DACSEIS
IST-2000-26057

Workpackage 12

Variance Estimation in Complex Surveys

Deliverable 12.1

2004

List of contributors:

Paul Knottnerus, Harm Jan Boonstra, Statistics Netherlands;
Seppo Laaksonen, Pauli Ollila, Statistics Finland;
Chris Skinner, Yves Berger, University of Southampton;
Anthony Davison, Sylvain Sardy, EPFL;
Andreas Quatember, JKU / IFAS Linz;
Ralf Münnich, Kersten Magg, Katrin Schmidt, Rolf Wiegert, University of Tübingen

Main responsibility:

Ralf Münnich, University of Tübingen (co-ordinator)

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

The deliverable D12.1 aims to give an overview of the DACSEIS research project, its workpackages, and its main achievements. It is designed to guide the interested user through the widespread research activities of the DACSEIS teams, the numerous high quality reports as well as through the structure and the different fields that yield the basis for the *recommended practice manual* and the final report.

The outcome of the project is the result of a three-years research of the entire DACSEIS team and the exchange with internal and external expert groups, especially in Luxembourg. Furthermore, project contribution has been effected on several conferences and in contact with end-user groups.

The workpackage overviews were contributed by:

1	Ralf Münnich, Rolf Wiegert, University of Tübingen
2	Andreas Quatember, University of Linz
3	Ralf Münnich, Rolf Wiegert, University of Tübingen
4	Pauli Ollila, Statistics Finland
5	Anthony Davison and Sylvain Sardy, Ecole Polytechnique Féderale de Lausanne
6	Yves Berger, University of Southampton
7	Paul Knottnerus, Statistics Netherlands
8	Chris Skinner, University of Southampton
9	Yves Berger, University of Southampton
10	Ralf Münnich, Rolf Wiegert, University of Tübingen
11	Seppo Laaksonen, Statistics Finland, Chris Skinner, University of Southampton
12	Ralf Münnich, Rolf Wiegert, University of Tübingen

Ralf Münnich University of Tübingen

Contents

List of Figures

List of Tables

Chapter 1

Overview of the DACSEIS Project

1.1 Data Quality

In the past years, precision of data and subsequently the measurement of data quality has become an increasingly challenging issue with the growing complexity of the surveys. This problem has become even more acute due to the increased number of error sources that now have to be considered, especially in a European context where international comparisons need to be made. Within these sample surveys, many features and peculiarities, e. g. the complexity of the design, nonresponse rates and behaviour, have to be taken into account. Even more sophisticated is the measurement of the data quality in the context of investigating the different aspects of errors.

When dealing with data quality, many different facets can be considered. Following the Eurostat definition of quality in statistics, the following aspects play an important role (cf.Eurostat, 2001a, Eurostat, 2001b, Linden and Grünewald, 2001, and Münnich and Wiegert, 2003):

- Relevance of statistical concepts;
- Accuracy of estimates;
- Timeliness and punctuality of data dissemination;
- Accessability and clarity of information;
- Comparability of statistics;
- Coherence;
- Completeness.

The DACSEIS project focuses on the accuracy of estimates. In this context, the measurement of accuracy is an integral part of evaluating data quality of survey data.

The European Statistics System faces the difficult task of organising a harmonised and reliable European data base dealing with a heterogeneous environment. This induces a quite strong need for research developing methods and tools which allow for the preparation of data of comparable quality standards and valuable precision.

Although investigating data quality and its measurement has been of interest for a longer time, it has recently become increasingly important. Therefore, many studies and projects have already been setup, focusing on special aspects of data quality (e.g. SUP.COM 97-06, 97-14, and 98-16; c.f. http://europa.eu.int/comm/eurostat/research/). Unfortunately, little effort has been undertaken in finding adequate applications of the methods in complex surveys which calls for an exchange of technology from theory into practice. Thus, a very important issue is to investigate the above mentioned *accuracy of estimates* in a practical environment, for instance simulations of populations of important applied surveys that allow for an easy application of more sophisticated research methods. This helps find suitable recommendations for the usage of these methods in applied surveys.

The importance of the assessment of quality can be drawn from the fact that all major statistical offices, e.g. the US Bureau of the Census, Statistics Canada, as well as Eurostat and its European member states, do research in measuring quality. Furthermore, Eurostat has its own *Working Group on Assessment of Quality in Statistics* and the *Task Force Variance Estimation*, which stresses the importance of further research in this field. DACSEIS particularly focuses on the research of variance estimation methods in complex surveys. The surveys of interest are mostly carried out by the National Statistical Institutes (NSI) to gain various types of data for a wide range of socio-economic problems.

1.2 Aim of the Project

A modern country with its administration and its economy will be positively influenced by and probably depend on an efficient organisation of official statistics. These official statistics have to preserve a reliable set of variables and data for diagnostic and analytical purposes to permanently get information about the recent economic and socio-demographic conditions in the Union. This is an important task for the EU in the context of its current status, but also in the context of a stepwise enlargement of the EU.

European statistics face the difficult task of creating a harmonised and reliable socio-economic database for the economy in a united Europe with special emphasis on different national surveys and their international comparability. The definitions of units and variables used in the member states of the EU need to be standardised and the quality of the data gained from complex surveys e. g. household, population surveys and especially the labour force surveys should be made more homogeneous and comparable with respect to different quality components.

The core of the problem is to obtain applicable methods for variance estimation in complex multi-purpose sampling schemes. Simulated universes, reflecting the relevant national surveys and their complex characteristics and properties, will be generated. They are the basis of the analysis of the precision of the variance estimation methods regarding the influence on special designs and conditions of gaining data. A catalogue of recommended methods should enable the user to estimate variances effectively and reliably with comparable standards. A list providing criteria to be checked with certain complex design will facilitate the determination of a suitable variance estimation method for any specific estimation problem.

This usage of recommended methods can be a nucleus of a harmonised and standardised European quality management system in statistics. To fulfil this task, all relevant variance estimation methods currently available should be analysed, classified, evaluated and perhaps improved. This will be accomplished by theoretical research and by a realistic Monte Carlo study relevant for national surveys.

To speak generally, errors in data may have unfortunate consequences in economic and social analysis to an extent yet unknown. Therefore, research within DACSEIS is primarily a tool to advance the provision of reliable information on data quality and their dominating components.

The main goal of the project is to analyse the *accuracy of estimates* while taking into consideration different aspects of practical needs, like nonresponse rates and response behaviour, imputation, rotation schemes and applicability of methods in large scale universes. Additionally, an important task in this context is to develop efficient methods to combine data of surveys and registers. These methods are useful in reducing the response burden and may help improve the data quality, especially when dealing with rare events in small areas. A challenging issue in this case is the variance estimation.

The accuracy itself can obviously have many aspects that should be inspected. To find out the accurate procedure for measuring data quality means dealing with a large variety of different aspects which should be considered simultaneously. However, a univariate measure can obviously not sufficiently cover these different appearances of accuracy.

A thorough Monte Carlo simulation study is integral part of DACSEIS to allow for an investigation of survey procedures identical to those surveys which are practically applied in the team member countries. With respect to disclosure control, DACSEIS will use adequately simulated universes for its research.

These considerations show the necessity of analysing the different components of accuracy and its relevant measures. Covering for instance estimators of variance, biases of the estimators, their mean squared errors, and complete simulated distributions of estimators as well as the influence of special properties of complex survey designs on the methods applied. In this context, the influence of nonresponse and imputation on accuracy is also an important goal which was studied thoroughly within DACSEIS.

A final result of the work will be the dissemination of the results of the research by a so called *recommended practice manual* that will be built up for all potential users as a database.

1.3 Structure of the Project

The research of the project is organised in 11 workpackages (WP) plus a final report. As shown in Figure 1.1, the investigation of quality issues as described earlier in the paper, is done in WP1 *Variance estimation in complex surveys*. This workpackage also builds the connection between all workpackages and yields the basis for the results in the form of *best practice recommendations*. To accomplish this challenging objective, different surveys will have to be investigated with respect to the underlying universes, their corresponding sample survey, as well as their peculiarities, e. g. nonresponse rates and behaviour, weighting schemes etc. The surveys to be examined within WP2 *Structures and analysis of relevant national surveys* basically cover household surveys. They are a very important input for WP3, in which survey data and sampling routines were developed. These play an important role for the *Monte Carlo simulation study of European surveys* aiming at elaborating the point and variance estimation procedures of interest. The methods to be investigated in *practical situations* are described within the methodological WPs. These practical situations will be simulated from the surveys mentioned above as well as from artificial surveys to find out the behaviour of the variance estimators within a *synthetic, but realistic* environment. The conclusions from this Monte Carlo simulation study will be summarised in the *recommended practice manual* as part of the final report.

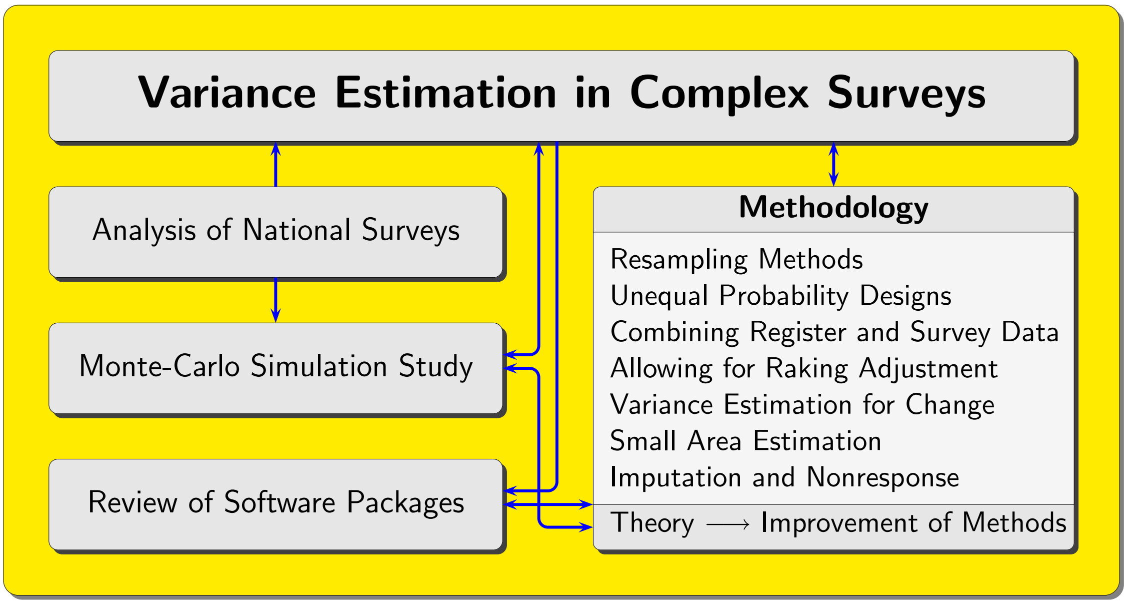

Figure 1.1: Outline of the structure of the workpackages within the DACSEIS project

The methodology mainly focuses on variances estimation of sampling errors for cross-sectional and to some extent also for longitudinal data. The methodological framework of variance estimation techniques can be seen from Figure 1.1 and will be described in more detail below. Additional work will be done by the examination of variance estimation methods for non-sampling errors with respect to missing values and their imputation.

Figure 1.1 presents in graphical form the structure and interaction of the relevant theoretical and practical parts of the DACSEIS variance estimation study. Details of the workpackages will follow below. The main part is the connection between WP1 and the recommended practice manual as part of the final report, the *variance estimation in complex surveys*. The methodology, which can be drawn from Figure 1.1, aims to investigate the state-of-the-art variance estimation methods in strong connection with practice. This is achieved by a large Monte-Carlo simulation study which allows for a thorough investigation of the influence of non-response on the estimates.

An overview of software packages which are used in the NSIs including a deep evaluation concludes the research while investigating the presence of the variance estimation methods for their availability in the packages.

The following chapter gives an overview of the different workpackages and the main research achievements of the project.

Chapter 2

Overview of the Achievements of the Project

2.1 Workpackage 1: Variance Estimation in Complex Surveys

This fundamental workpackage is divided into the two deliverables D1.1 and D1.2. Deliverable D1.1 has been arranged in five parts and two appendices. The subsections particulary deal with classification of variance estimation methods, evaluation criteria, overview of individual and household surveys, metadata for the dissemination of results, and non-response modeling for the simulation study. The appendices contain an overview of metadata and code dissemination as well as summary tables of the questionnaire.

Deliverable D1.2 focuses on a comparison of variance estimation methods based on a Monte-Carlo simulation study. This study uses close-to-reality survey data from six national surveys, the Dutch and the Finnish LFS, the Austrian and German Microcensus, the Swiss HBS as well as the German Income and Expenditure survey (EVS). The surveys themselves are described in workpackage 2, the corresponding data generation processes in workpackage 3. Additionally, special simulations are dealt with in another chapter of D1.2.

Chapter 1 of deliverable D1.1 comprises a detailed classification of variance estimation methods as a presentation of the state of the art. These methods cover direct variance estimators, linearization methods, and resampling methods. Details of the methodology which constitutes an important basis for the simulation study can be gathered from the respective workpackages.

At first, the direct variance estimation method is discussed in opposition to the classical design-based approach. To linear estimators with fixed weights not depending on the sample direct methods of variance estimation can be applied without too much difficulty. Another topic of this chapter is the discussion of estimators which are reliable and approximately unbiased with respect to the randomization mechanism used to generating the sample. However, the application to specific estimators and sampling designs will be complex in practice, in particular because of the difficulties arising from determining second order inclusion probabilities. To guarantee an overall analysis, stratified sampling and multistage designs as well as multi-phase sampling were examined in this context, too. Purpose and main achievement of this part of the study was to gain a set of decision criteria and suitable formulae for the simulations.

Linearization methods enable direct methods to be extended to non-linear estimators which are of interest in estimating variances. The theoretical methods proposed in literature were recorded and suitable methods were selected and compiled for the purpose of the simulation study. Keywords for those methods are Taylor series expansion, Jackknife linearization and calibration methods. Beyond, resampling methods were analyzed, its various features discussed and integrated into the context.

Chapter 1 presents an overview of present methods of variance estimation. The properties and performances of these theoretical estimators have been tested under practical conditions.

In Chapter 2 a list of criteria is presented suitable for evaluating point and variance estimation methods. It contains

- theoretical measures of accuracy,
- measurement of accuracy of variance estimates and
- empirical measurement of the accuracy.

These measures create problems as far as the practical implementation is concerned. In this chapter an operable basis for dealing with these obstacles is provided including the special context of non-response, too.

These evaluation methods play a very important role in practice and are an integral part in the dissemination of the simulation results under deliverable D12.2. They enable the user to investigate differences between the estimation methods of interest more deeply having regard to theoretical criteria and their measurability in concrete samples.

In the following Chapter 3 a thorough overview of the results of individual and household surveys in European official statistics is given. The results of this evaluation are presented in form of graphs and associated tables. This information about the surveys conducted by the NSIs at present is an instructive part of D1.1. Additionally, in Appendix B one can find the questionnaire itself and summary tables of the 22 evaluated questionnaires.

The subsequent Chapter 4 introduces metadata for the dissemination of DACSEIS results. These metadata are essential tools for the dissemination of the results in an operable form and to enable the user to apply the achievements of the DACSEIS project. The Appendix A contains an example for metadata and code dissemination.

The concluding Chapter 5 deals with definitions and applications of non-response modeling for the simulation study. Especially non-response for the Austrian Microcensus, the Dutch LFS, the Finnish LFS, the German Microcensus and the Swiss HBS are analyzed in a detailed manner for operable modeling in the simulation study. Tables and Graphs complete and illustrate the results. The whole chapter furnishes a lot of new and relevant scientific knowledge and many useful recommendations for the user´s practice.

The purpose of deliverable D1.2 is to prepare the ground for the large Monte-Carlo simulation study resulting in a thorough investigation of the state of the art methodology in variance estimation in terms of

- accuracy,
- computer efficiency, and
- practicability

of the methods. The conclusions of the results yield the recommended practice manual on variance estimation methods.

This deliverable is divided in two parts, the first one containing the description of the Monte-Carlo study and the second one presenting selected results of the simulations. Within the first chapter, the simulation study including its set-ups is described in detail. It also includes the programming environment under R, the connection of survey data and simulation tasks, as well as the final implementation of the simulation study to investigate the estimation methodology.

The second chapter presents three selected simulations. The first section refers to the repeated weighting in the Dutch LFS as a detailed investigation of the results under workpackage 7. Within the second section, the special features of the two-stage-design in the Austrian Microcensus are elaborated. Finally, the third section concludes with special simulations on the weighting in the Swiss HBS. The Appendix includes the codes from the different chapters and sections.

The main simulation study contains the comparative investigation of the methodology in the context of the six DACSEIS surveys. These will be presented in deliverable D12.2 in electronic form and in deliverable D12.3 as recommendations for their use in practice.

2.2 Workpackage 2: Structure and Analysis of Relevant National Surveys

One of the main targets of the DACSEIS project is the standardisation and harmonisation of variance estimation methods used to calculate the sampling errors in national surveys conducted by the National Statistical Institutes. To be able to generate the universes and to reproduce the relevant surveys by simulations these surveys had to be described. This was done in workpackage 2 for the five relevant Labour Force Surveys (=LFSs): The Microcensus of Statistics Austria, the LFS of Statistics Finland, the Microcensus of the German Statistische Bundesamt, the LFS of the CBS Netherlands and the LFS in the United Kingdom, which has been carried out in Great Britain by the Office for National Statistics. Also the German Sample Survey of Income and Expenditure and the Swiss Household Budget Survey conducted by the Swiss Federal Statistical Office were included.

For this purpose all partners involved in this workpackage delivered information of their national surveys. The whole report then included general remarks on each survey, the definitions of the underlying populations, the most important facts regarding the survey process, the description of the sampling frames, the sampling methods and the weighting procedures used in these surveys at the beginning of the project in March 2001. It also comprised a short overview over the known sources for the occurrence of a nonsampling error and - of course - a description of the methods for variance estimation currently used.

2.2.1 The Labour Force Surveys

Starting with the sampling frames used in the five LFSs, the comparison showed big differences between the different countries: The basis from which the main sampling frame for the Austrian Microcensus is built is the dwellings stock of the Austrian Housing census, which is conducted every ten years. Therefore all inhabited and uninhabited dwellings are the sampling units of this survey. The sampling frame for the Finnish LFS is built from the quite up-to-date Central Population Register. Therefore persons are the sampling units in Finland. In Germany the Population Census and the census of buildings and housing in West-Germany and the Population Register Statistics in East-Germany are used to build the sampling frame, which then offers buildings as the sampling units for the German Microcensus. The LFS of the Netherlands has a sampling frame, that is based on the Geographical Base Register and the register of houses in Amsterdam. These two registers are combined into one list of addresses, which then are the sampling units of the Dutch LFS. The LFS of the UK uses a subfile of the Postcode Address File, the telephone directory in a small area in the North of Great Britain and a valuation list for rating purposes in Northern Ireland to have a list out of which households as sampling units can be drawn.

Coming to the next facts for the description of the various LFSs Table 2.1 compares the intervals, in which the surveys are conducted and the approximate sample sizes (measured in responding persons).

Table 2.1: Some facts of the five included LFSs

Country	**Interval**	**Appr. sample size**
Austria	annually	53,000
Finland	monthly	10,000
Germany	annually	790,000
Netherlands	monthly	11,000
United Kingdom	quarterly	140,000

The quarterly Austrian Microcensus, of which the LFS is a part of once in a year, includes therefore about 0.7 % of the Austrian population. In Finland this percentage is 0.2 per month, in Germany, where the LFS is integrated in the German Microcensus, it is 1 % annually, in the Netherlands 0.07 % monthly and in the United Kingdom it is 0.25 % quarterly.

In each country a rotating panel system is used. This means, that in Austria for instance, the survey units stay in the quarterly Microcensus for eight successive times, meaning that one eighth of the sampling units is substituted by new ones each time, but only for two LFSs (because the LFS is once a year part of the Microcensus; see Table 2.1). In Finland and the UK there are five so called waves, in Germany and the Netherlands there are four.

Looking now at the different sampling methods used in our surveys, we find that two countries use rather simple ones: In Finland by means of a systematic selection of persons out of the Central Population Register we get a stratified random sample of these sampling units with proportional allocation of the sample on the geographical regions of Finland. In the UK it is more or less the same with households as sample units.

In the Dutch LFS like in the UK we have a geographically stratified sample of households, but within the strata a two stage sampling design is used with municipalities as PSUs and households as SSUs.

In Germany a stratified random sample (with stratum variables region and size of building) of clusters of households is conducted. These clusters are defined as buildings, parts of big buildings or some small buildings. The clusters are selected by a semi-systematic procedure, which leads to the above mentioned stratified sample.

The most complex sampling procedure is used in Austria. The dwellings of the Austrian sampling frame are partitioned into two groups: Part A with the dwellings of large (mainly urban) municipalities and part B with the dwellings of small (mainly rural) municipalities. Within these two strata different sample methods are used, that can roughly be described as stratified random in Part A resp. stratified two stage sampling of dwellings in Part B of the universe. In Part A the stratification of the dwellings is done with the federal states and some dwelling characteristics, whereas in Part B the stratification is done by federal states and some municipality characteristics. Within these strata as PSUs municipalities and then within these PSUs dwellings are drawn as SSUs.

In all of these surveys a weighting procedure in the estimation process is included with step one to compensate for nonresponse and step two to adjust the sampling distributions of certain variables to their population distributions. For the calibration of step two of the process in all countries more or less the same variables are used. The used variables in this iterative process are sex, age, region, nationality, ethnicity and marital status (only in the Netherlands) and reference week (only in Finland).

Comparing the non-response rates, we can find, that two of the surveys have non-response rates of about 10 to 15 %. That are Austria (there differing from the other countries nonresponse occurs only as the nonresponse of all persons in sampled inhabited dwellings) and Finland. In Germany this rate is significant lower with only about 3 %, whereas in the UK it is beyond 25 and in the Netherlands between 40 and 50 %.

Looking at last at the variance estimation methods used by the National Offices to estimate the variance of a total estimator, we can find that in Austria due to the fact, that for the consumers of the Microcensus-results it is sufficient to get a rough idea of the size of this error, the published sampling error is estimated by the use of the formulas for stratified random sampling, which leads - as we know now after the simulation studies - to an underestimation of the true variance. In the Finnish LFS the sampling variance is approximated by Statistics Finland by the variance estimator for the general regression estimator. The variance of a total is currently estimated in Germany by a formula, that takes into account the selection of clusters of persons and the stratification of these clusters, but does not include the effects of regional stratification and bounded estimation. Statistics Netherlands does not estimate variances for the LFS results at all. At last in the UK the standard errors of total estimates are estimated by multiplying the simple random sample errors with an estimated design effect.

2.2.2 The Household Budget Surveys

There are two surveys of the family of surveys of the household budget in the DACSEIS project. The German Sample Survey of Income and Expenditure and the Swiss Household Budget Survey. The main characteristics - the approximate sample sizes and the used sampling methods - of these two surveys, that before the start of the project were at last conducted in 1998, are included in Table 2.2. Of course, for the purpose of an HBS, households are the sampling units for both surveys.

Table 2.2: Some facts of the two included HBSs

Country	**Appr. sample size**	**Sampling method**
Germany	74,000	Quota sampling with combined quotas
Switzerland	9,000	Stratified random sampling

The German sampling frame itself is - as it is built from the private Microcensus households with a monthly net income of less than 17,500 Euro - a sample of the population of households of Germany. The selection of households for the quota method is done by the office using region, household type, social status and the household's net income as quota variables and disproportional to size allocation of the total sample number to the combined quota cells.

In Switzerland the telephone register is used as sampling frame. The total sample number is allocated disproportionally to the strata in Switzerland, too.

The weighting is done in the following way in these two surveys: Disregarding the methodological reservations for quota sampling, in the German sample survey at first the nonresponses are compensated within each quota cell and then the disproportional to strata sample is adjusted to the distribution of the combined quota variables in the Microcensus.

In the Swiss HBS at first inclusion probabilities and at second response probabilities are calculated for households using two different levels of response (level 1 means, that the household gave only the information about size, socio-economic group and nationality of reference person and level 2, that the household completed the participation) and with these factors the answers of the different households are weighted.

As always in market and opinion research one cannot calculate Nonresponse rates for a quota sample. So this cannot be done for the German Sample Survey of Income and Expenditure, too. In the Swiss HBS the Nonresponse rate was 50 % at level 1 of response and 70 % at level 2 of the survey.

For the Swiss HBS no variance estimation method was used so far. For the German survey for some important variables the variances were estimated by the use of the formulas for stratified random sampling based on the assumption that the purposive quota sample can be treated like a stratified random sample.

2.3 Workpackage 3: Monte-Carlo Simulation Study of European Surveys

Workpackage 3 contains the results of two deliverables D3.1 and D3.2. They form the essential basis of the DACSEIS study which mainly depends on two items:

- a thorough study of estimators and variance estimators,
- universes as a basis of the Monte-Carlo simulations to test the estimators and variance estimators in practical complex environments.

In order to make the large simulation study possible, adequate universes had to be generated which allow for applying the sampling schemes in practice. The universes to be considered consist of household and individual data; they include labour force and related surveys as well as household budget and consumption surveys. The surveys of interest were the Dutch and Finish labour force survey (LFS), the Austrian and German Microcensus, the Swiss household and budget survey (HBS) as well as the German Income and Expenditure survey (EVS). Detailed information on these surveys can be taken from the report on workpackage 2 of the DASCEIS project.

The purpose of this workpackage 3 was to provide the basis-structure for the Monte-Carlo simulation study which is placed at workpackage 1 interacting with the methodology from the workpackages 5 to 11. It consists of the generation of adequate universes and suitable interfaces for the national sampling schemes, the estimators and variance estimators, as well as the possible inclusion of different peculiarities of the national surveys. The universes had to be created from sample data since no adequate universe data, e. g. census data, were available. The simulation study itself followed a general flow chart which can be seen in Figure 1.1, Chapter 1 of WP 3 and the corresponding description.

Within this WP3 adequate *true* universes had to be generated. One major conflict of this task had to be solved; respecting the conflicting targets below its solution can only be an approximation:

- The best possible mechanism to create the universes should allow for rebuilding marginal distributions as well as interaction between variables.
- This mechanism should also allow for heterogeneities between subgroups, especially for regional aspects.
- Pure replication of units, which seems to be suitable for the first two items, should be avoided because this leads to a undue small variability of units within smaller subgroups.

- The use of too many variables on the microlevel should not end up in non-disclosure difficulties.

The interaction of these items and their observation seems to lead to an unsolvable conflict. However, the aim was to find best recommendations for the practical use of variance estimators in a practical environment. Therefore, several additional plausible assumptions were made. Still, these assumptions should not have any negative influence on the generation process.

Another section describes the general model of creating the universes for household and individual surveys. A very important topic in this context was to consider the national figures like size, regions, and population characteristics and to arrange them in a suitable manner. The survey specific implementations including the variable lists and some important figures of the data are presented as well.

The main achievements of this Chapter 3 consist in setting up and generating adequate universes for the simulations which are applicable tools to analyze the properties of estimates in a practical environment. The results of this recent work will have great impact on the effort to get improved data quality.

2.4 Workpackage 4: Review of Software Packages for Variance Estimation

The software evaluation of DACSEIS included eleven software (R, S-Plus, SAS, SPSS, Stata, Bascula, Clan, Genesees, Poulpe, Sudaan, WesVar). The properties and features of these software were described in general, but R, S-Plus and Genesees were excluded from the final, more detailed study. Although having potential (especially R) for very advanced program construction regarding survey sampling, R and S-Plus software weren't actually designed for any survey problems. At the beginning of the DACSEIS project the predecessor of Genesees, i.e. GSSE (developed by Statistics Italy), was not available for purchase in general, so it was excluded from the tests. SAS, SPSS and Stata were classified as general software and Bascula, Clan, Poulpe, Sudaan and WesVar were described as advanced software.

Three sample data were selected from the DACSEIS universes representing Finnish Labour Force Survey (FLFS), Swiss Household Budget Survey (SHBS) and German Microcensus (GMC). Those data were used for software evaluation purposes, following the design and estimation requirements in harmony with the structure of the simulation studies. However, the testing procedure included only such methods which were available in the software and produced proper variance estimates taking the survey design into account. Thus e.g. the imputation methods were not tested, because there were no software to deal with them in that sense. Two estimators were used for every survey: a Horvitz-Thompson estimator and a GREG/calibration estimator utilising auxiliary information from the population level.

Every software provided the same estimates of the parameters to be studied, so in that sense the software are equal. The variance estimators based on the Taylor linearisation in various software were at the same level as well. Furthermore, the resampling variance estimators were equal or nearly equal to the linearisation estimators, except the artificial Balanced Repeated Replication in Bascula, which gave variance estimates differing greatly from others (note that the Taylor-based variance estimates of Bascula were parallel with the results of other software). The parameters to be estimated here were rather simple (total of unemployed and unemployment rate for FLFS and GMC; mean income and mean expenditure for SHBS), so there should be rather marginal differences between the variance estimation methods with these large data sets. The differences between the software came in three main criteria: 1) how advanced the software is regarding the designs and the estimation methods, 2) what properties there are available serving the tasks needed in surveys, 3) how fast the software is and how the large data sets can be dealt with.

Concerning the test situation, the sampling designs in question (stratified simple random sampling without replacement, stratified cluster sampling) were available in all software. Every software produces the Horvitz-Thompson estimator with a variance estimator for these parameters to be estimated (provided that the size of the data set is not too big). The GREG estimators with proper variance estimation could be obtained only in Bascula, Clan and Poulpe. In addition to these three software containing also calibration there was a raking method in WesVar for making a calibration estimator.

One important factor is how the data preparations (e.g. stratum sizes, number of respondents in strata, finite population corrections, weights, classifications and/or recoding) can be conducted. The required preparations in full were possible only in SAS, SPSS and Stata. Recoding and classification were available in Bascula, Sudaan and WesVar. Otherwise the preparations had to be done in another environment (in SAS in this test pattern).

The GMC data sets provided a good test for the speed and memory use of the software. Even for the smallest regions the replication weight production for primary sampling units in WesVar caused considerably longer processing times than in other software. As a limit, WesVar could deal with Schleswig-Holstein with 29140 records, 35 strata and 1680 primary sampling units. Correspondingly, the largest data set for Stata was Niedersachsen with 74334 records. It is possible to extend the memory of Stata with some special definitions. In general, Stata performed the calculations rather fast. SAS-based software, Bascula and SPSS could be utilised for the full GMC data (831004 observations). SAS (proc surveymeans), SPSS and Sudaan conducted the task sufficiently, although SAS (surveymeans) had memory difficulties with some domain categorisations, at least in the test computer. The advanced software doing additional calculations for more complex demands required more time. Clan performed rather well but Bascula and Poulpe were time-consuming with the full GMC data.

The replicate weights made the WesVar data sets very large even at the level of a few thousand observations (e.g. for 5000 observations of FLFS 442 Mb). Poulpe has a structure which is prepared for very complex design and estimation processes; it produced a large number of different data sets and in total they reserved a lot of space, though much less than WesVar. In Bascula the survey data set in the Blaise form was rather large. Also Clan made additional data sets for estimation purposes, but the amount of memory they reserved was much less than in Poulpe. In practice the SAS survey data set is bigger than the corresponding Stata data set and the SPSS data set. Proc Surveymeans of SAS and Sudaan did not produce any additional data sets with these tasks.

In addition to the tests pattern for three DACSEIS data, a simplified test was conducted with the GMC data sets in order to test the speed of R coding and SAS-related survey software (Proc surveymeans, Sudaan, Clan). The task was reduced to Horvitz-Thompson estimation at the general level as if the sample was selected as element sampling. The results with different data combinations revealed that with this task R was faster in preparations and SAS in estimation. R couldn't deal with the GMC data as a whole which was due to limited memory. The estimation was possible when the two largest regions were removed. Sudaan (Descript procedure) was slightly slower than SAS Surveymeans. As stated before, Clan conducts many additional data phases in SAS and thus its calculations took longer than for the others. Still the performance of Clan was quite fair.

2.5 Workpackage 5: Resampling Methods for Variance Estimation

Standard textbooks contain formulae for the variances of a wide variety of sample survey estimators. The complexity of modern sample survey procedures, however, often does not allow the derivation of such formulae or undermines their accuracy, and an alternative is the use of resampling methods. An important contribution made by workpackage 5 is the identification of resampling methods suitable for use with the complex surveys described in workpackage 2, the comparison of their performance with other methods of variance estimation, and the provision of `R` code to compute them for the major simulation studies performed as part of DACSEIS. Deliverable D5.1 contains a more comprehensive version of the discussion below, with copious references and a summary of previously performed simulations, including a brief account of work performed by BOONSTRA and NIEUWENBROEK (2003) as part of DACSEIS. Deliverable D5.2 contains a description of the key methods chosen for implementation and the corresponding `R` code.

Consider complete response for a stratified single stage unequal probability sampling scheme without replacement, with N units divided into H strata, from which n units are sampled. Let n_h be the number of units sampled from the N_h population units in stratum h, and let π_{hi} be the inclusion probability for unit i of this stratum. Thus the total numbers of units in the population and in the sample are $N = \sum_{h=1}^{H} N_h$ and $n = \sum_{h=1}^{H} n_h$. Let x_{hi} and y_{hi} be variables that have been measured on the units, where y_{hi} is the scalar response of interest and x_{hi} is a $q \times 1$ vector of auxiliary variables, which may be continuous, categorical, or both.

Parameters of the finite population can be classified into two groups, depending on whether they are smooth (e.g., total, ratio of totals, temporal change in the ratio) or non-smooth (e.g., median, other quantiles) functions of the finite population responses. Estimation of these parameters is based on the data from the n sampled units and on their inclusion probabilities under the sampling design. The most important estimator of a total is the Horvitz-Thompson estimator $\widehat{\tau} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \omega_{hi} y_{hi} = \omega^{\mathrm{T}} y$, where y is the $n \times 1$ vector of sampled responses and ω is the $n \times 1$ vector of their weights $\omega_{hi} = 1/\pi_{hi}$, the inverse inclusion probabilities. The exact variance of this estimator involves the joint inclusion probabilities for the different units and can be used to approximate the variances of other estimators which are functions of totals, such as the ratio.

In many cases population totals are known for some of the auxiliary variables x, and this information can be used to increase precision of estimation, in particular by making some allowance for unit non-response. Suppose that q_C marginals of the q auxiliary variables are known, with $q_C \leq q$, let c be the $q_C \times 1$ vector of known marginals, and let X_C denote the $n \times q$ matrix of auxiliary variables whose marginal total for the entire population is known to equal c. Then the quality of the Horvitz-Thompson estimator can be improved by calibrating the weights w_{hi} to be as close as possible to the original weights ω, subject to the constraint that the weighted auxiliary variables match the marginals. In one simple case the calibrated weights equal

$$w = \omega + \Omega X_C (X_C^{\mathrm{T}} \Omega X_C)^{-1} (c - X_C^{\mathrm{T}} \omega), \tag{2.1}$$

where Ω denotes the diagonal matrix whose elements are the ω_{hi}, and the corresponding calibrated Horvitz-Thompson estimator of the total is $\widehat{\tau} = w^{\mathrm{T}} y$.

In practice observations are often missing due to an unknown non-response mechanism. There are two main types of non-response: unit nonresponse, which is often dealt with by calibration or another form of weight adjustment; and item nonresponse, which can be dealt with by imputation of the missing data. The imputation model is used to predict the missing responses for units with item non-response, and is generally constructed using those respondents with complete data. Such models and the imputed values they provide are of two types: deterministic, meaning that the model imputes a deterministic function of the observed data; and stochastic, meaning that the model imputes a random function of the observed data. 'Hot deck' imputation is one common form of stochastic imputation.

A common deterministic approach to imputation of missing responses based on the corresponding vectors x_{hi} of auxiliary variables is to use a linear model across strata, $Y_{hi} = x_{hi}^{\mathrm{T}} \beta + \epsilon_{hi}$, a different linear model for each stratum, $Y_{hi} = x_{hi}^{\mathrm{T}} \beta_h + \epsilon_{hi}, \quad h = 1, \ldots, H$, or one of its generalizations, such as robust or logistic regression. Let $z_{hi} = I(y_{hi} \neq \texttt{NA})$ be the indicator random variable corresponding to observed response, let $Z = \mathrm{diag}(z)$ be the $n \times n$ diagonal matrix of these indicators, and let X be the $n \times q$ matrix that contains the auxiliary variables corresponding to both respondents and nonrespondents. Also let $\widehat{y} = X\widehat{\beta}$ represent the $n \times 1$ vector of fitted values from the regression model used for imputation. Then (2.1) implies that the calibrated and imputed Horvitz-Thompson estimator may be written as

$$\begin{aligned} \widehat{\tau} &= w^{\mathrm{T}} \{ Zy + (I - Z)\widehat{y} \} \\ &= \omega^{\mathrm{T}} Zy + (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} X_C^{\mathrm{T}} \Omega Z y \\ &\quad + \omega^{\mathrm{T}} (I - Z) X \widehat{\beta} + (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} X_C^{\mathrm{T}} \Omega (I - Z) X \widehat{\beta}. \end{aligned} \tag{2.2}$$

Computation of the variance of the calibrated and imputed Horvitz-Thompson estimator as if the imputed responses $\widehat{y}$ were true responses can lead to considerable underestimation of the true variance, so the variance estimation technique must reflect the variance inflation due to imputation. This is relatively easily accomplished using a resampling method, because an estimator such as (2.2) is simply treated like any other estimator. A simpler approach is to use a standard linearization formula as if the imputed values had been observed, but this can result in severe underestimation of the variance and is not recommended. A powerful variant is multiple imputation, whose implementation in the DACSEIS setting is described in deliverable D11.2.

When an estimator can be expressed as a differentiable function $\widehat{\theta} = g(\widehat{\tau})$ of a vector of linear estimators $\widehat{\tau}$, e.g., the ratio estimator $\widehat{\theta} = g(\widehat{\tau}_{Y_1}, \widehat{\tau}_{Y_2})$ with $g(x_1, x_2) = x_1/x_2$, linear Taylor series expansion of g about the population mean τ of $\widehat{\tau}$ yields

$$\widehat{\theta} = g(\widehat{\tau}) \doteq g(\tau) + \nabla g(\tau)^{\mathrm{T}}(\widehat{\tau} - \tau),$$

whose variance is $\nabla g(\tau)^{\mathrm{T}}\mathrm{var}(\widehat{\tau})\nabla g(\tau)$. The variance estimator is obtained by replacing unknowns in this formula with the estimators $\nabla g(\widehat{\tau})$ and the empirical covariance matrix for $\widehat{\tau}$. A similar expression is derived for stratified samples.

A more general approach to linearization for a compactly differentiable statistic $\widehat{\theta}$ is though the infinitesimal jackknife. The extra generality stems from use of a von Mises rather than Taylor series expansion of the statistic $\widehat{\theta}$, enabling theoretical variance formulae to be obtained for estimators such as the sample median and other quantiles. The linearization variance estimate requires theoretical derivations which we undertook as part of deliverable D5.2, in order for comparisons to be made with other variance estimators in the presence of missing data.

The jackknife involves the systematic deletion of groups of units at a time, the recomputation of the statistic with each group deleted in turn, and then the combination of all these recalculated statistics. In survey practice this generally entails too many computations, and groups of units must be deleted. Such schemes are less computationally heavy than the usual bootstrap, but they tend to give unstable variance estimates.

Balanced half-sampling was originally developed for stratified multistage designs with two primary sampling units drawn with replacement in the first stage. It is generally applied using groups of units, in some cases using artifical strata introduced in order to stabilise the resulting variance estimates. It has some theoretical advantages over the jackknife - for example it can be used to estimate variances of non-smooth statistics - but is prone to the same type of computational difficulties.

The bootstrap idea is to mimic how the original data were generated. Like the balanced repeated replication and the jackknife methods, the bootstrap involves recomputing the statistic, now using resampling from an estimated population to obtain bootstrap samples. For stratified data, the resampling is performed independently within each stratum. The standard bootstrap uses sampling with replacement, corresponding to independent sampling from an original population, but this must be modified for use with survey data. The method is therefore rather computer intensive, but on the other hand gives consistent variance estimators for medians and other estimators based on quantiles, and has been adapted for use with calibrated and imputed estimators.

As part of deliverable D5.1, a simulation based on data from the Swiss Household Budget Survey was used to compare resampling and other variance estimation methods for the calibrated and imputed Horvitz-Thompson estimator, based on complete data from $N = 9275$ households in $H = 7$ strata of various sizes. Also available on each household is a set of 14 auxiliary variables, of which 10 population margins are known. For the simulation,

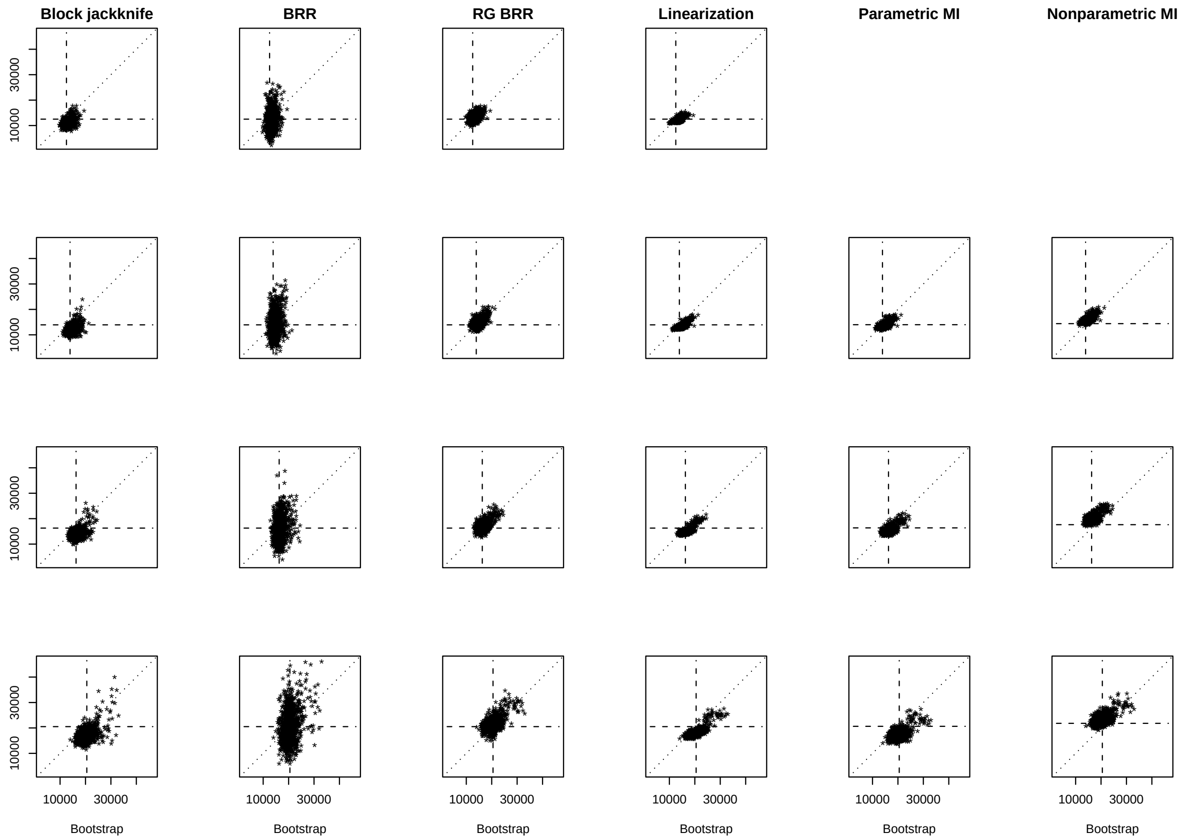

Figure 2.1: Comparison of resampling standard errors in the presence of calibration and imputation, as a function of the proportion of missing data; from top to bottom 0%, 20%, 40%, 60% item non-response. The dashed lines are the true sampling standard errors, obtained from a larger simulation, and the dotted line shows $x = y$. Simulation based on the 1998 Swiss Household Budget Survey.

we consider the $N = 9275$ households as the whole population, for which we know the total expenditure, and perform stratified random sampling without replacement and with equal inclusion probabilities 1/8 within 6 strata, and 3/8 in the other stratum, giving a sample size of 1332. Item non-response for the response variable is applied using a uniform probability of missingness across the entire sample. On each of the 500 samples simulated, we then calculate the calibrated and imputed Horvitz-Thomson estimates, and apply various estimation techniques to obtain variances for them. Figure 2.1 compares the performances of these variance estimation techniques for missingness rates of 0%, 20%, 40%, and 60%. The panels at the right of the figure generally show less variability in the variance estimates, with linearization performing particularly well in this study, followed by the bootstrap and by the multiple imputation methods - which however tend to give more biased variance estimates when the level of missingness is high.

These and other simulation results computed as part of workpackage 5 suggest that bootstrap and linearization variance estimators are most promising for use with complex surveys. The bootstrap has the advantage of being a general-purpose tool which can be applied without much tuning in many situations, and which can be used for both smooth and non-smooth statistics. Moreover and unlike the jackknife or balanced repeated replication, the number of recomputations needed is to a large extent controlled by the user rather than determined by the method. Its main disadvantages are its computational burden, particularly when used with imputation, and the fact that special programming is needed if it is applied to situations with large sampling fraction. Jackknife linearization demands special computation of influence function values adapted for particular circumstances, but it involves no resampling and so is much quicker than the bootstrap. Balanced repeated replication and the jackknife are almost competitive in some cases, but overall they perform worse than the other methods, and tuning seems to be needed to get the best performance from them. Multiple imputation is often a better alternative, and its computational requirements are generally relatively modest.

2.6 Workpackage 6: Variance Estimation for Unequal Probability Designs

Survey sampling textbooks often refer to the Sen-Yates-Grundy (Sen, 1953; Yates and Grundy, 1953) variance estimator for use with without-replacement unequal probability designs. This estimator is rarely implemented, because of the complexity of determining joint inclusion probabilities. In practice, the variance is usually estimated by simpler variance estimators such as the Hansen and Hurwitz (1943) variance estimator; which could leads to overestimation of the variance for large sampling fraction that are common in business surveys.

Unequal probability sampling was first suggested by Hansen and Hurwitz (1943) in the context of with-replacement sampling. Horvitz and Thompson (1952) developed the corresponding theory for sampling without-replacement. Variance estimation for sampling with-replacement is straightforward (Hansen and Hurwitz, 1943). However, for sampling without-replacement, the design unbiased Sen-Yates-Grundy variance estimator is hard to compute due to the joint inclusion probabilities. Although exact computation of these probabilities is possible with specific sampling designs, their calculation becomes practically impossible when the sample size is large. It is also currently inconceivable to provide these probabilities in released data-sets, as the set of joint inclusion probabilities is a series of $n(n-1)/2$ values; where $n$denotes the sample size. Moreover, standard statistical packages like SPSS, SAS, and STATA do not deal with these probabilities. Specialized software like SUDAAN needs to be used. However, even SUDAAN does not include actual computation of these probabilities. They need to be provided by the user.

We will consider alternative variance estimators that depend on the first-order inclusion probabilities only and are usually more accurate than the Hansen-Hurwitz estimator for large sampling fraction. The Hájek (1964) and the Brewer (2002) variance estimators are free of joint inclusion probabilities estimators for high entropy sampling design. The systematic sampling design is not a high entropy sampling design. Berger (2003) showed how the Hájek variance estimator can be extended to accommodate this sampling design. These estimators can be computed using the Splus library varianceht available at

`http://www.socstats.soton.ac.uk/staff/berger/unequal.html`

Berger (2004b) shows how weighted least squared can be used to compute the Hájek variance estimator. Standard softwares can be used to compute these estimators.

In this workpackage, we compare these estimators via simulation based upon the German Income and Consumption Universe (ICS). The ICS survey is a quota sample selected in each German federal state. The ICS data does not contain individual, but only household information. Based on these data, populations or universes have been created (see workpackage 3). The target parameters are 23 totals estimated by the Horvitz and Thompson (1952) estimator. We propose to select unequal systematic samples per federal states. Systematic sampling is widely used by statistical offices and it can be considered as an approximation of quota sampling. With a linear model, we generate a size variable correlated with the variable expenditure and with a coefficient of correlation of 0.6. The first-order inclusion probabilities are proportional to this size variable.

In each German federal state, we have selected 10 000 systematic samples with unequal probabilities. We suppose that each German state is a single stratum. We will compare the empirical sampling distribution of the simple random sampling variance estimator, the Hansen-Hurwitz variance estimator, the Hájek variance estimator, the Brewer variance estimator and the modified Hájek estimator (Berger, 2003) that take the systematic sampling design into account. We did not consider the Sen-Yates-Grundy variance estimator as this estimator is highly biased and not recommended for systematic sampling.

From this simulation, we concluded that the simple random sampling variance estimator has a different distribution; this is also the less accurate estimator. The other estimators have similar distributions and accuracy. Their distributions are also highly skewed, although unbiased. This can be explained by the small sampling fraction. With larger sampling fraction, we would obtain different results. The modified Berger's modified estimator can be slightly better.

In conclusion, variance with unequal probability sampling without replacement can be easily estimated with the Hájek (1964) variance estimator. This estimator can be easily computed as it is a weighted sum of residuals. This estimator expression is computationally simpler than the Sen-Yates-Grundy variance estimator and does not require computation of joint-inclusion probabilities.

2.7 Workpackage 7: Variance Estimation in the Case of Combining Register and Survey Data

2.7.1 Introduction

In the classical way of survey estimation, the set of weights is held constant per survey; see, e.g., Särndal *et al.* (1992). Such a unique set of weights for each survey makes it easy to compose various tables from the same survey. However, a consequence of the classical approach is that multidimensional tables from two or more surveys which have a variable in common, may have different numerical values for the same variable. The main aim of workpackage 7 is to develop an alternative weighting procedure that leads to a set of tables which are numerically consistent even when they are based on different surveys. In particular, this includes the derivation of the corresponding variance formulas.

The outline of this overview is as follows. Based on a questionnaire and some additional sources Section 2 summarizes briefly the various manners in which the NSI's are using auxiliary information from available registers, including the weighting schemes they apply. Section 3 describes the repeated weighting (RW) procedure for obtaining numerically consistent tables and the corresponding variance formulas. The applicability in The Netherlands and other European countries is discussed. Finally, Section 4 presents the results of a simulation study that was carried out in order to get insight into the performance of the RW estimator and the variance estimators proposed in this workpackage.

2.7.2 Use of register data in the European countries

In 2002 a questionnaire on the use of register data for Labour Force Surveys (LFS) was completed by 16 European NSI's. For a number of reasons nine countries didn't use register data in 2002 for the LFS. The main reasons mentioned in the questionnaire were: matching key problems, the law, and absence of a good population register. In contrast, seven countries used registers in combination with the LFS. The main goals for using registers are stratification, poststratification, calibration, nonresponse correction, bias reduction, regression estimator, imputation, sample selection, and special weighting procedures. The most important registers used by the NSI's are the registers on population, unemployment, job-seekers, education, and labour classes. For the countries using registers there were almost no matching problems or problems with ambiguous definitions of a variable. Although not all countries use registers, all of them make use of weighting schemes for the Labour Force Surveys. The most important variables included in the weighting scheme are sex, age, and region. Some other variables are marital status, employment status, nationality, and labour class; for further details on the questionnaire, see DACSEIS deliverable D7.1.

2.7.3 The Repeated Weighting (RW) estimator

This section describes briefly the RW estimation procedure. Throughout this section we assume that (i) the reference period of the surveys is the same, (ii) there is no definitional difference of a variable that is included in two tables from different surveys or registers, and (iii) 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; for more details and refinements, see DACSEIS deliverable D7.2.

The RW estimator

The main aim of the RW estimation procedure is to obtain a set of tables, which are mutually consistent in a numerical sense. Regarding a given reference period, a set of target tables is defined. Globally speaking, the proposed RW estimation procedure consists of the following three steps.

Step 1. Ordering of the tables
Order the tables in such a way that the margins of a multidimensional table are always estimated first, possibly from a register. In general, this means that according to the so-called splitting up procedure tables with less variables are estimated before tables with more variables. Also tables with categorical variables from a less fine classification are estimated first; for further details, see D7.2. The splitting up procedure is a practical way of dealing with the order problem, i.e., the problem that estimation results depend on the ordering of the tables to be estimated.

Step 2. Regression estimation of the tables
Estimate each table from the most appropriate data set. In general, this will be the largest survey or a combination of surveys, also called a block and denoted by B. Apply the general regression estimator to block B. Let a_i denote the vector of auxiliary variables for the ith element in the regression. Then the regression estimator of a (vectorized) table Γ can be written as

$$\begin{aligned}
\widehat{t}_{\Gamma}^{w(B)} &= \widehat{t}_{\Gamma}^{d(B)} + (\widehat{B}_{\Gamma;a}^{d(B)})^t (t_a - \widehat{t}_a^{d(B)}) \\
&= \sum_{i \in B} w_i^{(B)} \Gamma_i \\
\widehat{B}_{\Gamma;a}^{d(B)} &= (\sum_{i \in B} d_i^{(B)} a_i a_i^t)^{-1} \sum_{i \in B} d_i^{(B)} a_i \Gamma_i^t \\
w_i^{(B)} &= d_i^{(B)} \{1 + a_i^t (\sum_{i \in B} d_i^{(B)} a_i a_i^t)^{-1} (t_a - \widehat{t}_a^{d(B)})\} \\
\widehat{t}_y^{d(B)} &\equiv \sum_{i \in B} d_i^{(B)} y_i \qquad \text{for an arbitrary variable } y \\
d_i^{(B)} &= \tfrac{\lambda_i}{\pi_i} \qquad (0 \leq \lambda_i \leq 1).
\end{aligned}$$

The π_i stand for the first order inclusion probabilities. The actual values of the λ_i depend on the sizes and reliabilities of the (mutually disjoint) samples constituting block B. Per sample the λ_i are constant and for a given sample included in B the corresponding λ-value stands for the relative weight of that sample in the actual data block B. The $d_i^{(B)}$ can be seen as the *starting* weights in block B for table Γ. A well-known property of the block *regression* weights $w_i^{(B)}$ is that they satisfy the calibration equations

$$\sum_{i \in B} w_i^{(B)} a_i = t_a.$$

Step 3. The re-weighting procedure
When for a certain table the block weights $w_i^{(B)}$ lead to a margin that is numerically inconsistent with an estimate of that margin in a preceding estimated table of the set, that table should be reweighted. Such an inconsistency may occur when, for instance, such a margin was not included in the vector a_i of auxiliary variables or, in terms of calibration, the margin was not included in the calibration equations for the underlying block B; see Deville and Särndal (1992). By re-weighting we mean an adjustment of the original block weights $w_i^{(B)}$ for this specific table so that the margins of the re-weighted table are in line with the margins from the preceding estimated tables and/or registers.

That is, the RW estimator is defined in a recursive way by

$$
\begin{aligned}
\widehat{t}_{\Gamma}^{RW} &= \widehat{t}_{\Gamma}^{w(B)} + (\widehat{B}_{\Gamma;m}^{w(B)})^t(\widehat{t}_m^{RW} - \widehat{t}_m^{w(B)}) \\
&= \sum_{i \in B} r_i^{(B)} \Gamma_i \\
\widehat{B}_{\Gamma;m}^{w(B)} &= (\sum_{i \in B} w_i^{(B)} m_i m_i^t)^{-1} \sum_{i \in B} w_i^{(B)} m_i \Gamma_i^t \\
r_i^{(B)} &= w_i^{(B)} \{1 + m_i^t (\sum_{i \in B} w_i^{(B)} m_i m_i^t)^{-1} (\widehat{t}_m^{RW} - \widehat{t}_m^{w(B)})\}.
\end{aligned}
$$

where m is a vector consisting of margins of the present table Γ. The elements in $\widehat{t}_m^{RW}$ are estimates from a previous table or known counts from a register. By construction the r_i satisfy the corresponding consistency equations. That is,

$$\sum_{i \in B} r_i^{(B)} m_i = \widehat{t}_m^{RW}.$$

Currently, Statistics Netherlands is implementing the RW estimator in their regular estimation process to obtain consistency among tables. From the definition of the RW estimator it is clear that this estimator can also be applied in other European countries under the assumptions mentioned in the beginning of this section. In fact, the $r_i^{(B)}$ are only a cosmetic adjustment of the commonly used regression weights $w_i^{(B)}$ for obtaining numerical consistency in the table at hand. Furthermore, for implementing this RW estimation procedure it is important to have an appropriate metadata system underlying the micro databases from the surveys and available registers. That is, there is a software tool needed for the collection process of tables related to a given target table Γ. Such a tool is certainly necessary when there are many multidimensional tables to be estimated with variables with many different (hierarchical) classifications or variables like income in either categorical or quantitative form.

The variance of the RW estimator

In order to derive a formula for the variance of the RW estimator, consider the case with one register and two mutually disjoint, independent samples S_1 and S_2. For deriving the variance formulas for the RW estimator three basic principles are used. First, note that similar to the standard regression estimator the random character of $\widehat{B}_{\Gamma;m}^{w(B)}$ can be ignored for large samples. Hence, we insert $B_{\Gamma;m}$ in the formulas. Second, approximating $\widehat{t}_{\Gamma}^{RW}$ by

$$\widehat{t}_{\Gamma}^{RW} \approx \widehat{t}_{\Gamma}^{w(B)} + (B_{\Gamma;m})^t(\widehat{t}_m^{RW} - \widehat{t}_m^{w(B)}),$$

it is not difficult to see by induction that the RW estimator always can be written as a linear combination of regression estimators. Third, by construction any regression estimator, based on S_1, S_2 or their union, can be written as a linear combination of ordinary HT estimators from S_1 and S_2 plus a constant, provided that the estimated regression matrices are replaced by their counterparts from the population. Hence, denoting these linear combinations of HT estimators in S_1 and S_2 by the vectors $\widehat{t}^{HT}_{e_1(\Gamma)}$ and $\widehat{t}^{HT}_{e_2(\Gamma)}$, respectively, we can write $\widehat{t}^{RW}_{\Gamma}$ as

$$\begin{aligned}\widehat{t}^{RW}_{\Gamma} &= \widehat{t}^{HT}_{e_1(\Gamma)} + \widehat{t}^{HT}_{e_2(\Gamma)} + \text{ constant} \\ &= \sum_{i\in S_1} \tfrac{e_{1i}(\Gamma)}{\pi_{1i}} + \sum_{i\in S_2} \tfrac{e_{2i}(\Gamma)}{\pi_{2i}} + \text{ constant.}\end{aligned}$$

In DACSEIS deliverable D7.2 it is pointed out how the variables $e_{ki}(\Gamma)$, also called superresiduals, can be determined from the superresiduals $e_{ki}(m)$ corresponding to the margins m of table Γ $(k = 1, 2)$. Hence, table by table the required superresiduals can be determined recursively. Similar to the ordinary regression estimator, the (approximate) covariance matrix of $\widehat{t}^{RW}_{\Gamma}$ can be estimated in the standard manner by

$$V\hat{a}r(\widehat{t}^{RW}_{\Gamma}) = \sum_{k=1}^{2} \sum_{i,j\in S_k} (\tfrac{\pi_{k,ij}-\pi_{ki}\pi_{kj}}{\pi_{k,ij}}) \tfrac{\widehat{e}_{ki}(\Gamma)\widehat{e}^{t}_{kj}(\Gamma)}{\pi_{ki}\pi_{kj}},$$

where for the calculation of the estimated superresiduals $\widehat{e}_{ki}$ the estimated regression matrices are used.

2.7.4 A simulation study

One of the tables analysed in the simulation study was the table $SEX \times MST \times EMPL$ or, for short, SME. The variables sex and marital status (MST; 3 categories) are known from a register while the variable employment ($EMPL$; 3 categories) is observed in an SRS sample (S) for $n = 500$ as well as $n = 5000$. The simulation study was based on a part (188,216 persons) of the region Noord-Brabant from the Dutch universe created for the DACSEIS project; see DACSEIS deliverables D3.1 and D3.2.

In the simulations we carried out, the overall weighting scheme is $SEX \times MUN + AGE$, where MUN is a municipality variable with 10 categories while AGE is a variable with 6 categories. Hence, the vector a_i of auxiliaries consists in this case of 25 $(= 2 \times 10 + 6 - 1)$ variables. The SME table, estimated by means of the standard regression estimator, is consistent with the SEX counts from the register because this variable is included in the weighting scheme. However, the SME table thus obtained is numerically inconsistent with the M counts from the register. Therefore, according to the so-called splitting up procedure, the tables S, E, M, SM, SE, and ME should be estimated first; note that the tables S, M and SM can be counted from the register. Furthermore, the regression estimator of the table SE is already consistent with the register because S is included in the vector a_i of auxiliaries. In contrast, the regression estimator of table ME is not numerically consistent with the register, because M was not included in a_i. The re-weighting scheme for the table ME consists of 6 variables (margins), i.e., three for M and three for E; note that one is redundant. In the re-weighting procedure for the table ME the vector $\widehat{t}^{(RW)}_{M}$ consists of the corresponding counts from the register while the vector $\widehat{t}^{(RW)}_{E}$ stands for the ordinary regression estimator from S. That is,

$$\widehat{t}^{(RW)}_{m;ME} = \begin{pmatrix} t_M \\ \widehat{t}^{w(S)}_{E} \end{pmatrix}.$$

The re-weighted table ME, thus obtained, is denoted by $\widehat{t}^{(RW)}_{ME}$. For the table SME the vector $\widehat{t}^{(RW)}_{m}$ now takes the form

$$\widehat{t}^{(RW)}_{m;SME} = \begin{pmatrix} t_{SM} \\ \widehat{t}^{w(S)}_{SE} \\ \widehat{t}^{(RW)}_{ME} \end{pmatrix}.$$

The main findings from the simulation study on the RW estimator of the table SME in D7.3 are (i) even for small samples ($n = 500$) the relative RMSE (root mean square error) of the RW estimator never exceeded that of the regression estimator, (ii) for $n = 500$ in some cells of the table the relative RMSE of the RW estimator and the regression estimator exceed that of the HT estimator but never for $n = 5000$, (iii) for increasing n the relative (negative) bias of the proposed variance estimator of the RW estimator is decreasing similar to that of the standard variance estimator of the regression estimator, and (iv) the differences between the performances of the variance estimator described here and an alternative, simplified variance approximation, not discussed here, are small; for more results, see DACSEIS deliverables D7.2 and D7.3.

2.8 Workpackage 8: Allowing for Raking Adjustment

Raking ratio estimation is a form of calibration estimation which makes use of auxiliary information on population counts within the categories of several categorical variables. It may be viewed as a generalization of poststratification, where only a single categorical variable is used. Given auxiliary information on population counts, raking ratio estimation may be viewed as an alternative to generalised regression estimation (GREG). Raking ratio estimation appears to have a more well-established history of applications in many national statistical institutes (NSIs), perhaps because of its ease of computation, involving repeated use of standard post-stratification adjustments. In some NSIs, GREG has tended to replace raking ratio estimation. One reason is that the GREG can be expressed in closed form and computed in one step, whereas the computation of a raking ratio estimator is iterative. Perhaps a more important reason is that GREG can handle a wider class of forms of auxiliary information, including population totals of continuous variables, whereas raking is restricted to the use of population counts in the categories of discrete variables. Nevertheless, raking ratio continues to be used widely in NSIs in many countries, e.g. the USA and the UK. One advantage is that it always produces positive weights, whereas GREG requires modification to meet this condition. In addition, raking may reduce non-response bias more than GREG under certain assumptions. Although GREG and raking often produce similar estimates and are asymptotically equivalent under certain strong conditions, their properties still require further comparison, especially in the presence of unit non-response when these conditions will not hold in general.

This workpackage considers two forms of raking ratio estimation, the classical estimator obtained by the application of iterative proportional fitting as well as an estimator, which may be interpreted as a maximum likelihood estimator within a certain framework. The GREG estimator is also considered as a benchmark for comparison. The primary aim is to investigate alternative variance estimation approaches for the raking ratio estimators. In particular, we focus on linearization variance estimators and consider the choice between using design weights and raking weights both to weight the residuals and to weight the estimated regression coefficients when computing the residuals. We study the properties of these alternative variance estimators both with and without unit non-response. Data from the Great Britain Labour Force Survey (LFS) and the German Sample Survey of Income and Expenditure (SIE), two national surveys, are used to evaluate the properties of these estimators in simulation studies. Both a multiplicative and an additive non-response model are considered in the simulations. The complex designs used for both the LFS and the SIE are mimicked as far as possible in this investigation.

The simulation study shows little difference between the bias or variance properties of the three calibration estimators considered: the GREG estimator, the classical raking estimator and the maximum likelihood raking estimator. Some small differences in the distribution of extreme weights are observed. A few negative weights are observed for the GREG estimator, whereas weights are necessarily positive for both raking estimators. Some very large weights are observed for the maximum likelihood raking estimator, suggesting no advantage in this method, despite the fact that one might have expected it to demonstrate different bias properties in the presence of non-response.

Amongst the variance estimators, the main finding is the contrast between the 'standard' linearization variance estimator which weights residuals by the design weight and the 'jackknife linearization' variance estimator which weights residuals by the calibrated weight. It is found that the latter variance estimator tends always to have reduced bias and that this effect is very marked in the presence of non-response, when the former estimator could be severely biased. The bias of the jackknife linearization variance estimator is generally small and the coverage level of the associated confidence intervals is generally close to the nominal coverage.

Alternative ways of weighting the regression coefficients when calculating the residuals in the linearization variance estimator are considered but little effect is observed and there is no evidence that this choice is important.

In general, the findings for the categorical variables in the British Labour Force Survey are remarkably similar to the findings for the continuous variables in the German Income and Expenditure survey.

2.9 Workpackage 9: Variance Estimation for Change

Most surveys are continuing surveys; that is, repeated monthly, quarterly, annually or with some other fixed frequency. An important reason for doing this is to estimate the manner in which population parameters change from one wave (or survey period) to the next. There is considerable interest in estimation of changes between two waves of a survey (Smith *et al.*, 2003), for example change in the number of unemployed or in the unemployment rate.

Variance estimates of change are usually computed from estimates of variance and covariance between cross-sectional estimators at different waves. Variance and covariance estimation would be relatively straightforward if the same sample is selected at both waves. Unfortunately, this is rarely the case, as samples at different waves are usually overlapping sets of units.

In this workpackage, we considered three variance estimators: the Kish variance estimator (KISH, 1965), the Tam variance estimator (TAM, 1984) and a novel variance estimator (BERGER, 2004a). These estimators can be computed using the Splus library rot available at

http://www.socstats.soton.ac.uk/staff/berger/change.html

This library is also available for R. We compare these estimators by a simulation-based approaches based on the 2000 Finish Labour Force Survey. Variance estimation of change needs to take account of rotation schemes. Three rotation schemes have been considered: rotation with simple random sampling, rotation group sampling and rotation with systematic sampling.

The variance of change between two totals is the sum of the variances of the two totals subtracted by twice the covariance. Standard estimators can be used to estimate the variances. The covariance can be estimated from the correlation estimated from the matched sample (KISH, 1965, p. 457). However, if a large correlation is slightly over-estimated, the resulting estimator for the variance of change can under-estimate the variance significantly (BERGER, 2004a).

We propose an alternative estimator for the covariance. The proposed variance estimator is based on the Hájek approach (HÁJEK, 1964). First, we estimate the covariance unconditionally, that is, assuming the sample sizes random. Second, in order to capture the fixed size feature of the sampling scheme, we derive the conditional covariance given the numbers of units caught in the first wave sample, in the second wave sample and in the matched sample. An estimator for the covariance is an estimator of this conditional covariance. BERGER (2004a) showed how to extend the proposed estimator for variance estimation of change between estimators which are function of totals using Taylor linearization (e.g. ANDERSSON and NORDBERG, 1994).

The Finnish Labour Force Survey is a systematic sample from the Central Population Register. The individuals from this register are the sampling units. Based on these data, a pseudo Universes of 3 900 000 individuals has been created (see workpackage 3). Data are available for both waves (February 2000 and May 2000). For each individual, we know the labour force status and the education level at both waves. The region specifies the stratification. The parameters of interest are the total number of individuals in each labour force status and education level.

1000 samples are selected by stratum-by-stratum rotation schemes (rotation with simple random sampling, rotation group sampling and rotation with systematic sampling). For each sample selected, total non-response will be generated randomly according to probabilities of response. These probabilities will not be used for variance and point estimation. The response mechanism gives a response rate between 85% and 88% on both waves.

The proposed variance estimator (BERGER, 2004a) is the most accurate variance estimator especially with systematic sampling. The other estimators have a large negative bias. This is probably due to the fact that the correlation is over-estimated (see Berger, 2004a). Furthermore, it still gives unbiased estimates when the change is negligible. The other estimator proposed have large negative bias which can be explain by an overestimation of the correlation (see Berger, 2004a). As far as the method of rotation is concerned, the series of simulation suggests that the rotation with systematic sampling gives more precise point estimates and better variance estimates.

2.10 Workpackage 10: Variance Estimation for Small Area Estimates

The purpose of workpackage 10 is to investigate small area estimation methods in a practical environment. In contrary to the other methodological workpackages, the main focus is laid on applicability of standard small area estimators on German data. This results from a special co-operation between the FP5 projects EURAREA (cf. `http://www.statistics.gov.uk/methods_quality/eurarea/`) and DACSEIS. Following the recommendations from the EURAREA project, workpackage 10 investigates the performance of several small area estimators, synthetic and empirical best unbiased predictors, in comparison to classical estimators such as the Horvitz-Thompson and the GREG on the German Microcensus.

In modern survey sampling, small area estimation becomes of increasing importance, because of the growing demand for reliable small area statistics. Sample sizes in small subpopulations are rarely large enough for direct estimators to provide adequate precision.

This leads to the need to *borrow strength*, that is to use data with complementary information from some auxiliary variables of related areas or registers. By this procedure the effective sample size can be increased and the estimation precision can be enhanced.

Workpackage 10 consists of a short description of these estimators and in the following of special problems in DACSEIS concerning the German Microcensus data. Four very different types of small area classifications are distinguished and elaborated in a simulation study comparatively. The design of the small areas was chosen in order to find out peculiarities in the application of the methodology. Additionally to the simulations on the German Microcensus, the Finnish experience on real datasets was added in order to enable the user to examine the performance of the estimators in a different environment. The Finnish study is also part of the EURAREA project.

On the basis of selected tasks of the simulation study the different estimators are evaluated. As far as the Finnish data are concerned the analysis is distinguished into two parts. The first deals with the estimation of three different kinds of target variables: disposable income, unemployment and household composition. The variable mentioned last is the proportion of the single-person households. In that context estimates were calculated for NUTS3 and NUTS4 regions using $R = 500$ replications with a sample size of 12,000 individuals. The main part of these results is described in Office of National Statistics (2003). The second part, mainly described in workpackage 10, consists of the comparison of the performance of enhanced estimators (EBLUPs with time and area effect). This study is based on 1,000 samples of size 2,000. The results of that simulation study are specified by several measures such as relative bias, relative root mean squared error or confidence interval coverage rate.

The German part includes various simulations concerning different kinds of small area categorisations like the classifications by regional strata, house size classes, combination of the first two ones and regional strata divided by eight. The target variables unemployment, household composition and income were estimated using different kinds of covariates. In that context the simulation study included $R = 1,000$ replications.

As an additional task in workpackage 10, a study on the influence of non-response on the accuracy of the methods used was performed. Two methods for correcting for non-response were applied, multiple imputation on the one had and a calibration estimator with non-response correction on the other hand.

The state of the art in that field is integrated in the EURAREA reports. The DACSEIS report includes the particular simulation study on the German Microcensus data as well as the related Finnish experience. Some results of the obtained estimation quality were gained by simulations in a practical environment. The achievements of these chapters are of high relevance; they provide for practical usage in the DACSEIS Recommended Practice Manual (RPM).

2.11 Workpackage 11: Imputation and Non-Response

Sample surveys are subject to both *unit non-response* and *item non-response.* Unit non-response arises when no survey data are collected for a unit. Item non-response arises when some data are collected for a unit but values of some items are missing.

Different approaches to point estimation may be adopted in the presence of non-response. Some methods just ignore the non-response. In the case of unit non-response this will usually involve treating the set of responding units as if it were the selected sample. In the case of item non-response this may involve deleting units which have missing values on any of the variables used in a particular analysis (*available cases analysis*) or deleting units which have missing values on any of the survey variables (*complete cases analysis*). Such approaches may be subject to bias and, in general, do not make most efficient use of the data.

Weighting and *imputation* are the two main methods used to correct for bias due to non-response and to make efficient use of data. Weighting is classically used to treat the problem of unit non-response, whereas imputation is classically used to treat problems of item non-response.

Weighting is a 'unit-level' adjustment, providing a common form of adjustment for all analyses based a common set of responding units and is thus natural for the treatment of unit non-response. It is less practical to use weighting to treat item non-response, since a different method of weighting would be required for estimates based upon different sets of variables.

In contrast to weighting, imputation is a variable-specific adjustment and is thus natural to treat missing data in a given variable. Imputation tends to become more complicated and time consuming to implement the more the variables are treated and thus it is not usually considered as a practical solution for unit non-response in a survey measuring many variables.

In a classical frequentist framework for statistical inference, it is usual to summarise the properties of a point estimator in a sample survey in terms of its bias and variance. The presence of non-response will usually introduce bias into point estimators. A primary purpose of weighting and imputation methods is to reduce this bias. In addition to affecting bias, non-response will also affect the variance of a point estimator. The focus of this workpackage is on variance rather than bias. Particular attention is given to the question: what variance? Variances may be defined with respect to a number of stochastic mechanisms, including the sampling design, the non-response mechanism, models and stochastic features of an imputation method. Different statistical frameworks and ways of defining the variance are discussed.

In addition to considering alternative ways of defining variances in the presence of non-response, attention is given to the assessment of the impact of non-response on the variance. A new measure of variance inflation, the *neff*, is introduced.

Given an appropriate definition of the variance of a point estimator in the presence of non-response the main focus of the workpackage is on methods for estimating this variance. A number of existing methods are reviewed and some new methods are introduced with an emphasis on the estimation of variances for point estimators based upon imputed data. Only brief attention is given to variance estimation for point estimators which use weighting to compensate for non-response.

The jackknife method of variance estimation is discussed in workpackage 5. Its practical application in complex surveys usually involves dropping primary sampling units in turn. This approach has traditionally been dependent upon the sampling design being stratified multi-stage. In workpackage 6 the jackknife method is generalised to handle any complex design with unequal probabilities. In this workpackage the generalized method of workpackage 6 is further extended to handle imputed data.

Variance estimation in presence of imputed data depends upon the form of the imputation. A basic distinction is between single and multiple imputation. Single imputation is the traditional approach in national statistical offices, where each missing value is replaced by just a single imputed value. In contrast, the multiple imputation approach involves creating multiple imputed datasets, where the imputed values may differ between datasets. Pseudo code is developed for both of these cases and for a range of settings in each case.

The standard approach to multiple imputation employs a simple variance estimator, defined in terms of the standard point estimators and variance estimators which may be computed from the complete (imputed) datasets. The validity of this variance estimator depends upon the method of imputation obeying certain conditions. These conditions and the associated inference are most simply expressed in a Bayesian framework. This workpackage also proposes a non-Bayesian approach to multiple imputation which does not require these conditions to hold.

2.12 Workpackage 12: Final Report

The final report of the DACSEIS project is devided in three deliverables including overviews and main achievements of all analytical and computational work done during DACSEIS.

The first deliverable gives an overview of all workpackages and their main achievements. Deliverables D12.2 and D12.3 aim at presenting the DACSEIS simulation study results including the recommendations for the practical use of the variance estimation methodology derived from the simulation study. The deliverable D12.2 is arranged as an electronic platform and contains all major simulation results in the form of a hypertext manual. The hypertext manual covers all simulation tasks including the graphs and measures for the point and variance estimators, special comparisons for estimators and non-response rates, as well as an overview of the methodology and survey specifications.

The recommended practice manual, deliverable D12.3, finally provides recommendations for the use of variance estimation methods in practice as well as further recommendations gained from the DACSEIS research. The first chapter summarizes selected simulation tasks and comparisons for the six surveys examined in DACSEIS. The second chapter presents the main recommendations for the variance estimation methodology. These cover recommendations for the different groups of methods under non-response, direct methods with weighting, methods under single imputation, and multiple imputation. The methods are finally compared in order to help the end-user to find the optimal routines for his own applications.

Finally, the deliverable gives recommendations on the applicability of repeated weighting methods in other surveys than the Dutch LFS, on combining results from different national surveys, e. g. in a European context, as well as on the selection of appropriate software packages for variance estimation.

Chapter 3

Summary and Outlook

The DACSEIS research has supported a wide range of innovations in variance estimation methods. Special emphasis was put on the practical needs and the applicability of the methodology with respect to the needs of the European Statistical System.

A widespread as well as detailed Monte-Carlo study has enabled an empirical evaluation of variance estimation methods in realistic European settings for household and individual surveys. This enabled to investigate the variance estimation methodology under different conditions in order to assess the robustness of the different methods.

From these investigations, recommended practices were evolved in order to facilitate the end-user from NSIs or other statistical institutes applying the methodology appropriately.

DACSEIS has kept closely in touch with NSIs, indeed six of ten DACSEIS partners are NSIs, in order to support and emphasise the practicability of an innovative methodology. In fact, the project has been designed to facilitate the implementation of the recommended variance estimation methods at Eurostat and in the NSIs.

The DACSEIS research has benefitted from cooperation between universities and NSIs in several European countries. This, together with other projects such as EUREDIT and EURAREA, has led to a strengthening of the human capacity and general infrastructure in Europe for methdological research in official statistics. This strenthening could, however, be continued and extended for an effective support of the European Research Area in the field of Statistics.

References

Andersson, C. and Nordberg, L. (1994): A method for variance estimation of non-linear function of totals in surveys - theory and a software implementation. *Journal of Official Statistics* **10**, 395–405.

Berger, Y. G. (2003): A modified Hájek variance estimator for systematic sampling. *Statistics in Transition* **6**, 5–21.

Berger, Y. G. (2004a): *Variance Estimation for Measures of Change in Probability Sampling.* To appear in the Canadian Journal of Statistics in December 2004.

Berger, Y. G. (2004b): A simple variance estimator for unequal probability sampling without replacement. *Journal of Applied Statistics* **31**, 305–315.

Boonstra, H. J. H. and Nieuwenbroek, N. (2003): *An empirical comparison of BRR and linearization variance estimators.* DACSEIS.

Brewer, K. R. W. (2002): *Combined Survey Sampling Inference (Weighing Basu's Elephants).* Arnold publishers.

Deville, J. C. and Särndal, C. E. (1992): Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376–382.

Eurostat (2001a): Working group on assessment of quality in statistics. Eurostat: Doc. Eurostat/A4/Quality/00/General/Definition.

Eurostat (2001b): Working group on assessment of quality in statistics. Eurostat: Doc. Eurostat/A4/Quality/00/General/Standard.

Hájek, J. (1964): Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics* **35**, 1491–1523.

Hansen, M. H. and Hurwitz, W. N. (1943): On the theory of sampling from a finite population. *Annals of Mathematical Statistics* **14**, 333–362.

Horvitz, D. G. and Thompson, D. J. (1952): A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.

Kish, L. (1965): *Survey Sampling.* New York: John Wiley.

Linden, H. and Grünewald, W. (2001): Quality measurement - eurostat experiences. Proceedings of Statistics Canada Symposium 2001.

Münnich, R. and Wiegert, R. (2003): Datenqualität am beispiel des deutschen mikrozensus. *Baden-Württemberg in Wort und Zahl* **51**, 185–190.

Office of National Statistics (2003): Report on the performance of the ßtandardësti-mators. P. Heady and R. Lehtonen. Version1. 0/300503. EURAREA working paper IST-2000-26290.

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

Sen, P. K. (1953): On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics* **5**, 119–127.

Smith, P., Pont, M. and Jones, T. (2003): Developments in business survey methodology in the Office for National Statistics, 1994-2000. *Journal of the Royal Statistical Society, Series D* **52**, 1–30.

Tam, S. M. (1984): On covariances from overlapping samples. *The American Statistician* **38**, 288–289.

Yates, F. and Grundy, P. M. (1953): Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society Serie B* **1**, 253–261.