DACSEIS
IST-2000-26057

Workpackage 12

The DACSEIS Recommended Practice Manual

Deliverables 12.2 and 12.3

2004

List of contributors:

Paul Knottnerus, CBS;
Anthony Davison, Sylvain Sardy, EPFL;
Chris Skinner, University of Southampton;
Pauli Ollila, Statistics Finland;
Ralf Münnich, University of Tübingen

Main responsibility:

Anthony Davison, EPFL;
Ralf Münnich, University of Tübingen;
Chris Skinner, University of Southampton

IST–2000–26057–DACSEIS

The DACSEIS research project is financially supported within the IST programme of the European Commission. Research activities take place in close collaboration with Eurostat.

http://europa.eu.int/comm/eurostat/research/

http://www.cordis.lu/ist/

http://www.dacseis.de/

Preface

Deliverable D12.3 presents *best practice recommendations* on the use of variance estimators derived from DACSEIS research. It is intended to guide the interested user through the many results of the simulation study performed by the DACSEIS teams, and to furnish recommendations on the use of repeated weighting, criteria for combining results from national surveys, and on software packages for application of existing variance estimation methods.

The recommended practice manual D12.3 is accompanied by an electronic version as deliverable D12.2, which contains the simulation results from the main simulation study and from specialised simulations from several workpackages.

Chapter 1 was written by Anthony Davison, Sylvain Sardy, and Ralf Münnich. Anthony Davison, Chris Skinner, and Ralf Münnich are responsible for Chapter 2. Paul Knottnerus contributed Sections 3.1 and 3.2, and Pauli Ollila Section 3.3. Special thanks go to Kersten Magg for his contributions to the appendix.

The work presented as deliverable D12.3 was highly influenced by fruitful discussions within the DACSEIS team, and also on many occasions where DACSEIS results were presented. The authors thank all who contributed to the success of the project, particularly the officials of Eurostat and the European Commission for their co-operation and firm support of the project.

On behalf of the DACSEIS team,
Ralf Münnich (co-ordinator DACSEIS), Tübingen, July 2004

Contents

List of Figures

Chapter 1

Selected Results from the National Survey Simulations

The DACSEIS simulation study was conducted on the six DACSEIS survey files which are based on close-to-reality data. Each of the universe files was used in the general simulation study but has also its own specialised purpose. Within this chapter, selected simulation results are presented and described, in order to present representative and interesting results from the different surveys.

An overview to the surveys, the corresponding sampling schemes, and their peculiarities can be drawn from the report on workpackage 2. The survey data are described in the workpackage 3 report. The set-up of the Monte-Carlo study can be drawn from Chapter 1 of deliverable D1.2. A detailed overview of the simulation tasks conducted in the DACSEIS simulation study can be found in Appendix A, which also includes a description of how to read the simulation tasks in the electronic version of the DACSEIS recommended practice manual (RPM) (deliverable D12.2) where all results are presented. Further, it contains references to the detailed descriptions of the estimators and surveys, including all characteristics in the DACSEIS deliverables.

The types of graph in the following subsections are all equivalent and so are described in detail for the Austrian Microcensus, as well in Appendix A.

1.1 Austrian Microcensus

In the Austrian Microcensus, mainly Horvitz–Thompson related estimators were applied. The examples in the recommended practice manual show comparisons between Burgenland (BGL), where the entire two-stage scheme has to be considered, and Vorarlberg (VBG), for which a stratified sampling scheme is used. The target variable is the total number of employed; details of the set-ups can be drawn from deliverable D12.2.

Figure 1.1 shows the influence of the different imputation methods on the point and variance estimators. The standard bootstrap with Shao/Sitter correction (cf. deliverables D5.1 and D5.2 or Shao and Sitter, 1996) tends to overestimate the true variance in Vorarlberg, whereas a small underestimation was observed in Burgenland. In both cases

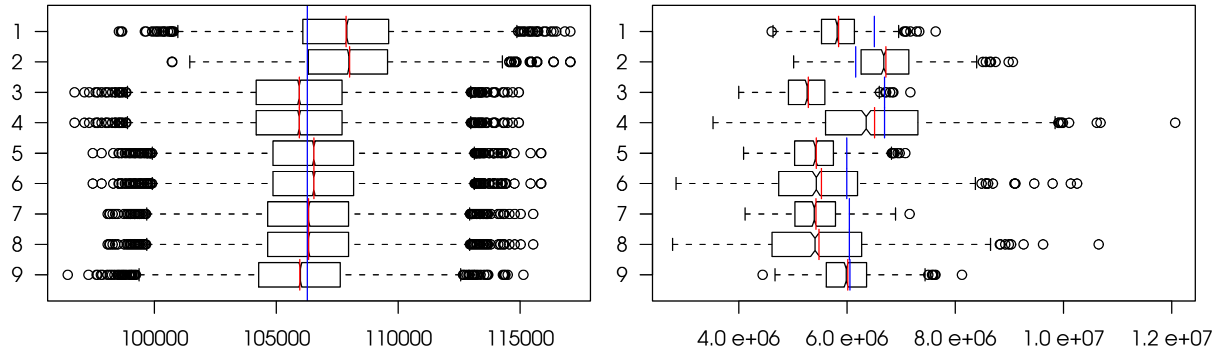

Figure 1.1: Point (left) and variance (right) estimators in Vorarlberg (employed; variable set 27; 25% non-response rate)

the best variance estimation under non-response and imputation is given using multiple imputation with logit imputation. In terms of unbiasedness of the estimators, the `SIlfs4` imputation gave the best results within this comparison.

The estimators applied are the Horvitz–Thompson estimator with two weighting schemes (1) and (2), the Horvitz–Thompson estimator with logit imputation, no non-response correction (3) and Shao/Sitter-correction (4), with `SIlfs1` imputation (5) and (6), and with `SIlfs4` imputation (7) and (8), and with multiple imputation (9).

The blue line in the left panel of Figure 1.1 indicates the true total in the synthetic AMC universe. In the case of the variance estimator, the simulated variance of the point estimator in indicated by the blue line, which is the reference value for the variance estimator. The red lines indicate the corresponding estimates.

The corresponding values of the standard measures are shown in Figure 1.2. The left panel presents the relative root mean squared error (RRMSE) (blue) of the true values and of the simulated values (green) and the average of the estimated relative standard error (red). The right panel presents the achieved coverage rates within the simulation for 90% (red) and 95% (blue) confidence intervals. In both graphs, only small biases, due to underestimation, of point and variance estimator caused deviations from the expected results. In general the estimated relative standard error overestimated the true relative MSE.

Even for 40% non-response one obtains similar results. The Shao/Sitter correction works well in all cases and adjusts for the huge underestimation when no correction is applied.

The logit imputation, for single and multiple imputation, did not show convergence problems. However, the variance estimator under single imputation showed slightly higher variability in comparison to multiple imputation.

The specialized weighting scheme to correct for non-response shows adequate variance estimators but is severely biased.

The bootstrap Shao/Sitter is very slow — more than 200 seconds were typically needed for each simulation run. The specialised weighting is about equally as fast as single imputation, without resampling, all taking less than 10 seconds. Multiple imputation took close to 50 seconds per simulation.

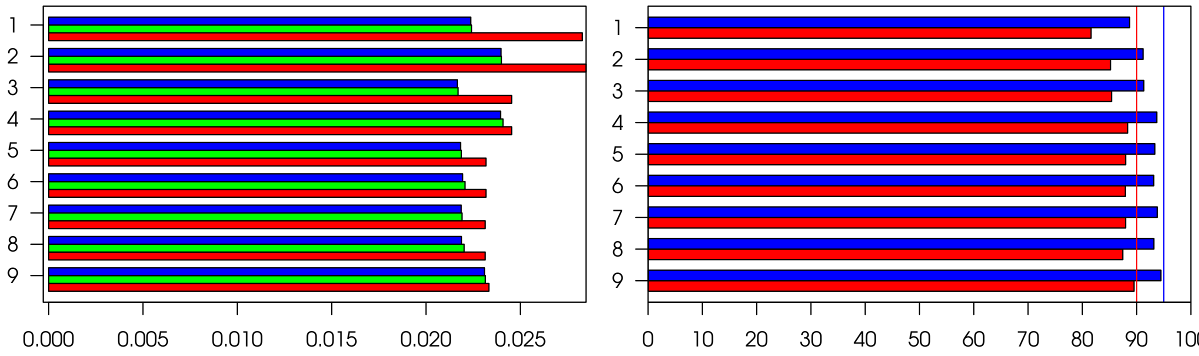

Figure 1.2: Point and variance estimators in Vorarlberg

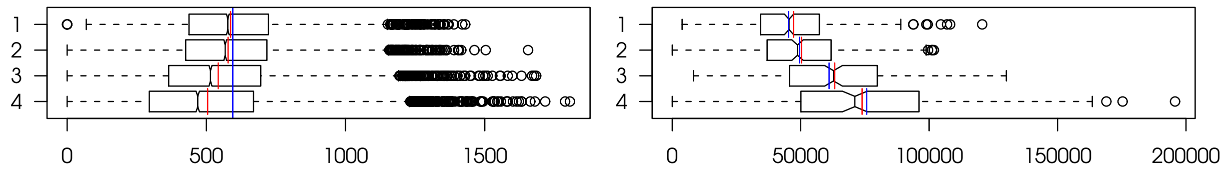

Figure 1.3: Simulation results for Horvitz–Thompson estimation of the total of unemployed with `SIlfs4` imputation and bootstrap Shao/Sitter in Noord-Brabant; comparison of the estimators with 5, 10, 25, and 40% non-response

1.2 Dutch Labour Force Survey

Multiple imputation (MI) fails badly in a subregion of of Noord-Brabant, either with the GREG or with the Horvitz–Thompson estimators; the MI variances can severely under- or over-estimate the true variance of the method. The same simulation performed significantly better in Amsterdam (a subregion of North-Holland), in terms of finding adequate variance estimates. However, multiple imputation is outperformed by raking and calibration methods on the one hand and single imputation with bootstrap Shao/Sitter on the other hand.

Raking works well, giving the best point and variance estimates overall in the simulation on the Dutch Labour Force Survey.

In the Dutch Labour Force Survey some difficulties arose with biased results. It seems that here the most sophisticated non-response mechanism badly influenced the estimation results. However, the variance estimators turned out to be appropriate.

The following example shows the appropriateness of the Shao/Sitter correction applied to the Horvitz–Thompson estimator with `SIlfs4` imputation; see Figure 1.3. The point estimator tends to be biased with considerable non-response rates. The variance estimator yields very good results for all the non-response rates we applied.

The bias of the estimator causes some loss for the standard measures (cf. section 1.1) — see Figure 1.4.

The repeated weighting estimators are presented in workpackage 7 and the corresponding simulation results are given in Section 2.1 of deliverable D1.2.

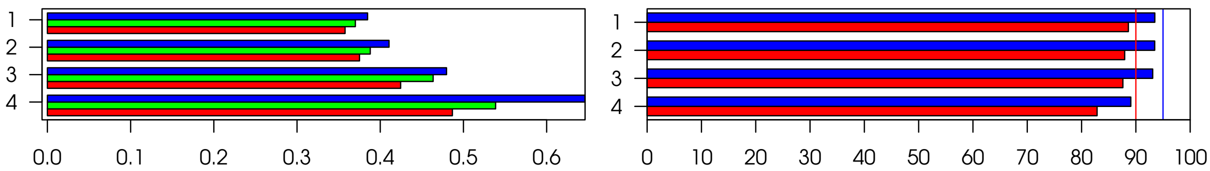

Figure 1.4: Simulation results for Horvitz–Thompson estimation of the total of unemployed with `SIlfs4` imputation and bootstrap Shao/Sitter in Noord-Brabant; comparison of the estimators with 5, 10, 25, and 40% non-response

1.3 Finnish Labour Force Survey

The main aim in using the Finnish Labour Force Survey, for which data were available in three waves, was to evaluate the different variance estimators for change developed under workpackage 9. A detailed description of the estimators and the simulation results is given in deliverable D9.1. The entire simulation results can be viewed in the recommended practice manual D12.2.

In addition to variance estimation for change, three different non-response mechanisms were implemented. However, there are almost no differences in the accuracy of the estimators obtaining use the three mechanisms implemented.

The variance estimators show similar results in the other surveys. However, the point estimators are slightly biased in the present simulations. The direct variance estimator of the standard calibration estimator badly underestimates the true variance without non-response correction. The Shao/Sitter bootstrap again yields very good variance estimation results, but again the computation times make alternatives attractive if they are available. Applying `SIlfs1` uses approximately 100 seconds per run, whereas `SIlfs4` needs almost 400 seconds. The complexity of the imputation in connection with the sample size the Shao/Sitter correction plays a key role in the applicability of these resampling methods.

Surprisingly, the delete-d-jackknife does not perform well in the Finnish Labour Force Survey. This phenomenon was observed in most discrete-type simulations with high non-response rates. Even if further investigation seem necessary, one may conclude that it is more difficult to choose suitable parameters for implementing the delete-d-jackknife. Moreover the bootstrap outperformed the jackknife in all the simulations.

1.4 German Sample Survey of Income and Expenditure

The following example compares ratio imputation with linear regression imputation in connection with Horvitz–Thompson and GREG estimators for the total income in Saarland. For each combination, no non-response correction, bootstrap Shao/Sitter, and delete-d-jackknife with Rao/Shao correction (RAO and SHAO, 1992) was applied (Horvitz–Thompson, ratio: 1–3; Horvitz–Thompson, linear regression: 4–6; corresponding GREG,

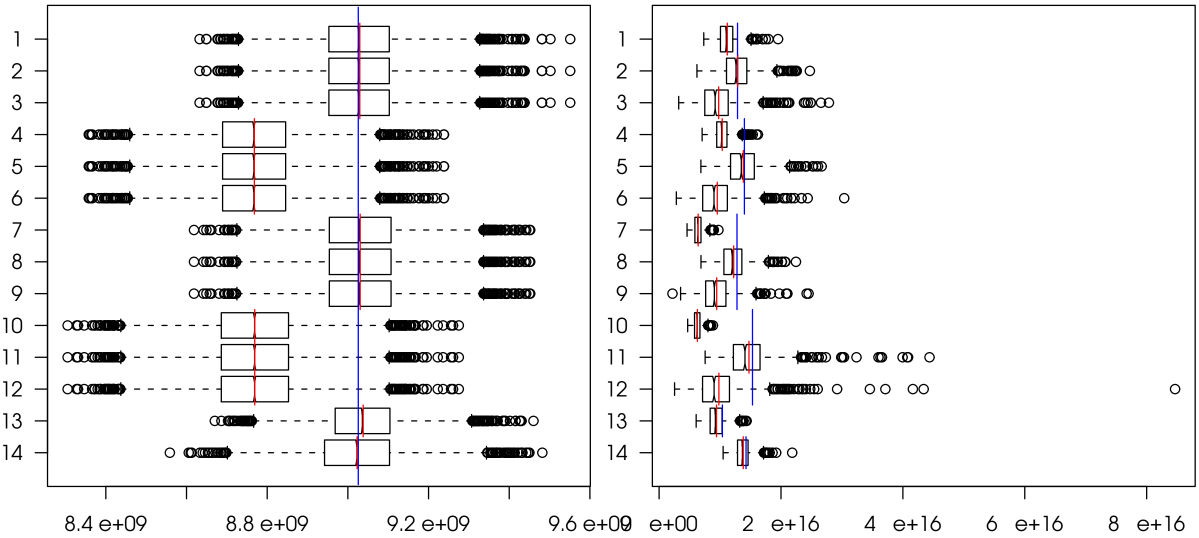

Figure 1.5: Point and variance estimation results in Saarland with 25% non-response for Horvitz–Thompson and GREG estimation with ratio and regression imputation, compared to raking and Lundstrøm–Särndal.

7–12). In order to complete the comparison, the Lundstrøm–Särndal formula (13) and the raking with jackknife linearisation (14) was applied.

In the case of the German Sample Survey of Income and Expenditure no non-response mechanism was available, so the figures from the Swiss Household and Budget Survey were (suitably) adapted to the German data.

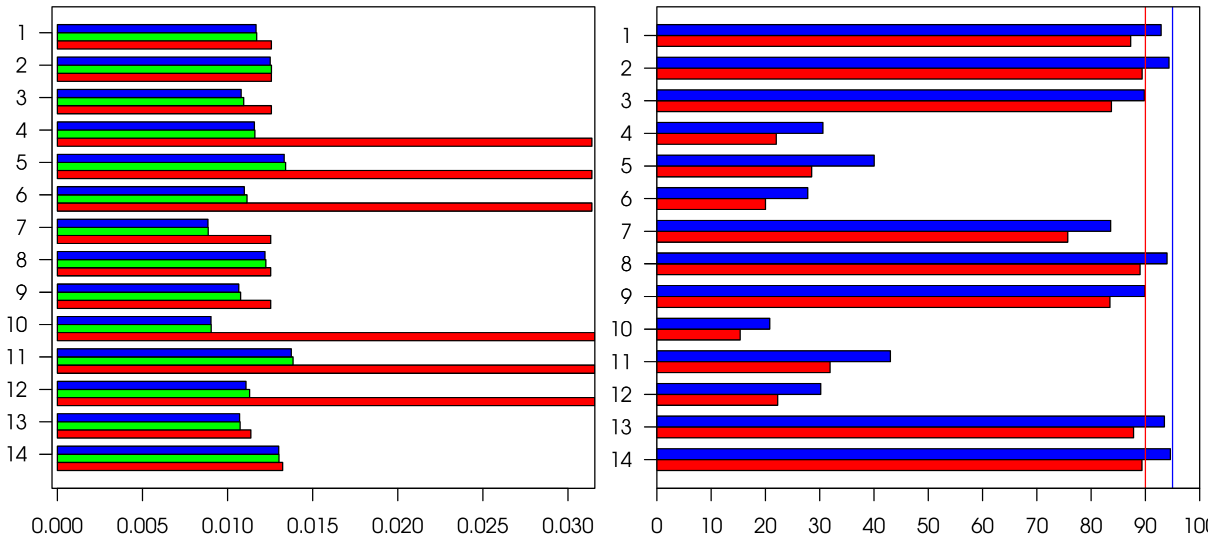

Figure 1.6: Measures in Saarland with 25% non-response, for Horvitz–Thompson and GREG estimation with ratio and regression imputation in comparison to raking and Lundstrøm–Särndal variance estimation.

Figure 1.5 suggests that we should prefer ratio imputation to regression imputation. Here, the bootstrap Shao/Sitter with ratio imputation seems to yield results comparable to raking with a linearized variance estimator. The Lundstrøm–Särndal estimator seems to produce slightly smaller variances than the other estimators, but has a small bias.

The delete-*d*-jackknife was again outperformed by the bootstrap variance estimator.

1.5 German Microcensus

There are many simulation results for the German Microcensus. The main comparisons were performed on the federal state Saarland (SAL), some using subgroups defined according to the strata on regional level (sub1) or house size level (sub5). The subgrouping was necessary in order to obtain results for the very time-consuming resampling-based variance estimators, but also yields valuable information for the other estimators.

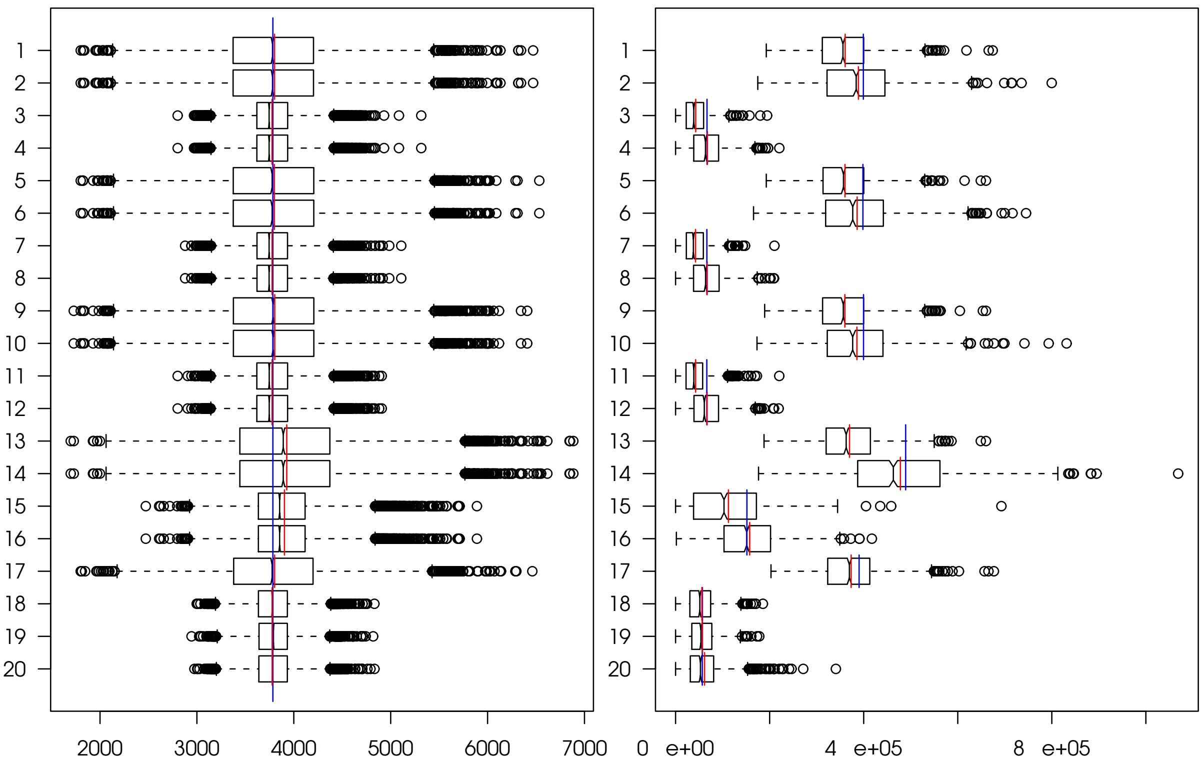

Figure 1.7: Point and variance estimation results for the Horvitz–Thompson and GREG estimators of the total unemployed in Saarland with 25% non-response, with several single imputation methods compared to multiple imputation, raking, and the Lundstrøm–Särndal method.

The main results are similar to those for the other surveys. The following example compares the performance of the single imputation methods `SIlfs1` (1–4), `SIlfs2` (5–8), `SIlfs3` (9–12) and `SIlfs4` (13–16), each with Horvitz–Thompson and GREG (without and with bootstrap non-response correction), a modified multiple imputation with a linear regression nearest neighbour technique (cf. Münnich and Rässler, 2004; HT:

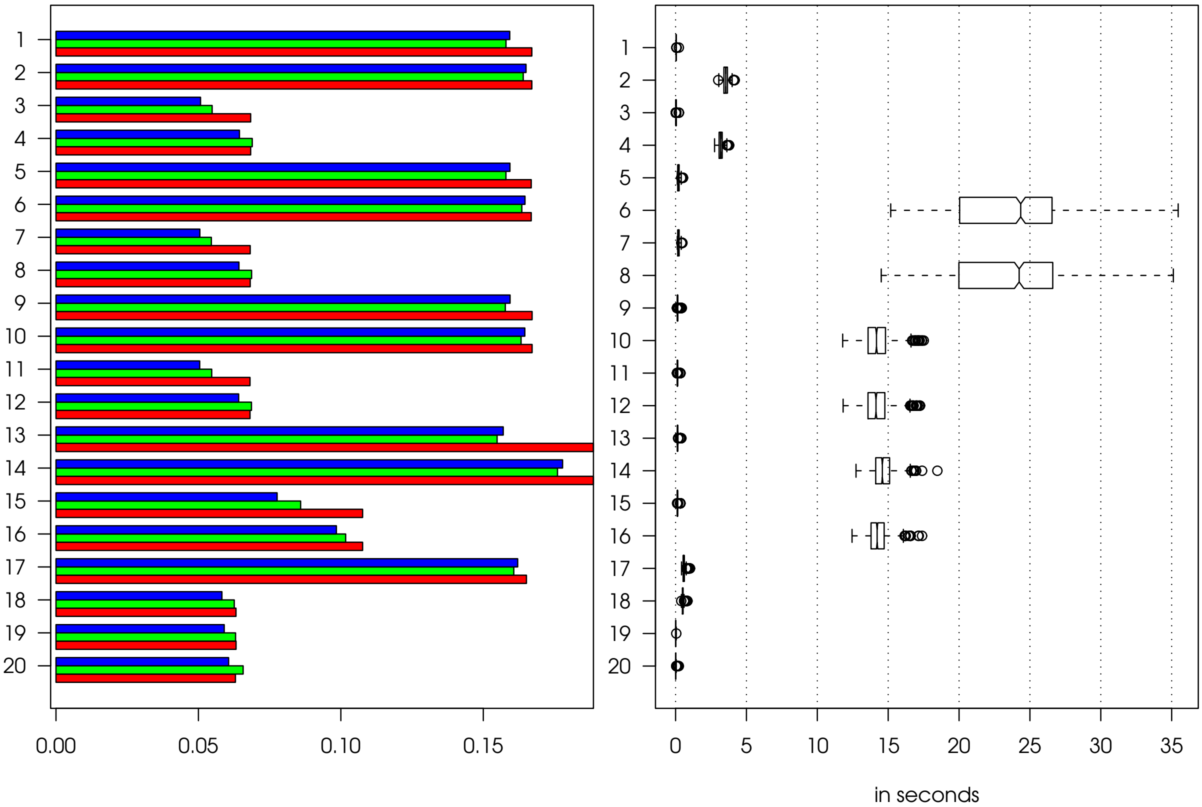

Figure 1.8: Measures and time consumption (variance estimates) for the Horvitz–Thompson and GREG estimators with several single imputation methods in comparison to multiple imputation, raking and Lundstrøm–Särndal methods; data for Saarland with 25% non-response.

17, GREG: 18), a raking with jackknife linearisation variance estimator (19), and the Lundstrøm–Särndal formula (20).

The bootstrap Shao/Sitter works generally fine for all the various imputation methods. However, in this example it tends to slightly underestimate the true variance when Horvitz–Thompson estimators are used. The other methods seem to perform similarly to the Shao/Sitter approach applied to the GREG estimator. However, small advantages can be seen for the raking-with-linearisation variance estimator and multiple imputation, while the Lundstrøm–Särndal formula shows in several cases a slightly larger variance of the variance estimator.

The logit imputation failed in this subgroup due to severe problems in estimating the parameters of the underlying logit model. This was observed for both single and multiple imputation, and can lead to very poor point and variance estimates. Further, the jackknife variance estimators which were omitted in the above graphs again overestimated the true variance severely, especially with higher non-response rates.

The measures from this example are exactly as expected. The computer efficiency also turns out as expected. However, huge differences occur, depending on the imputation model. In real applications, it would be appropriate to spend considerable effort in finding a suitable imputation model.

1.6 Swiss Household and Budget Survey

In the Swiss Household and Budget Survey, similar results were obtained in comparison to the other surveys. However, some differences for point estimators were observed which depend on the choice of the nine implemented non-response mechanism; for details see deliverable D1.1, Section 5.5. To overcome the non-response bias, a special weighting scheme was implemented. This study can be drawn from Section 2.3 in deliverable D1.2.

The jackknife is unstable but the bootstrap, linearisation, and Lundstrøm–Särndal formulae are all very comparable in terms of stability. The Lundstrøm–Särndal is quickest by a factor of three, but otherwise not preferable to linearisation. This was true for all levels of missingness.

Multiple imputation is performed using the linearisation formula once the imputation data were available, with five imputations. It gave results comparable with the Lundstrøm–Särndal formula and linearisation in terms of bias and stability. The computational cost was quite low.

Chapter 2

DACSEIS Recommended Practices

2.1 Recommendations for the Use of the Variance Estimation Methods

This section contains the main conclusions from the DACSEIS simulation studies, which are given as recommendations later. The following sections are based on the results of the simulation study and so make reference to the results from the previous chapter and the electronic *recommended practice manual*.

2.1.1 Analytical Formulae

The calibration and raking estimators including the appropriate variance estimators (cf. deliverables D8.1 and D8.2) developed as part of workpackage 8 generally provided the best estimators of totals. The `Horvitz-Thompson` estimator formula with GREG g-weights including the jackknife linearisation variance estimator seems to be best of these overall: such variance estimators are generally fastest to compute and seem to have no disadvantage in terms of lack of stability or of coverage error. The version of this with maximum likelihood calibration performs about the same as the others in terms of variance and confidence interval coverage, but is distinctly slower; there seems to be no reason to use it in preference to the faster essentially equivalent procedures, unless positive weights are regarded as essential.

The Lundstrøm–Särndal formula often behaves well and is comparable to the aforementioned calibration methods with linearized variance estimators. In the few cases where the variance estimation formulae for the raking estimators behave relatively poorly, the Lundstrøm–Särndal formula can provide an improvement. The Swiss Household and Budget Survey results suggest that jackknife linearisation gives similar results to the Lundstrøm–Särndal formula, but for that survey the Lundstrøm–Särndal formula was three times faster.

2.1.2 Multiple Imputation

The performance of multiple imputation depends heavily on the imputation model used. In particular, the logit model used for imputation of categorical variables can yield variance estimates that are very poor; our simulations suggest that this method of imputation should be used carefully. The problem generally arises in small samples with very few observations and is the result of poorly converging maximum likelihood estimates for the logit model — compare MÜNNICH and RÄSSLER (2004) or the *recommended practice manual*, the German Microcensus with the target variable *unemployed* in Saarland against the subregions `sub1` and `sub5`, or the target variable *unemployed female, age 65+*.

In cases where multiple imputation with the logit model produces stable variance estimates, the coverage of confidence intervals is generally correct, though with a slight tendency towards undercoverage due to underestimation of the variance. Increasing the number of imputations beyond 15 seems to make no difference to the variances, but reducing the number of imputations as low as five can yield slightly less stable variance estimates. The number of imputations has little discernable effect on the coverage of confidence intervals.

2.1.3 Resampling Variance Estimators

A disadvantage of the bootstrap is that compared to analytical formulae it is computationally much slower. On the other hand it generally produces adequate variance estimates even in the presence of appreciable non-response. Bootstrap variance estimates tend to be larger than the formula-based equivalents, thereby somewhat offsetting the tendency of the formulae towards undercoverage of confidence intervals.

There seems to be no reason to prefer the delete-d jackknife to the bootstrap. The latter is quicker to compute, produces more stable estimators of variance, and is less sensitive to the choice of imputation procedure when single imputation is used. Especially with the very heterogeneous strata in the German Microcensus, the delete-d jackknife was very sensitive to the imputation method and generally produced strong overestimates of the variances. Balanced repeated replication was eliminated from consideration at an early stage, since preliminary simulations revealed its high computational cost and lack of stability.

Figures 1.1, 1.5, and 1.7 show the effect of using different imputation schemes on the accuracy of the different variance estimates.

2.1.4 Influence of the Non-response Rate on Results

From the previous chapter, one might conclude that raising the non-response rate leads to an increase of the true variance of the estimator. However, the extent of increase including the effects on the different methods may vary from case to case — see also the discussion on the effective sample size and variance inflation in deliverable D11.1. Raking and calibration estimators with linearisation variance estimator seem less sensitive to the non-response rate than those resampling-based variance estimators which use single imputation. However, this effect depends on the single imputation routine used. The same effect can be found with multiple imputation, again partly because of problems in getting reasonable estimates for use in the logit imputation.

2.2 Conclusions and Recommendations

The DACSEIS simulation study includes a wide range of comparative studies for point and variance estimation methods as shown above and in the electronic *recommended practice manual*.

The main distinction was made in terms of treating non-response by

- weighting,
- single imputation, or
- multiple imputation.

In general, weighting is used for unit non-response and imputation is used for item non-response, but they are often not viewed as alternatives in household surveys.

In the DACSEIS simulation study, the focus was put on one target variable such that all three cases were directly competing with each other. Under these circumstances, one may conclude that raking and calibration estimators with classical residual or jackknife linearisation variance estimators should be preferred to resampling variance estimators or multiple imputation purely for computational reasons.

In order to apply multiple imputation, a proper imputation method must be available, in addition to an appropriate variance estimator for full datasets. Resampling methods generally require some information about the replicate structure of the data.

The overall conclusion appears to be that if (calibration) weighting can be used to handle non-response then the associated direct and linearisation variance estimation methods are satisfactory. On the other hand, if imputation is needed then variance estimation is more difficult. Variance estimation for single imputed data is often complex to implement for a given singly imputed dataset — for example, a new variance estimator may have to be constructed for each new point estimator, or the imputation method may need to be repeated on different subsamples. In contrast, the multiple imputation variance estimator is simple to implement once the multiple imputed datasets are constructed. The difficulty is in implementing proper multiple imputation in the face of complex sampling designs.

Chapter 3

Further DACSEIS Recommendations

3.1 Repeated Weighting for European Surveys

At present Statistics Netherlands uses the repeated weighting (RW) estimator in its regular estimation process to obtain consistency among tables. From the definition of the RW estimator, also given in Section 2.2.1 of this final report, it is clear that this estimator can also be applied in other European countries, e.g., in case of a register and a sample provided that:

1. the reference periods of the register and the survey are the same;
2. the register and the survey refer to the same population;
3. for the common variables in register and sample there are no definitional differences;
4. the categorical variables have hierarchical classifications, i.e., a class of a relatively fine classification always belongs to exactly one class of a less fine classification; and
5. there are enough survey data available so that the sample size for each (table) cell is large enough.

These assumptions illustrate that the harmonization of surveys and registers plays a key role during all phases of the survey design process for the RW estimator. For implementing this RW estimation procedure it is also important to have an appropriate metadata system underlying the micro databases from the surveys and available registers. For instance, Statistics Netherlands has a software tool for the collection of tables related to a given target table. Such a tool is necessary when there are many multidimensional tables to be estimated with variables with many different (hierarchical) classifications or variables such as income in either categorical or quantitative form.

Technically speaking, repeated weighting amounts to a further cosmetic adjustment of the commonly-used regression weights w_i resulting in new final weights r_i. This cosmetic adjustment resembles the adjustment in the regression or calibration estimator used by many national statistics institutes (NSIs), where the regression weights w_i can be seen as an adjustment of the starting weights $(1/\pi_i)$. The consequence is that the final weights may vary from table to table; for further details, see DACSEIS deliverable D7.2.

3.2 Criteria for Combining National Surveys

When a survey is intended to cover the whole of the resident population from the Member States of the European Union, a number of conditions have to be taken into account. First, the reference period of the relevant national surveys should be the same. Second, the population underlying each national survey in the reference period is the same and defined uniformly for all national statistical institutes with respect to persons living in residential homes, hospitals, institutions, etc. Third, there are no definitional differences among the variables. Obviously, variables may have a different meaning for different surveys. For each NSI separately it is already an arduous task to realize a certain level of harmonization of the definitions. Often this lack of harmonization is more or less due to a stovepipe model for the organisation of the production process within statistics offices. Attaining uniform definitions for all NSIs is even more difficult. In order to trace the national definitional differences, it might be useful to define a European standard variable for each concept and, next, to describe the differences between the European standard variable and the corresponding national variables. In some cases this may lead to a simple mathematical translation of a national variable into the standard variable. Fourth, the NSIs should use identical classifications for categorical variables. For aggregating tables with categorical variables, it is needed that the categorical variables have hierarchical classifications, i.e., a class of a relatively fine classification always belongs to exactly one class of a coarser classification. Fifth, the national surveys have to be available within a European framework according to an agreed coding system so that a modern statistical use is possible. Finally, some other factors which may have an impact on the differences between national surveys are mode effects (questionnaire, personal interview, telephone, or internet), treatment of non-response, imputation strategy, and the coverage errors in the sampling frame compared to the target population.

These conditions illustrate that harmonization of the underlying national survey designs plays a key role for further aggregating the various national surveys. Assuming that the conditions mentioned above are satisfied, it is not difficult to derive the variance for an aggregated European survey, provided that the variance estimators of the underlying national surveys are available. Let C be the number of Member States of the European Union. Then in obvious notation we have for the estimator of the covariance matrix of a European aggregated frequency table in vectorized form

$$\widehat{t}_{EUR} = \sum_{c=1}^{C} \widehat{t}_c, \quad \widehat{\mathrm{Cov}}(\widehat{t}_{EUR}) = \sum_{c=1}^{C} \widehat{\mathrm{Cov}}(\widehat{t}_c),$$

where we used independence of the national surveys. Likewise, the variance for estimated ratios and (weighted) means can be estimated. In fact, the formulae resemble those for a stratified sample with the countries as strata.

3.3 Recommendations on Software for Variance Estimation

SAS survey procedures, SPSS 'complex samples' and Stata 'survey data analysis' can be recommended for simple survey tasks. Stata (when compared with the other two) has most advanced modelling tools taking the sampling design into account. SAS, SPSS and Stata can deal with hundreds of thousands of observations. All three software packages can be considered user-friendly: the required structures are rather easily manipulated in menus (SPSS, Stata) or using a command line (SAS, SPSS).

When there are various needs for data analysis of complex surveys, Sudaan is a good alternative. It deals with more complex sampling designs than SAS, SPSS and Stata, and has a much wider range of analysis tools for surveys. Also large tasks were performed rather easily in Sudaan. However, Sudaan requires some preliminary preparations (e.g. weights, stratum sizes) for the data to be used in calculations: another computer package is needed for this. The most natural choice is SAS, because there is a SAS-callable version of Sudaan. Sudaan follows the structure of SAS programming.

WesVar emphasises ease and general applicability of resampling methods at the expense of speed. The software is very versatile (including for example poststratification and raking); functions of great complexity can be included in the estimation process. However, WesVar is not at its best with large scale surveys, for example those having several thousand primary sampling units; then both speed and memory comparison results were poor when compared with other packages. As in Sudaan, WesVar requires preliminary preparation of the data, which must be performed external to WesVar. Despite some minor inconviniencies while using WesVar, the software is user-friendly with its menus/screens.

Bascula is a part of Blaise (an integrated system for survey processing), and it might not be reasonable to purchase Blaise only for the use of Bascula. If Blaise is available, Bascula provides an advanced weighting tool (linear or multiplicative weighting) with abilities for proper variance estimation based on Taylor linearisation, though the Balanced Repeated Replications method of Bascula provides rather unstable variance estimates. However, the calculations can be conducted only for the total, the mean and the ratio. Bascula is slower than SAS or SPSS. Bascula also requires external preparation of the survey data. When the basic order of the weight and estimate calculations of Bascula is understood, the operations can be carried out quite easily.

Clan and Poulpe are free-of-charge macro packages, which can be used in SAS. They are the most advanced software in our comparisons, so far as the sampling designs and especially the estimation methods based on auxiliary information are concerned. In addition to standard sampling designs, more complex designs such as network sampling, rotational designs, and two-phase sampling can be dealt with. Weights based on both calibration and GREG estimation can be applied if the auxiliary information used in modelling is provided by the user. Clan also calculates weights based on the theory of the GREG estimator. Poulpe does not create weights; there is another macro package, Calmar, available for calibration. The speciality of Poulpe is the definition of a sampling tree, in which all the levels of sampling can be expressed in such a way that the system accurately calculates stratified multilevel designs with various selection methods. Both packages are intended for sophisticated users; the level of theoretical knowledge needed for operations is quite high. Although logical and rather tight, the program code required in Clan is not easily understood unless the user is experienced with SAS code and to some extent the SAS macro language as well (although the well-prepared manual helps). Poulpe is constructed in French and the English documentation is insufficient. This is a considerable obstacle to a more widespread use of Poulpe. Furthermore, its structure should be developed in a more user-friendly direction, as at present the variety of different properties and aspects makes Poulpe hard to understand as a whole. Both packages calculate more than the ordinary survey software, thus the processing times are longer, especially for Poulpe.

Chapter 4

Conclusions

After all the recommendations given above, one should not forget that the simulation study was conducted on DACSEIS household and individual data in the presence of item non-response on the target variable alone. The study could of course be enriched by further research, for example using business data with severe outliers or non-response models with data *missing not at random* in order to further suit the practical needs of NSIs and other potential users.

In the DACSEIS case, the non-response was mainly applied to the estimation variable. More difficult non-response patterns, such as item non-response on many variables, possibly with non-monotone patterns, may lead to certain other preferences. Under these circumstances, imputation may be advantageous relative to classical estimation techniques that need full response on the auxiliary variable. Additionally, one should keep in mind the specific problems on small areas and domains raised only to a little extent in workpackage 10.

As the process of harmonization in Europe proceeds and the need for comparable standards grows, one may also wish for more appropriate comparability of statistics. However, legal and political aspects pose obstacles for statistical needs. An obvious example are the different legal conditions for the use of register data in Europe, which have a huge influence on the quality of the register data and data quality in general. Any recommendations in this area will have to be seen from different aspects and impacted by local administration and legislation.

Appendix A

Overview of the Simulation Study

The electronic recommended practice manual (deliverable D12.2) contains results for a wide range of simulation tasks, mainly in the field of variance estimation methods under non-response. The evaluation was conducted using different universes, non-response mechanisms rates, estimation purposes, target variables, auxiliary variables, regional subclassifications, as well as different subpopulations of interest.

Additionally to the presentation of the individual results, comparisons of different point and variance estimators for the same simulation task as well as for different non-response rates were conducted. In the case of the German Microcensus further comparisons with regard to small area simulation results were added.

In order to achieve an overview of the work conducted, the following sections summarise the simulation tasks, give a description of the abbreviations used, and should serve as a reference for further reading of the DACSEIS deliverables.

The graphical presentation in the DACSEIS recommended practice manual follows a general scheme. The point and variance estimators are displayed using histograms and boxplots; normality is assessed using normal quantile plots.

The comparisons include a comparative boxplot for the point and variance estimators. Blue lines indicate the true values and the red lines the estimated values. Moreover three important measures are presented as bar charts: the relative root MSE (blue) of the true values and of the simulated values (green), and the average of the estimated relative standard error (red). Finally, the coverage rates of confidence intervals are presented for 90% (red) and 95% (blue) levels.

Details on the small area estimation graphs are given in workpackage 10 (deliverable D10.1+2).

A.1 Description of the Task Names

To identify the simulation tasks, the simulation task names will be uniquely identified. Considering as example `GMC.1.T.ELO.BAW.4.1.sub0.HT.MI.logit` the logical parts of the task name will be described as follows:

Task Component	*Description*
GMC	German Microcensus
1	Non-response mechanism 1
T	Estimation variable as total
ELO	Estimation variable, here ELO unemployed
BAW	Code for federal state
4	Specification of the auxiliary variables (estimation and imputation)
1	Specification of the non-response rate
sub0	Information about considered subpopulation
HT	Horvitz-Thompson estimator
MI	Estimation with multiple imputation
logit	Estimation with a logit-routine

A.2 Description of the Abbreviations

The simulation study contains many abbreviations concerning the computation, which are summarised in the following tables. The abbreviations correspond to all of the components of the task names mentioned in the previous chapter.

A.2.1 The DACSEIS Pseudo Universes

The study contains six pseudo universes. Workpackage 2 gives a thorough overview of the underlying surveys, and a detailed description of the pseudo-universes including their mode of generation can be found in workpackage 3. Table A.1 explains the abbreviations used in the simulation.

Abbreviation	Universe	Country	Kind of Survey
DLFS	Dutch Labour Force Survey	Netherlands	Labour Force Surveys
FLFS	Finnish Labour Force Survey	Finland	
GMC	German Microcensus	Germany	Microcensus Surveys
AMC	Austrian Microcensus	Austria	
EVS	German Sample Survey of Income and Expenditure	Germany	Household Budget Surveys
HBS	Swiss Household Budget Surveys	Switzerland	

Table A.1: Overview of the six pseudo universes used in the simulation study

A.2.2 Non-response Mechanisms

Non-response mechanisms were generally applied to the samples. The details of the mechanisms can be drawn from section 5 of deliverable 1.1. In the case of the German Sample Survey of Income and Expenditure no realistic mechanisms were available due to its being a quota sample. In order to use the data in the DACSEIS context, Swiss non-response mechanisms were adapted to the German data.

Abbreviation	Universe and Sample Description
HBS.1	HBS, unit-NR simple weighting model, item-NR MCAR model
HBS.2	HBS, unit-NR simple weighting model, item-NR intermediate model
HBS.3	HBS, unit-NR simple weighting model, item-NR MAR model
HBS.4	HBS, unit-NR intermediate weighting model, item-NR MCAR model
HBS.5	HBS, unit-NR intermediate weighting model, item-NR intermediate model
HBS.6	HBS, unit-NR intermediate weighting model, item-NR MAR model
HBS.7	HBS, unit-NR complex weighting model, item-NR MCAR model
HBS.8	HBS, unit-NR complex weighting model, item-NR intermediate model
HBS.9	HBS, unit-NR complex weighting model, item-NR MAR model
DLFS.4	DLFS, more complicated item-NR NMAR model
FLFS.1	FLFS, unit-NR MAR model
FLFS.2	FLFS, more complex unit-NR MAR model
FLFS.3	fLFS, unit-NR NMAR model
FLFS.10	FLFS, rotation with simple random sampling without replacement (Tam's method), NR mechanism deliverable 9.1
FLFS.11	FLFS, rotation with systematic sampling (used by British LFS), NR mechanism deliverable 9.1
FLFS.12	FLFS, rotation group sampling (used by Statistics Canada), NR mechanism deliverable 9.1
EVS.1	EVS, unit-NR simple weighting model, item-NR MCAR model
EVS.2	EVS, unit-NR simple weighting model, item-NR intermediate model
EVS.3	EVS, unit-NR simple weighting model, item-NR MAR model
EVS.10	EVS, disproportional sampling design, full-response
EVS.11	EVS, disproportional sampling design with new weights, full-response
EVS.12	EVS, optimal sampling design, full-response
EVS.13	EVS, proportional sampling design, full-response
EVS.14	EVS, proportional sampling design with new weights, full-response
EVS.20	EVS, full-response
AMC.1	AMC, unit-NR MAR model
GMC.1	GMC, unit-NR MAR model

Table A.2: Overview of the non-response (NR) mechanisms used in the simulation study

A.2.3 Non-response Rates

Abbreviation	Non-response Rate
0	0% non-response
1	5% non-response
2	10% non-response
3	25% non-response
4	40% non-response
9	original non-response

Table A.3: Overview of the non-response rates used in the simulation study

A.2.4 Parameters of Interest

Abbreviation	Purpose
T	total estimator
M	mean estimator

Table A.4: Overview of the purposes used in the simulation study

A.2.5 Federal States

Abbreviation		Federal State
ALL		the whole country
AMC	BGL	Burgenland
	VBG	Vorarlberg
DLFS	NOB	Noord-Brabant
	NOH	Noord-Holland
EVS, GMC	BAW	Baden-Württemberg
	BAY	Bayern
	BER	Berlin
	BRA	Brandenburg
	BRE	Bremen
	HAM	Hamburg
	HES	Hessen
	MVP	Mecklenburg-Vorpommern
	NIE	Niedersachsen
	NRW	Nordrhein-Wesfalen
	RLP	Rheinland-Pfalz
	SAA	Sachsen-Anhalt
	SAC	Sachsen
	SAL	Saarland
	SWH	Schleswig-Holstein
	THN	Thüringen

Table A.5: Overview of the federal states used in the simulation study

A.2.6 Subpopulations

Abbreviation	Subpopulation
sub0	entire region
sub1	house size class 3
sub5	house size classes 1-3 in regional stratum 1 (GMC)
sub6	municipality 6 (Amsterdam)
sub15	municipalities 1-3
RS	regional strata (small area)
GGK	house size classes (small area)
RSxGGK	regional strata by house size classes (small area)
NUTS5	regional strata divided by 8 (small area)

Table A.6: Overview of the subpopulations used in the simulation study

A.2.7 Estimators

With regards to the aim of the DACSEIS simulation study, the estimator descriptions include whether it is a point and variance estimator as well as use of a weighting or imputation method.

The Horvitz–Thompson and GREG estimators can be found in any text book on survey sampling. In most cases they were applied to a stratified design. Additionally, raking and calibration estimators were applied which are described in deliverables D8.1 and D8.2 including suitable variance estimators. Additionally, the Lundstrøm–Särndal point and variance estimator is based on a general calibration approach using g-weights in a two-phase approach (cf. Lundstrøm and Särndal, 2002, chapter 6).

Resampling based variance estimators are discussed in deliverables D5.1 and D5.2. The deliverables include a description of the non-response correction by Rao/Shao (jackknife) and Shao/Sitter (bootstrap).

The single imputation methods `SIlfs1` to `SIlfs4`, `SIratio`, and `SIlinreg` are described in deliverable D11.2, chapter 2. The numbering of the LFS type imputation routines is harmonised with the numbers on page 17 in the report. `SIlogit` was applied in accordance with `MIlogit`. The multiple imputation routines were presented in chapter 3 in the same deliverable. The additional MI routing `MIlinbin` can be found in Münnich and Rässler (2004).

The repeated weighting methods can be found in deliverable D7.3. The unequal probability methods are described in deliverables D6.1 and D6.2. Variance estimation for change is presented in deliverables D9.1 and D9.2. Finally, the small area estimators are described in deliverable D10.1+2.

Task Name	Point Estimator	Variance Estimator	NR Correction
HTamc.direct	Horvitz-Thompson	direct	
HTimp.direct.gew1	Horvitz-Thompson, special AMC weights	direct	
HTimp.direct.gew2	Horvitz-Thompson, special AMC weights	direct	
HTimp.direct.SIlogit	Horvitz-Thompson	direct	Single Imputation (SI logit)
HTimp.boot.SIlogit	Horvitz-Thompson	bootstrap, 100 replications	Single Imputation (SI logit)
HTimp.boot.SIlogit0	Horvitz-Thompson	bootstrap, 50 replications	Single Imputation (SI logit)
HTimp.ddJK.SIlogit	Horvitz-Thompson	delete-d jackknife	Single Imputation (SI logit)
HTimp.direct.SIlfs1	Horvitz-Thompson	direct	Single Imputation (SI LFS 1)
HTimp.boot.SIlfs1	Horvitz-Thompson	bootstrap, 100 replications	Single Imputation (SI LFS 1)
HTimp.boot.SIlfs10	Horvitz-Thompson	bootstrap, 50 replications	Single Imputation (SI LFS 1)
HTimp.ddJK.SIlfs1	Horvitz-Thompson	delete-d jackknife	Single Imputation (SI LFS 1)
HTimp.direct.SIlfs2	Horvitz-Thompson	direct	Single Imputation (SI LFS 2)
HTimp.boot.SIlfs2	Horvitz-Thompson	bootstrap, 100 replications	Single Imputation (SI LFS 2)
HTimp.boot.SIlfs20	Horvitz-Thompson	bootstrap, 50 replications	Single Imputation (SI LFS 2)
HTimp.ddJK.SIlfs2	Horvitz-Thompson	delete-d jackknife	Single Imputation (SI LFS 2)
HTimp.direct.SIlfs3	Horvitz-Thompson	direct	Single Imputation (SI LFS 3)
HTimp.boot.SIlfs3	Horvitz-Thompson	bootstrap, 100 replications	Single Imputation (SI LFS 3)
HTimp.boot.SIlfs30	Horvitz-Thompson	bootstrap, 50 replications	Single Imputation (SI LFS 3)
HTimp.ddJK.SIlfs3	Horvitz-Thompson	delete-d jackknife	Single Imputation (SI LFS 3)
HTimp.direct.SIlfs4	Horvitz-Thompson	direct	Single Imputation (SI LFS 4)
HTimp.boot.SIlfs4	Horvitz-Thompson	bootstrap, 100 replications	Single Imputation (SI LFS 4)
HTimp.boot.SIlfs40	Horvitz-Thompson	bootstrap, 50 replications	Single Imputation (SI LFS 4)
HTimp.ddJK.SIlfs4	Horvitz-Thompson	delete-d jackknife	Single Imputation (SI LFS 4)

Table A.7: Overview of the estimators used in the simulation study, Table 1

Task Name	Point Estimator	Variance Estimator	NR Correction
HTimp.direct.SIratio	Horvitz-Thompson	direct	Single Imputation (SI ratio)
HTimp.boot.SIratio	Horvitz-Thompson	bootstrap, 100 replications	Single Imputation (SI ratio)
HTimp.boot.SIratio0	Horvitz-Thompson	bootstrap, 50 replications	Single Imputation (SI ratio)
HTimp.ddJK.SIratio	Horvitz-Thompson	delete-d jackknife	Single Imputation (SI ratio)
HTimp.direct.SIlinreg	Horvitz-Thompson	direct	Single Imputation (SI linear regression)
HTimp.boot.SIlinreg	Horvitz-Thompson	bootstrap, 100 replications	Single Imputation (SI linear regression)
HTimp.boot.SIlinreg0	Horvitz-Thompson	bootstrap, 50 replications	Single Imputation (SI linear regression)
HTimp.ddJK.SIlinreg	Horvitz-Thompson	delete-d jackknife	Single Imputation (SI linear regression)
HT.calSFSO.boot.SIratio	Horvitz-Thompson, SFSO weights	bootstrap, 100 replications	Single Imputation (SI ratio)
HT.calSFSO.ddJK.SIratio	Horvitz-Thompson, SFSO weights	delete-d jackknife	Single Imputation (SI ratio)
HT.calS.direct.SIlinreg	Horvitz-Thompson, calibrated	direct	Single Imputation (SI linear regression)
HT.calSFSO.boot.SIlinreg	Horvitz-Thompson, SFSO weights	bootstrap, 100 replications	Single Imputation (SI linear regression)
HT.calSFSO.ddJK.SIlinreg	Horvitz-Thompson, SFSO weights	delete-d jackknife	Single Imputation (SI linear regression)
HT.MI.logit	Horvitz-Thompson	MI Inference	Multiple Imputation (MI logit, 30 repl.)
HT.MI.logit0	Horvitz-Thompson	MI Inference	Multiple Imputation (MI logit, 5 repl.)
HT.MI.logit1	Horvitz-Thompson	MI Inference	Multiple Imputation (MI logit, 15 repl.)
HT.MI.linbin	Horvitz-Thompson	MI Inference	Multiple Imputation (MI linbin, 30 repl.)
HT.MI.linbin0	Horvitz-Thompson	MI Inference	Multiple Imputation (MI linbin, 5 repl.)
HT.MI.linbin1	Horvitz-Thompson	MI Inference	Multiple Imputation (MI linbin, 15 repl.)
HT.MI.linreg	Horvitz-Thompson	MI Inference	Multiple Imputation (MI linear regression)
HT.MI.linreg0	Horvitz-Thompson	MI Inference	Multiple Imputation (MI linear regression, 5 replications)
HT.MI.linreg1	Horvitz-Thompson	MI Inference	Multiple Imputation (MI linear regression, 15 replications)
HTcal.MI.linreg	Horvitz-Thompson,	MI Inference calibrated	Multiple Imputation (MI linear regression)

Table A.8: Overview of the estimators used in the simulation study, Table 2

Task Name	Point Estimator	Variance Estimator	NR Correction
HTraking.direct	Horvitz-Thompson, calibrated	direct	Weigting/Calibration (exponential calibration function)
HTrakingWOR.direct	Horvitz-Thompson, calibrated	jackknife linearisation	Weigting/Calibration (exponential calibration function)
HTrakingJK.direct	Horvitz-Thompson, calibrated	jackknife linearisation	Weigting/Calibration (exponential calibration function)
HTrakingJKWOR.direct	Horvitz-Thompson, calibrated	jackknife linearisation	Weigting/Calibration (exponential calibration function)
HTMLraking.direct	Horvitz-Thompson, calibrated	direct	Weigting/Calibration (MInimum entropy distance)
HTMLrakingWOR.direct	Horvitz-Thompson, calibrated	direct	Weigting/Calibration (MInimum entropy distance)
HTMLrakingJK.direct	Horvitz-Thompson, calibrated	jackknife linearisation	Weigting/Calibration (MInimum entropy distance)
HTMLrakingJKWOR.direct	Horvitz-Thompson, calibrated	jackknife linearisation	Weigting/Calibration (MInimum entropy distance)
HTgreg.direct	GREG	direct	Weigting/Calibration (linear calibration function)
HTgregWOR.direct	GREG	direct	Weigting/Calibration (linear calibration function)
HTgregJK.direct	GREG	jackknife linearisation	Weigting/Calibration (linear calibration function)
HTgregJKWOR.direct	GREG	jackknife linearisation	Weigting/Calibration (linear calibration function)
gCALNR.direct	Horvitz-Thompson, calibrated	direct	Weighting (Calibration approach)
gGREGimp.direct.SIlogit	GREG	direct	Single Imputation (SI logit)
gGREGimp.boot.SIlogit	GREG	bootstrap, 100 replications	Single Imputation (SI logit)
gGREGimp.boot.SIlogit0	GREG	bootstrap, 50 replications	Single Imputation (SI logit)
gGREGimp.ddJK.SIlogit	GREG	delete-d jackknife	Single Imputation (SI logit)

Table A.9: Overview of the estimators used in the simulation study, Table 3

Task Name	Point Estimator	Variance Estimator	NR Correction
gGREGimp.direct.SIlfs1	GREG	direct	Single Imputation (SI LFS 1)
gGREGimp.boot.SIlfs1	GREG	bootstrap, 100 replications	Single Imputation (SI LFS 1)
gGREGimp.boot.SIlfs10	GREG	bootstrap, 50 replications	Single Imputation (SI LFS 1)
gGREGimp.ddJK.SIlfs1	GREG	delete-d jackknife	Single Imputation (SI LFS 1)
gGREGimp.direct.SIlfs2	GREG	direct	Single Imputation (SI LFS 2)
gGREGimp.boot.SIlfs2	GREG	bootstrap, 100 replications	Single Imputation (SI LFS 2)
gGREGimp.boot.SIlfs20	GREG	bootstrap, 50 replications	Single Imputation (SI LFS 2)
gGREGimp.ddJK.SIlfs2	GREG	delete-d jackknife	Single Imputation (SI LFS 2)
gGREGimp.direct.SIlfs3	GREG	direct	Single Imputation (SI LFS 3)
gGREGimp.boot.SIlfs3	GREG	bootstrap, 100 replications	Single Imputation (SI LFS 3)
gGREGimp.boot.SIlfs30	GREG	bootstrap, 50 replications	Single Imputation (SI LFS 3)
gGREGimp.ddJK.SIlfs3	GREG	delete-d jackknife	Single Imputation (SI LFS 3)
gGREGimp.direct.SIlfs4	GREG	direct	Single Imputation (SI LFS 4)
gGREGimp.boot.SIlfs4	GREG	bootstrap, 100 replications	Single Imputation (SI LFS 4)
gGREGimp.boot.SIlfs40	GREG	bootstrap, 50 replications	Single Imputation (SI LFS 4)
gGREGimp.ddJK.SIlfs4	GREG	delete-d jackknife	Single Imputation (SI LFS 4)
gGREGimp.direct.SIratio	GREG	direct	Single Imputation (SI ratio)
gGREGimp.boot.SIratio	GREG	bootstrap, 100 replications	Single Imputation (SI ratio)
gGREGimp.boot.SIratio0	GREG	bootstrap, 50 replications	Single Imputation (SI ratio)
gGREGimp.ddJK.SIratio	GREG	delete-d jackknife	Single Imputation (SI ratio)
gGREGimp.direct.SIlinreg	GREG	direct	Single Imputation (SI linear regression)
gGREGimp.boot.SIlinreg	GREG	bootstrap, 100 replications	Single Imputation (SI linear regression)
gGREGimp.boot.SIlinreg0	GREG	bootstrap, 50 replications	Single Imputation (SI linear regression)
gGREGimp.ddJK.SIlinreg	GREG	delete-d jackknife	Single Imputation (SI linear regression)

Table A.10: Overview of the estimators used in the simulation study, Table 4

Task Name	Point Estimator	Variance Estimator	NR Correction
gGREG.MI.logit	GREG	MI Inference	Multiple Imputation (MI logit, 30 repl.)
gGREG.MI.logit0	GREG	MI Inference	Multiple Imputation (MI logit, 5 repl.)
gGREG.MI.logit1	GREG	MI Inference	Multiple Imputation (MI logit, 15 repl.)
gGREG.MI.linbin	GREG	MI Inference	Multiple Imputation (MI linbin, 30 repl.)
gGREG.MI.linbin0	GREG	MI Inference	Multiple Imputation (MI linbin, 5 repl.)
gGREG.MI.linbin1	GREG	MI Inference	Multiple Imputation (MI linbin, 15 repl.)
gGREG.MI.linreg	GREG	MI Inference	Multiple Imputation (MI linear regression, 30 replications)
gGREG.MI.linreg0	GREG	MI Inference	Multiple Imputation (MI linear regression, 5 replications)
gGREG.MI.linreg1	GREG	MI Inference	Multiple Imputation (MI linear regression, 15 replications)
gCALimp.direct.SIlfs1	Horvitz-Thompson, calibrated	direct	Single Imputation (SI LFS 1)
gCALimp.boot.SIlfs1	Horvitz-Thompson, calibrated	bootstrap, 100 replications	Single Imputation (SI LFS 1)
gCALimp.ddJK.SIlfs1	Horvitz-Thompson, calibrated	delete-d jackknife	Single Imputation (SI LFS 1)
gCALimp.direct.SIlfs4	Horvitz-Thompson, calibrated	direct	Single Imputation (SI LFS 4)
gCALimp.boot.SIlfs4	Horvitz-Thompson, calibrated	bootstrap, 100 replications	Single Imputation (SI LFS 4)
gCALimp.ddJK.SIlfs4	Horvitz-Thompson, calibrated	delete-d jackknife	Single Imputation (SI LFS 4)
gCALimp.direct.SIratio	Horvitz-Thompson, calibrated	direct	Single Imputation (SI ratio)
gCALimp.boot.SIratio	Horvitz-Thompson, calibrated	bootstrap, 100 replications	Single Imputation (SI ratio)
gCALimp.ddJK.SIratio	Horvitz-Thompson, calibrated	delete-d jackknife	Single Imputation (SI ratio)
gCALimp.direct.SIlinreg	Horvitz-Thompson, calibrated	direct	Single Imputation (SI linear regression)
gCALimp.boot.SIlinreg	Horvitz-Thompson, calibrated	bootstrap, 100 replications	Single Imputation (SI linear regression)
gCALimp.ddJK.SIlinreg	Horvitz-Thompson, calibrated	delete-d jackknife	Single Imputation (SI linear regression)

Table A.11: Overview of the estimators used in the simulation study, Table 5

Task Name	Point Estimator	Variance Estimator	NR Correction
HT.VHT.Direct	Horvitz-Thompson	direct	
RE.VRE.Direct	Regression Estimator	linearisation variance estimation (without g-weights)	
RE.VREg.Direct	Regression Estimator	linearisation variance estimation	
RWE.VRWE.Direct	Repeated Weighting Estimator	linearisation variance estimation	
RWE.SVRWE.Direct	Repeated Weighting Estimator	linearisation variance estimation (simplified estimator)	
direct4c.Kish	Horvitz-Thompson difference	Variance estimator for change between two population totals	
direct4c.Tam	Horvitz-Thompson difference	Variance estimator for change between two population totals	
direct4c.Berger	Horvitz-Thompson difference	Variance estimator for change between two population totals	
directUP.Berger	Horvitz-Thompson	Variance estimator for unequal probability design	
directUP.Brewer	Horvitz-Thompson	Variance estimator for unequal probability design	
directUP.Hajek	Horvitz-Thompson	Variance estimator for unequal probability design	
directUP.Hansen	Horvitz-Thompson	Variance estimator for unequal probability design	
directUP.SRSWR	Horvitz-Thompson	Variance estimator based on simple random sampling	
Small Area Estimators			
nsm.direct	National Sample Mean	direct; MSE	
direct.mean.direct	EBLUP B	direct (MSE)	
greg.mean.direct syntha.mean.direct	GREG SYNTH A	direct (MSE) direct (MSE)	
synthb.mean.direct	SYNTH B	direct (MSE)	
eblupa1.mean.direct	EBLUP A	direct (MSE)	
eblupa2.mean.direct	EBLUP A	direct (MSE)	
eblupb1.mean.direct	EBLUP B	direct (MSE)	
eblupb2.mean.direct	EBLUP B	direct (MSE)	

Table A.12: Overview of the estimators used in the simulation study, Table 6

A.2.8 Estimation Variables

Abbreviation	Target Variable
ELO	unemployed
ELOM	unemployed men
ELOA1	unemployed aged (14,24)
ELOA1F	unemployed women aged (14,24)
ELOA1M	unemployed men aged (14,24)
ELOA2	unemployed aged (25,44)
ELOA2F	unemployed women aged (25,44)
ELOA2M	unemployed men aged (25,44)
ELOA3	unemployed aged (45,64)
ELOA3F	unemployed women aged (45,64)
ELOA3M	unemployed men aged (45,64)
ELOA4	unemployed aged (65+)
ELOA4F	unemployed women aged (65+)
ELOA4M	unemployed men aged (65+)
ELOFEB	unemployed women aged (14+)
ELOMEB	unemployed men aged (14+)
ELOEB	unemployed aged (14+)
INC	net income
HHC	proportion of single-person households
ERW	employment
ED0	highest level of education - no answer
ED1	highest level of education - upper secondary education
ED2	highest level of education - post secondary
ED3	highest level of education - 5B-programmes
ED4	highest level of education - 5A-programmes
ED5	highest level of education - second stage
ED6	highest level of education - level unknown
LS0	labour force characteristics - employed
LS1	labour force characteristics - unemployed
LS2	labour force characteristics - conscripts
LS3	labour force characteristics - students
LS4	labour force characteristics - disabled
LS5	labour force characteristics - pensioners
LS6	labour force characteristics - domestic work
LS7	labour force characteristics - others
T01	population
T02	women
T03	people aged (15,24)
T04	people aged (25,34)
T05	people aged (35,44)

Table A.13: Overview of the target variables used in the simulation study, Table 1

Abbreviation	Target Variable
T06	people aged (45,54)
T07	people aged (55,64)
T08	other European
T09	non European
T10	women aged (15,24)
T11	women aged (25,34)
T12	women aged (35,44)
T13	women aged (45,54)
T14	women aged (55,64)
T15	other European women
T16	non European women
T17	other European aged (15,24)
T18	other European aged (25,34)
T19	other European aged (35,44)
T20	other European aged (45,54)
T21	other European aged (55,64)
T22	non European aged (15,24)
T23	non European aged (25,34)
T24	non European aged (35,44)
T25	non European aged (45,54)
T26	non European aged (55,64)
T27	other European women aged (15,24)
T28	other European women aged (25,34)
T29	other European women aged (35,44)
T30	other European women aged (45,54)
T31	other European women aged (55,64)
T32	non European women aged (15,24)
T33	non European women aged (25,34)
T34	non European women aged (35,44)
T35	non European women aged (45,54)
T36	non European women aged (55,64)
T37	people divorced or widowed
T38	people unmarried
T39	unemployed labour force
T40	non labour force

Table A.14: Overview of the target variables used in the simulation study, Table 2

Abbreviation	Target Variable
T41	women divorced or widowed
T42	women unmarried
T43	unemployed labour force women
T44	non labour force women
T45	unemployed, divorced or widowed labour force
T46	unemployed unmarried labour force
T47	non labour force divorced or widowed
T48	non labour force unmarried
T49	unemployed, divorced or widowed labour force women
T50	unemployed unmarried labour force women
T51	non labour force divorced or widowed women
T52	non labour force unmarried women
T53	household with 1 person per household
T54	household with 2 person per household
T55	household with 3 person per household
T56	household with 4 person per household
T57	household with 5 person per household
T58	household with 6 person per household
T59	household with 7 person per household
T60	household with 8 person per household
T61	household with 9 or more person per household
T62	household self employed
T63	household civil servant or military
T64	household employee
T65	household worker
T66	household unemployed, pensioner, students, other
T67	household of other type of household
T68	household mother/father alone with 1 child
T69	household mother/father alone with 2 or more children
T70	household couple with 1 child - spouse employed
T71	household couple with 1 child - spouse unemployed
T72	household couple with 2 or more children - spouse employed
T73	household couple with 2 or more children - spouse unemployed
T74	household net income
T75	household expenditure

Table A.15: Overview of the target variables used in the simulation study, Table 3

A.2.9 Auxiliary Variables

The auxiliary variables are divided into two parts. We distinguish between auxiliary variables concerning estimation and imputation. A number of auxiliary variable stand for special variables with respect to estimation and also imputation. Tables A.16–A.19 show the different combinations of auxiliary variables.

Abbreviation	Auxiliary Variable for Estimation
1	registered unemployed, German men, German women, non German men
2	people aged (25,65), men, registered unemployed, German
3	people aged (25,64), men, registered unemployed, German
4	people aged 24 and less, men, registered unemployed, German
5	registered unemployed
6	German men, German women, non German men
7	registered unemployed, German men, German women, non German men
8	registered unemployed men, German men, German women, non German men
9	registered unemployed men, German men, German women, non German men
10	registered unemployed men, German men, German women, non German men
11	registered unemployed men, German men, German women, non German men
12	registered unemployed men, German men
13	registered unemployed men, strata
14	registered unemployed, German men, German women, non German men
15	registered unemployed, German men, German women, non German men
16	registered unemployed, German men, German women, non German men
17	registered unemployed, German men, German women, non German men
18	registered unemployed, German men, German women, non German men
19	registered unemployed, German men, German women, non German men
20	registered unemployed, German men, German women, non German men
21	registered unemployed, German men, German women, non German men
22	-
23	age groups(6) * gender + region
24	men aged (15,24), men aged (25,34), men aged (35,44), men aged (45,54), men aged (55,64), men aged (65+), women aged (15,24), women aged (25,34), women aged (35,44), women aged (45,54), women aged (55,64), married, divorced or widowed

Table A.16: Overview of the auxiliary variables for estimation used in the simulation study, Table 1

Abbreviation	Auxiliary Variable for Estimation
25	expenditure
26	-
27	-
28	-
29	type of household + socio economic status of the household
30	registered unemployed
31	people aged (25,65), men, highest education (university degree)
32	people aged (25,65), men, income
33	people aged (25,65), men, highest education (university degree), unemployed, number of person per household
34	number of persons per household (1-5), socio economic status of the household (0-3), income
35	household with 2 persons, couple with 2 or more children, employee, income
36	income
37	income, socio economic status of the household (0-3)
38	expenditure, (unemployed, pensioners, students and others)
39	gender (2 categories) * municipality (10 categories) * age (6 categories)
40	gender (2 categories) * municipality (10 categories) * age (6 categories)
41	gender (2) * age (6) + region
43	registered unemployed, German men, German women, non German men
44	registered unemployed, German men, German women, non German men
45	registered unemployed, German men, German women, non German men
46	registered unemployed, German men, German women, non German men
47	registered unemployed, German men, German women, non German men
48	registered unemployed, German men, German women, non German men
49	registered unemployed, German men, German women, non German men

Table A.17: Overview of the auxiliary variables for estimation used in the simulation study, Table 2

Abbreviation	Auxiliary Variable for Imputation
1	registered unemployed, German men, German women, non German men, age classes, nonresponse classes
2	people aged (25,65), men, registered unemployed, German, nonresponse classes, strata
3	people aged (25,64), men, registered unemployed, German, nonresponse classes, strata
4	people aged 24 and less, men, registered unemployed, German, nonresponse classes, strata
5	registered unemployed
6	German men, German women, non German men, age classes, nonresponse classes
7	people aged (25,54), men, German, highest education (5,6)
8	registered unemployed men
9	registered unemployed men, German men, German women, non German men, age classes, nonresponse classes
10	registered unemployed men, age classes
11	registered unemployed men, strata
12	registered unemployed men, regional strata, house size classes
13	registered unemployed men, regional strata, house size classes
14	registered unemployed women aged (14,24)
15	registered unemployed women aged (25,44)
16	registered unemployed women aged (45,64)
17	registered unemployed women aged (65+)
18	registered unemployed men aged (14,24)
19	registered unemployed men aged (25,44)
20	registered unemployed men aged (45,64)
21	registered unemployed men aged (65+)
22	education
23	education
24	married, people aged (25,54), Dutch

Table A.18: Overview of the auxiliary variables for imputation used in the simulation study, Table 1

Abbreviation	Auxiliary Variable for Imputation
25	expenditure
26	people aged (0-24;65+) and house size (1+), people aged (25,39) and house size (1+), people aged (40,54) and house size (1+), people aged (55,64) and house size (1+)
27	house size (0), people aged (65+) and house size (1+), men aged (25,39) and house size (1+), women aged (25,39) and house size (1+), men aged (40,54) and house size (1+), women aged (40,54) and house size (1+), men aged (55,64) and house size (1+), women aged (55,64) and house size (1+)
28	expenditure
29	expenditure
30	-
31	-
32	-
33	-
34	-
35	-
36	-
37	-
38	-
39	-
40	-
41	-
43	registered unemployed aged (14,24)
44	registered unemployed aged (25,44)
45	registered unemployed aged (45,64)
46	registered unemployed aged (65+)
47	registered unemployed women aged (14+)
48	registered unemployed men aged (14+)
49	registered unemployed aged (14+)

Table A.19: Overview of the auxiliary variables for imputation used in the simulation study, Table 2

References

Lundstrøm, S. and Särndal, C. E. (2002): *Estimation in the Presence of Nonresponse and Frame Imperfections.* Statistics Sweden.

Münnich, R. and Rässler, S. (2004): Variance estimation unter multiple imputation. Proceedings of the Q2004 conference.

Rao, J. N. K. and Shao, J. (1992): Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811–822.

Shao, J. and Sitter, R. R. (1996): Bootstrap for imputed survey data. *Journal of the American Statistical Association* **91**, 1278–1288.