <- NULL
resDev3 <- NULL; resRos <- NULL; resBer <- NULL
```

```
#Set up the desired inclusion probabilities.
pi_i <- n*(pop[,2])/sum(pop[,2])
enum <- NULL  #A variable to store the units included with certainty
#Subpopulation to contain all units which are not included with certainty
pop_sub <- pop

#Units are included with certainty if pi_i is greater than 1.
#The pi_i's are then recalculated for the remaining units.
#This process is repeated until all pi_i's are between 0 and 1.
while(sum(pi_i >= 1)>0){
#Combine all the units which are included with certainty
enum <- rbind(enum,pop_sub[which(pi_i>=1),])

#Redefine the population and pi_i to only include the units,
#which are not included with certainty
pop_sub <- pop_sub[-which(pi_i>=1),]

#Recalculate pi_i's with the units included with certainty excluded.
pi_i <- (n-length(enum[,1]))*(pop_sub[,2])/sum(pop_sub[,2])
}

#The number of units to be sampled if some are included with certainty
n_samp <- n-length(enum[,1])

r <- 50000  #Number of simulations to run
for(R in c(1:r)) {
#Progress output to monitor the running of the code
if(round(R/10000)-R/10000 == 0){cat(R,"\n")}

#Take a sample using RANSYS sampling
#Indicator variable.
delta <- matrix(rep(0,length(pop_sub[,1])),ncol=1)

frame <- cbind(c(1:length(pop_sub[,1])),pi_i)
#Selects a random order for the units
samp <- sample(c(1:length(pop_sub[,1])))
frame <- frame[samp,] #Places the units in the above random order
#Place 0 into first position and determine W_i.
frame <- rbind(rep(0,length(frame[1,])),frame)
frame <- cbind(frame,cumsum(frame[,2]))

j <- 0
u <- runif(1,0,1) #generates random starting point
d_samp <- NULL

#Determine which units satisfy step iv in Algorithm 1.
for(i in c(2:(length(pop_sub[,1])+1))){
if(frame[i-1,3] < u & u <= frame[i,3]){
d_samp <- c(d_samp,frame[i,1])
```

```
j <- j + 1
u <- u + 1
} }

#Place an indicator value of 1 for the units selected.
delta[d_samp,] <- 1
#Retain sampled units and their corresponding pi_i's
pop_samp <- cbind(pop_sub,pi_i,inv_pi = pi_i^(-1),delta)
pop_samp <- pop_samp[delta==1,]

#Determine the HTE of the total excluding the units included with
#certainty, as they should not be used in the variance estimators.
y_HT <- sum(pop_samp[,1]*pop_samp[,4])

w.samp <- 1-pop_samp[,3] #Define the remainder weights as 1-pi_i

#Brewer Family Estimators:
#Calculate the appropriate coefficients for Brewer variances.
#The enumerated units are excluded.
c1 <-(n_samp-1)/(n_samp-sum(pi_i^2)/n_samp)
c2 <-(n_samp-1)/(n_samp-pop_samp[,3])
c3 <-(n_samp-1)/(n_samp-2*pop_samp[,3]+sum(pi_i^2)/n_samp)
d <- (2*n_samp-1)*pop_samp[,3]/(n_samp-1)
c4 <-(n_samp-1)/(n_samp-d+sum(pi_i^2)/(n_samp-1))
  y_bar <- y_HT/n_samp

#Brewer Variance estimation formula
varc1<-sum((1/c1-pop_samp[,3])*(pop_samp[,1]*pop_samp[,4]-y_bar)^2)
varc2<-sum((1/c2-pop_samp[,3])*(pop_samp[,1]*pop_samp[,4]-y_bar)^2)
varc3<-sum((1/c3-pop_samp[,3])*(pop_samp[,1]*pop_samp[,4]-y_bar)^2)
varc4<-sum((1/c4-pop_samp[,3])*(pop_samp[,1]*pop_samp[,4]-y_bar)^2)

#Dev3 or BR-Dev
a_i <- (w.samp)/sum(w.samp); A <- 1/(1-sum(a_i^2))
varDev3 <- A*sum(w.samp*(pop_samp[,1]/pop_samp[,3]-y_HT/n_samp)^2)

#Hajek-Deville Family Estimators:
#Dev1
#Calculate the appropriate coefficient, cDev.
#The enumerated units are excluded.
cDev <- w.samp*n_samp/(n_samp-1)
Y_star <- pop_samp[,3]*sum(cDev*pop_samp[,1]/pop_samp[,3])/sum(cDev)
varDev <- sum(cDev/(pop_samp[,3]^2)*(pop_samp[,1]-Y_star)^2)

#Dev2
cDev2 <- w.samp*(1-sum((w.samp/sum(w.samp))^2))^(-1)
Y_s <- pop_samp[,3]*sum(cDev2*pop_samp[,1]/pop_samp[,3])/sum(cDev2)
varDev2 <- sum(cDev2/(pop_samp[,3]^2)*(pop_samp[,1]-Y_s)^2)
```

```
#Rosen Estimator
A2.1 <- sum(pop_samp[,1]*(w.samp/(pop_samp[,3]^2))*log10(w.samp))
A2.2 <- sum(w.samp/pop_samp[,3]*log10(w.samp))
A_s2 <- A2.1/A2.2; f <- n_samp/(n_samp-1)
varRos <- f*sum(w.samp*(pop_samp[,1]/pop_samp[,3]-A_s2)^2)

#Berger Estimator
g_k <- w.samp*n_samp/(n_samp-1)*sum(w.samp)/sum(pi_i*(1-pi_i))
Y_star_g <- pop_samp[,3]*sum(g_k*pop_samp[,1]/pop_samp[,3])/sum(g_k)
varBer <- sum(g_k/(pop_samp[,3]^2)*(pop_samp[,1]-Y_star_g)^2)

#Store the appropriate values
#The units included with certainty are now added to the HTE of the total.
resSYG <- rbind(resSYG,c(R,y_HT+sum(enum[,1]),varSYG))
resc1 <- rbind(resc1,c(R,y_HT+sum(enum[,1]),varc1))
resc2 <- rbind(resc2,c(R,y_HT+sum(enum[,1]),varc2))
resc3 <- rbind(resc3,c(R,y_HT+sum(enum[,1]),varc3))
resc4 <- rbind(resc4,c(R,y_HT+sum(enum[,1]),varc4))
resDev <- rbind(resDev,c(R,y_HT+sum(enum[,1]),varDev))
resDev2 <- rbind(resDev2,c(R,y_HT+sum(enum[,1]),varDev2))
resDev3 <- rbind(resDev3,c(R,y_HT+sum(enum[,1]),varDev3))
resRos <- rbind(resRos,c(R,y_HT+sum(enum[,1]),varRos))
resBer <- rbind(resBer,c(R,y_HT+sum(enum[,1]),varBer))
}

#Save the results for each sample size
save(file=filenamec1,resc1); save(file=filenamec2,resc2)
save(file=filenamec3,resc3); save(file=filenamec4,resc4)
save(file=filenameDev,resDev); save(file=filenameDev2,resDev2)
save(file=filenameDev3,resDev3); save(file=filenameRos,resRos)
save(file=filenameBer,resBer)
}
```

A.3 CPS Code

```
#Simulates 50,000 samples by CPS
#and determines the variance estimates for each sample.

#Read in the appropriate data set.
pop <- read.table(file= "H:Data\\mu284.txt",sep ="\t", header = T)

#Obtain the Y and X variable from the population.
#This example represents Population (a)
pop <- pop[-c(16,114,137),c(4,2)]

#The simulations is conducted for all three sample sizes considered.
for(n in c(10,20,40)){
cat(n,"\n",date(),"\n") #progress output
#Note Dev3 is the BR-Dev variance estimator in this code
```

```
#variables to store relevant information
pop_samp <- NULL; num_reject <- NULL; resSYG <- NULL; resc1 <- NULL
resc2 <- NULL; resc3 <- NULL; resc4 <- NULL; resDev <- NULL
resDev2 <- NULL; resDev3 <- NULL; resRos <- NULL; resBer <- NULL

#files to store all the relevant results
filenameSYG<-paste("H:CPS\\Pop a\\Results.SYG",".",n,".save",sep="")
filenamec1<-paste("H:CPS\\Pop a\\Results.c1",".",n,".save",sep="")
filenamec2<-paste("H:CPS\\Pop a\\Results.c2",".",n,".save",sep="")
filenamec3<-paste("H:CPS\\Pop a\\Results.c3",".",n,".save",sep="")
filenamec4<-paste("H:CPS\\Pop a\\Results.c4",".",n,".save",sep="")
filenameDev<-paste("H:CPS\\Pop a\\Results.Dev",".",n,".save",sep="")
filenameDev2<-paste("H:CPS\\Pop a\\Results.Dev2",".",n,".save",sep="")
filenameDev3<-paste("H:CPS\\Pop a\\Results.Dev3",".",n,".save",sep="")
filenameRos<-paste("H:CPS\\Pop a\\Results.Ros",".",n,".save",sep="")
filenameBer<-paste("H:CPS\\Pop a\\Results.Ber",".",n,".save",sep="")
filenameRej<-paste("H:CPS\\Pop a\\Rejected",".",n,".save",sep="")

#Set up the desired inclusion probabilities
pi_i <- n*(pop[,2])/sum(pop[,2])
enum <- NULL #A variable to store the units included with certainty

#Subpopulation to contain all units which are not included with certainty
pop_sub <- pop

#Units are included with certainty if their pi_i is greater than 1.
#The pi_i's are then recalculated for the remaining units.
#This process is repeated until all the pi_i are between 0 and 1.
while(sum(pi_i >= 1)>0){
#Combine all the units which are included with certainty
enum <- rbind(enum,pop_sub[which(pi_i>=1),])

#Redefine the population and pi_i to only include the units,
#that are not included with certainty
pop_sub <- pop_sub[-which(pi_i>=1),]

#Recalculate pi_i's with the units included with certainty excluded.
pi_i <- (n-length(enum[,1]))*(pop_sub[,2])/sum(pop_sub[,2])
}

#The number of units to be sampled if some are included with certainty
n_samp <- n-length(enum[,1])

#Determine the Poisson Sampling working probabilities, p_tilde,
#and the joint inclusion probabilities from pi_i.
p_tilde <- UPMEpiktildefrompik(pi_i)
pij <- UPmaxentropypi2(pi_i)
```

```
r <- 50000 #Number of simulations to run
for(R in c(1:r)) {
#progress output to monitor the running of the code
if(round(R/2000)-R/2000 == 0){cat(R,"\n")}

#Take a sample using CPS
N <- length(pi_i); delta <-NULL
k <- 0 #counts the number of samples rejected.
#Keep sampling until the sample has n_samp units.
while(sum(delta) != n_samp){
delta <- rbinom(N,1,p_tilde); k <- k+1}

#retain only the sampled units and their corresponding pi's
pop_samp <- cbind(pop_sub,pi_i,inv_pi = pi_i^(-1),delta)
pop_samp <- pop_samp[delta==1,]
pij_samp <- pij[delta==1,delta==1]

#Determine the HTE of the total excluding the units included with
#certainty, as they should not be used in the variance estimators.
y_HT <- sum(pop_samp[,1]*pop_samp[,4])

#SYG variance estimator:
syg_sub <- 0
for(i in c(1:(n_samp-1))){
for(j in ((i+1):n_samp)){
k2 <-(pop_samp[i,3]*pop_samp[j,3]-pij_samp[i,j])/pij_samp[i,j]
k3 <-(pop_samp[i,1]*pop_samp[i,4]-pop_samp[j,1]*pop_samp[j,4])^2
syg_sub <- syg_sub + k2/k3
} }
varSYG <- syg_sub

#Brewer Family Estimators:
#Calculate the appropriate coefficients for Brewer variances.
#The enumerated units are excluded.
c1 <-(n_samp-1)/(n_samp-sum(pi_i^2)/n_samp)
c2 <-(n_samp-1)/(n_samp-pop_samp[,3])
c3 <-(n_samp-1)/(n_samp-2*pop_samp[,3]+sum(pi_i^2)/n_samp)
d <- (2*n_samp-1)*pop_samp[,3]/(n_samp-1)
c4 <-(n_samp-1)/(n_samp-d+sum(pi_i^2)/(n_samp-1))
y_bar <- y_HT/n_samp

#Brewer Variance estimation formula
varc1 <-sum((1/c1-pop_samp[,3])*(pop_samp[,1]*pop_samp[,4]-y_bar)^2)
varc2 <-sum((1/c2-pop_samp[,3])*(pop_samp[,1]*pop_samp[,4]-y_bar)^2)
varc3 <-sum((1/c3-pop_samp[,3])*(pop_samp[,1]*pop_samp[,4]-y_bar)^2)
varc4 <-sum((1/c4-pop_samp[,3])*(pop_samp[,1]*pop_samp[,4]-y_bar)^2)

#Dev3 or BR-Dev
a_i <- (w.samp)/sum(w.samp); A <- 1/(1-sum(a_i^2))
```

```
varDev3 <- A*sum(w.samp*(pop_samp[,1]/pop_samp[,3]-y_HT/n_samp)^2)

#Hajek-Deville Family Estimators:
#Dev1
#Calculate the appropriate coefficient, cDev.
#The enumerated units are excluded.
cDev <- w.samp*n_samp/(n_samp-1)
Y_star <- pop_samp[,3]*sum(cDev*pop_samp[,1]/pop_samp[,3])/sum(cDev)
varDev <- sum(cDev/(pop_samp[,3]^2)*(pop_samp[,1]-Y_star)^2)

#Dev2
cDev2 <- w.samp*(1-sum((w.samp/sum(w.samp))^2))^(-1)
Y_s <- pop_samp[,3]*sum(cDev2*pop_samp[,1]/pop_samp[,3])/sum(cDev2)
varDev2 <- sum(cDev2/(pop_samp[,3]^2)*(pop_samp[,1]-Y_star2)^2)

#Rosen Estimator
A2.1 <- sum(pop_samp[,1]*(w.samp/(pop_samp[,3]^2))*log10(w.samp))
A2.2 <- sum(w.samp/pop_samp[,3]*log10(w.samp))
A_s2 <- A2.1/A2.2;  f <- n_samp/(n_samp-1)
varRos <- f*sum(w.samp*(pop_samp[,1]/pop_samp[,3]-A_s2)^2)

#Berger Estimator
g_k <- w.samp*n_samp/(n_samp-1)*sum(w.samp)/sum(pi_i*(1-pi_i))
Y_star_g <- pop_samp[,3]*sum(g_k*pop_samp[,1]/pop_samp[,3])/sum(g_k)
varBer <- sum(g_k/(pop_samp[,3]^2)*(pop_samp[,1]-Y_star_g)^2)

#Store the appropriate values
#The units included with certainty are now added to the HTE of the total.
resSYG <- rbind(resSYG,c(R,y_HT+sum(enum[,1]),varSYG))
resc1 <- rbind(resc1,c(R,y_HT+sum(enum[,1]),varc1))
resc2 <- rbind(resc2,c(R,y_HT+sum(enum[,1]),varc2))
resc3 <- rbind(resc3,c(R,y_HT+sum(enum[,1]),varc3))
resc4 <- rbind(resc4,c(R,y_HT+sum(enum[,1]),varc4))
resDev <- rbind(resDev,c(R,y_HT+sum(enum[,1]),varDev))
resDev2 <- rbind(resDev2,c(R,y_HT+sum(enum[,1]),varDev2))
resDev3 <- rbind(resDev3,c(R,y_HT+sum(enum[,1]),varDev3))
resRos <- rbind(resRos,c(R,y_HT+sum(enum[,1]),varRos))
resBer <- rbind(resBer,c(R,y_HT+sum(enum[,1]),varBer))
num_reject <- rbind(num_reject,c(R,k))
}
#Save the results for each sample size
save(file=filenameSYG,resSYG); save(file=filenamec1,resc1)
save(file=filenamec2,resc2); save(file=filenamec3,resc3)
save(file=filenamec4,resc4); save(file=filenameDev,resDev)
save(file=filenameDev2,resDev2); save(file=filenameDev3,resDev3)
save(file=filenameRos,resRos); save(file=filenameBer,resBer)
save(file=filenameRej,num_reject)
}
```

A.4 Aires' Algorithm

```
#This function determines the joint inclusion probabilities
#using Aires' Algorithm.
#pi_i is a vector of the desired inclusion probabilities.
#p is a vector of the working probabilities determined from pi_i.
#The R sampling package was used to determine p.
#see Aires' Algorithm in section 2.1.2
second_order <- function(p,pi_i){

N <- length(p); gamma <- p/(1-p)
gamma_eq <- NULL
pi_mat <- matrix(0,nrow = N-1,ncol=N)

#Calculate the joint inclusion probabilities, pi_ij, for all i
#and only for j >i as pi_ij is symmetric.
for(i in c(1:(N-1))){

for(j in c((i+1):N)){
#Note due to rounding errors the condition of gamma[i]==gamma[j],
#was tested as the absolute difference being less than 10^(-10).
if(abs(gamma[i] - gamma[j]) > 10^(-10)){
pi_ij <- (gamma[i]*pi_i[j]-gamma[j]*pi_i[i])/(gamma[i]-gamma[j])
}
else{pi_ij <- NA
gamma_eq <- rbind(gamma_eq,c(i,j)) }
pi_mat[i,j] <- pi_ij
} }

#Determine the inclusion probabilities when gamma[i]==gamma[j].
for(r in unique(gamma_eq[,1])){
i <- r
j <- gamma_eq[which(gamma_eq[,1]==i),2]

if(i == 1){vec <-  pi_mat[i,c((i+1):N)]}
else {vec <- c(pi_mat[i,c((i+1):N)],pi_mat[c(1:(i-1)),i])}
k <- sum(is.na(vec))

pi_mat[i,j] <- ((n-1)*pi_i[i] - sum(vec,na.rm=TRUE))/k
}
#returns the joint inclusion probabilities
return(pi_mat)
}
```

B Effect of 10,000 simulations

	n=10			n=20		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
$\hat{V}_{SYG}$	0.0580	0.0582	0.0426	0.1024	0.1126	0.0904
$\hat{\hat{V}}_{BR1}$	0.0000	0.0000	0.0000	0.0039	0.0016	0.0031
$\hat{\hat{V}}_{BR2}$	0.0091	0.0092	0.0085	0.0136	0.0117	0.0143
$\hat{\hat{V}}_{BR3}$	0.0184	0.0185	0.0171	0.0234	0.0219	0.0255
$\hat{\hat{V}}_{BR4}$	0.0194	0.0195	0.0180	0.0239	0.0224	0.0261
$\hat{\hat{V}}_{BR-Dev}$	0.0127	0.0124	0.0119	0.0214	0.0191	0.0224
$\hat{\hat{V}}_{Dev1}$	0.0056	0.0056	0.0050	0.0000	0.0000	0.0000
$\hat{\hat{V}}_{Dev2}$	0.0091	0.0088	0.0084	0.0077	0.0072	0.0080
$\hat{\hat{V}}_{Ros}$	0.0065	0.0065	0.0059	0.0033	0.0029	0.0035
$\hat{\hat{V}}_{Ber}$	0.0532	0.0535	0.0381	0.0928	0.1029	0.0806

Table B.1: Profile of the MSEs for three trials of 10,000 simulations

C Effect of 50,000 simulations

C.1 Population (a)

	n=10			n=20			n=40		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
$\hat{\hat{V}}_{BR1}$	0.0220	0.0207	0.0217	0.2720	0.2729	0.2714	1.7172	1.7118	1.7146
$\hat{\hat{V}}_{BR2}$	0.0809	0.0808	0.0807	0.3391	0.3400	0.3393	1.7984	1.7919	1.7951
$\hat{\hat{V}}_{BR3}$	0.1398	0.1408	0.1398	0.4061	0.4072	0.4072	1.8796	1.8720	1.8755
$\hat{\hat{V}}_{BR4}$	0.1463	0.1475	0.1463	0.4096	0.4107	0.4108	1.8816	1.8741	1.8776
$\hat{\hat{V}}_{BR-Dev}$	0.1115	0.1116	0.1113	0.4110	0.4118	0.4112	2.0026	1.9957	1.9988
$\hat{\hat{V}}_{Dev1}$	0	0	0	0	0	0	0	0	0
$\hat{\hat{V}}_{Dev2}$	0.0306	0.0308	0.0305	0.0717	0.0715	0.0716	0.2004	0.2000	0.2000
$\hat{\hat{V}}_{Ros}$	0.0193	0.0193	0.0193	0.0759	0.0761	0.0760	0.3084	0.3072	0.3075
$\hat{\hat{V}}_{Ber}$	0.0708	0.0607	0.0751	0.0709	0.0531	0.0986	0.0796	0.0544	0.0859

Table C.1: Profile of the RBs for three trials of 50,000 simulations - population (a)

	n=10			n=20			n=40		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
$\hat{\hat{V}}_{BR1}$	0	0	0	0.0075	0.0062	0.0071	0.0415	0.0398	0.0398
$\hat{\hat{V}}_{BR2}$	0.0085	0.0088	0.0084	0.0178	0.0165	0.0177	0.0543	0.0521	0.0520
$\hat{\hat{V}}_{BR3}$	0.0172	0.0176	0.0169	0.0282	0.0269	0.0285	0.0672	0.0647	0.0644
$\hat{\hat{V}}_{BR4}$	0.0181	0.0186	0.0179	0.0287	0.0275	0.0291	0.0676	0.0650	0.0647
$\hat{\hat{V}}_{BR-Dev}$	0.0121	0.0123	0.0119	0.0262	0.0245	0.0260	0.0734	0.0709	0.0705
$\hat{\hat{V}}_{Dev1}$	0.0040	0.0043	0.0042	0	0	0	0	0	0
$\hat{\hat{V}}_{Dev2}$	0.0075	0.0079	0.0077	0.0082	0.0078	0.0081	0.0142	0.0138	0.0136
$\hat{\hat{V}}_{Ros}$	0.0051	0.0054	0.0053	0.0043	0.0040	0.0043	0.0105	0.0101	0.0101
$\hat{\hat{V}}_{Ber}$	0.0472	0.0427	0.0414	0.0761	0.0798	0.0877	0.1725	0.1649	0.1722

Table C.2: Profile of the MSEs for three trials of 50,000 simulations - population (a)

	Trial 1			Trial 2			Trial 3		
	10	20	40	10	20	40	10	20	40
$\hat{\tilde{V}}_{BR1}$	0.0460	-0.0764	-0.0040	-0.1376	-0.5060	-0.2102	-0.5040	-0.0879	-0.2422
$\hat{\tilde{V}}_{BR2}$	0.1049	-0.0094	0.0772	-0.0775	-0.4389	-0.1301	-0.4450	-0.0200	-0.1617
$\hat{\tilde{V}}_{BR3}$	0.1637	0.0577	0.1584	-0.0175	-0.3717	-0.0500	-0.3860	0.0480	-0.0812
$\hat{\tilde{V}}_{BR4}$	0.1703	0.0612	0.1605	-0.0108	-0.3682	-0.0479	-0.3794	0.0515	-0.0792
$\hat{\tilde{V}}_{BR-Dev}$	0.1355	0.0626	0.2814	-0.0467	-0.3671	0.0737	-0.4145	0.0519	0.0421
$\hat{\tilde{V}}_{Dev1}$	0.0240	-0.3484	-1.7212	-0.1583	-0.7789	-1.9220	-0.5257	-0.3593	-1.9567
$\hat{\tilde{V}}_{Dev2}$	0.0546	-0.2767	-1.5208	-0.1275	-0.7074	-1.7219	-0.4953	-0.2877	-1.7568
$\hat{\tilde{V}}_{Ros}$	0.0433	-0.2725	-1.4128	-0.1390	-0.7028	-1.6148	-0.5065	-0.2833	-1.6493
$\hat{\tilde{V}}_{Ber}$	0.0948	-0.2775	-1.6416	-0.0976	-0.7258	-1.8676	-0.4506	-0.2607	-1.8708

Table C.3: RBs (%) for three trials of 50,000 simulations - population (a)

	Trial 1			Trial 2			Trial 3		
	10 (10^{12})	20 (10^{11})	40 (10^{10})	10 (10^{12})	20 (10^{11})	40 (10^{10})	10 (10^{12})	20 (10^{11})	40 (10^{10})
$\hat{\tilde{V}}_{BR1}$	6.3950	6.2152	4.7785	6.3181	6.0992	4.7934	6.2248	6.1471	4.7640
$\hat{\tilde{V}}_{BR2}$	6.4036	6.2254	4.7913	6.3269	6.1095	4.8058	6.2332	6.1578	4.7762
$\hat{\tilde{V}}_{BR3}$	6.4122	6.2358	4.8042	6.3357	6.1199	4.8183	6.2417	6.1686	4.7886
$\hat{\tilde{V}}_{BR4}$	6.4132	6.2364	4.8046	6.3367	6.1205	4.8186	6.2427	6.1691	4.7889
$\hat{\tilde{V}}_{BR-Dev}$	6.4071	6.2339	4.8104	6.3304	6.1175	4.8245	6.2366	6.1661	4.7947
$\hat{\tilde{V}}_{Dev1}$	6.3991	6.2076	4.7370	6.3224	6.0930	4.7536	6.2290	6.1400	4.7242
$\hat{\tilde{V}}_{Dev2}$	6.4026	6.2159	4.7512	6.3260	6.1008	4.7674	6.2324	6.1481	4.7378
$\hat{\tilde{V}}_{Ros}$	6.4002	6.2119	4.7475	6.3235	6.0970	4.7638	6.2301	6.1443	4.7343
$\hat{\tilde{V}}_{Ber}$	6.4422	6.2837	4.9095	6.3608	6.1728	4.9186	6.2662	6.2277	4.8964

Table C.4: MSEs for three trials of 50,000 simulations - population (a)

C.2 Population (d)

	n=10			n=20			n=40		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
$\hat{\hat{V}}_{BR1}$	0	0	0	0	0	0	0	0	0
$\hat{\hat{V}}_{BR2}$	2.8614	2.8522	2.8659	1.7172	1.7266	1.7200	0.8584	0.8616	0.8602
$\hat{\hat{V}}_{BR3}$	5.7229	5.7043	5.7318	3.4344	3.4532	3.4399	1.7168	1.7231	1.7203
$\hat{\hat{V}}_{BR4}$	6.0805	6.0608	6.0900	3.5775	3.5971	3.5832	1.7541	1.7606	1.7577
$\hat{\hat{V}}_{BR-Dev}$	5.2929	5.2825	5.3017	3.2821	3.2970	3.2834	2.1855	2.1920	2.1903
$\hat{\hat{V}}_{Dev1}$	0.9790	0.9633	1.0032	0.8290	0.8424	0.8451	0.1824	0.1762	0.1750
$\hat{\hat{V}}_{Dev2}$	3.3579	3.3407	3.3869	2.3786	2.3975	2.3935	1.5003	1.4972	1.4957
$\hat{\hat{V}}_{Ros}$	1.2877	1.2740	1.3095	0.9447	0.9577	0.9589	0.3101	0.3058	0.3048
$\hat{\hat{V}}_{Ber}$	4.6109	4.4331	4.6001	2.3553	2.3633	2.3796	1.0409	1.0112	0.9793

Table C.5: Profile of the RBs for three trials of 50,000 simulations - population (d)

	n=10			n=20			n=40		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
$\hat{\hat{V}}_{BR1}$	0	0	0	0	0	0	0	0	0
$\hat{\hat{V}}_{BR2}$	0.3094	0.3018	0.3064	0.0992	0.0999	0.0991	0.1155	0.1144	0.1152
$\hat{\hat{V}}_{BR3}$	0.6377	0.6222	0.6317	0.2039	0.2055	0.2038	0.2371	0.2349	0.2364
$\hat{\hat{V}}_{BR4}$	0.6800	0.6635	0.6737	0.2129	0.2145	0.2128	0.2426	0.2403	0.2418
$\hat{\hat{V}}_{BR-Dev}$	0.5209	0.5099	0.5159	0.1677	0.1701	0.1680	0.2151	0.2137	0.2151
$\hat{\hat{V}}_{Dev1}$	0.2478	0.2415	0.2449	0.1026	0.1029	0.1029	0.0811	0.0778	0.0802
$\hat{\hat{V}}_{Dev2}$	0.4514	0.4417	0.4465	0.1697	0.1716	0.1703	0.1750	0.1712	0.1743
$\hat{\hat{V}}_{Ros}$	0.2605	0.2542	0.2577	0.1039	0.1042	0.1041	0.0851	0.0822	0.0843
$\hat{\hat{V}}_{Ber}$	1.0522	0.9957	1.0465	0.3938	0.3806	0.3832	0.8632	0.8341	0.8295

Table C.6: Profile of the MSEs for three trials of 50,000 simulations - population (d)

	Trial 1			Trial 2			Trial 3		
	10	20	40	10	20	40	10	20	40
$\hat{\tilde{V}}_{BR1}$	-5.8835	-4.2768	-1.7398	-6.1058	-3.8232	-1.5922	-5.8158	-4.1370	-1.6812
$\hat{\tilde{V}}_{BR2}$	-3.0221	-2.5596	-0.8813	-3.2536	-2.0966	-0.7306	-2.9499	-2.4170	-0.8210
$\hat{\tilde{V}}_{BR3}$	-0.1606	-0.8424	-0.0229	-0.4015	-0.3700	0.1310	-0.0840	-0.6971	0.0392
$\hat{\tilde{V}}_{BR4}$	0.1970	-0.6993	0.0144	-0.0449	-0.2261	0.1684	0.2743	-0.5538	0.0766
$\hat{\tilde{V}}_{BR-Dev}$	-0.5906	-0.9946	0.4458	-0.8233	-0.5262	0.5998	-0.5140	-0.8535	0.5091
$\hat{\tilde{V}}_{Dev1}$	-4.9045	-3.4478	-1.5573	-5.1425	-2.9808	-1.4160	-4.8126	-3.2919	-1.5061
$\hat{\tilde{V}}_{Dev2}$	-2.5256	-1.8982	-0.2395	-2.7651	-1.4257	-0.0950	-2.4288	-1.7434	-0.1854
$\hat{\tilde{V}}_{Ros}$	-4.5958	-3.3321	-1.4296	-4.8317	-2.8655	-1.2864	-4.5062	-3.1781	-1.3764
$\hat{\tilde{V}}_{Ber}$	-1.2726	-1.9215	-0.6988	-1.6727	-1.4599	-0.5809	-1.2156	-1.7574	-0.7019

Table C.7: RBs (%) for three trials of 50,000 simulations - population (d)

	Trial 1			Trial 2			Trial 3		
	10 (10^{20})	20 (10^{19})	40 (10^{17})	10 (10^{20})	20 (10^{19})	40 (10^{17})	10 (10^{20})	20 (10^{19})	40 (10^{17})
$\hat{\tilde{V}}_{BR1}$	4.1502	2.2413	3.7613	4.0615	2.2436	3.7120	4.1097	2.2360	3.7415
$\hat{\tilde{V}}_{BR2}$	4.4596	2.3405	3.8769	4.3633	2.3436	3.8265	4.4162	2.3351	3.8567
$\hat{\tilde{V}}_{BR3}$	4.7879	2.4452	3.9984	4.6836	2.4491	3.9470	4.7414	2.4398	3.9779
$\hat{\tilde{V}}_{BR4}$	4.8303	2.4542	4.0039	4.7250	2.4582	3.9523	4.7834	2.4488	3.9833
$\hat{\tilde{V}}_{BR-Dev}$	4.6711	2.4091	3.9764	4.5714	2.4138	3.9257	4.6256	2.4040	3.9566
$\hat{\tilde{V}}_{Dev1}$	4.3980	2.3440	3.8424	4.3030	2.3465	3.7899	4.3546	2.3388	3.8217
$\hat{\tilde{V}}_{Dev2}$	4.6016	2.4111	3.9363	4.5032	2.4152	3.8833	4.5563	2.4063	3.9158
$\hat{\tilde{V}}_{Ros}$	4.4108	2.3452	3.8464	4.3157	2.3478	3.7942	4.3675	2.3400	3.8258
$\hat{\tilde{V}}_{Ber}$	5.2025	2.6352	4.6245	5.0571	2.6242	4.5461	5.1563	2.6192	4.5710

Table C.8: MSEs for three trials of 50,000 simulations - population (d)

C.3 Variances of the Mean Squared Errors

	Population (a)			Population (d)		
	10	20	40	10	20	40
$\hat{V}_{BR1}$	0.007269	0.003398	0.000216	0.001975	0.000016	0.000615
$\hat{V}_{BR2}$	0.007279	0.003392	0.000218	0.002328	0.000018	0.000643
$\hat{V}_{BR3}$	0.007290	0.003386	0.000221	0.002728	0.000022	0.000672
$\hat{V}_{BR4}$	0.007291	0.003386	0.000221	0.002781	0.000022	0.000673
$\hat{V}_{BR-Dev}$	0.007289	0.003416	0.000221	0.002494	0.000024	0.000653
$\hat{V}_{Dev1}$	0.007252	0.003321	0.000217	0.002261	0.000015	0.000700
$\hat{V}_{Dev2}$	0.007261	0.003345	0.000220	0.002427	0.000020	0.000715
$\hat{V}_{Ros}$	0.007259	0.003338	0.000217	0.002265	0.000016	0.000692
$\hat{V}_{Ber}$	0.007761	0.003079	0.000124	0.005515	0.000067	0.001603

Table C.9: Variances of the MSEs across the three trials of 50,000 simulations

D Relative Biases - Population (b)

	RANSYS		
	10	20	40
$\hat{V}_{SYG}$	0.0000	0.0000	0.0000
$\hat{\hat{V}}_{BR1}$	-0.1955	0.2312	-0.5642
$\hat{\hat{V}}_{BR2}$	-0.1309	0.3072	-0.4748
$\hat{\hat{V}}_{BR3}$	-0.0663	0.3831	-0.3854
$\hat{\hat{V}}_{BR4}$	-0.0591	0.3871	-0.3831
$\hat{\hat{V}}_{BR-Dev}$	-0.1005	0.3787	-0.2737
$\hat{\hat{V}}_{Dev1}$	-0.1934	0.0615	-1.7103
$\hat{\hat{V}}_{Dev2}$	-0.1630	0.1329	-1.5118
$\hat{\hat{V}}_{Ros}$	-0.1785	0.1167	-1.4933
$\hat{\hat{V}}_{Ber}$	-0.1215	0.1360	-1.6148

Table D.1: Relative biases (%) using Brewer and Donadio's formula (see equation (2.36)) - population (b)

E Entropy and Variance

E.1 Sampling Design - Second Small Population

Sample	Estimate	CPS	RANSYS	OSYS
(1,2,3)	1531.3889	0.0009	0.0010	0
(1,2,4)	1614.1667	0.0019	0.0033	0
(1,2,5)	1781.4116	0.0299	0.0635	0
(1,3,4)	1779.7222	0.0033	0.0034	0.0138
(1,3,5)	1946.9671	0.0513	0.0626	0.1884
(1,4,5)	2029.7449	0.1140	0.0663	0
(2,3,4)	1903.8889	0.0073	0.0057	0
(2,3,5)	2071.1338	0.1125	0.0659	0.0066
(2,4,5)	2153.9116	0.2501	0.2629	0.3955
(3,4,5)	2319.4671	0.4287	0.4655	0.3957

Table E.1: Sampling design - second small population having a total of 2175.00

E.2 Sampling Design - Third Small Population

Sample	Estimate	CPS	RANSYS	OSYS
(1,2,3)	1539.9060	0.0003	0.0049	0
(1,2,4)	1568.4940	0.0005	0.0045	0
(1,2,5)	1564.1490	0.0014	0.0024	0
(1,2,6)	1535.6170	0.0027	0.0022	0
(1,3,4)	1532.7700	0.0015	0.0023	0
(1,3,5)	1528.4240	0.0038	0.0095	0
(1,3,6)	1499.8930	0.0074	0.0095	0
(1,4,5)	1557.0130	0.0066	0.0094	0.0691
(1,4,6)	1528.4810	0.0129	0.0096	0
(1,5,6)	1524.1360	0.0330	0.0164	0
(2,3,4)	1527.7280	0.0037	0.0059	0
(2,3,5)	1523.3820	0.0096	0.0125	0
(2,3,6)	1494.8510	0.0187	0.0389	0
(2,4,5)	1551.9710	0.0167	0.0123	0.0201
(2,4,6)	1523.4390	0.0327	0.0390	0.1495
(2,5,6)	1519.0940	0.0836	0.0465	0
(3,4,5)	1516.2470	0.0459	0.0266	0
(3,4,6)	1487.7150	0.0896	0.0530	0.0203
(3,5,6)	1483.3700	0.2295	0.2451	0.3910
(4,5,6)	1511.9580	0.3999	0.4494	0.3500

Table E.2: Sampling design - third small population having a total of 1506.00

E.3 Joint Inclusion Probabilities

The joint inclusion probabilities for the first small population:

$$\pi_{ij}(CPS) = \begin{pmatrix} 0.3474 & 0.1595 & 0.1860 & 0.1439 & 0.2054 \\ & 0.6316 & 0.3836 & 0.3035 & 0.4166 \\ & & 0.6947 & 0.3499 & 0.4700 \\ & & & 0.5895 & 0.3816 \\ & & & & 0.7368 \end{pmatrix}$$

$$\pi_{ij}(RANSYS) = \begin{pmatrix} 0.3474 & 0.1519 & 0.1800 & 0.1507 & 0.2074 \\ & 0.6316 & 0.3927 & 0.3012 & 0.4217 \\ & & 0.6947 & 0.3492 & 0.4689 \\ & & & 0.5895 & 0.3764 \\ & & & & 0.7368 \end{pmatrix}$$

$$\pi_{ij}(OSYS) = \begin{pmatrix} 0.3474 & 0.3444 & 0.2605 & 0.0000 & 0.0839 \\ & 0.6316 & 0.3232 & 0.2226 & 0.3692 \\ & & 0.6947 & 0.3703 & 0.4330 \\ & & & 0.5895 & 0.5929 \\ & & & & 0.7368 \end{pmatrix}$$

The joint inclusion probabilities for the second small population:

$$\pi_{ij}(CPS) = \begin{pmatrix} 0.2013 & 0.0327 & 0.0555 & 0.1193 & 0.1952 \\ & 0.4027 & 0.1207 & 0.2593 & 0.3926 \\ & & 0.6040 & 0.4393 & 0.5925 \\ & & & 0.8054 & 0.7928 \\ & & & & 0.9866 \end{pmatrix}$$

$$\pi_{ij}(RANSYS) = \begin{pmatrix} 0.2013 & 0.0678 & 0.0670 & 0.0729 & 0.1924 \\ & 0.4027 & 0.0726 & 0.2718 & 0.3923 \\ & & 0.6040 & 0.4745 & 0.5939 \\ & & & 0.8054 & 0.7946 \\ & & & & 0.9866 \end{pmatrix}$$

$$\pi_{ij}(OSYS) = \begin{pmatrix} 0.2013 & 0 & 0.20219 & 0.0138 & 0.18844 \\ & 0.4027 & 0.00655 & 0.3955 & 0.40208 \\ & & 0.6040 & 0.4095 & 0.59072 \\ & & & 0.8054 & 0.79126 \\ & & & & 0.9866 \end{pmatrix}$$

F Pairwise Entropy

F.1 Population $N = 8$ and $n = 2$

Label	Y	X	π_i	$\tilde{p}$
1	95	90	0.0952	0.1122
2	95	100	0.1058	0.1236
3	125	120	0.1270	0.1459
4	132	140	0.1481	0.1675
5	170	160	0.1693	0.1884
6	200	180	0.1905	0.2087
7	460	500	0.5291	0.4820
8	550	600	0.6349	0.5718
Total	1827	1890		

Table F.1: Population N=8 and n=2

Sample	Estimate	CPS	RANSYS	OSYS
(1,2)	1895.2500	0.0041	0.0048	0
(1,3)	1981.8750	0.0050	0.0062	0
(1,4)	1888.5000	0.0058	0.0075	0
(1,5)	2001.5630	0.0068	0.0093	0
(1,6)	2047.5000	0.0077	0.0111	0
(1,7)	1866.9000	0.0271	0.0229	0.0949
(1,8)	1863.7500	0.0388	0.0334	0
(2,3)	1882.1250	0.0055	0.0074	0
(2,4)	1788.7500	0.0065	0.0090	0
(2,5)	1901.8130	0.0075	0.0099	0
(2,6)	1947.7500	0.0086	0.0126	0
(2,7)	1767.1500	0.0302	0.0252	0.1057
(2,8)	1764.0000	0.0433	0.0378	0
(3,4)	1875.3750	0.0079	0.0104	0
(3,5)	1988.4380	0.0091	0.0132	0
(3,6)	2034.3750	0.0104	0.0149	0
(3,7)	1853.7750	0.0366	0.0298	0.1266
(3,8)	1850.6250	0.0525	0.0447	0
(4,5)	1895.0630	0.0107	0.0159	0
(4,6)	1941.0000	0.0122	0.0181	0
(4,7)	1760.4000	0.0431	0.0338	0.0367
(4,8)	1757.2500	0.0618	0.0544	0.1118
(5,6)	2054.0630	0.0141	0.0211	0
(5,7)	1873.4630	0.0497	0.0396	0
(5,8)	1870.3130	0.0713	0.0604	0.1697
(6,7)	1919.4000	0.0565	0.0439	0
(6,8)	1916.2500	0.0811	0.0680	0.1907
(7,8)	1735.6500	0.2860	0.3349	0.1639

Table F.2: Sampling design - population N=8 and n=2