Simulation studies of Repeated weighting

Discussion paper 03008

Coen van Duin and Vincent Snijders

Statistics Netherlands

Voorburg/Heerlen, october 2003

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2002–2003	= 2002 to 2003 inclusive
2002/2003	= average of 2002 up to and including 2003
2002/'03	= crop year, financial year, school year etc. beginning in 2002 and ending in 2003

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Prinses Beatrixlaan 428
2273 XZ Voorburg
The Netherlands

Printed by
Statistics Netherlands - Facility Services

Cover design
WAT ontwerpers, Utrecht

Information
E-mail: infoservice@cbs.nl

Where to order
E-mail: verkoop@cbs.nl

Internet
http://www.cbs.nl

ISSN: 1572-0314
Key figure: X-10
Production code: 6008303008

Statistics Netherlands

Simulation studies of repeated weighting

Summary: Repeated weighting provides a method to obtain sets of table estimates with numerically consistent margins from combinations of registers and surveys. It is based on repeated application of the regression estimator and generates a new set of weights for each table which is estimated. Repeated weighting is implemented in the prototype software package VRD. This report describes the results of five simulations in which various aspects of repeated weighting were tested. The differences in accuracy between the repeated weighting- and the standard regression estimator were found to be small. When correctly implemented, repeated weighting consistently yielded a smaller standard deviation. In certain cases, a very limited increase in bias compared to standard weighting was found. The VRD estimator for the variance was found to be reliable only for cells of sufficient size and with a low enough variance of the weights.

Keywords: consistent estimates, combining registers and surveys, regression estimator, simulations, variance estimation

1. Introduction

Recently, a method has been developed at Statistics Netherlands to obtain numerically consistent estimates of tables from a combination of registers and surveys; see Kroese and Renssen (1999) and Houbiers *et al.* (2003). This method, repeated weighting (RW), is based on repeated application of the regression estimator. It produces a new set of calibration weights for each table estimate, where the calibration restrictions depend on the table margins that are already estimated. In addition to improving consistency, this method may also produce estimates that are more accurate than those obtained with a standard regression estimator, since it makes wider use of auxiliary information from other surveys. Repeated weighting is implemented in a prototype software package called VRD, see Houbiers and Snijders (2002). VRD estimates both the tables and their variances. The RW estimator and the variance formula are discussed in Knottnerus (2003 A).

This report describes simulations that were carried out to investigate the properties of the RW estimator. The aim was to compare the accuracy of the RW estimator with that of a standard regression estimator and to test the variances calculated by VRD. Five separate simulations were performed. In each case, data was used from a register and from two independently drawn surveys. The first simulation mimics a situation where one of the two surveys is very large and the two surveys have a sizeable overlap. The other four simulations focus on situations where there are two relatively small, non-overlapping surveys.

Section 2 provides an introduction to repeated weighting. The general simulation procedure is explained in section 3. Sections 4 and 5 describe specifics for the two types of simulations. Section 6 discusses the results, which are presented in detail in the appendix.

2. Repeated weighting

This section gives a short introduction to repeated weighting. It mainly focuses on those elements important to the simulations.

2.1 Combining registers and surveys

Repeated weighting was developed for situations where information on one or several variables is available from more than a single data source. A basic example is shown in figure 1a. There are two surveys with non-overlapping samples s_1 and s_2, drawn from the same population[1]. The first survey provides information on the weekly working hours H and the monthly wage W of the sampled persons. The second includes these variables and, in addition, the education level E. H and E are classification variables, while W is a quantitative variable. The gender G of each person in the population is known from a register.

From this set of microdata, three rectangular "blocks" can be created. A block is a subset of data in which for each sampling unit the same variables are available. These blocks are shown schematically in figure 1b. They are numbered consecutively with decreasing size. Block 1 is the register, which contains data on the entire population. Only G is available for this block. Block 2 contains the union of s_1 and s_2. For these sampling units, G, H and W are available. Block 4 contains the units in s_2 and the variables G, H, W, and E.

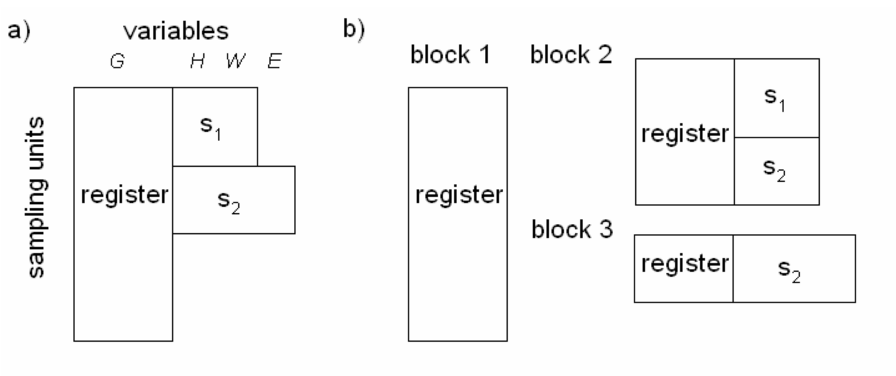

Figure 1 a) Schematic representation of a microdata set for which repeated weighting can be used. b) The blocks that can be created from this dataset.

---

[1] In this example, we assume for simplicity that the two samples are strictly non-overlapping. In practice, if one has two independently drawn samples that are both much smaller than the

These blocks represent the largest subsets of data from which tables can be estimated. In order to reduce the variance of the table estimates, each table is obtained from the largest suitable data block. The frequency table [G x H] x 1, for instance, is estimated from block 2, though one could also estimate it from block 3. The table [G x E] x 1, on the other hand, can only be obtained from block 3.

2.2 Standard weighting

When standard weighting (SW) is used, tables are estimated using a single set of weights per block. These weights represent the starting point for repeated weighting and are therefore called *starting weights*. We denote them by $d_i^{b_k}$, where b_k indicates block k. Typically, the starting weights correct for non-response and unequal inclusion probabilities and have been calibrated on register variables to adjust for sampling fluctuations. In general, even variables from other blocks may have been included in their weighting scheme. However, this is not currently common practice at Statistics Netherlands and we will not consider that possibility here.

The starting weights for the register are of course equal to one. For blocks which contain data from the register plus a single survey s_k, the starting weights are the survey weights $d_i^{s_k}$, possibly calibrated on one or more register variables. In the above example, the starting weights could include a calibration on G. For block 3, the weights would then be given by

$$d_i^{b_3} = d_i^{s_2} \phi_i (G; s_2), \qquad (1)$$

where $\phi_i (G, s_k)$ denotes a correction weight.

The survey weights $d_i^{s_2}$ which appear in (1) consist of an inclusion weight and a correction factor for non-response. Also, certain calibrations on register variables may already be included in these weights. We are assuming that the survey weights have at least been calibrated on the population total N.

Block 2 contains elements from both surveys. Its starting weights are obtained in two steps. First, the survey weights are rescaled by a sample-dependent factor to ensure that the block weights reproduce the population total

$$d_i^{b_2(0)} = \begin{cases} \lambda_1 d_i^{s_1}, \text{for } i \in s_1 \\ \lambda_2 d_i^{s_2}, \text{for } i \in s_2 \end{cases}, \qquad (2)$$

where the λ's are positive factors which satisfy the constraint

$$\lambda_1 + \lambda_2 = 1. \qquad (3)$$

---

population, the overlap between these two samples will be very small and one can still follow the same procedures as in this example.

This constraint ensures that the zeroth order block weights $d_i^{b_k(0)}$ yield the population size when summed over the block, assuming that the survey weights do so when summed over their samples.

The factors λ should be chosen in such a way that the variance of estimators from block 2 is minimized. Assuming independent and non-overlapping samples s_1 and s_2 in figure 1, we can write

$$\begin{aligned} \mathrm{Var}(\hat{y}_{b_2}) &= \mathrm{Var}(\lambda_1 \hat{y}_{s_1} + \lambda_2 \hat{y}_{s_2}) \\ &= \lambda_1^2 \,\mathrm{Var}(\hat{y}_{s_1}) + \lambda_2^2 \mathrm{Var}(\hat{y}_{s_2}) \\ &= (\lambda_1^2 + (1-\lambda_1)^2 r)\mathrm{Var}(\hat{y}_{s_1}), \end{aligned} \tag{4}$$

where $\hat{y}_{b_k}$ denotes the estimator of some quantity Y from block b_k and $\hat{y}_{s_\mathrm{k}}$ the estimator of the same quantity from sample s_k. On the last line, r indicates the ratio of the variances from s_2 and s_1. Minimizing this expression with respect to λ_1, we obtain

$$\lambda_1 = \frac{r}{r+1} . \tag{5}$$

For simple random sampling, the ratio of the variances is the inverse ratio of the sample sizes: $r = V_2 / V_1 = n_1 / n_2$. This yields the factors

$$\lambda_1 = \frac{n_1}{n_1+n_2}, \lambda_2 = \frac{n_2}{n_1+n_2} \quad \text{(simple random sampling).} \tag{6}$$

If a non-trivial sampling scheme is used, the correlation between the variable and the survey weights starts to play a role and r will in general be different for different estimators. However, Kish (1992) argues that it is often a good approximation to assume that variables and weights are uncorrelated. In that case, the variance scales with the inverse of an effective sample size $n_{\mathrm{eff.}}$, which is given by

$$n_{\mathrm{eff.}} = \frac{n}{1+L}, \tag{7}$$

where L is the square of the coefficient of variation of the survey weights.

$$L = \frac{\frac{1}{n}\sum_{i\in s}(d_i^s)^2 - \left(\frac{1}{n}\sum_{i\in s} d_i^s\right)^2}{\left(\frac{1}{n}\sum_{i\in s} d_i^s\right)^2} . \tag{8}$$

Because L is positive, $n_{\mathrm{eff.}} \leq n$. The sample size is therefore effectively reduced by the variability of the weights. Using (7), we obtain the same expression for the λ's as for s.r.s. (6), but with n_k replaced by $n_{\mathrm{eff.},k}$.

After generating the zeroth order block weights (2) in this way, the starting weights are obtained by calibration on register variables, in this case G.

$$d_i^{b_2} = d_i^{b_2(0)}\phi_i(G,b_2)\,.$$

2.3 Table estimation

Tables estimated with starting weights may suffer from numerically inconsistent margins. This problem can arise when two tables are obtained from a different block. Consider the frequency tables [G x H] x 1 and [H x E] x 1. The first is estimated from block 2, the second from block 3. Since they are estimated from different sets of data, using weights that are designed only to reproduce the marginal table [G] x 1, they will likely have a different marginal table [H] x 1.

This problem is solved by repeated weighting. The basic principle is to recalibrate the weights with which a table is estimated on the margins it has in common with tables estimated earlier. This may result in different sets of weights for different tables estimated from the same block.

Table estimates obtained from such a procedure depend on the set of tables which is estimated and on the order in which the estimates are performed. These factors therefore need to be fixed in some way. In the simulation study, two procedures for achieving this are compared: minimal repeated weighting and splitting up.

2.4 Minimal repeated weighting

2.4.1 Frequency tables

We first discuss the minimal repeated weighting procedure for the case that only frequency tables are estimated.

The first step is to specify the full set of output (frequency) tables one wants to estimate. Next, certain marginal tables of these output tables are added to this set. There are two cases in which a marginal table is added:

- it can be estimated from a larger block than the "parent" table,
- it is obtained from the same block as the parent table, but it can be consistently estimated with starting weights.

A table can always be consistently estimated with starting weights if all its marginal tables are obtained from the same block as the table itself. It can still be consistently estimated with starting weights if one or more of these marginal tables are obtained from another (larger) block, provided that these are tables that have been included in the calibration scheme of the starting weights. Since we are only considering cases where the starting weights contain calibrations on register tables, these marginal tables will have to be obtained from the register.

After the marginal tables have been added to the set, the process is repeated for *their* marginal tables. This continues until no more tables need to be added according to the above criteria. The set is subsequently ordered (the procedure for which is described below) and the tables are estimated one after the other. Each table is calibrated on the margins it has in common with tables that have been estimated before.

The set is ordered first by increasing block number, second by increasing dimensionality. Within groups of tables from the same block, tables that can be consistently estimated with starting weights are moved to the front. This procedure does not always fully specify the order of the estimates and different choices can lead to (slightly) different results. This is a specific problem of the minimal repeated weighting procedure, which we come back to in section 2.4.4.

The reweighting process ensures that all shared margins of the tables in the set are estimated consistently. Furthermore, the variance of the table estimates may be improved compared to standard weighting, since margins that are estimated from larger data blocks are used as calibration restrictions, and these margins have a positive correlation with the table interior. An example of minimal repeated weighting is discussed in section 2.4.3.

2.4.2 Quantitative tables

When quantitative tables are estimated, an extra consistency requirement comes into play.

Suppose we want to estimate the quantitative tables $[H \times E] \times M$ and $[G \times H] \times M$ and the frequency table $[H \times E] \times 1$. Consistency is required between the two quantitative tables, because they share the marginal table $[H\ ] \times M$. In addition, consistency is required between $[H \times E] \times M$ and $[H \times E] \times 1$. The reason is that the ratio of these two tables yields an estimate of the average monthly wage by working hours and education level. For this estimate to be accurate, the weights with which the quantitative table is obtained should correctly reproduce the total number of elements per table cell. Especially when the quantitative table and the underlying frequency table are not obtained from the same block, unreliable averages may result if the quantitative table is not calibrated on the frequency table.

This second consistency requirement is met by adding to the set the underlying frequency tables of all quantitative tables. Subsequently, the marginal tables are added to this set as discussed before. The tables are ordered, with the additional restriction that a quantitative table must always be estimated *after* its frequency table, which ensures that the quantitative table will be calibrated on its underlying frequency table. This restriction does not violate the principle that tables from larger blocks are estimated first, since the frequency table is always obtained from the same block as the quantitative table or from a larger one.

2.4.3 Example

Suppose we want to estimate the tables $[H \times E] \times W$ and $[H \times E] \times 1$. With minimal reweighting, the following table estimates are generated

Table	Block	Recalibrations
W	2	starting weights (no recalibrations)
$[H] \times 1$	2	starting weights
$[H] \times W$	2	$[H] \times 1 + W$*
$[E] \times 1$	3	starting weights
$[H \times E] \times 1$	3	$[H] \times 1 + [E] \times 1$
$[H \times E] \times W$	3	$[H \times E] \times 1 + [H] \times W$

The asterisk in the calibrations column indicates that the calibration of $[H] \times W$ on $[H] \times 1 + W$ has no effect. This is because $[H] \times 1$ and W are estimated with starting weights from the same block as $[H] \times W$ and, consequently, the three tables are already consistent. Note that the marginals $[H] \times 1$ and $[H] \times W$ were included in the set because they can be obtained from a larger block than their parent tables, while W and $[E] \times 1$ were included because they can be estimated with starting weights.[2]

2.4.4 Drawbacks of minimal repeated weighting

There are two disadvantages related to minimal repeated weighting. One was mentioned earlier: the order in which the tables are estimated is not always unambiguously determined and a different order will yield a slightly different result. Another disadvantage is that the outcome of a table estimate depends on what other output tables one is estimating at the same time. Suppose that, in the previous example, we would not have been interested only in $[H \times E] \times W$ and $[H \times E] \times 1$, but also in $[E] \times W$. Using the same minimal repeated weighting procedure, we would have arrived at the following estimates

Table	Block	Recalibrations
W	2	starting weights
$[H] \times 1$	2	starting weights
$[H] \times W$	2	$[H] \times 1 + W$ *
$[E] \times 1$	3	starting weights
$[H \times E] \times 1$	3	$[H] \times 1 + [E] \times 1$
$[E] \times W$	3	$[E] \times 1 + W$
$[H \times E] \times W$	3	$[H \times E] \times 1 + [H] \times W + [E] \times W$

The table $[H \times E] \times W$ now has a different weighting scheme than in the earlier example. Apparently, estimates of tables can change when a new output table is

---

[2] The table estimate W is in fact redundant, in the sense that exactly the same estimates for the other tables are obtained whether it is included or not, see section 4.2.2.

added. For table estimates of sufficient accuracy, this change should be very small, but it might still be considered a problem.

2.5 Splitting up procedure

The splitting up procedure avoids these problems by introducing *all* marginal tables into the set, regardless of how these tables are estimated. In this way, table weights are always calibrated on all possible margins. As a result, different choices for ordering the tables can no longer result in different calibration schemes for the table weights (provided that these choices comply with the rules for ordering the tables that were mentioned earlier). Also, the estimate for an output table has become independent of the other output tables one chooses to estimate at the same time.

This is easily checked for the above example. One indeed finds that, if splitting up is used, the set of estimates becomes the one shown in section 2.4.4, also when [*E*] x *W* is not specified as output table.

A drawback of the splitting up procedure is that it requires, in general, more table estimates and more calibrations than minimal repeated weighting. This makes it less economical. For table cells with a small size, there is also the risk of a reduced accuracy of the estimates compared to standard weighting. The calibrations, which are implemented using the regression estimator, make use of estimated regression coefficients. If many of these coefficients are estimated based on a small sample, their inaccuracies, though relatively small individually, may collectively result in a less stable estimator than if no calibrations were performed. Moreover, splitting up introduces additional calibrations on margins that are estimated from the same block as the table itself, albeit with a different calibration scheme, and these calibrations do not necessarily increase the accuracy of the table estimate.

2.6 Variance estimation

Under certain conditions, VRD can provide an estimate for the variance of tables obtained with repeated weighting. The details of the formula implemented in VRD are explained in Snijders and Houbiers (2002). Here, we simply give the formula and discuss the assumptions on which it is based.

The variance estimator implemented in VRD has the following form

$$\hat{\mathrm{V}}\left(\hat{Y}_{T_\alpha}^{\mathrm{RW}}\right)=\sum_k\left(\frac{n_k}{n_k-1}\sum_{i\in s_k}\left(\frac{z_{T_\alpha,i}^{s_k}}{\pi_i^{s_k}}\right)^2-\frac{1}{n_k-1}\left(\sum_{i\in s_k}\frac{z_{T_\alpha,i}^{s_k}}{\pi_i^{s_k}}\right)^2\right), \tag{9}$$

where $\hat{Y}_{T_\alpha}^{\mathrm{RW}}$ denotes the RW estimator of the variable Y in cell α of table $\boldsymbol{T}$, k labels the samples which are involved in this estimate and $\pi_i^{s_k}$ is the inclusion probability for element i in sample s_k. The residuals $z_{T^\alpha,i}^{s_k}$ are given by

$$
\begin{aligned}
z_{T_\alpha,i}^{s_k} &= \sum_{t\in \text{ set } \boldsymbol{T}} \sum_{\gamma} M_{t_\gamma}^{s_k,T_\alpha} y_i^{t_\alpha} \\
&= \lambda_k y_i^{T_\alpha} + \sum_{\gamma} M_{t_{1,\gamma}}^{s_k,T_\alpha} y_i^{t_{1,\gamma}} + \sum_{\gamma} M_{t_{2,\gamma}}^{s_k,T_\alpha} y_i^{t_{2,\gamma}} + \dots,
\end{aligned} \tag{10}
$$

where "set ***T***" contains ***T*** and all tables which are used as calibration restrictions for the estimate of table ***T***. These tables $\boldsymbol{t}_m \neq \boldsymbol{T}$ may appear in the calibration scheme of the survey weights $d_i^{s_k}$ or of the starting weights $d_i^{b_k}$, or they may be tables that are estimated before ***T*** in the repeated weighting procedure and with which ***T*** has a margin in common. The index γ refers to the cells of the tables $\boldsymbol{t}_m$. The matrix elements $M_{t_\gamma}^{s_k,T_\alpha}$ contain the paramaters λ_k and estimated regression coefficients that are used in the calibrations. For frequency tables, the variables $y_i^{t_\gamma}$ have a value of one if element i is part of the subpopulation defined by the table cell t^γ and zero otherwise. For quantitative tables, $y_i^{t_\gamma}$ is the score of the quantitative variable for element i if the element is in the table cell and zero otherwise.

The following approximations are used in (9)

- The formula assumes a sampling scheme with replacement (Hansen Hurwitz estimator). It is approximately correct for sampling *without* replacement as long as all samples are small compared to the population ($n_k << N$).
- The samples s_k are assumed to have no elements in common.
- The matrix ***M*** is obtained from sample data and therefore has a nonzero variance. This variance is neglected in (9). For simple random sampling, the contribution that is neglected in this way is smaller than (9) by a factor of $1/\sqrt{n_\mathrm{k}}$ and the approximation is therefore asymptotically correct for large n_k. For non-trivial sampling schemes, we expect the approximation to be valid for $n_{\mathrm{eff.},k} >>1$.

It follows from these approximations that (9) can be used in cases where $n_k << N$, $n_{\mathrm{eff.},k} >>1$ and the overlap between samples is much smaller than the sample sizes. It should be noted that the RW estimator can still be used in cases where these restrictions are not satisfied. However, equation (9) is then no longer expected to yield a good estimate of its variance.

An additional approximation is involved in the estimation of variances of ratios of table estimators. Consider the table "average monthly wage by education level", which is given by

$$
[E]\times \mathrm{Av}(W)^\alpha = \frac{[E]\times W^\alpha}{[E]\times 1^\alpha},
$$

where α is a cell index for the table. We denote the estimator of this table by $\hat{t}^{\alpha}_{\mathrm{Av}(W)}$. To obtain its variance, we expand to lowest order in the relative variances of its numerator ($\hat{t}^{\alpha}_{W}$) and denominator ($\hat{t}^{\alpha}_{1}$)

$$\mathrm{Var}\left(\hat{t}^{\alpha}_{\mathrm{Av}(W)}\right) \approx \frac{\mathrm{Var}\left(\hat{t}^{\alpha}_{W}\right)}{\hat{t}^{\alpha\,2}_{1}} + \frac{\hat{t}^{\alpha\,2}_{W}\,\mathrm{Var}\left(\hat{t}^{\alpha}_{1}\right)}{\hat{t}^{\alpha\,4}_{1}} - 2\,\frac{\hat{t}^{\alpha}_{W}\,\mathrm{Cov}\left(\hat{t}^{\alpha}_{1}, \hat{t}^{\alpha}_{W}\right)}{\hat{t}^{\alpha\,3}_{1}}. \qquad (11)$$

The variances in (11) are estimated using equation (9). To obtain the covariance, we estimate the variance of the sum of the two tables using (9) and then use the identity Cov(x,y)=(1/2)(Var(x+y)-Var(x)-Var(y)). Eq. (11) is valid if the variances of the table estimators $\hat{t}^{\alpha}_{W}$ and $\hat{t}^{\alpha}_{1}$ are small compared to the square of their values.

2.7 Structure of Earnings Survey

Repeated weighting is currently used at Statistics Netherlands for the Structure of Earning Survey (SES). To test RW in a realistic setting, one of the simulations is constructed to mimic the SES. We therefore briefly discuss how RW is implemented for this particular survey.

2.7.1 Data set

The SES is concerned with properties of jobs. This survey combines data from three different sources.

- A register containing data on 6.6 million jobs. The simulation uses the classification variables gender (*G*), age (*A*) and business activity (*B*) from this register.
- The Employment and Wages Survey (EWS), which covers a subpopulation of 2.8 million jobs. We use the classification weekly working hours (*H*) and the quantitative variable monthly wage (*W*) from this survey.
- The Labour Force Survey (LFS), whose sample contains approximately a hundred thousand records. From this survey, only the classification variable Education level (*E*) is used.

The structure of the data set is shown in figure 2. There are four data blocks: the first contains all sampling units, the second those in the EWS, the third those in the LFS, the fourth those in the cross section of EWS and LFS. The LFS has an overlap of about 50% with the EWS, so block 4 contains approximately 50 000 units.

2.7.2 Starting weights

The starting weights for blocks 2 and 3 are obtained by the procedure described in section 2.2 for blocks containing data from a single survey. The weights are calibrated on the register variables *G*, *A* and *B*, using the weighting scheme [$G \times A$] $\times$ 1 + [B] $\times$ 1.

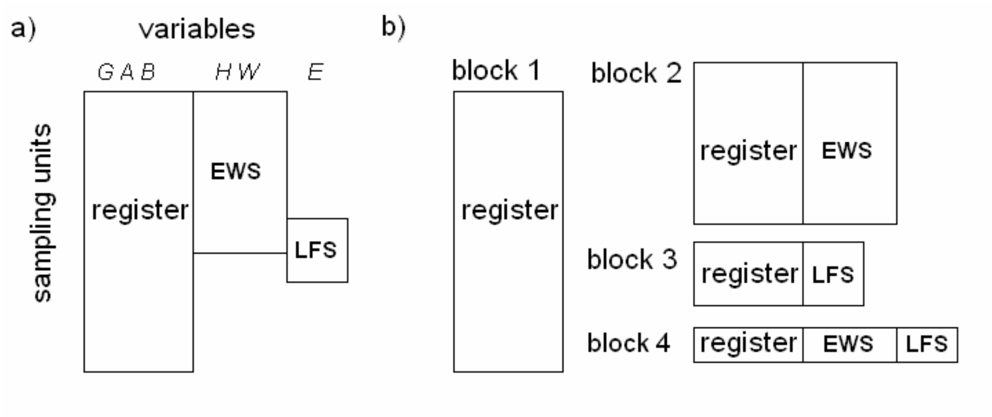

Figure 2:a) Microdata structure of the SES, only the variables that are used in the simulation are shown here, b) blocks used in obtaining estimates from the SES.

For block 4, the situation is different from the one discussed in section 2.2, since one now has a cross section of two samples rather than a union. The approach is in this case to treat block 4 as if it were a third sample. For the survey weights of this sample, the products of the survey weights of EWS and LFS are used

$$d_i^{\mathrm{EWS} \cap \mathrm{LFS}} = d_i^{\mathrm{EWS}} d_i^{\mathrm{LFS}} . \qquad (12)$$

This choice is based on a probabilistic interpretation of the weights. If the EWS/LFS survey weight of an element represents the inverse probability for that element to be in the EWS/LFS sample, then the product of the EWS and LFS survey weights represents the inverse probability for the element to be in *both* of these samples. Since the survey weights include corrections for non-response, this "probability to be in a sample" should be interpreted as an inclusion-and-response probability. In this argument, we neglect the fact that the survey weights also contain corrections for sampling fluctuations, which, strictly speaking, is not consistent with a probabilistic interpretation of the weights. These corrections, however, are typically small.

Unfortunately, the LFS and, especially, the EWS survey weights have a wide distribution. Because of this, the product in eq. (11) results in some "outlier" survey weights for block 4 with very high values. These outliers can lead to unstable estimates. As an ad hoc solution to this problem, a maximum value of 6000 is imposed on the weights. Subsequently, the weights are rescaled by a constant factor to ensure that they correctly reproduce the total population size N. This procedure yields the zeroth order block weights for block 4

$$d_i^{b_4(0)} = \mathrm{Min}(d_i^{\mathrm{EWS} \cap \mathrm{LFS}}, 6000) \frac{N}{\sum_{i \in b_4} \mathrm{Min}(d_i^{\mathrm{EWS} \cap \mathrm{LFS}}, 6000)} . \qquad (13)$$

As was the case for the other blocks, these zeroth order weights are calibrated on [*G* x *A*] x 1 + [*B*] x 1 to produce the starting weights $d_i^{b_4}$.

2.7.3 Variance estimates

The variance formula which was discussed in section 2.6 cannot be directly applied to the SES. Firstly, due to complications involved in incorporating the data from the Labour Force Survey into the Structure of Earnings Survey, the inclusion probabilities $\pi_i^{s_k}$ are not known. The survey weights $d_i^{s_k}$ are known, however, and it was decided to use these in equation (9) instead of the inverse inclusion probabilities. This means that the calibrations which are incorporated in the survey weights are not considered in the variance estimation. Secondly, the condition that the samples s_k should be small compared to the population and have little mutual overlap, under which equation (9) was derived, is not met for the SES. The EWS sample covers nearly half the population and EWS and LFS have a large cross section. In order to overcome this second problem, the following procedure was proposed (Houbiers, 2002).

- Treat all tables that are estimated from block 2 as register tables. The variance of these table estimates is assumed to be zero, since block 2 covers nearly half the population.
- Identify as s_1 the LFS sample.
- Identify as s_2 the cross-section of the LFS and EWS samples. Use the zeroth-order block weights of block 4, eq. (13), as survey weights for s_2.

Note that the samples s_1 and s_2 coincide with blocks 3 and 4. The starting weights for block 3 are derived from the survey weights of s_1, those of block 4 from the survey weights of s_2.

The samples s_1 and s_2 are not independent and the variance estimator therefore does not decompose into a sum of independent contributions from the two samples, as in equation (9). Ignoring this at first, we let VRD compute a first approximation $V^{(0)}$ to the variance using this equation anyway. In the process, the residuals $z_{T^\alpha,i}^{s_k}$ are calculated by the program. It is shown in (Houbiers, 2002) that one can generate an improved variance estimate $V^{(1)}$ which does properly take into account the overlap between the two samples using these residuals. This is achieved by removing from s_1 those units which are also in s_2 and by performing the following transformation on the residuals in s_2

$$z_i^{s_2} \rightarrow z_i^{s_2} + \frac{d_i^{s_1}}{d_i^{s_2}} z_i^{s_1} \quad (i \in s_2).$$

The improved estimate $V^{(1)}$ is obtained by using these new residuals and the reduced sample s_1 in (9).

In the simulation which mimics the SES, both $V^{(0)}$ and $V^{(1)}$ were calculated for the target table [G x H x E] x Av(W). It was found that the relative difference between the two variances was at most of the order 10^{-5}. Recently, it has been pointed out that the fact the $V^{(0)}$ and $V^{(1)}$ are nearly identical can be explained following a line of argument set out in (Knottnerus, 2001). Given the extremely small difference between the two variance estimates, only the results for $V^{(0)}$ are presented in this report. It is however important to stress that under different circumstances than the ones in the SES, for instance if more surveys were included in the data set, or if different calibration schemes were used for the different blocks, the overlap between samples could play an important role in the variance, in which case it *would* be necessary to use $V^{(1)}$ rather than $V^{(0}$.

3. The simulations

Five separate simulations were carried out to test the properties of the repeated weighting estimator under different circumstances. The aim was to answer a number of questions regarding repeated weighting.

The first question concerns the bias. Repeated weighting involves repeated application of the regression estimator. Weights for new estimates are calibrated on the outcome of earlier estimates. Since the regression estimator is only asymptotically unbiased, one could worry whether repeated weighting leads to an accumulation of bias. To test this, the bias is estimated from the simulation results both for standard weighting, minimal repeated weighting and splitting up.

The second question concerns the variance of the repeated weighting estimator. Application of the regression estimator results in an improved precision only if the auxiliary variable on which one is calibrating is known with greater accuracy than the estimate of this variable using the uncalibrated weights. This condition is not always met in repeated weighting. Especially when the splitting up procedure is used, one often calibrates tables on marginal tables which are estimated from the same data as the table itself, but with different weights. In those cases, the estimate on which one calibrates could be worse than the one obtained with uncalibrated weights. Furthermore, even if the tables are calibrated on a more accurate table estimate, the variance could still increase if the estimated regression coefficients have large fluctuations, which can happen for small (effective) sample sizes.

To address these concerns, we investigate whether the variances of the repeated weighting estimates, as obtained from the simulation, are significantly lower or higher than those from standard weighting and how their difference depends on the (effective) cell and sample size. Also, we check whether the splitting up procedure has an adverse effect on the variance compared to minimal repeated weighting.

Finally, the simulations offer the possibility to verify the standard deviation estimates that are computed by VRD. As has been discussed in sections 2.6 and

2.7.3, a number of approximations are involved in these estimates. The simulation results are used to check under what conditions these approximations are valid.

3.1 General procedure

The simulation procedure resembles the application of the bootstrap to survey sampling which is discussed in Särndal *et al.* (1992). It determines the variance of a quantity by first creating an artificial population from a real survey sample and then drawing repeatedly samples from this artificial population, estimating the quantity of interest from each of these samples. The variance of the estimates gives an estimatation of the variance of the quantity. In the procedure discussed in Särndal *et al.*, the samples are drawn with the scheme of the survey sample on which the population is based. Here, we also uses schemes other than that of the original sample.

The artificial population is based on an EWS and an LFS sample from the Structure of Earnings Survey (see sections 2.7). From this population, a register is created by selecting a few of its variables. In each simulation run, two samples are drawn from the population. Together with the register, they form a microdata set from which table estimates are made. After 600 runs simulation averages and variances are calculated.

3.2 Artificial population

The artificial population is based on block 2 of the SES data set over the year 2000. Five variables from block 2 are included in the population file: *G, A, B, H* and *W*, as well as the EWS survey weight d_i^{EWS}. To this are added from block 3 the variable *E* and the LFS survey weight d_i^{LFS}.

Since these last two variables are not available for most elements in block 2, they need to be imputed. A value of *E* is imputed for each of the records in block 2 that is not in block 3. Probabilistic imputation is used, where the probability for each value to be imputed is determined from the distribution of *E* by gender, age and ethnicity in the population of jobs (as estimated from the LFS). A value of d_i^{LFS} is imputed for all records in block 2, including those for which d_i^{LFS} is known from the LFS sample. This is done to avoid creating an atypical group of elements in the population, with weights d_i^{LFS} that have different properties from those of the other elements. First, the continuous variable d_i^{LFS} in block 3 is converted into a set-valued variable by replacing each value with the average of the octile to which that value belongs. Next, d_i^{LFS} is imputed for the records in block 2 using the probability distribution of this new variable by gender and ethnicity.[3]

[3] Age is not used, since it only has a weak correlation with d_i^{LFS}.

After these imputations, records in block 2 with starting weights larger than one are duplicated to produce a file which contains as many records as there are jobs in the actual population (6.4 million). Non-integer weights are interpreted probabilistically. Thus, from an EWS record with $d_i^{b_2}$ = 4.2, at least four record are generated in the final population and a fifth with a probability of 0.2. There is a limited number of EWS records for which the starting weight are slightly less than one. Those records each have a probability of $1\text{-}d_i^{b_2}$ to be excluded from the population.

The final population file has 6.4 million records and contains the variables *G, A, B, H, W* and *E* and the survey weights d_i^{LFS} and d_i^{EWS} .

3.3 Simulation estimates and margins

The following quantities are obtained from the simulations.

- The bias $B_{\boldsymbol{T}}$. This is computed for standard weighting (SW; estimation with starting weights), for minimal RW and for splitting up (RW+).
- The percentage of estimates P_T^{VRD} which lie inside the 95% confidence interval obtained from VRD's standard deviation estimator. Because computing standard deviations in VRD is time-consuming, this quantity is only obtained for the RW estimator.
- The standard deviation $\mathrm{S}_{\boldsymbol{T}}$ of the set of table estimates over the different simulation runs. This is also obtained for SW, RW and RW+.

It is briefly discussed how these quantities and their confidence intervals are estimated.

3.3.1 Bias

The bias of a table estimator $\hat{\boldsymbol{T}}$ is estimated by the difference of its simulation average $\hat{\boldsymbol{T}}_{\text{sim.}}$ and the true population value $\boldsymbol{T}$.

$$\hat{B}_T^{\text{sim.}} = \hat{\boldsymbol{T}}_{\text{sim.}} - \boldsymbol{T} = \frac{1}{R}\sum\nolimits_r \hat{\boldsymbol{T}}_r - \boldsymbol{T} \,, \tag{14}$$

where $\hat{\boldsymbol{T}}_{r.}$ denotes the table estimate in run *r* and *R* is the number of simulation runs (*R*=600).

Since $\boldsymbol{T}$ is a population constant, the variance of $\hat{B}_{\boldsymbol{T}}^{\text{sim.}}$ is the same as that of $\hat{\boldsymbol{T}}_{\text{sim.}}$, which is estimated by

$$\hat{\mathrm{S}}^2\left[\hat{\boldsymbol{T}}_{\text{sim.}}\right] = \frac{\hat{\mathrm{S}}_{\boldsymbol{T}}^2}{R} \,, \tag{15}$$

where $S^2_{\boldsymbol{T}}$ is the variance of the table estimator $\hat{\boldsymbol{T}}$. Since $\hat{\boldsymbol{T}}_{\text{sim.}}$ is an average over R independent table estimates, its variance is smaller than that of $\hat{\boldsymbol{T}}$ by a factor R. The variance estimator of the table is given by

$$\hat{S}^2_{\boldsymbol{T}} = \frac{1}{R-1}\sum_r \left(\hat{\boldsymbol{T}}_r - \hat{\boldsymbol{T}}_{\text{sim.}}\right)^2 . \tag{16}$$

Assuming normality, the 95% confidence interval of the bias is given by $\left[\hat{B}^{\text{sim.}}_T - 1.96\,\hat{S}[\hat{B}^{\text{sim.}}_T], \hat{B}^{\text{sim.}}_T + 1.96\,\hat{S}[\hat{B}^{\text{sim.}}_T]\right]$, where the standard deviation of $\hat{B}^{\text{sim.}}_{\boldsymbol{T}}$ is obtained from (15) and (16)

$$\hat{S}\left[\hat{B}^{\text{sim.}}_T\right] = \frac{\hat{S}_T}{\sqrt{R}} = \sqrt{\frac{1}{R(R-1)}\sum_r \left(\hat{\boldsymbol{T}}_r - \hat{\boldsymbol{T}}_{\text{sim.}}\right)^2} . \tag{17}$$

It follows from our definition of the confidence interval that a bias is found to differ significantly from zero if it lies outside the interval $\left[-1.96\,\hat{S}[\hat{B}^{\text{sim.}}_T], 1.96\,\hat{S}[\hat{B}^{\text{sim.}}_T]\right]$, or, equivalently, if its absolute value exceeds 1.96 times its simulation standard deviation. Using (17), we can express this criterion in term of the bias ratio $\hat{B}^{\text{sim.}}_T / \hat{S}_T$: a bias is found to differ significantly from zero if the absolute value of its ratio exceeds $1.96/R^{1/2}=0.08$ (600 simulation runs).

3.3.2 Coverage probability of the VRD confidence interval

The coverage probability is obtained by determining for each table cell the fraction of the simulation runs for which the cell estimate deviates less from the population value than the margin on the estimate which was computed by VRD

$$\hat{P}^{\text{VRD}}_{T^\alpha} = \frac{1}{R}\sum_r \Theta\left(1.96\,\hat{S}^\alpha_{\text{VRD},r} - \left|\hat{T}^\alpha_r - T^\alpha\right|\right), \tag{18}$$

where $\hat{S}^\alpha_{\text{VRD},r}$ is the VRD standard deviation estimate for cell α on run r. The Θ-function is one if its argument is positive, zero if it is negative. $\hat{P}^{\text{VRD}}_{T^\alpha}$ can be interpreted as a simple random sampling estimator of a frequency variable from a sample with R elements. We therefore estimate its standard deviation by

$$\hat{S}\left[\hat{P}^{VRD}_{T^\alpha}\right] = \sqrt{\frac{\hat{P}^{VRD}_{T^\alpha}\left(1-\hat{P}^{VRD}_{T^\alpha}\right)}{R-1}} . \tag{19}$$

3.3.3 Accuracy of simulation variance

Assuming normality, the standard deviation of the variance estimate $\hat{S}^2_{\boldsymbol{T}}$ is given by (see Knottnerus, 2003 B, page 298)

$$\mathrm{S}\left[\hat{\mathrm{S}}_T^2\right] = \mathrm{S}_T^2 \sqrt{\frac{2}{R-1}} \; . \tag{20}$$

The relative standard deviation of $\hat{\mathrm{S}}_T^2$ only depends on the number of runs. Since 600 runs are used in these simulations, S_T^2 is determined with a relative margin of plus or minus 11%. The relative standard deviation of $\hat{\mathrm{S}}_T$ is half that of $\hat{\mathrm{S}}_T^2$. $\mathrm{S}_{\boldsymbol{T}}$ is therefore known with a margin of 6%.

3.3.4 Difference of variances

The variance S_T^2 is estimated by the simulation average of the quantity

$$\Delta_r^2 = \frac{R}{R-1}\left(\hat{\boldsymbol{T}}_r - \hat{\boldsymbol{T}}_{\text{sim.}}\right)^2 , \tag{21}$$

see eq. (16). In order to test whether two weighting methods have different variances, we check whether the quantity

$$\hat{\mathrm{D}}_{1,2} = \hat{\mathrm{S}}_T^{(1)^2} - \hat{\mathrm{S}}_T^{(2)^2} = \frac{1}{R}\sum_r \left(\Delta_r^{(1)^2} - \Delta_r^{(2)^2}\right) \tag{22}$$

differs significantly from zero. If the variances for the two methods were estimated independently, the standard deviation of $\hat{\mathrm{D}}_{1,2}$ could be obtained from the standard deviations of $\hat{\mathrm{S}}_T^{(1)^2}$ and $\hat{\mathrm{S}}_T^{(2)^2}$ using (20). However, since both variance estimates are obtained from the same simulation runs, the correlation of $\Delta_r^{(1)}$ and $\Delta_r^{(2)}$ is important and we have to determine the margin of (22) directly. Following the same procedure as for the bias, one arrives at the expression

$$\hat{\mathrm{S}}\left[\hat{\mathrm{D}}_{1,2}\right] = \sqrt{\frac{1}{R(R-1)}\sum_r \left(\Delta_r^{(1)^2} - \Delta_r^{(2)^2} - \hat{\mathrm{D}}_{1,2}\right)^2} \; . \tag{23}$$

Assuming normality, the difference $\hat{\mathrm{D}}_{1,2}$ of two variances is considered significant if its absolute value exceeds 1.96 times its standard deviation (23).

4. Simulation 1: overlapping samples

4.1 Simulation scheme

The microdata set for simulation 1 has the same structure as that of the SES, see section 2.7.1. It consists of a register, a large EWS-sample which covers nearly half the population and a relatively small LFS-sample which has a 50% overlap with the EWS sample. The register contains the variables *G*, *A* and *B*, the EWS sample the variables *H* and *W* and the LFS sample the variables *E*. The EWS sample is drawn from the artificial population with the inverse survey weights d_i^{EWS}. The LFS sample is drawn with the inverse of the imputed survey weights d_i^{LFS}.[4] From these survey weights, the starting weights for blocks 2 through 4 are calculated as discussed in section 2.7.2, using the calibration model $[G \times A] \times 1 + [B] \times 1$.

Poisson sampling is used to draw the samples. As a result, the sample sizes n_k fluctuate between different draws with a relative standard deviation of $(n_k)^{-1/2}$. Poisson sampling was used because it requires less computation time than fixed-size sampling techniques. For a simple Horvitz-Thompson estimator using uncalibrated weights, the fluctuations in n_k would result in increased variances compared to fixed size sampling. However, all estimators used in this simulation include a calibration on the total population size N, which removes most of this effect, see Särndal *et al.* (1992), page 87.[5]

4.2 Table set

The simulation focuses on the table *average monthly wage by gender, working hours and education level,* which we denote by $[G \times H \times E] \times \mathrm{Av}(W)$. This table is the ratio of the quantitative table $[G \times H \times E] \times W$ and the frequency table $[G \times H \times E] \times 1$. It is a more detailed version of the SES output table $[G \times E] \times \mathrm{Av}(W)$. The extra classification H is included to create a larger difference between the calibration schemes for standard weighting, minimal repeated weighting and splitting up.

4.2.1 Minimal repeated weighting

For the case of minimal repeated weighting, the following table set is generated from the output tables $[G \times H \times E] \times W$ and $[G \times H \times E] \times 1$.

Table	Block	Recalibrations
$[G \times H] \times 1$	2	starting weights
$[G \times H] \times W$	2	starting weights
$[G \times E] \times 1$	3	starting weights
$[G \times H \times E] \times 1$	4	$[G \times H] \times 1 + [G \times E] \times 1$
$[G \times H \times E] \times W$	4	$[G \times H] \times W + [G \times H \times E] \times 1$

---

[4] Notice that the samples are drawn with *inverse survey weights,* not with inclusion probabilities. As was mentioned in section 2.7.3, the inclusion probabilities for the SES are not known. We therefore treat the survey weights as inclusion weights, ignoring the effects of non-response and of possible calibrations included in the survey weights.

[5] It can be shown that the adjusted Poisson estimator has, for $1 \ll n \ll N$, approximately the same variance as a (fixed size) Hansen-Hurwitz estimator which is calibrated on the population size. Since the VRD variance formula (section 2.6) was derived for a Hansen-Hurwitz estimator, we expect that the accuracy of the variance estimates is not affected by the fact that Poisson sampling is used instead of a fixed-size sampling method.

The first three tables are obtained without repeated weighting, since they can be estimated consistently with each other using the starting weights. The last two tables are esimated with repeated weighting;

4.2.2 *Splitting up*

When the splitting up procedure is used, the same output tables generate the set

Table	**Block**	**Recalibrations**
[*G* x *H*] x 1	2	starting weights
[*G* x *H*] x *W*	2	starting weights
[*G* x *E*] x 1	3	starting weights
[*H* x *E*] x 1	4	[*H*] x 1 + [*E*] x 1
[*G* x *H* x *E*] x 1	4	[*G* x *H*] x 1+ [*G* x *E*] x 1+ [*H* x *E*] x 1
[*E*] x *W*	4	[*E*] x 1 + *W*
[*G* x *E*] x *W*	4	[*G*] x *W* + [*E*] x *W* +[*G* x *E*] x 1
[*H* x *E*] x *W*	4	[*H*] x *W* + [*E*] x *W* + [*H* x *E*] x 1
[*G* x *H* x *E*] x *W*	4	[*G* x *H*] x *W* + [*G* x *E*] x *W* + [*H* x *E*] x *W* + [*G* x *H* x *E*] x 1

The tables [*H*] x 1 , *W*, [*E*] x 1, [*H*] x *W* and [*G*] x *W* are omitted from this set because they are "redundant". Including them would not change the weights for any of the other tables in the set.[6]

4.3 Properties of the data blocks

The EWS sample contains approximately 2.8 million elements. Its survey weights have values between 1 and 300, with an average of 2.3. The sample has a large value of L: 11.1 (see eq. (8)). Consequently, the EWS sample has an effective size of $n/(1+L) \approx 230{,}000$. The LFS sample, which has approximately 100,000 elements, has a more even weight distribution. Its lowest weight is 29, its highest 152 and it has an L of 0.3.

The target tables [*G* x *H* x *E*] x *W* and [*G* x *H* x *E*] x 1 are both obtained from block 4, which contains the cross section of the EWS and LFS samples. This block

[6] This is the case because there appear higher dimensional tables in the set of which the omitted tables are margins and which, furthermore, are estimated from the same block and with the same weights as the omitted tables. The presence of these tables ensures that the calibrations on the omitted tables are carried out, even if they are not themselves included in the set.

contains approximately 45,000 elements. Its zeroth-order block weights (13) have an L of 9.5. As a result, the effective size of this block is only about 4300.

Table *1* shows the cell size and the effective cell size for the cross-tabulation $G \times H \times E$ from block 4. Since the size of the block varies between different runs, the table shows expectation values. The number of observations varies from 12 to over 7000 per cell. The effective cell size is lower than this by a factor of one plus the relative weight variance in the cell. This factor varies from 8.8 to 12.1 for the different cells.

Clearly, the extremely low values of $n_{\text{eff.,cell}}$ for some cells would be problematic if this were a real output table. Estimates from these cells are expected to be very unreliable. For the purpose of performance evaluation, however, it is useful to have such a wide range of n_{cell} and $n_{\text{eff.,cell}}$, since it allows us to test repeated weighting under varying circumstances.

Table 1: Real and effective cell size for the cross-tabulation G x H x E from block 4.

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	n_{cell}	29	101	84	239	1818	43	180	296	566	478
	$n_{\text{eff., cell}}$	3	9	9	27	178	4	17	32	55	46
Lower secondary	n_{cell}	46	157	126	377	3025	71	270	481	905	768
	$n_{\text{eff., cell}}$	4	14	12	41	292	7	25	50	91	73
Upper secondary C	n_{cell}	35	121	83	172	1115	75	276	382	714	650
	$n_{\text{eff., cell}}$	3	10	8	18	106	7	24	39	69	61
Upper secondary A/B	n_{cell}	27	91	60	132	896	56	189	250	479	517
	$n_{\text{eff., cell}}$	2	8	6	14	86	5	16	25	46	47
Tertiary C	n_{cell}	80	273	241	859	7549	200	759	1442	2840	2645
	$n_{\text{eff., cell}}$	7	24	24	92	736	18	67	146	290	248
Tertiary B	n_{cell}	27	96	98	394	3509	93	369	749	1495	1333
	$n_{\text{eff., cell}}$	3	8	10	43	343	8	34	77	146	128
Tertiary A and higher	n_{cell}	12	40	47	210	1904	26	106	242	493	455
	$n_{\text{eff., cell}}$	1	3	5	24	190	3	9	24	49	43

5. Simulation 2: non-overlapping samples

5.1 Simulation scheme

Simulation 1 tested repeated weighting in a situation where the dataset contained two overlapping samples, one of them large, and where one of the blocks consisted of the cross-section of these two samples. In addition, four other simulations have been carried out: simulations 2.1-2.4. These simulations test repeated weighting in situations where the dataset contains two small, non-overlapping samples and one of the blocks contains the union of these samples. In such a situation, the criteria for using the VRD variance estimator are satisfied (see section 2.6), which was not the case in simulation 1.

The dataset used in these four simulations is of the type discussed in section 2.1. The two samples each contain approximately a hundred thousand elements. They are drawn from the same artificial population that is used in simulation 1. The register contains the same variables as in simulation 1: G, A and B. For sample s_1, the variables H and W are available, while s_2 contains data on H, W and E. Three data blocks are formed from this set, see figure 1. The starting weights for these blocks are calculated as discussed in section 2.2, with the difference that the calibration scheme [G x A] x 1 + [B] x 1 is used instead of [G] x 1. This same calibration scheme was used for the starting weights in simulation 1.

In the first and simplest of the four simulations (simulation 2.1) the two samples are drawn with simple random sampling and each has a fixed size of n=100,000. To ensure zero overlap between the two samples, they are drawn in a two stage process. First, the 200,000 elements in block 2 are drawn. Next, 100,000 units from this block are randomly assigned to s_1, the remaining 100,000 elements are assigned to s_2. Thus, the size of the two samples and of their union is the same in every simulation run.

In simulations 2.2 through 2.4, a non-trivial sampling scheme is used which is derived from that of the SES. One sample is drawn with the inverse LFS survey weights, the other with the inverse of adjusted EWS survey weights $d_i^{\text{EWS}*}$. The weights are adjusted to ensure that the sample size is approximately 100,000 units. Furthermore, their distribution is made more narrow to reduce the difference between the real and effective sample size, see section 5.3.2. Poisson sampling is used, so the sample- and block sizes fluctuate between draws.

In simulation 2.2, s_1 is drawn with the inverse $d_i^{\text{EWS}*}$-weights and s_2 with the inverse of d_i^{LFS}. The starting weights for block 2 are calculated using factors λ_k based on the real, not the effective sample size, see section 2.2. Since both samples have nearly the same size, this results in λ_1=1- $\lambda_2 \approx$ ½,. Simulation 2.3 follows the same procedure, but now the λ_k are based on the effective sample size $n_{eff.}$, which takes into account the difference in the relative weight variance between the two samples. This results in $\lambda_1 \approx 0.19$, $\lambda_2 \approx 0.81$. In simulation 2.4, the inclusion probabilities for s_1 and s_2 are interchanged. The λ's are again based on the effective sample sizes, which now yields $\lambda_1 \approx 0.81$, $\lambda_2 \approx 0.19$.

5.2 Table set

Simulations 2.1-2.4 again focus on the output table [G x H x E] x Av(W). Even though the structure of the data set in these simulations is different from that in simulation 1, it turns out that the same tables have to be estimated, with the same calibration schemes, see sections 4.2.1 and 4.2.2. Tables that are estimated from block 4 in simulation 1 are estimated from block 3 in simulations 2.1-2.4.

5.3 Properties of the data blocks

5.3.1 Simulation 2.1

In simulation 2.1, the samples s_1 and s_2 both have a fixed size of n=100,000. Because simple random sampling is used, the variance of their weights is equal to zero and the effective cell size is equal to the real cell size.

Both output tables are obtained from block 3. Table 2 shows the cell size for the cross-tabulation G x H x E from this block. There are no extremely small cell sizes as was the case for simulation 1 (see table 1). The smalles cell has 51 elements.

Table 2: Cell size for G x H x E from block 3 in simulation 2.1.

Gender x Working hours	Male					Female				
Education Level	<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	101	309	199	577	4332	135	486	566	1105	936
Lower secondary	153	470	322	850	7400	205	687	881	1649	1505
Upper secondary C	135	381	221	456	2985	240	758	720	1376	1361
Upper secondary A/B	98	268	150	354	2419	173	517	488	941	1109
Tertiary C	276	859	578	1899	17806	619	2006	2647	5124	5226
Tertiary B	92	309	245	826	8006	295	982	1377	2711	2513
Tertiary A and higher	51	142	119	437	4299	66	258	454	888	868

5.3.2 Simulations 2.2 -2.4

One of the samples in simulations 2.2-2.4 is drawn using adjusted EWS-weights $d_i^{\text{EWS*}}$. The mean value of these weights is 28 times that of the EWS-weights, reducing the sample size from 2,8 milion to 100,000. Furthermore, the adjusted weights have a narrower distribution than the original EWS-weights. The EWS* samples have an average L of 4.6, compared to 11.1 for the EWS samples. This L is still much larger than the value of 0.3 for the LFS-samples, but the difference between the two samples is not as extreme as it was in simulation 1. The effective size of the EWS* sample is approximately 18,000, that of the LFS sample 77,000.

Tables 3 and 4 show the real and effective cell sizes for G x H x E from the LFS- and EWS*-sample. The LFS sample is expected to yield the more accurate estimates, because of its larger effective size.

The output tables are obtained from block 3. Below, we list which sample's elements appear in block 3 for simulations 2.2-2.4. Furthermore, we give the values of the parameters λ_k. Since the λ's are different for different simulation runs, the values listed here are expectation values.

	λ_1	λ_2	block 3=s_2
Simulation 2.2	0.50	0.50	LFS sample
Simulation 2.3	0.19	0.81	LFS sample
Simulation 2.4	0.81	0.19	EWS* sample

Simulations 2.2 and 2.3 are identical except for the way in which the parameters λ are calculated. In simulation 2.3, the quality difference between the two samples is taken into account, in simulation 2.2, it is not. We therefore expect to obtain more accurate RW estimates in simulation 2.3.

Simulations 2.3 and 2.4 differ in the role which the two samples play. In 2.3, the output tables are obtained from the "good" LFS sample, while the "poor" EWS* sample is used only to obtain table estimates from block 2 for use in calibrations.

Because of its large λ, the LFS sample yields the dominant contribution to estimates from block 2. Since the starting weights estimates are also obtained from the LFS sample, we expect to find little difference between the RW- and SW-estimates in this simulation. In simulation 2.4, the output tables are obtained from the EWS*-sample and the starting weights variances will therefore be larger than those in simulation 2.3. The calibrations on tables from block 2, which is dominated by the LFS sample, are now expected to have a large effect and considerable differences between SW- and RW-estimates are expected.

Table 3: Real and effective cell size for G x H x E from the LFS sample

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	n_{cell}	98	298	185	535	4175	138	480	581	1103	935
	$n_{\text{eff., cell}}$	78	226	144	410	3245	109	367	449	849	716
Lower secondary	n_{cell}	146	464	322	860	7358	214	714	915	1709	1580
	$n_{\text{eff., cell}}$	116	360	251	668	5753	165	556	712	1341	1236
Upper secondary C	n_{cell}	134	349	216	447	2921	259	770	761	1402	1415
	$n_{\text{eff., cell}}$	104	275	170	345	2276	198	601	590	1092	1090
Upper secondary A/B	n_{cell}	94	264	141	339	2411	185	532	505	979	1157
	$n_{\text{eff., cell}}$	76	206	110	262	1870	141	416	393	762	891
Tertiary C	n_{cell}	288	881	568	1877	17927	650	2099	2763	5370	5485
	$n_{\text{eff., cell}}$	220	682	447	1471	14060	513	1632	2163	4184	4276
Tertiary B	n_{cell}	92	309	236	820	8055	330	1038	1452	2834	2608
	$n_{\text{eff., cell}}$	73	243	184	641	6308	250	808	1138	2213	2032
Tertiary A and higher	n_{cell}	46	141	119	445	4329	67	263	481	922	893
	$n_{\text{eff., cell}}$	37	111	90	344	3387	53	205	373	725	702

Table 4: Real and effective cell size for G x H x E from the adjusted EWS sample

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	n_{cell}	71	244	201	577	4302	100	414	655	1268	1083
	$n_{\text{eff., cell}}$	12	36	44	95	857	20	75	137	237	190
Lower secondary	n_{cell}	109	368	295	875	7033	159	601	1042	1977	1682
	$n_{\text{eff., cell}}$	19	57	50	155	1335	29	107	213	385	289
Upper secondary C	n_{cell}	83	290	195	408	2630	169	616	835	1568	1445
	$n_{\text{eff., cell}}$	10	39	31	70	484	27	91	171	297	224
Upper secondary A/B	n_{cell}	63	215	141	316	2124	124	421	546	1049	1142
	$n_{\text{eff., cell}}$	8	31	23	56	381	19	63	107	189	183
Tertiary C	n_{cell}	186	633	552	1975	17404	447	1691	3116	6167	5788
	$n_{\text{eff., cell}}$	28	88	91	345	3349	71	279	594	1198	1021
Tertiary B	n_{cell}	66	224	230	901	8090	208	818	1625	3255	2919
	$n_{\text{eff., cell}}$	11	29	47	172	1553	29	139	315	603	554
Tertiary A and higher	n_{cell}	28	94	108	483	4385	57	232	524	1072	995
	$n_{\text{eff., cell}}$	3	14	16	81	872	12	42	88	207	197

6. Results and discussion

The results for the five simulations are presented in the appendix. Tables 6-10 show the relative biases of the SW, RW and RW+ estimators. Tables 11-15 show the standard deviations of these estimators. Figures 3-7 compare the standard deviations of the SW and RW and the RW and RW+ estimators. Figures 8-12 show the coverage probability of the VRD confidence interval as a function of the relative SW standard deviation of the cell. In figure 13, the coverage probability is plotted versus the effective cell size for all five simulations. All these results are obtained for the target table $[G \times H \times E] \times \text{Av}(W)$.

6.1 Bias

The biases of the SW, RW and RW+ estimators for the five simulations are shown in tables 6-10. They are expressed as a percentage of the population value for Av(W) in the cell, which is the same in every simulation. Values for the bias which differ significantly from zero are indicated in grey. In these grey cells, the bias ratio exceeds 0.08 (see section 3.3.1).

Throughout the simulations, little difference is seen between the RW and the RW+ estimator as far as the bias is concerned, but some differences are found between these two repeated weighting estimators and the standard regression estimator (SW). Over all, the standard regression estimator is seen to yield a somewhat smaller bias, although, throughout the whole range of (effective) cell sizes, also many cases occur where repeated weighting does better. The largest differences between the SW and RW/RW+ bias, of the order of a few percent of the population value, are found in simulations 1 and 2.4 for the very worst cells ($n_{\text{eff.,cell}}\approx 10$ and smaller). Cells of such poor quality only occur in these two simulations. For the cells with larger effective sizes, the difference in the relative bias between the SW- and the RW/RW+ estimators are of the order of 0.1% or smaller.

To determine the relevance of these changes in the bias, it is useful to consider what their impact is on the root mean square error (rmse). The rmse is the square root of the quadratic sum of bias and standard deviation. If the SW bias in the rmse is replaced by the RW or RW+ bias, a relative change of at most 5% of the rmse is found for effective cell sizes of about 10 and smaller. For effective sizes between 10 and 1000, the changes are (much) smaller than 1%. For effective cell size larger than 1000, increases of the rmse of up to 4% are again found for a small number of cells in simulation 2.3. However, these changes are very small in absolute terms, since the rmse is only some tens of a percent of the cell value in these cases.

Altogether, we find no indications for a strong accumulation of bias as a result of repeated weighting. There is some increase in bias, but even for the very worst cells with effective sizes of about ten or less, this only results in an increase of the rmse of at most 5%.

6.2 Standard deviation

Below, in table 5, we summarize the results for the standard deviation S_T in the five simulations. This table is based on figures 3-8 in the appendix. The first column indicates whether minimal repeated weighting results grosso modo in a smaller or larger standard deviation than standard weighting. The third column compares splitting up to minimal repeated weighting in the same way. The second and fourth column give an indication of the relative magnitude of the differences in the standard deviations for RW compared to SW and for RW+ compared to RW.

Table 5:Overview of the differences in the standard deviation of the SW, RW and RW+ estimators for the five simulations

	S_T^{RW} compared to S_T^{SW}		S_T^{RW+} compared to S_T^{RW}	
	trend	scale of rel. differences	trend	scale of rel. differences
simulation 1	smaller	10%	larger	1%
simulation 2.1	smaller	1%	similar	0.1%
simulation 2.2	larger	10%	similar	1%
simulation 2.3	smaller	1%	similar	0,1%
simulation 2.4	smaller	10%	larger	1%

In simulations 1, 2.1, 2.3 and 2.4, minimal repeated weighting yields a more accurate estimator than standard weighting. In simulation 2.2, it yields a less accurate one. Apparently, the choice of λ_k in this simulation results in estimates from block 2 which are less accurate than those from block 3. The calibrations on these unreliable estimates lead to an RW estimator which has a larger standard deviation than the SW estimator. In simulation 2.3, which is identical to simulation 2.2 except for the fact that the λ_k are determined from the effective rather than the real sample sizes,[7] repeated weighting does yield an improved estimator. This demonstrates the importance of taking into account not only the samples' sizes but also the variances of their weights when determining the parameters λ_k.

In figures 3-7 in the appendix, the relative difference between the SW and RW standard deviations is plotted against the relative SW standard deviation of the cell, which scales with $1/\sqrt{n_{\text{eff.,cell}}}$ for each of the simulations.[8] In simulations 1, 2,1, 2.3 and 2.4, an increase of the standard deviation in RW compared to SW is only found for a handful of cells with a small effective size. A significant increase is found in three cases, twice for cells with an effective size of 3, once for a cell with an effective size of 12.

We conclude that, if the parameters λ_k are chosen correctly, repeated weighting yields a smaller standard deviation than standard weighting. This improvement is found throughout, except for cells with very small effective sizes (n_{eff}. ≈ 10 or smaller).

[7] For a given simulation run *r*, the samples used in simulations 2.2 and 2.3 are identical. As a result, the SW estimates for the target table, which are obtained from block 3, are exactly the same in both simulations.

[8] The relative SW standard deviation can be fitted to $C_0 + C_1/\sqrt{n_{\text{eff.,cell}}}$ with R^2=0.8-0.9 for the different simulations. The intersect C_0 is small: between -0.009 and +0.006. The slope C_1 is of order 1 (between 0.6 and 0.9).

For all five simulations, the differences between the results for splitting up and minimal repeated weighting are considerably smaller than between those for minimal repeated weighting and standard weighting. Apparently, the additional calibrations that are performed in the splitting up procedure have little effect on the standard deviation of the estimator. This can be understood from the fact that the most influential calibrations will be those on tables estimated from larger blocks than the target table, and those calibrations are already included in the minimal reweighting scheme. Since additional calibrations introduce extra fluctuations into the estimator (due to the variance in the estimates of the regression coefficients) a small adverse effect of the splitting up procedure on the precision could be expected.

This deterioration of the precision is indeed found in simulations 1 and 2.4. That it is found for these simulations and not for the others could be because the effective size of the block from which the target table is obtained is smallest in simulations 1 and 2.4 (4,300 and 18,000 compared to 77,000 and 100,000). The small effective block size leads to larger fluctuations in the estimates of the regression coefficients that are used in the RW+ estimator. This in turn results in an increased standard deviation. Both in simulation 1 and 2.4, the increase in standard deviation in RW+ compared to RW is found to be worst for the cells with the smallest effective size (or largest relative SW standard deviation). This dependence could be explained in the same way: the estimates of regression coefficients that refer to the smallest cells have the largest fluctuations.

An alternative explanation for the fact that an increase in the RW+ standard deviation is only seen in simulations 1 and 2.4 is that, in these simulations, the differences between RW and RW+ are of the order of 1%, compared to 0.1% in simulations 2.1 and 2.3 Possibly, the RW+ estimator is less accurate than the RW estimator also for these other simulations, but the differences are too small to show up in the simulation results. This argument cannot explain why no clear deterioration is seen in simulation 2.2, where the simulation results should be accurate enough to show the effect.

With regard to the splitting up procedure, we conclude that it yields estimates which are of a similar precision as those from minimal repeated weighting. When an increase in the standard deviation is found, this increase is worst for the cells whose effective sizes are smallest.

6.3 Coverage probability

Figures 8-12 show the coverage probability of the VRD confidence interval $P_{\text{cell}}^{\text{VRD}}$ for the different simulations. These results are obtained for the minimal repeated weighting estimator only. The nominal value for the coverage probability is 95%. It is found that this value is only achieved for the best cells. Plotted against the relative SW standard deviation of the cell, we find a linear reduction of the coverage probability for relative standard deviation larger than about 0.03-0.05 (the onset point varies somewhat for the different simulations). For simulation 1, nearly all cells are in the region where a linear reduction of the coverage probability is seen.

For simulation 2.1, most cells are in the “save” region and have values for $P_{\text{cell}}^{\text{VRD}}$ that do not differ significantly from 95% (equation (19) is used to test for statistical significance). In the results for simulations 2.2-2.4 one can see the cross-over from a roughly constant value for $P_{\text{cell}}^{\text{VRD}}$ at or near 95% to a linear reduction of the probability at higher relative standard deviations.

In figure 13, $P_{\text{cell}}^{\text{VRD}}$ is plotted against the effective cell size for all five simulations. The results collapse fairly well onto a single curve, suggesting that $P_{\text{cell}}^{\text{VRD}}$ is mainly determined by the effective cell size. We should bear in mind that all these points refer to estimates of the same table, using weights that have undergone the same calibrations in each of the simulations, so possibly the type of variable which is estimated and the specific calibrations which are performed are additional factors determining $P_{\text{cell}}^{\text{VRD}}$. Given this particular table and these particular calibrations, the VRD confidence interval is found to be reliable only for effective cell sizes larger than 50-100.

Most likely, this problem is due to the linearization technique used by the VRD variance estimator, on which the confidence interval is based. By comparing the variance determined from the simulation, $\hat{S}_T^2$, with the simulation average of the VRD variance estimates $\hat{S}_{\text{VRD},r}^2$, it can be seen that the variance estimator has a negative bias which increases as the effective cell size decreases. Since the same trend is found for simulation 1 and simulations 2.1-2.4, even though the variance is estimated in a different way in simulation 1 (see section 2.7.3), the main cause for the negative bias has to lie in an approximation which is used in the variance estimator for each of the simulations. For this reason, we conclude that it is the linearization procedure which causes these problems: the variance is underestimated because the variances of the matrices $\boldsymbol{M}$, which contain the estimated regression coefficients, are neglected (see section 2.6). Since the regression coefficients are determined less accurately for cells of poorer quality, this would explain the $n_{\text{eff.,cell}}$-dependence of $P_{\text{cell}}^{\text{VRD}}$.[9]

Note that this same problem occurs for variance estimates for the standard regression estimator, see Särndal *et al.* (1992), page 280. Arguably, the problem is more serious for repeated weighting, since RW estimates can involve a very large number of calibrations, some of which use other (RW) estimates as restrictions. The difference between the RW- and the SW variance estimator in this respect is a subject for further research. It would also be useful to investigate whether one can

---

[9] We have verified that, if the standard deviation $\hat{S}_T$ is used to determine the confidence interval rather than the VRD-estimate for the standard deviation $\hat{S}_{\text{VRD},r}$, the cell-size dependence of the coverage probability disappears and we do obtain the nominal value of 95%. This clearly shows that the problem is caused by the negative bias of the variance estimator and not, for instance, by deviations from the normal distribution for small cells, which would change the relation between standard deviation and confidence interval.

specify in general, so also for table estimates with different calibration schemes than were considered here, under what conditions the variance estimator can be relied upon. A further step would be to try and correct for the negative bias in the variance estimator.

6.4 Summary of conclusions

We give a brief summary of the conclusions from the previous paragraphs.

- Repeated weighting results in only a small increase in bias for cells of sufficient quality ($n_{\text{eff.,cell}} > 10$). The splitting up procedure yields no additional bias compared to minimal repeated weighting
- The parameters λ in the starting weights should reflect the relative quality of the constituent samples of a data block. To accurately compare the quality of two samples, it is important to consider not only the sample sizes but also the coefficient of variation of the samples' weights.
- If the parameters λ are chosen properly, minimal repeated weighting yields a reduced standard deviation compared to a standard regression estimator which uses only auxiliary information from a register.
- The splitting up procedure may result in an increase in the standard deviation compared to minimal repeated weighting. However, this increase is much smaller than the gain in precision compared to the standard regression estimator.
- The VRD variance estimator underestimates the variance for small effective cell sizes. This is probably due to the linearization procedure, in which the variances of the estimated regression coefficients are neglected. This approximation is also commonly used when determining the variance of a standard regression estimator. For the target table considered here, the VRD variance estimator becomes reliable for effective cell sizes larger than 50-100.

Appendix: results for the table [*G x H x E*] x Av(*W*)

Bias

Tables 6 through 10 show the relative bias $\hat{B}_{T}^{sim.}/\boldsymbol{T}$ of the SW, RW and RW+ estimators of the table [*G x H x E*] x Av(*W*) for the various simulations. Significant biases are indicated in grey.

Table 6: Relative bias for simulation 1

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	-2,8%	-0,6%	-0,2%	1,1%	0,2%	-1,8%	1,5%	0,7%	0,7%	0,2%
	RW	2,1%	2,3%	0,6%	0,9%	0,0%	1,5%	1,8%	0,7%	0,5%	0,2%
	RW+	0,9%	2,0%	0,6%	1,0%	0,0%	0,5%	1,4%	0,5%	0,4%	0,0%
Lower secondary	SW	-2,1%	-0,9%	0,0%	0,8%	0,4%	-3,3%	0,0%	0,5%	0,4%	0,6%
	RW	-1,0%	1,1%	0,7%	0,5%	0,3%	0,7%	0,0%	0,1%	0,1%	0,3%
	RW+	-1,5%	1,0%	0,6%	0,6%	0,3%	0,2%	-0,3%	0,2%	0,2%	0,4%
Upper secondary C	SW	-2,1%	-2,8%	-2,0%	0,4%	0,4%	-3,1%	1,6%	-0,2%	-0,1%	-0,1%
	RW	-1,4%	-0,7%	-1,3%	0,2%	0,1%	-4,8%	1,1%	-0,4%	-0,4%	-0,1%
	RW+	-1,0%	-0,7%	-1,3%	0,4%	0,1%	-4,7%	1,3%	-0,3%	-0,3%	-0,3%
Upper secondary A/B	SW	-1,8%	-0,2%	-1,8%	1,0%	-0,3%	2,5%	-0,6%	-0,3%	0,8%	1,1%
	RW	-0,1%	2,2%	-0,8%	0,7%	-0,4%	4,1%	-0,7%	-0,3%	0,5%	0,8%
	RW+	-0,5%	2,0%	-1,0%	0,8%	-0,4%	3,9%	-0,6%	-0,3%	0,5%	0,9%
Tertiary C	SW	-1,8%	-3,0%	-0,4%	0,5%	0,2%	-1,5%	0,8%	0,2%	0,3%	0,4%
	RW	-0,8%	-2,1%	0,8%	0,3%	0,0%	-0,6%	0,8%	0,1%	0,1%	0,2%
	RW+	-0,6%	-1,8%	0,8%	0,3%	0,1%	-0,1%	0,7%	0,1%	0,1%	0,2%
Tertiary B	SW	1,7%	1,2%	0,6%	0,6%	0,1%	0,0%	-0,5%	0,6%	0,3%	-0,3%
	RW	5,8%	4,7%	1,5%	0,4%	-0,1%	0,2%	-1,0%	0,6%	0,0%	-0,3%
	RW+	4,9%	4,8%	1,4%	0,4%	-0,1%	-0,3%	-0,8%	0,6%	0,0%	-0,3%
Tertiary A and higher	SW	4,9%	0,4%	8,7%	1,7%	-0,2%	3,7%	-0,4%	0,6%	0,6%	0,0%
	RW	10,7%	3,8%	9,5%	1,4%	-0,4%	9,1%	-2,0%	0,6%	0,4%	-0,2%
	RW+	9,3%	3,2%	9,8%	1,5%	-0,4%	8,1%	-2,0%	0,5%	0,3%	-0,1%

Table 7:Relative bias for simulation 2.1

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	-0,31%	-0,16%	0,29%	-0,02%	-0,03%	-0,21%	0,06%	0,00%	0,04%	0,05%
	RW	-0,19%	-0,16%	0,24%	0,01%	-0,02%	-0,17%	0,03%	-0,02%	0,02%	0,06%
	RW+	-0,21%	-0,16%	0,24%	0,01%	-0,02%	-0,21%	0,03%	-0,02%	0,02%	0,06%
Lower secondary	SW	-0,11%	0,04%	-0,40%	-0,04%	0,00%	0,60%	-0,06%	0,05%	0,00%	-0,06%
	RW	0,10%	0,07%	-0,41%	-0,01%	0,01%	0,61%	-0,08%	0,02%	-0,02%	-0,05%
	RW+	0,09%	0,06%	-0,41%	-0,02%	0,01%	0,59%	-0,09%	0,02%	-0,02%	-0,05%
Upper secondary C	SW	-0,07%	-0,27%	0,18%	0,00%	0,01%	-0,29%	-0,01%	0,06%	0,04%	-0,04%
	RW	0,07%	-0,24%	0,18%	0,03%	0,02%	-0,27%	-0,04%	0,04%	0,02%	-0,03%
	RW+	0,06%	-0,24%	0,18%	0,03%	0,02%	-0,26%	-0,03%	0,04%	0,02%	-0,03%
Upper secondary A/B	SW	-0,36%	-0,23%	-0,29%	0,09%	-0,10%	0,17%	0,28%	0,06%	0,00%	-0,10%
	RW	-0,22%	-0,20%	-0,30%	0,11%	-0,08%	0,22%	0,26%	0,04%	-0,03%	-0,08%
	RW+	-0,23%	-0,20%	-0,30%	0,12%	-0,08%	0,21%	0,26%	0,04%	-0,03%	-0,09%
Tertiary C	SW	-0,07%	-0,20%	0,10%	0,02%	-0,03%	-0,05%	0,02%	0,05%	0,03%	-0,03%
	RW	0,09%	-0,18%	0,10%	0,04%	-0,01%	-0,03%	0,00%	0,03%	0,01%	-0,02%
	RW+	0,10%	-0,18%	0,09%	0,04%	-0,01%	-0,01%	-0,01%	0,03%	0,01%	-0,02%
Tertiary B	SW	-0,62%	-0,09%	-0,01%	0,02%	0,02%	-0,08%	0,04%	-0,06%	0,01%	0,00%
	RW	-0,36%	-0,05%	-0,03%	0,04%	0,03%	-0,10%	0,01%	-0,08%	-0,01%	0,01%
	RW+	-0,36%	-0,05%	-0,03%	0,05%	0,03%	-0,09%	0,02%	-0,08%	-0,01%	0,01%
Tertiary A and higher	SW	-0,07%	0,03%	0,07%	0,13%	0,00%	-0,02%	0,19%	-0,05%	0,13%	0,05%
	RW	0,04%	0,07%	0,02%	0,16%	0,01%	0,03%	0,10%	-0,07%	0,11%	0,06%
	RW+	0,03%	0,09%	0,03%	0,15%	0,01%	0,01%	0,12%	-0,06%	0,11%	0,06%

Table 8: Relative bias for simulation 2.2

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	0,38%	0,29%	0,49%	-0,09%	-0,05%	-0,11%	0,01%	0,14%	-0,02%	-0,05%
	RW	0,21%	0,07%	0,45%	-0,29%	-0,05%	-0,27%	-0,10%	0,09%	-0,08%	-0,05%
	RW+	0,22%	0,05%	0,45%	-0,28%	-0,05%	-0,31%	-0,11%	0,08%	-0,08%	-0,06%
Lower secondary	SW	0,21%	-0,07%	0,10%	0,07%	-0,09%	0,32%	0,12%	-0,07%	-0,02%	-0,07%
	RW	0,05%	-0,24%	0,11%	-0,12%	-0,09%	0,02%	0,04%	-0,14%	-0,08%	-0,08%
	RW+	0,03%	-0,26%	0,09%	-0,11%	-0,09%	-0,03%	0,03%	-0,14%	-0,07%	-0,08%
Upper secondary C	SW	0,01%	0,06%	0,26%	0,21%	-0,12%	0,48%	-0,10%	-0,02%	-0,08%	-0,07%
	RW	-0,28%	-0,11%	0,27%	0,01%	-0,13%	0,05%	-0,24%	-0,08%	-0,14%	-0,08%
	RW+	-0,29%	-0,11%	0,27%	0,02%	-0,12%	0,05%	-0,23%	-0,08%	-0,13%	-0,07%
Upper secondary A/B	SW	-0,40%	0,24%	-0,62%	0,19%	0,03%	0,27%	0,21%	0,00%	0,01%	-0,10%
	RW	-0,62%	0,16%	-0,65%	0,00%	0,03%	0,06%	0,10%	-0,06%	-0,05%	-0,11%
	RW+	-0,60%	0,14%	-0,65%	0,01%	0,04%	0,08%	0,08%	-0,05%	-0,05%	-0,11%
Tertiary C	SW	0,08%	0,07%	0,40%	0,05%	-0,08%	0,34%	0,09%	-0,02%	-0,04%	-0,02%
	RW	-0,15%	-0,23%	0,38%	-0,13%	-0,08%	0,04%	-0,01%	-0,08%	-0,09%	-0,03%
	RW+	-0,12%	-0,22%	0,39%	-0,14%	-0,08%	0,04%	-0,02%	-0,07%	-0,10%	-0,03%
Tertiary B	SW	0,49%	-0,13%	-0,01%	-0,04%	-0,08%	0,22%	-0,18%	0,05%	-0,02%	-0,02%
	RW	0,12%	-0,32%	-0,05%	-0,23%	-0,08%	-0,09%	-0,30%	0,01%	-0,08%	-0,03%
	RW+	0,17%	-0,33%	-0,04%	-0,23%	-0,09%	-0,04%	-0,30%	0,01%	-0,08%	-0,02%
Tertiary A and higher	SW	-1,40%	0,00%	-0,21%	0,04%	-0,08%	-0,39%	0,18%	0,14%	-0,11%	-0,03%
	RW	-1,38%	-0,30%	-0,21%	-0,14%	-0,08%	-0,56%	-0,01%	0,09%	-0,16%	-0,03%
	RW+	-1,52%	-0,27%	-0,22%	-0,16%	-0,09%	-0,64%	0,03%	0,08%	-0,16%	-0,03%

Table 9: Relative bias for simulation 2.3

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	0,38%	0,29%	0,49%	-0,09%	-0,05%	-0,11%	0,01%	0,14%	-0,02%	-0,05%
	RW	0,28%	0,10%	0,25%	-0,35%	-0,12%	-0,19%	-0,11%	0,02%	-0,15%	-0,11%
	RW+	0,28%	0,08%	0,24%	-0,35%	-0,12%	-0,20%	-0,12%	0,01%	-0,15%	-0,11%
Lower secondary	SW	0,21%	-0,07%	0,10%	0,07%	-0,09%	0,32%	0,12%	-0,07%	-0,02%	-0,07%
	RW	0,10%	-0,24%	-0,08%	-0,18%	-0,15%	0,18%	0,02%	-0,22%	-0,15%	-0,13%
	RW+	0,10%	-0,25%	-0,09%	-0,18%	-0,15%	0,16%	0,01%	-0,22%	-0,14%	-0,13%
Upper secondary C	SW	0,01%	0,06%	0,26%	0,21%	-0,12%	0,48%	-0,10%	-0,02%	-0,08%	-0,07%
	RW	-0,15%	-0,09%	0,12%	-0,05%	-0,19%	0,29%	-0,26%	-0,18%	-0,21%	-0,13%
	RW+	-0,15%	-0,08%	0,12%	-0,03%	-0,18%	0,29%	-0,25%	-0,17%	-0,19%	-0,11%
Upper secondary A/B	SW	-0,40%	0,24%	-0,62%	0,19%	0,03%	0,27%	0,21%	0,00%	0,01%	-0,10%
	RW	-0,52%	0,14%	-0,77%	-0,06%	-0,03%	0,17%	0,07%	-0,14%	-0,12%	-0,16%
	RW+	-0,51%	0,14%	-0,76%	-0,04%	-0,02%	0,18%	0,07%	-0,13%	-0,11%	-0,15%
Tertiary C	SW	0,08%	0,07%	0,40%	0,05%	-0,08%	0,34%	0,09%	-0,02%	-0,04%	-0,02%
	RW	-0,06%	-0,17%	0,22%	-0,19%	-0,14%	0,20%	-0,03%	-0,15%	-0,16%	-0,08%
	RW+	-0,05%	-0,17%	0,22%	-0,19%	-0,14%	0,20%	-0,04%	-0,15%	-0,16%	-0,08%
Tertiary B	SW	0,49%	-0,13%	-0,01%	-0,04%	-0,08%	0,22%	-0,18%	0,05%	-0,02%	-0,02%
	RW	0,30%	-0,28%	-0,22%	-0,27%	-0,14%	0,08%	-0,33%	-0,06%	-0,14%	-0,07%
	RW+	0,31%	-0,28%	-0,22%	-0,28%	-0,14%	0,09%	-0,32%	-0,06%	-0,15%	-0,08%
Tertiary A and higher	SW	-1,40%	0,00%	-0,21%	0,04%	-0,08%	-0,39%	0,18%	0,14%	-0,11%	-0,03%
	RW	-1,43%	-0,21%	-0,40%	-0,19%	-0,14%	-0,47%	-0,01%	0,02%	-0,22%	-0,08%
	RW+	-1,48%	-0,20%	-0,41%	-0,21%	-0,15%	-0,50%	0,01%	0,01%	-0,23%	-0,09%

Table 10: Relative bias for simulation 2.4

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	0,50%	-0,59%	0,64%	0,31%	0,04%	-0,30%	-0,13%	0,12%	-0,07%	0,15%
	RW	1,85%	-0,45%	0,02%	0,14%	-0,14%	0,89%	-0,07%	-0,02%	-0,23%	0,06%
	RW+	1,49%	-0,42%	0,09%	0,15%	-0,14%	0,53%	-0,10%	-0,03%	-0,25%	0,03%
Lower secondary	SW	-1,31%	-0,55%	-0,13%	-0,18%	0,05%	0,39%	-0,16%	0,12%	0,24%	-0,04%
	RW	-1,61%	-0,34%	-0,51%	-0,35%	-0,12%	0,98%	-0,09%	-0,12%	0,05%	-0,18%
	RW+	-1,40%	-0,43%	-0,48%	-0,36%	-0,11%	0,92%	-0,17%	-0,08%	0,04%	-0,18%
Upper secondary C	SW	-1,51%	0,49%	0,46%	0,36%	-0,07%	-0,39%	0,05%	0,11%	0,17%	0,04%
	RW	-1,62%	0,91%	0,18%	0,19%	-0,27%	-0,70%	-0,07%	-0,11%	0,01%	-0,09%
	RW+	-1,54%	0,82%	0,14%	0,20%	-0,26%	-0,44%	-0,07%	-0,10%	0,04%	-0,05%
Upper secondary A/B	SW	-0,47%	0,29%	0,17%	0,52%	-0,09%	0,05%	0,73%	-0,08%	0,10%	0,43%
	RW	0,39%	0,71%	-0,21%	0,33%	-0,24%	0,79%	0,64%	-0,23%	-0,07%	0,27%
	RW+	0,21%	0,63%	-0,24%	0,35%	-0,23%	0,61%	0,63%	-0,24%	-0,05%	0,31%
Tertiary C	SW	-0,47%	-0,54%	0,47%	0,06%	0,04%	-0,77%	-0,07%	0,08%	-0,03%	0,03%
	RW	-0,51%	-0,62%	0,17%	-0,07%	-0,10%	0,17%	-0,03%	-0,08%	-0,19%	-0,10%
	RW+	-0,47%	-0,52%	0,14%	-0,08%	-0,10%	0,25%	-0,01%	-0,09%	-0,19%	-0,11%
Tertiary B	SW	1,00%	-0,56%	1,37%	0,02%	0,03%	0,09%	0,08%	0,11%	-0,04%	-0,06%
	RW	2,71%	0,03%	0,93%	-0,11%	-0,14%	-0,52%	0,02%	0,00%	-0,21%	-0,17%
	RW+	2,43%	0,03%	0,92%	-0,09%	-0,14%	-0,55%	0,03%	-0,01%	-0,21%	-0,18%
Tertiary A and higher	SW	2,03%	1,02%	1,23%	0,48%	0,03%	0,53%	-0,46%	0,10%	0,19%	0,09%
	RW	3,23%	1,46%	0,63%	0,35%	-0,14%	1,72%	-0,90%	-0,10%	0,03%	-0,03%
	RW+	2,92%	1,32%	0,67%	0,34%	-0,13%	1,35%	-0,85%	-0,09%	0,03%	-0,03%

Standard deviation

Tables 11-15 show the relative standard deviation $\hat{S}_T / \boldsymbol{T}$ for the SW, RW and RW+ estimators of the table [G x H x E] x Av(W) in the five simulations. Significant differences between the RW and RW+ results are indicated by a grey RW and RW+ field, significant differences between SW and both RW and RW+ are indicated by a grey SW field.

Figures 3-8 a) show the relative difference between the standard deviation of the SW and RW estimator, $\left(\hat{S}_{T^\alpha}^{SW} - \hat{S}_{T^\alpha}^{RW}\right) / \hat{S}_{T^\alpha}^{SW}$, for the 70 cells of [G x H x E] x Av(W). This quantity is plotted against the relative SW standard deviation $\hat{S}_{T^\alpha}^{SW} / T^\alpha$ of the cells. The relative SW standard deviation can be considered as a measure for the quality of the cell. Figures 3-8 b) show the relative difference between the standard deviation of the RW and RW+ estimator for the cells, plotted against the same quantity. In both sets of figures, boxes indicate significant differences.

Table 11: Relative standard deviation for simulation 1

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	29,2%	27,5%	24,0%	11,1%	3,9%	31,7%	12,9%	7,9%	5,4%	5,8%
	RW	33,1%	27,0%	21,2%	10,3%	3,6%	36,3%	12,9%	7,5%	5,1%	5,7%
	RW+	34,9%	27,0%	22,2%	10,5%	3,6%	38,2%	13,3%	7,6%	5,2%	5,8%
Lower secondary	SW	37,9%	27,5%	17,5%	8,7%	2,8%	24,0%	11,2%	7,6%	4,8%	5,6%
	RW	28,1%	21,7%	15,5%	8,0%	2,7%	25,5%	10,3%	6,1%	4,3%	4,6%
	RW+	30,3%	22,7%	16,0%	8,1%	2,6%	27,2%	10,8%	6,4%	4,4%	5,0%
Upper secondary C	SW	44,6%	24,7%	16,9%	12,2%	5,7%	42,5%	17,1%	8,6%	5,1%	5,2%
	RW	33,3%	23,6%	16,7%	11,9%	4,9%	27,5%	14,9%	7,7%	4,9%	4,8%
	RW+	38,4%	23,9%	17,3%	12,2%	4,9%	29,3%	15,7%	7,8%	4,9%	5,1%
Upper secondary A/B	SW	41,2%	25,7%	20,7%	13,5%	5,1%	41,6%	20,1%	9,2%	6,1%	7,4%
	RW	40,3%	25,5%	20,9%	12,7%	4,9%	37,7%	18,8%	9,0%	5,9%	6,4%
	RW+	40,8%	25,9%	21,4%	12,9%	4,8%	40,2%	19,7%	9,0%	6,0%	6,8%
Tertiary C	SW	30,8%	21,6%	11,8%	5,6%	1,6%	21,3%	8,1%	4,0%	2,4%	2,9%
	RW	21,2%	14,2%	10,2%	4,4%	1,3%	15,2%	6,6%	3,2%	2,0%	2,0%
	RW+	23,7%	14,7%	10,4%	4,5%	1,3%	16,5%	7,1%	3,2%	2,0%	2,2%
Tertiary B	SW	39,3%	23,2%	17,8%	7,1%	2,4%	39,2%	13,0%	5,2%	3,5%	3,4%
	RW	39,1%	22,7%	16,5%	6,7%	2,2%	22,5%	10,8%	4,7%	3,1%	3,1%
	RW+	42,9%	23,4%	16,7%	6,9%	2,2%	23,5%	11,4%	4,8%	3,2%	3,2%
Tertiary A and higher	SW	40,7%	27,8%	25,1%	10,1%	3,3%	27,1%	24,6%	9,3%	5,4%	6,6%
	RW	67,5%	28,0%	24,4%	9,4%	3,2%	35,5%	14,7%	9,0%	5,3%	5,6%
	RW+	67,6%	28,7%	25,5%	9,8%	3,1%	36,9%	16,6%	9,1%	5,4%	6,2%

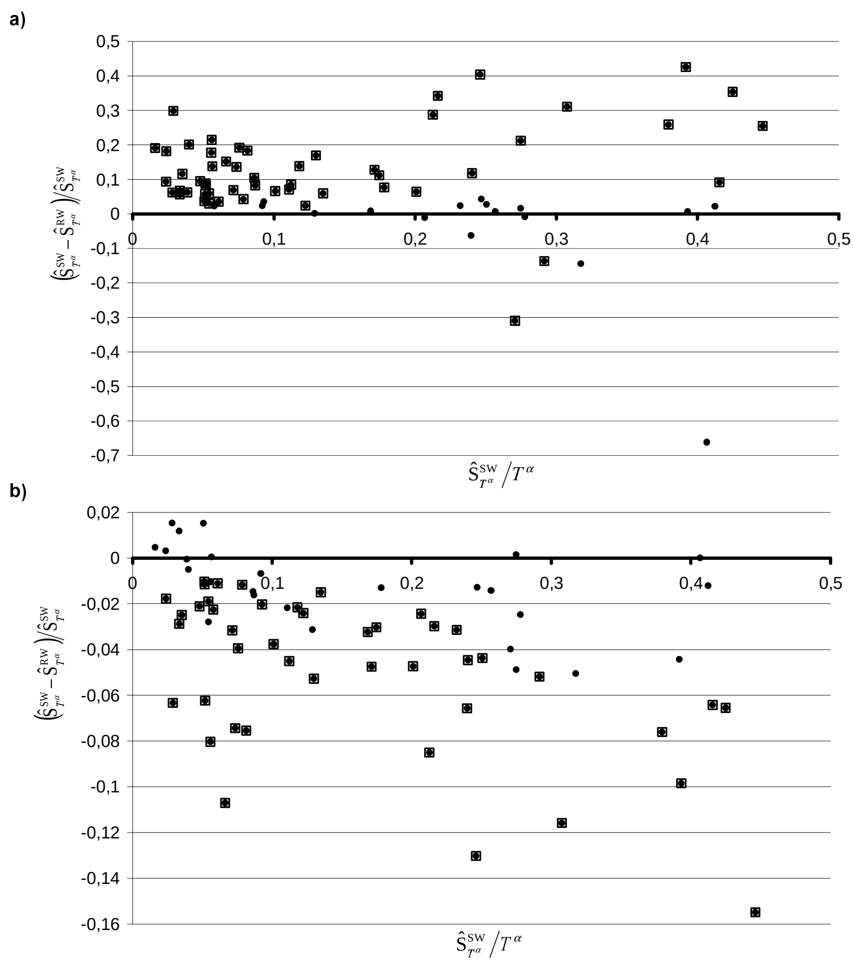

Figure 3: Relative difference of the standard deviations for the SW and RW estimator (a) and the RW and RW+ estimator (b), simulation 1

Table 12: Relative standard deviation for simulation 2.1

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	7,85%	5,36%	5,78%	2,74%	0,80%	7,00%	3,27%	1,98%	1,33%	1,44%
	RW	7,73%	5,17%	5,48%	2,68%	0,79%	6,89%	3,20%	1,96%	1,30%	1,42%
	RW+	7,73%	5,17%	5,49%	2,68%	0,79%	6,88%	3,21%	1,96%	1,30%	1,42%
Lower secondary	SW	6,91%	4,30%	4,24%	2,27%	0,58%	6,24%	2,47%	1,59%	1,07%	1,12%
	RW	6,85%	4,14%	4,11%	2,21%	0,56%	5,99%	2,41%	1,54%	1,03%	1,08%
	RW+	6,85%	4,13%	4,10%	2,21%	0,56%	6,00%	2,42%	1,54%	1,03%	1,09%
Upper secondary C	SW	8,26%	4,58%	5,27%	3,11%	1,09%	6,88%	2,92%	1,95%	1,29%	1,22%
	RW	7,96%	4,52%	5,23%	3,05%	1,08%	6,85%	2,82%	1,92%	1,28%	1,20%
	RW+	7,93%	4,51%	5,24%	3,05%	1,08%	6,86%	2,83%	1,91%	1,27%	1,20%
Upper secondary A/B	SW	8,34%	4,99%	6,68%	3,47%	1,05%	7,43%	3,33%	2,34%	1,55%	1,33%
	RW	8,22%	4,96%	6,60%	3,41%	1,05%	7,34%	3,28%	2,31%	1,52%	1,31%
	RW+	8,21%	4,96%	6,59%	3,42%	1,05%	7,35%	3,28%	2,32%	1,52%	1,31%
Tertiary C	SW	5,48%	3,26%	2,75%	1,36%	0,34%	3,55%	1,42%	0,89%	0,59%	0,62%
	RW	4,94%	2,94%	2,63%	1,22%	0,33%	3,23%	1,30%	0,80%	0,54%	0,55%
	RW+	4,93%	2,94%	2,63%	1,22%	0,33%	3,23%	1,30%	0,80%	0,54%	0,55%
Tertiary B	SW	9,23%	4,25%	4,60%	1,89%	0,54%	5,42%	2,10%	1,08%	0,83%	0,84%
	RW	9,21%	4,13%	4,47%	1,84%	0,52%	5,11%	1,99%	1,03%	0,78%	0,81%
	RW+	9,22%	4,13%	4,46%	1,84%	0,52%	5,09%	2,00%	1,03%	0,78%	0,81%
Tertiary A and higher	SW	10,23%	6,73%	6,27%	2,49%	0,77%	9,18%	4,54%	1,92%	1,38%	1,31%
	RW	10,06%	6,64%	6,09%	2,43%	0,75%	9,22%	4,46%	1,90%	1,36%	1,31%
	RW+	10,07%	6,63%	6,10%	2,43%	0,76%	9,24%	4,47%	1,91%	1,36%	1,31%

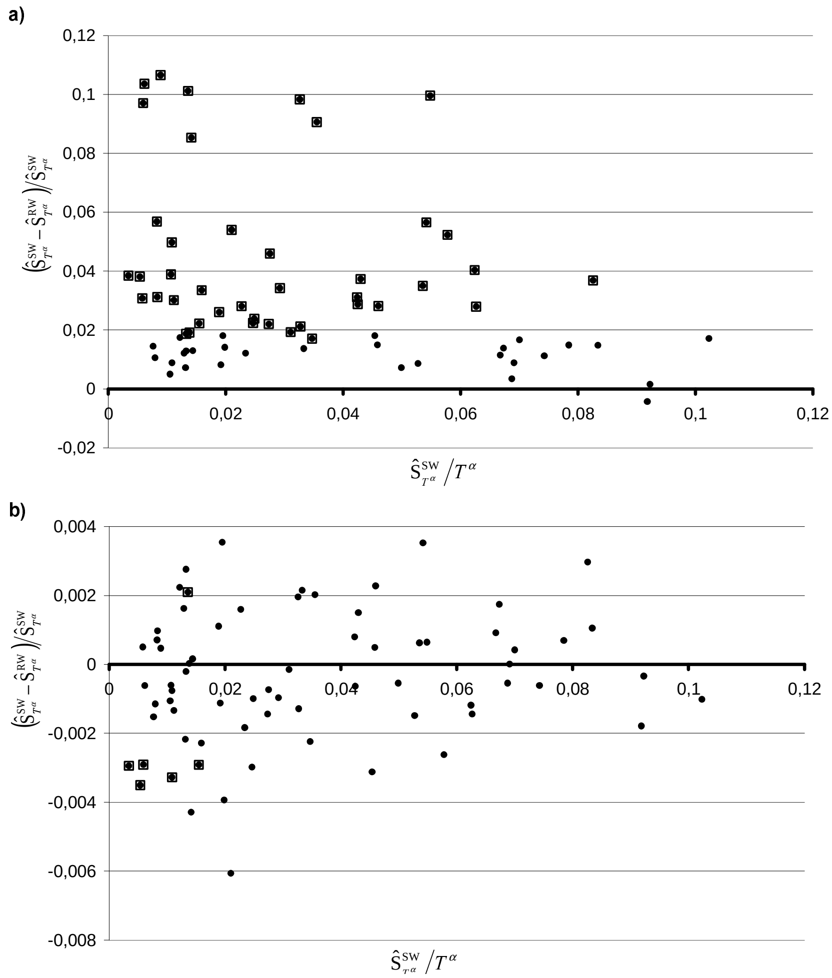

Figure 4: Relative difference of the standard deviations for the SW and RW estimator (a) and the RW and RW+ estimator (b), simulation 2.1

Table 13: Relative standard deviation for simulation 2.2

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	9,22%	6,03%	6,67%	3,16%	0,91%	7,52%	3,29%	2,30%	1,54%	1,67%
	RW	9,62%	6,57%	6,77%	3,18%	0,94%	7,93%	3,31%	2,29%	1,55%	1,74%
	RW+	9,50%	6,42%	6,71%	3,17%	0,94%	7,93%	3,36%	2,31%	1,56%	1,77%
Lower secondary	SW	7,19%	5,12%	5,13%	2,37%	0,67%	8,40%	2,73%	1,76%	1,21%	1,31%
	RW	8,25%	5,71%	5,38%	2,39%	0,70%	8,92%	2,83%	1,79%	1,22%	1,33%
	RW+	8,09%	5,56%	5,34%	2,38%	0,69%	9,00%	2,87%	1,80%	1,22%	1,33%
Upper secondary C	SW	8,85%	5,28%	6,04%	3,65%	1,22%	7,45%	2,94%	2,21%	1,46%	1,41%
	RW	9,70%	5,40%	6,17%	3,65%	1,21%	8,45%	3,17%	2,26%	1,47%	1,44%
	RW+	9,83%	5,49%	6,21%	3,66%	1,21%	8,33%	3,12%	2,27%	1,47%	1,44%
Upper secondary A/B	SW	9,51%	5,78%	7,15%	3,95%	1,24%	7,95%	3,87%	2,63%	1,73%	1,54%
	RW	9,78%	6,11%	7,13%	3,95%	1,25%	8,08%	4,04%	2,62%	1,71%	1,57%
	RW+	9,73%	6,21%	7,12%	3,95%	1,27%	8,13%	4,00%	2,64%	1,72%	1,56%
Tertiary C	SW	5,95%	3,82%	3,44%	1,51%	0,38%	4,41%	1,68%	0,97%	0,68%	0,64%
	RW	7,46%	4,36%	3,50%	1,53%	0,43%	5,15%	1,87%	1,05%	0,69%	0,69%
	RW+	7,59%	4,30%	3,53%	1,54%	0,43%	5,18%	1,94%	1,04%	0,69%	0,69%
Tertiary B	SW	10,04%	4,96%	5,49%	2,23%	0,64%	5,34%	2,44%	1,27%	0,95%	0,89%
	RW	11,08%	5,22%	5,36%	2,21%	0,66%	6,04%	2,65%	1,30%	0,95%	0,93%
	RW+	11,29%	5,39%	5,39%	2,24%	0,66%	5,97%	2,61%	1,30%	0,94%	0,94%
Tertiary A and higher	SW	12,44%	8,30%	7,16%	3,04%	0,87%	11,10%	5,50%	2,26%	1,53%	1,53%
	RW	13,56%	8,19%	7,18%	3,03%	0,86%	11,33%	5,57%	2,26%	1,56%	1,53%
	RW+	12,91%	8,20%	7,15%	3,03%	0,86%	11,39%	5,57%	2,27%	1,58%	1,54%

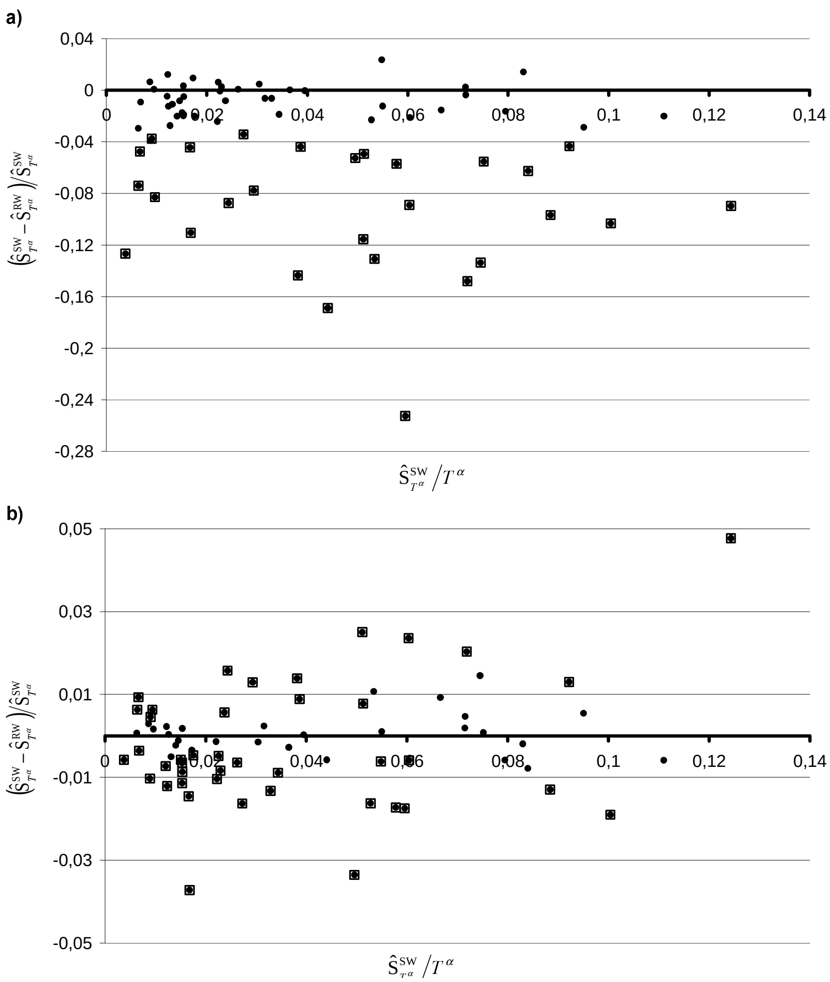

Figure 5: Relative difference of the standard deviations for the SW and RW estimator (a) and the RW and RW+ estimator (b), simulation 2.2

Table 14: Relative standard deviation for simulation 2.3

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	9,22%	6,03%	6,67%	3,16%	0,91%	7,52%	3,29%	2,30%	1,54%	1,67%
	RW	9,15%	6,04%	6,54%	3,12%	0,91%	7,49%	3,23%	2,26%	1,52%	1,67%
	RW+	9,14%	6,01%	6,54%	3,12%	0,91%	7,48%	3,24%	2,26%	1,53%	1,67%
Lower secondary	SW	7,19%	5,12%	5,13%	2,37%	0,67%	8,40%	2,73%	1,76%	1,21%	1,31%
	RW	7,26%	5,12%	5,08%	2,32%	0,67%	8,37%	2,70%	1,72%	1,19%	1,29%
	RW+	7,23%	5,10%	5,07%	2,33%	0,66%	8,38%	2,70%	1,72%	1,19%	1,29%
Upper secondary C	SW	8,85%	5,28%	6,04%	3,65%	1,22%	7,45%	2,94%	2,21%	1,46%	1,41%
	RW	8,79%	5,18%	5,99%	3,61%	1,21%	7,40%	2,93%	2,18%	1,45%	1,39%
	RW+	8,81%	5,18%	6,00%	3,61%	1,20%	7,38%	2,93%	2,19%	1,45%	1,39%
Upper secondary A/B	SW	9,51%	5,78%	7,15%	3,95%	1,24%	7,95%	3,87%	2,63%	1,73%	1,54%
	RW	9,40%	5,82%	7,07%	3,90%	1,23%	7,84%	3,87%	2,60%	1,70%	1,52%
	RW+	9,40%	5,84%	7,05%	3,90%	1,23%	7,85%	3,87%	2,61%	1,71%	1,52%
Tertiary C	SW	5,95%	3,82%	3,44%	1,51%	0,38%	4,41%	1,68%	0,97%	0,68%	0,64%
	RW	5,99%	3,63%	3,29%	1,44%	0,38%	4,34%	1,65%	0,95%	0,66%	0,62%
	RW+	6,01%	3,63%	3,29%	1,44%	0,38%	4,34%	1,65%	0,95%	0,67%	0,62%
Tertiary B	SW	10,04%	4,96%	5,49%	2,23%	0,64%	5,34%	2,44%	1,27%	0,95%	0,89%
	RW	10,02%	4,90%	5,29%	2,18%	0,63%	5,28%	2,41%	1,24%	0,92%	0,88%
	RW+	10,05%	4,92%	5,29%	2,18%	0,63%	5,29%	2,42%	1,24%	0,93%	0,88%
Tertiary A and higher	SW	12,44%	8,30%	7,16%	3,04%	0,87%	11,10%	5,50%	2,26%	1,53%	1,53%
	RW	12,60%	8,11%	7,04%	3,00%	0,86%	11,10%	5,39%	2,23%	1,53%	1,51%
	RW+	12,47%	8,11%	7,03%	3,00%	0,86%	11,10%	5,41%	2,23%	1,54%	1,51%

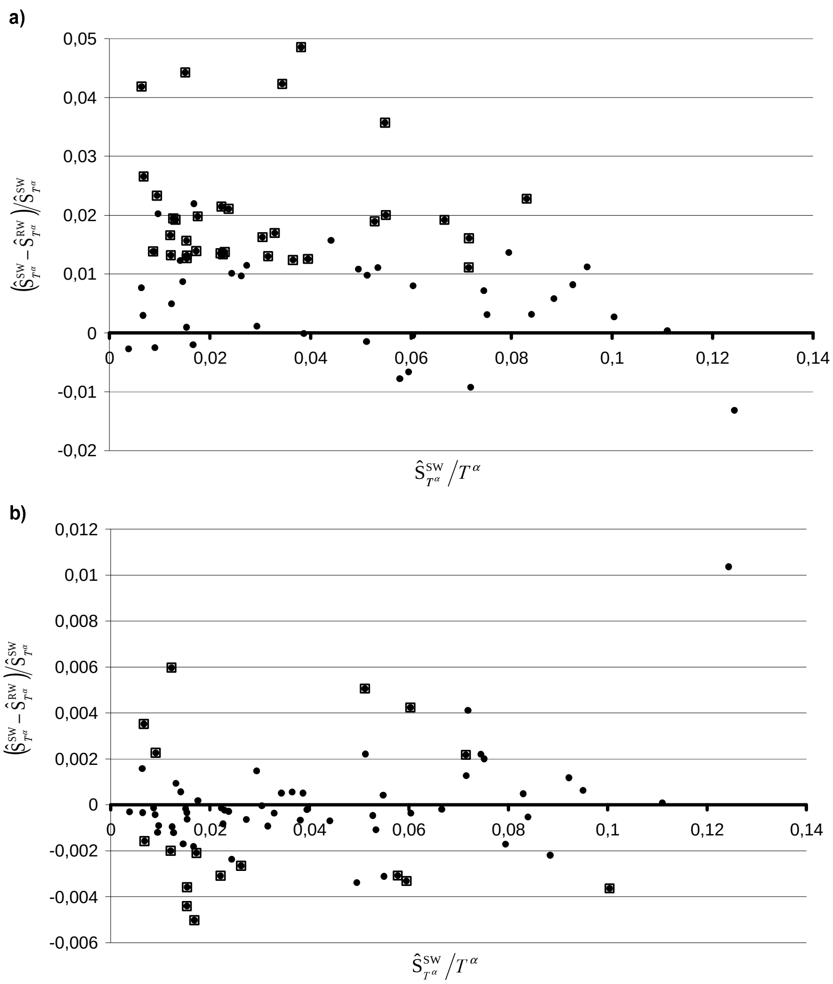

Figure 6: Relative difference of the standard deviation for the SW and RW estimator (a) and the RW and RW+ estimator (b), simulation 2.3

Table 15: Relative standard deviation for simulation 2.4

Gender x Working hours		Male					Female				
Education Level		<4	4-12	12-20	20-35	35+	<4	4-12	12-20	20-35	35+
Primary or less	SW	18,81%	14,03%	13,12%	5,77%	1,75%	15,81%	6,32%	3,89%	2,47%	2,84%
	RW	19,01%	13,18%	11,87%	5,52%	1,72%	16,39%	6,07%	3,79%	2,42%	2,81%
	RW+	18,93%	13,18%	12,00%	5,53%	1,71%	16,45%	6,13%	3,84%	2,43%	2,83%
Lower secondary	SW	23,92%	12,99%	8,73%	4,08%	1,29%	18,65%	4,83%	3,90%	2,18%	2,67%
	RW	19,38%	12,10%	8,10%	3,87%	1,21%	17,72%	4,73%	3,48%	2,05%	2,53%
	RW+	20,40%	12,12%	8,10%	3,88%	1,21%	17,76%	4,80%	3,54%	2,06%	2,54%
Upper secondary C	SW	24,13%	12,68%	10,31%	6,06%	2,43%	25,48%	7,86%	4,49%	2,45%	3,11%
	RW	20,64%	12,39%	9,92%	5,99%	2,33%	21,09%	7,32%	4,26%	2,31%	2,89%
	RW+	21,54%	12,36%	9,98%	6,01%	2,32%	21,79%	7,37%	4,28%	2,32%	2,92%
Upper secondary A/B	SW	22,85%	12,19%	15,03%	6,92%	2,42%	21,08%	11,20%	4,26%	2,91%	3,41%
	RW	22,32%	12,02%	14,46%	6,69%	2,38%	20,51%	10,44%	4,17%	2,82%	3,14%
	RW+	22,59%	12,04%	14,60%	6,72%	2,37%	20,58%	10,58%	4,18%	2,83%	3,20%
Tertiary C	SW	16,86%	10,26%	6,76%	2,68%	0,78%	10,59%	3,50%	1,97%	1,20%	1,44%
	RW	12,61%	7,84%	5,75%	2,31%	0,67%	9,00%	3,10%	1,64%	1,01%	1,14%
	RW+	13,12%	7,87%	5,80%	2,31%	0,67%	9,13%	3,15%	1,65%	1,02%	1,15%
Tertiary B	SW	21,28%	11,46%	9,31%	3,82%	1,22%	20,37%	6,45%	2,29%	1,64%	1,74%
	RW	21,36%	11,06%	8,74%	3,63%	1,09%	15,51%	5,59%	2,18%	1,54%	1,60%
	RW+	21,76%	11,23%	8,76%	3,66%	1,09%	15,81%	5,68%	2,19%	1,55%	1,63%
Tertiary A and higher	SW	25,95%	18,38%	14,40%	5,09%	1,86%	16,21%	12,20%	5,45%	2,69%	3,17%
	RW	26,83%	18,07%	13,81%	4,98%	1,74%	17,00%	10,41%	5,01%	2,67%	3,02%
	RW+	26,34%	18,07%	13,77%	5,01%	1,74%	17,01%	10,82%	5,07%	2,68%	3,04%

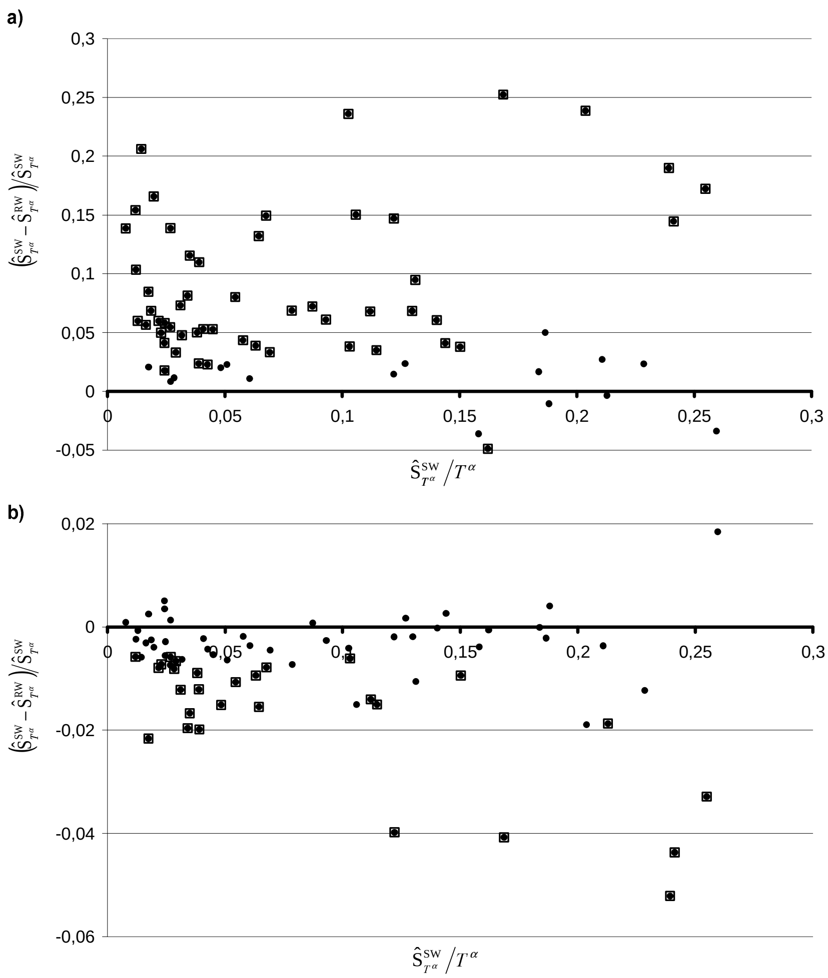

Figure 7: Relative difference of the standard deviations for the SW and RW estimator (a) and the RW and RW+ estimator (b), simulation 2.4

VRD coverage probability

Figures 8-12 show, for each cell of the table [$G \times H \times E$] x Av(W), what percentage $P_{\text{cell}}^{\text{VRD}}$ of the 600 estimates made in the simulation were accurate to within the margin calculated by VRD. The coverage probability $P_{\text{cell}}^{\text{VRD}}$ is plotted versus the relative SW standard deviation of the cell. Boxes indicate significant deviations of $P_{\text{cell}}^{\text{VRD}}$ from the nominal value of 95%. In figure 13, the coverage probability for all five simulations is plotted as a function of the effective cell size.

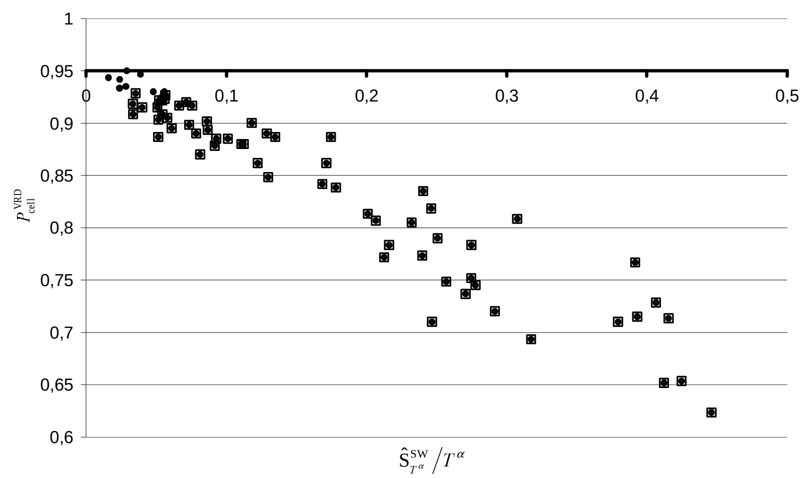

Figure 8: VRD coverage probability for simulation 1

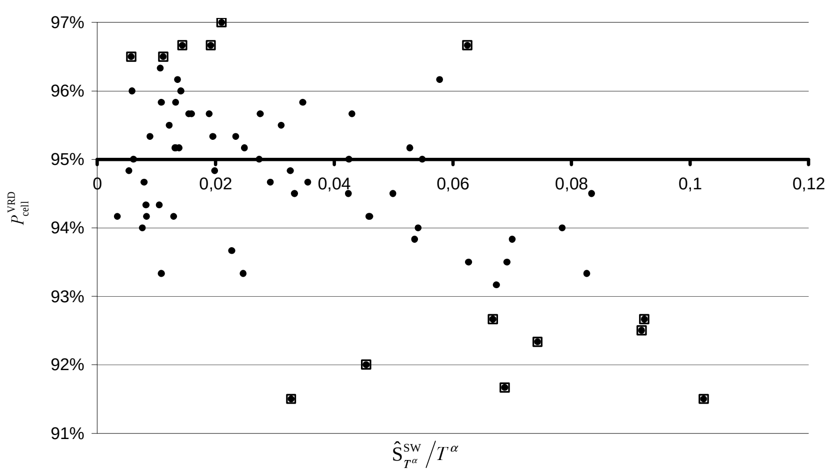

Figure 9: VRD coverage probability for simulation 2.1

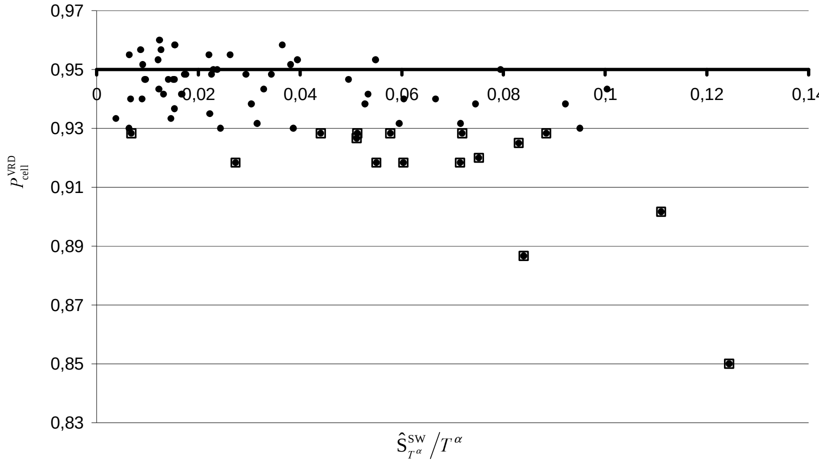

Figure 10: VRD coverage probability for simulation 2.2

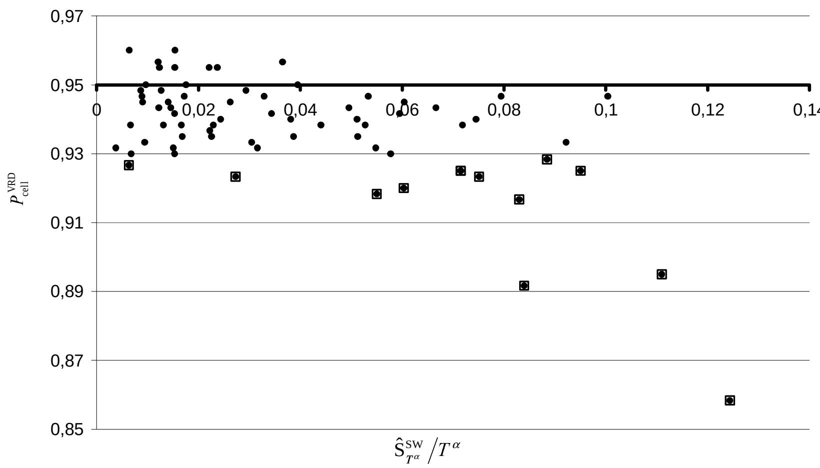

Figure 11: VRD coverage probability for simulation 2.3

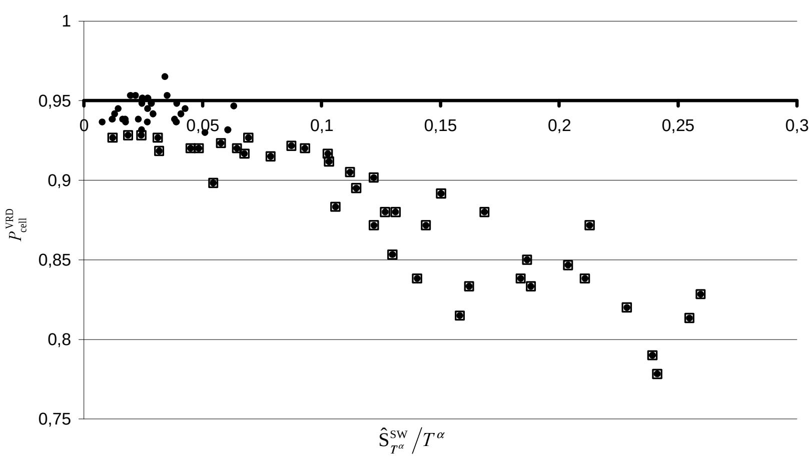

Figure 12: VRD coverage probability for simulation 2.4

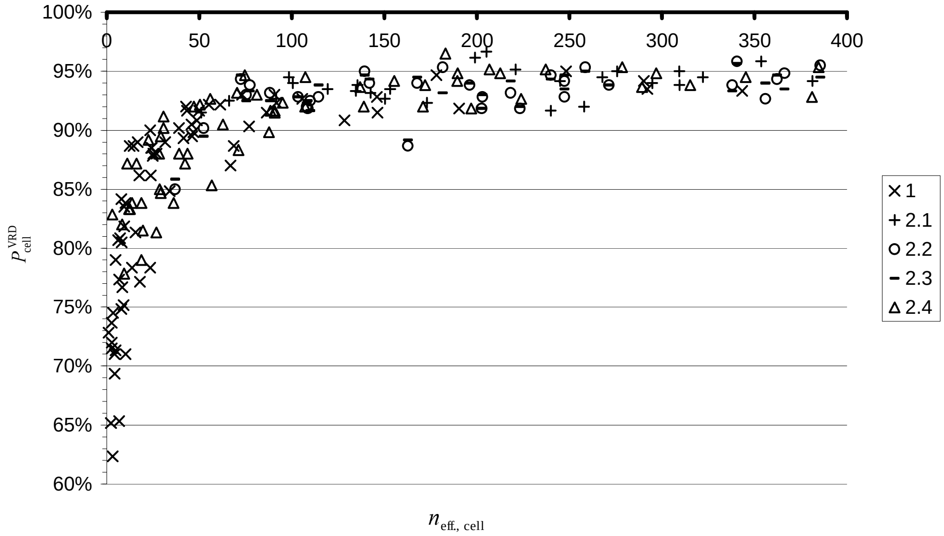

Figure 13: VRD coverage probability versus $n_{eff.,cell}$ for all five simulations

References

Houbiers, M. (2002), Variantieschatting bij herhaald wegen –de LSO-tabellen (Variance estimation for repeated weighting –the SES tables), Research paper, Statistics Netherlands.

Houbiers, M., Knottnerus, P., Kroese, A.H., Renssen, R.H. and Snijders, V. (2003). Estimating consistent table sets: position paper on repeated weighting, Discussion paper, Statistics Netherlands, in preparation.

Houbiers, M. and V. Snijders (2002), Gebruikershandleiding VRD 1.4 (User's manual VRD 1.4), Research paper, Statistics Netherlands.

Kish, L. (1992), Weighting for Unequal P_i, Journal of Official Statistics, 8, 183-200.

Knottnerus, P. (2001), Varianties bij herhaald wegen (Variances in repeated weighting), Research paper, Statistics Netherlands.

Knottnerus, P. (2003 A) , Repeated weighting and the variance of its estimator, paper presented at the 54th session of the International Statistics Institute (ISI), Berlin.

Knottnerus, P. (2003 B) Sample Survey Theory, Some Pythagorean Perspectives, Springer-Verlag.

Kroese, A.H. and Renssen, R.H. (1999), Weighting and imputation at Statistics Netherlands, Proceedings of the IASS Conference on Small Area Estimation, Riga, pp.190-120.

Särndal, C., B. Swensson and J. Wretman (1992), Model Assisted Survey Sampling, Springer-Verlag.

Snijders, V. and M. Houbiers (2002), Variantieschatting bij herhaald wegen voor VRD 1.4 (Variance estimation for repeated weighting with VRD 1.4), Research paper, Statistics Netherlands .