EUROPEAN COMMISSION
Community Research

SEVENTH FRAMEWORK PROGRAMME

AMELI
Advanced Methodology for European Laeken Indicators

Deliverable 3.1

Variance Estimation for Complex Surveys

Version: 2011

Christian Bruch, Ralf Münnich and Stefan Zins

The project **FP7–SSH–2007–217322 AMELI** is supported by European Commission funding from the Seventh Framework Programme for Research.

http://ameli.surveystatistics.net/

Contributors to Deliverable 3.1

Chapter 1: Christian Bruch, Ralf Münnich and Stefan Zins, University of Trier.

Chapter 2: Ralf Münnich and Stefan Zins, University of Trier.

Chapter 3: Christian Bruch, Ralf Münnich and Stefan Zins, University of Trier.

Chapter 4: Christian Bruch, Ralf Münnich and Stefan Zins, University of Trier.

Chapter 5: Christian Bruch, Ralf Münnich and Stefan Zins, University of Trier.

Main responsibility

Christian Bruch and Ralf Münnich and Stefan Zins, University of Trier

Evaluators

Internal expert: Monique Graf, Swiss Federal Statistical Office.

Aim and Objectives of Deliverable 3.1

Within the European Statistical System (ESS) more and more quality reports have to be produced for the different surveys. One component of these reports contains information on the accuracy of the estimates. In general, this information is based on variance components which have to be estimated from the same sample. Hence, variance estimation techniques play an important role in modern survey statistics.

Measuring the accuracy of estimates such as totals, means or proportions generally requires to apply the appropriate variance estimation methodology. This allows for deriving standard errors, confidence intervals and other similar measures as well as design effects which help measuring the impact of sampling designs on the outcome of the estimates. In general, two different methodologies can be applied for variance estimation of non-linear estimators, linearization and resampling methods. The efficiency of the variance estimators is mainly based on the properties of the point estimators as well as of the possible skewness of the variable of interest. Further, complex sampling designs may question the appropriateness of the variance estimation methodology and may lead to unusual conclusions which are of major practical impact.

The methodology captured in work package 3 covers two of the most frequently encountered problems of variance estimation in modern sample surveys. The first one is the presence of complex sampling designs, these issues are mainly dealt with in deliverable 3.1. The second one is the necessity of using complex (point) estimators, e.g. if poverty indicators need to be estimated or their change over time, which is the main focus of deliverable 3.2 (see Münnich and Zins, 2011).

The aim of deliverable 3.1 is to give an overview of modern variance estimation methods. The efficiency of the variance estimation methods can only be tested in a practical environment which is achieved within a large scale simulation study.

The deliverable is structured in the following way. The introduction provides a framework with some much needed nomenclature and gives also a short summary on well known variance estimators. Thereby elements of complex sample survey designs such as stratification and multistage sampling are already introduced. Chapter 2 is dedicated to the problem of variance estimator in the presence of unequal probability sampling. It is for instance not uncommon in household surveys, such as in EU-SILC, to sample households with probability proportional to size, i.e. the number of persons living in the particular household. Chapter 3 gives an overview on the existing resampling methods and deals with their implementation in the presence of complex sampling designs, in particular with multi-stage sampling and unequal selection probabilities. Chapter 4 includes the results from a small simulation study, which analyses explicitly the effects of the cluster sampling at the first stage in two-stage sampling designs. That is, we examined the influence on variance estimates of large variations in the size of the Primary Sampling Units (PSUs), the mean value of the PSUs and the variance within the PSUs.

Contents

Chapter 1

Introduction

In the European Statistical System quality reports gain increasingly in importance. Especially the accuracy measurement of point estimators like totals, means or ratios plays a major role. This can be done by using measures like the root mean square error or providing confidence intervals for estimates. A central component of all criteria is the variance of the point estimator which is also an accuracy criterion by itself (cf. MÜNNICH, 2008, p. 319 f.). In practice the true value of the variance is unknown and therefore has to be estimated from the sample. The complexity of the variance estimation depends highly on the sampling designs or parameters like the size of the Primary Sampling Units (PSUs) or the homogeneity of the units within and between the PSUs. The aim of this deliverable is to present different variance estimation methods which can be used for different sampling designs and constellations of parameters.

First, we introduce a general framework which serves as the basis of our analysis. Let us consider the finite population $\mathcal{U}$ of N identifiable units, so that they can be represented by integers 1,2, ..., N, $\mathcal{U} = \{1, \ldots, N\}$. Now we want to draw a sample of n units from $\mathcal{U}$ by means of random sample without replacement. A sample s can be defined by the following column vector (cf. TILLÉ, 2006, p. 8)

$$\mathfrak{s} = (\mathcal{I}_1, \ldots, \mathcal{I}_i, \ldots, \mathcal{I}_N)^T \in \{0,1\}^N \ , \tag{1.1}$$

where

$$\mathcal{I}_i = \begin{cases} 1 & \text{if} \quad i \in s \\ 0 & \text{if} \quad i \notin s \end{cases} \quad .$$

Now a design specifies a probability distribution function $p(.)$ on

$$\mathcal{S}_n := \{\mathfrak{s} \in \{0,1\}^n | \sum_{i \in \mathcal{U}} \mathcal{I}_i = n\} \ ,$$

i.e on all non-empty subsets s of n distinct units from U, (cf. TILLÉ, 2006, p. 10 and 14) hence,

$$\sum_{s \in \mathcal{S}_n} p(s) = 1 \qquad \text{and} \qquad p(s) \geq 0 \quad \forall\, s \in \mathcal{S}_n \qquad .$$

If $p(.)$ satisfies the above condition, we call it a fixed size sampling design without replacement, (cf. BERGER, 1998a, p. 149f.). A sample S is then drawn by means of $p(.)$

such that $Pr(S = s) = p(s)$. The probability of inclusion for the i-th element in $\mathcal{U}$ is denoted by π_i, with $\pi_i = \sum_{s \ni i} p(s)$ and the second order inclusion probability π_{ij}, with $\pi_{ij} = \sum_{s \ni \{i,j\}} p(s)$. We can now use the Horvitz-Thompson estimator

$$\hat{\tau} = \sum_{i \in s} y_i \cdot \frac{1}{\pi_i}\,, \tag{1.2}$$

to estimate the total $\tau = \sum_{i \in \mathcal{U}} y_i$.

Two basic sampling designs of this paper are stratified random sampling and two stage cluster sampling. Further sampling designs in this paper are mainly combinations of this sampling designs. In case of stratified random sampling the population is divided into H strata h $(h = 1, \ldots, H)$ of size N_h with $N_1 + N_2 + ... + N_h = N$. In each stratum a sample of size n_h is drawn. When the parameter of interest is a total value the variance of this estimator is computed by:

$$\mathrm{V}\left(\hat{\tau}_{StrRS}\right) = \sum_{h=1}^{H} \mathrm{V}\left(\hat{\tau}_h\right) = \sum_{h=1}^{H} N_h^2 \cdot \left(\frac{N_h - n_h}{N_h}\right) \cdot \frac{\sigma_h^2}{n_h}, \tag{1.3}$$

with

$$\sigma_h^2 = \frac{1}{N_h - 1} \cdot \sum_{i=1}^{N_h} \left(y_{hi} - \overline{y}_h\right)^2.$$

This variance can be estimated by:

$$\hat{\mathrm{V}}\left(\hat{\tau}_{StrRS}\right) = \sum_{h=1}^{H} N_h^2 \cdot \left(\frac{N_h - n_h}{N_h}\right) \cdot \frac{s_h^2}{n_h}, \tag{1.4}$$

with

$$s_h^2 = \frac{1}{n_h - 1} \cdot \sum_{i=1}^{n_h} \left(y_{hi} - \overline{y}_h\right)^2$$

(cf. Lohr, 1999, p. 99 ff.).

When a two stage cluster sampling is used the variance of the point estimator is defined by

$$\mathrm{V}\left(\hat{\theta}\right) = \mathrm{V}_1\left[\mathrm{E}_2\left(\hat{\theta}\right)\right] + \mathrm{E}_1\left[\mathrm{V}_2\left(\hat{\theta}\right)\right] \tag{1.5}$$

where E_1 indicates the expected value over all possible units of the first stage and E_2 the expected value over all possible units of the second stage for a given set of units of the first stage. V_1 is the variance over the drawn PSUs and $V_2\left(\hat{\theta}\right)$ the variance over all possible Secondary Sampling Units (SSUs) for a given set of selected PSUs (cf. COCHRAN, 1977, p. 275 f.).

Hence, in case of two stage cluster sampling the variance of a point estimator consists of two components: the first component is the variance of the several units of the first stage and the second component is the variance of the SSUs within the several PSUs (cf. LOHR, 1999, p. 147).

The variance of a total value can be estimated by

$$\hat{V}\left(\hat{\tau}_{2St}\right) = N'^2 \cdot \left(\frac{N' - n'}{N'}\right) \cdot \frac{s_e^2}{n'} + \frac{N'}{n'} \cdot \sum_{i=1}^{n'} N_i''^2 \cdot \left(\frac{N_i'' - n_i''}{N_i''}\right) \cdot \frac{s_i^2}{n_i''} \tag{1.6}$$

with

$$s_e^2 = \frac{1}{n' - 1} \cdot \sum_{i=1}^{n'} \left(\hat{\tau}_i - \frac{\hat{\tau}}{N'}\right)^2 \tag{1.7}$$

and

$$s_i^2 = \frac{1}{n_i'' - 1} \cdot \sum_{q=1}^{n_i''} \left(y_{iq} - \overline{y}_i\right)^2$$

(cf. LOHR, 1999, p. 147).

N' indicates the number of PSUs in the universe at stage 1 and n' their number in the sample. N_i'' indicates the number of SSUs in the universe in a certain PSU i at stage 2 and n_i'' the number of drawn SSUs in the PSU i. This procedure can easily be extended to sampling designs with more than two stages.

The direct variance estimator of both, a two stage design with stratification at the first stage is given by:

$$\hat{V}\left(\hat{\tau}_{2St}\right) = \sum_{h=1}^{H} N_h'^2 \cdot \left(\frac{N_h' - n_h'}{N_h'}\right) \cdot \frac{s_e^2}{n_h'} + \sum_{h=1}^{H} \frac{N_h'}{n_h'} \cdot \sum_{i=1}^{n_h'} N_{hi}''^2 \cdot \left(\frac{N_{hi}'' - n_{hi}''}{N_{hi}''}\right) \cdot \frac{s_{hi}^2}{n_{hi}''}$$

with

$$s_e^2 = \frac{1}{n_h^{'} - 1} \cdot \sum_{i=1}^{n_h^{'}} \left(\hat{\tau}_{hi} - \frac{\hat{\tau}_h}{N_h^{'}} \right)^2$$

and

$$s_{hi}^2 = \frac{1}{n_{hi}^{''} - 1} \cdot \sum_{q=1}^{n_{hi}^{''}} \left(y_{hiq} - \overline{y}_{hi} \right)^2$$

(cf. for an example of a stratified three stage design PRESTON, 2009, p. 228).

Chapter 2

Variance Estimators and their Approximation

2.1 Variance of the Horvitz-Thompson Estimator

The variance of $\hat{\tau}$, as defined in (1.2), is calculated by

$$\mathrm{V}\left(\hat{\tau}\right) = \sum_{i\in\mathcal{U}} \pi_i(1-\pi_i)\left(\frac{y_i}{\pi_i}\right)^2 + 2\cdot\sum_{i=1}^{N}\sum_{j>i}^{N}(\pi_{ij}-\pi_i\cdot\pi_j)\cdot\frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}\,. \tag{2.1}$$

For the derivation of the above variance see Cochran (1977), p. 260. An unbiased estimator for the variance in (2.1) is given by

$$\hat{\mathrm{V}}\left(\hat{\tau}\right) = \sum_{i\in s}(1-\pi_i)\cdot\left(\frac{y_i}{\pi_i}\right)^2 + 2\cdot\sum_{i=1}^{n}\sum_{j>i}^{n}(1-\frac{\pi_i\cdot\pi_j}{\pi_{ij}})\cdot\frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}\,. \tag{2.2}$$

The justification of the variance estimator in (2.2) can be found in Särndal et al. (1992), p. 44. Although they use another notation for (2.2), $\hat{\mathrm{V}}(\hat{\tau}) = \sum\sum_s \frac{1}{\pi_{ij}}\cdot\left(\frac{\pi_{ij}}{\pi_i\cdot\pi_j}-1\right)y_iy_j$, with $\sum\sum_s a_{ij} = \sum_{i\in s}a_{ii} + \sum_{i\in s}\sum_{j\in s}a_{ij}$ and $i\neq j$, which is equivalent to (2.2). In case of fixed size designs Yates and Grundy (1953), p. 257, established an alternative to the variance estimator in (2.2). Thereafter the variance in (2.1) can also be written as

$$\begin{aligned}\mathrm{V}\left(\hat{\tau}\right) &= -\frac{1}{2}\sum_{i}^{N}\sum_{j\neq i}^{N}(\pi_{ij}-\pi_i\cdot\pi_j)\cdot\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \\ &= \sum_{i=1}^{N}\sum_{j<i}^{N}(\pi_i\cdot\pi_j-\pi_{ij})\cdot\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2\,,\end{aligned} \tag{2.3}$$

and its unbiased estimate is given by

$$\hat{\mathrm{V}}\left(\hat{\tau}\right) = \sum_{i=1}^{n}\sum_{j<i}^{n}\frac{\pi_i\cdot\pi_j-\pi_{ij}}{\pi_{ij}}\cdot\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2\,, \tag{2.4}$$

(cf. Cochran, 1977, p. 260f.). For (2.4) and (2.2) to be unbiased, a necessary and sufficient condition is that $\pi_{ij} > 0$, $\forall\ i,l \in \mathcal{U}$. However, for both variance estimates to be strictly non negative a sufficient condition is that $\pi_i \cdot \pi_j \geq \pi_{ij}$, $\forall\ i,j \in \mathcal{U}$, $i \neq j$, (cf. Tillé, 1996, p. 179).

With the two variances estimators (2.4) and (2.2) it is possible to handle all common sample designs. In practice, however, there are drawbacks with these estimators. First, the double sum which appears in both formulas involves a fast growing mass of terms which demands that these variance estimators be calculated recursively, (see, Deville, 1999, p. 194). Second, the second order inclusion probabilities π_{ij} are difficult to obtain for most samples designs applied in practice. Only in rare cases with simple schemes they are easily calulated, e.g. for a simple random sample (SRS) design without replacement (WoR) the π_{ij}s are given by

$$\pi_{ij} = \begin{cases} \dfrac{n}{N} & i = j \wedge i, j \leq N \\ \dfrac{n \cdot (n-1)}{N \cdot (N-1)} & \text{else.} \end{cases} . \tag{2.5}$$

The probabilities result from the hypergeometric distribution, i.e. the number of samples including the i-th element $\binom{N-1}{n-1}$ respectively the i-th and j-th element $\binom{N-2}{n-2}$ divided by the number of all possible samples $\binom{N}{n}$. Inserting the above inclusion probabilities into (2.4) yields

$$\begin{aligned} V(\hat{\tau}) &= N^2 \cdot \sum_{i}^{N} \sum_{j<i}^{N} \left(\frac{n^2}{N^2} - \frac{n \cdot (n-1)}{N \cdot (N-1)} \right) \cdot \left(\frac{y_i}{n/N} - \frac{y_j}{n/N} \right)^2 \\ &= N^2 \cdot \frac{N-n}{N \cdot n} \cdot \frac{1}{N-1} \cdot \frac{1}{N} \cdot \sum_{i}^{N} \sum_{j<i}^{N} (y_i - y_j)^2 \\ &= N^2 \cdot \frac{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \frac{1}{N} \cdot \sum_{j=1}^{N} y_j)^2}{n} \cdot \frac{N-n}{N-1} \\ &= N^2 \cdot \frac{\sigma_y^2}{n} \cdot \frac{N-n}{N-1} , \end{aligned} \tag{2.6}$$

where σ_y^2 denotes the variance of variable y, (cf. Münnich, 2005, p. 55). Substituting σ_y^2 by a suitable estimate $\hat{\sigma}_y^2$, in case of SRS by $\frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - \frac{1}{n} \cdot \sum_{j=1}^{n} y_j)^2$, leads to the variance estimate $\hat{V}(\hat{\tau})$.

2.2 Variance Estimation when Sampling With Replacement

To avoid the calculation of the double sum in (2.4) or (2.2) in the presence of more complex designs different approximations can be found in the literature that only use π_i but dispense with the π_{ij}, (cf. Berger and Skinner, 2004). One way to evade the computation of this large number of cross-products is to use variance estimation for

samples drawn with replacement. Thus, we treat the sample as it would have been drawn by more simple design than the true one. The estimation of τ from a sample drawn with replacement can be done by the Hansen-Hurwitz estimator, (see SÄRNDAL et al., 1992, p. 51f.).

$$\hat{\tau}_{HH} = \frac{1}{n}\sum_{i=1}^{n} Z_i \ , \tag{2.7}$$

where Z_i is a random variable such that

$$Z_i = \frac{y_k}{p_k}$$

that is, if the k-th element is selected in the i-th draw and p_k is the selection proability of the k-th element. Further, we have the following relationship between the inclusion probability π_k and p_k

$$\begin{aligned}\pi_k &= 1 - (1-p_k)^n \ , \\ &= np_k + \sum_{l=2}^{\infty} \binom{n}{k} (-p_k)^l \ ,\end{aligned}$$

and if p_k is small, which is reasonable if N is large, then $\pi_k \doteq np_k$. It follows that

$$\forall\, i = 1, \ldots, n; \qquad \Pr\left(Z_i = \frac{y_k}{p_k}\right) = p_k; \qquad k = 1, \ldots, N \, .$$

The estimator in (2.7) is unbiased for the population total τ, because $\mathrm{E}(Z_i) = \sum_{k\in\mathcal{U}} \frac{y_k}{p_k} \cdot p_k = \tau$ and its variance is

$$\mathrm{V}\left(\hat{\tau}_{HH}\right) = \sum_{k\in U}\left(\frac{y_k}{p_k} - \tau\right)^2 \cdot p_k \cdot \frac{1}{n} \, . \tag{2.8}$$

An unbiased estimator of the variance in (2.8) is given by

$$\hat{\mathrm{V}}\left(\hat{\tau}_{HH}\right) = \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\left(Z_i - \hat{\tau}_{HH}\right)^2 \ , \tag{2.9}$$

the proof of (2.8) and (2.9) can be found in SÄRNDAL et al. (1992), p. 52. The variance estimate in (2.9) can now be used to obtain a simplified variance estimation for (1.2), (cf. SÄRNDAL et al., 1992, p. 422)

$$\hat{\mathrm{V}}_0\left(\hat{\tau}\right) = \frac{1}{n\cdot(n-1)} \cdot \sum_{k\in s}\left(\frac{y_k}{p_k} - \hat{\tau}\right)^2 \ , \tag{2.10}$$

where p_k is taken as $p_k = \frac{\pi_k}{n}$ for $k \in s$. By using (2.10) instead of (2.4) or (2.2) we calculate the variance estimator as if the sample was drawn with replacement, whereas it was actually drawn without replacement. But this computational simplification that results from getting rid of the π_{ij} has the drawback that $\hat{\mathrm{V}}_0$ will not be without bias for

$V(\tau)$, (see Särndal et al., 1992, p. 422). However, this bias which is for any fixed sample size design given by

$$E\left(\hat{V}_0(\hat{\tau})\right) - V(\hat{\tau}) = \frac{n}{n-1} \cdot (V_0(\hat{\tau}) - V(\hat{\tau})) \ , \tag{2.11}$$

with $V(\hat{\tau})$ given in (2.1) or (2.3), will be positive in cases where sampling without replacement results in a lower sampling variance than sampling with replacement, as it is the case for SRS. Thus, the use of (2.10) will lead to the construction of over conservative confidence intervals, whereas for unequal proability sampling this need not to be the case, (cf. Gabler, 1990, p. 64f.). A sufficient conditon for $V_0(\hat{\tau}) - V(\hat{\tau})$ to be nonnegative is

$$\pi_{ij} \geq \pi_i \cdot \pi_j \cdot (1 - \frac{1}{n}) \ , \qquad \forall \quad i,j \in \mathcal{U}, \, i \neq j \tag{2.12}$$

which can be seen from rewriting (2.8) in following way

$$\mathrm{V}\left(\hat{\tau}_{HH}\right) = \sum_{k=1}^{N} \frac{y_k^2}{n p_k} - \frac{\tau^2}{n} = \frac{1}{n} \sum_{k=1}^{N} \sum_{l<k}^{N} \left(\frac{y_k}{p_k} - \frac{y_l}{p_l}\right)^2 p_k \cdot p_l \ , \tag{2.13}$$

having in mind that $p_k = \frac{\pi_k}{n}$ it follows that (2.13) is always greater than (2.3) if condition (2.12) is met, e.g. in cases of SRS WoR.

Although condition (2.12) is sufficient, it is not a necessary one. A weaker condition, yet still not a necessary one, is given be Gabler (1990), p. 69f., which states that if

$$\sum_{k=1}^{N} \min_{1 \leq l \leq N} \frac{\pi_{kl}}{\pi_l} \leq n - 1 \ ,$$

then $V_0(\hat{\tau}) \geq V(\hat{\tau})$, (see also Gabler, 1984, for another sufficient condition).

Thus, we conclude that in some cases (2.9) might serve as an easy to compute estimate for an upper bound for the sampling variance when sampling without replacement. This simplification is, however, limited by the fact that we need to know the π_{ij}'s to be sure that (2.8) is greater than (2.1) or (2.3).

2.3 Variance Approximations

To begin of this section we start with an alternative formulation for the variance of the Horvitz-Thompson estimator for fixed sample sizes (2.3), (cf. Brewer, 2003, p. 149). This approch will lead to a first approximation of (2.3) and show to what extent approximations of the second order inclusion probabilities are needed. First we begin with some general properties of the inclusion probabilities, which should be respected by proposed variance approximations, (cf. Tillé, 1996, p. 184):

$$\sum_{\substack{i \in \mathcal{U} \\ i \neq j}} \pi_{ij} = \pi_j \cdot (n-1) \tag{2.14a}$$

$$\sum_{j\in\mathcal{U}}\sum_{\substack{i\in\mathcal{U}\\ i\neq j}}\pi_{ij} = n\cdot(n-1) \tag{2.14b}$$

$$\sum_{j\in\mathcal{U}}\sum_{\substack{i\in\mathcal{U}\\ i\neq j}}\pi_j\pi_i = n^2 - \sum_{j\in\mathcal{U}}\pi_j^2 \tag{2.14c}$$

$$\sum_{\substack{i\in\mathcal{U}\\ i\neq j}}(\pi_i\pi_j - \pi_{ij}) = \pi_j(1-\pi_j) \tag{2.14d}$$

Now we may rewrite (2.3) in the following way,

$$\begin{aligned}\mathrm{V}(\hat{\tau}) =& \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}(\pi_i\cdot\pi_j - \pi_{ij})\cdot\left(\frac{y_i}{\pi_i} - \frac{\tau}{n}\right)^2 - \left(\frac{y_j}{\pi_j} - \frac{\tau}{n}\right)^2 \\ =& \sum_{i=1}^{N}\sum_{j\neq i}^{N}(\pi_i\cdot\pi_j - \pi_{ij})\cdot\left(\frac{y_i}{\pi_i} - \frac{\tau}{n}\right)^2 \\ & - \sum_{i=1}^{N}\sum_{j\neq i}^{N}(\pi_i\cdot\pi_j - \pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{\tau}{n}\right)\cdot\left(\frac{y_j}{\pi_j} - \frac{\tau}{n}\right),\end{aligned} \tag{2.15}$$

and by relation (2.14d)

$$\begin{aligned}\mathrm{V}(\hat{\tau}) =& \sum_{i=1}^{N}\pi_i(1-\pi_i)\cdot\left(\frac{y_i}{\pi_i} - \frac{\tau}{n}\right)^2 \\ & - \sum_{i=1}^{N}\sum_{j\neq i}^{N}(\pi_i\cdot\pi_j - \pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{\tau}{n}\right)\cdot\left(\frac{y_j}{\pi_j} - \frac{\tau}{n}\right).\end{aligned} \tag{2.16}$$

The first term in (2.16) equals (2.9) for $p_k = \pi_k/n$ with a finite population correction term $(1-\pi_i)$ and constitutes the major part of the variance of $\hat{\tau}$ (cf., BREWER, 2003, p. 149f.). The importance of the second term in (2.16) depends on the sampling design $p(.)$. For the second term to be negligiable compared to the first term, $p(.)$ must be of the form that $\pi_{ij} \approx \pi_i\cdot\pi_j$. For example this is justifiable in case of SRS, (see (2.5)). However, if the second term in (2.16) is expected to become important one can look for more suitable approximations of π_{ij}, which lead to proxy variances of $\hat{\tau}$ having the simple form of the first term in (2.16). This will be the main concern for rest of the section. To do so we start by considering Poisson sampling.

Poisson sampling is a design for which the variance of (1.2) can be derived as easily as for the Hansen-Hurwitz estimator, where only the π_i's have to be known. Because in Poisson sample all elements of the random vector $\mathfrak{s}$ in (1.1) are independent random variables. Poisson sampling is defined by

$$p_{\text{poiss}}(\tilde{s}) = \prod_{i\in\tilde{s}} p_i \prod_{\mathcal{U}\setminus\tilde{s}} 1 - p_i\ ,$$

with numbers $0 \leq p_i \leq 1\,\forall\, i \in \mathcal{U}$ (see HÁJEK, 1981, p. 54). Poisson sampling has a random sample size $\tilde{n}$, with $E(\tilde{n}) = \sum_{i=1}^{N} p_i$ and $\mathrm{V}(\tilde{n}) = \sum_{i=1}^{N} p_i(1-p_i)$. Further we have

first-order inclusion probabilities, $\tilde{\pi}_i = p_i \ \forall i \in \mathcal{U}$ and second-order inclusion probabilities $\tilde{\pi}_{ij} = p_i p_j \ i, j \in \mathcal{U}, i \neq j$. Thus, if $\tilde{s}$ is a sample drawn after $p_{\text{poiss}}(.)$, then

$$\mathrm{V}_{\text{poiss}}(\hat{\tau}) = \mathrm{V}_{\text{poiss}} \left(\sum_{i \in \tilde{s}} \frac{y_i}{\pi_i} \right) = \sum_{i \in \mathcal{U}} \frac{y_i^2}{\pi_i^2} \cdot \tilde{\pi}_i (1 - \tilde{\pi}_i) \ , \tag{2.17}$$

where $\mathrm{V}_{\text{poiss}}$ denotes the variance under the design $p_{\text{poiss}}(.)$. Note that (2.17) contains also π_i because it is the variance of estimator (1.2) computed under Poisson sampling, (cf. Deville and Tillé, 2005, p. 573). To have a fixed sample size we make $p_{\text{poiss}}(.)$ a conditional Poisson sampling design given that its sample size $\tilde{n}$ is fixed to n. Heuristically we can describe such a fix size Poisson sampling design as

$$p_{\text{poiss}}(\tilde{s} | \tilde{n} = n) = \frac{p_{\text{poiss}}(\tilde{s})}{P(\tilde{s} \in \mathcal{S}_n)} \ ,$$

where $p_{\text{poiss}}(\tilde{s} | \tilde{n} = n) = 0$ if $\tilde{s} \notin \mathcal{S}_n$ and $P(\tilde{s} \in \mathcal{S}_n)$ denotes the probability under design p_{poiss} that the size of $\tilde{s}$ is n (see Berger, 2004b, p. 454). The problem connected with the conditioning on $P(\tilde{s} \in \mathcal{S}_n)$ is to reevaluate the $\tilde{\pi}_i$'s. They can be computed exactly (see Hájek, 1981, chapter. 14), but in Hájek (1964, p. 1508ff) it is shown that we can approximate $\tilde{\pi}_i$ by π_i.

We may then write for the variance of $\hat{\tau}$

$$\mathrm{V}(\hat{\tau}) = \mathrm{V}_{\text{poiss}} \left(\hat{\tau} | \tilde{n} = n \right) \ ,$$

(cf. Matei and Tillé, 2005, p. 548). If we assume that through Poisson sampling the couple $\{\hat{\tau}, \tilde{n}\}$ is bivariate normal distributed (cf. Deville and Tillé, 2005, p. 573) then, by exploiting the linear relation between $\hat{\tau}$ and $\tilde{n}$, it is possible to write

$$\mathrm{V}_{\text{poiss}} \left(\hat{\tau} | \tilde{n} = n \right) = \mathrm{V}_{\text{poiss}} (\hat{\tau} + (n - \tilde{n}) \cdot \beta) \ , \tag{2.18}$$

where

$$\beta = \frac{\mathrm{Cov}_{\text{poiss}}(\tilde{n}, \hat{\tau})}{\mathrm{V}_{\text{poiss}}(\tilde{n})} \ ,$$

and

$$\mathrm{Cov}_{\text{poiss}}(\tilde{n}, \hat{\tau}) = \sum_{i \in \mathcal{U}} \tilde{\pi}_i (1 - \tilde{\pi}_i) \cdot \frac{y_i}{\pi_i} \ ,$$
$$\mathrm{V}_{\text{poiss}}(\tilde{n}) = \sum_{i \in \mathcal{U}} \tilde{\pi}_i (1 - \tilde{\pi}_i) \ .$$

If we define $b_i = \tilde{\pi}_i (1 - \tilde{\pi}_i)$ we get the following variance approximation formula

$$\mathrm{V}_{\text{approx}} (\hat{\tau}) = \sum_{i \in \mathcal{U}} b_i \cdot \epsilon_i^2 \ , \tag{2.19}$$

where

$$\epsilon_i = \frac{y_i}{\pi_i} - \beta.$$

Approximation (2.19) is valid for the class of maximum entropy designs, which are designs that maximise the quantity

$$-\sum_{s\in\mathcal{S}} p(s)\log p(s)\ ,$$

under condition $\pi_i = \sum_{s\ni i} p(s)$, which is met by conditional Poisson samplign $\tilde{p}(.)$ (cf. MATEI and TILLÉ, 2005).

Depending on the choice of b_i there exist numerous variants of approximation (2.19). An analysis of the literature on different values of b_i can be found in MATEI and TILLÉ (2005).

2.3.1 Hajek Approximation

HÁJEK (1964), p. 1509, proposed the following relation between the probabilities π_i and π_{ij}

$$\pi_i\pi_j - \pi_{ij} = d^{-1}\cdot\pi_i\,(1-\pi_i)\cdot\pi_j\cdot(1-\pi_j)\,[1+o(1)]\ , \tag{2.20}$$

where $d = \sum_{i\in\mathcal{U}}\pi_i\cdot(1-\pi_i)$ and $o(1)\to 0$ if $d\to\infty$. The relation above is vaild for rejective sampling, which he defines as a conditional Poisson sampling or conditional sampling with replacement, (cf. HÁJEK, 1981, p. 66f.). Thus, we can write:

$$\pi_{ij}\approx\pi_i\pi_j\cdot\left(1-(1-\pi_i)\cdot(1-\pi_j)\cdot d^{-1}\right)\qquad 1\le i\ne j\le N\ . \tag{2.21}$$

Inserting (2.21) into (2.3) gives the following approximation for $\mathrm{V}\,(\hat{\tau})$

$$\mathrm{V}_{\mathrm{Haj}}(\hat{\tau}) = \sum_{i\in\mathcal{U}}\pi_i(1-\pi_i)\cdot\left(\frac{y_i}{\pi_i}-B\right)^2\ , \tag{2.22}$$

where

$$B = \frac{\sum_{i\in\mathcal{U}}\frac{y_i}{\pi_i}\cdot\pi_i(1-\pi_i)}{\sum_{i\in\mathcal{U}}\pi_i(1-\pi_i)}\ .$$

Now the right-hand side of formula (2.20) may be approximated by the prodcut c_ic_j where $c_i = \pi_i(1-\pi_i)\cdot\sqrt{d}^{-1}$. But because this does not satisfy the relation (2.14d) let c_i be set to $c_i = \pi_i(1-\lambda_i)\cdot\sqrt{d_\lambda}^{-1}$ and $d_\lambda = \sum_{j=1}^{N}\pi_j(1-\lambda_j)$. HÁJEK (1981), p. 27, gives an approximation of the $\pi_i(1-\lambda_i)$ terms, which can be described, in line with the general approximation in (2.19), in the following way

$$_{\mathrm{HAJEK}}b_i^* = \pi_i(1-\lambda_i) = \pi_i(1-\pi_i)\cdot\left[1-\frac{\pi_i(1-\lambda_i)}{d_\lambda}\right]^{-1}\ . \tag{2.23}$$

A value for $\pi_i(1-\lambda_i)$ is then obtained by iteration. For the first iteration inserting $\pi_i(1-\lambda_i) = \pi_i(1-\pi_i)$ on the right hand side. The obtained value is then again insterted on the right hand side of (2.23), this process is repeated until the results become stable as $\pi_i(1-\lambda_i)$ converges to $\pi_i(1-\pi_i)$ (cf. HÁJEK, 1981, p. 76). As a compromise between

simplicity and precision Hájek (1981) proposed for small π_i's the following approximation for b_i

$$_{\text{HAJEK}}b_i = \pi_i(1 - \frac{n-1}{n}\pi_i) \; . \tag{2.24}$$

For $\pi_i = \frac{n}{N} \; \forall i \in \mathcal{U}$ (2.22) becomes the standard variance estimator (2.6), (cf. Berger, 2003, p. 9). Note that if $\frac{\tau}{n}$ is substituted by B in (2.22) equals the first term in (2.16).

Furthermore, Berger (1998a), showed that (2.21) can also be used for a class of highly randomized or high entropy sampling designs (cf. Berger, 2004a, p. 307), which includes amongst others the RaoSampford sampling design (see, Sampford, 1967).

Fixed-Point Approximation

Rewrite the general approximation in (2.19) as

$$\mathrm{V}_{\text{approx}}\left(\hat{\tau}\right) = \sum_{i\in\mathcal{U}} \frac{y_i^2}{\pi_i^2}\left(b_i - \frac{b_i^2}{\sum_{j\in\mathcal{U}} b_j}\right) - \frac{1}{\sum_{j\in\mathcal{U}} b_j}\sum_{i\in\mathcal{U}}\sum_{j\neq i}\frac{y_i y_j}{\pi_i\pi_j}\cdot b_i b_j \; , \tag{2.25}$$

Deville and Tillé (2005) proposed, by comparing (2.25) with (2.1), to obtain a most accurate approximation of the variance of $\hat{\tau}$ by solving the following equation system to find another approximation of b_i,

$$b_i - \frac{b_i^2}{\sum_{j\in\mathcal{U}} b_j} = \pi_i(1-\pi_i) \; . \tag{2.26}$$

Because (2.26) is a non linear equation system the b_i can be approximated by iteration, (see also Tillé, 2006, p. 139f.). To solve, (2.26) Deville and Tillé (2005) suggested a fixed-point algorithm by using the recurrence equation until convergence

$$b_i^{(k)} = \frac{\left[b_i^{(k-1)}\right]^2}{\sum_{i\in\mathcal{U}} b_i^{(k-1)}} + \pi_i\left(1-\pi_i\right) \qquad \text{for} \quad k = 0, 1, 2, 3, \ldots \tag{2.27}$$

and as starting value $b_i^{(0)}$

$$b_i^{(0)} = \pi_i(1-\pi_i)\cdot\frac{N}{N-1}.$$

Though there the following necessary conditon has to hold in order to find a unique solution to (2.27),

$$\max_{1\leq i\leq N}\frac{\pi_i(1-\pi_i)}{\sum_{j\in\mathcal{U}}\pi_j(1-\pi_j)} < \frac{1}{2} \; ,$$

(cf. Deville and Tillé, 2005, p. 575). If the process does not converge Tillé (2006), suggestes to use the values $b_i^{(1)}$ obtained after one iteration, where

$$b_i^{(1)} = \pi_i(1-\pi_i)\cdot\left[\frac{N\cdot\pi_i(1-\pi_i)}{(N-1)\cdot\sum_{j\in\mathcal{U}}\pi_j(1-\pi_j)} + 1\right] \; .$$

Note that the proceeding described above is quite simular to Hájek (1981) iterative approach to solving (2.23).

Adjustment on Marginal Totals

In order to find an approximation which respects the constraint (2.14a), TILLÉ (1996) applies the Iterative Proportional Fitting Procedure to the matrix $\mathbf{A} = [a_{ij}]$, where

$$a_{ij} = \begin{cases} \pi_i \pi_j & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} .$$

Matrix $\mathbf{A}$ is adjusted on the marginals totals $\tau_b = (\tau_{b1}, \ldots, \tau_{bi}, \ldots, \tau_{bN})$ and $\tau_b^{'}$ where $\tau_{bi} = \pi_i \cdot (n-1)$. This can be achieved be applying the following algorithm:

$$\begin{aligned} \eta_i^{(0)} &= \pi_i, \quad \forall i \in \mathcal{U}, \\ \eta_i^{(2k-1)} &= \frac{(n-1)\pi_i}{\eta^{(2k-2)} - \eta_i^{(2k-2)}} \quad \text{and} \\ \eta_i^{(2k)} &= \eta_i^{2k-1} \left[\frac{n(n-1)}{\left(\eta^{2k-1}\right)^2 - \sum_{i \in \mathcal{U}} \left(\eta_i^{2k-1}\right)^2} \right]^{1/2} \end{aligned}$$

where

$$\eta^{(k)} = \sum_{i \in \mathcal{U}} \eta_i^k, \quad \text{for} \quad k = 0, 1, 2, 3, \ldots \quad ,$$

(cf., TILLÉ, 1996, p. 185). Then the coefficients η_i are used to approximate the second order inclusion probabilities, i.e. $\pi_{ij} \approx \eta_i \eta_j$.

2.3.2 Brewer Approximation

There exists another class of approximate expressions for the variance of $\hat{\tau}$ proposed by BREWER (2003) and BREWER and DONADIO (2003). They use an approximation for the π_{ij} derived by HARTLEY and RAO (1962) for a randomized systematic sampling scheme with unequal inclusion probabilities of the following form

$$\pi_{ij} \approx \frac{1}{2} \pi_i \pi_j \left(c_i^* + c_j^* \right) \ , \tag{2.28}$$

where

$$c_i^* = \frac{(n-1)}{n} \left(1 - n^{-2} \sum_{i \in U} \pi_i^2 + 2 \frac{\pi_i}{n} \right) .$$

An approximation for the variance in (2.3) is obtained by developing the second term in (2.16) with (2.28), which yields

$$\sum_{i=1}^{N} \pi_i^2 (1 - c_i^*) \cdot \left(\frac{y_i}{\pi_i} - \frac{\tau}{n} \right)^2 . \tag{2.29}$$

Adding to the term above the first term in (2.16) gives the following approximation

$$\mathrm{V}_{\mathrm{Brew}}(\hat{\tau}) = \sum_{i}^{N} \pi_i (1 - c_i^* \pi_i) \cdot \left(\frac{y_i}{\pi_i} - \frac{\tau}{n} \right)^2 . \tag{2.30}$$

2.3.3 Estimators for Variance Approximations

Whereas the previews section gave an obverview of possible approximations to variance (2.3), this section introduces some estimators for these approximations. Following the typology of Matei and Tillé (2005) we will distinguish between two types of estimators for variance approximations, estimators of *type 1* that require knowledge of π_i for all $i \in s$ and estimators of *type 2* requiring knowledge of π_i for all $i \in \mathcal{U}$.

Type 1 Variances Estimates

From the general variance approximation (2.22) a plausible estimator can be written as

$$\hat{\mathrm{V}}_{\text{approx}}(\hat{\tau}) = \sum_{i \in s} \frac{\hat{b}_i}{\pi_i^2} \cdot e_i^2 \,, \tag{2.31}$$

where

$$e_i = y_i - \hat{B} \tag{2.32}$$

$$\text{and} \quad \hat{B} = \frac{\sum_{j \in s} \frac{y_j}{\pi_j} \cdot \hat{b}_j}{\sum_{j \in s} \hat{b}_j} \cdot \pi_i \,. \tag{2.33}$$

Depending on the the choice of $\hat{b}_i$ numerous estimates have been proposed in the literature, (see, Matei and Tillé, 2005, Cap. 4). A simple value for $\hat{b}_i$ could be

$${}_1\hat{b}_i = (1 - \pi_i)\frac{n}{n-1} \tag{2.34}$$

which lead for SRS to the standard variance estimator of (2.6). A more complex choice of $\hat{b}_i$ proposed by Deville, 1999 is given by

$${}_2\hat{b}_i = (1 - \pi_i)\left[1 - \sum_{j \in s} \frac{1 - \pi_k}{\sum_{j \in s} 1 - \pi_j}\right] \,. \tag{2.35}$$

Also Fixed-Point Approximation can be used to obtain a value for $\hat{b}_i$. The same algorithm is applied as described in section (2.3.1) except that the right hand side of equation (2.26) and second term on right hand side of (2.27) are multiplied by π_i^{-1}. The initial value for the algoritm would be

$${}_3\hat{b}_i^{(0)} = (1 - \pi_i) \cdot \frac{n}{n-1} \,,$$

and a necessary conditon that a solution exists is

$$\max_{1 \le i \le n} \frac{(1 - \pi_i)}{\sum_{j \in s}(1 - \pi_j)} < \frac{1}{2} \,,$$

(cf. Tillé, 2006, p. 141f). If the process does not converge Tillé (2006) suggests to use the values ${}_3\hat{b}_i^{(1)}$ obtained after one iteration:

$${}_3\hat{b}_i^{(1)} = (1 - \pi_i) \cdot \left[\frac{n \cdot (1 - \pi_i)}{(n-1) \cdot \sum_{j \in s}(1 - \pi_j)} + 1\right] \,.$$

Rosén suggested an estimator (cf. ROSÉN, 1997, p. 167f.) with $\hat{b}_i = {}_1\hat{b}_i$, but substituting $\hat{B}$ in (2.32) by

$$ {}_r\hat{B} = \frac{\sum_{j\in s} \frac{y_j}{\pi_j} \cdot \frac{1-\pi_j}{\pi_j} \ln(1-\pi_j)}{\sum_{j\in s} \frac{1-\pi_j}{\pi_j} \ln(1-\pi_j)} \cdot \pi_i \ . $$

In accordance with variance approximation (2.30) BREWER (2003) and BREWER and DONADIO (2003) constructed a variance estimator which they claim to be design-unbiased under SRS. This variance estimator is given by

$$ \hat{\mathrm{V}}_{\mathrm{Brew}}(\hat{\tau}) = \sum_i^n \frac{\left[\frac{1}{c_i^*} - 1\right]}{\pi_i^2} \cdot \left(y_i - \frac{\hat{\tau}}{n} \cdot \pi_i\right)^2 \ , \tag{2.36} $$

where (2.36) is the corresponding Horvitz-Thompson estimator to the sum in (2.30) corrected by a factor c_i^{*-1}, (cf. BREWER and DONADIO, 2003, p. 6). BREWER (2003) proposed a selection of choices for $\hat{c}_i^*$. One of them belongs to the category of Type 1 estimators:

$$ {}_1\hat{c}_i^* = \frac{n-1}{n-\pi_i} \ . $$

The implicit motivation for the choice of ${}_1\hat{c}_i^*$ is given by the relation (2.14d), (see BREWER and DONADIO, 2003, p. 5).

Type 2 Variances Estimates

From the work of BERGER (1998b) the following value of $\hat{b}_i$ can be suggested, (see MATEI and TILLÉ, 2005, p. 554):

$$ {}_4\hat{b}_i = (1-\pi_i)\frac{n}{n-1}\left[\frac{\sum_{j\in s} 1-\pi_k}{\sum_{j\in\mathcal{U}} \pi_k(1-\pi_j)}\right] \ . \tag{2.37} $$

Till's approach (cf. TILLÉ, 1996, p. 184f.) of approximating the π_{ij}'s as described in section (2.3.1) can be used to derive an estimate of the form

$$ \begin{aligned} \hat{\mathrm{V}}_{\mathrm{Tille}}(\hat{\tau}) = &\left[\sum_{j\in s} {}_5\hat{b}_i\right] \sum_{i\in s} \frac{{}_5\hat{b}_i}{\pi_i^2} \left(y_i - \frac{\sum_{j\in s} \frac{y_j}{\pi_j} \cdot {}_5\hat{b}_j}{\sum_{j\in s} {}_5\hat{b}_j} \cdot \pi_i\right)^2 \\ &- \frac{n}{\pi_i^2} \sum_{i\in s} \left(y_i - \frac{\hat{\tau}}{n} \cdot \pi_i\right)^2 \ , \end{aligned} \tag{2.38} $$

where

$$ {}_5\hat{b}_i = \frac{\pi_i}{\eta_i} \ , $$

with η_i obtained by the alogrithm presented in section (2.3.1). Note that the second term in (2.38) corresponds to the estimator in (2.31) with $\hat{b}_i = 1 \ \forall \ i \in \mathcal{U}$.

Finally, there are three variants of estimate (2.36) which belong to category Type 2 estimators, (see Brewer, 2003, p. 152f and 156f.):

$$
\begin{aligned}
{}_2\hat{c}^* &= \frac{n-1}{n - \frac{\sum_{j\in\mathcal{U}} \pi_j^2}{n}} \,, \\
{}_3\hat{c}^* &= \frac{\frac{n-1}{n}}{1 - 2\frac{\pi_i}{n} + \frac{\sum_{j\in\mathcal{U}} \pi_j^2}{n^2}} \,, \\
{}_4\hat{c}^* &= \frac{\frac{n-1}{n}}{1 - \frac{(2n-1)\pi_i}{n(n-1)} + \frac{\sum_{j\in\mathcal{U}} \pi_j^2}{n(n-1)}} \,.
\end{aligned}
$$

The value ${}_2\hat{c}^*$ is deduced by comparing relations (2.14b) and (2.14c). The more complex estimator using ${}_3\hat{c}^*$ and ${}_4\hat{c}^*$ stem form c_i^* used in approximation (2.28), where ${}_4\hat{c}^*$ represents a supposed improvement over ${}_3\hat{c}^*$ in terms of unbiasedness for $\pi_i = \frac{n}{N}\ \forall\ i \in \mathcal{U}$, that is under SRS.

Chapter 3

Resampling Methods

The resampling method of variance estimation requires to select two or more (sub-)samples from a given population, or possibly a sample, and computing a separate estimate of the population parameter of interest from each sample. Variance estimation is done from the combination of all samples. The resampling methods differ in the way they generate the subsamples and with that in the treatment of complex survey designs (cf. MÜNNICH, 2005, p. 69).

3.1 Random Groups

The basic idea behind this method is to divide the sample into R groups which are then randomly selected, usually using the same sampling design, (cf. WOLTER, 2007, p. 21). An estimation of the population parameter of interest is carried out for each of the subsamples and the whole sample as well. The variance estimation is based on these repeated estimates. In the context of random groups we can differentiate between two cases. In the first case, the subsamples are mutually independent, while in the second case there is a dependency between random groups, (cf. WOLTER, 2007, p. 21). A random group variance estimator in case of independency can be derived as follows. Let τ be the population parameter of interest and $\hat{\tau}_r$ its corresponding estimate on basis of the r-th subsample. Then we can define the point estimator for τ using the method of random groups by

$$\hat{\tau}_{RG} = \frac{1}{R} \cdot \sum_{r=1}^{R} \hat{\tau}_r. \tag{3.1}$$

An unbiased variance estimator for $\mathrm{V}(\hat{\tau}_{RG})$ is stated by

$$\hat{\mathrm{V}}(\hat{\tau}_{RG}) = \frac{1}{R} \cdot \frac{1}{R-1} \cdot \sum_{r=1}^{R} (\hat{\tau}_r - \hat{\tau}_{RG})^2 . \tag{3.2}$$

Instead of computing $\hat{\tau}$ on the basis of the whole sample we take $\hat{\tau}_{RG}$ as a point estimator and using $\hat{V}(\hat{\tau}_{RG})$ as its corresponding variance estimator (cf. Wolter, 2007, p. 21 ff., Münnich, 2005, p. 70 f.). In case of nonindependent subsamples the random group estimator (3.1) produces biased estimates. An investigation of the magnitude and sign of the bias is given in Wolter (2007, Section 2.4). In surveys the method of random groups works well with respect to large samples of Primary Sampling Units (PSUs) (cf. Rust, 1985, p. 384). But as mentioned in Münnich (2005, p. 71) the method plays a minor part in practice due to their lack of efficiency and the advancements which are described in the following sections.

3.2 Balanced Repeated Replication

In presence of a stratified simple random sample (StrSRS) with two PSUs per stratum the method of random groups leads to an unstable variance estimation in practice. For such case balanced repeated replication (BRR) was developed (cf. Lohr, 1999, p. 298 f., Särndal et al., 1992, p. 430). This method allows a faster convergence than the method of random groups. The procedure goes back to McCarthy (1966, 1969) and was originally called balanced half samples (cf. Münnich, 2005, p. 71).

In the basic model each of the H strata consists of two PSUs. Only one element per stratum is drawn, so we take a half sample. As a result 2^H half samples are possible and with a large number of strata this could lead to a huge computational effort. To reduce the complexity we try to select a *balanced* set of half samples R with $R \ll 2^H$. Now let this set of R half samples by defined through a $R \times H$ matrix **RH** with the (r,h)-th element $\delta_{rh} = +1$ or -1, indicating whether the PSU from the h-th stratum selected for the r-th half sample is the first or the second PSU. Then the set of R half-samples is said to be *balanced* if

$$\sum_{r=1}^{R} \delta_{rh}\delta_{rh'} = 0 \qquad \forall\, h \neq h' \,, \tag{3.3}$$

(cf. Shao et al., 1998, p. 822, Wolter, 2007, p. 107ff., Münnich, 2008, p. 325f.). A balanced matrix **RH** can be obtained by using a $R \times R$ Hadamard matrix. The rows of a Hadamard matrix denote the half samples and the columns the strata, where $H+1 \leq R \leq H+4$, (cf. Münnich, 2008, p. 325 f., Shao et al., 1998, p. 822). Because its rows and columns are mutually orthogonal it fulfills condition (3.3) hence the half samples are drawn mutually independent. An example for an Hadamard matrix of dimension 8×8:

$$\mathbf{RH} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \tag{3.4}$$

(cf. DAVISON and SARDY, 2004, p. 18).

The weights of the chosen elements have to be adjusted and multiplied by 2 (cf. SHAO et al., 1998, p. 822). As before by applying the method of random groups the population parameter of interest $\hat{\tau}_r$ is computed on the base of each replication, here, by using balanced repeated replication on each half sample. Then a standard BRR variance estimator is given by

$$\hat{\mathrm{V}}_{BRR}(\hat{\tau}) = \frac{1}{R} \cdot \sum_{r=1}^{R} (\hat{\tau}_r - \hat{\tau}_{StrRS})^2 , \tag{3.5}$$

where $\hat{\tau}_{StrRS}$ is the estimated total in case of stratified random sampling (StrRS) (cf. MÜNNICH, 2005, p. 74).

The basic model, as described before, doesn't reflect a realistic scenario. In practice the elements per stratum are larger than two ($n_h > 2$) and associated with a smaller number of strata. Therefore the basic model has to be extended. One possibility is to divide the PSUs in a stratum h randomly into two groups of sizes $n_{h,1} = \lfloor n_h/2 \rfloor$ and $n_{h,2} = n_h - n_{h,1}$. For a possibility to handle the case when n_h is odd see KOVACEVIC and YUNG (1997, p. 45). A set of R replicates continues to be balanced, if in any two strata, pairs of groups have the same frequency of appearing in the R-replication (cf. WOLTER, 2007, p. 128 ff., DAVISON and SARDY, 2004, p. 18 f., RAO and SHAO, 1996, p. 343 f., MÜNNICH, 2008, p. 326, MÜNNICH, 2005, p. 77, SHAO et al., 1998, p. 824).

The weights of the r-th replicate have to be adjusted:

$$w_{h,i}^{(r)} := \begin{cases} w_{hi} \cdot \left[1 + \left\{ \dfrac{n_{h,2} \cdot (1 - f_h)}{n_{h,1}} \right\}^{1/2} \right], & \delta_{rh} = 1, \\ w_{hi} \cdot \left[1 - \left\{ \dfrac{n_{h,1} \cdot (1 - f_h)}{n_{h,2}} \right\}^{1/2} \right], & \delta_{rh} = -1, \end{cases} \tag{3.6}$$

where $(1 - f_h)$ is a correction when the sampling design is without replacement (cf. DAVISON and SARDY, 2004, p. 18, MÜNNICH, 2005, p. 77).

The variance estimator of the grouped balanced repeated replication (GBRR) is also adequate to the basic model and unbiased:

$$\hat{\mathrm{V}}_{GBRR}(\hat{\tau}) = \frac{1}{R} \cdot \sum_{r=1}^{R} (\hat{\tau}_r - \hat{\tau})^2 , \tag{3.7}$$

(cf. MÜNNICH, 2005, p. 77).

The BRR variance estimator can lead to instable variance estimates for two reasons (cf. e.g. Davison and Sardy, 2004, p. 19, Davison and Sardy, 2007, p. 375, Valliant, 1987, p. 80). The first reason results from (3.5) whose distribution is $c\chi^2_{R-1}$. Thus, the coefficient of variation is roughly $(2/R)^{1/2}$, where c is a constant and > 0. The coefficient of variation increases in case of a small R or equivalently a small number of strata H. The second reason is the high sensitivity of some statistics when the weights are perturbed. Then the calculation of all the replicate estimates $\hat{\tau}_1, \ldots, \hat{\tau}_r$ can be impossible (cf. Davison and Sardy, 2004, p. 19).

For that reason Fay (1989) suggests a scheme which makes the weighting milder by choosing a factor ε, where $0 < \varepsilon \leq 1$. The weights result from:

$$w_{h,i}^{(r)} := \begin{cases} w_{hi} \cdot \left[1 + \varepsilon \cdot \left\{\frac{n_{h,2} \cdot (1 - f_h)}{n_{h,1}}\right\}^{1/2}\right], & \delta_{rh} = 1, \\ w_{hi} \cdot \left[1 - \varepsilon \cdot \left\{\frac{n_{h,1} \cdot (1 - f_h)}{n_{h,2}}\right\}^{1/2}\right], & \delta_{rh} = -1. \end{cases} \tag{3.8}$$

The resulting variance estimator is defined by

$$\hat{\mathrm{V}}_{GBRR,\varepsilon}(\hat{\tau}) = \frac{1}{R \cdot \varepsilon^2} \cdot \sum_{r=1}^{R} (\hat{\tau}_{r,\varepsilon} - \hat{\tau})^2, \tag{3.9}$$

(cf. Davison and Sardy, 2004, p. 19, Münnich, 2005, p. 77 f.).

Another proposal is given by Rao and Shao (1996, p. 344). They suggest to repeat the random grouping T times and to compute (3.7) or (3.9) each time and averaging over the T repetitions:

$$\hat{\mathrm{V}}_{RG}(\hat{\tau}) = \frac{1}{T} \cdot \sum_{t=1}^{T} \hat{\mathrm{V}}_{GBRR}(\hat{\tau}), \tag{3.10}$$

(cf. Rao and Shao, 1996, p. 344).

Wolter (2007, p. 130 f.) recommends to subdivide the real stratum h (for $h = 1, , H$) into m_h artificial strata with two elements per stratum and to apply the basic model of balanced repeated replication (cf. Münnich, 2005, p. 78).

For stratified multistage designs half samples are often taken only at the first stage (for more details see Shao et al., 1998, p. 822 ff., Shao and Tu, 1995, p. 241 ff., Rao et al., 1992, p. 213 f., Rust and Rao, 1996, p. 289 ff., Rao and Shao, 1999, p. 405 ff.).

3.3 Jackknife

Another resampling method of great importance is the jackknife. This procedure was first introduced by QUENOUILLE (1949) as a technique which can reduce the estimation bias of a serial correlation coefficient, but he generalized the method in his 1956 paper (cf. QUENOUILLE, 1956) with respect to an infinite population context. TUKEY (1958) proposed to treat this subsample estimator as an independent and identically distributed random variable which can be used as a variance estimator (cf. WOLTER, 2007, p. 151, for the following explanations see SHAO and TU, 1995, p. 4 ff.). As mentioned before $\hat{\tau} = \hat{\tau}(Y_1, \ldots, Y_n)$ describes an estimator of an unknown population parameter τ of interest like total values. $\hat{\tau}_{-i} = \hat{\tau}(Y_1, \ldots Y_{i-1}, Y_{i+1}, \ldots, Y_n)$ is the same statistic, but based on observations by omitting one of the elements. The jackknife bias estimator of Quenouille is defined by

$$\hat{b}_{d1JK} = (n-1) \cdot \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \hat{\tau}_{-i} - \hat{\tau} \right). \tag{3.11}$$

The jackknife estimator of τ which reduces the bias results from

$$\hat{\tau}_{d1JK} = \hat{\tau} - \hat{b}_{d1JK} = n \cdot \hat{\tau} - \frac{(n-1)}{n} \cdot \sum_{i=1}^{n} \hat{\tau}_{-i}. \tag{3.12}$$

This estimator can also be written as

$$\hat{\tau}_{d1JK} = \frac{1}{n} \cdot \sum_{i=1}^{n} \hat{\tau}'_i. \tag{3.13}$$

$\hat{\tau}'_i = n \cdot \hat{\tau} - (n-1) \cdot \hat{\tau}_{-i}$ (where $i = 1, \ldots, n$) are defined by Tukey as the jackknife pseudovalues. He made the conjecture that the pseudovalues are nearly independent and identically distributed and show approximately the same variance as $\sqrt{n} \cdot \hat{\tau}$. So $\mathrm{V}(\sqrt{n} \cdot \hat{\tau})$ can be estimated by the sample variance on the basis of $\hat{\tau}'_1, \ldots, \hat{\tau}'_n$ and with that to estimate $\mathrm{V}(\hat{\tau})$ by

$$\begin{aligned} \hat{\mathrm{V}}_{d1JK}(\hat{\tau}) &= \frac{1}{n \cdot (n-1)} \cdot \sum_{i=1}^{n} \left(\hat{\tau}'_i - \frac{1}{n} \cdot \sum_{j=1}^{n} \hat{\tau}'_j \right)^2 \\ &= \frac{n-1}{n} \cdot \sum_{i=1}^{n} \left(\hat{\tau}_{-i} - \frac{1}{n} \cdot \sum_{j=1}^{n} \hat{\tau}_{-j} \right)^2. \end{aligned} \tag{3.14}$$

This is called the (delete-1) jackknife variance estimator for $\hat{\tau}$ (cf. SHAO and TU, 1995, p. 6). As shown in formula (3.14) the delete-1 jackknife estimates the variance by omitting

one element of the observed sample in each replication and by computing the population parameter of interest on the basis of these replications. Each element is deleted one time and in doing so n replications result. This could lead to a huge computational effort with an increasing sample size. In case of a stratified sampling design the sample weights have to be adjusted (cf. MÜNNICH, 2008, p. 326). When for example element i in stratum h is omitted, then the weights for element j in stratum k also have to be recalculated:

$$w_{k,j}^{(hi)} := \begin{cases} w_{kj}, & \text{if } k \neq h, \\ w_{hj} \cdot n_h / (n_h - 1), & \text{if } k = h, j \neq i, \\ 0, & k = h, j = i, \end{cases} \tag{3.15}$$

(cf. DAVISON and SARDY, 2004, p. 14).

In such sample designs the delete-1 jackknife variance estimator is defined as

$$\hat{\mathrm{V}}_{d1JK,strat}(\hat{\tau}) = \sum_{h=1}^{H} \frac{(1 - f_h) \cdot (n_h - 1)}{n_h} \sum_{i=1}^{n_h} \left(\hat{\tau}_{h,-i} - \overline{\hat{\tau}_h}\right)^2. \tag{3.16}$$

As mentioned before $-i$ implies that element i is omitted. f_h is the sampling fraction in the h-stratum when the sampling design is without replacement and $\overline{\hat{\tau}_h}$ is the average of the delete-1 jackknife estimator of the population parameter of interest in stratum h (cf. MÜNNICH, 2008, p. 326).

3.3.1 Delete-1 jackknife for complex survey designs

An example for a delete-1 jackknife for unistage stratified samples with unequal probabilities is given in BERGER (2007). In assuming that our interesting population parameter θ is obtained by a function of means $\theta = g(\mu_1, \ldots, \mu_O)$, where O is the number of survey variables. The estimation of θ is done by the substitution estimator $\hat{\theta} = g(\hat{\mu_1}, \ldots, \hat{\mu_O})$ and for $\hat{\mu_O} = \sum_{i \in s} w_i^{ha} \cdot y_i$ the so called Hájek estimator is used with $w_i^{ha} = 1/\pi_i \left(\sum_{v \in s} 1/\pi_v\right)$.

A generalized jackknife which is applicable for unequal probability sampling and stratification is given in CAMPBELL (1980) and is defined by

$$\hat{\mathrm{V}}_{jack}\left(\hat{\theta}\right) = \sum_{i \in s} \sum_{j \in s} \frac{\pi_{ij} - \pi_i \cdot \pi_j}{\pi_{ij}} \cdot u_i^g \cdot u_j^g \tag{3.17}$$

with $u_i^g = \left(1 - w_i^{ha}\right) \cdot \left(\hat{\theta} - \hat{\theta}_{-i}\right)$ (cf. BERGER, 2007, p. 955). According to BERGER (2007) this generalized jackknife shows some practical disadvantages as in the case of the Horvitz-Thompson variance estimator (cf. chapter 2, p. 6). Therefore he proposes another jackknife for stratified sampling with unequal probabilities by substituting the second order inclusion probabilities by the Hájek approximation (see section 2.3.1). This variance estimator is given by

$$\hat{\mathrm{V}}_{jack}\left(\hat{\theta}\right) = \sum_{h=1}^{H} \sum_{i \in s_h} \tilde{c}_{hi} \cdot \left(u_i^g - \sum_{j \in s_h} \tilde{\nu}_{hj} \cdot u_i^g \right)^2, \tag{3.18}$$

where $\tilde{c}_{hi} = \frac{n_h}{(n_h - 1)} \cdot (1 - \pi_i), (i \in s_h)$ and $\tilde{\nu}_{hi} = \lambda_h \cdot \tilde{c}_{hi} \cdot \delta_{hi}$. When $i \in s_h$ δ_{hi} is 1 and 0 otherwise. The quantity λ_h is chosen that $\sum_{i \in s} \tilde{\nu}_{hi} = 1$; thus $\lambda_{hi} = \left(\sum_{i \in s_h} \tilde{c}_{hi}\right)^{-1}$, where $\tilde{c}_{hi}$ consists of two terms. The first term $n_h / (n_h - 1)$ is a correction of the degrees of freedom and the second term $(1 - \pi_i)$ is the varying finite population correction which is useful when the sampling fraction is large. The advantage of this jackknife estimator is that the computation of the exact second order inclusion probabilities is not necessary and the estimator has always positive values (cf. BERGER, 2007, p. 955 f.). Under unequal probability sampling BERGER (2007) shows that his proposed jackknife variance estimator is consistent under certain conditions (for more details see BERGER, 2007, p. 956 ff.).

In case of multistage sampling the question is at which stage the unities should be deleted. In the literature often the units at the first stage, so the PSUs, are omitted (cf. RAO et al., 1992, p. 211 ff., RUST and RAO, 1996, p. 287 ff.). Then a single replication is obtained by removing all units which belong to a certain PSU of the original sample. In case of stratified two stage sampling the weights of the elements of the last stage of the remaining PSUs in the stratum of the deleted PSUs have to be adjusted. By deleting PSU j in stratum k the population parameter of interest in a certain replication is computed by:

$$\widehat{\tau}_{(-j)} = \sum_{(hiq) \in s} w_{hiq} \cdot y_{hiq} + \sum_{(kiq) \in s} \{ \frac{n_k^{'}}{n_k^{'} - 1} \cdot w_{kiq} \} \cdot y_{kiq},$$

where $n_k^{'}$ is the number of PSUs in stratum k and h $\neq$ k and i $\neq$ j. w_{hiq} is the weight of the element q in PSU i in stratum h. The first term refers to the elements which are not in stratum k. The adjustment of the weights of the elements in stratum k is done in the second term. Then a customary delete-1 cluster jackknife variance is applied with the estimates of the several replications:

$$\hat{\mathrm{V}}_{d1JK}(\hat{\tau}) = \sum_{k=1}^{H} \frac{(1 - f_k) \cdot (n_k^{'} - 1)}{n_k^{'}} \sum_{j=1}^{n_k^{'}} (\hat{\tau}_{-j} - \hat{\tau})^2$$

(cf. RAO et al., 1992, p. 211 ff., RUST and RAO, 1996, p. 287 ff.).

Another possibility is to apply the delete-1 jackknife at all stages and with that to delete the elements at the several stages. In case of sampling designs with many stages this can lead to a huge effort. Under certain conditions like small sampling fractions or homogeneous Ultimate Sampling Units (USUs) in the several unities, the exclusive consideration of the first stage can be sufficient (cf. SÄRNDAL et al., 1992, p. 139 f. or LOHR, 1999, p. 209 f. for the direct estimator).

3.3.2 Extensions of the delete-1 jackknife method

Besides its possible computational effort the delete-1 jackknife variance estimator shows another disadvantage. This method is asymptotically consistent, if the estimator is smooth. But in case of nonlinear estimators, e.g. sample quantiles, the jackknife could lead to inconsistent variance estimators (cf. Shao and Wu, 1989, p. 1176). The proof for the inconsistent variance estimation of the median with the delete-1 jackknife method is given in Efron (1982, p. 16). The delete-d jackknife can be used to solve this problem (cf. Shao and Wu, 1989, p. 1176, for the following explanations to delete-d jackknife see Shao and Tu, 1995, Section 2.3 and 5.2).

This extension of the ordinary jackknife is characterized by deleting more elements d from the given sample with the size n. The population parameter of interest is computed by the remaining values:

$$\hat{\tau}_{-s_u} = \hat{\tau}_{-s_u}\left(Y_i, i \in s_u^c\right), \tag{3.19}$$

where s_u describes a subset of $\{1, \ldots, n\}$ and has the size d. s_u^c indicates the complement of s_u. The delete-d jackknife variance estimator is defined by

$$\hat{\mathrm{V}}_{ddJK}\left(\hat{\tau}\right) = \frac{n-d}{d \cdot D} \cdot \sum_{u=1}^{D} \left(\hat{\tau}_{-s_u} - \frac{1}{D} \cdot \sum_{v=1}^{D} \hat{\tau}_{-s_v} \right)^2, \tag{3.20}$$

where $\mathcal{S}$ contains all possible subsets of $\{1, \ldots, n\}$ with size d and D describes their total number in $\mathcal{S}$ which is $D = \binom{n}{d}$. If d is large it may be very extensive to compute the delete-d jackknife variance estimator. Then it is the best way to consider only a subset of $\mathcal{S}$. The question is how to choose these subsets. Shao and Tu (1995) suggest two different methods. The first is called balanced subsampling. Here, the selection of the subsets is systematic. $\mathcal{M} = \{s_1, \ldots, s_m\}$ includes the m selected subsets in $\mathcal{S}$ satisfying two assumptions which are essentially those for balanced repeated replication. At first, it is necessary that the frequency of being in a subset in $\mathcal{M}$ is equal for all elements $(i = 1, \ldots, n)$. The second assumption is that the frequency of being in a subset in $\mathcal{M}$ is equal in respect of all pairs of elements (i, j), where $1 \leq i < j \leq n$. Each subset can be regarded as a block, so $\mathcal{M}$ is called a balanced incomplete block design (BIBD) (cf. John, 1971). An approximation of the delete-d jackknife variance estimator on the base of balanced subsampling is given by

$$\hat{\mathrm{V}}_{bddJK}\left(\hat{\tau}\right) = \frac{n-d}{d \cdot m} \cdot \sum_{u=1}^{m} \left(\hat{\tau}_{-s_u} - \frac{1}{m} \cdot \sum_{v=1}^{m} \hat{\tau}_{-s_v} \right)^2, \tag{3.21}$$

where $n \leq m \leq N$.

The selection of subsets with balanced subsampling is still very extensive due to the enumeration of balanced subsets. The second method has the advantage of being less complex and is called random subsampling. Here, we draw $\{s_1, \ldots, s_m\}$ from $\mathcal{S}$ using

simple random sampling with replacement and the variance is estimated by formula (3.21) by omitting these subsets from the sample $\mathcal{S}$.

Simple random sampling without replacement is also possible. If m is much smaller than D, simple random sampling with replacement and without replacement leads to similar results. SHAO and TU (1995) proposed to use simple random sampling with replacement, because of its simplicity.

In case of a stratified random sampling design the delete-d jackknife variance estimator can also be applied. But then it is necessary to delete d elements within the strata. The several strata are divided into m_h disjoint groups l of size d, where $m_h \cdot d = n_h$. For the replications these groups are systematically deleted and the delete-d jackknife variance estimator is defined by

$$\hat{\mathrm{V}}_{ddJK,strat}\left(\hat{\tau}\right) = \sum_{h=1}^{H} \frac{(1-f_h)\cdot(n_h-d)}{m_h} \cdot \sum_{l=1}^{m_h} \left(\hat{\tau}_{h,-l} - \frac{1}{m_h} \cdot \sum_{o=1}^{m_h} \hat{\tau}_{h,-o} \right)^2 . \tag{3.22}$$

$\hat{\tau}_{h,-l}$ is the estimator in stratum h which is calculated by deleting group l of size d from the observation of the stratum h (cf. MÜNNICH, 2005, p. 85). By omitting a block of elements with indexes $i \in \mathcal{K}$ in stratum h the weights for element j in stratum k have to be adjusted:

$$w_{k,j}^{(h\mathcal{K})} := \begin{cases} w_{kj}, & \text{if } k \neq h, \\ w_{hj} \cdot n_h / \left(n_h - \|\mathcal{K}\|\right), & \text{if } k = h, j \notin \mathcal{K}, \\ 0, & k = h, j \in \mathcal{K}, \end{cases} \tag{3.23}$$

where $\|\mathcal{K}\|$ indicates the number of elements in $\mathcal{K}$, with that $\|\mathcal{K}\| = d$ (cf. DAVISON and SARDY, 2004, p. 15).

With respect to the consistency and asymptotic properties it can be said that, if an estimator is less smooth, a large d is necessary. If the estimator is smooth, d can be small (cf. SHAO and WU, 1989, p. 1176 and for the examination of consistency and asymptotic properties of the delete-d jackknife see Chapter 3).

Another possibility to reduce the complexity is using the delete-a-group jackknife. Within the scope of this method the obtained sample is divided into G groups. This can be done randomly (cf. SHAO and TU, 1995, p. 195) or as described in KOTT (2001b) and KOTT (2001a). Here, the PSUs are ordered an appropriate manner, for instance in case of stratified random sampling by stratum (cf. KOTT, 2001b, p. 3). Within a stratum the units can be ordered randomly. Afterwards the PSUs are matched systematically to the groups. For example the selection can be done by allocating the first, nineteenth, thirty-seventh, ... elements to group one, the second, twentieth, thirty-eighth, ... elements to group two and so on (cf. KOTT, 2001b, p. 3, KOTT, 2001a, p. 522). The groups should have nearly the same size and be mutually exclusive (cf. BIENIAS et al., 2003, p. 539). Instead of deleting one or d elements from the sample, one group g, $(g = 1, \ldots, G)$ in

each replication is omitted now. As before the statistic of interest is computed on each replicate and the delete-a-group jackknife variance estimator is stated as follows:

$$\hat{\mathrm{V}}_{dgJK}\left(\hat{\tau}\right) = \frac{G-1}{G} \cdot \sum_{g=1}^{G} \left(\hat{\tau}_{-g} - \frac{1}{G} \cdot \sum_{l=1}^{G} \hat{\tau}_{-l} \right)^2 , \tag{3.24}$$

(cf. SHAO and TU, 1995, p. 195).

$\hat{\tau}_{-g}$ is an estimator, for example the total value, computed by the replication without the elements corresponding to group g.

As before, in case of deleting one element i or d elements the weights have to be adjusted as well when using the delete-a-group jackknife variance estimator. This is shown in KOTT (2001a) for a stratified single-phase sampling design. If $\hat{\tau}$ is a weighted estimator of the form $\sum_{i=1}^{n} w_i \cdot y_i$, then the estimator $\hat{\tau}_{-g}$ is computed by $\hat{\tau}_{-g} = \sum_{i=1}^{n} w_{i(-g)} \cdot y_i$. When the PSU is in group g, then the weight $w_{i(-g)}$ of an element i takes the value zero. Otherwise $w_{i(-g)}$ is computed by adjusting w_i to consider that some $w_{i(-g)}$ are zero. These weights are calculated by $\left[n_h / \left(n_h - n_{hg}\right)\right] \cdot w_i$, where n_h is the number of PSUs in the stratum h. n_{hg} is the number of PSUs which are in the stratum h as well as in the group g (cf. KOTT, 2001a, p. 522). A wide discussion about the applicability of the delete-a-group jackknife for different estimation strategies which include sampling designs like stratified simple random sampling, systematic probability sampling or Poisson sampling is given in KOTT (2001b).

Another proposal to apply this method in case of a stratified design is given in RUST (1985). Here, the construction of groups is done in each stratum. The delete-a-group jackknife variance estimator is defined by

$$\hat{\mathrm{V}}_{dgJK,strat}\left(\hat{\tau}\right) = \sum_{h=1}^{H} \frac{\left(l_h - 1\right)}{l_h} \cdot \sum_{g=1}^{l_h} \left(\hat{\tau}_{-g} - \hat{\tau}\right)^2 , \tag{3.25}$$

where l_h is the number groups in stratum h $(h = 1, \ldots, H)$, $\hat{\tau}$ the estimated population parameter of interest and $\hat{\tau}_{-g}$ the same statistic which is calculated by deleting group g in stratum h. To generalize the estimator it is also possible to omit only a random selection of l_h which can be indicated as g_h. Then the delete-a-group jackknife variance estimator is stated by

$$\hat{\mathrm{V}}_{sdgJK,strat}\left(\hat{\tau}\right) = \sum_{h=1}^{H} \frac{\left(l_h - 1\right)}{g_h} \cdot \sum_{g=1}^{g_h} \left(\hat{\tau}_{-g} - \hat{\tau}\right)^2 , \tag{3.26}$$

(cf. RUST, 1985, p. 387).

As said before the delete-a-group jackknife especially provides computational advantages in comparison to the ordinary jackknife. In each replication a group of elements is omitted

instead of only one observation. As a result the number of replications is reduced from n, the number of observations in a sample, to the number of divided groups G (cf. MÜNNICH, 2005, p. 82). A disadvantage is related to one requirement of the method. So it is necessary that the sample size of the first-phase in each stratum is large. At least five units per stratum are required. Otherwise the true variance is overestimated and leads to a bias (cf. KOTT, 2001b, p. 1 f., KOTT, 2001a, p. 521 ff.). But in practice strata often have sample sizes lower than five. For these scenarios KOTT (2001a, p. 522) proposed an extended version of the ordinary delete-a-group jackknife (for more details see KOTT, 2001a, Chapter 3). As pointed out in RUST (1985, p. 387) the precision is maximized when each omitted group is of size one and is deleted one time. The larger the number of groups, the more precise is the estimation, but the more complex and extensive is the computation. The optimal number of groups depends on the underlying conditions like the complexity of the estimation (cf. BIENIAS et al., 2003, p. 542).

The delete-d jackknife, which was discussed before, can be regarded as a generalization of the delete-a-group jackknife. The latter divides the sample into mutually exclusive groups, whereas within the scope of the delete-d jackknife all groups of size d are possible (cf. MÜNNICH, 2005, p. 84).

3.4 Bootstrap

The last resampling method we discuss will be the bootstrap. EFRON (1979) introduced this method which in his opinion is more applicable and more dependable than the jackknife. Let $X_1, \ldots, X_n$ describe an independent and identically distributed sample of n observations from an unknown distribution F and $\hat{\tau}(X_1, \ldots, X_n)$ indicates a given statistic. In case of non independent $X_1, \ldots, X_n$ there are some problems by applying the bootstrap method. These problems are explained in the next section. The estimation of F is described as $\hat{F}$.

The bootstrap variance estimator results by substituting F in the theoretical formula of the variance defined as

$$\mathrm{V}(\hat{\tau}) = \int \left[\hat{\tau}(x) - \int \hat{\tau}(y)\, d \prod_{i=1}^{n} F(y_i) \right]^2 d \prod_{i=1}^{n} F(x_i) \tag{3.27}$$

by $\hat{F}$:

$$\begin{aligned} \hat{\mathrm{V}}_{boot}(\hat{\tau}) &= \int \left[\hat{\tau}(x) - \int \hat{\tau}(y)\, d \prod_{i=1}^{n} \hat{F}(y_i) \right]^2 d \prod_{i=1}^{n} \hat{F}(x_i) \\ &= \mathrm{V}_* \left[\hat{\tau}(X_1^*, \ldots, X_n^*) \,|X_1, \ldots, X_n \right], \end{aligned} \tag{3.28}$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_1)$ (cf. SHAO and TU, 1995, Section 1.2 and 1.4).

Formula (3.28) is called the theoretical bootstrap. $X_1^*, \ldots, X_n^*$ is a sample of independent and identically distributed elements and drawn from $\hat{F}$. This sample is denoted as the bootstrap sample and $\mathrm{V}_* [\cdot | X_1, \ldots, X_n]$ indicates the conditional variance for given $X_1, \ldots, X_n$ (cf. Shao and Tu, 1995, Section 1.2 and 1.4). Is $\hat{F}$ equal to F then

$$\mathrm{V}(\hat{\tau}) = \mathrm{V}_* [\hat{\tau}(X_1^*, \ldots, X_n^*) | X_1, \ldots, X_n]. \tag{3.29}$$

When $\hat{F} \neq F$ the equivalence of the variances is not given anymore. But the bootstrap variance $\hat{\mathrm{V}}_{boot}(\hat{\tau})$ can be taken as a variance estimator for $\hat{\tau}$ (cf. Shao, 2003, p. 380 f., Münnich, 2005, p. 88 f.).

3.4.1 The Monte Carlo bootstrap

Often in practice the theoretical bootstrap can not be written as an explicit function of $X_1, \ldots, X_n$ and as a result this approach is not applicable. To estimate the variance we can make use of Monte Carlo methods and approximate $\hat{\mathrm{V}}_{boot}(\hat{\tau})$ (cf. Shao and Tu, 1995, p. 10 f.). This is done by the Monte Carlo bootstrap. Within the scope of this method subsamples of size n are drawn from the original sample by simple random sampling with replacement. Overall, B subsamples are drawn and the population parameter of interest τ is calculated on these subsamples which is denoted by $\hat{\tau}_i^*$. The Monte Carlo bootstrap variance estimator results from

$$\hat{\mathrm{V}}_{boot,MC}(\hat{\tau}) = \frac{1}{B-1} \cdot \sum_{i=1}^{B} \left(\hat{\tau}_i^* - \frac{1}{B} \cdot \sum_{j=1}^{B} \hat{\tau}_j^* \right)^2, \tag{3.30}$$

(cf. Münnich, 2008, p. 326).

Due to the law of large numbers is $\hat{\mathrm{V}}_{boot}(\hat{\tau}) = \lim_{B \to \infty} \hat{\mathrm{V}}_{boot,MC}(\hat{\tau})$ (cf. Shao and Tu, 1995, p. 11).

In this context, an appropriate number of replications B must be chosen. With a larger B the goodness of the approximation increases but the same applies for the complexity and extent of the computation. The same statistic is evaluated $B+1$ times and this could lead to a huge effort (cf. Münnich, 2005, p. 91). But in general the bootstrap variance estimator doesn't need a lot of replications. In many cases 100 resamples are sufficient (cf. Münnich, 2008, p. 326).

Is the sampling design stratified, then a random subsample with replacement of size n_h is drawn in each stratum h ($h = 1, \ldots H$) from the original sample. This has no impact on the design weights because the drawing is independent between strata. As before an estimator for stratified random sampling is computed at each replication (cf. Münnich, 2005, p. 92 f.).

The bootstrap has a huge advantage in contrast to the ordinary delete-1 jackknife. Under acceptable conditions with respect to the underlying distribution the bootstrap variance estimator is consistent for nonlinear estimators (cf. SHAO and WU, 1989, p. 1176).

Otherwise, as mentioned above, the Monte Carlo bootstrap draws subsamples with replacement and requires an original sample whose observations are independent and identically distributed like they occur for SRS and StrSRS with replacement. But in surveys often a probability sampling design or a without replacement sample design is used which cause dependencies. When sampling without replacement, the Monte Carlo bootstrap can lead to a biased variance estimator because of the missing finite sampling correction. Furthermore, the approach has disadvantages in case of large sampling fractions and small sample sizes. Then the bootstrap variance estimator can be inconsistent (cf. DAVISON and SARDY, 2004, p. 21 f.). For that reason modified versions of the ordinary bootstrap are given in the literature, which will be discussed briefly.

The first modification is the without replacement bootstrap.The intention of this approach is to create a pseudopopulation by generating N/n replicates of the original sample. Afterwards B subsamples are drawn randomly from this population but without replacement. In case of stratified samples the same is done, but with regard to the several strata. This means that for each stratum the replication is done N_h/n_h times and in each stratum several samples are taken without replacement B times. In case of a noninteger N_h/n_h, the sizes of the pseudopopulations is randomly selected from:$\{\lfloor N_h \rfloor, \lceil N_h \rceil\}$. Furthermore, a pseudopopulation for each stratum requires more storage, which is hardly manageable when there are a lot of large strata (cf. DAVISON and SARDY, 2004, Section 6.1 and 6.2). For that reason some other proposals are given in the literature.

MCCARTHY and SNOWDEN (1985, p. 4) proposed the with-replacement bootstrap. Samples are drawn in each stratum with replacement, where the sample size in each stratum v_h is calculated by

$$v_h = \frac{(n_h - 1)}{(1 - f_h)}. \tag{3.31}$$

An advantage of this procedure is that it avoids creating a pseudopopulation which has to be stored. Furthermore the resulting variance estimator is unbiased and consistent. But if v_h is noninteger a randomization is also needed. (cf. DAVISON and SARDY, 2004, p. 23).

Another kind of bootstrap is the rescaling bootstrap which was suggested by RAO and WU (1988). When it is possible to write an estimator as a function of means $\tau = g(\overline{y})$ this method can be applied. Subsamples of size m_h are sampled with replacement, but the bootstrap sample is rescaled by

$$\tilde{y}_h^* = \overline{y}_h + \sqrt{m_h \cdot \frac{(1 - f_h)}{(n_h - 1)}} \cdot (\overline{y}_h^* - \overline{y}_h), \tag{3.32}$$

where $\overline{y}_h$ is the estimated mean in stratum h computed on the original sample and $\overline{y}_h^*$ is the estimated mean in stratum h calculated on the bootstrap sample. For m_h RAO and WU (1988) proposed choosing

$$m_h = \left\lfloor \frac{(1-f_h)\cdot(n_h-2)^2}{(1-2f_h)\cdot(n_h-1)} \right\rfloor. \tag{3.33}$$

Then the third order moments are matched and no problems arise when m_h is not an integer. The estimated population mean, which is computed on the rescaled bootstrap sample, is given by

$$\tilde{y}^* = \sum_{h=1}^{H} \frac{N_h}{N} \cdot \tilde{y}_h^*, \tag{3.34}$$

where N_h is the frequency of elements in the population in stratum h. Then the population parameter of interest is computed by

$$\hat{\tau}^* = g\left(\tilde{y}^*\right). \tag{3.35}$$

The whole process is repeated B times and the variance estimation is based on the several estimates $\hat{\tau}_i^*$. An advantage of this approach is that it avoids generating and storing a pseudopopulation (cf. Sitter, 1992b, p. 137, Davison and Sardy, 2004, p. 23).

Sitter (1992a) suggests the mirror-match bootstrap. This approach takes subsamples without replacement in each stratum separately of size n_h^*, where $1 \leq n_h^* < n_h$, altogether $k_h = \frac{n_h \cdot \left(1-f_h^*\right)}{n_h^* \cdot (1-f_h)}$ times. Sitter (1992a) propose $n_h^* = f_h \cdot n_h$. Then the third order moments are matched. When the statistics of interest are linear the mirror-match bootstrap is consistent. In case of a noninteger k_h a randomization between the bracketing integers is necessary (cf. Davison and Sardy, 2004, p. 23 Sitter, 1992a, p. 756 ff. Sitter, 1992b, p. 138).

The Monte Carlo bootstrap and the jackknife have in common that both procedures permit the computation of the bias. Therefore we can use

$$\hat{b}_{boot,MC} = \frac{1}{B} \cdot \sum_{i=1}^{B} \hat{\tau}_i^* - \hat{\tau}, \tag{3.36}$$

(cf. Efron, 1982, p. 33).

Another application of the Monte Carlo bootstrap is to compute confidence intervals. With recourse to the quantiles

$$z_i^* = \frac{\hat{\tau}_1^* - \hat{\tau}}{\sqrt{\hat{\mathrm{V}}_{boot,MC}\left(\hat{\tau}_1^*\right)}}, \ldots \frac{\hat{\tau}_B^* - \hat{\tau}}{\sqrt{\hat{\mathrm{V}}_{boot,MC}\left(\hat{\tau}_B^*\right)}} \tag{3.37}$$

we can ascertain the $\alpha/2$- and $(1-\alpha/2)$ quantile $z^*_{\alpha/2}$ and $z^*_{1-\alpha/2}$. $\hat{\mathrm{V}}_{boot,MC}(\hat{\tau})$ is defined by (3.30). The studentized bootstrap confidence interval is given by

$$\left[\hat{\tau} - \sqrt{\hat{\mathrm{V}}_{boot,MC}(\hat{\tau})} \cdot z^*_{1-\alpha/2}; \hat{\tau} - \sqrt{\hat{\mathrm{V}}_{boot,MC}(\hat{\tau})} \cdot z^*_{\alpha/2}\right] \tag{3.38}$$

(cf. DAVISON and HINKLEY, 1997, p. 27 ff., MÜNNICH, 2005, p. 92). In contrast to the variance estimation more replications are needed. To receive a high accuracy DAVISON and HINKLEY (1997) suggest more than 1000 replications. The bootstrap method can also be used for hypothesis testing with respect to unknown model parameters. A proposal for this issue in case of a probability sampling design is given in BEAUMONT and BOCCI (2009).

3.4.2 Bootstrap with complex survey designs

An example of a more complex survey design, using unequal probability sampling without replacement, is the Rao-Hartley-Cochran (RHC) sampling method (cf. RAO et al., 1962). Here the population of size N is divided into L groups (G_g) of size N_g with $g = 1, \ldots, L$. Then from the several groups one unit is drawn with the probability z_i/Z_g for group G_g, where $z_i = x_i/X$, $Z_g = \sum_{i \in G_g} z_i$, x_i describes some size measure of unit i and $X = \sum_{i=1}^{N} x_i$. An unbiased estimator of the population mean is given by $\hat{\bar{Y}} = \sum_{g=1}^{L} w_g \cdot y_g/L$, where $w_g = f/\pi_g$ and π_g denotes the inclusion probability of an unit which is drawn from group g and defined by z_g/Z_g. f indicates the inclusion probability in case of simple random sampling without replacement and is defined by n/N. Where, y_g and z_g are values which are chosen from group g, i.e. $y_g \in \{y_i\}_{j=1}^{G_g}$ and $z_g \in \{z_j\}_{j=1}^{G_g}$, with $\sum_{g=1}^{L} Z_g = 1$. An unbiased estimator for the variance of $\hat{\bar{Y}}$ is given by

$$\hat{\mathrm{V}}\left(\hat{\bar{Y}}\right) = \frac{\sum_{g=1}^{L} N_g^2 - N}{N^2 - \sum_{g=1}^{L} N_g^2} \cdot \sum_{g=1}^{L} Z_g \left(\frac{y_g}{N \cdot z_g} - \hat{\bar{Y}}\right)^2 \tag{3.39}$$

(cf. SITTER, 1992b, p. 142).

In general, it is possible to generate the subsamples by attaching to each unit from G_k in the original sample the probability P_k and then select a sample $\{*\mathbf{y}, *\mathbf{z}\}$ of size m with replacement with probabilities P_k from the original sample (cf. RAO and WU, 1988, p. 237). But for our purposes we will refer to the bootstrap in case of unequal probability sampling which is described in SITTER (1992b, p. 142) by generating a pseudopopulation and drawing random samples without replacement.

Let $\hat{Y} = N \cdot \hat{\bar{Y}} = \sum_{g=1}^{L} Z_g \cdot y_g/z_g$. First a replication for (z_g, y_g) is done $k_g = Z_g/z_g$ times for $g = 1, \ldots, L$ to generate a pseudopopulation as described at p. 30. Afterwards this pseudopopulation of size N^* is randomly divided into L^* groups $\left(\Gamma^*_g\right)$ of sizes N^*_g. In the next step one pair (z^*_i, y^*_i) is chosen randomly from each group with the probability z^*_i/Z^*_g, where $Z^*_g = \sum_{i \in \Gamma^*_g} z^*_i$. Then the population parameter of interest is computed by

$\hat{\Theta}^* = \hat{\Theta}(\mathbf{z}^*, \mathbf{y}^*)$. This process is repeated B times to get the estimates $\hat{\Theta}^*_1, ..., \hat{\Theta}^*_B$. As before the variance estimation is calculated on the several estimates (cf. Sitter, 1992b, p. 142). If the population parameter of interest is a total value instead of the mean, the variance estimator of the form (3.30) shows a small bias. Then it is possible to correct the variance estimate by an appropriate multiplicative factor or to choose appropriate values of n^* and N^*_g (for more details see Sitter, 1992b, p. 142).

A bootstrap for two-stage cluster sampling is given in Wolter (2007, p. 211). This bootstrap is applied at the first stage of the sampling design and when a PSU is drawn with replacement from the original sample all units of the following stages are also included in the bootstrap replicate (cf. Wolter, 2007, p. 211). A further proposal of a particular bootstrap for multistage designs is an extension of the rescaling bootstrap which was presented before and is given in Rao et al. (1992) or Rust and Rao (1996). In contrast to this bootstrap, the rescaling of the extended bootstrap is done to the survey weights rather than to the observed values. The extended bootstrap is appropriate for smooth statistics as well as non-smooth statistics (cf. Preston, 2009, p. 228). In detail, the approach can be described as follows: From the $n_h^{'}$ PSUs of the original sample $m_h^{'}$ PSUs are drawn with replacement. But for each of the B repetitions the survey weights are adjusted:

$$w^*_{hiq} = \left[\left(1 - \left(\frac{m_h^{'}}{n_h^{'} - 1}\right)^{1/2}\right) + \left(\frac{m_h^{'}}{n_h^{'} - 1}\right)^{1/2} \cdot \left(\frac{n_h^{'}}{m_h^{'}}\right) \cdot r_{hi}\right] \cdot w_{hiq}, \tag{3.40}$$

where r_{hi} describes the number of times a certain PSU appears in the subsample. By using these weights for the estimation of the population parameter of interest and repeating this procedure B times, the variance estimation is done on the B estimates as before. The value of $m_h^{'}$ has to be determined. An appropriate choice with only a little, if any loss in efficiency is $m_h^{'} = (n_h^{'} - 1)$. Then the calculation of bootstrap weights reduces to:

$$w^*_{hiq} = w_{hiq} \cdot \frac{n_h^{'}}{(n_h^{'} - 1)} \cdot r_{hi}$$

(cf. Rao et al., 1992, p. 214 f., Rust and Rao, 1996, p. 291 f.).

Another modification of the rescaling bootstrap for multistage designs is the without replacement rescaling bootstrap which is presented in Chipperfield and Preston (2007, p. 169 f.). An application of this bootstrap for stratified three stage sampling is given in Preston (2009, p. 228 ff.). But an extension to more stages is possible. The without replacement rescaling bootstrap is applied as follows: At the first stage $m_h^{'}$ PSUs are taken randomly without replacement from the $n_h^{'}$ PSUs from the original sample. The PSU bootstrap weights are computed by

$$w^{*'}_{hi} = \left(1 - \lambda_h^{'} + \lambda_h^{'} \cdot \frac{n_h^{'}}{m_h^{'}} \cdot \delta^{'}_{hi}\right) \cdot w^{'}_{hi}, \tag{3.41}$$

where $\lambda_h^{'} = \sqrt{m_h^{'} \cdot \frac{(1-f_h^{'})}{(n_h^{'}-m_h^{'})}}$, $f_h^{'} = \frac{n_{hi}^{'}}{N_h^{'}}$ and $\delta_{hi}^{'}$ is 1 when PSU i in stratum h is drawn and 0 otherwise.

At the second stage from the $n_{hi}^{''}$ SSUs within the PSUs $m_{hi}^{''}$ SSUs are taken randomly without replacement. The SSU bootstrap weights are calculated by

$$\begin{aligned} w_{hiq}^{*''} &= (1-\lambda_h^{'}+\lambda_h^{'} \cdot \frac{n_h^{'}}{m_h^{'}} \cdot \delta_{hi}^{'} \\ &- \lambda_{hi}^{''} \cdot \sqrt{\frac{n_h^{'}}{m_h^{'}}} \cdot \delta_{hi}^{'} + \lambda_{hi}^{''} \cdot \sqrt{\frac{n_h^{'}}{m_h^{'}} \cdot \delta_{hi}^{'} \cdot \frac{n_{hi}^{''}}{m_{hi}^{''}} \cdot \delta_{hiq}^{''}}) \cdot w_{hiq}^{''} \cdot \frac{w_{hi}^{'}}{w_{hi}^{*'}}, \end{aligned}$$

where $\lambda_{hi}^{''} = \sqrt{m_{hi}^{''} \cdot f_h^{'} \frac{(1-f_{hi}^{''})}{(n_{hi}^{''}-m_{hi}^{''})}}$, $f_h^{''} = \frac{n_{hi}^{''}}{N_h^{''}}$ and $\delta_{hiq}^{''}$ is 1 when SSU q in PSU i in stratum h is chosen and 0 otherwise.

Finally at the third stage from the $n_{hiq}^{'''}$ USUs within the SSUs $m_{hiq}^{'''}$ USUs are drawn randomly without replacement. As before the USU bootstrap weights have to be adjusted:

$$\begin{aligned} w_{hiqp}^{*'''} &= (1-\lambda_h^{'}+\lambda_h^{'} \cdot \frac{n_h^{'}}{m_h^{'}} \cdot \delta_{hi}^{'} - \lambda_{hi}^{''} \cdot \sqrt{\frac{n_h^{'}}{m_h^{'}}} \cdot \delta_{hi}^{'} + \lambda_{hi}^{''} \cdot \sqrt{\frac{n_h^{'}}{m_h^{'}}} \cdot \delta_{hi}^{'} \cdot \frac{n_{hi}^{''}}{m_{hi}^{''}} \cdot \delta_{hiq}^{''} \\ &- \lambda_{hiq}^{'''} \cdot \sqrt{\frac{n_h^{'}}{m_h^{'}}} \cdot \delta_{hi}^{'} \cdot \sqrt{\frac{n_{hi}^{''}}{m_{hi}^{''}}} \cdot \delta_{hiq}^{''} \\ &+ \lambda_{hiq}^{'''} \cdot \sqrt{\frac{n_h^{'}}{m_h^{'}}} \cdot \delta_{hi}^{'} \cdot \sqrt{\frac{n_{hi}^{''}}{m_{hi}^{''}}} \cdot \delta_{hiq}^{''} \cdot \frac{n_{hiq}^{'''}}{m_{hiq}^{'''}} \cdot \delta_{hiqp}^{'''}) \cdot w_{hiqp}^{'''} \cdot \frac{w_{hi}^{'}}{w_{hi}^{*'}} \cdot \frac{w_{hiq}^{''}}{w_{hiq}^{*''}}, \end{aligned}$$

where $\lambda_{hiq}^{'''} = \sqrt{m_{hiq}^{'''} \cdot f_h^{'} \cdot f_{hi}^{''} \cdot \frac{(1-f_{hiq}^{'''})}{(n_{hiq}^{'''}-m_{hiq}^{'''})}}$, $f_h^{'''} = \frac{n_{hi}^{'''}}{N_h^{'''}}$ and $\delta_{hiqp}^{'''}$ is 1 when USU p in SSU q in PSU i in stratum h is chosen and 0 otherwise.

By using these weights at the several stages for the estimation of the population parameter of interest the variance is calculated as above in case of the extended rescaling bootstrap (cf. PRESTON, 2009, p. 228 ff.). PRESTON (2009, p. 229) proposed the selection of $m_h^{'} = \lfloor n_h^{'}/2 \rfloor$, $m_{hi}^{''} = \lfloor n_{hi}^{''}/2 \rfloor$ and $m_{hiq}^{'''} = \lfloor n_{hiq}^{'''}/2 \rfloor$. This has the advantage that negative weights are avoided.

Another interesting bootstrap for such sampling schemes is given in FUNAOKA et al. (2006, p. 151 ff.) and is called the Bernoulli bootstrap. But this bootstrap is not presented in our paper.

3.4.3 Parametric bootstrap

Besides the Monte Carlo bootstrap there is another kind of bootstrap proposed in the literature which is called the parametric bootstrap. Here, we assume that our sample observations $y_1, \ldots, y_n$ follow a certain distribution F_ψ which depends on a certain parameter ψ. First, the parameter ψ is estimated by $\hat{\psi}$ on the basis of $y_1, \ldots, y_n$ (e.g. with the maximum likelihood method). When this estimated parameter is substituted into the model, the fitted model results, with $\hat{F} = F_{\hat{\psi}}$. The parametric bootstrap doesn't estimate the variance by drawing subsamples from the sample itself, but by sampling from the fitted distribution $\hat{F}$. So we get R replications and for each replication we estimate the population parameter of interest $\hat{\tau}_r^*$. The variance estimation with the parametric bootstrap is given by

$$\hat{\mathrm{V}}_{boot,para}\left(\hat{\tau}\right) = \frac{1}{R-1} \cdot \sum_{r=1}^{R} \left(\hat{\tau}_r^* - \frac{1}{R} \cdot \sum_{j=1}^{R} \hat{\tau}_r^* \right)^2 , \tag{3.42}$$

(cf. Davison and Hinkley, 1997, p. 15 f.).

As mentioned in Efron and Tibshirani (1993, p. 55 f.), the parametric bootstrap is useful in surveys where information about the underlying population is obtainable, especially their distribution, and for making comparisons with the nonparametric bootstrap.

An interesting proposal to review the accuracy of bootstrap estimates is illustrated in Efron (1992). Here, the jackknife is used to calculated standard errors for bootstrap estimates. The jackknife is well-suited for such applications, because the standard error estimated by the jackknife method can be computed immediately from the bootstrap replications and a further resampling is not necessary (cf. Efron, 1992, p. 83).

3.5 Discussion

Table 3.1 shows a comparison of the different resampling methods that have been presented in this section. It should give some recommendations of the suitability of the resampling methods in different areas of applications.

The BRR (3.6), the delete-d jackknife, the delete-a-group jackknife, the rescaling bootstrap (3.40) and the rescaling bootstrap without replacement (3.41) are not described explicitly for unequal probability sampling. For that reason these methods are not considered for such sampling designs (see row four). Sampling WR/ WoR indicates if variance estimation method is suitable for sampling with replacement (WR) or without replacement (WoR). The last row indicates whether the variance estimation method considers a finite population correction (FPC) or not. This is important when the samples are drawn without replacement and the sampling fraction is large.

The delete-a-group jackknife as described in (3.24) does not include a FPC. However, the FPC can be included in the same way as explained for the delete-1 jackknife in (3.16) or the delete-d jackknife in (3.22).

Table 3.1: Comparison of resampling methods

Method	BRR (Basic Model)	BRR (3.6)	Delete-1 Jackknife	Delete-d Jackknife	Delete-a-Group Jackknife	Monte Carlo Bootstrap	Rescaling Bootstrap (3.40)	Rescaling Bootstrap WoR (3.41)
Statistic	Smooth and non-smooth	Smooth and non-smooth	Only for smooth statistics	Smooth and non-smooth	Smooth and non-smooth	Smooth and non-smooth	Smooth and non-smooth	Smooth and non-smooth
Strati-fication	Only when 2 elements per stratum	Required	Appropriate	Appropriate	Appropriate	Appropriate	Appropriate	Appropriate
Unequal Probabil-ity Sampling	Wolter (2007, p. 113)	Not considered	Berger (2007)	Not considered	Not considered	The ordinary Monte Carlo Bootstrap may lead to biased variance estimates	Not considered	Not considered
Sampling WR/WoR	WR	WoR	WR/WoR	WR/WoR	WR/WoR	WR	WR	WoR
FPC	Not considered	Considered	Possible	Possible	Possible	Not considered	Not considered	Considered

Chapter 4

Simulation Study

The purpose of this simulation study is to evaluate variance estimators, especially those described in chapter (3). As a complex sampling design a two stage design is chosen. The main focus of the simulation is put on the comparison of the accuracy of different variance estimation methods with respect to the distribution of the study variable within the PSUs. The simulation study should investigate the influence of different parameters like the size of the PSUs or the homogeneity of the elements between and within the PSUs on different variance estimation methods. The simulation should reveal in which way specific parameter constellations affect the different variance estimation methods. The sampling design is simple (without stratification and without unequal probabilities) in order to isolate the effects of the clustering of sample elements in the PSUs.

The statistic of interest is the total. Samples are drawn by means of a two stage sampling design, at the first stage PSUs are selected by SRS and at the second stage SSUs are sampled by SRS within each PSU. Because of the absence of any stratification, the balanced repeated replication method as described in section (3.2) is not included. The same goes for the variance estimators and their approximations as described in chapter (2), as all sampling units are selected with equal probability. The quality of the variance estimation methods from chapter (2) and are investigated in Hulliger et al. (2011).

In the simulation study the direct estimation method as defined in formula (1.6), the delete-1 jackknife (cf. section 3.3), an ordinary Monte Carlo bootstrap (cf. section 3.4.1), the rescaling bootstrap and the rescaling bootstrap without replacement as described in section (3.4.2) are considered. For the direct estimation, the jackknife and the Monte Carlo bootstrap two versions are investigated. One version estimates the variance by considering both stages of the sampling design whereas the other version uses the first stage only. Under some conditions, such as a small variance at the second stage and small sampling fraction at the first stage it is also possible to estimate the variance of a two stage sampling design by only taking into account the variance due to the the first stage of sampling (cf. Särndal et al., 1992, p. 139 f., Lohr, 1999, p. 209 f.). For that reason, this version of the three methods is also considered to investigate the possibility to ignore the resampling of SSUs. The rescaling bootstrap is done only at the first stage, whereas the rescaling bootstrap without replacement includes both stages. To be sure that our variance estimates are reliable we chose the rather prudent number of 450 bootstrap replications.

The study was preformed using R, (see R DEVELOPMENT CORE TEAM, 2010).

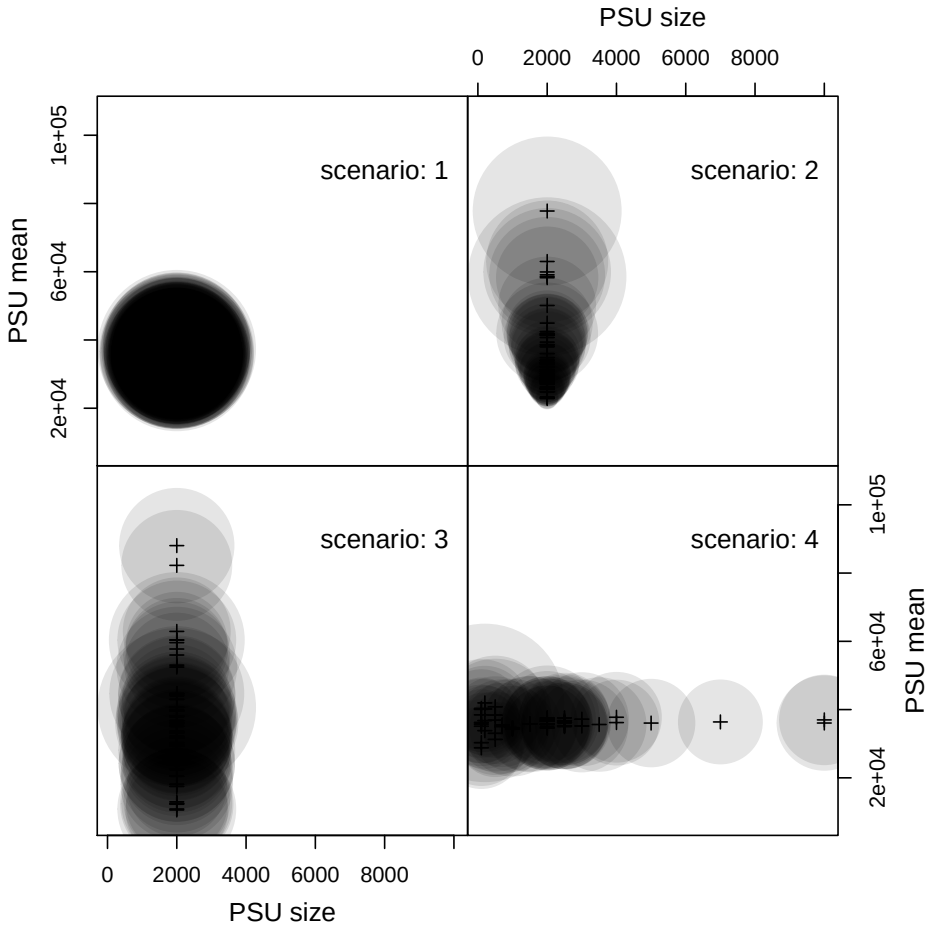

Figure 4.1: Scenario 1-4

To investigate the influence of different circumstances seven scenarios with different compositions of the PSUs are considered. The values of the study variable populating each PSU are generated from a log-normal distributions. The scenarios differ with respect to the mean and the variance of these log-normal distributions and the PSU size. The different scenarios are displayed in figure (4.1) and (4.2). The x-axis indicates the PSU size and the y-axis describes the PSU mean. The radii of the circles for the different PSUs correspond to the coefficient of variation within the particular PSU.

Scenario 1 describes a situation where the PSUs are of equal size, the mean of the different PSUs and the variances within the PSUs are similar. The other scenarios result by varying the three parameters. First, in scenario 2, the variances within the PSUs differ significantly and in scenario 3 the expected values vary between the PSUs. Scenario 4 constitutes a situation where the size of the different PSUs is highly unequal. Furthermore, it is possible to vary more than one parameter at a time. Hence, scenario 5 is characterized by different expected values and unequal sizes and scenario 6 by different variances within the PSUs and unequal sizes. In scenario 7 all three parameters vary greatly.

The observed values are generated using the `rlnorm` function in R. For scenario 1 the call was:

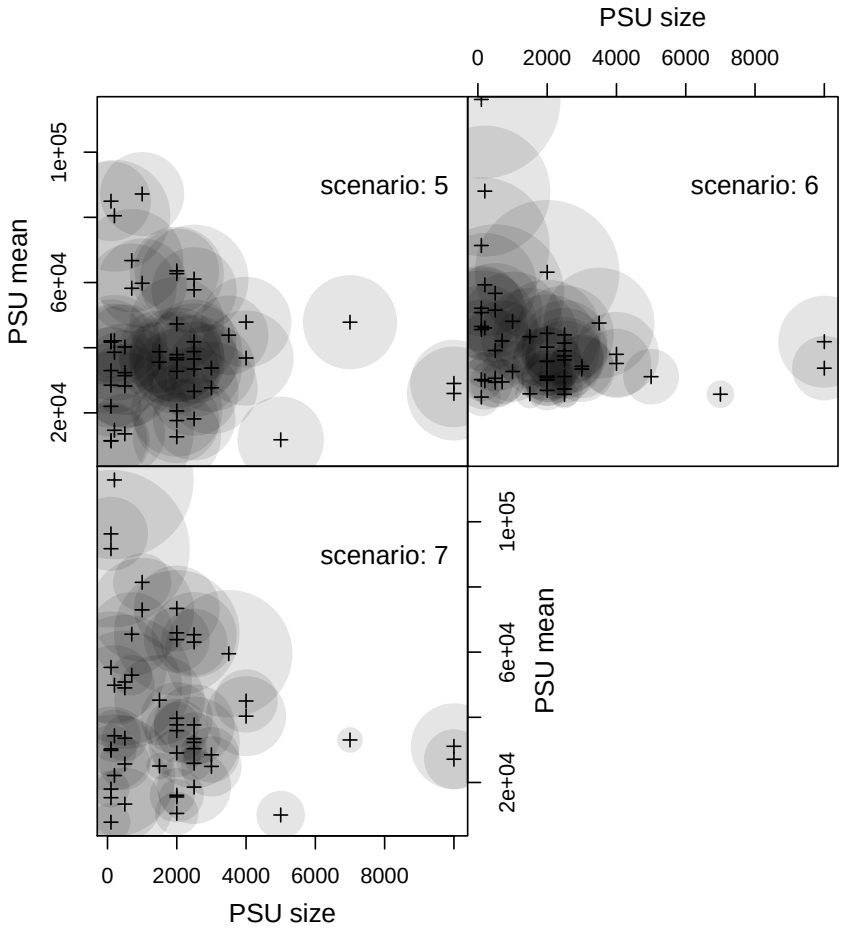

Figure 4.2: Scenario 5-7

```
income<-rlnorm(100000, meanlog = 10, sdlog = 1)
```

In scenarios with different expected values and variances the second and third parameter are varied between different PSUs. These different values are drawn from a normal distribution using the following R expressions:

```
EVPSU <- rnorm(50,10,0.5)

StDvPSU <- rnorm(50,1,0.25),
```

where the first parameter describes the number of PSU in the population. To indicate the PSUs a PSU indicator variable is used. In case of equal PSUs sizes with 50 PSUs, this variable is generated as follows:

```
psu<-rep(1:50,each=2000).
```

But in case of unequal PSU sizes, the PSU indicator is constructed by:

```
psugr<-c(rep(100,8),rep(2000,2),rep(1000,2),rep(500,5),
         rep(1500,2),rep(2000,8),rep(2500,8),rep(3000,2),
         rep(3500,1),rep(10000,2),rep(4000,2),rep(7000,1),
         rep(5000,1),rep(200,4),rep(700,2))

psu<-rep(1:50,psugr).
```

In the psugr vector the PSU size and the number of PSU with this particular size are defined.

The population in all scenarios consists of 50 PSUs and 100,000 SSUs. In the simulation study 10,000 samples of size 15 at the first stage and of size 1000 at the second stage are drawn repeatedly. The sample size at the second stage is allocated proportionally to the size of the PSUs. The variance estimators are computed for each sample to get information about the distribution of the different variance estimators. Thus, reliable statements about the quality of the different methods are possible. The relevant benchmark is the Monte Carlo variance of the point estimator, which is computed as the empirical variance of the 10,000 point estimates. The relative bias is used to compare the quality of the variance estimators in the different scenarios. This measure is defined as follows:

$$\text{rel. Bias} = \frac{\text{Estimated Value} - \text{Benchmark}}{\text{Benchmark}}. \tag{4.1}$$

The sample process is with equal probabilities and with that the total value in the presence of two sampling stages is estimated by:

$$\hat{\tau}_{2St} = \frac{N'}{n'} \cdot \sum_{i=1}^{n'} \sum_{q=1}^{n_i''} \frac{N_i''}{n_i''} \cdot y_{iq} \tag{4.2}$$

(cf. LOHR, 1999, p. 147).

The point estimator in the seven scenarios are given in figure (4.3).

The figure shows that the point estimator is unbiased in all scenarios. The true values of the different scenarios, indicated by the green lines, are equal between all scenarios. This is because after generating these values, they are adjusted such that their total is equal in all scenarios. In scenario 1 the point estimator has the lowest variance in contrast to the other scenarios, as with the variation of the parameters the variance of the point estimators increases. Another effect is the steady rise of the number of outliers.

The results of the different variance estimation methods are given in figure (4.4). The Monte Carlo variance is displayed by the green lines.

The relative bias is shown in figure (4.5).

In scenario 1, when no parameter is changed, the direct estimator (Dir), jackknife (Jack) and the rescaling bootstrap without replacement (RescWoR) lead to nearly unbiased

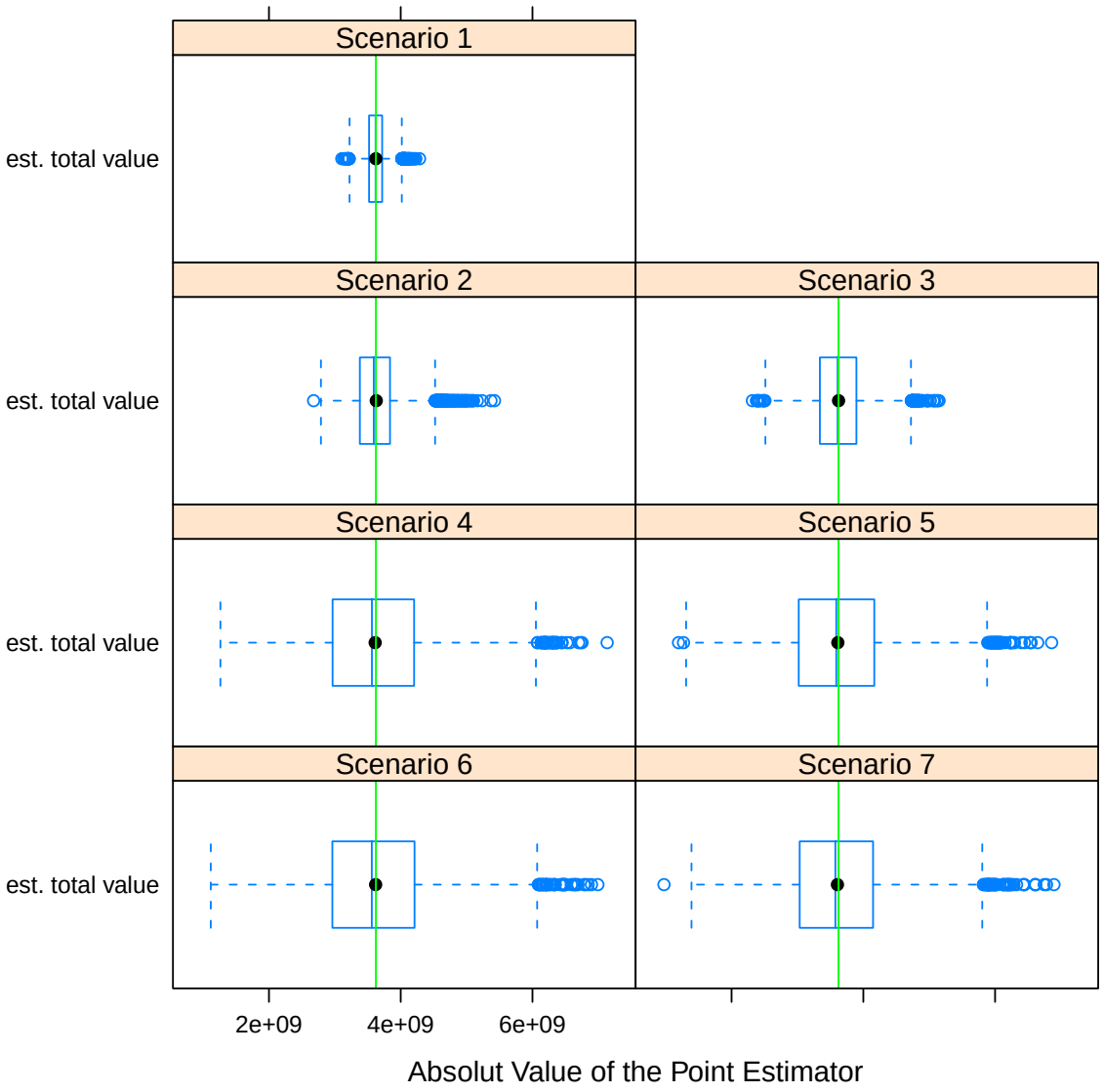

Figure 4.3: Absolute Value of the Point Estimators in the different scenarios

variance estimates. In contrast to this result, the direct estimator and the jackknife based on the first stage (Dir1 and Jack1) lead to a severe underestimation. The same applies for the rescaling bootstrap (Resc) which does not include the second stage unlike the rescaling bootstrap without replacement. This result may be attributed to the large sampling fraction at the first stage and a comparatively large variance at second stage. Here, the consideration of the second stage is necessary. The jackknife method and the direct method lead to same results because the statistic of interest is linear. The Monte Carlo bootstrap (MCBoot) leads to a huge overestimation as a result of the large sampling fraction at the first stage and the missing finite population correction of the approach. This method also underestimates the variance when only the first stage is considered (MCBoot1).

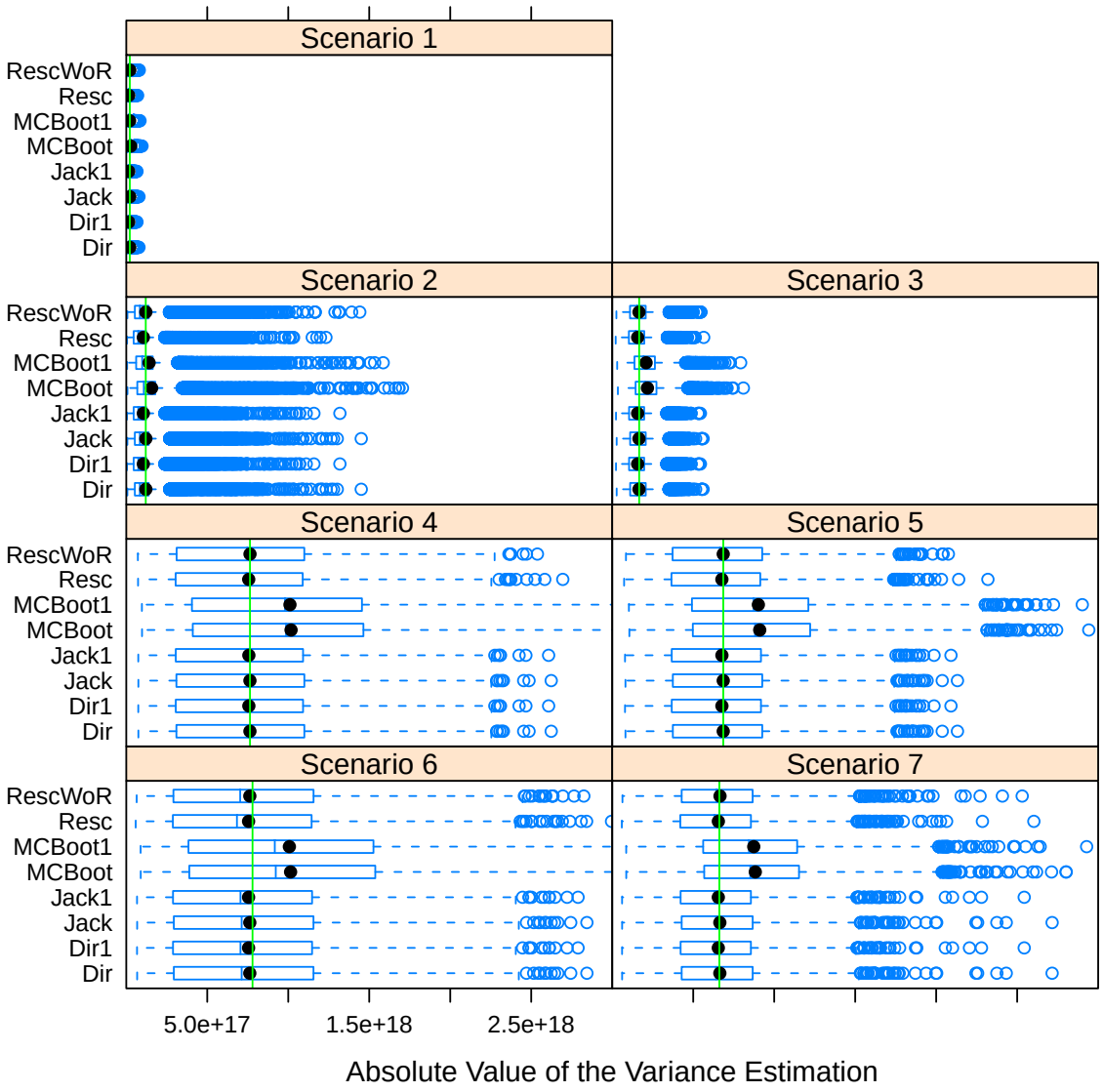

Figure 4.4: Absolute Value of the Variance Estimation Methods in the different scenarios

In scenario 2, the variance within the different PSUs is unequal. Due to high variances in some PSUs (cf. 4.1) the variance at the second stage increases as opposed to scenario 1. For that reason the methods without considering the second stage also show a significant underestimation. But due to the generation process also the variability of the mean of the PSUs rises and with that the variance at first stage opposite scenario 1. Thus, the relative bias of these methods is smaller than in this situation. The methods which consider the second stage lead to unbiased estimates. These methods yield, however, also a lot of outliers through the high variance within some PSUs.

A converse effect appears when only the expected value between the PSUs varies. Hence, in scenario 3, there is only a small difference between the methods which include all stages and the methods which only consider the first stage. This is a result of the varying

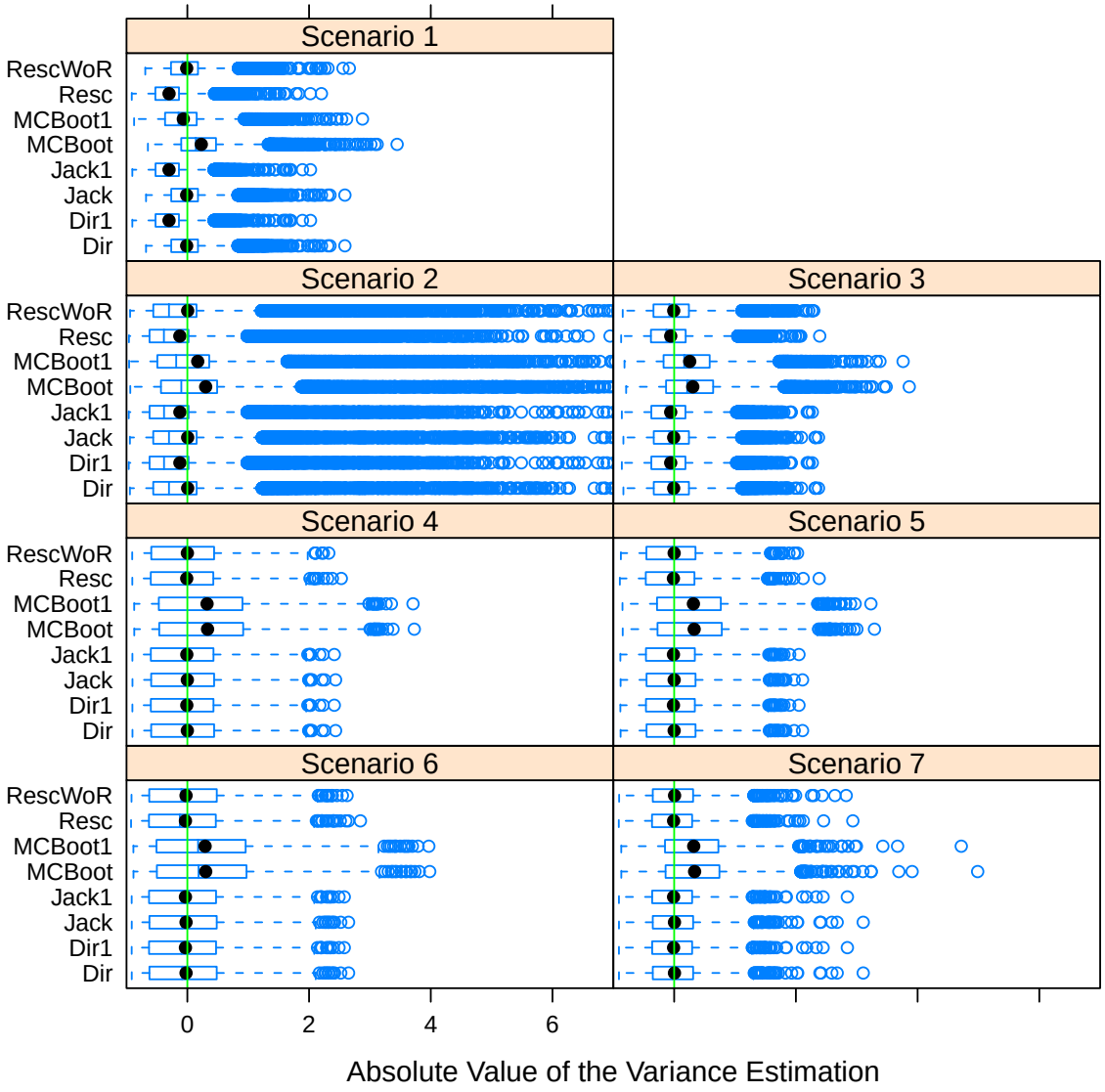

Figure 4.5: Relative Bias of the Variance Estimation Methods in the different scenarios

expected values between the PSUs which also increase the variance at the first stage. Thus, it may be sufficient to estimate only the variance at the first stage. The same applies for scenario 4 with unequal PSU sizes. But here the variance of the variances increases dramatically as a result of the unequal sizes and the large benchmark.

In cases where more than one parameter varies, there is also a huge influence on the variance of variances, especially in scenario 6. In scenario 7 contrary effects arise. The variance at the second stage increases through the varied variance within some PSUs, but also the variance at the first stage due to the unequal expected values between the PSUs.

Figure (4.6) and (4.7) shows the kernel density estimation of the relative bias of the variance estimation methods for the different scenarios. The jackknife is not included

because it yields the same results as the direct estimation methods. For the reason of clarity the Monte Carlo bootstrap, which is only based on the first stage, is also left out. In scenario 1 the distribution of the variance estimation methods is almost symmetric with only a few outliers. For most methods the estimates are lower than the benchmark, which is still represented by the green line. The shape of the distribution of the different methods is quite similar, but the distribution of the methods which include all stages is closer to the benchmark. As seen already in the boxplots above, the number of outliers is increasing with a varying variance between the PSUs and the distribution becomes flatter. The same applies for scenario 3 with unequal expected values between the PSUs. But here the number of outliers decrease and the distribution of the variance estimation methods except the Monte Carlo Bootstrap is quite similar. In scenarios with unequal PSU sizes the distribution of the variance estimators has two local maxima, which are dependent on the drawing of PSUs with large or low sizes. The shape of the distributions of the variance estimators in scenarios where more than one parameter is changed are a combination of these basic shapes. It seems that the effects of changing all parameters offset each other such that the distribution is almost normally distributed.

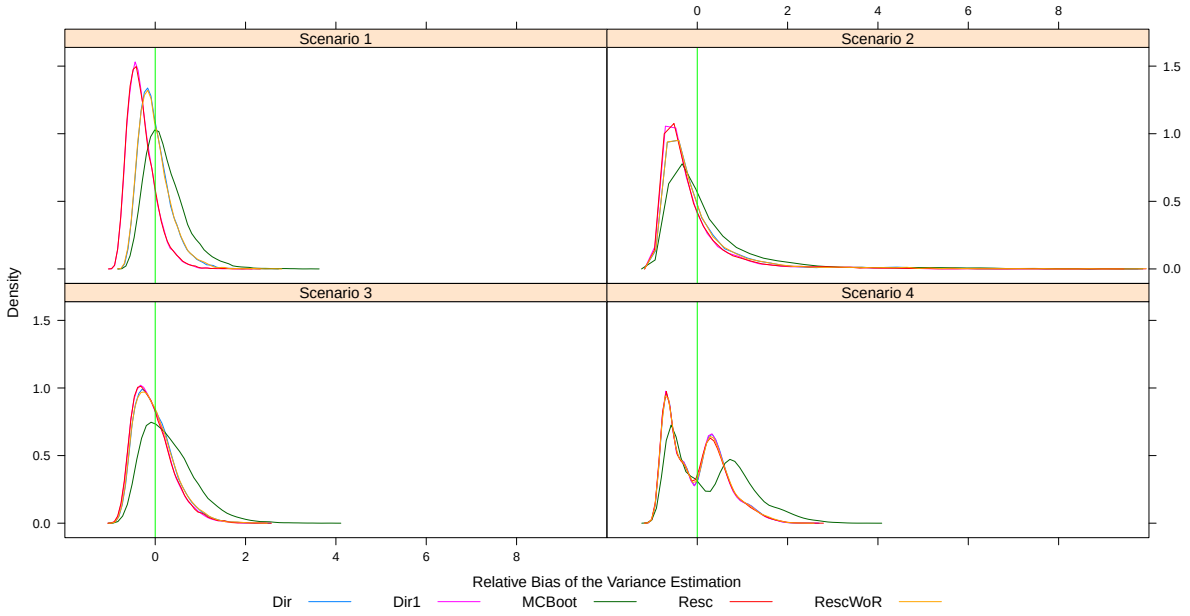

Figure 4.6: Kernel Density Estimation of the Relative Bias Of the Variance Estimation Methods scenario 1-4

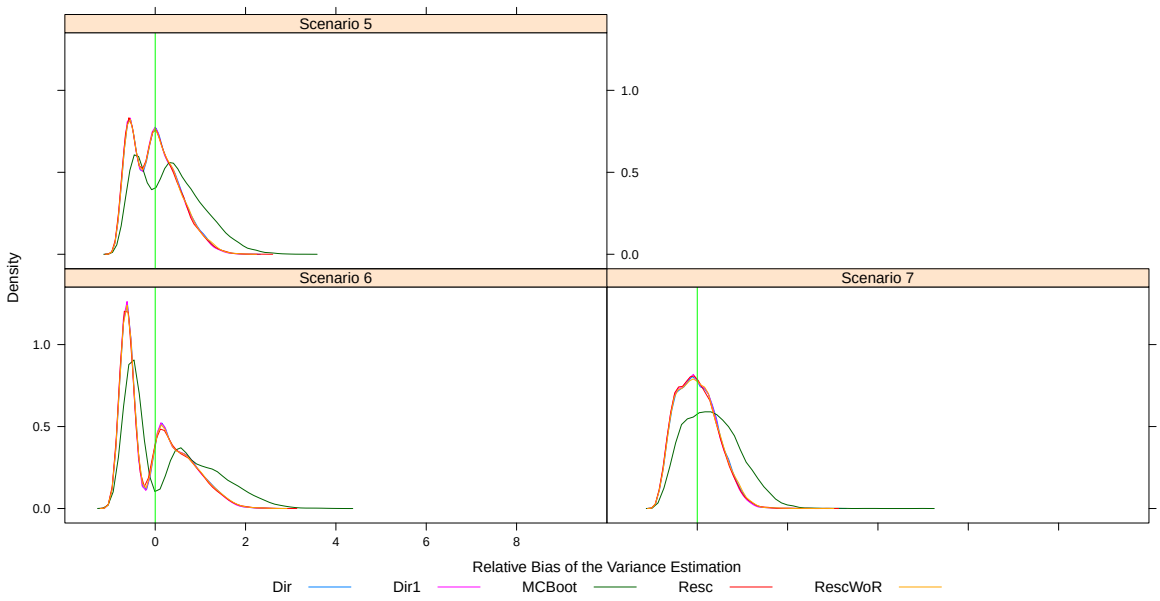

Figure 4.7: Kernel Density Estimation of the Relative Bias Of the Variance Estimation Methods scenario 5-7

Chapter 5

Summary and Outlook

The present work gives an overview of variance estimation in the presence of complex survey designs. The presented instruments overs most features of samples survey designs encountered in practice. However, to present a kind of cookbook approach to variance estimation remains difficult, because for selecting the appropriate variance estimator it also necessary to consider the properties of the statistic of interest. Resampling methods are in general defined independently from the point estimators, which makes them more suitable to set up a best-practice given the sampling design. Most direct variance estimates are developed for linear point estimators only (e.g. estimated totals). For complex estimators often approximate variance estimators were only developed, which is subject to the methodology presented in Münnich and Zins (2011).

Although we treat the problem of complex statistics and survey designs mostly separated from each other, there are of course interactions between them. A complex survey design may render a variance estimator suitable for complex estimator inept, because of its reliance on properties established by simple sampling design, e.g. iid observations of a variable of interest. A synthesis of both methodological branches is made in form of a large scale simulation study based on the AMELIA dataset (see Alfons et al., 2011), whose results are analyzed in deliverable D7.1 (see Hulliger et al., 2011, chapter 9). Also the simulation study in Chapter 4 highlights the problems of variance estimation in case of a multistage sampling design, that are induced through cluster sampling cluster.

The practical relevance of variance estimation gains strongly in importance. This can be seen due to the fact that in 2010, a Directors of Methodology Task Force on accuracy measurement in household surveys was installed in order to elaborate recommendations on the use of the different variance estimation methods. The report is going to be published by Eurostat in 2011.

The deliverable shows a greater treatment of the topic variance estimation in the literature but a wide range of issues is still not or not well covered. An example is the variance estimation of change on the base of rotational panels. Another important application of variance estimation is the area of design effects (for details on design effect see Kish, 1965 and for methods to estimate the design effect see for instance Gabler et al., 1999, Gabler et al., 2006 or Ganninger, 2009) which is entirely based on variance estimation methods.

Bibliography

Alfons, A., **Filzmoser, P.**, **Hulliger, B.**, **Kolb, J.-P.**, **Kraft, S.**, **Münnich, R.** and **Templ, M.** (**2011**): *Synthetic Data Generation of SILC Data*. Research Project Report WP6 – D6.2, FP7-SSH-2007-217322 AMELI.
URL http://ameli.surveystatistics.net

Beaumont, J.-F. and **Bocci, C.** (**2009**): *A practical bootstrap method for testing hypotheses from survey data*. Survey Methodology, 35 (1), pp. 25 – 35.

Berger, Y. and **Skinner, C.** (**2004**): *Variance Estimation for Unequal Probability Designs*. Technical report, DACSEIS deliverable D6.1, http://www.dacseis.de.

Berger, Y. G. (**1998**a): *Rate of convergence for asymptotic variance of the Horvitz-Thompson estimator*. Journal of Statistical Planning and Inference, 74, pp. 149–168.

Berger, Y. G. (**1998**b): *Variance estimation Using List Sequential Scheme for Unequal Probability Sampling*. Journal of Official Statistics, 14 (3), p. 315 – 323.

Berger, Y. G. (**2003**): *Rate of convergence for asymptotic variance of the Horvitz-Thompson estimator*. Statistics in transition, 6 (1), pp. 5–21.

Berger, Y. G. (**2004**a): *A Simple Variance Estimator for Unequal Probability Sampling without Replacement*. Journal of Applied Statistics, 31 (3), pp. 305–315.

Berger, Y. G. (**2004**b): *Variance estimation for measures of change in probability sampling*. The Canadian Journal of Statistics. La Revue Canadienne de Statistique, 32 (4), pp. 451–467.

Berger, Y. G. (**2007**): *A jackknife variance estimator for unistage stratified samples with unequal probabilities*. Biometrika, 94 (4), pp. 953 – 964.

Bienias, J. L., **Kott, P. S.** and **Evans, D. A.** (**2003**): *Applying the delete-a-group jackknife variance estimator to analyses of data from a complex longitudinal survey*. Technical report, Proceedings of the Annual Meeting of the American Statistical Association, Section on Survey Research Methods.
URL http://www.amstat.org/Sections/Srms/Proceedings/y2003/Files/JSM2003-000818.pdf

Brewer, K. (**2003**): Combined Survey Sampling Inference. New York: Arnold.

Brewer, K. and **Donadio, M.** (**2003**): *The High Entropy Variance of the Horvitz-Thompson Estimator*. Survey Methodology, 29 (2), pp. 189 – 196.

Campbell, C. (**1980**): *A different view of finite population estimation*. Proc. Surv. Res. Meth. Sect. Am. Statist. Assoc., pp. 319 – 324.

Chipperfield, J. and **Preston, J.** (**2007**): *Efficient bootstrap for business surveys*. Survey Methodology, 33 (2), pp. 167 – 172.

Cochran, W. G. (**1977**): Sampling Techniques. New York: Wiley.

Davison, A. and **Sardy, S.** (**2007**): *Resampling Variance Estimation in Surveys with Missing Data*. Journal of Official Statistics, 23 (3), pp. 371 – 386.

Davison, A. C. and **Hinkley, D. V.** (**1997**): Bootstrap methods and their application. Cambridge: Cambridge University Press.

Davison, A. C. and **Sardy, S.** (**2004**): *Resampling Methods for Variance Estimation*. Technical report, DACSEIS deliverable D5.1, http://www.dacseis.de.

Deville, J.-C. (**1999**): *Variance estimation for complex statistics and estimators: Linearization and residual techniques*. Survey Methodology, 25 (2), pp. 193– 203.

Deville, J.-C. and **Tillé, Y.** (**2005**): *Variance approximation under balanced sampling*. Journal of Statistical Planning and Inference, 18 (2), pp. 569 – 591, ISSN 0378-3758.

Efron, B. (**1979**): *Bootstrap methods: Another look at the jackknife*. The Annals of Statistics, 7 (1), pp. 1 – 26.

Efron, B. (**1982**): The jackknife, the bootstrap and other resampling plans. Number 38 in CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia: Society for Industrial and applied mathematics.

Efron, B. (**1992**): *Jackknife-after-bootstrap standard errors and influence functions*. Journal of the Royal Statistical Society, Series B, 54 (1), pp. 83 – 127.

Efron, B. and **Tibshirani, R. J.** (**1993**): An introduction to the bootstrap. Monographs on Statistics and Applied Probability 57, New York: Chapman & Hall.

Fay, R. E. (**1989**): *Theory and Application of Replicate Weighting for Variance Calculations*. Survey Research Methods Section, ASA, p. 212 – 217, American Statistical Association.

Funaoka, F., **Saigo, H.**, **Sitter, R. R.** and **Toida, T.** (**2006**): *Bernoulli bootstrap for stratified multistage sampling*. Survey Methodology, 32 (2), pp. 151 – 156.

Gabler, S. (**1984**): *On unequal probability sampling: sufficient conditions for the superiority of sampling without replacement*. Biometrika, 71 (1), pp. 171–175.

Gabler, S. (**1990**): Minimax solutions in sampling from finite populations, *Lecture Notes in Statistics*, vol. 64. New York: Springer-Verlag, vi+132 pp.

Gabler, S., **Häder, S.** and **Lahiri, P.** (**1999**): *A model based justification of Kish's formula for design effects for weighting and clustering*. Survey Methodology, 25 (1), pp. 105 – 106.

Gabler, S., **Häder, S.** and **Lynn, P.** (**2006**): *Design effects for multiple design surveys.* Survey Methodology, 32 (1), pp. 115 – 120.

Ganninger, M. (**2009**): Design Effects: Model-based versus Design-based Approach. Number 3 GESIS-Schriftenreihe, Bonn: GESIS.

Hartley, H. O. and **Rao, J. N. K.** (**1962**): *Sampling with unequal probabilities and without replacement.* Ann. Math. Statist., 33, pp. 350 – 374.

Hájek, J. (**1964**): *Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population.* The Annals of Mathematical Statistics, 35 (4), pp. 1491–1523.

Hájek, J. (**1981**): Sampling from a finite population. New York: Dekker.

Hulliger, B., **Alfons, A.**, **Bruch, C.**, **Filzmoser, P.**, **Graf, M.**, **Kolb, J.-P.**, **Lehtonen, R.**, **Lussmann, D.**, **Meraner, A.**, **Münnich, R.**, **Nedyalkova, D.**, **Schoch, T.**, **Templ, M.**, **Valaste, M.**, **Veijanen, A.** and **Zins, S.** (**2011**): *Report on the Simulation Results.* Research Project Report WP7 – D7.1, FP7-SSH-2007-217322 AMELI.
URL http://ameli.surveystatistics.net

John, P. W. M. (**1971**): Statistical Design and analysis of experiments. New York: Macmillan.

Kish, L. (**1965**): Survey Sampling. John Wiley & Sons.

Kott, P. S. (**2001**a): *The delete-a-group jackknife.* Journal of Official Statistics, 17 (4), pp. 521 – 526.

Kott, P. S. (**2001**b): *Using the delete-a-group jackknife variance estimator in NASS surveys.* Technical report, NASS Research Report 98-01 (revised July 2001).

Kovacevic, M. S. and **Yung, W.** (**1997**): *Variance Estimation for Measures of Income Inequality.* Survey Methodology, 23 (1), pp. 41 – 52.

Lohr, S. (**1999**): Sampling: Design and Analysis. Pacific Grove: Duxbury Press.

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 (4), pp. 543 – 570.

McCarthy, P. J. (**1966**): *Replication: An Approach to the Analysis Data from Complex Surveys.* Technical report, Public Health Service Publication 14, National Center for Health Statistics, Public Health Service, Washington D. C.: U.S. Government Printing Office, series 2.

McCarthy, P. J. (**1969**): *Pseudo-Replication: Half Samples.* Review of the International Statistical Institute, 37 (3), pp. 239 – 264.

McCarthy, P. J. and **Snowden, C. B.** (**1985**): The bootstrap and finite population sampling. Data Evaluation and Methods Research. Series 2, no. 95, Vital and health statistics, National Center for Health Statistics.

Münnich, R. (**2005**): *Datenqualität in komplexen Stichprobenerhebungen*, unpublished thesis, university Tübingen.

Münnich, R. (**2008**): *Varianzschätzung in komplexen Erhebungen*. Austrian Journal of Statistics, 37 (3&4), pp. 319 – 334.

Münnich, R. and **Zins, S.** (**2011**): *Variance Estimation for Indicators of Poverty and Social Exclusion*. Research Project Report WP3 – D3.2, FP7-SSH-2007-217322 AMELI. URL http://ameli.surveystatistics.net

Preston, J. (**2009**): *Rescaled bootstrap for stratified multistage sampling*. Survey Methodology, 35 (2), pp. 227 – 234.

Quenouille, M. H. (**1949**): *Approximate tests of correlation in time-series*. Journal of the Royal Statistical Society, Series B, 11 (1), pp. 68 – 84.

Quenouille, M. H. (**1956**): *Notes on bias in estimation*. Biometrika, 43 (3 - 4), pp. 353 – 360.

R Development Core Team (**2010**): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL http://www.R-project.org/

Rao, J. N. K., **Hartley, H. O.** and **Cochran, W. G.** (**1962**): *On a simple procedure of unequal probability sampling without replacement*. J. Roy. Statist. Soc. Ser. B, 24, pp. 482–491.

Rao, J. N. K. and **Shao, J.** (**1996**): *On balanced half-sample variance estimation in stratified random sampling*. Journal of the American Statistical Association, 91 (433), pp. 343 – 348.

Rao, J. N. K. and **Shao, J.** (**1999**): *Modified balanced repeated replication for complex survey data*. Biometrika, 86 (2), pp. 403 – 415.

Rao, J. N. K. and **Wu, C. F. J.** (**1988**): *Resampling inference with complex survey data*. Journal of the American Statistical Association, 83 (401), pp. 231 – 241.

Rao, J. N. K., **Wu, C. F. J.** and **Yue, K.** (**1992**): *Some recent work on resampling methods for complex surveys*. Survey Methodology, 18 (2), pp. 209 – 217.

Rosén, B. (**1997**): *On sampling with probability proportional to size*. J. Statist. Plann. Inference, 62 (2), pp. 159–191, ISSN 0378-3758, doi:10.1016/S0378-3758(96)00186-3.
URL http://dx.doi.org/10.1016/S0378-3758(96)00186-3

Rust, K. (**1985**): *Variance estimation for complex estimators in sample surveys*. Journal of Official Statistics, 1 (4), pp. 381 – 397.

Rust, K. and **Rao, J. N. K.** (**1996**): *Variance estimation for complex surveys using replication techniques*. Statistical Methods in Medical Research, 5, pp. 283 – 310.

Sampford, M. R. (**1967**): *On Sampling Without Replacement with Unequal Probabilities of Selection*. Biometrika, 54 (3/4), pp. 499–513, ISSN 00063444.

Shao, J. (**2003**): Mathematical Statistics. New York: Springer, second ed.

Shao, J., **Chen, Y.** and **Chen, Y.** (**1998**): *Balanced repeated replication for stratified multistage survey data under imputation.* Journal of the American Statistical Association, 93 (442), pp. 819 – 831.

Shao, J. and **Tu, D.** (**1995**): The jackknife and bootstrap. New York: Springer.

Shao, J. and **Wu, C. F. J.** (**1989**): *A general theory for jackknife variance estimation.* The Annals of Statistics, 17 (3), pp. 1176 – 1197.

Sitter, R. R. (**1992a**): *A resampling procedure for complex survey data.* Journal of the American Statistical Association, 87, pp. 755 – 765.

Sitter, R. R. (**1992b**): *Comparing three bootstrap methods for survey data.* The Canadian Journal of Statistics, 20 (2), pp. 135 – 154.

Särndal, C.-E., **Swensson, B.** and **Wretman, J.** (**1992**): Model Assisted Survey Sampling. New York: Springer-Verlag.

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

Tillé, Y. (**2006**): Sampling algorithms. Springer Series in Statistics, New York: Springer.

Tukey, J. W. (**1958**): *Bias and confidence in not quite large samples (Abstract).* Annals of Mathematical Statistics, 29, p. 614.

Valliant, R. (**1987**): *Some prediction properties of balanced half-sample variance estimators in single-stage sampling.* Journal of the Royal Statistical Society. Series B (Methodological), 49 (1), pp. 68–81.

Wolter, K. M. (**2007**): Introduction to variance estimation. New York: Springer.

Yates, F. and **Grundy, P. M.** (**1953**): *Selection Without Replacement from Within Strata with Probability Proportional to Size.* Journal of the Royal Statistical Society. Series B (Methodological), 15 (2), pp. 253–261, ISSN 00359246.